INFSYS

RESEARCH

REPORT

INSTITUT FÜR INFORMATIONSSYSTEME

ABTEILUNG WISSENSBASIERTE SYSTEME

COMPLEXITY RESULTS FOR STRUCTURE-BASED CAUSALITY

Thomas Eiter and Thomas Lukasiewicz

INFSYS RESEARCH REPORT 1843-01-01

JANUARY 2001; NOVEMBER 2001

Institut für Informationssysteme
Abtg. Wissensbasierte Systeme
Technische Universität Wien
Favoritenstraße 9-11
A-1040 Wien, Austria
Tel: +43-1-58801-18405
Fax: +43-1-58801-18493
sek@kr.tuwien.ac.at
www.kr.tuwien.ac.at

TECHNISCHE UNIVERSITÄT WIEN

INFSYS RESEARCH REPORT
INFSYS RESEARCH REPORT 1843-01-01, JANUARY 2001; NOVEMBER 2001

COMPLEXITY RESULTS FOR STRUCTURE-BASED CAUSALITY

Thomas Eiter [1] and Thomas Lukasiewicz [1]

Abstract. We give a precise picture of the computational complexity of causal relationships in Pearl's structural models, where we focus on causality between variables, event causality, and probabilistic causality. As for causality between variables, we consider the notions of causal irrelevance, cause, cause in a context, direct cause, and indirect cause. As for event causality, we analyze the complexity of the notions of necessary and possible cause, and of the sophisticated notions of weak and actual cause by Halpern and Pearl. In the course of this, we also prove an open conjecture by Halpern and Pearl, and establish other semantic results. We then analyze the complexity of the probabilistic notions of probabilistic causal irrelevance, likely causes of events, and occurrences of events despite other events. Moreover, we consider decision and optimization problems involving counterfactual formulas. To our knowledge, no complexity aspects of causal relationships in the structural-model approach have been considered so far, and our results shed light on this issue.

Keywords: Causal model, probabilistic causal model, causality between variables, event causality, probabilistic causality, weak cause, actual cause, complexity, counting hierarchy.

---

[1] Institut und Ludwig Wittgenstein Labor für Informationssysteme, Technische Universität Wien, Favoritenstraße 9-11, A-1040 Vienna, Austria. E-mail: {eiter, lukasiewicz}@kr.tuwien.ac.at.

Acknowledgements: This work has been partially supported by the Austrian Science Fund under project Z29-INF and by a DFG grant.

A short and preliminary version of this paper appeared in: *Proceedings 7th International Joint Conference on Artificial Intelligence (IJCAI-01)*, pp. 35–40, 2001.

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 nonmonotonic formalisms (especially in the context of logic programming) and those that evolved from (or that are closely related to) the area of Bayesian networks.

A representative of the former is Geffner's modal nonmonotonic logic for handling causal knowledge [10, 11], 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, e.g. [13, 22, 23, 24, 27, 36]; see especially the work by Turner [36] and the references therein for an overview.

A representative of the latter is Pearl's approach to modeling causality by structural equations [1, 9, 30, 31], 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 [1]. Moreover, in a recent work, Halpern [15] gave an axiomatization of reasoning about causal formulas in the structural-model approach, and explored its computational aspects. Furthermore, 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 [9]. In particular, it allows an elegant definition of the important notions of actual causation and causal explanation [16, 17, 18].

Roughly speaking, the main idea behind the structural-model approach is that the world is modeled by random variables, which may causally influence each other. The variables are divided into background variables, which are influenced by factors outside the model, and observable variables, which are influenced by background and observable variables. This latter influence is described by functions for the observable variables. The following is a simple example due to Halpern and Pearl [16, 17, 18], which illustrates the structural-model approach.

Example 1.1 (arsonists) 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 B, which describe the observable situation, where A_i has the value 1 iff the arsonist i drops the match, and B 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 B 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.

$$\begin{tikzcd}
U_1 \arrow[r] & A_1 \arrow[dr] & \\
& & F \\
U_2 \arrow[r] & A_2 \arrow[ur] &
\end{tikzcd}$$

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 [31]), 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 [15] 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 of this paper can be summarized as follows (a review of the complexity classes mentioned is provided in Section 2.4):

- We analyze the complexity of deciding causal relationships between variables in the structural-model approach. We consider the notions of causal irrelevance, cause, cause in a context, direct cause, and indirect cause, which have been described in [9, 30, 31]. It turns out that deciding these notions has a complexity among NP, $\mathrm{co\text{-}NP}$, and D^{P}, and that hardness holds even in restricted cases.

- We analyze the complexity of deciding causal relationships between events. We consider the notions of necessary and possible cause [9, 30, 31]. and the sophisticated notions of weak and actual cause by Halpern and Pearl [16]. It turns out that deciding necessary and possible causes is complete for D^{P} and NP, respectively, while deciding weak and actual causes is Σ_2^P-complete in the general case, and NP-complete in the case of binary variables.

- We prove some semantic results related to the notions of actual and weak cause. In detail, we prove an open conjecture by Halpern and Pearl [16], which says that actual all causes are primitive events. Furthermore, we give a new characterization of weak cause for the case of binary variables.

- We analyze the complexity of probabilistic causal relationships. We consider the notions of probabilistic causal irrelevance, and, in slight generalizations, of likely causes of events, and of occurrences of events despite other events [9, 30, 31]. It turns out that deciding probabilistic causal irrelevance is complete for the class $\mathbf{C}_=$, while deciding the other two notions of probabilistic causality is complete for the class $\mathbf{C}$. Thus, deciding these probabilistic causal relationships is harder than $\mathrm{co\text{-}NP}$. Note that few $\mathbf{C}_=$-complete problems, and none in AI, were known.

- We analyze the complexity of some decision and function computation problems involving counterfactual formulas. We show that deciding whether the probability of a causal formula is at least α and whether the conditional probability over two causal formulas is at least α is complete for $\mathbf{C}$. Furthermore, we show that computing marginal probabilities of causal formulas is complete for $\#$P. We remark that our conditional probabilities over causal formulas are a generalization of Balke and Pearl's counterfactual queries [1, 30, 31].

Our results draw a precise picture of the complexity of structural causality. They give useful insight for implementing decision procedures for causal reasoning in the structural-model approach and for identifying tractable cases of such causal reasoning. Our results show that causal relationships may be exploited in counterfactual reasoning [1, 30, 31]. The results of this paper also proved useful for analyzing the complexity of reasoning about Halpern and Pearl's explanations [16, 18], which are based on the notions of weak and actual cause; see our companion paper [6].

The rest of this paper is organized as follows. Section 2 contains some preliminaries on causal and probabilistic causal models, on their representation for computation, and on the complexity classes that we encounter in this paper. In Section 3, we analyze the complexity of causal relationships between variables. Section 4 concentrates on the complexity of causal relationships between events, including Halpern and Pearl's actual and weak cause. In Section 5, we then analyze the complexity of probabilistic causal relationships. The final Section 6 discusses the presented results and gives an outlook on future research.

In order to increase readability, some proofs and technical details have been moved to Appendices A–C.

2 Preliminaries

In this section, we give some technical preliminaries. We first recall causal and probabilistic causal models. We then discuss some aspects on their representation, and we finally describe the complexity classes that appear in our results.

2.1 Causal models and probabilistic causal models

We star with recalling structure-based causal and probabilistic causal models; for a rich background, see especially [1, 9, 30, 31, 15]. Roughly speaking, the main idea behind structure-based causal models is that the world is modeled by random variables, which may have a causal influence on each other. The variables are divided into exogenous variables, which are influenced by factors outside the model, and endogenous variables, which are influenced by exogenous and endogenous variables. This latter influence is described by structural equations for the endogenous variables.

More formally, we assume a set of *random variables*. Capital letters U, V, W, etc. denote variables and sets of variables. Each variable X may take on *values* from a nonempty finite *domain* $D(X)$. 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 $x(X_i) \in D(X_i)$; for $X = \emptyset$, the unique value is the empty mapping $\emptyset$. The *domain* of X, denoted $D(X)$, is the set of all values for X. Lower case letters x, y, z, etc. denote values for the sets of variables X, Y, Z, etc., respectively. Assignments of values to variables $X = x$ are often abbreviated by the value x. For $Y \subseteq X$ and $x \in D(X)$, we use $x|Y$ to denote the restriction of x to Y. For sets of variables X, Y and values $x \in D(X), y \in D(Y)$, 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 $x(X_i)$. Furthermore, we often identify the values 0 and 1 with the classical truth values $\mathbf{false}$ and $\mathbf{true}$, respectively.

We are now ready to define 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 with $U \cap V = \emptyset$, 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 of X to each value of the *parents* $PA_X \subseteq U \cup V \setminus \{X\}$ of X. Every value $u \in D(U)$ is also called a *context*. The parent relationship between the variables of $M = (U, V, F)$ is expressed by the *causal graph* for M, which is the directed graph that has $U \cup V$ as the set of nodes, and a directed edge from X to Y iff X is a parent of Y, for all variables $X, Y \in U \cup V$.

We focus here on the principal class of *recursive* causal models $M = (U, V, F)$; as argued in [16], we do not lose much generality by concentrating on recursive causal models. A causal model $M = (U, V, F)$ is *recursive*, if and only if its causal graph is a directed acyclic graph (dag). Equivalently, there exists a total ordering $\prec$ on V such that $Y \in PA_X$ implies $Y \prec X$, for all $X, Y \in V$. In recursive causal 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 following, M is reserved for denoting a recursive causal model.

Example 2.1 (arsonists continued) In our introductory example, the causal model $M=(U,V,F)$ is given by $U=\{U_1,U_2\}$, $V=\{A_1,A_2,B\}$, and $F=\{F_{A_1},F_{A_2},F_B\}$, where $F_{A_1}=U_1$, $F_{A_2}=U_2$, and $F_B=1$ iff $A_1=1$ or $A_2=1$. The causal graph for M is shown in Fig. 1. As this graph is acyclic, M is recursive.

In a causal model, we may set endogenous variables X to a value x by an "external action". More formally, for any causal model $M=(U,V,F)$, set of variables $X\subseteq V$, and value $x\in D(X)$, the causal model $M_{X=x}=(U,V,F_{X=x})$, where

$$F_{X=x} \quad = \quad \{F_Y \mid Y\in V\setminus X\}\cup\{F_{X'}=x(X') \mid X'\in X\}\,,$$

is a *submodel* of M. We use M_x and F_x to abbreviate $M_{X=x}$ and $F_{X=x}$, respectively, if X is understood from the context. Similarly, for $Y\subseteq V$, we write $Y_x(u)$ to abbreviate $Y_{M_x}(u)$.

We next add probabilistic uncertainty to causal models, where contexts serve as possible worlds. That is, we add a probability distribution on the set of all contexts of a causal model. More formally, a *probabilistic causal model* (M,P) consists of a causal model $M=(U,V,F)$ and a probability function P on $D(U)$.

Example 2.2 (arsonists continued) In our running example, a probabilistic causal model (M,P) may be given by the uniform distribution P over $D(U)$. Thus, $P(u)=0.25$ for each context $u\in D(U)$.

2.2 Model representation for computation

For computational purposes, we need a suitable representation of causal models. Different such representations are possible, and as long as they are polynomial-time intertranslatable, the representation details do not matter. Our aim here is to choose a representation which, on the one hand, admits an expressive class of causal models, and on the other hand, does not bear intractability for simple problems which we expect to be polynomial.

For this reason, we assume in this paper the following representation of causal models $M=(U,V,F)$ and probabilistic causal models (M,P), respectively:

1. Each function $F_X\colon D(PA_X)\to D(X)$, $X\in V$, is computable in polynomial time.

2. 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.

3. 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 $u\in D(U)$.

Assumption (1) leaves the precise representation of F_X open; it could be the code of a procedure which, on input of an arbitrary value v for PA_X, computes the output $F_X(v)$ in polynomial time, or a Boolean circuit which computes, from a binary encoding of v, the output value. These two representations are in a sense expressive for P, since they allow for representing any polynomial-time computable function F_X using polynomial-time effort; in practice, however, we might often have less expressive representations. For example, F_X might be encoded as a Boolean formula ϕ over equality atoms $val(Y)=y$, where $val(Y)$ is a function ranging over the variable domain $D(Y)$ for Y from $PA_X\cup\{X\}$ and $y\in D(Y)$ is a value for Y, such that the output value is given by $val(X)$ in a model for ϕ where for each $Y\in PA_X$, $val(Y)$ takes the value of Y as in v. Another possible representation, which is frequently used in the context of Bayesian networks, is by a table that lists the output values $F_X(v)$ for all $v\in D(PA_X)$; in fact, this amounts to a restricted form of the previous formula-based representation. Intuitively, a more restrictive form of representation requires a larger encoding of the same function. In particular, tables can be exponentially larger than an equivalent formula or piece of code, and thus increase the input size. However, as discussed in Section 2.3 below, there is no difference in our complexity results regardless of whether we use tables, formulas, or procedure code for representing the functions.

The assumption (2) takes into account that in general, we may have natural names of the values of a domain, such as e.g., *red*, *green*, *blue* for the respective colors, or *Monday*, *Tuesday*, $\ldots$, *Sunday* for the days of the week. It is, of course, possible to map the values $v_1,\ldots,v_k$ of $D(X)$ to a particular interval of integers, e.g. $[0,\ldots,k-1]$, such that $D(X)$ is fully described by the size k of the domain; this encoding will, for large domains, result in an exponential saving with respect to the space which is needed for enumerating the domain. The reason why we do not assume this domain representation is that, besides the use of enumerative specifications in general, some problems on causal models that are tractable under enumerative representation might become hard under "implicit" domain representation. For example, simply telling whether there exists some value v such that $F_X(v)$ takes on a particular value from $D(X)$ is NP-complete, if we assume that F_X is represented by procedural code or, equivalently, by a Boolean circuit, as follows from well-known results (cf. [29]). However, for all the problems that we consider, except for probabilistic likely causes, the form of the domain representation (enumerative or implicit) does not affect our complexity results; in particular, hard instances show already up if the domains contain few (in fact, at most three) values.

Assumption (3) may look restrictive, and therefore needs some explanation. The parameter b there helps in a normalized representation of probabilities through integers, but is not essential and may be eliminated in a standard computational environment. Namely, the probability function $P(u)$ is, by assumption, computable in time $|u|^k$ for some fixed $k \geq 0$, and thus $P(u)$ occupies at most $|u|^k$ bits. By scaling with a suitable integer b, we can turn each $P(u)$, represented in a standard number format, into an integer $f(u) = P(u) * b$ in polynomial time; if we disregard exponents in a floating point representation, then $b = 2^{|u|^k}$ would be suitable. Thus, we can transform, without loss of generality, any polynomial-time computable probability function $P(u)$ in polynomial time into an equivalent pair (f, b). The latter, however, relieves us from low level details of representation.

The following proposition is immediate.

Proposition 2.3 *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.*

2.3 Restricted classes of models

We pay particular attention to the classes of causal and probabilistic causal models in which the variable domains and the input degree of the functions F_X are subject to bounds. We say that a causal model M (resp., probabilistic causal model (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, the different representations that we have discussed in Section 2.2 are equivalent under polynomial-time computations. That is, we can replace procedural code for computing F_X in polynomial time by an equivalent formula or table that lists all function values, and vice versa. (Observe that an efficient transformation into a table would not be possible if domains are implicitly represented.) As we shall see, the hardness parts of our complexity results are established for bounded models. Thus, the precise form of function representation does not matter for our complexity results.

2.4 Complexity classes

We assume that the reader has some familiarity with the basic concept and notions of complexity theory, such as P, NP, complete problems and polynomial time transformations; for a background, we refer to [29, 20]. The main decisional complexity classes that we encounter in the rest of the paper are shown in Fig. 2, where arrows denote containment. The classes P, NP, co-NP, Σ_2^P, and Π_2^P are from the Polynomial Hierarchy (PH), which is contained in PSPACE. The class $\mathrm{D^P} = \{L \cap L' \mid L \in \mathrm{NP}, L' \in \text{co-NP}\}$ is the "conjunction" of NP and co-NP. The class **C** is from the Counting Hierarchy (CH) of complexity classes [35]. Informally, **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, where computing $f(|I|)$ and checking each proof J_i can be done in polynomial time. The class **C** is known to coincide with the famous class PP (probabilistic P) [12], which contains the problems decidable by a polynomial-time Turing machine that accepts an input iff the majority of its runs halt in an accepting state.

$$\begin{tikzcd}
& \text{co-NP} \arrow[r] \arrow[dr] & \mathbf{C}_= \arrow[dr] & \Sigma_2^P = \mathrm{NP}^{\mathrm{NP}} \\
\mathrm{P} \arrow[ur] \arrow[dr] & & \mathrm{D^P} \arrow[ur] \arrow[r] \arrow[dr] & \mathbf{C} = \mathrm{PP} \\
& \mathrm{NP} \arrow[r] \arrow[ur] & \text{co-}\mathbf{C}_= \arrow[ur] & \Pi_2^P = \text{co-NP}^{\mathrm{NP}}
\end{tikzcd}$$

Figure 2: Containment between complexity classes

The class $\mathbf{C}_=$ is a variant of **C**, where "exactly $f(|I|)$" replaces "at least $f(|I|)$". While this difference may seem to be marginal, **C** and $\mathbf{C}_=$ have quite different properties [35]. Intuitively, $\mathbf{C}_=$ is an extension of co-NP, and has many properties of this class. Both **C** and $\mathbf{C}_=$ are contained in PSPACE, and it is widely believed that these classes are not contained in PH. To our knowledge, no problems in AI, or any other applied field of computer science, which are complete for $\mathbf{C}_=$ have been reported so far. For more details on $\mathbf{C}_=$ and **C**, we refer to [35, 37]. The functional analog of these classes is the well-known class #P, which is the class of all functions f (from strings to the nonnegative integers) for which there exists a nondeterministic polynomial-time Turing machine T such that for every input string w, it holds that $f(w)$ is the number of accepting runs of T on w; for more details on #P, we refer to [20, 29].

We remark that all classes C in Fig. 2 are closed under polynomial-time reductions, i.e., if a problem Π has a polynomial time transformation into a problem Π' from C, then also Π belongs to C. Furthermore, each C has complete problems under polynomial-time transformations, including canonical variants of the satisfiability problem (SAT), i.e., deciding satisfiability of a Boolean formula ϕ. The latter is well-known NP-complete, while its complement, deciding unsatisfiability of ϕ (equivalently, whether the negation of ϕ is a tautology) is co-NP-complete; deciding, given to Boolean formulas ϕ_1 and ϕ_2, whether ϕ_1 is satisfiable and ϕ_2 is unsatisfiable is complete for $\mathrm{D^P}$. The classes Σ_2^P and Π_2^P have complete problems in terms of deciding the truth of quantified Boolean formulas (QBFs) $\exists A \forall B\, \phi$ and $\forall A \exists B\, \phi$, respectively. The classes **C** and $\mathbf{C}_=$ have complete problems in terms of counting versions of SAT. Namely, deciding whether a Boolean formula ϕ has at least m satisfying assignments, where m is part of the input, is complete for **C**, and similarly deciding whether ϕ has exactly m satisfying assignments is complete for $\mathbf{C}_=$ (cf. [37, 29]). It is not hard show from this that the problems remain hard even if we set $m = 2^{n-1}$, where n is the number of variables (and thus m may be dropped from the input). We refer to these variants as GE-HALFSAT and

HALFSAT, respectively.[1]

For all problems in the previous paragraph, hardness for the respective classes holds if in addition ϕ is in conjunctive normal form (CNF), except for tautology checking of a Boolean formula ϕ and for deciding $\forall A \exists B \phi$, where ϕ is in disjunctive normal form (DNF). In our proofs, we often make use of this restriction.

3 Causality between Variables

In this section, we analyze the complexity of deciding causal relationships between variables due to Galles and Pearl [9]; see also [30, 31]. We consider the notions of causal irrelevance, cause, cause in a context, direct cause, and indirect cause.

3.1 Definitions

We now recall the notions of causal irrelevance, cause, cause in a context, direct cause, and indirect cause from [9]. Let $M = (U, V, F)$ be a causal model, and let $X, Y, Z \subseteq V$ be sets of endogenous variables such that $X, Y \neq \emptyset$. Then,

- *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 $w \in D(W)$, it holds $Y_{xzw}(u) = Y_{x'zw}(u)$.
- *X is a cause of Y* iff $x, x' \in D(X)$ and $u \in D(U)$ exist such that $Y_x(u) \neq Y_{x'}(u)$.
- *X is a cause of Y in the context $Z = z$*, where $z \in D(Z)$, iff there exist values $x, x' \in D(X)$ and $u \in D(U)$ such that $Y_{xz}(u) \neq Y_{x'z}(u)$.
- *X is a direct cause* of Y iff there exist values $x, x' \in D(X)$, $u \in D(U)$, and $z \in D(V \setminus X \cup Y)$ such that $Y_{xz}(u) \neq Y_{x'z}(u)$.
- *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 to illustrate these causal relationships.

Example 3.1 (arsonists continued) In our running example, A_1 is not causally irrelevant to B. For instance, if we set $A_2 \in V \setminus \{A_1, B\}$ to 0, and A_1 to 0 and 1, then B 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 B, but not a cause of B 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 B.

3.2 Results

Our results on the complexity of deciding the above notions of causality are summarized in Table 1. In detail, deciding causal irrelevance is co-NP-complete, while deciding its semantically complementary notions of cause, cause in a context, and direct cause is NP-complete. Moreover, deciding indirect cause, which is the logical conjunction of cause and the complement of direct cause, is D^{P}-complete. 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.

[1] A popular **C**-complete variant of SAT is MAJSAT, i.e., deciding whether *more than* half of the assignments, i.e., at least $2^{n-1} + 1$ many satisfy ϕ [29]. MAJSAT is easily reduced to GE-HALFSAT, which we use for uniformity here.

Table 1: Complexity of Causality between Variables

Problem	Complexity
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	$\mathrm{D^P}$-complete

The following result shows that deciding causal irrelevance is co-NP-complete. Here, we have membership in co-NP, as the complement of causal irrelevance can be decided by guessing and checking in polynomial time. Hardness for co-NP is shown by a reduction from the co-NP-complete problem of deciding whether a given propositional formula in 3DNF is a tautology.

Theorem 3.2 *Given a causal model $M = (U, V, F)$ 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. The problem is in co-NP, as a set of variables $W \subseteq V \setminus X \cup Y \cup Z$ 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, by Proposition 2.3.

We prove 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 $k, n \geq 1$, is a tautology.

(a) We first construct $M = (U, V, F)$ and $X, Y, Z \subseteq V$ such that (1)–(3) and (4a) are satisfied, and that X is causally irrelevant to Y given Z iff ϕ is a tautology.

We define the causal model $M = (U, V, F)$ as follows. The exogenous and endogenous variables are defined by $U = \{A_1, \ldots, A_n\}$ and $V = \{A, D_1, \ldots, D_k\}$, respectively, where $D(S) = \{0, 1\}$ for all variables $S \in U \cup V$. The functions $F = \{F_S \mid S \in V\}$ are defined as follows:

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

The corresponding causal graph is shown in Fig. 3. We define $X, Y, Z \subseteq V$ by $X = \{A\}$, $Y = \{D_k\}$, and $Z = \emptyset$. The values 0 and 1 of X are denoted by x_0 and x_1, respectively. Observe that (1)–(3) and (4a) are satisfied.

It can now be shown that X is causally irrelevant to Y given Z iff ϕ is a tautology (see Appendix A). Informally, any assignment w to some nonempty $W \subseteq \{D_1, \ldots, D_{k-1}\}$ always yields the same value of Y in M_{xzw} and $M_{x'zw}$. Hence, we can assume $W = \emptyset$. Under any $u \in D(U)$, if X is set to x_1, then Y becomes

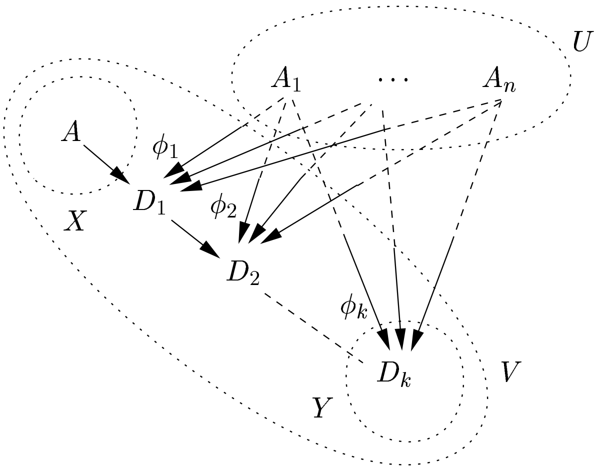

Figure 3: Causal Graph for co-NP-Hardness of Causal Irrelevance for (1)–(3) and (4a)

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.

(b) We next construct $M = (U, V, F)$ and $X, Y, Z \subseteq V$ such that (1)–(3) and (4b) are satisfied, and that X is causally irrelevant to Y given Z iff ϕ is a tautology.

We define the causal model $M = (U, V, F)$ as follows. The exogenous and endogenous variables are defined by $U = \{A_1, \ldots, A_n\}$ and $V = \{A, D_1, \ldots, D_k, B\}$, respectively, where $D(S) = \{0, 1\}$ for all variables $S \in U \cup V$. The functions $F = \{F_S \mid S \in V\}$ are defined as follows:

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

The corresponding causal graph is shown in Fig. 4. We define $X, Y, Z \subseteq V$ by $X = \{A\}$, $Y = \{D_1, \ldots, D_k, B\}$, and $Z = \emptyset$. We write $Y^1, \ldots, Y^{k+1}$ to denote $D_1, \ldots, D_k, B$. The values 0 and 1 of X are denoted by x_0 and x_1, respectively. Observe that (1)–(3) and (4b) are satisfied. It can now be shown that X is causally irrelevant to Y given Z iff ϕ is a tautology (see Appendix A). $\Box$

The following result shows that deciding the notions of cause and cause in a context is NP-complete. Here, it is easy to see that these problems can be solved by guessing and checking in polynomial time. Hardness for NP is shown by a reduction from the NP-complete problem of deciding whether a propositional formula in 3DNF is not a tautology, using a construction similar to the one illustrated in Fig. 4.

Theorem 3.3 *(a) Given a causal model $M = (U, V, F)$ and $X, Y \subseteq V$ such that $X, Y \neq \emptyset$, deciding whether X is a cause of Y is* NP*-complete. Hardness holds even if (1) M is binary and bounded, and (2) X, Y are singletons.*

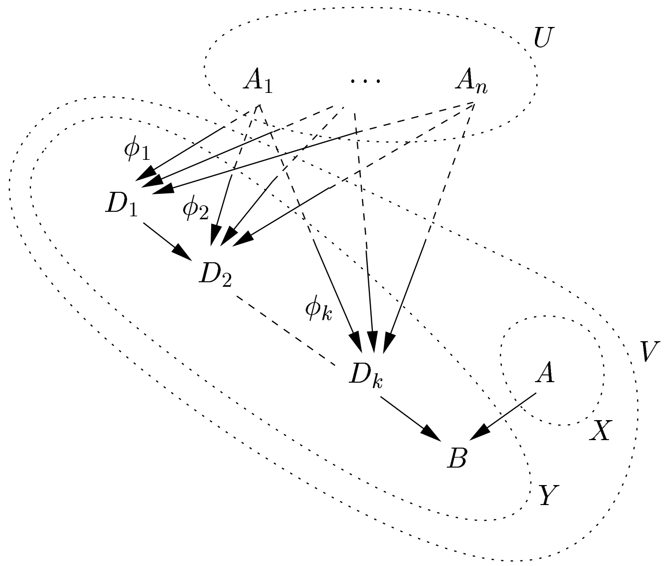

Figure 4: Causal Graph for co-NP-Hardness of Causal Irrelevance for (1)–(3) and (4b)

(b) Given a causal model $M = (U, V, F)$, $X, Y, Z \subseteq V$ such that $X, Y \neq \emptyset$, and $z \in D(Z)$, deciding whether 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, Y are singletons.*

The next theorem shows that deciding direct and indirect cause is NP- and D^{P}-complete, respectively. Here, the NP-completeness result is shown similarly as in the proof of Theorem 3.3. The D^{P}-membership result follows easily from (a) and Theorem 3.3 (a), and the D^{P}-hardness result is shown by a reduction from the following D^{P}-complete problem. Given two propositional formulas in 3DNF α and β, decide whether α is a tautology and β is not a tautology. Roughly speaking, the construction is a combination of the two constructions shown in Figs. 3 and 4.

Theorem 3.4 *(a) Given a causal model $M = (U, V, F)$ and $X, Y \subseteq V$ such that $X, Y \neq \emptyset$, deciding whether X is a direct cause of Y is* NP*-complete. Hardness holds even if (1) M is binary and bounded, (2) X is a singleton, and (3) $V = X \cup Y$.*
(b) Given a causal model $M = (U, V, F)$ and $X, Y \subseteq V$ such that $X, Y \neq \emptyset$, deciding whether X is an indirect cause of Y is D^{P}*-complete. Hardness holds even if (1) M is binary and bounded, and (2) X is a singleton.*

4 Event Causality

In this section, we analyze the complexity of deciding causal relationships between events. In particular, we consider the notions of necessary and possible cause due to Galles and Pearl [9], see also [30, 31], and the notions of weak and actual cause by Halpern and Pearl [16, 17, 18], which are inspired by Pearl's causal beams [31].

4.1 Definitions

We now recall the notions of necessary and possible cause from [9] and the notions of weak and actual cause from [16, 17, 18]. We first define events and the truth of events in a causal model $M = (U, V, F)$ under a context $u \in D(U)$.

A *primitive event* is an expression of the form $Y = y$, where Y is an endogenous variable and y is a value for Y. The set of *events* is the closure of the set of primitive events under the Boolean operators $\neg$ 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 a context $u \in D(U)$, denoted $(M, u) \models \phi$, is inductively defined as follows:

- $(M, u) \models Y = y$ 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 $(M, u) \models \psi$.

Further operators $\rightarrow$ and $\vee$ are defined as usual, i.e., $\phi \rightarrow \psi$ and $\phi \vee \psi$ stand for $\neg\phi \vee \psi$ and $\neg(\neg\phi \wedge \neg\psi)$, respectively. We write $\phi(u)$ to abbreviate $(M, u) \models \phi$. For $X \subseteq V$ and $x \in D(X)$, we use $\phi_x(u)$ 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 $X_1 = x_1 \wedge \cdots \wedge X_k = x_k$.

The following result follows immediately from Proposition 2.3.

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

We are now ready to define the notions of necessary and possible cause (which are slightly more general than in [9]). Let $M = (U, V, F)$ be a causal model, and let $X \subseteq V$ and $x \in D(X)$. Let ϕ be an event. Then,

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

We next recall the notions of weak and actual cause from [16, 17, 18]. 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 endogenous 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)** $\phi_{xw\hat{z}}(u)$ 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 minimality condition is satisfied:

AC3. X is minimal with AC1 and AC2. That is, for every $X' \subset X$, it holds that $X' = x|X'$ is not a weak cause of ϕ under u.

The following example illustrates these causal relationships.

Example 4.2 (arsonists continued) In our running example, $A_1 = 1$, $A_2 = 1$, and $A_1 = 1 \wedge A_2 = 1$ always cause $B = 1$. For instance, let us show that $A_1 = 1$ always causes $B = 1$: (i) if A_1 is set to 1, then B 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 B 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 context $u = (1, 1)$ in which both arsonists intend to start a fire. Then, $A_1 = 1$, $A_2 = 1$, and $A_1 = 1 \wedge A_2 = 1$ are weak causes of $B = 1$. For instance, let us show that $A_1 = 1$ is a weak cause of $B = 1$: (AC1) both A_1 and B is 1 under u, (AC2 (a)) if both A_1 and A_2 are set to 0, then B has the value 0, and (AC2 (b)) if A_1 is set to 1 and A_2 to 0, then B is 1. In fact, $A_1 = 1$ and $A_2 = 1$ are actual causes of $B = 1$, while $A_1 = 1 \wedge A_2 = 1$ is not an actual cause of $B = 1$.

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, where we assume a syntactic restriction to binary causal models. In detail, in both the general and the binary case, deciding necessary cause is $\mathrm{D^P}$-complete, while deciding possible cause is only NP-complete. Roughly, necessary cause is the conjunction of a universal quantification and an existential one, while possible cause involves only an existential quantification. Furthermore, deciding weak and actual cause is Σ_2^P-complete in the general case and NP-complete in the binary case. Roughly, weak and actual causes involve an existential quantification followed by a universal one, where the latter can be removed in the binary case. Moreover, actual causes are shown to be always primitive, and thus can be identified in constant time among the weak causes. 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	$\mathrm{D^P}$-complete	$\mathrm{D^P}$-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 result shows that deciding necessary and possible cause is complete for $\mathrm{D^P}$ and NP, respectively. Here, membership is easily proved using Propositions 2.3 and 4.1. The NP-hardness result is proved as in the proof of Theorem 3.3, while the $\mathrm{D^P}$-hardness result is shown by a reduction from deciding, given two propositional formulas in 3DNF α and β, whether α is a tautology and β is not a tautology. Roughly, the construction suitably combines two instances of the construction in Fig. 3.

Theorem 4.3 *(a) Given a causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, deciding whether $X = x$ always causes ϕ is $\mathrm{D^P}$-complete. Hardness holds even if M is binary and bounded, and X is a singleton.*

(b) Given a causal model $M=(U,V,F)$, $X\subseteq V$, $x\in D(X)$, $u\in D(U)$, and an event ψ, deciding whether $X=x$ may have caused ψ is NP*-complete. Hardness holds even if M is binary and bounded, X is a singleton, and ψ is primitive.*

The next theorem shows that deciding weak cause is Σ_2^P-complete in the general case. We sketch the main ideas behind its proof, which is technically quite involved.

Theorem 4.4 *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, $|D(S)|\leq 3$ for all $S\in U\cup V$, and either M is bounded or ϕ is primitive.*

Proof. 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 Propositions 2.3 and 4.1, in AC1, deciding whether $X(u)=x$ and $\phi(u)$ hold is polynomial. Moreover, in AC2, some W, $\overline{x}$, and w as required can be guessed and verified in polynomial time with an NP-oracle (needed for (b)). In summary, checking AC1 and AC2 is in Σ_2^P.

Hardness for Σ_2^P is shown by a polynomial transformation from the following standard Σ_2^P-complete problem [29]. Given 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 $B=\{B_1,\ldots,B_n\}$, decide whether Φ is true.

We now define $M=(U,V,F)$, $X\subseteq V$, $x\in D(X)$, $u\in D(U)$, and ϕ as stated such that $X=x$ is a weak cause of ϕ under u iff Φ is true.

We define the causal model $M=(U,V,F)$ as follows. The exogenous and endogenous variables are $U=\{E\}$ and $V=A\cup B\cup\{C,G,H\}$, respectively, where $D(S)=\{0,1,2\}$ 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 $F=\{F_S\mid S\in V\}$ are then defined as follows:

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

We define $X=\{G\}$ and $Y=\{H\}$. Let $u\in D(U)$ and $x=0$, and 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 (see Appendix B). Roughly speaking, the existential and universal quantification over A and B in Φ is expressed by the existential quantification over $W\subseteq V\setminus X$ and $w\in D(W)$ in AC2 and the universal quantification over $\hat{Z}\subseteq V\setminus(X\cup W)$ in AC2 (b), respectively. We then especially have to ensure that (i) $W\cap B=\emptyset$ 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).

Observe that X is a singleton, $|D(S)|\leq 3$ for all $S\in U\cup V$, and ϕ is primitive. To show Σ_2^P-hardness for the case that X is a singleton, $|D(S)|\leq 3$ for all $S\in U\cup V$, and M is bounded, define $M'=(U,V\setminus\{H\},F\setminus\{F_H\})$. Then, $X=x$ is a weak cause of ϕ' under u in M' iff Φ is true. $\Box$

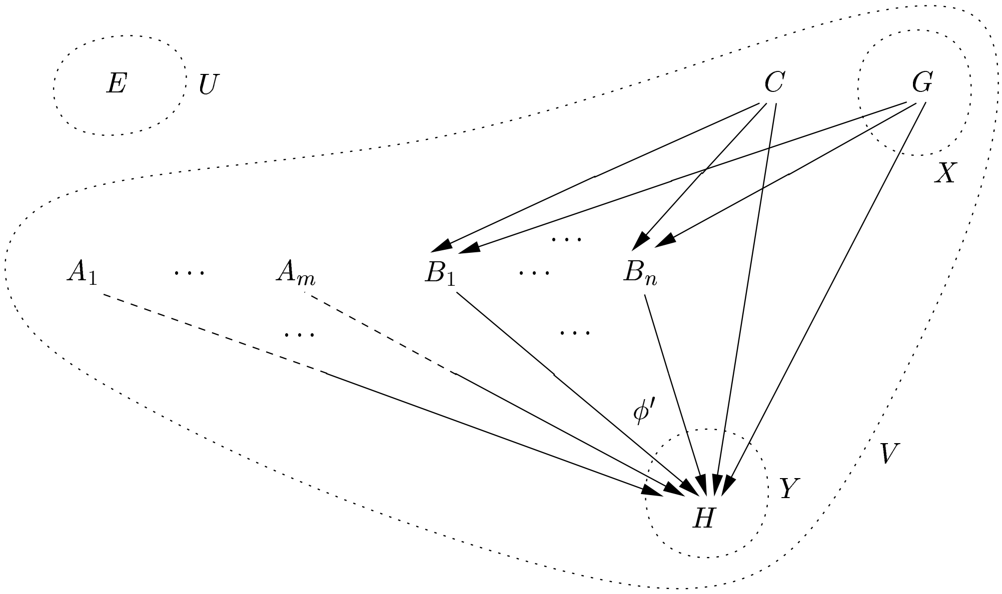

Figure 5: Causal Graph for Σ_2^P-Hardness of Weak Cause

The just sketched proof of Σ_2^P-hardness makes use of *non-binary* causal models. Thus, we may ask whether deciding weak cause in the binary case has a lower complexity. Indeed, the following semantic result shows that in the binary case, AC2 can be expressed in a different way, and the new characterization then implies that deciding weak cause is in NP in the binary case.

Theorem 4.5 *Let $M = (U, V, F)$ be binary, let $X \subseteq V$ and $x \in D(X)$, and let ϕ be an event. Then, $X = x$ is a weak cause of ϕ under $u \in D(U)$ iff AC1 and the following condition AC2′ hold:*

AC2′. *Some set of endogenous 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)** $\phi_{xw}(u)$,
- **(c)** $Z_{xw}(u) = Z(u)$ *for* $Z = V \setminus (X \cup W)$.

Proof. 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) (see Appendix B). More precisely, any singleton $S \in V \setminus (X \cup W)$ 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 cause in the binary case is NP-complete. More formally, this is expressed by the following theorem.

Theorem 4.6 *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 cause. We first prove the following *Splitting Theorem*, which says that if $X=x$ is a weak cause of an event ϕ under a context u, then for any splitting of $X=x$ into two events $X_0=x_0$ and $X_1=x_1$, at least one among $X_0=x_0$ and $X_1=x_1$ is another weak cause of ϕ under u.

Theorem 4.7 (Splitting Theorem) *Let $M=(U,V,F)$ be a causal model, let $X \subseteq V$ and $x \in D(X)$, and let ϕ be an event. Let $\{X_0, X_1\}$ be a partition of X (that is, $X_0, X_1 \neq \emptyset$, $X_0 \cup X_1 = X$, and $X_0 \cap X_1 = \emptyset$), and let $x_0 = x|X_0$ and $x_1 = x|X_1$. If $X=x$ is a weak cause of ϕ under u, then either (1) $X_0=x_0$ is a weak cause of ϕ under u, or (2) $X_1=x_1$ is a weak cause of ϕ under u.*

Proof. Let $X=x$ be a weak cause of ϕ under u. That is, AC1–AC2 hold. In particular, in 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) $\neg\phi_{\overline{x}w}(u)$, and (b) $\phi_{xw\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. Let $\overline{x}_0 = \overline{x}|X_0$ and $\overline{x}_1 = \overline{x}|X_1$. We now consider two cases:

(1) Assume that $\phi_{x_0\overline{x}_1w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. Informally, we can then move the variables in X_1 to the W-part of AC2 assigning them $\overline{x}_1$. That is, we have (a) $\neg\phi_{\overline{x}'w'}(u)$, and (b) $\phi_{x'w'\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X' \cup W')$ and $\hat{z} = \hat{Z}(u)$, where $X' = X_0$, $W' = X_1 \cup W$, $x' = x_0$, $\overline{x}' = \overline{x}_0$, and $w' = \overline{x}_1 w$. That is, $X'=x'$ is a weak cause of ϕ under u. That is, $X_0=x_0$ is a weak cause of ϕ under u.

(2) Assume that $\neg\phi_{x_0\overline{x}_1w\hat{z}}(u)$ for some $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. Informally, we can then move the variables in X_0 and $\hat{Z}$ to the W-part of AC2 assigning them x_0 and $\hat{z}$, respectively. That is, we have (a) $\neg\phi_{\overline{x}'w'}(u)$, and (b) $\phi_{x'w'\hat{z}'}(u)$ for all $\hat{Z}' \subseteq V \setminus (X' \cup W')$ and $\hat{z}' = \hat{Z}'(u)$, where $X' = X_1$, $W' = X_0 \cup W \cup \hat{Z}$, $x' = x_1$, $\overline{x}' = \overline{x}_1$, and $w' = x_0 w \hat{z}$ (note that each instance of AC2 (b) for $X'=x'$ is an instance of AC2 (b) for $X=x$). That is, $X'=x'$ is a weak cause of ϕ under u. That is, $X_1=x_1$ is a weak cause of ϕ under u. □

Example 4.8 (arsonists continued) In our running example, only from the fact that $A_1=1 \wedge A_2=1$ is a weak cause of $B=1$, we know by the Splitting Theorem that either $A_1=1$ is a weak cause of $B=1$, or that $A_2=1$ is a weak cause of $B=1$.

As a corollary, it follows that every actual cause is primitive, as otherwise it would contain a smaller weak cause and thus violate the minimality condition AC3. This proves an open conjecture by Halpern and Pearl [16], which has been proved independently by Hopkins [19], with a more involved argumentation. Note that the following corollary also holds for the setting of possibly infinite domains and sets of endogenous variables in [16].

Corollary 4.9 (conjecture by Halpern and Pearl [16]) *Let $M=(U,V,F)$ be a causal model, let $X \subseteq V$ and $x \in D(X)$, and let ϕ be an event. If $X=x$ is an actual cause of ϕ under u, then X is a singleton.*

This shows that $X=x$ 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 a corollary of Theorem 4.4 and Corollary 4.9, 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 4.10 *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, $|D(S)| \leq 3$ for all $S \in U \cup V$, and M is bounded or ϕ is primitive.*

Moreover, as a corollary of Theorem 4.6 and Corollary 4.9, deciding whether $X = x$ is an actual cause of ϕ under u is NP-complete in the binary case.

Corollary 4.11 *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

In this section, we analyze the computational complexity of probabilistic causal relationships. In detail, we consider the problem of deciding probabilistic causal irrelevance, and some decision and function computation problems involving counterfactual formulas. These problems are due to Galles and Pearl [9]; see also [30, 31].

5.1 Definitions

We now recall causal formulas from [15], which generalize the counterfactual formulas in [9, 30, 31], and we define their probability in probabilistic causal models. We then recall the notions of probabilistic causal irrelevance, of likely causes of events, and of occurrences of events despite other events from [9].

A *basic causal formula* is an expression $[X_1 \leftarrow x_1, \ldots, X_k \leftarrow x_k]\, \phi$, where ϕ is an event, $X_1, \ldots, X_k$ are pairwise distinct endogenous variables, $x_i \in D(X_i)$ for all $i \in \{1, \ldots, k\}$, and $k \geq 0$. The set of *causal formulas* is the closure of the set of basic causal formulas under the Boolean operators $\neg$ and $\wedge$. For $X = \{X_1, \ldots, X_k\}$ and $x = x_1 \ldots x_k$, we use $[X \leftarrow x]\, \phi$ to abbreviate $[X_1 \leftarrow x_1, \ldots, X_k \leftarrow x_k]\, \phi$. We then use $Y_x = y$ to abbreviate $[X \leftarrow x]\, Y = y$. Note that such formulas $Y_x = y$ are called *counterfactual formulas* in [9, 30, 31].

The *truth* of a causal formula ψ in a causal model $M = (U, V, F)$ under $u \in D(U)$, denoted $(M, u) \models \psi$, is inductively defined as follows:

- $(M, u) \models [X \leftarrow x]\, \phi$ iff $\phi_x(u)$ in M,
- $(M, u) \models \neg\phi$ iff $(M, u) \models \phi$ does not hold,
- $(M, u) \models \phi \wedge \psi$ iff $(M, u) \models \phi$ and $(M, u) \models \psi$.

Further connectives $\rightarrow$ and $\vee$ for causal formulas are defined as usual, i.e., $\phi \rightarrow \psi$ and $\phi \vee \psi$ stand for $\neg\phi \vee \psi$ and $\neg(\neg\phi \wedge \neg\psi)$, respectively. The following result is immediate by Proposition 4.1.

Proposition 5.1 *Given a context $u \in D(U)$ and a causal formula ψ, deciding whether $(M, u) \models \psi$ holds can be done in polynomial time.*

The *probability* of a causal formula ψ in a probabilistic causal model (M, P), where $M = (U, V, F)$, denoted $P(\psi)$, is defined as follows:

$$P(\psi) = \sum_{u \in D(U),\, (M,u) \models \psi} P(u)\,.$$

For causal formulas ϕ, ψ with $P(\phi) > 0$, the *conditional probability of* ψ *given* ϕ, denoted $P(\psi \mid \phi)$, is defined as $P(\psi \wedge \phi) \,/\, P(\phi)$. Thus, if ϕ is true in all contexts, then $P(\psi \mid \phi) = P(\psi)$.

We remark that our probabilities of causal formulas generalize the probabilities of conjunctions of counterfactual formulas $Y_x = y$ and events $Y = y$ in [9, 30, 31]. Furthermore, our conditional probabilities over causal formulas are a generalization of the probabilities of counterfactual queries as described in [1, 30, 31].

Example 5.2 (arsonists continued) Let (M, P) be the probabilistic causal model of our running example. Then, $P([A_1 = 0]\, B = 1 \mid A_1 = 1 \wedge A_2 = 0)$ represents the probability that the forest would have burned down, if arsonist 1 did not start a fire, given that in the real world arsonist 1 started a fire and arsonist 2 did not.

We are now ready to define the concepts of probabilistic causal irrelevance, of likely causes of events, and of occurrences of events despite other events.

- For $X, Y, Z \subseteq V$ with $X, Y \neq \emptyset$, we say X *is probabilistically causally irrelevant to* Y *given* Z, denoted $(X \not\rightarrow Y \mid Z)_P$, iff $P(Y_{xz} = y) = P(Y_{x'z} = y)$ for all $x, x' \in D(X)$, $y \in D(Y)$, and $z \in D(Z)$. Intuitively, once the value of Z is fixed at z, changing X between any two values will not change the probability of Y.
- For $X \subseteq V$, $x \in D(X)$, an event ϕ, and $\alpha \in [0, 1]$, we say $X = x$ *is a likely cause of* ϕ *with goodness* α, iff (i) ϕ is observed (which implies that $\phi(u)$ for some $u \in D(U)$), and (ii) $P([X \leftarrow x]\, \phi \wedge [X \leftarrow x']\, \neg\phi \mid \phi) \geq \alpha$ for some $x' \in D(X) \setminus \{x\}$.
- For $X \subseteq V$, $x \in D(X)$, an event ϕ, and $\alpha \in [0, 1]$, we say ϕ *occurred despite* $X = x$ *with goodness* α, iff (i) ϕ and $X = x$ are observed (which implies that $\phi(u)$ and $X(u) = x$ for some $u \in D(U)$), and (ii) $P([X \leftarrow x]\, \phi) \leq 1 - \alpha$.

Note that the last two definitions are slightly more general than in [9, 30, 31], as ϕ can be any event here and is not restricted to only primitive events.

Example 5.3 (arsonists continued) Let $M = (U, V, F)$ be the causal model defined in Example 2.1, and let P be the uniform distribution over $D(U)$. Denote by a_1 (resp., $\overline{a}_1$) the value 1 (resp., 0) of A_1. Then, A_1 is not probabilistically causally irrelevant to B, as $P(B_{a_1} = 1) = 1 \neq 0.5 = P(B_{\overline{a}_1} = 1)$. Informally, the actions of arsonist 1 are not probabilistically causally irrelevant to the state of the forest.

Consider now the probability function P on $D(U)$ that is given by $P(u_{1,1}) = 0.1$, $P(u_{1,0}) = 0.7$, $P(u_{0,1}) = 0.1$, and $P(u_{0,0}) = 0.1$, where the contexts $u_{i,j} \in D(U)$ are defined by $u_{i,j}(U_1) = i$ and $u_{i,j}(U_2) = j$ for all $i, j \in \{0, 1\}$. Assume that the event $B = 1$ is observed. Then, $A_1 = 1$ is a likely cause of $B = 1$ with goodness $7/9$, as $P(B_{a_1} = 1 \wedge B_{\overline{a}_1} = 0 \mid B = 1) \geq 7/9$. Furthermore, suppose next that the two events $B = 1$ and $A_1 = 0$ are observed. Then, $B = 1$ occurred despite $A_1 = 0$ with goodness 0.8, as $P(B_{\overline{a}_1} = 1) = 0.2 \leq 1 - 0.8$.

5.2 Results

In addition to the probabilistic causality notions in [9, 30, 31] that we introduced in the previous subsection, we also analyze the complexity of deciding whether the probability of a causal formula is at least α, and whether the conditional probability of a causal formula given a causal formula is at least α. We also analyze the complexity of computing the probability of a causal formula. Our complexity results are summarized in Table 3. In detail, deciding probabilistic causal irrelevance is $\mathrm{C}_=\mathrm{P}$-complete and computing the probability of a causal formula is #P-complete, while the other four problems are all **C**-complete. We remark that for all these six problems, hardness holds even if M is binary and bounded, and P is the uniform distribution. Furthermore, we would obtain similar results for causal formulas if "at least" is replaced by "at most".

The above results are nontrivial and need some explanations. Firstly, the problems of testing likely causes of events, occurrences of events despite other events, whether the probability of a causal formula is at least α, and whether the conditional probability over two causal formulas is at least α are harder than co-NP, and thus not polynomially reducible to SAT-testing. Moreover, these problems cannot be reduced to any solver for problems that are located in the Polynomial Hierarchy. On the other hand, they are solvable in polynomial space, and thus reducible to a QBF solver (e.g. [4, 32, 8]) in polynomial time. Furthermore, testing probabilistic causal irrelevance is "easier" than **C**-complete problems, which could perhaps help in finding polynomial time (randomized) approximation algorithms for this problem.

We remark that for computing the conditional probability $P(\psi \mid \phi)$ over two causal formulas, our result on computing $P(\phi)$ for a single causal formula ϕ implies that this problem can be described as the ratio f_1/f_2 of two #P-computable functions f_1 and f_2, respectively. However, a normalized representation, similar as the one that we have chosen for elementary probabilities in Assumption (3) of Section 2.2, seems not straightforward, and so is a representation of $P(\psi \mid \phi)$ in #P. We can, however, polynomially reduce the problem (in the sense for functions, see e.g. [29]) to a function g in #P, such that the desired $P(\psi \mid \phi)$ is easily read off from the value of $g(I)$ at the original problem input I.

Table 3: Complexity of Probabilistic Causality

Problem	Complexity
X is probabilistically causally irrelevant to Y given Z	$\mathbf{C}_{=}$-complete
$X = x$ is a likely cause of ϕ with goodness α	**C**-complete
ϕ occurred despite $X = x$ with goodness α	**C**-complete
probability of a causal formula is at least α	**C**-complete
conditional probability over two causal formulas is at least α	**C**-complete
probability of a causal formula	#P-complete

The following theorem shows that deciding probabilistic causal irrelevance is complete for $\mathbf{C}_{=}$. Here, the membership part is proved by a reduction to an exponential number of instances of the $\mathbf{C}_{=}$-complete problem EQUALRUN: 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. This locates the problem in the complexity class $\forall\mathbf{C}_{=}$, which is is a generalization of $\mathbf{C}_{=}$ similar to Π_2^P for co-NP [35]: A problem P is in $\forall\mathbf{C}_{=}$, if there exists a problem P' in $\mathbf{C}_{=}$ such that I is a Yes-instance of P iff for every string J of size polynomial in the size of I, it holds that I, J is a Yes-instance of P'. Thus, deciding probabilistic causal irrelevance is in the class $\forall\mathbf{C}_{=}$, which somewhat surprisingly coincides with $\mathbf{C}_{=}$. The hardness part is shown by a reduction from the $\mathbf{C}_{=}$-complete problem HALFSAT.

Theorem 5.4 *Given a probabilistic causal model (M, P), where $M = (U, V, F)$, and $X, Y, Z \subseteq V$ with $X, Y \neq \emptyset$, deciding whether $(X \not\rightarrow Y \mid Z)_P$ is complete for $\mathbf{C}_{=}$. Hardness holds even if M is binary and* *bounded, Z is empty, X, Y are singletons, and P is the uniform distribution.*

Proof. We first show membership in $\mathbf{C}_=$. Recall that $(X \not\rightarrow Y \mid Z)_P$ holds iff $P(Y_{xz} = y) = P(Y_{x'z} = y)$ for all values x', x, y, and z. We now show that deciding whether $P(Y_{xz} = y) = P(Y_{x'z} = y)$ holds, for given x, x', y, and z, can be transformed in polynomial time to the following problem EQUALRUN, which is in $\mathbf{C}_=$ by Lemma 5.5: 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.

Lemma 5.5 (cf. [28, 14]) EQUALRUN *is in* $\mathbf{C}_=$.

Let P be given by (f, b) as described in Section 2.2. Let T_1 and T_2 be NP Turing machines which on input x, x', y, z nondeterministically select an $u \in D(U)$ and 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 is that the problem is in the class $\forall\mathbf{C}_=$. However, the following nontrivial result is known:

Lemma 5.6 (cf. [28, 14]) $\forall\mathbf{C}_= = \mathbf{C}_=$*, that is, the classes coincide.*

Thus, our reduction in fact proves membership in $\mathbf{C}_=$.

Hardness for $\mathbf{C}_=$ is shown by a reduction from HALFSAT, which is $\mathbf{C}_=$-complete by Lemma 5.7 (see Appendix C): Given a Boolean formula ϕ on the atoms $A_1, \ldots, A_n$ with $n \geq 1$, decide whether exactly half of the truth assignments to $A_1, \ldots, A_n$ satisfy ϕ.

Lemma 5.7 HALFSAT *is complete for* $\mathbf{C}_=$.

We now construct a probabilistic causal model (M, P), where $M = (U, V, F)$, and $X, Y, Z \subseteq V$ with $X, Y \neq \emptyset$, such that $(X \not\rightarrow Y \mid Z)_P$ iff exactly 2^{n-1} truth assignments to $A_1, \ldots, A_n$ satisfy ϕ. The construction is based on similar ideas as the construction shown in Fig. 4, but more involved.

We define M as follows. The exogenous and endogenous variables are defined by $U = \{A_1, \ldots, A_n\}$ and $V = \{A, B\} \cup \{D_\alpha \mid \alpha \text{ is a subformula of } \phi\}$, respectively, where $D(S) = \{0, 1\}$ for all $S \in U \cup V$. The functions $F = \{F_S \mid S \in V\}$ are defined as follows:

- $F_{D_{A_i}} = A_i$ for every $i \in \{1, \ldots, n\}$,
- $F_{D_\alpha} = \neg D_\beta$ for every subformula of ϕ of the form $\alpha = \neg\beta$,
- $F_{D_\alpha} = D_\beta \wedge D_\gamma$ for every subformula of ϕ of the form $\alpha = \beta \wedge \gamma$,
- $F_A = 1$,
- $F_B = D_\phi \equiv A$.

We define $P(u) = 2^{-n}$ for all $u \in D(U)$. We define $X = \{A\}$, $Y = \{B\}$, and $Z = \emptyset$. The values 0 and 1 of X are denoted by x_0 and x_1, respectively. We now show that $(X \not\rightarrow Y \mid Z)_P$ iff exactly 2^{n-1} truth assignments to $A_1, \ldots, A_n$ satisfy ϕ.

Denote by m_1 (resp., m_0) the number of truth assignments in which ϕ is true (resp., false). For every $u \in D(U)$, it holds that $Y_{x_0}(u) = 1$ iff ϕ is false under u, and that $Y_{x_1}(u) = 1$ iff ϕ is true under

u. It thus follows $P(Y_{x_0}=1)=m_0\cdot 2^{-n}$ and $P(Y_{x_1}=1)=m_1\cdot 2^{-n}$. Moreover, for every $u\in D(U)$, it holds that $Y_{x_0}(u)=0$ iff ϕ is true under u, and that $Y_{x_1}(u)=0$ iff ϕ is false under u. Hence, it follows $P(Y_{x_0}=0)=m_1\cdot 2^{-n}$ and $P(Y_{x_1}=0)=m_0\cdot 2^{-n}$. In summary, this shows that $P(Y_x=y)=P(Y_{x'}=y)$ for all $x,x'\in D(X)$ and $y\in D(Y)$ iff $m_0=m_1$. That is, $(X\not\rightarrow Y\mid Z)_P$ iff ϕ is satisfied by exactly 2^{n-1} truth assignments.

Notice that M is binary and bounded, Z is empty, X,Y are singletons, and P is the uniform distribution. Thus, hardness holds even in this restricted setting. □

The next theorem shows that deciding whether the probability of a causal formula is at least α is complete for **C**. Here, the membership result is shown by a reduction to deciding whether an NP Turing machine has at least m accepting runs. The hardness result is proved by a reduction from GE-HALFSAT.

Theorem 5.8 *Given a probabilistic causal model (M,P), where $M=(U,V,F)$, a causal formula ψ, and $\alpha\geq 0$, deciding whether $P(\psi)\geq\alpha$ is* **C***-complete. Hardness holds even if M is binary and bounded, P is the uniform distribution, ψ is a counterfactual formula of the form $Y_x=y$ with singletons X,Y, and $\alpha=0.5$.*

Proof. We first show membership in **C**. Let P be given by (f,b) as described in Section 2.2. Let the NP Turing machine T have input ψ and nondeterministically generate all $u\in D(U)$. Then, for each u, let T generate $f(u)$ paths, and decide whether $(M,u)\models\psi$ holds, which can be done in polynomial time by Proposition 5.1. On each of these paths, T accepts if $(M,u)\models\psi$, otherwise it rejects. Then, $P(\psi)\geq\alpha$ iff T has at least $\alpha\cdot b$ accepting paths.

C-Hardness is shown by a reduction from GE-HALFSAT, which is **C**-complete by Lemma 5.9 (see Appendix C): Given a Boolean formula ϕ on the atoms $A_1,\ldots,A_n$ with $n\geq 1$, decide whether at least half of the truth assignments to $A_1,\ldots,A_n$ satisfy ϕ.

Lemma 5.9 GE-HALFSAT *is complete for* **C**.

We define the causal model $M=(U,V,F)$, the probability distribution P, and the sets of variables $X,Y\subseteq V$ as in the proof of Theorem 5.4. Denote by x_1 the value 1 of X, and define ψ as $Y_{x_1}=1$. Observe that M is binary and bounded, ψ is a counterfactual formula of the form $Y_x=y$, where X,Y are singletons, and P is the uniform distribution. As argued in the proof of Theorem 5.4, for every $u\in D(U)$, it holds that $Y_{x_1}(u)=1$ iff ϕ is true under u. Hence, $P(\psi)=P(Y_{x_1}=1)=m\cdot 2^{-n}$, where m is the number of truth assignments that satisfy ϕ. Thus, $P(\psi)\geq 0.5$ iff at least half of the truth assignments to $A_1,\ldots,A_n$ satisfy ϕ. □

The next theorem shows that also deciding occurrences of events despite other events is complete for **C**. Here, membership follows from Theorem 5.8, using that NP is contained in **C**, and that **C** is closed under polynomial-time conjunctive reductions [3]. Hardness is proved similarly as in the proof of Theorem 5.8.

Theorem 5.10 *Given a probabilistic causal model (M,P), where $M=(U,V,F)$, $X\subseteq V$, $x\in D(X)$, an event ψ, and $\alpha\in[0,1]$, deciding whether ψ occurred despite $X=x$ with goodness α is* **C***-complete. Hardness holds even if M is binary and bounded, P is the uniform distribution, X is a singleton, ψ is primitive, and $\alpha=0.5$.*

The following result states a **C**-completeness result similar to Theorem 5.8 for conditional probabilities $P(\psi\mid\phi)$ over causal formulas. However, we assume here that α is rational. Since probabilities $P(\phi)$ over causal formulas amount to a special case, hardness is immediate by Theorem 5.8.

Theorem 5.11 *Given a probabilistic causal model (M, P), where $M = (U, V, F)$, causal formulas ϕ and ψ, and a rational number $\alpha \in [0, 1]$, deciding whether $P(\psi \wedge \phi) \geq \alpha \cdot P(\phi)$ is* **C***-complete. Hardness holds even if M is binary and bounded, P is the uniform distribution, ψ is a counterfactual formula $Y_x = y$ with singletons X and Y, $\phi = \top$, and $\alpha = 0.5$.*

Proof. We first show membership in **C**. Let P be given by (f, b) as described in Section 2.2. We can compute in polynomial time integers $t > 0$ and $s \in [0, t]$ such that $\alpha = s \,/\, t$. Let the NP Turing machines T_1 and T_2 be constructed as in the proof of Theorem 5.8 such that they have exactly $b \cdot P(\psi \wedge \phi)$ and $b \cdot P(\phi)$ accepting runs,respectively. Let the NP Turing machine ${T_1}'$ (resp., ${T_2}'$) be obtained from T_1 (resp., T_2) by nondeterministically generating t (resp., s) paths for each run r of T_1 (resp., T_2) that corresponds to some $u \in D(U)$ with $f(u) > 0$, where each new run in ${T_1}'$ (resp., ${T_2}'$) accepts iff its old run r in T_1 (resp., T_2) accepts. Thus, the number of accepting runs of ${T_1}'$ (resp., ${T_2}'$) is $t \cdot b \cdot P(\psi \wedge \phi)$ (resp., $s \cdot b \cdot P(\phi)$). Let the NP Turing machine ${T_2}''$ be obtained from ${T_2}'$ by turning each accepting run into a non-accepting one, and vice versa, and by then adding $(t - s) \cdot b$ accepting runs. Observe now that ${T_1}'$ and ${T_2}''$ have the same number m of runs. Then, ${T_2}''$ has $m - s \cdot b \cdot P(\phi)$ accepting runs. Let the NP Turing machine T be obtained from ${T_1}'$ and ${T_2}''$ by nondeterministically entering either a run of ${T_1}'$ or a run of ${T_2}''$. Then, half of the runs of T are accepting iff $m - s \cdot b \cdot P(\phi) + t \cdot b \cdot P(\psi \wedge \phi) \geq m$, that is, $P(\psi \wedge \phi) \geq \alpha \cdot P(\phi)$.

Hardness for **C** is immediate by Theorem 5.8. □

The following theorem shows that deciding likely causes of events is **C**-complete. The membership part follows from Theorem 5.11, using that NP is contained in **C**, and that **C** is closed under polynomial-time conjunctive and disjunctive reductions [3]. The hardness part is proved similarly as in the proof of Theorem 5.8.

Theorem 5.12 *Given a probabilistic causal model (M, P), where $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, an event ψ, and a rational number $\alpha \in [0, 1]$, deciding whether $X = x$ is a likely cause of ψ with goodness α is complete for* **C**. *Hardness holds even if M is binary and bounded, P is the uniform distribution, X is a singleton, ψ is primitive, and $\alpha = 0.5$.*

We remark that for the result in the previous theorem, Assumption (2) in Section 2.2 about model representation, i.e., that domains are enumerated, is relevant. In fact, under implicit domain representation, the complexity of this problem increases to the class $\exists$**C** in the Counting Hierarchy, which coincides with the class $\mathrm{NP}^{\mathrm{PP}}$, i.e., nondeterministic polynomial time with an oracle for problems in PP [35] (see, e.g., [25] for other occurrences of this complexity class in AI problems). We leave the proof of $\exists$**C**-completeness, which can be obtained by minor adjustments of the proof of Theorem 5.11, to the interested reader.

The following result shows that computing the probability of a causal formula is complete for #P. The proof is similar to the proof of Theorem 5.8.

Theorem 5.13 *Given a probabilistic causal model (M, P), where $M = (U, V, F)$ and P is given by (f, b) as described in Section 2.2, and a causal formula ψ, computing $b \cdot P(\psi)$ is complete for* #P. *Hardness holds even if M is binary and bounded, P is the uniform distribution, and ψ is a counterfactual formula of the form $Y_x = y$ with singletons X, Y.*

Proof. We first show membership in #P. Let P be given by (f, b) as described in Section 2.2. Let the NP Turing machine T be defined as in the proof of Theorem 5.8. It then holds $P(\psi) = m \,/\, b$, where m is the number of accepting paths of T.

Hardness for #P is shown by a reduction from the following #P-complete problem #SAT. Given a Boolean formula ϕ on the atoms $A_1, \ldots, A_n$ with $n \geq 1$, compute the number of truth assignments to $A_1, \ldots, A_n$ that satisfy ϕ.

We define the causal model $M = (U, V, F)$, the probability distribution P, and the causal formula ψ as in the proof of Theorem 5.8. Let $(f, b) = (1, 2^n)$. It then holds $P(\psi) = m \cdot 2^{-n}$, where m is the number of truth assignments that satisfy ϕ. $\Box$

6 Discussion and Conclusion

In this section, we summarize our results and discuss related work. We then describe some applications of our results and give an outlook on future research.

6.1 Summary

In this paper, we have studied the computational complexity of causal relationships in Pearl's structural-model approach. We have considered several different such relationships from [9, 30, 31, 16, 17, 18], which can be classified into notions of causality between variables, notions of event causality, and notions of probabilistic causality. In the course of our analysis, we have also established some novel semantic results about weak and actual causes as defined by Halpern and Pearl [16], in particular, settling an open conjecture. The notions of causality between variables and of event causality are located in classes at the low end of the Polynomial Hierarchy, and are with the exception of weak and actual causes complete for NP or $\mathrm{co\text{-}NP}$, or only mildly harder. The notions of probabilistic causality are characterized by counting classes corresponding to NP, in particular, by $\mathbf{C}$ and $\mathbf{C}_{=}$. All notions of causality that we have considered are intractable, where the more sophisticated notions of weak and actual cause, and the notions of probabilistic causality have the highest complexity (Σ_2^P and $\mathbf{C}$, respectively).

The results for the simpler notions of event causality ("always causes" and "may cause") are similar to the results for causality between variables; this is intuitively explained by the fact that all these notions reduce to (possibly combined) tests for the existence of certain hypothetical scenarios (given by causal submodels), which can be guessed and then checked in polynomial time. The more sophisticated notions of event causality (weak and actual causality), however, have much more involved conditions, which request to visit a (possibly exponential) number of further scenarios to prove that a guess for a hypothetical scenario is suitable. Because of the complexity gap, we cannot efficiently transform weak and actual causality to simpler notions of event and variable causality. In other terms, weak and actual causality are more expressive, as they are capable of representing harder problems. Compared to notions of probabilistic causality, weak and actual causes have neither lower nor higher complexity, while all the other notions of event and variable causality have lower complexity. Among the probabilistic causal relationships, probabilistic irrelevance has the lowest complexity ($\mathbf{C}_{=}$-complete); even if intractable, it may have its uses (see below).

6.2 Related work

To our knowledge, there is no previous work in the literature on the computational complexity of causal relationships in the structural-model approach. The presented complexity results on probabilistic causal relationships are in some sense related to the complexity of probabilistic inference, in particular, in Bayesian networks and in probabilistic logic. As for Bayesian networks, Cooper [5] showed that deciding $P(X = x) > 0$

is NP-complete, and by the results of Roth [33], deciding $P(X=x) > \alpha$ for a rational number α is **C**-complete, while computing the exact probability $P(X=x)$ is #P-complete (assuming our representation). Note that by applying the techniques used in this paper, we can easily derive that evaluating probabilistic conditional statements $P(X=x \mid Y=y) \geq \alpha$ in Bayesian networks is **C**-complete. Only weakly related is work on the complexity of maximum a posteriori explanations (MAPs) in Bayesian networks [34], which are assignments to all variables extending given partial assignments such that the probability is maximum. Computing a MAP may be viewed as a classical constraint optimization problem over contexts, where no special relationships between variables or contexts are of interest. This is quite different from the various notions of causality that we have considered here. Other less related work concerns reasoning about conditional probability statements in probabilistic logic. Here, the results by Fagin et al. [7] imply that deciding whether $P(\psi \mid \phi) \in [\alpha, \beta]$ for all models P of a finite set of conditional probability statements is co-NP-complete, while Lukasiewicz [26] has shown that computing the tightest interval $[\alpha, \beta]$ such that $P(\psi \mid \phi) \in [\alpha, \beta]$ for all models P of a finite set of conditional probability statements is $\mathrm{FP}^{\mathrm{NP}}$-complete, i.e., complete for the class of functions computable in polynomial time with an NP oracle.

6.3 Applications of results

Our complexity results show that, similar to independencies [31], deterministic and probabilistic causal relationships might be used to simplify the evaluation of probabilistic counterfactuals [1], which are special causal formulas: Besides the simple form $P(Y_x = y)$, a more involved form is $P(Z_{x'} = z \mid X = x \wedge Y = y)$, which reads "if X is x and Y is y in the real world, and if X were x', what is then the probability that Z is z ?". 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 (#P-hardness). Furthermore, tests with complexity at most D^{P} can be polynomially reduced to at most two calls to a SAT-solver (e.g. [2, 39]); note that reducing problems to satisfiability checking proved useful in other contexts [21].

Our complexity results for the notion of weak cause proved useful for analyzing the computational aspects of explanations and partial explanations as introduced by Halpern and Pearl [16], which are crucially based on the notion of weak cause. Roughly speaking, an explanation is a minimal expression $X = x$ that justifies an event ϕ with respect to a *set* of contexts, through $X = x$ being a weak cause of ϕ in all contexts where $X = x$ holds; see [16, 18] for a precise definition. As shown in a companion paper [6], the conceptually more involved notion of explanation has also higher computational complexity than weak cause, and resides at the third level of the polynomial hierarchy. Moreover, the restriction to binary causal models makes the complexity drop by one level, as in the case of weak causality.

6.4 Open issues

The work in this paper provides a first step to detailed understanding of the computational aspects of causality, as defined on Pearl's causal models. While we have presented some basic results, several issues remain open at this point, and settling them requires further efforts. 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 this purpose, suitable classes of causal and probabilistic causal models need to be isolated, which should be efficiently recognizable. These classes may be defined in terms of conditions on the causal graph associated with a causal model. For example, probabilistic causal irrelevance in stable causal models [9] can be tested in polynomial time, as it coincides with path interception in their causal graphs. Other obvious conditions that work for some notions of causality can be obtained by imposing bounds on the depth of the causal graph.

Another issue is the development of algorithms and implementations. For this, transformations of problems on causality to other problems in knowledge representation and reasoning might be considered. In particular, it would be interesting to see whether existing computational logic systems can be profitable used as an implementation framework for this purpose.

Acknowledgments

This work has been partially supported by the Austrian Science Fund under 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. Furthermore, we are thankful to the reviewers of the IJCAI-01 abstract of this paper, whose constructive comments were very helpful to improve our work.

A Appendix: Proofs for Section 3

Proof of Theorem 3.2 (continued). It remains to show that X is causally irrelevant to Y given Z iff ϕ is a tautology.

(a) For every nonempty set of variables $W \subseteq V \setminus X \cup Y \cup Z = \{D_1, \ldots, D_{k-1}\}$ and every $w \in D(W)$, it holds $Y_{xw}(u) = Y_{x'w}(u)$ for all $x, x' \in D(X)$ and $u \in D(U)$. Assume now $W = \emptyset$. Then, $Y_{x_1}(u) = 1$, and $Y_{x_0}(u)$ is the truth value of ϕ under $u \in D(U)$. Thus, $Y_x(u) = Y_{x'}(u)$ for all $u \in D(U)$ and $x, x' \in D(X)$ iff ϕ is a tautology. In summary, $Y_{xw}(u) = Y_{x'w}(u)$ for all $W \subseteq V \setminus X \cup Y \cup Z$ and all values $u \in D(U)$, $x, x' \in D(X)$, and $w \in D(W)$ iff ϕ is a tautology. That is, X is causally irrelevant to Y given Z iff ϕ is a tautology.

(b) Since $V = X \cup Y \cup Z$, it holds $W = \emptyset$. Observe that $Y^i_x(u) = Y^i(u) = Y^i_{x'}(u)$ for all $i \in \{1, \ldots, k\}$, $x, x' \in D(X)$, and $u \in D(U)$. Moreover, $Y^{k+1}_{x_1}(u) = 1$, and $Y^{k+1}_{x_0}(u)$ is set to the truth value of ϕ under the truth assignment given by $u \in D(U)$. Thus, $Y^{k+1}_{x_0}(u) = Y^{k+1}_{x_1}(u)$ for all $u \in D(U)$ iff ϕ is a tautology. In summary, this shows that $Y_x(u) = Y_{x'}(u)$ for all $u \in D(U)$ and $x, x' \in D(X)$ iff ϕ is a tautology. That is, X is causally irrelevant to Y given Z iff ϕ is a tautology. $\Box$

Proof of Theorem 3.3. The problem in (a) (resp., (b)) is in NP, since values $x, x' \in D(X)$ and $u \in D(U)$ such that $Y_x(u) \neq Y_{x'}(u)$ (resp., $Y_{xz}(u) \neq Y_{x'z}(u)$) can be guessed and verified in polynomial time, by Proposition 2.3.

We next show NP-hardness. As X is a cause of Y iff X is a cause of Y in the context $Z = \emptyset$, the problem in (a) is a special case of the one in (b). It is thus sufficient to show NP-hardness for (a). We give a polynomial transformation from the NP-complete problem of deciding whether a given propositional formula in 3DNF $\phi = \phi_1 \vee \cdots \vee \phi_k$ on the variables $A_1, \ldots, A_n$, where $k, n \geq 1$, is not a tautology.

We now construct $M = (U, V, F)$ and $X, Y \subseteq V$, where M is binary and bounded, and X, Y are singletons, such that X is a cause of Y iff ϕ is not a tautology.

We define the causal model $M = (U, V, F)$ as in the proof of Theorem 3.2 for the case (1)–(3) and (4b). The corresponding causal graph is shown in Fig. 4. We then define the sets of endogenous variables $X, Y \subseteq V$ by $X = \{A\}$ and $Y = \{B\}$. The values 0 and 1 of X are denoted by x_0 and x_1, respectively.

Then, $Y_{x_1}(u) = 1$ for all $u \in D(U)$, and $Y_{x_0}(u)$ is set to the truth value of ϕ under the truth assignment given by $u \in D(U)$. Thus, $Y_{x_0}(u) = Y_{x_1}(u)$ for all $u \in D(U)$ iff ϕ is a tautology. That is, $Y_x(u) = Y_{x'}(u)$

for all $x, x' \in D(X)$ and $u \in D(U)$ iff ϕ is a tautology. That is, $Y_x(u) \neq Y_{x'}(u)$ for some $x, x' \in D(X)$ and $u \in D(U)$ iff ϕ is not a tautology. That is, X is a cause of Y iff ϕ is not a tautology. $\square$

Proof of Theorem 3.4. (a) The problem is in NP, as values $x, x' \in D(X)$, $u \in D(U)$, and $z \in D(V \setminus X \cup Y)$ such that $Y_{xz}(u) \neq Y_{x'z}(u)$ can be guessed and verified in polynomial time, by Proposition 2.3.

To show NP-hardness, we give a polynomial transformation from the following NP-complete problem. Given a propositional formula in 3DNF $\phi = \phi_1 \vee \cdots \vee \phi_k$ on the variables $A_1, \ldots, A_n$, where $k, n \geq 1$, decide whether ϕ is not a tautology.

We define the causal model $M = (U, V, F)$ and the sets of endogenous variables $X, Y \subseteq V$ as in the proof of Theorem 3.2 for the case (1)–(3) and (4b). As shown there, $Y_x(u) = Y_{x'}(u)$ for all $x, x' \in D(X)$ and $u \in D(U)$ iff ϕ is a tautology. That is, $Y_x(u) \neq Y_{x'}(u)$ for some $x, x' \in D(X)$ and $u \in D(U)$ iff ϕ is not a tautology. That is, X is a direct cause of Y iff ϕ is not a tautology.

As M is binary and bounded, X is a singleton, and $V = X \cup Y$, hardness holds even in this restricted case.

(b) The problem is in $\mathrm{D^P}$, as it can be expressed as the logical conjunction of a problem in NP and a problem in co-NP, by (a) and Theorem 3.3 (a).

Hardness for $\mathrm{D^P}$ is shown by a polynomial transformation from the following $\mathrm{D^P}$-complete problem [29]. Given two propositional formulas in 3DNF $\alpha = \alpha_1 \vee \cdots \vee \alpha_k$ and $\beta = \beta_1 \vee \cdots \vee \beta_l$ on the variables $A_1, \ldots, A_m$ and $B_1, \ldots, B_n$, respectively (where $k, l, m, n \geq 1$), decide whether α is a tautology and β is not a tautology. Without loss of generality, α and β do not share any variables.

We now construct a causal model $M = (U, V, F)$ and $X, Y \subseteq V$, where M is binary and bounded, and X is a singleton, such that X is an indirect cause of Y iff α is a tautology and β is not a tautology.

We define the causal model $M = (U, V, F)$ as follows. The exogenous and endogenous variables are defined by $U = \{B_1, \ldots, B_n\}$ and $V = \{A, A_1, \ldots, A_m, B, C_1, \ldots, C_{k+1}, D_1, \ldots, D_l\}$, respectively, where $D(S) = \{0, 1\}$ for all $S \in U \cup V$. The functions $F = \{F_S \mid S \in V\}$ are defined as follows:

- $F_S = 1$ for all $S \in \{A, A_1, \ldots, A_m, B\}$,
- $F_{C_1} = A \vee \alpha_1$,
- $F_{C_i} = C_{i-1} \vee \alpha_i$ for all $i \in \{2, \ldots, k\}$,
- $F_{C_{k+1}} = C_k \vee B$,
- $F_{D_1} = B \vee \beta_1$,
- $F_{D_i} = D_{i-1} \vee \beta_i$ for all $i \in \{2, \ldots, l\}$.

The corresponding causal graph is shown in Fig. 6. We define the sets of variables $X, Y \subseteq V$ by $X = \{B\}$ and $Y = \{C_1, \ldots, C_{k+1}, D_l\}$. Then, $Z = V \setminus (X \cup Y) = \{A, A_1, \ldots, A_m, D_1, \ldots, D_{l-1}\}$. We use $Y^1, \ldots, Y^{k+1}, Y^0$ to denote $C_1, \ldots, C_{k+1}, D_l$. The values 0 and 1 of X are denoted by x_0 and x_1, respectively.

We now prove that X is an indirect cause of Y iff α is a tautology and β is not a tautology. More precisely, we first show that X is not a direct cause of Y iff α is a tautology. We then prove that X is a cause of Y iff β is not a tautology.

Observe that $Y^0_{x_1 z}(u) = Y^0_z(u) = Y^0_{x_0 z}(u)$ for all $z \in D(Z)$ and $u \in D(U)$. Furthermore, $Y^i_{x_1 z}(u) = Y^i_z(u) = Y^i_{x_0 z}(u)$ for all $i \in \{1, \ldots, k\}$, $z \in D(Z)$, and $u \in D(U)$. Finally, $Y^{k+1}_{x_1 z}(u) = 1$ for all $z \in D(Z)$ and $u \in D(U)$, and $Y^{k+1}_{x_0 z}(u)$ is set to the truth value of $A \vee \alpha$ under $z \in D(Z)$, for all $u \in D(U)$. Hence, $Y^{k+1}_{x_1 z}(u) = Y^{k+1}_{x_0 z}(u)$ for all $z \in D(Z)$ and $u \in D(U)$ iff α is a tautology. In summary, $Y_{xz}(u) = Y_{x'z}(u)$ for

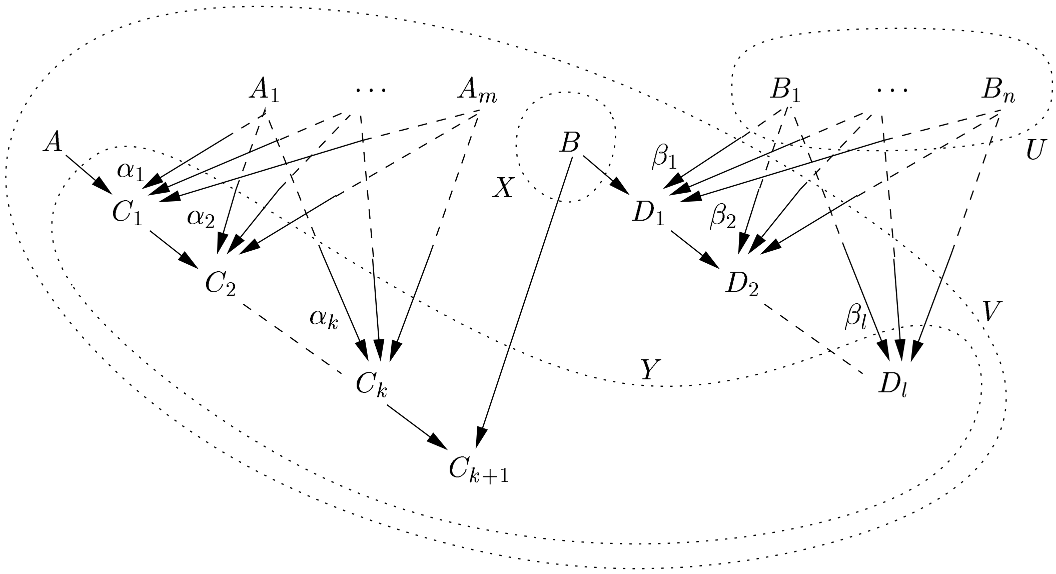

Figure 6: Causal Graph for D^{P}-Hardness of Indirect Cause

all $x, x' \in D(X)$, $z \in D(Z)$, and $u \in D(U)$ iff α is a tautology. That is, X is not a direct cause of Y iff α is a tautology.

Observe then that $Y^i_{x_1}(u) = 1 = Y^i_{x_0}(u)$ for all $i \in \{1, \ldots, k{+}1\}$ and $u \in D(U)$. Furthermore, $Y^0_{x_1}(u) = 1$ for all $u \in D(U)$, and $Y^0_{x_0}(u)$ is set to the truth value of β under $u \in D(U)$. Thus, $Y^0_{x_0}(u) = Y^0_{x_1}(u)$ for all $u \in D(U)$ iff β is a tautology. In summary, $Y_x(u) = Y_{x'}(u)$ for all $x, x' \in D(X)$ and $u \in D(U)$ iff β is a tautology. That is, $Y_x(u) \neq Y_{x'}(u)$ for some $x, x' \in D(X)$ and $u \in D(U)$ iff β is not a tautology. That is, X is a cause of Y iff β is not a tautology. $\Box$

B Appendix: Proofs for Section 4

Proof of Theorem 4.3. (a) The problem is in D^{P}, as it can be expressed as the logical conjunction of a problem in co-NP and a problem in NP. In detail, a context $u \in D(U)$ such that $\neg\phi_x(u)$ can be guessed and verified in polynomial time, by Proposition 4.1. Hence, deciding whether (i) holds is in co-NP. Moreover, two values $x' \in D(X)$ and $u' \in D(U)$ such that $x' \neq x$ and $\neg\phi_{x'}(u')$ can also be guessed and verified in polynomial time, by Proposition 4.1. Thus, deciding whether (ii) holds is in NP. In summary, deciding whether (i) and (ii) holds is D^{P}.

Hardness for D^{P} is shown by a polynomial transformation from the following D^{P}-complete problem [29]. Given two propositional formulas in 3DNF $\alpha = \alpha_1 \vee \cdots \vee \alpha_k$ and $\beta = \beta_1 \vee \cdots \vee \beta_l$ on the variables $A_1, \ldots, A_m$ and $B_1, \ldots, B_n$, respectively (where $k, l, m, n \geq 1$), decide whether α is a tautology and β is not a tautology. Without loss of generality, α and β do not share any variables.

We now construct $M = (U, V, F)$, $X, Y \subseteq V$, $x \in D(X)$, and $y \in D(Y)$, where M is binary and bounded, and X is a singleton, such that $X = x$ always causes $Y = y$ iff α is a tautology and β is not a tautology.

We define the causal model $M = (U, V, F)$ as follows. The exogenous and endogenous variables are defined by $U = \{A_1, \ldots, A_m, B_1, \ldots, B_n\}$ and $V = \{A, C_1, \ldots, C_k, D_1, \ldots, D_l\}$, respectively, where $D(S) =$

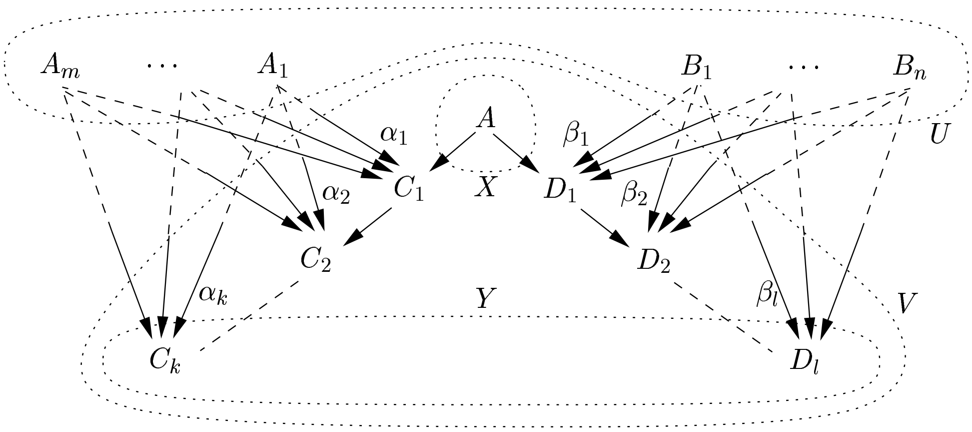

Figure 7: Causal Graph for $\mathrm{D^P}$-Hardness of Necessary Cause

$\{0,1\}$ for all $S \in U \cup V$. The functions $F = \{F_S \mid S \in V\}$ are defined as follows:

- $F_A = 1$,
- $F_{C_1} = A \vee \alpha_1$,
- $F_{C_i} = C_{i-1} \vee \alpha_i$ for all $i \in \{2, \ldots, k\}$,
- $F_{D_1} = \neg A \vee \beta_1$,
- $F_{D_i} = D_{i-1} \vee \beta_i$ for all $i \in \{2, \ldots, l\}$.

The corresponding causal graph is shown in Fig. 7. We define the endogenous variables $X, Y \subseteq V$ by $X = \{A\}$ and $Y = \{C_k, D_l\}$. We write Y^0 and Y^1 to denote C_k and D_l, respectively. The values 0 and 1 of X are denoted by x_0 and x_1, respectively. Let $y \in D(Y)$ be defined by $y(Y^0) = y(Y^1) = 1$. We now prove that $X = x_0$ always causes $Y = y$ iff α is a tautology and β is not a tautology.

We have $Y^1_{x_0}(u) = 1$ for all $u \in D(U)$, and $Y^0_{x_0}(u) = 1$ for all $u \in D(U)$ iff α is a tautology. Hence, $Y_{x_0}(u) = y$ for all $u \in D(U)$ iff α is a tautology.

Moreover, $Y^0_{x_1}(u) = 1$ for all $u \in D(U)$, and $Y^1_{x_1}(u) \neq 1$ for some $u \in D(U)$ iff β is not a tautology. Thus, $Y_{x_1}(u) \neq y$ for some $u \in D(U)$ iff β is not a tautology.

In summary, $Y_{x_0}(u) = y$ for all $u \in D(U)$ and $Y_{x_1}(u') \neq y$ for some $u' \in D(U)$, iff α is a tautology and β is not a tautology. That is, $X = x_0$ always causes $Y = y$ iff α is a tautology and β is not a tautology.

(b) The problem is in NP, as some values $x' \in D(X)$ and $u \in D(U)$ such that $x \neq x'$, $X(u) = x$, $\psi(u)$, and $\neg\psi_{x'}(u)$ can be guessed and verified in polynomial time, by Propositions 2.3 and 4.1.

Hardness for NP is shown by a polynomial transformation from the following NP-complete problem. Given a propositional formula in 3DNF $\phi = \phi_1 \vee \cdots \vee \phi_k$ on the variables $A_1, \ldots, A_n$, where $k, n \geq 1$, decide whether ϕ is not a tautology.

We define the causal model $M = (U, V, F)$, the sets of variables $X, Y \subseteq V$, and the values $x_0, x_1 \in D(X)$ as in the proof of Theorem 3.3. Observe that M is binary and bounded, and that X, Y are singletons. We now show that $X = x_1$ may have caused $Y = 1$ iff ϕ is not a tautology.

Observe that $x_1 \neq x_0$, and that $X(u) = x_1$ and $Y(u) = 1$ for all $u \in D(U)$. As argued in the proof of Theorem 3.3, $Y_{x_0}(u) = 1$ for all $u \in D(U)$ iff ϕ is a tautology. Equivalently, $Y_{x_0}(u) \neq 1$ for some $u \in D(U)$ iff ϕ is not a tautology. In summary, $X = x_1$ may have caused $Y = 1$ iff ϕ is not a tautology. $\Box$

Proof of Theorem 4.4 (continued). It remains to show that $X = x$ is a weak cause of ϕ under u iff Φ is true. Clearly, $X(u) = 0$ and $Y(u) = 1$. That is, AC1 holds. We now show that AC2 holds iff Φ is true.

($\Rightarrow$) Assume AC2 holds. That is, some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that $Y_{\overline{x}w}(u) = 0$ and $Y_{xw\hat{z}}(u) = 1$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. Notice that $\overline{x} \neq x$ and $H \notin W$, as otherwise $Y_{\overline{x}w}(u) = Y_{xw}(u)$. Moreover, $C \in W$ and $w(C) = 1$, as otherwise $Y_{\overline{x}w}(u) = 1$. Observe then that $B \cap W = \emptyset$, as otherwise $Y_{\overline{x}w}(u) = 1$ if $S \in B \cap W$ and $w(S) \in \{0, 1\}$, and $Y_{xw}(u) = 0$ if $S \in B \cap W$ and $w(S) = 2$. This shows $W \setminus \{C\} \subseteq A$. Observe that $S_{xw}(u) = 0$ for all $S \in A \setminus W$. Define now the truth assignment I to the variables in A by $I(S) = \mathtt{true}$, iff $S \in W$ and $w(S) = 1$. Notice then that $S_{xw}(u) = 1$ and $S(u) = 0$, for all $S \in B$. Given any truth assignment J to the variables in B, let $\hat{Z}$ be the set of all $S \in B$ with $J(S) = \mathbf{false}$, and let $\hat{z} = \hat{Z}(u)$. In summary, for all $S \in A$, we have $S_{xw\hat{z}}(u) = 0$ iff $I(S) = \mathbf{false}$. Moreover, for all $S \in B$, we have $S_{xw\hat{z}}(u) \in \{0, 1\}$, and $S_{xw\hat{z}}(u) = 0$ iff $J(S) = \mathbf{false}$. As AC2 holds, we have $Y_{xw\hat{z}}(u) = 1$. That is, $\phi'_{xw\hat{z}}(u)$ is true, or equivalently, $\gamma'_{xw\hat{z}}(u)$ is true, which is equivalent to γ being true in $I \cup J$. In summary, a truth assignment I to the variables in A exists such that γ is true in $I \cup J$ for all truth assignments J to the variables in B. That is, Φ is true.

($\Leftarrow$) Assume Φ is true. That is, there exists a truth assignment I to the variables in A such that γ is true in $I \cup J$ for all truth assignments J to the variables in B. Define $\overline{x} = 1$, $W = \{C\} \cup A$, $w(C) = 1$, and $w(S) = 1$ iff $I(S) = \mathtt{true}$, for all $S \in A$. We now show that AC2 (a) and (b) hold. As $S_{\overline{x}w}(u) = 2$ for all $S \in B$, it follows $Y_{\overline{x}w}(u) = 0$. That is, AC2 (a) holds. Consider next any $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. Assume first that $Y \subseteq \hat{Z}$. As $Y(u) = 1$, it then follows $Y_{xw\hat{z}}(u) = 1$. Assume next that $\hat{Z} \subseteq B$. Define the truth assignment J to the variables in B by $J(S) = \mathbf{false}$ iff $S \in \hat{Z}$. In summary, for all $S \in A$, we now have $S_{xw\hat{z}}(u) = 0$ iff $I(S) = \mathbf{false}$. Moreover, for all $S \in B$, we have $S_{xw\hat{z}}(u) \in \{0, 1\}$, and $S_{xw\hat{z}}(u) = 0$ iff $J(S) = \mathbf{false}$ (as $S_{xw}(u) = 1$ and $S(u) = 0$). As γ is true in $I \cup J$, it thus follows that $\gamma'_{xw\hat{z}}(u)$ holds. As $S_{xw\hat{z}}(u) \in \{0, 1\}$ for all $S \in B$, we then get $Y_{xw\hat{z}}(u) = 1$. In summary, some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that $Y_{xw\hat{z}}(u) = 1$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. That is, AC2 (b) holds. $\Box$

Proof of Theorem 4.5 (continued). We show that for binary causal models M, condition AC2 is equivalent to the following condition AC2$''$. By AC2$''$ (c), setting $\hat{Z}'$ to $\hat{z}'$ is immaterial in $\phi_{xw'\hat{z}'}(u)$ of AC2$''$ (b). Thus, AC2$''$ is in fact equivalent to AC2$'$. This then proves the theorem.

AC2$''$. Some set of endogenous 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) $\phi_{xw'\hat{z}'}(u)$ for all $\hat{Z}' \subseteq V \setminus (X \cup W')$ and $\hat{z}' = \hat{Z}'(u)$,

(c) $Z'_{xw'}(u) = Z'(u)$ for $Z' = V \setminus (X \cup W')$.

($\Leftarrow$) Trivially, AC2$''$ implies AC2.

($\Rightarrow$) Assume that AC2 holds. Then, define $Z = V \setminus (X \cup W)$. Moreover, define the values $\overline{z}, z \in D(Z)$ by $\overline{z} = Z_{\overline{x}w}(u)$ and $z = Z_{xw}(u)$. Let $Z^\star$ denote the set of all variables $I \in Z$ such that either (α) $\overline{z}(I) = z(I)$ or (β) $\overline{z}(I) \neq z(I)$ and $I(u) \neq z(I)$. Note now that for binary variables I, condition (β) implies $\overline{z}(I) = I(u)$. Observe also that $I \in Z \setminus Z^\star$ implies $I(u) = z(I)$. Define now $W' = W \cup Z^\star$ and $Z' = Z \setminus Z^\star$, and define the values $\overline{x}' \in D(X)$, $w' \in D(W')$ by $\overline{x}' = \overline{x}$, $w'|W = w$, and $w'|Z^\star = \overline{z}|Z^\star$. It is now easy to verify that

AC2$''$ (a) and (b) hold. However, condition AC2$''$ (c) does not necessarily hold. But we can clearly iterate the just described construction until AC2$''$ (c) holds. $\Box$

Proof of Theorem 4.6. We first show membership in NP. By Theorem 4.5, it holds that $X = x$ is a weak cause of ϕ under u iff (i) $X(u) = x$ and $\phi(u)$, and (ii) 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 $Z = V \setminus (X \cup W)$. By Propositions 2.3 and 4.1, deciding (i) is polynomial. In (ii), some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ such that (a) to (c) hold can be guessed and verified in polynomial time, by Propositions 2.3 and 4.1. In summary, deciding whether (i) and (ii) holds is in NP.

To show NP-hardness, we give a polynomial transformation from the following NP-complete problem. Given a propositional formula in 3DNF $\phi = \phi_1 \vee \cdots \vee \phi_k$ on the variables $A_1, \ldots, A_n$, where $k, n \geq 1$, decide whether ϕ is not a tautology.

We now construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and a primitive event ϕ such that $X = x$ is a weak cause of ϕ under u iff ϕ is not a tautology.

We define the causal model $M = (U, V, F)$ as follows. The exogenous and endogenous variables are defined by $U = \{C\}$ and $V = \{A, A_1, \ldots, A_n, D_1, \ldots, D_k\}$, respectively, where $D(S) = \{0, 1\}$ for all $S \in U \cup V$. The functions $F = \{F_S \mid S \in V\}$ are defined as follows:

- $F_S = 1$ for all $S \in \{A, A_1, \ldots, A_n\}$,
- $F_{D_1} = A \vee \phi_1$,
- $F_{D_i} = D_{i-1} \vee \phi_i$ for all $i \in \{2, \ldots, k\}$.

The corresponding causal graph is similar to the one shown in Fig. 3. We define $X = \{A\}$, $Y = \{D_k\}$, and $T = \{A_1, \ldots, A_n\}$. The values 0 and 1 of X are denoted by x_0 and x_1, respectively. Let $x = x_1$ and $u \in D(U)$. Let ϕ be defined as $Y = 1$. We now show that $X = x$ is a weak cause of ϕ under u iff ϕ is not a tautology.

Observe first that for all $W \subseteq V \setminus X$ with $W \not\subseteq T$, $w \in D(W)$, and $u \in D(U)$, we have $Y_{xw}(u) = Y_{x'w}(u)$ for all $x, x' \in D(X)$. For all $W \subseteq V \setminus X$ with $W \subseteq T$, $w \in D(W)$, and $u \in D(U)$, we have $Y_{x_1 w}(u) = 1$, and $Y_{x_0 w}(u)$ is set to the truth value of ϕ under the truth assignment $I \colon T \to \{0, 1\}$ given by $I(S) = w(S)$ for all $S \in W$ and by $I(S) = 1$ for all $S \in T \setminus W$.

By Theorem 4.5, $X = x_1$ is a weak cause of ϕ under u iff (i) $X(u) = x_1$ and $Y(u) = 1$, and (ii) some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that $Y_{\overline{x}w}(u) \neq 1$, $Y_{x_1 w}(u) = 1$, and $Z_{x_1 w}(u) = Z(u)$ for $Z = V \setminus (X \cup W)$. Obviously, (ii) implies $\overline{x} = x_0$. Clearly, (i) holds. By the argumentation above, (ii) is equivalent to the existence of some $W \subseteq T$ and $w \in D(W)$ such that $Y_{x_0 w}(u) \neq 1$ and $Z_{x_1 w}(u) = Z(u)$ for $Z = V \setminus (X \cup W)$. Observe then that $S_{x_1 w}(u) = 1 = S(u)$ for all $S \in Z = (T \setminus W) \cup \{D_1, \ldots, D_k\}$. Hence, (ii) is in fact equivalent to the existence of some $W \subseteq T$ and $w \in D(W)$ such that $Y_{x_0 w}(u) \neq 1$. By the argumentation above, this condition is in turn equivalent to ϕ not being a tautology.

In summary, $X = x$ is a weak cause of ϕ under u iff ϕ is not a tautology. Observe that M is bounded, X is a singleton, and ϕ is primitive. $\Box$

C Appendix: Proofs for Section 5

Proof of Lemma 5.7. The problem is in $\mathrm{C}_=$, as it is a special case of the following $\mathrm{C}_=$-complete problem E-SAT, mentioned in the discussion of [37]: Given a Boolean formula ϕ and an integer m, decide whether ϕ has exactly m satisfying truth assignments. Hardness holds even if ϕ is a CNF.

Hardness for $\mathbf{C}_=$ is shown by a polynomial transformation from E-SAT, using similar techniques as in [38, 35]. Let ϕ be a Boolean formula on the atoms $x_1, \ldots, x_n$ with $n \geq 1$, and let m be an integer. We now construct a Boolean formula ϕ' on the atoms $x_0, x_1, \ldots, x_n$ such that ϕ has exactly m satisfying truth assignments to the atoms $x_1, \ldots, x_n$ iff ϕ' has exactly 2^n satisfying truth assignments to the atoms $x_0, x_1, \ldots, x_n$.

If $m = 0$, then ϕ has no satisfying truth assignments to $x_1, \ldots, x_n$ iff $\phi' = x_0 \vee \phi$ has exactly 2^n satisfying truth assignments to $x_0, x_1, \ldots, x_n$. Otherwise, let $b_1 \cdots b_n$ be the binary representation of $2^n - m$, that is, $2^n - m = \sum_{i=1}^{n} b_i \cdot 2^{i-1}$. We identify 0 and 1 with the propositional constants *false* and *true*, respectively. We define $\phi' = (x_0 \vee \phi_0) \wedge (\neg x_0 \vee \phi)$, where:

$$\phi_0 = \bigvee_{i=1}^{n} \left(b_i \wedge \neg x_i \wedge \left(\bigwedge_{j=i+1}^{n} b_j \equiv x_j \right) \right) .$$

Roughly, a truth assignment I to the atoms $x_1, \ldots, x_n$ satisfies ϕ_0 iff it is lexicographically smaller than $b_1 \cdots b_n$, where the truth values **false** and **true** are identified with 0 and 1, respectively. Thus, ϕ_0 has exactly $2^n - m$ satisfying truth assignments to the atoms $x_1, \ldots, x_n$. Hence, ϕ has exactly m satisfying truth assignments to $x_1, \ldots, x_n$ iff ϕ' has exactly 2^n satisfying truth assignments to $x_0, x_1, \ldots, x_n$.

If we start from a DNF ϕ, then the construction yields a formula ϕ' which is easily rewritten to a DNF. Obviously, $\mathbf{C}_=$-hardness of E-SAT for CNFs implies also $\mathbf{C}_=$-hardness for DNFs, and ϕ' is a Yes-Instance of HALFSAT iff $\neg\phi'$ is a Yes-Instance of HALFSAT. It follows that HALFSAT is $\mathbf{C}_=$-hard for both CNFs and DNFs ϕ. □

Proof of Lemma 5.9. The problem is in **C**, as it is a special case of the following **C**-complete problem GE-SAT: Given a Boolean formula ϕ and an integer m, decide whether ϕ has at least m satisfying truth assignments.

Hardness for **C** is shown by a polynomial transformation from GE-SAT. Let ϕ be a Boolean formula on the atoms $x_1, \ldots, x_n$ with $n \geq 1$, and let m be an integer. Let the Boolean formula ϕ' on the atoms $x_0, x_1, \ldots, x_n$ be defined as in the proof of Lemma 5.7. Then, ϕ has at least m satisfying truth assignments to $x_1, \ldots, x_n$ iff ϕ' has at least 2^n satisfying truth assignments to $x_0, x_1, \ldots, x_n$. Hardness under restriction to CNFs and DNFs ϕ follows by similar arguments as in Lemma 5.7. □

Proof of Theorem 5.10. As for membership in **C**, recall that ψ occurred despite $X = x$ with goodness α, iff (i) $\psi(u)$ and $X(u) = x$ for some $u \in D(U)$, and (ii) $P([X \leftarrow x]\, \psi) \leq 1 - \alpha$, that is, $P([X \leftarrow x]\, \neg\psi) \geq \alpha$. Guessing some $u \in D(U)$ and verifying that $\psi(u)$ and $X(u) = x$ hold can be done in polynomial time by Proposition 4.1. That is, deciding whether (i) holds is in NP and thus by Fig. 2 in **C**. By Theorem 5.8, deciding whether (ii) holds is in **C**. As **C** is closed under polynomial-time conjunctive reductions [3], the problem is in **C**.

Hardness for **C** is shown by a reduction from GE-HALFSAT. Let ϕ be a Boolean formula on the atoms $A_1, \ldots, A_n$ with $n \geq 1$. Without loss of generality, ϕ is not a tautology.

We define the causal model $M = (U, V, F)$, the probability distribution P, and the sets of variables $X, Y \subseteq V$ as in the proof of Theorem 5.4. Denote by x_1 the value 1 of X, and define ψ as $Y = 0$. Let $\alpha = 0.5$. Observe that M is binary and bounded, P is the uniform distribution, X is a singleton, and ψ is primitive.

By the proof of Theorem 5.4, for every $u \in D(U)$, it holds that $Y_{x_1}(u) = 1$ iff ϕ is true under u. Hence, we have $P([X \leftarrow x_1]\, \neg\psi) = m \cdot 2^{-n}$, and thus $P([X \leftarrow x_1]\, \psi) = 1 - m \cdot 2^{-n}$, where m is the number of satisfying truth assignments of ϕ. Hence, (ii) $P([X \leftarrow x_1]\, \psi) \leq 1 - \alpha$ holds, iff at least 2^{n-1} truth assignments to $A_1, \ldots, A_n$ satisfy ϕ. As ϕ is not a tautology, (i) $\psi(u)$ and $X(u) = x_1$ for some $u \in D(U)$. In summary, ψ occurred despite $X = x$ with goodness α, iff at least half of the truth assignments to $A_1, \ldots, A_n$ satisfy ϕ. $\Box$

Proof of Theorem 5.12. As for membership in **C**, recall that $X = x$ is a likely cause of ψ with goodness α, iff (i) $\phi(u)$ and $f(u) > 0$ for some $u \in D(U)$, and (ii) $P([X \leftarrow x]\, \phi \wedge [X \leftarrow x']\, \neg\phi \mid \phi) \geq \alpha$ for some $x' \in D(X) \setminus \{x\}$. Guessing some $u \in D(U)$ and verifying that $\phi(u)$ and $f(u) > 0$ holds is polynomial, by Proposition 4.1 and Assumption (3) in Section 2.2. Hence, deciding whether (i) holds is in NP, and thus by Fig. 2 also in **C**. By Theorem 5.11, deciding whether $P([X \leftarrow x]\, \phi \wedge [X \leftarrow x']\, \neg\phi \mid \phi) \geq \alpha$ for some fixed $x' \in D(X) \setminus \{x\}$ is in **C**. As **C** is closed under polynomial-time disjunctive reductions [3], and by Assumption (2) in Section 2.2, deciding whether (ii) holds is in **C**. In summary, as **C** is closed under polynomial-time conjunctive reductions [3], deciding whether (i) and (ii) holds is in **C**.

Hardness for **C** is shown by a reduction from GE-HALFSAT. Let ϕ be a Boolean formula on the atoms $A_1, \ldots, A_n$ with $n \geq 1$. Let the causal model $M = (U, V, F)$ be defined as in the proof of Theorem 5.4, except that the function $F_B = D_\phi \equiv A$ is replaced by $F_B = \neg D_\phi \vee A$. Let P be the uniform distribution over $D(U)$. Let $X = \{A\}$ and $Y = \{B\}$. Denote by x_0 and x_1 the values 0 and 1, respectively, of X, and define ψ as $Y = 1$. Let $\alpha = 0.5$. Observe that M is binary and bounded, X is a singleton, and ψ is primitive.

We now show that $X = x_1$ is a likely cause of ψ with goodness α iff at least half of the truth assignments to $A_1, \ldots, A_n$ satisfy ϕ. Observe first that for every $u \in D(U)$, it holds $(M, u) \models \psi$ and $(M, u) \models [X \leftarrow x_1]\, \psi$. Thus, $\psi(u)$ and $f(u) > 0$ for all $u \in D(U)$. Moreover, (ii) $P([X \leftarrow x_1]\, \psi \wedge [X \leftarrow x']\, \neg\psi \mid \psi) \geq \alpha$ for some $x' \in D(X) \setminus \{x_1\}$ is equivalent to $P(Y_{x_0} = 0) \geq \alpha$. Notice now that $(M, u) \models Y_{x_0} = 0$ iff ϕ is true under u. Hence, $P(Y_{x_0} = 0) = m \cdot 2^{-n}$, where m is the number of satisfying truth assignments of ϕ. Thus, $P(Y_{x_0} = 0) \geq \alpha$ iff at least 2^{n-1} truth assignments to $A_1, \ldots, A_n$ satisfy ϕ. In summary, $X = x_1$ is a likely cause of ψ with goodness α, iff at least half of the truth assignments to $A_1, \ldots, A_n$ satisfy ϕ. $\Box$

References

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

[2] R. Bayardo and R. Schrag. Using CSP look-back techniques to solve real-world SAT instances. In *Proceedings AAAI-97*, pages 203–208, 1997.

[3] R. Beigel, N. Reingold, and D. A. Spielman. PP is closed under intersection. *Journal of Computer and System Sciences*, 50(2):191–202, 1995.

[4] M. Cadoli, A. Giovanardi, and M. Schaerf. An algorithm to evaluate quantified boolean formulae. In *Proceedings AAAI-98*, pages 262–267, 1998.

[5] G. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[6] T. Eiter and T. Lukasiewicz. Complexity results for explanations in the structural-model approach. Technical Report INFSYS RR-1843-01-08, Institut für Informationssysteme, TU Wien, 2001.

[7] R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87:78–128, 1990.

[8] R. Feldmann, B. Monien, and S. Schamberger. A distributed algorithm to evaluate quantified boolean formulae. In *Proceedings AAAI-00*, pages 285–290, 2000.

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

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

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

[12] J. Gill. Computational complexity of probabilistic turing machines. *SIAM Journal on Computing*, 6:675–695, 1975.

[13] E. Giunchiglia and V. Lifschitz. An action language based on causal explanation: Preliminary report. In *Proceedings AAAI-98*, pages 623–630, 1998.

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

[15] J. Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.

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

[17] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach – Part I: Causes. In *Proceedings UAI-01*, pages 194–202, 2001.

[18] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach – Part II: Explanations. In *Proceedings IJCAI-01*, pages 27–34, 2001.

[19] M. Hopkins. A proof of the conjunctive cause conjecture. Unpublished, 2001.

[20] D. S. Johnson. A catalog of complexity classes. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume A, chapter 2, pages 67–161. MIT Press, Cambridge, MA, 1990.

[21] H. Kautz and B. Selman. Planning as satisfiability. In *Proceedings ECAI-92*, pages 359–363, 1992.

[22] V. Lifschitz. Situation Calculus and Causal Logic. In *Proceedings KR-98*, pages 536–546, 1998.

[23] F. Lin. Embracing causality in specifying the indirect effects of actions. In *Proceedings AAAI-95*, pages 1985–1993, 1995.

[24] F. Lin and Y. Shoham. Provably correct theories of action. *J. ACM*, 42(2):293–320, 1995.

[25] M. L. Littman, J. Goldsmith, and M. Mundhenk. The computational complexity of probabilistic planning. *Journal of Artificial Intelligence Research*, 9:1–36, 1998.

[26] T. Lukasiewicz. Probabilistic logic programming with conditional constraints. *ACM Transactions on Computational Logic*, 2(3):289–339, 2001.

[27] N. McCain and H. Turner. Causal theories of actions and change. In *Proceedings AAAI-97*, pages 460–465, 1997.

[28] M. Ogiwara. A characterization of $\mathrm{P}^{\mathrm{C=P}}$. *The Transactions of the Institute of Electronics, Information and Communication Engineers*, E75-D:44–49, 1992.

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

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

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

[32] J. Rintanen. Improvements to the evaluation of quantified boolean formulae. In *Proceedings IJCAI-99*, pages 1192–1197, 1999.

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

[34] S. E. Shimony. Finding MAPs for belief networks is NP-hard. *Artificial Intelligence*, 68:399–410, 1994.

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

[36] H. Turner. A logic of universal causation. *Artificial Intelligence*, 113:87–123, 1999.

[37] K. W. Wagner. The complexity of combinatorial problems with succinct input representation. *Acta Informatica*, 23:325–356, 1986.

[38] K. W. Wagner. More complicated questions about maxima and minima, and some closures of NP. In *Proceedings ICALP-86*, pages 434–443, 1986. Also *Theoretical Computer Science*, 51:53–80, 1987.

[39] H. Zhang. SATO: An efficient propositional prover. In *Proceedings CADE-97*, pages 272–275, 1997.