Complexity Results for Explanations in the Structural-Model Approach ⋆

Thomas Eiter [a,*] and Thomas Lukasiewicz [b,a]

[a] *Institut für Informationssysteme, Technische Universität Wien*
Favoritenstraße 9-11, 1040 Vienna, Austria

[b] *Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza"*
Via Salaria 113, 00198 Rome, Italy

Abstract

We analyze the computational complexity of Halpern and Pearl's (causal) explanations in the structural-model approach, which are based on their notions of weak and actual cause. In particular, we give a precise picture of the complexity of deciding explanations, α-partial explanations, and partial explanations, and of computing the explanatory power of partial explanations. Moreover, we analyze the complexity of deciding whether an explanation or an α-partial explanation over certain variables exists. We also analyze the complexity of deciding explanations and partial explanations in the case of succinctly represented context sets, the complexity of deciding explanations in the general case of situations, and the complexity of deciding subsumption and equivalence between causal models. All complexity results are derived for the general case, as well as for the restriction to the case of binary causal models, in which all endogenous variables may take only two values. To our knowledge, no complexity results for explanations in the structural-model approach have been derived so far. Our results give insight into the computational structure of Halpern and Pearl's explanations, and pave the way for efficient algorithms and implementations.

Key words: Causal model, probabilistic causal model, weak cause, explanation, α-partial explanation, partial explanation, explanatory power, complexity.

⋆ This paper is an extended and revised version of a paper that appeared in: *Proceedings of the 8th International Conference on Principles of Knowledge Representation and Reasoning (KR 2002)*, pp. 49–60, Morgan Kaufmann, 2002.

* Corresponding author.
Email addresses: eiter@kr.tuwien.ac.at (Thomas Eiter), lukasiewicz@dis.uniroma1.it (Thomas Lukasiewicz).

1 Introduction

The automatic generation of explanations plays an important role in many AI areas like planning, diagnosis, natural language processing, and probabilistic inference. Notions of explanations have been studied quite extensively in the literature, see especially [28,21,45] for philosophical work, and [33,48,29] for work in AI that is related to Bayesian networks. A critical examination of such approaches from the viewpoint of explanations in probabilistic systems is given in [6].

In a recent paper [25,27], Halpern and Pearl introduced an elegant definition of causal explanation in the structural-model approach, which is based on their notions of weak and actual cause [25,26]. They showed that this notion of causal explanation models well many problematic examples in the literature. The main idea is that an explanation is a fact that is not known for certain but, if found to be true, would constitute a cause of the fact to be explained, regardless of the agent's initial uncertainty. An important note is that Halpern and Pearl's notion of causal explanation is very different from the concepts of causal explanation which have been considered in other works in AI, e.g. in [37,38,22].

Informally, the basic 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 [25,26,27], which illustrates the structural-model approach.

Example 1.1 (*Arsonists*) Suppose two arsonists lit matches in different parts of a dry forest, and both cause trees to start burning. Assume now either match by itself suffices to burn down the whole forest. We may model such a scenario in the structural-model framework 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 is 1 iff arsonist i intends to start a fire. We then have three binary variables A_1, A_2, and B, which describe the observable situation, where A_i is 1 iff arsonist i drops the match, and B is 1 iff the whole forest burns down. The causal dependencies between these variables are expressed by functions, which say that the value of A_i is given by the value of U_i, and that B is 1 iff either A_1 or A_2 is 1. These dependencies can be graphically represented as in Fig. 1.

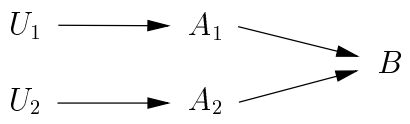

Figure 1. Causal Graph

Causes and explanations for events, such as $B = 1$ (the whole forest burns down), are defined by considering the values of variables in the above model and certain hypothetical variants (see Sections 2.2, 3.1, and 4.1). For example, arsonist 1 starting a fire is a (weak and an actual) cause of the whole forest burning down under every possible context in which arsonist 1 intends to start a fire. Moreover, arsonist 1 starting a fire is an explanation of the whole forest burning down relative to the set of all possible contexts in which either arsonist intends to start a fire. □

For more examples and extensive background on structural causal models, we refer especially to [2,20,40,41,24].

While the semantic aspects of explanations in the structural-model approach have been thoroughly studied in [25,27], a study of their computational properties is missing so far. In their papers, Halpern and Pearl were not concerned with algorithms for computing explanations, and thus the issue of how explanations can be (as efficiently as possible) computed remains to be considered. An important step towards resolving this issue is an analysis of the computational complexity of explanations. However, no complexity results for explanations, apart from trivial intractability results which are inherited from Boolean functions, were known, and a characterization of the complexity of explanations was open.

In this paper, we aim at filling this gap by giving a precise account of the complexity of explanations in structural causal models. It continues and extends the work in [14,18] on the complexity of actual and weak causes, which are a stepping stone for defining explanations. As for computation in the structural-model approach, Hopkins [30] recently explored search-based strategies for computing actual causes (i.e., minimal weak causes) in both the general and restricted settings. However, he did not pay much attention to complexity issues, and did not provide a detailed analysis of the intrinsic complexity of actual causes, nor did he address the computation of explanations on top of weak causes.

The main contributions of this paper can be summarized as follows (a review of the mentioned complexity classes is given in Section 2.3):

- We determine the complexity of (full) explanations in the structural-model approach [25,27]. We consider the problems of recognizing explanations and of deciding whether an explanation over certain variables exists. As it turns out, these problems are complete for $\mathrm{D}_2^{\mathrm{P}}$ and Σ_3^P, respectively, in the unrestricted case, and complete for D^{P} and Σ_2^P, respectively, in the binary case. Thus, recognition and existence of explanations reside, loosely speaking, at the second and the third level of the well-known Polynomial Hierarchy.
- We then determine the complexity of partial explanations in the structural-model approach [25,27], which relax full explanations in a probabilistic setting. We consider the problems of recognizing α-partial / partial explanations, of deciding whether an α-partial explanation over certain variables exists, and of comput-

ing the explanatory power of partial explanations. These problems turn out to be complete for $\mathrm{P}_{\|}^{\Sigma_2^P}$, Σ_3^P, and $\mathrm{FP}_{\|}^{\Sigma_2^P}$, respectively, in the unrestricted case, and complete for $\mathrm{P}_{\|}^{\mathrm{NP}}$, Σ_2^P, and $\mathrm{FP}_{\|}^{\mathrm{NP}}$, respectively, in the binary case.

- Furthermore, we analyze the complexity of explanations and partial explanations in a setting where context sets are succinctly represented. In the standard setting, the contexts $u_1, u_2, \ldots, u_n$ which ought to be respected for forming an explanation are simply enumerated in the problem input. In another (natural) representation, the contexts are given by a membership function $\chi(u)$, which on input of a context u tells whether u ought to be respected or not. This form of representation is more succinct than simple context enumeration in general, and may lead to exponential savings in storage for the context set of interest. However, this is traded for a significant increase in the complexity of explanations. More precisely, we show that recognizing explanations and partial explanations is complete for Π_4^P in the unrestricted case, and complete for Π_3^P in the binary case.
- Finally, we analyze the complexity of explanations in the generalization of contexts to situations, which are pairs (M, u) of a causal model M and a context u [25,27]; here, also uncertainty about the causal model, and not only about the context which applies to the actual scenario can be modeled. We consider the problems of recognizing explanations and deciding explanation existence. We find that for the recognition problem, moving from contexts to situations results in a complexity increase; as we show, this problem is Π_3^P-complete both in the unrestricted and the binary case. For the existence problem, no complexity increase happens in general, i.e., the problem remains Σ_3^P-complete, but for the binary case, in which the problem becomes Σ_3^P-complete.
- In our analysis of explanations for situations, we encounter and resolve problems on structural causal models which are interesting in their own right. Namely, we consider the problems of subsumption and equivalence between causal models M_1 and M_2 modulo the language of causal formulas [25,27]. That is, given M_1 and M_2, is it true that each causal formula ϕ which holds on M_1 also holds on M_2 (denoted $M_1 \leq M_2$), respectively that M_1 and M_2 model the same set of causal formulas (denoted $M_1 \equiv M_2$), and thus are indistinguishable in the language of causal formulas. As we show, both deciding $M_1 \leq M_2$ and deciding $M_1 \equiv M_2$ is Π_3^P-complete, in the unrestricted and, noticeably, also in the binary case. Both membership in Π_3^P and hardness for Π_3^P are not immediate, and require suitable auxiliary results which help to distinguish causal models.

Our results in the present paper draw a precise picture of the complexity of explanations in the structural-model approach, and are valuable and important in several respects:

- First and foremost, they provide a handle in understanding the computational nature of explanations and the intrinsic difficulties which are at the heart of their

computation. They must be reflected somehow in the worst-case behavior of "optimal" algorithms solving the problem. In this way, our results contribute in paving the way for efficient algorithms and for implementations of explanations in the structural-model approach.

- Second, the insight into sources of complexity which make the problems intractable provides a starting point for identifying cases of lower complexity, and in particular of tractable cases. While we do not purse this issue here, results on this can be found in [15,16].
- Third, the results are useful in comparing Halpern and Pearl's notion of causal explanation with other notions of explanations (e.g., abductive explanations [35], [47,12] and maximum a posteriori explanations, alias most probable explanations in Bayesian networks [33,34]), and allow to assess the existence of efficient mappings between different frameworks for generating explanations.

The rest of this paper is organized as follows. Section 2 provides some preliminaries on structure-based causal models, the notion of weak cause, and the complexity classes that we encounter in this paper. In Section 3, we analyze the complexity of full explanations in the structural-model approach. Section 4 concentrates on the complexity of partial explanations. In Section 5, we then analyze the complexity of explanations in the case of succinctly represented context sets. Section 6 deals with the complexity of explanations and of related problems in the general case of situations. In Section 7, we discuss related work on other frameworks of explanations, and compare our results to complexity results for them. Section 8 gives a discussion of the results, in particular of implications for algorithms, and provides some concluding remarks, including an outlook on future research issues.

While several of the results are intuitive, their proofs (in particular, the hardness parts) are nontrivial and technically quite involved. Thus, in order not to distract from the flow of reading, some technical details are moved to Appendices A–D.

2 Preliminaries

In this section, we give some technical preliminaries. We first recall structure-based causal models and the notion of weak cause by Halpern and Pearl [25,26]. We then describe the complexity classes that appear in our results.

2.1 Causal Models

We start with recalling structure-based causal models; for a rich background, see especially [2,20,40,41,24]. 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 finite set of *random variables*. Capital letters U, V, W, etc. denote variables and sets of variables. Each variable X_i may take on *values* from a nonempty finite *domain* $D(X_i)$. A *value* for a set of variables $X = \{X_1, \ldots, X_n\}$ is a mapping $x : X \to D(X_1) \cup \cdots \cup D(X_n)$ 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 $x \in D(X)$, denote by $x|Y$ the restriction of x to Y. For disjoint sets of variables X, Y and values $x \in D(X), y \in D(Y)$, denote by xy the union of x and y. As usual, we often identify singletons $\{X_i\}$ with X_i and their values x with $x(X_i)$. We often identify the values 0 and 1 with the classical truth values **false** and **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 $F_X : D(PA_X) \to D(X)$ 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 $X, Y \in U \cup V$. A causal model $M = (U, V, F)$ is *binary* iff $|D(X)| = 2$ for all $X \in V$.

We focus here on the principal class of *recursive* causal models $M = (U, V, F)$; as argued in [25], we do not lose much generality by concentrating on recursive causal models. A causal model $M = (U, V, F)$ is *recursive*, if its causal graph is a directed acyclic graph. 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 endogenous

variables $X \subseteq V$, and value $x \in D(X)$, the causal model $M_{X=x} = (U, V \setminus X, F_{X=x})$, where $F_{X=x} = \{F'_Y \mid Y \in V \setminus X\}$ and each F'_Y is obtained from F_Y by setting X to 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 a set of endogenous variables $Y \subseteq V$, we write $Y_x(u)$ to abbreviate $Y_{M_x}(u)$.

As for computation, we assume that in causal models $M = (U, V, F)$, where $F = \{F_X \mid X \in V\}$, every function $F_X : D(PA_X) \to D(X)$ with $X \in V$ is computable in polynomial time. The following proposition is then immediate.

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

2.2 Weak Causes

We now recall the notion of weak cause from [25,26]. 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 [1] and y is a value for Y. The set of *events* is the closure of the set of primitive events under the Boolean operations $\neg$ and $\wedge$ (that is, every primitive event is an event, and if ϕ and ψ are events, then also $\neg\phi$ and $\phi \wedge \psi$).

The *truth* of an event ϕ in a causal model $M = (U, V, F)$ under 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 $\vee$ and $\rightarrow$ are defined as usual, i.e., $\phi \vee \psi$ and $\phi \rightarrow \psi$ stand for $\neg(\neg\phi \wedge \neg\psi)$ and $\neg\phi \vee \psi$, respectively. We write $\phi(u)$ to abbreviate $(M, u) \models \phi$. For $X \subseteq V$ and $x \in D(X)$, we use $\phi_x(u)$ as an abbreviation of $(M_x, u) \models \phi$. For $X = \{X_1, \ldots, X_k\} \subseteq V$ with $k \geq 1$ and $x_i \in D(X_i)$, we use $X = x_1 \cdots x_k$ to abbreviate $X_1 = x_1 \wedge \ldots \wedge X_k = x_k$.

The following result follows immediately from Proposition 2.2.

Proposition 2.3 *Let $X \subseteq V$ and $x \in D(X)$. Given $u \in D(U)$ and an event ϕ, deciding whether $\phi(u)$ and $\phi_x(u)$ (given x) hold can be done in polynomial time.*

[1] Note that [17] also admitted exogenous variables in primitive events, while [25,26] does not. This does not affect the complexity of explanations in the basic setting, but has some consequences for the generalization to situations, as discussed in Section 6.

We are now ready to recall the notion of weak cause [25,26]. Let $M = (U, V, F)$ be a causal model. 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 iff the following conditions hold:

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

AC2. Some $W \subseteq V \setminus X$ and some $\overline{x} \in D(X)$ 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 $\hat{z} = \hat{Z}(u)$.

The following example illustrates the notion of weak cause.

Example 2.4 (*Arsonists continued*) Consider the context $u_{1,1}{=}(1, 1)$ 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. Moreover, $A_1 = 1$ (resp., $A_2 = 1$) is the only weak cause of $B = 1$ under the context $u_{1,0} = (1, 0)$ (resp., $u_{0,1} = (0, 1)$) in which only arsonist 1 (resp., 2) intends to start a fire. □

The following proposition characterizes irrelevant variables in weak causes.

Proposition 2.5 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$, let ϕ be an event, and let $u \in D(U)$. Let $X_0 \in X$ such that in the causal graph for M, it holds that X_0 is not a predecessor of any variable in ϕ. Let $X' = X \setminus \{X_0\}$ and $x' = x|X'$. Then, $X = x$ is a weak cause of ϕ under u iff (i) $X_0(u) = x(X_0)$ and (ii) $X' = x'$ is a weak cause of ϕ under u.*

Proof. ($\Rightarrow$) Assume that $X = x$ is a weak cause of ϕ under u. That is, (AC1) $X(u) = x$ and $\phi(u)$ hold, and (AC2) some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, $w \in D(W)$ 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 $\hat{z} = \hat{Z}(u)$. In particular, (i) holds. Moreover, $X'(u) = x'$ and $\phi(u)$ hold. Since X_0 is no predecessor of any variable in ϕ, it also follows that (a) $\neg\phi_{\overline{x}'w'}(u)$ and (b) $\phi_{x'w'\hat{z}}(u)$ hold for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$, where $\overline{x}' = \overline{x}|X'$, $w' = wx_0$, and $x_0 = x(X_0)$. This shows that $X' = x'$ is a weak cause of ϕ under u, and thus also (ii) holds.

($\Leftarrow$) Assume that (i) and (ii) hold. Thus, $X' = x'$ is a weak cause of ϕ under u. That is, (AC1) $X'(u) = x'$ and $\phi(u)$ hold, and (AC2) some $W \subseteq V \setminus X'$, $\overline{x}' \in D(X')$, $w \in D(W)$ exist such that (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 $\hat{z} = \hat{Z}(u)$. Since also $X_0(u) = x(X_0)$, it holds that $X(u) = x$ and $\phi(u)$. Furthermore, as X_0 is no predecessor of any variable in ϕ, it follows that (a) $\neg\phi_{\overline{x}'x_0w'}(u)$ and (b) $\phi_{x'x_0w'\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$, where $w' = w|(W \setminus \{X_0\})$ and $x_0 = x(X_0)$. Hence, $X = x$ is a weak cause of ϕ under u. □

We finally recall a result from [14,18], which shows that deciding weak cause is

complete for Σ_2^P (resp., NP) in the general (resp., binary) case. Note that this result holds also when the domain $D(X) = \{1, \ldots, n_X\}$ of each variable $X \in U \cup V$ is implicitly specified by $n_X \geq 1$.

Theorem 2.6 (see [14,18]) *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 $X = x$ is a weak cause of ϕ under u is complete for Σ_2^P (resp., NP) in the general (resp., binary) case.*

2.3 Complexity Classes

We assume that the reader has some elementary background in complexity theory, and is familiar with the concepts of polynomial-time solvability, NP, polynomial-time transformations among problems, and hardness resp. completeness of a problem for a complexity class, as can be found e.g. in [31,32,39]. We now briefly recall the complexity classes that we encounter in this paper.

We recall that the Polynomial Hierarchy (PH) contains the classes $\Delta_1^{\mathrm{P}} = \mathrm{P}$, $\Sigma_1^P = \mathrm{NP}$, $\Pi_1^P = \text{co-NP}$, $\Delta_{k+1}^{\mathrm{P}} = \mathrm{P}^{\Sigma_k^P}$, $\Sigma_{k+1}^P = \mathrm{NP}^{\Sigma_k^P}$, and $\Pi_k^P = \text{co-}\Sigma_k^P$, for all $k \geq 1$. They informally model solving problems with an oracle for some other problem, which belongs to the class shown in the superscript. For our concerns, it is convenient to assume that the oracle can be called on instances of different problems from a suite of fixed problems in the class, which does not affect the computational power. Each call to the oracle costs unit time. In particular, $\Delta_2^{\mathrm{P}} = \mathrm{P}^{\mathrm{NP}}$ (resp., $\Sigma_2^P = \mathrm{NP}^{\mathrm{NP}}$) denotes the class of problems decidable in polynomial time with the help of an NP oracle on a deterministic (resp., a nondeterministic) Turing machine. Intuitively, the computational power of the classes in PH increases with each level k.

From these classes, further complexity classes have been derived. The class $\mathrm{D}_k^{\mathrm{P}} = \{L \times L' \mid L \in \Sigma_k^P, L' \in \Pi_k^P\}$, $k \geq 1$, is the "conjunction" of Σ_k^P and Π_k^P; in particular, $\mathrm{D}_1^{\mathrm{P}}$ is the familiar class D^{P}. The letter D is mnemonic for the fact that a problem in $\mathrm{D}_k^{\mathrm{P}}$ can be seen as the difference between two problems in Σ_k^P formally viewed as sets of Yes-instances. Such problems can be easily solved with two calls to a Σ_k^P oracle, but are intuitively easier since the calls can be made independently of each other. On the other hand, the problems are at least as difficult as both Σ_k^P-complete and Π_k^P-complete problems.

The class $\mathrm{P}_{\|}^{\Sigma_k^P}$, $k \geq 1$, contains the decision problems which can be solved in polynomial time with one round of *parallel* calls to a Σ_k^P oracle rather than calls in arbitrary order, and is part of the Refined PH [51]. Intuitively, constraining the use of the oracle to one parallel call of all queries restricts the computational power, since the outcome of an oracle call can not be taken into account for other oracle calls. It is known that constraining to parallel oracle queries is tantamount to re-

stricting the number of oracle calls, in arbitrary order, to $O(\log n)$ many, where n is the size of the problem input, cf. [51].

According to the current belief in complexity theory, Fig. 2 shows a strict hierarchy of inclusions.

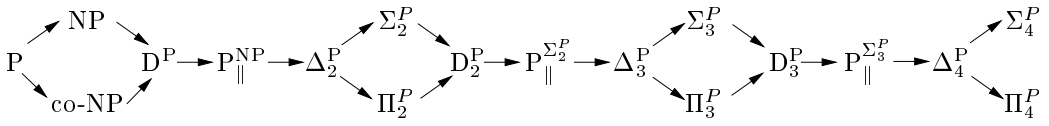

Figure 2. Containment between Complexity Classes

For classifying problems that compute an output value (e.g., the set of atoms that are entailed by a classical formula ϕ), function classes similar to the classes above have been introduced (cf. [46,31]). Among these are FP, $\mathrm{FP}_{\|}^{\mathrm{NP}} = \mathrm{FP}_{\|}^{\Sigma_1^P}$, and $\mathrm{FP}_{\|}^{\Sigma_k^P}$, which are the functional analogs of P, $\mathrm{P}_{\|}^{\mathrm{NP}} = \mathrm{P}_{\|}^{\Sigma_1^P}$, and $\mathrm{P}^{\Sigma_k^P}$, respectively. For further background on these complexity classes, we refer to [31,32,39,46,51].

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 is $\mathrm{co\text{-}NP}$-complete; deciding, given two Boolean formulas ϕ_1 and ϕ_2, whether ϕ_1 is satisfiable and ϕ_2 is unsatisfiable is complete for $\mathrm{D^P}$. More generally, a complete problem for the class Σ_k^P (resp., Π_k^P), $k \geq 1$, is to decide the validity of a given quantified Boolean formula (QBF) $Q_1 X_1 Q_2 X_2 \ldots Q_k X_k\ \phi$, where $X_1, X_2, \ldots, X_k$ are sets of variables, $Q_1 Q_2 \ldots Q_k$ is a sequence of alternating quantifiers $\exists$ and $\forall$ such that $Q_1 = \exists$ (resp., $Q_1 = \forall$), and ϕ is a Boolean formula over the variables in $X_1 \cup X_2 \cup \cdots \cup X_k$. Deciding, given two QBFs Φ_1 and Φ_2 of the form $Q_1 X_1 Q_2 X_2 \ldots Q_k X_k\ \phi$ with $Q_1 = \exists$, whether Φ_1 is valid and Φ_2 is not valid is complete for the class $\mathrm{D}_k^{\mathrm{P}}$. Finally, given l such QBFs $\Phi_1, \ldots, \Phi_l$, deciding whether the number of valid formulas among $\Phi_1, \ldots, \Phi_l$ is even (resp., computing the set of all valid formulas among $\Phi_1, \ldots, \Phi_l$) is complete for $\mathrm{P}_{\|}^{\Sigma_k^P}$ (resp., $\mathrm{FP}_{\|}^{\Sigma_k^P}$).

In this paper, unless stated otherwise, completeness for a decision class is with respect to standard polynomial-time transformations. Completeness for a function class is understood in terms of a natural generalization of polynomial time transformations: The problem P_1 reduces to P_2, if there are polynomial time functions f and g such that for each instance I_1 of P_1, the output for I_1 is given by $g(I_1, P_2(f(I_1)))$ [2]; see [46,31] for formal details. In case of P and FP, completeness is understood in terms of reductions that can be computed in logarithmic space.

[2] Note that the first argument of g allows to access the original problem instance I_1.

3 Explanations

In this section, we analyze the complexity of (full) explanations in the structural-model approach due to Halpern and Pearl [25,27]. We consider the problems of recognizing explanations and of deciding whether an explanation over certain variables exists. We consider the general as well as the restriction to the binary case.

3.1 Definitions

We now recall the concept of (full) explanation from [25,27]. Intuitively, an explanation of an observed event ϕ is a minimal conjunction of primitive events that causes ϕ even when there is uncertainty about the actual situation at hand. The agent's epistemic state is given by a set of possible contexts $u \in D(U)$, which describes all the possible scenarios for the actual situation.

More formally, let $M = (U, V, F)$ be a causal model, let $X \subseteq V$ and $x \in D(X)$, let ϕ be an event, and let $\mathcal{C} \subseteq D(U)$ be a set of contexts. Then, $X = x$ is an *explanation* of ϕ relative to $\mathcal{C}$, if the following conditions hold:

EX1. $\phi(u)$ holds for every context $u \in \mathcal{C}$.

EX2. $X = x$ is a weak cause of ϕ under every $u \in \mathcal{C}$ such that $X(u) = x$.

EX3. X is minimal. That is, for every $X' \subset X$, some $u \in \mathcal{C}$ exists such that $X'(u) = x|X'$ and $X' = x|X'$ is not a weak cause of ϕ under u.

EX4. $X(u) = x$ for some $u \in \mathcal{C}$, and $X(u') \neq x$ for some $u' \in \mathcal{C}$.

The following example illustrates the above notion of explanation.

Example 3.1 (*Arsonists continued*) Consider the set of contexts $\mathcal{C} = \{u_{1,1}, u_{1,0}, u_{0,1}\}$. Then, both $A_1 = 1$ and $A_2 = 1$ are explanations of $B = 1$ relative to $\mathcal{C}$, since (EX1) $B(u_{1,1}) = B(u_{1,0}) = B(u_{0,1}) = 1$, (EX2) $A_1 = 1$ (resp., $A_2 = 1$) is a weak cause of $B = 1$ under $u_{1,1}$ and $u_{1,0}$ (resp., $u_{1,1}$ and $u_{0,1}$), (EX3) A_1 and A_2 are obviously minimal, and (EX4) $A_1(u_{1,1}) = 1$ and $A_1(u_{0,1}) \neq 1$ (resp., $A_2(u_{1,1}) = 1$ and $A_2(u_{1,0}) \neq 1$). Furthermore, $A_1 = 1 \wedge A_2 = 1$ is not an explanation of $B = 1$ relative to $\mathcal{C}$, as here, the minimality condition EX3 is violated. □

3.2 Results

In our complexity analysis, we focus on the following problems, which are major tasks in explanation-based causal reasoning:

Explanation: Given $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, an event ϕ, and a set of contexts $\mathcal{C} \subseteq D(U)$, decide whether $X = x$ is an explanation of ϕ relative to $\mathcal{C}$.

Explanation Existence: Given $M = (U, V, F)$, $X \subseteq V$, an event ϕ, and a set of contexts $\mathcal{C} \subseteq D(U)$, decide whether some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$.

The first problem, Explanation, is the recognition of an explanation. It emerges directly from the definition of explanation in Section 3.1 and captures its intrinsic complexity. The second problem, Explanation Existence, is associated with the important task of finding an explanation for an event ϕ. Similar as in other frameworks for explanations (e.g. [35,47]), the set X focuses attention to a subset of the variables, in terms of which the explanation must be formed. Finding explanations is certainly the central task of a causal-reasoning system built for applications in practice, and thus this problem deserves special attention. We analyze the complexity of these problems for the general as well as the binary case, where M is restricted to binary causal models (i.e., each endogenous variable may take only two values).

Our complexity results on these two problems for the general and the binary case are summarized in Table 1. In detail, the problem Explanation is complete for the class $\mathrm{D}_2^{\mathrm{P}}$ (resp., D^{P}) in the general (resp., binary) case, while the problem Explanation Existence is complete for Σ_3^P (resp., Σ_2^P) in the general (resp., binary) case. It thus turns out that finding explanations is at the third level of PH. Hence, explanations are harder to compute than weak causes, which lie at the second level of PH [14]. On the other hand, recognizing explanations is only mildly harder than recognizing weak causes, which is Σ_2^P-complete.

Table 1
Complexity of Explanations

Problem	general case	binary case
Explanation	$\mathrm{D}_2^{\mathrm{P}}$-complete	D^{P}-complete
Explanation Existence	Σ_3^P-complete	Σ_2^P-complete

We now show how the complexity results in Table 1 can be formally derived. In order not to distract from the flow of reading, we present the main parts and key ideas behind constructions, and move some technical details to Appendix A.

The following result shows that deciding explanations is $\mathrm{D}_2^{\mathrm{P}}$-complete in the general case. The problem is in $\mathrm{D}_2^{\mathrm{P}}$, as condition EX2 amounts to a conjunction of a linear number of problems in Σ_2^P, and EX3 to the negation of such a problem; EX1 and EX4 are easily checked. Thus, by usual techniques, the explanation check can be reduced to a conjunction of problems in Σ_2^P and Π_2^P. Hardness for $\mathrm{D}_2^{\mathrm{P}}$ is shown by a reduction from the $\mathrm{D}_2^{\mathrm{P}}$-complete problem of deciding, given a pair (Φ_1, Φ_2) of QBFs, whether Φ_1 is valid and Φ_2 is not valid.

Theorem 3.2 *Explanation is* $\mathrm{D}_2^{\mathrm{P}}$*-complete.*

Proof. As for membership in $\mathrm{D}_2^{\mathrm{P}}$, recall that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff EX1–EX4 hold. Deciding in EX1 whether $\phi(u)$ for every $u \in \mathcal{C}$ and in EX4 whether $X(u) = x$ and $X(u') \neq x$ for some $u, u' \in \mathcal{C}$ is polynomial. In EX2, the set $\mathcal{C}'$ of all $u \in \mathcal{C}$ such that $X(u) = x$ is polynomially computable. By Theorem 2.6 and as Σ_2^P is closed under polynomially many conjunctions, deciding whether $X = x$ is a weak cause of ϕ under every $u \in \mathcal{C}'$ is in Σ_2^P. In EX3, guessing some $X' \subset X$ and checking that $X' = x|X'$ is a weak cause of ϕ under every $u \in \mathcal{C}$ such that $X'(u) = x|X'$ is in Σ_2^P. Thus, deciding EX3 is in Π_2^P. In summary, deciding whether $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ is in $\mathrm{D}_2^{\mathrm{P}}$.

Hardness for $\mathrm{D}_2^{\mathrm{P}}$ is shown by a reduction from deciding, given a pair (Φ_1, Φ_2) of QBFs $\Phi_i = \exists A_i \forall B_i\, \gamma_i$ with $i \in \{1, 2\}$, where each γ_i is a propositional formula on the variables $A_i = \{A_{i,1}, \ldots, A_{i,m_i}\}$ and $B_i = \{B_{i,1}, \ldots, B_{i,n_i}\}$, whether Φ_1 is valid and Φ_2 is not valid. We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, and ϕ as required such that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff Φ_1 is valid and Φ_2 is not valid.

Roughly speaking, the main idea behind this construction is as follows. We construct $M_1 = (U, V_1, F_1)$ and $M_2 = (U, V_2, F_2)$ and two events ϕ_1 and ϕ_2 such that (i) $V_1 \cap V_2 = \{G\}$ (that is, G is the only endogenous variable that M_1 and M_2 have in common), and (ii) for every $u \in D(U)$, it holds that $G = 0$ is a weak cause of ϕ_i under u in M_i iff Φ_i is valid (see Fig. 3, left side). The causal model M is the union of M_1 and M_2, enlarged by the additional endogenous variables G' and H (see Fig. 3, right side). We then construct ϕ and $u_1, u_2 \in D(U)$ such that ϕ is under u_1 and u_2 equivalent to ϕ_1 and ϕ_2, respectively. Finally, the construction is such that $G = 0 \wedge G' = 0$ is an explanation of ϕ relative to $\mathcal{C} = \{u_1, u_2\}$ in M, iff (a) $G = 0$ is a weak cause of ϕ_1 under u_1 in M_1, and (b) $G = 0$ is not a weak cause of ϕ_2 under u_2 in M_2, where (a) (resp., (b)) is encoded in EX2 (resp., EX3). That is, $G = 0 \wedge G' = 0$ is an explanation of ϕ relative to $\mathcal{C}$ in M, iff Φ_1 is valid and Φ_2 is not valid.

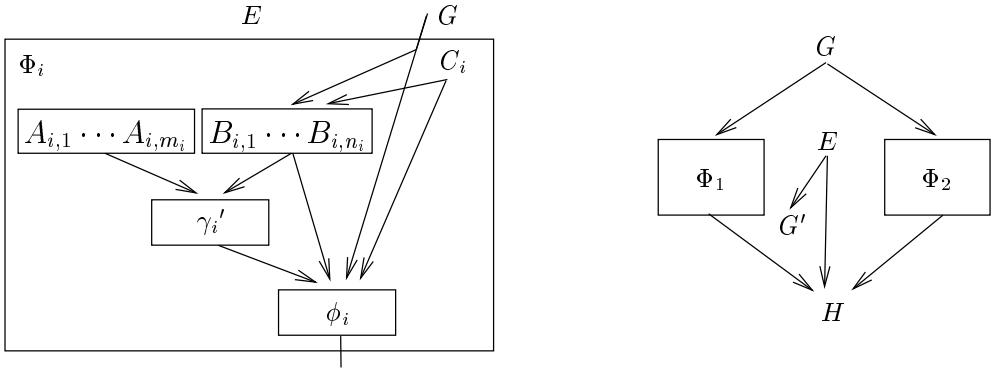

Figure 3. Schematic Construction for Evaluating two QBFs Φ_1 and Φ_2

More formally, for every $i \in \{1,2\}$, the causal model $M_i = (U, V_i, F_i)$ is defined by $U{=}\{E\}$ and $V_i{=}A_i \cup B_i \cup \{G, C_i\}$, where $D(S) = \{0,1,2\}$ for all $S \in B_i$, and $D(S) = \{0,1\}$ for all $S \in U \cup V_i \setminus B_i$. Moreover, we define

$$\phi_i = (\gamma_i' \wedge \bigwedge_{S \in B_i} S \neq 2) \vee (C_i = 0) \vee (G = 1 \wedge C_i = 1 \wedge \bigvee_{S \in B_i} S \neq 2),$$

where γ_i' is obtained from γ_i by replacing each $S \in A_i \cup B_i$ by "$S = 1$". The functions in $F_i = \{F_S^i \mid S \in V_i\}$ are defined as follows:

- $F_S^i = 0$ for all $S \in A_i \cup \{G, C_i\}$,
- $F_S^i = G + C_i$ for all $S \in B_i$.

As shown in [14,18], for every $i \in \{1,2\}$ and $u \in D(U)$, it holds that $G = 0$ is a weak cause of ϕ_i under u in M_i iff Φ_i is valid.

The causal model $M = (U, V, F)$ is now defined by $V = V_1 \cup V_2 \cup \{G', H\}$, where $D(G') = D(H) = \{0,1\}$, and $F = F_1 \cup F_2 \cup \{F_{G'} = E$, $F_H = 1$ iff $(E = 0 \wedge \phi_1) \vee (E = 1 \wedge \phi_2)$ is true$\}$. Let ϕ be defined as $H = 1$, and let $u_1, u_2 \in D(U)$ be defined by $u_1(E) = 0$ and $u_2(E) = 1$. Observe that ϕ is primitive.

For every $i \in \{1,2\}$ and $u \in D(U)$, it holds that $G = 0$ is a weak cause of ϕ_i under u in M iff Φ_i is valid. Hence, for every $i \in \{1,2\}$,

(i) $G = 0$ is a weak cause of ϕ under u_i in M iff Φ_i is valid.

By Proposition 2.5, the following statements hold:

(ii) $G = 0$ is a weak cause of ϕ under u_1 in M iff
$G = 0 \wedge G' = 0$ is a weak cause of ϕ under u_1 in M.

(iii) $G' = 0$ is not a weak cause of ϕ under u_1 in M.

Using these results, we now show that $G = 0 \wedge G' = 0$ is an explanation of ϕ relative to $\mathcal{C} = \{u_1, u_2\}$ iff Φ_1 is valid and Φ_2 is not valid.

($\Rightarrow$) Assume that $G = 0 \wedge G' = 0$ is an explanation of ϕ relative to $\mathcal{C}$. In particular, by EX2, $G = 0 \wedge G' = 0$ is a weak cause of ϕ under u_1. Moreover, by EX3, $G = 0$ is either not a weak cause of ϕ under u_1, or not a weak cause of ϕ under u_2. By (ii), $G = 0$ is a weak cause of ϕ under u_1. Thus, $G = 0$ is not a weak cause of ϕ under u_2. By (i), Φ_1 is valid, and Φ_2 is not valid.

($\Leftarrow$) Assume that Φ_1 is valid and Φ_2 is not valid. We first show that EX1 holds. As $C_i(u) = 0$ for all $i \in \{1,2\}$ and $u \in \mathcal{C}$, we get $\phi_i(u)$ for all $i \in \{1,2\}$ and $u \in \mathcal{C}$. Thus, $\phi(u)$ for all $u \in \mathcal{C}$. To see that EX4 holds, observe that $G(u_1) = G'(u_1) = 0$, while $G(u_2) = 0$ and $G'(u_2) = 1$. We next show that EX2 holds. By (i), $G = 0$ is a weak cause of ϕ under u_1. By (ii), it follows that $G = 0 \wedge G' = 0$ is a weak cause

of ϕ under u_1. We now show that EX3 holds. By (i), $G = 0$ is not a weak cause of ϕ under u_2. By (iii), $G' = 0$ is not a weak cause of ϕ under u_1. □

The following theorem shows that deciding whether an explanation over certain variables exists is Σ_3^P-complete. Here, the Σ_3^P upper bound is straightforward by the Σ_2^P upper bound of recognizing explanations, and a standard guess and check argument. The Σ_3^P-hardness of Explanation Existence stems from a subtlety in the definition of explanation. From satisfaction of EX1, EX2 and EX4 for $X = x$ we can *not* conclude that some $X' = x'$ contained in $X = x$ exists which will satisfy EX1-EX4; if we minimize $X = x$ so as to satisfy EX3, the resulting $X' = x'$ may violate EX4. It is this interplay of the conditions which makes this problem difficult, and the proofs of the hardness results nontrivial.

Theorem 3.3 *Explanation Existence is Σ_3^P-complete.*

Proof (sketch). As for membership in Σ_3^P, observe that the problem can be reduced to guessing some $X' \subseteq X$ and $x' \in D(X')$, and verifying that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$. By Theorem 3.2, this can be done in polynomial time with two calls to a Σ_2^P-oracle. Thus, the problem is in Σ_3^P.

Hardness for Σ_3^P is shown by a reduction from deciding whether a given QBF $\Phi = \exists B \forall C \exists D \gamma$ is valid, where γ is a propositional formula on the variables $B \cup C \cup D$. We construct $M = (U, V, F)$, $X \subseteq V$, $\mathcal{C} \subseteq D(U)$, and ϕ such that Φ is valid iff some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$. Roughly, the main idea is to encode the quantifier "$\exists B$" in guessing some $X' \subseteq X$, and "$\forall C\ \exists D\ \gamma$" in checking the complement of a weak cause in EX3. Note that the construction is technically involved; it is schematically shown in Fig. 4. □

In the binary case, the complexity of all considered problems drops by one level in PH; this parallels the drop of the complexity of weak causes from Σ_2^P to NP in the binary case [14]. The membership parts can be derived analogous as in the general case, and the hardness parts by slight adaptations of the constructions in the proofs, where certain subcomponents for weak cause testing are modularly replaced. The following two results show that recognizing explanations (resp., deciding the existence of explanations) is complete for $\mathrm{D^P}$ (resp., Σ_2^P) in the binary case.

Theorem 3.4 *Explanation is $\mathrm{D^P}$-complete in the binary case.*

Proof. As for membership in $\mathrm{D^P}$, recall that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff EX1–EX4 hold. By the proof of Theorem 3.2, checking EX1 and EX4 is polynomial. Moreover, in EX2, the set $\mathcal{C}'$ of all $u \in \mathcal{C}$ such that $X(u) = x$ is polynomially computable. By Theorem 2.6, deciding whether $X = x$ is a weak cause of ϕ under every $u \in \mathcal{C}'$ is in NP in the binary case. In EX3, guessing some $X' \subset X$ and checking that $X' = x|X'$ is a weak cause of ϕ under every $u \in \mathcal{C}$ with $X'(u) = x|X'$ is in NP in the binary case. Thus, the complementary problem of deciding EX3 is

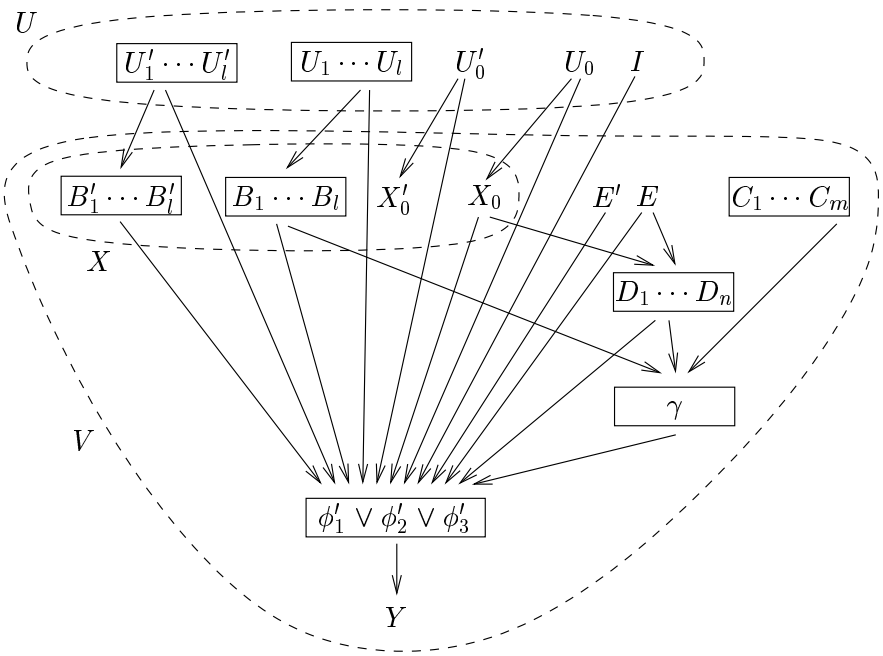

Figure 4. Schematic Construction for Evaluating a QBF $\Phi = \exists B \, \forall C \, \exists D \, \gamma$

in co-NP in the binary case. In summary, deciding whether $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ is in D^{P} in the binary case.

Hardness for D^{P} is shown by a reduction from the following D^{P}-complete problem. Given two propositional formulas in 3DNF $\alpha_1 = \alpha_{1,1} \vee \cdots \vee \alpha_{1,k_1}$ and $\alpha_2 = \alpha_{2,1} \vee \cdots \vee \alpha_{2,k_2}$ on the variables $A_1 = \{A_{1,1}, \ldots, A_{1,n_1}\}$ and $A_2 = \{A_{2,1}, \ldots, A_{2,n_2}\}$, respectively, where $k_1, k_2, n_1, n_2 \geq 1$, decide whether α_1 is not a tautology and α_2 is a tautology. Without loss of generality, $A_1 \cap A_2 = \emptyset$, and $k_1, k_2 \geq 2$.

We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, and ϕ such that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff α_1 is not a tautology and α_2 is a tautology. The construction is similar to the one in the proof of Theorem 3.2. Roughly, we replace the part for Σ_2^P-hardness of deciding weak cause in the general case by a new part for NP-hardness of deciding weak cause in the binary case.

More formally, for every $i \in \{1, 2\}$, we define the causal model $M_i = (U, V_i, F_i)$ as follows. The exogenous and endogenous variables are given by $U = \{E\}$ and $V_i = A_i \cup \{G, D_{i,1}, \ldots, D_{i,k_i-1}\}$, respectively, where $D(S) = \{0, 1\}$ for all $S \in U \cup V_i$. The functions $F_i = \{F_S^i \mid S \in V_i\}$ are defined by:

- $F_S^i = 1$ for all $S \in A_i \cup \{G\}$,
- $F_{D_{i,1}}^i = G \vee \alpha_{i,1}$,
- $F_{D_{i,j}}^i = D_{i,j-1} \vee \alpha_{i,j}$ for all $j \in \{2, \ldots, k_i - 1\}$.

Let $\phi_i = D_{i,k-1} \vee \alpha_{i,k}$. As shown in [14,18], for every $i \in \{1,2\}$ and $u \in D(U)$, it holds that $G=1$ is a weak cause of ϕ_i under u in M_i iff α_i is not a tautology.

The causal model $M=(U,V,F)$ is now defined by $V = V_1 \cup V_2 \cup \{G', H\}$ and $F = F_1 \cup F_2 \cup \{F_{G'} = E, F_H = 1$ iff $(E=0 \wedge \phi_1) \vee (E=1 \wedge \phi_2)$ is true$\}$. Let ϕ be defined as $H=1$, and let $u_1, u_2 \in D(U)$ be defined by $u_1(E)=1$ and $u_2(E)=0$. Observe that ϕ is primitive.

By a similar line of argumentation as in the proof of Theorem 3.2, it follows that $G=1 \wedge G'=1$ is an explanation of ϕ relative to $\mathcal{C} = \{u_1, u_2\}$ iff α_1 is not a tautology and α_2 is a tautology. □

Theorem 3.5 *Explanation Existence is Σ_2^P-complete in the binary case.*

Proof (sketch). As for membership in Σ_2^P, by Theorem 3.4, guessing some $X' \subseteq X$ and $x' \in D(X')$, and verifying that $X'=x'$ is an explanation of ϕ relative to $\mathcal{C}$ can be done in polynomial time with two NP-oracle calls in the binary case. This shows that Explanation Existence is in Σ_2^P in the binary case.

Hardness for Σ_2^P is shown by a reduction from the following Σ_2^P-complete problem. Given a QBF $\Phi = \exists B \, \forall C \, \gamma$, where γ is a propositional formula on the variables $B=\{B_1, \ldots, B_l\}$ and $C=\{C_1, \ldots, C_m\}$, decide whether Φ is valid. We construct $M=(U,V,F)$, $X \subseteq V$, $\mathcal{C} \subseteq D(U)$, and ϕ such that Φ is valid iff some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X'=x'$ is an explanation of ϕ relative to $\mathcal{C}$. The construction is similar to the one in the proof of Theorem 3.3. Roughly, we replace the part for Σ_2^P-hardness of deciding weak cause in the general case by a new part for NP-hardness of deciding weak cause in the binary case. □

4 Partial Explanations and Explanatory Power

In this section, we analyze the complexity of partial explanations in the structural-model approach due to Halpern and Pearl [25,27]. We consider the problems of recognizing α-partial / partial explanations and of deciding whether an α-partial explanation over certain variables exists. Furthermore, we consider the problem of computing the explanatory power of a partial explanation. All complexity results are derived for the general as well as the binary case.

4.1 Definitions

We now recall the notions of α-partial / partial explanations and of explanatory power of partial explanations [25,27]. Roughly, the main idea behind partial expla-

nations is to generalize the notion of explanation of Section 3.1 to a setting where additionally a probability distribution over the set of possible contexts is given.

Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$. Let ϕ be an event, and let $\mathcal{C} \subseteq D(U)$ be such that $\phi(u)$ for all $u \in \mathcal{C}$. We use the expression $\mathcal{C}^{\phi}_{X=x}$ to denote the unique largest subset $\mathcal{C}'$ of $\mathcal{C}$ such that $X = x$ is an explanation of ϕ relative to $\mathcal{C}'$. The following proposition shows that $\mathcal{C}^{\phi}_{X=x}$ is defined, if there exists a subset $\mathcal{C}'$ of $\mathcal{C}$ such that $X = x$ is an explanation of ϕ relative to $\mathcal{C}'$. It also gives a useful characterization of $\mathcal{C}^{\phi}_{X=x}$.

Proposition 4.1 *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 $\mathcal{C} \subseteq D(U)$ be such that $\phi(u)$ for all $u \in \mathcal{C}$. If $X = x$ is an explanation of ϕ relative to some $\mathcal{C}' \subseteq \mathcal{C}$, then $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (i) $X(u) \neq x$, or (ii) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u.*

Proof. Clearly, $\mathcal{C}^{\phi}_{X=x}$ does not contain any $u \in \mathcal{C}$ such that $X(u) = x$ and that $X = x$ is not a weak cause of ϕ under u, as otherwise EX2 would be violated. Hence, $\mathcal{C}^{\phi}_{X=x}$ is a subset of the set of all $u \in \mathcal{C}$ such that either (i) or (ii). Assume now that some $u' \in \mathcal{C}$ with $X(u') \neq x$ does not belong to $\mathcal{C}^{\phi}_{X=x}$. Then, $X = x$ is an explanation of ϕ relative to $\mathcal{C}' = \mathcal{C}^{\phi}_{X=x} \cup \{u'\}$. But this contradicts $\mathcal{C}^{\phi}_{X=x}$ being the largest such $\mathcal{C}'$. Assume next that some $u' \in \mathcal{C}$ such that $X(u') = x$ and that $X = x$ is a weak cause of ϕ under u' does not belong to $\mathcal{C}^{\phi}_{X=x}$. Then, $X = x$ is an explanation of ϕ relative to $\mathcal{C}' = \mathcal{C}^{\phi}_{X=x} \cup \{u'\}$. But this contradicts again $\mathcal{C}^{\phi}_{X=x}$ being the largest such $\mathcal{C}'$. Hence, $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (i) or (ii). □

Let P be a probability function on $\mathcal{C}$, and define

$$P(\mathcal{C}^{\phi}_{X=x} \mid X = x) \;=\; \sum_{\substack{u \in \mathcal{C}^{\phi}_{X=x} \\ X(u) = x}} P(u) \;\; / \sum_{\substack{u \in \mathcal{C} \\ X(u) = x}} P(u)\,.$$

Then, $X = x$ is called an *α-partial explanation* of ϕ relative to $(\mathcal{C}, P)$ iff $\mathcal{C}^{\phi}_{X=x}$ is defined and $P(\mathcal{C}^{\phi}_{X=x} \mid X = x) \geq \alpha$. We say $X = x$ is a *partial explanation* of ϕ relative to $(\mathcal{C}, P)$ iff $X = x$ is an α-partial explanation for some $\alpha > 0$; furthermore, $P(\mathcal{C}^{\phi}_{X=x} \mid X = x)$ is called its *explanatory power* (or *goodness*).

Example 4.2 (*Arsonists continued*) Consider the set of contexts $\mathcal{C} = \{u_{1,1}, u_{1,0}, u_{0,1}\}$, and let P be the uniform distribution over $\mathcal{C}$. Then, both $A_1 = 1$ and $A_2 = 1$ are 1-partial explanations of $B = 1$. That is, both $A_1 = 1$ and $A_2 = 1$ are partial explanations of $B = 1$ with explanatory power 1. □

As for computation, we assume that the above probability functions P are computable in polynomial time.

4.2 Results

In our analysis, we consider the following important problems related to partial explanations and their explanatory power:

α-Partial Explanation: Given $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, an event ϕ, a set of contexts $\mathcal{C} \subseteq D(U)$ such that $\phi(u)$ for all $u \in \mathcal{C}$, a probability function P on $\mathcal{C}$, and $\alpha \geq 0$, decide whether $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$.

α-Partial Explanation Existence: Given $M = (U, V, F)$, $X \subseteq V$, an event ϕ, a set of contexts $\mathcal{C} \subseteq D(U)$ such that $\phi(u)$ for all $u \in \mathcal{C}$, a probability function P on $\mathcal{C}$, and $\alpha \geq 0$, decide whether some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X' = x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$.

Partial Explanation: Given $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, an event ϕ, a set of contexts $\mathcal{C} \subseteq D(U)$ such that $\phi(u)$ for all $u \in \mathcal{C}$, a probability function P on $\mathcal{C}$, decide whether $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$.

Explanatory Power: Given $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, an event ϕ, $\mathcal{C} \subseteq D(U)$, and a probability function P on $\mathcal{C}$, where (i) $\phi(u)$ for all $u \in \mathcal{C}$, and (i) $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$, compute the explanatory power of $X = x$.

The problems α-Partial / Partial Explanation and α-Partial Explanation Existence can be viewed as relaxations of Explanation and Explanation Existence, respectively, in a probabilistic context. Explanatory Power is the problem of computing the "goodness" of a partial explanation $X = x$, given by the coverage of the cases where $X = x$ is true in the contexts $\mathcal{C}$. This information can be used to rank partial explanations and single out "best" ones.

Our complexity results on these problems for the general and the binary case are summarized in Table 2. In detail, recognizing α-partial / partial explanations is complete for $\mathrm{P}_{\|}^{\Sigma_2^P}$ (resp., $\mathrm{P}_{\|}^{\mathrm{NP}}$) in the general (resp., binary) case, while deciding the existence of α-partial explanations is complete for Σ_3^P (resp., Σ_2^P). Furthermore, computing the explanatory power of a partial explanation is complete for $\mathrm{FP}_{\|}^{\Sigma_2^P}$ (resp., $\mathrm{FP}_{\|}^{\mathrm{NP}}$) in the general (resp., binary) case. Hence, finding α-partial explanations has the same complexity as finding full explanations, while recognizing α-partial / partial explanations is mildly harder than recognizing full explanations.

The following result shows that recognizing α-partial explanations is $\mathrm{P}_{\|}^{\Sigma_2^P}$-complete. Roughly, to recognize an α-partial / partial explanation, we need to know the set of contexts $\mathcal{C}_{X=x}^{\phi}$. By exploiting the basic characterization result in Proposition 4.1, it can be computed efficiently with parallel calls to a Σ_2^P oracle. Once $\mathcal{C}_{X=x}^{\phi}$ is known, we need to check whether $X = x$ is an explanation relative to it, the rest is

Table 2
Complexity of Partial Explanations and Explanatory Power

Problem	general case	binary case
α-Partial Explanation	$\mathrm{P}_{\parallel}^{\Sigma_2^P}$-complete	$\mathrm{P}_{\parallel}^{\mathrm{NP}}$-complete
α-Partial Explanation Existence	Σ_3^P-complete	Σ_2^P-complete
Partial Explanation	$\mathrm{P}_{\parallel}^{\Sigma_2^P}$-complete	$\mathrm{P}_{\parallel}^{\mathrm{NP}}$-complete
Explanatory Power	$\mathrm{FP}_{\parallel}^{\Sigma_2^P}$-complete	$\mathrm{FP}_{\parallel}^{\mathrm{NP}}$-complete

easy. Thus, the complexity of these problems lies here in the computation of $\mathcal{C}^{\phi}_{X=x}$.

Theorem 4.3 *α-Partial Explanation is $\mathrm{P}_{\parallel}^{\Sigma_2^P}$-complete.*

Proof (sketch). We first prove membership in $\mathrm{P}_{\parallel}^{\Sigma_2^P}$. Recall that $X=x$ is an α-partial explanation of ϕ relative to $(\mathcal{C},P)$ iff (a) $X=x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$, and (b) $P(\mathcal{C}^{\phi}_{X=x} \mid X=x) \geq \alpha$. By Proposition 4.1, $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (i) $X(u) \neq x$, or (ii) $X(u)=x$ and $X=x$ is a weak cause of ϕ under u. As deciding (i) is polynomial, and deciding (ii) is in Σ_2^P, by Theorem 2.6, computing $\mathcal{C}^{\phi}_{X=x}$ is in $\mathrm{FP}_{\parallel}^{\Sigma_2^P}$. Once $\mathcal{C}^{\phi}_{X=x}$ is given, deciding (a) is possible with two Σ_2^P-oracle calls, by Theorem 3.2, and deciding (b) is polynomial. It is now well-known that two rounds of parallel Σ_2^P-oracle queries in a polynomial-time computation can be replaced by a single one [3]. Hence, the problem is in $\mathrm{P}_{\parallel}^{\Sigma_2^P}$.

Hardness for $\mathrm{P}_{\parallel}^{\Sigma_2^P}$ is shown by a reduction from deciding, given k QBFs $\Phi_i = \exists A_i \forall B_i\, \gamma_i$ with $i \in \{1,\ldots,k\}$, where each γ_i is a propositional formula on the variables $A_i = \{A_{i,1},\ldots,A_{i,m_i}\}$ and $B_i = \{B_{i,1},\ldots,B_{i,n_i}\}$, whether the number of valid formulas among $\Phi_1,\ldots,\Phi_k$ is even. Without loss of generality, $A_1 \cup B_1,\ldots,A_k \cup B_k$ are pairwise disjoint, Φ_1 is valid, and for each $j \in \{2,\ldots,k\}$, the validity of Φ_j implies the validity of Φ_{j-1} [51]. We construct $M=(U,V,F)$, $X \subseteq V$, $x \in D(X)$, ϕ, $\mathcal{C} \subseteq D(U)$, P, and α such that $X=x$ is an α-partial explanation of ϕ relative to $(\mathcal{C},P)$ iff the number of valid formulas among $\Phi_1,\ldots,\Phi_k$ is even. Roughly, the main idea behind this construction is as follows. For each Φ_i, we construct an instance of weak cause, that is, $M_i=(U_i,V_i,F_i)$, $X_i \subseteq V_i$, $x_i \in D(X_i)$, $u_i \in D(U_i)$ and an event ϕ_i, such that $X_i=x_i$ is a weak cause of ϕ_i under u_i in M_i iff Φ_i is valid. Then, M is the union of all M_i, enlarged by additional variables (see Fig. 5), and we define $X=X_1 \cup \cdots \cup X_k$ and $x=x_1 \ldots x_k$. By setting P to the uniform distribution over $\mathcal{C}$ and $\alpha = 1 \,/\, |\mathcal{C}|$, we obtain that $X=x$ is an α-partial explanation of ϕ relative to $(\mathcal{C},P)$ iff $X=x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$. The latter is made to hold iff the number of valid formulas among the Φ_i's is even. In detail, EX3 is violated iff i is even, Φ_i is not valid, and Φ_{i-1} is valid. □

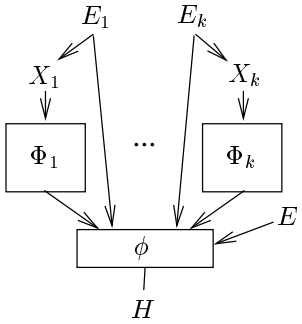

Figure 5. Schematic Construction for Evaluating k QBFs $\Phi_1, \ldots, \Phi_k$

The following theorem shows that deciding the existence of α-partial explanations is complete for Σ_3^P. Here, the Σ_3^P upper bound follows from the $\mathrm{P}_{||}^{\Sigma_2^P}$ upper bound of recognizing α-partial explanations by a standard guess and check argument. The Σ_3^P-hardness is inherited from the Σ_3^P-hardness of Explanation Existence.

Theorem 4.4 *α-Partial Explanation Existence is Σ_3^P-complete.*

Proof. We first prove membership in Σ_3^P. By Theorem 4.3, deciding whether $X'{=}x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ is in $\mathrm{P}_{||}^{\Sigma_2^P}$. Hence, guessing some $X' \subseteq X$ and $x' \in D(X')$, and deciding whether $X' = x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ is in Σ_3^P.

Hardness for Σ_3^P is shown by a reduction from Explanation Existence (see Theorem 3.3). Given an instance of it, let P be the uniform distribution on $\mathcal{C}$, and let $\alpha = 1$. Then, $X' = x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$. □

The next theorem shows that deciding partial explanations is $\mathrm{P}_{||}^{\Sigma_2^P}$-complete. The membership part is proved similarly as in the proof of Theorem 4.3. The hardness part follows easily from the hardness result in Theorem 4.3.

Theorem 4.5 *Partial Explanation is $\mathrm{P}_{||}^{\Sigma_2^P}$-complete.*

Proof. As for membership in $\mathrm{P}_{||}^{\Sigma_2^P}$, recall that $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$, and (b) $\mathcal{C}^{\phi}_{X=x}$ contains some u such that $X(u) = x$ and $P(u) > 0$. By the proof of Theorem 4.3, computing $\mathcal{C}^{\phi}_{X=x}$ is in $\mathrm{FP}_{||}^{\Sigma_2^P}$. Once $\mathcal{C}^{\phi}_{X=x}$ is given, checking (a) is in $\mathrm{D}_2^{\mathrm{P}}$ by Theorem 3.2, and checking (b) is polynomial. As two rounds of parallel Σ_2^P-oracle queries in a polynomial-time computation can be replaced by a single one [3], deciding whether $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ is in $\mathrm{P}_{||}^{\Sigma_2^P}$.

We next show $\mathrm{P}_{\|}^{\Sigma_2^P}$-hardness. If P is the uniform distribution over $\mathcal{C}$, then $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff $X = x$ is a $\frac{1}{|\mathcal{C}|}$-partial explanation of ϕ relative to $(\mathcal{C}, P)$. By the proof of Theorem 4.3, deciding the latter is complete for $\mathrm{P}_{\|}^{\Sigma_2^P}$. Thus, deciding whether $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ is $\mathrm{P}_{\|}^{\Sigma_2^P}$-hard, and hardness holds even if P is the uniform distribution over $\mathcal{C}$. □

The following result shows that computing the explanatory power of a partial explanation is $\mathrm{FP}_{\|}^{\Sigma_2^P}$-complete. Here, the membership part is proved similarly as in the proof of Theorem 4.3. The hardness part is shown by a reduction from computing all valid QBFs among k given QBFs $\Phi = \exists A \forall B \gamma$.

Theorem 4.6 *Explanatory Power is* $\mathrm{FP}_{\|}^{\Sigma_2^P}$*-complete.*

Proof (sketch). We first prove membership in $\mathrm{FP}_{\|}^{\Sigma_2^P}$. Let $X = x$ be a partial explanation of ϕ relative to $(\mathcal{C}, P)$. To compute its explanatory power, we compute first $\mathcal{C}_{X=x}^{\phi}$ and then $P(\mathcal{C}_{X=x}^{\phi} \mid X = x)$. By the proof of Theorem 4.3, the former is in $\mathrm{FP}_{\|}^{\Sigma_2^P}$, while the latter is polynomial. In summary, the problem is in $\mathrm{FP}_{\|}^{\Sigma_2^P}$.

Hardness for $\mathrm{FP}_{\|}^{\Sigma_2^P}$ is shown by a reduction from computing, given k QBFs $\Phi_i = \exists A_i \forall B_i \gamma_i$ with $i \in \{1, \ldots, k\}$, where each γ_i is a propositional formula on the variables $A_i = \{A_{i,1}, \ldots, A_{i,m_i}\}$ and $B_i = \{B_{i,1}, \ldots, B_{i,n_i}\}$, the vector $(v_1, \ldots, v_k) \in \{0,1\}^k$ such that $v_i = 1$ iff Φ_i is valid, for all $i \in \{1, \ldots, k\}$. Without loss of generality, $A_1 \cup B_1, \ldots, A_k \cup B_k$ are pairwise disjoint, and Φ_1 is valid. Roughly speaking, the main idea is to construct a problem instance such that $(v_1, \ldots, v_k)$ is the bit-vector representation of the explanatory power of $X = x$. The construction is similar to the one in the proof of Theorem 4.3, which is illustrated in Fig. 5. For each Φ_i, we construct $M_i = (U_i, V_i, F_i)$, $X_i \subseteq V_i$, $x_i \in D(X_i)$, $u_i \in D(U_i)$, and an event ϕ_i such that $X_i = x_i$ is a weak cause of ϕ_i under u_i in M_i iff Φ_i is valid. These models are then combined in M such that $u_i \in \mathcal{C}_{X=x}^{\phi}$ iff Φ_i is valid. Defining $P(u_i) = 2^{i-1} \,/\, (2^k - 1)$ for all $i \in \{1, \ldots, k\}$ completes the reduction. □

5 Succinct Representation

Our complexity results in Sections 3 and 4 (as summarized in Tables 1 and 2) assume that the set of contexts $\mathcal{C}$ is enumerated in the input. However, $\mathcal{C}$ may contain exponentially many contexts. Hence, a descriptive representation can be much more compact and desirable in practice. In the *succinct representation* setting, we thus assume that $\mathcal{C}$ is given by a tractable membership function $\chi_{\mathcal{C}}(u)$. That is, on input of $u \in D(U)$, function $\chi_{\mathcal{C}}(u)$ reports in polynomial time whether $u \in \mathcal{C}$ holds. This

includes, e.g., descriptions of $\mathcal{C}$ in terms of propositional formulas β over U such that the models of β describe the contexts in $\mathcal{C}$.

Table 3 shows our complexity results for some of the problems in Sections 3 and 4 in the setting where contexts are succinctly represented. More precisely, recognizing explanations and partial explanations in the case of succinct context sets is complete for Π_4^P (resp., Π_3^P) in the general (resp., binary) case.

Table 3
Complexity of Explanations and Partial Explanations: Succinct Representation

Problem	general case	binary case
Explanation	Π_4^P-complete	Π_3^P-complete
Partial Explanation	Π_4^P-complete	Π_3^P-complete

Thus, it turns out that succinct representation increases the complexity of Explanation and Partial Explanation drastically. Intuitively, in this case checking a property for all contexts in $\mathcal{C}$ becomes much harder, since there seems no better way than guessing the "right" context witnessing or disproving the property. The complexity increase by two levels in PH stems from the fact that condition EX3 involves two nested checks of properties for all contexts in $\mathcal{C}$. This dominates the complexity of EX1, EX2, and EX4 and leads to Π_4^P complexity.

For α-Partial Explanation, we have similar effects. Worse, we need to calculate sums of probabilities over succinctly represented context sets. This leads us outside PH: It requires us to solve problems which are at least as hard as deciding whether a given propositional CNF β has $\geq k$ models, where k is in the input. This problem is, as generally believed, not in PH. We refrain from a detailed analysis of computing α-partial explanations here. A complexity increase for Explanation Existence under succinct context sets to Σ_5^P is plausible, though we have not analyzed it; note that already the Π_4^P-hardness proof for Explanation is rather involved.

The following result shows that deciding explanation is Π_4^P-complete for succinct context sets. Here, membership in Π_4^P follows from the fact that checking EX1, EX2, EX3, and EX4 is in $\mathrm{co\text{-}NP}$, Π_3^P, Π_4^P, and NP, respectively, for succinct context sets. Hardness for Π_4^P is shown by a reduction from deciding whether a given QBF $\Phi = \forall A \exists B \forall C \exists D \gamma$ is valid, which is essentially encoded in condition EX3.

Theorem 5.1 *Explanation is Π_4^P-complete for succinct context sets.*

Proof (sketch). Recall that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff EX1–EX4 hold. Under succinct context sets, in EX1, deciding $\phi(u)$ for all $u \in \mathcal{C}$ is in $\mathrm{co\text{-}NP}$. In EX4, deciding whether $X(u) = x$ and $X(u') \neq x$ hold for some $u, u' \in \mathcal{C}$ is in NP. By Theorem 2.6, deciding whether $X = x$ is a weak cause of ϕ under every

$u \in \mathcal{C}$ with $X(u) = x$ in EX2 is in Π_3^P. Thus, deciding whether some $X' \subset X$ exists such that $X' = x|X'$ is a weak cause of ϕ under every $u \in \mathcal{C}$ with $X'(u) = x|X'$ is in Σ_4^P. That is, deciding EX3 is in Π_4^P. In summary, deciding whether EX1–EX4 hold is in Π_4^P under succinct context sets.

Hardness for Π_4^P is shown by a reduction from deciding whether a given QBF $\Phi = \forall A \exists B \forall C \exists D \gamma$ is valid, where γ is a propositional formula on the variables $A \cup B \cup C \cup D$. We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, ϕ, and $\mathcal{C} \subseteq D(U)$ such that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff Φ is valid. This construction is schematically illustrated in Fig. 6. Roughly, the main idea is to encode Φ in EX3, where the quantifier "$\forall A$" is represented by considering all $X' \subset X$, the quantifier "$\exists B$" is expressed by finding some $u \in D(U)$, and $\forall C\ \exists D\ \gamma$ is expressed by checking the complement of a weak cause. □

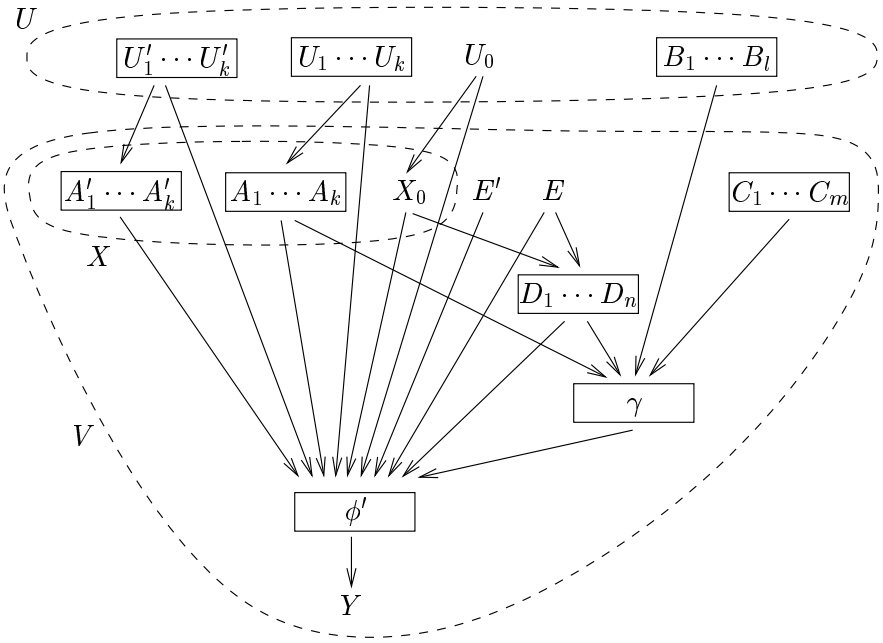

Figure 6. Schematic Construction for Evaluating a QBF $\Phi = \forall A\ \exists B\ \forall C\ \exists D\ \gamma$

The next result shows that under succinct context sets, also deciding partial explanation is Π_4^P-complete. Here, membership in Π_4^P can be proved similarly as in the proof of Theorem 5.1, using additionally Proposition 4.1. Hardness for Π_4^P easily follows from an extension of the hardness part in the proof of Theorem 5.1, where we additionally assume the uniform distribution P on the set of contexts.

Theorem 5.2 *Partial Explanation is Π_4^P-complete for succinct context sets.*

Proof. As for membership in Π_4^P, recall that $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}^\phi_{X=x}$, and (b) $\mathcal{C}^\phi_{X=x}$ contains some u such that $X(u) = x$ and $P(u) > 0$. By Proposition 4.1, $\mathcal{C}^\phi_{X=x}$ is the

set of all $u \in \mathcal{C}$ such that either (i) $X(u) \neq x$, or (ii) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. To check that (a) holds, we check that EX1–EX4 hold. Clearly, EX1 and EX2 always hold. The complement of EX3 says that some $X' \subset X$ exists such that for every $u \in \mathcal{C}$ it holds that $X'(u) = x|X'$ and $u \in \mathcal{C}^{\phi}_{X=x}$ implies that $X' = x|X'$ is a weak cause of ϕ under u. That is, some $X' \subset X$ exists such that for every $u \in \mathcal{C}$, it holds either (a) $X'(u) \neq x|X'$, or (b) $X(u) = x$ and $X = x$ is not a weak cause of ϕ under u, or (c) $X' = x|X'$ is a weak cause of ϕ under u. As deciding whether $X = x$ (resp., $X' = x|X'$) is a weak cause of ϕ under u is in Σ^P_2, deciding whether EX3 does not hold is in Σ^P_4. That is, deciding whether EX3 holds is in Π^P_4. EX4 says that some $u, u' \in \mathcal{C}^{\phi}_{X=x}$ exist such that $X(u) \neq x$ and $X(u') = x$. Equivalently, some $u, u' \in \mathcal{C}$ exist such that $X(u) \neq x$, and $X(u') = x$ and $X = x$ is a weak cause of ϕ under u'. Thus, deciding whether EX4 holds is in Σ^P_2. In summary, checking (a) is in Π^P_4. Finally, (b) says that some $u \in \mathcal{C}$ exists such that $X(u) = x$, $P(u) > 0$, and $X = x$ is a weak cause of ϕ under u. Thus, checking (b) is in Σ^P_2. In summary, deciding whether (a) and (b) hold is in Π^P_4.

Hardness for Π^P_4 is shown by a reduction from the Π^P_4-complete problem of deciding whether a QBF $\Phi = \forall A \exists B \forall C \exists D \gamma$ is valid, where γ is a propositional formula on the variables $A \cup B \cup C \cup D$.

Let $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, ϕ, and $\mathcal{C} \subseteq D(U)$ be defined as in the proof of Theorem 5.1, and let P be the uniform distribution over $\mathcal{C}$. By the proof of Theorem 5.1, $(*)$ $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff Φ is valid. Furthermore, ϕ is primitive, $\phi(u)$ for all $u \in \mathcal{C}$, and for every $u \in \mathcal{C}$, either (i) $X(u) \neq x$, or (ii) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u.

By Proposition 4.1, $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}$, and (b) $\mathcal{C}$ contains some u such that $X(u) = x$ and $P(u) > 0$. Here, (a) implies (b). By $(*)$, it follows that $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff Φ is valid. □

6 Generalization: Situations

In this section, we analyze the complexity of recognizing explanations and of deciding the existence of explanations in the general case of situations [25,27]. In the course of this, we also analyze the complexity of checking subsumption and equivalence between causal models.

We now recall the concept of explanation for the case of situations [25,27]. Intuitively, an agent may also be uncertain about the causal model, and not only about the context that applies to the actual situation at hand. Thus, in the general case of situations, the agent's epistemic state consists of a set of pairs (M, u), called *situations*, where M is a causal model and u is a context. Before defining explanations for situations, we first define causal formulas and their truth and validity.

A *basic causal formula* is an expression of the form $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\,\phi$, where ϕ is an event, $Y_1, \ldots, Y_k$ are pairwise distinct endogenous variables, $y_i \in D(Y_i)$ 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 operations $\neg$ and $\wedge$. For $Y = \{Y_1, \ldots, Y_k\}$ and $y = y_1 \ldots y_k$, we use $[Y \leftarrow y]\,\phi$ to abbreviate $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\,\phi$. As usual, we use $\phi \vee \psi$ and $\top$ to abbreviate $\neg(\neg\phi \wedge \neg\psi)$ and $\phi \vee \neg\phi$, respectively. The *truth* of a causal formula ψ in $M = (U, V, F)$ under $u \in D(U)$, denoted $(M, u) \models \psi$, is inductively defined by:

- $(M, u) \models [Y \leftarrow y]\,\phi$ iff $\phi_y(u)$ 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 $(M, u) \models \psi$.

We say ψ is *valid* in $M = (U, V, F)$, denoted $M \models \psi$, if $(M, u) \models \psi$ for all $u \in D(U)$. By $Th(M)$ we denote the set of all causal formulas which are valid in M.

The following result, whose easy proof is omitted, shows that deciding validity is co-NP-complete. Roughly, this result is immediate by the fact that checking $M \models \psi$ amounts to checking $(M, u) \models \psi$ for each of the in general exponentially many contexts u in $D(U)$.

Proposition 6.1 *Given a causal model $M = (U, V, F)$ and a causal formula ψ, deciding whether $M \models \psi$ is* co-NP*-complete.*

We are now ready to define situations, and explanations relative to situations as follows. A *situation* $S = (M, u)$ consists of a causal model $M = (U, V, F)$ and a context $u \in D(U)$. Informally, rather than having explanations of the form $X = x$ relative to a set of contexts $\mathcal{C}$, where X is a set of endogenous variables and $x \in D(X)$, we now generalize to explanations of the form $(\psi, X = x)$ relative to a set of situations $\mathcal{S}$, where ψ is a causal formula that restricts the causal models to be considered from $\mathcal{S}$.

Before we give a formal definition, we introduce some useful notation. Let for any set of situations $\mathcal{S}$ and causal formulas ψ and ψ' denote $\psi \models_{\mathcal{S}} \psi'$ that $M \models \psi$

implies $M \models \psi'$, for all $(M,u) \in \mathcal{S}$, and let $\psi \equiv_{\mathcal{S}} \psi'$ denote $\psi \models_{\mathcal{S}} \psi' \wedge \psi' \models_{\mathcal{S}} \psi$, i.e., equivalence of ψ and ψ' on the causal models occurring in $\mathcal{S}$.

Let then ψ be a causal formula, let X be a set of endogenous variables, and let $x \in D(X)$. Furthermore, let ϕ be an event, and let $\mathcal{S}$ be a set of situations. Then, $(\psi, X = x)$ is an *explanation of ϕ relative to $\mathcal{S}$*, if the following conditions hold:

ES1. $(M,u) \models \phi$ for every situation $(M,u) \in \mathcal{S}$.

ES2. $X = x$ is a weak cause of ϕ under u in M, for every $(M,u) \in \mathcal{S}$ such that $(M,u) \models X = x$ and $M \models \psi$.

ES3. $(\psi, X = x)$ is minimal. That is, there is no $(\psi', X' = x') \not\approx_{\mathcal{S}} (\psi, X = x)$ satisfying ES2 such that (i) $\psi \models_{\mathcal{S}} \psi'$ and (ii) $X' \subseteq X$ and $x' = x|X'$.

ES4. $(M,u) \models X = x$ for some $(M,u) \in \mathcal{S}$, and $(M',u') \models \neg(X = x)$ for some $(M',u') \in \mathcal{S}$.

In ES3, $(\psi', X' = x') \not\approx_{\mathcal{S}} (\psi, X = x)$ means that either $\psi \not\equiv_{\mathcal{S}} \psi'$, i.e., ψ' and ψ are not equivalent on the causal models in $\mathcal{S}$, or that $X' = x'$ and $X = x$ are different.

Observe that the notion of explanation for sets of contexts is a special case of the notion of explanation for sets of situations, as $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ in M iff $(\top, X = x)$ is an explanation of ϕ relative to $\{(M,u) \mid u \in \mathcal{C}\}$.

The following example illustrates explanations relative to situations.

Example 6.2 (*Arsonists continued*) Consider the causal model $M = (U, V, F)$ of the running example given in Example 2.1. Let the causal model $M' = (U, V, F')$ be identical to M except that the function $F'_B \in F'$ is now defined by $F'_B = 1$ iff $A_1 = 1$ and $A_2 = 1$. Informally, this says that the whole forest burns down iff both arsonists start a fire. Let the causal model $M'' = (U'', V, F'')$ be identical to M except that now the set of exogenous variables is given by $U'' = \{U_1, U_2, U_3\}$ and the function $F''_B \in F''$ is defined by $F''_B = 1$ iff $A_1 = 1$ or $A_2 = 1$ or $U_3 = 1$. Intuitively, the whole forest burns down iff either some arsonist starts a fire or some other event results into a fire (which is expressed by the exogenous variable U_3). Let the causal formula ψ be given by $[A_1 \leftarrow 0,\ A_2 \leftarrow 0]\, B = 0$ (intuitively, the whole forest burns down only if one of the two arsonists starts a fire), and let the context $u_{1,1,1} \in D(U'')$ be defined by $u_{1,1,1}(U_i) = 1$ for all $i \in \{1,2,3\}$ (intuitively, both arsonists intend to start a fire, and there is also some other event that results into a fire).

Then, both $(\psi, A_1 = 1)$ and $(\psi, A_2 = 1)$ are explanations of $B = 1$ relative to the set of situations $\mathcal{S} = \{(M, u_{1,1}), (M, u_{0,1}), (M, u_{1,0}), (M', u_{1,1}), (M'', u_{1,1,1})\}$, since (ES1) $S \models B = 1$ for all $S \in \mathcal{S}$, (ES2) $A_1 = 1$ (resp., $A_2 = 1$) is a weak cause of $B = 1$ relative to every $S \in \{(M, u_{1,1}), (M, u_{1,0}), (M', u_{1,1})\}$ (resp., $S \in \{(M, u_{1,1}), (M, u_{0,1}), (M', u_{1,1})\}$), (ES3) $A_1 = 1$ (resp., $A_2 = 1$) is trivially minimal, and ψ also cannot be weakened to some ψ' such that $M'' \models \psi'$, since $A_1 = 1$ (resp., $A_2 = 1$) is *not* a weak cause of $B = 1$ relative to $(M'', u_{1,1,1})$, and (ES4) $A_1(u_{1,1}) = 1$ and

$A_1(u_{0,1}) \neq 1$ (resp., $A_2(u_{1,1}) = 1$ and $A_2(u_{1,0}) \neq 1$) in M. $\Box$

We next define the concepts of subsumption and equivalence between causal models. We say that a causal model $M = (U, V, F)$ *subsumes* a collection of causal models $M_1, M_2, \ldots, M_n$, where $M_i = (U_i, V_i, F_i)$ with $V = V_i$, $i \in \{1, \ldots, n\}$, denoted $M_1, M_2, \ldots, M_n \leq M$, iff for all causal formulas ϕ on the variables in V, it holds that $M_i \models \phi$, for all $i \in \{1, \ldots, n\}$, implies $M \models \phi$, that is, $\bigcap_{i=1}^{n} Th(M_i) \subseteq Th(M)$. Two causal models $M_1 = (U_1, V_1, F_1)$ and $M_2 = (U_2, V_2, F_2)$, where $V_1 = V_2$, are *equivalent*, denoted $M_1 \equiv M_2$, iff $M_1 \leq M_2$ and $M_2 \leq M_1$. That is, M_1 and M_2 are equivalent iff $Th(M_1) = Th(M_2)$. In other words, M_1 and M_2 are indiscernible in the language of causal formulas.

The following result provides a characterization of the failure of subsumption of a collection of causal models by some causal model. This characterization is particularly useful for assessing the computational complexity of deciding this relationship as well as of deciding equivalence of causal models.

Theorem 6.3 *Let $M = (U, V, F)$ and $M_i = (U_i, V, F_i)$, $1 \leq i \leq n$, be causal models. Then, $M_1, M_2, \ldots, M_n \not\leq M$ iff the following property holds:*

$(*)$ *There exists some $u \in D(U)$ such that for every $i \in \{1, \ldots, n\}$ and for every $u_i \in D(U_i)$, there exists some causal formula $[Y \leftarrow y]\, X = x$, where Y is a (possibly empty) set of endogenous variables and X is a single variable, such that (i) $(M, u) \not\models [Y \leftarrow y]\, X = x$ and (ii) $(M_i, u_i) \models [Y \leftarrow y]\, X = x$.*

Proof. ($\Rightarrow$) Suppose $M_1, M_2, \ldots, M_n \not\leq M$, that is, $T = \bigcap_{i=1}^{n} Th(M_i) \not\subseteq Th(M)$. Let $\phi \in T \setminus Th(M)$ be an arbitrary formula. As $\phi \notin Th(M)$, there exists some context $u \in D(U)$ such that $(M, u) \not\models \phi$, while $(M_i, u_i) \models \phi$ for all $i \in \{1, \ldots, n\}$ and $u_i \in D(U_i)$. As easily seen, for all recursive causal models $M' = (U', V', F')$ and $u' \in D(U')$, the following holds (cf. also [24]):

- $(M', u') \models [Y \leftarrow y]\, \neg\psi$ iff $(M', u') \models \neg[Y \leftarrow y]\, \psi$;
- $(M', u') \models [Y \leftarrow y]\, (\psi_1 \wedge \psi_2)$ iff $(M', u') \models [Y \leftarrow y]\, \psi_1 \wedge [Y \leftarrow y]\, \psi_2$.

Therefore, ϕ is equivalent to a Boolean combination of causal formulas of the form $[Y' \leftarrow y']\, X' = x'$, where Y' is a (possibly empty) set of endogenous variables and X' is a single variable. Moreover, as the domain of every variable is finite, we can equivalently rewrite ϕ into a disjunctive normal form

$$\bigvee_{j \in J} \Big(\bigwedge_{k \in K_j} [Y_{j,k} \leftarrow y_{j,k}]\, X_{j,k} = x_{j,k} \Big),$$

where each $X_{j,k}$ is a single variable. Since $(M_i, u_i) \models \phi$, it follows that $(M_i, u_i) \models [Y_{j,k} \leftarrow y_{j,k}]\, X_{j,k} = x_{j,k}$ for some $j = j_0$ and all $k \in K_{j_0}$; on the other hand, since $(M, u) \not\models \phi$, some $k_0 \in K_{j_0}$ exists such that $(M, u) \not\models [Y_{j_0,k_0} \leftarrow y_{j_0,k_0}]\, X_{j_0,k_0} = x_{j_0,k_0}$. As $(M_i, u_i) \models [Y_{j_0,k_0} \leftarrow y_{j_0,k_0}]\, X_{j_0,k_0} = x_{j_0,k_0}$, this proves property $(*)$.

($\Leftarrow$) Suppose that $(*)$ holds. Let ϕ be the disjunction of all formulas $[Y \leftarrow y]\, X = x$ for all $i \in \{1, \ldots, n\}$ and u_i as in $(*)$. Then, $(M_i, u_i) \models \phi$ for all $i \in \{1, \ldots, n\}$ and $u_i \in D(U_i)$, while $(M, u) \not\models \phi$ by construction. This shows that $\bigcap_{i=1}^{n} Th(M_i) \not\subseteq Th(M)$, that is, $M_1, M_2, \ldots, M_n \not\leq M$. □

We remark that a similar result would hold for causal models with arbitrary (finite and/or infinite variable domains), if also causal formulas $[Y \leftarrow y]\, X \neq x$, where $X = x$ is a primitive event, are allowed in Theorem 6.3.

6.2 Results

Our complexity results for the case of situations are summarized in Table 4. We consider the problem of recognizing explanations, which turns out to be complete for Π_3^P in the general and the binary case. Furthermore, we consider the problem of deciding the existence of explanations, which is shown to be complete for Σ_3^P in the general and the binary case. We also consider the problems of deciding subsumption and equivalence between causal models, which are shown to be complete for Π_3^P in the general and the binary case.

Table 4
Complexity of Explanations: Situations

Problem	general case	binary case
Explanation	Π_3^P-complete	Π_3^P-complete
Explanation Existence	Σ_3^P-complete	Σ_3^P-complete

Notice that by a standard guess and check argument, Π_3^P membership of Explanation for situations implies a Σ_4^P upper bound for deciding the existence of an explanation for situations, in a sensible formulation of the problem (see below). Moreover, as explanations for contexts are a special case of explanations for situations, the Σ_3^P lower bound of Explanation Existence in the case of contexts immediately implies a Σ_3^P lower bound of Explanation Existence in the case of situations.

As we show, this lower bound is in fact complemented with a Σ_3^P upper bound, which means that deciding the existence of explanations for situations is not harder than for contexts. On the other hand, the problem is already Σ_3^P-hard for binary models. This is explained by subsumption checks which implicitly occur in forming an explanation for situations, whose complexity dominates the complexity of explanations in the binary case.

We exploit the characterization of subsumption in Theorem 6.3 to derive the following complexity result on checking subsumption between causal models.

Theorem 6.4 *Given causal models $M = (U, V, F)$ and $M_i = (U_i, V, F_i)$, $1 \leq i \leq k$, deciding whether $M_1, M_2 \ldots, M_k \leq M$ is Π_3^P-complete. Hardness holds even if $k = 1$, that is, for pairs of causal models.*

Proof. We first prove membership in Π_3^P. By Theorem 6.3, to show that $M_1, \ldots, M_k \not\leq M$, we can guess some $u \in D(U)$ and then check that for every $i \in \{1, \ldots, k\}$ and $u_i \in D(U_i)$, there exists some causal formula $[Y \leftarrow y]\, X = x$, where Y is a (possibly empty) set of endogenous variables and X is a single variable, such that (i) $(M, u) \not\models [Y \leftarrow y]\, X = x$ and (ii) $(M_i, u_i) \models [Y \leftarrow y]\, X = x$. This can be done in nondeterministic polynomial time, using a Σ_2^P-oracle. Thus, the problem is in Π_3^P.

Hardness for Π_3^P for $k = 1$ is shown by a reduction from deciding whether a given QBF $\Phi = \forall B\, \exists C\, \forall D\, \gamma$ is valid, where $\gamma = \gamma(B, C, D)$ is a propositional formula on the variables $B = \{B_1, \ldots, B_l\}$, $C = \{C_1, \ldots, C_m\}$, and $D = \{D_1, \ldots, D_n\}$.

We now construct two causal models $M = (U, V, F)$ and $M_1 = (U_1, V, F_1)$ such that $M_1 \leq M$ iff Φ is valid. The sets of exogenous and endogenous variables are defined by $U = U_1 = B \cup C$ and $V = D \cup W \cup \{Z\}$, respectively, where $W = \{W_1, \ldots, W_l\}$ and $D(X) = \{0, 1\}$ for all $X \in U \cup V$. The functions $F = \{F_X \mid X \in V\}$ and $F_1 = \{F_X^1 \mid X \in V\}$ are defined by $F_X = F_X^1 = 0$ for all $X \in V \setminus \{Z\}$, $F_Z = \bigvee_{i=1}^{l} (B_i \neq W_i)$, and $F_Z^1 = \bigvee_{i=1}^{l} (B_i \neq W_i) \vee \neg\gamma$ (see Fig. 7).

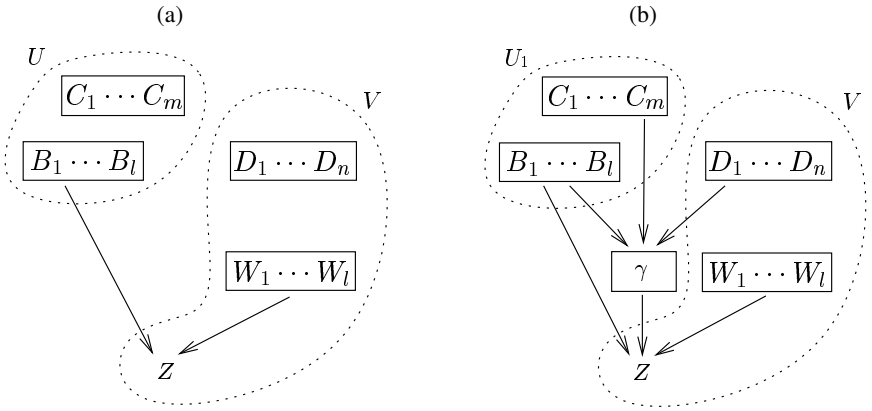

Figure 7. Causal Models (a) $M = (U, V, F)$ and (b) $M_1 = (U_1, V, F_1)$

We now prove that Φ is valid iff $M_1 \leq M_2$. It can be shown that Φ is not valid iff $(*)$ some $u \in D(U)$ exists such that for every $u_1 \in D(U_1)$, there exists a causal formula $[Y \leftarrow y]\, X = x$, where $Y \subseteq V$ and $X \in V$, such that (i) $(M, u) \not\models [Y \leftarrow y]\, X = x$ and (ii) $(M_1, u_1) \models [Y \leftarrow y]\, X = x$ (see Appendix D). By Theorem 6.3, this proves that Φ is valid iff $M_1 \leq M_2$. □

By an extension to the proof of Theorem 6.4, we obtain the following complexity result on testing equivalence between causal models.

Theorem 6.5 *Given two causal models $M_1 = (U_1, V, F_1)$ and $M_2 = (U_2, V, F_2)$, deciding whether $M_1 \equiv M_2$ is Π_3^P-complete.*

Proof. We first show membership in Π_3^P. Recall that $M_1 \equiv M_2$ iff $M_1 \leq M_2$ and $M_2 \leq M_1$. By Theorem 6.4, deciding whether $M_1 \leq M_2$ (resp., $M_2 \leq M_1$) holds is in Π_3^P. Thus, as Π_3^P is closed under conjunction, the problem is in Π_3^P.

For the Π_3^P-hardness part, we give a reduction from deciding $M_1 \leq M_2$. Roughly speaking, we construct a causal model M such that $Th(M) = Th(M_1) \cap Th(M_2)$. Then, $M_1 \leq M_2$ iff $M \equiv M_1$, which proves the result.

We construct the causal model $M = (U, V, F)$ as follows. Assume, without loss of generality, that M_1 and M_2 are such that the union $G(M_1) \cup G(M_2) = (V \cup U_1 \cup U_2, E_1 \cup E_2)$ of their causal graphs $G(M_1) = (V \cup U_1, E_1)$ and $G(M_2) = (V \cup U_2, E_2)$ is a directed acyclic graph; note that M and M_1 in the proof of Theorem 6.4 have this property.

The set of exogenous variables is given by $U = U_1 \cup U_2 \cup \{U_0\}$, where U_0 is a fresh exogenous variable with domain $D(U_0) = \{1, 2\}$. The functions $F = \{F_X \mid X \in V\}$ are constructed from the functions $F_1 = \{F_X^1 \mid X \in V\}$ and $F_2 = \{F_X^2 \mid X \in V\}$ as follows. For each $X \in V$, let the parents PA_X of X in M be the union PA_X^1 and PA_X^2 of the parents of X in M_1 and M_2, respectively, plus U_0, and define $F_X(x) = F_X^1(x|PA_X^1)$ if $x|U_0 = 1$, and $F_X(x) = F_X^2(x|PA_X^2)$ if $x|U_0 = 2$. That is, if the U_0-component of x is $i \in \{1, 2\}$, then the value of F_X is the value of the function F_X^i for X in the model M_i on x projected to the parents of X.

Notice that M is a recursive causal model, because its causal graph $G(M) = (U \cup V, E_1 \cup E_2 \cup \{U_0 \rightarrow X \mid X \in V\})$ is a directed acyclic graph. Clearly, for every causal formula on V, it holds that $M \models \phi$ iff $M_1 \models \phi$ and $M_2 \models \phi$. Thus, $Th(M) = Th(M_1) \cap Th(M_2)$, as desired. As M can be built in polynomial time from M_1 and M_2, the result follows. □

We finally address the issue of recognizing explanations relative to a set of situations $\mathcal{S}$. In that, we make use of the following lemma, which is helpful in checking the minimality condition ES3.

Lemma 6.6 *Let $\mathcal{M}$ and $\mathcal{M}' = \{M_1, \ldots, M_n\}$ be sets of causal models such that $\mathcal{M}' \subseteq \mathcal{M}$. Then, there exists a causal formula ϕ defining $\mathcal{M}'$ in $\mathcal{M}$, that is, $\mathcal{M}' = \{M \in \mathcal{M} \mid M \models \phi\}$, iff $M_1, \ldots, M_n \not\leq M$ holds for every $M \in \mathcal{M} \setminus \mathcal{M}'$.*

Proof. ($\Rightarrow$) Let ϕ define $\mathcal{M}'$, and assume towards a contradiction that there exists some $M \in \mathcal{M} \setminus \mathcal{M}'$ such that $M_1 \ldots, M_n \leq M$. Since $\phi \in \bigcap_{i=1}^n Th(M_i)$, it follows that $\phi \in Th(M)$, which contradicts that ϕ defines $\mathcal{M}'$.

($\Leftarrow$) Suppose that $M_1 \ldots, M_n \not\leq M$ holds for every model $M \in \mathcal{M} \setminus \mathcal{M}'$. Hence,

there exists a formula $\phi_M \in \bigcap_{i=1}^{n} Th(M_i)$ such that $\phi_M \notin Th(M)$. Consequently, the formula $\phi = \bigwedge_{M \in \mathcal{M} \setminus \mathcal{M}'} \phi_M$ defines $\mathcal{M}'$, that is, for every $M \in \mathcal{M}$, it holds that $M \in \mathcal{M}'$ iff $M \models \phi$. $\Box$

We are now ready to analyze the complexity of recognizing explanations in the case of situations. The following theorem shows that this problem is Π_3^P-complete. Here, Π_3^P-hardness is inherited from the Π_3^P-hardness of subsumption checking. Notice that for binary causal models, the complexity of recognizing explanations is the same, as subsumption checking is Π_3^P-hard already for binary causal models.

Theorem 6.7 *Given a causal formula ψ, a set of endogenous variables X, a value $x \in D(X)$, an event ϕ, and a set of situations $\mathcal{S}$, deciding whether $(\psi, X = x)$ is an* explanation *of ϕ relative to $\mathcal{S}$ is Π_3^P-complete.*

Proof. We first prove membership in Π_3^P. Recall that $(\psi, X = x)$ is an explanation of ϕ relative to $\mathcal{S}$ iff ES1–ES4 hold. Let $\mathcal{M}$ denote the set of all causal models M such that $(M, u) \in \mathcal{S}$ for some context u.

By Proposition 2.3, in ES1, deciding whether $(M, u) \models \phi$ for all $(M, u) \in \mathcal{S}$ is polynomial, and in ES4, deciding whether $(M, u) \models X = x$ and $(M', u') \models \neg(X = x)$ for some $(M, u), (M', u') \in \mathcal{S}$ is polynomial.

In ES2, we decide whether for every $(M, u) \in \mathcal{S}$, it holds (a) $(M, u) \models \neg(X = x)$, or (b) $(M, u') \models \neg\psi$ for some context u' in M, or (c) $X = x$ is a weak cause of ϕ under u in M. By Proposition 2.3, (a) is polynomial and (b) is in NP. By Theorem 2.6, (c) is in Σ_2^P. In summary, deciding whether ES2 holds is in Σ_2^P.

In ES3, we apply Lemma 6.6: To disprove ES3, we may guess some $X' \subseteq X$ and some $\mathcal{M}' = \{M_1, \ldots, M_n\} \subseteq \mathcal{M}$ such that the following holds: (i) $\{M \in \mathcal{M} \mid M \models \psi\} \subseteq \mathcal{M}'$, (ii) $M_1, \ldots, M_n \not\leq M$ for all $M \in \mathcal{M} \setminus \mathcal{M}'$, (iii) $X' \neq X$ or $\{M \in \mathcal{M} \mid M \models \psi\} \neq \mathcal{M}'$, and (iv) for all $(M, u) \in \mathcal{S}$, either (a) $(M, u) \models \neg(X' = x|X')$, or (b) $M \notin \mathcal{M}'$, or (c) $X' = x|X'$ is a weak cause of ϕ under u in M. Tasks (i), (iii), and (iv) are clearly solvable in polynomial time with a Σ_2^P oracle. As for (ii), by Theorem 6.4, checking whether $M_1, \ldots, M_n \not\leq M$ holds for each $M \in \mathcal{M} \setminus \mathcal{M}'$ can be done in nondeterministic polynomial time with a Σ_2^P-oracle. This implies that deciding whether ES3 holds is in Π_3^P. In summary, deciding whether ES1–ES4 hold is in Π_3^P.

Hardness for Π_3^P is shown by a reduction from the problem of deciding subsumption between causal models, which is Π_3^P-complete by Theorem 6.4: Given two causal models $M_1 = (U_1, V, F_1)$ and $M_2 = (U_2, V, F_2)$, decide whether $M_1 \leq M_2$. By the proof of Theorem 6.4, we can assume that $U_1 = U_2 = U$.

We now construct a causal formula ψ, a set of endogenous variables X, a value $x \in D(X)$, an event ϕ, and a set of situations $\mathcal{S}$, such that $(\psi, X = x)$ is an explanation of ϕ relative to $\mathcal{S}$ iff $M_1 \leq M_2$.

The set of situations is defined by $\mathcal{S} = \{S_i = (M_i, u_i) \mid 3 \leq i \leq 6\}$, where the causal models $M_i{=}(U_i, V_i, F_i)$ and the contexts u_i are given as follows. For $i{\in}\{3, \ldots, 6\}$, the sets of exogenous and endogenous variables are defined by $U_i = U \cup \{U_0\}$ and $V_i = V \cup \{X_0, Y, T\}$, respectively, where $D(X){=}\{0, 1\}$ for all $X{\in}\{U_0, X_0, Y, T\}$. For $i \in \{3, \ldots, 6\}$, the functions $F_i = \{F^i_X \mid X \in V_i\}$ are defined as follows:

- $F_3 = \{F^3_{X_0} = 0,\ F^3_Y = (U_0 = 0) \wedge (X_0 = 1),\ F^3_T = 1\} \cup F_1$;
- $F_4 = \{F^4_{X_0} = 0,\ F^4_Y = (U_0 = 0) \wedge (X_0 = 1),\ F^4_T = 1\} \cup F_2$;
- $F_5 = \{F^5_{X_0} = 0,\ F^5_Y = X_0,\ F^5_T = 0\} \cup \{F^5_X = 0 \mid X \in V\}$;
- $F_6 = \{F^6_{X_0} = 1,\ F^6_Y = 0,\ F^6_T = 0\} \cup \{F^6_X = 0 \mid X \in V\}$.

The contexts $u_3, \ldots, u_6$ are arbitrary such that $u_3(U_0) = 0$ and $u_4(U_0) = 1$.

Observe now that $X_0 = 0$ is a weak cause of $Y = 0$ under u_3 in M_3, while $X_0 = 0$ is not a weak cause of $Y = 0$ under u_4 in M_4 (but $X_0(u_4) = 0$ in M_4). Moreover, notice that $X_0 = 0$ is a weak cause of $Y = 0$ under u_5 in M_5, while $X_0 = 0$ is not a weak cause of $Y = 0$ under u_6 in M_6 (as $X_0(u_6) \neq 0$ in M_6).

Intuitively, if we want to form an explanation $(\psi, X_0 = 0)$ for $Y = 0$, the situation (M_6, u_6) serves, together with the situation (M_5, u_5), as a witness to the property ES4. By minimality of an explanation, we must have M_5 selected by ψ, since $X_0 = 0$ is in M_5 a weak cause for $Y = 0$ in context u_5. Furthermore, M_3 may be selected; this, however, is only possible if it does not require to select also M_4 by subsumption, as (M_4, u_4) spoils the condition ES2. That is, M_3 may not be selected, just if $M_3 \leq M_4$ holds, which is equivalent to $M_1 \leq M_2$. Thus, if we have a causal formula ψ which selects precisely M_5 and M_6, then the candidate $(\psi, X_0 = 0)$ is an explanation just if $M_1 \leq M_2$ holds.

We now show that $(T = 0, X_0 = 0)$ is an explanation of $Y = 0$ relative to $\mathcal{S}$ iff $M_1 \leq M_2$ holds. Indeed, it is easily checked that by construction, ES1, ES2, and ES4 hold. If $M_1 \not\leq M_2$, then ES3 is violated, as $(T = 0 \vee \phi, X_0 = 0)$ satisfies ES2, where $\phi \in Th(M_1) \setminus Th(M_2)$ is arbitrary, and $\{M \in \mathcal{M} \mid M \models T = 0 \vee \phi\} = \{M_3, M_5, M_6\} \supset \{M_5, M_6\} = \{M \in \mathcal{M} \mid M \models T = 0\}$, where $\mathcal{M} = \{M_i \mid 3 \leq i \leq 6\}$. Conversely, if ES3 is violated, then $(\psi', X' = x') \not\approx_{\mathcal{S}} (T = 0, X_0 = 0)$ means that $X' = x'$ coincides with $X_0 = 0$ and thus $M_3 \models \psi'$ and $M_4 \not\models \psi'$ must hold (as $X_0 = 0$ is not a weak cause of $Y = 0$ under u_4 in M_4, but $X_0(u_4) = 0$ in M_4). Thus, $M_1 \not\leq M_2$ holds.

As the above reduction is polynomial, this shows Π^P_3-hardness. □

Let us now turn to the issue of deciding the existence of explanations in the general case of situations. This problem has to be carefully defined, since otherwise simple (and perhaps unintended) explanations may be found.

It is not difficult to see that if an event ϕ satisfies ES1 for a set of situations $\mathcal{S}$,

and if X_0 is variable and x_0 a value for X_0 such that ES4 holds for $X_0 = x_0$, that then some explanation of form $(\psi, X_0 = x_0)$ for ϕ w.r.t. $\mathcal{S}$ exists. This implies that given a set of variables X to build explanations using them for ϕ w.r.t. $\mathcal{S}$, deciding whether some explanation exists is possible in polynomial time.

A more sensible formulation of Explanation Existence for the case of situations is the following.

Explanation Existence (for situations): Given a finite set of situations $\mathcal{S}$, a set of endogenous variables X, a causal formula ψ, and an event ϕ, decide whether a causal formula ψ' with $\psi \models_{\mathcal{S}} \psi'$, $X' \subseteq X$, and $x' \in D(X')$ exist such that $(\psi', X' = x')$ is an explanation of ϕ relative to $\mathcal{S}$.

Here, the causal formula ψ is a positive selection condition for causal models in ES2, such that each causal model M satisfying ψ must be respected and the event $X = x$ must be a weak cause of ϕ under u for every situation $(M, u) \in \mathcal{S}$ such that $(M, u) \models X = x$ and $M \models \psi$. A weakening of ψ, that is, a cautious enlargement of the set of respected causal models is admissible, which amounts to adding alternative selection conditions. Before we analyze the complexity of Explanation Existence for situations, we introduce some terminology.

We call a pair $(\psi, X = x)$ a *pseudo-explanation* of an event ϕ relative to a set of situations $\mathcal{S}$, if $(\psi, X = x)$ satisfies conditions ES1, ES2, ES4, and the following weakened form of ES3:

ES3′. There is no $(\psi, X' = x') \not\approx_{\mathcal{S}} (\psi, X = x)$ satisfying ES2 such that $X' \subseteq X$ and $x' = x|X'$.

The following result is useful for determining the complexity of Explanation Existence.

Lemma 6.8 *Given a causal formula ψ, an event ϕ, a set of endogenous variables X, and a finite set of situations $\mathcal{S}$, there exists an explanation $(\psi', X' = x')$ of ϕ relative to $\mathcal{S}$ such that $\psi \models_{\mathcal{S}} \psi'$, $X' \subseteq X$, and $x' \in D(X')$ iff there exists a pseudo-explanation $(\psi', X' = x')$ of ϕ relative to $\mathcal{S}$ such that $\psi \models_{\mathcal{S}} \psi'$, $X' \subseteq X$, and $x' \in D(X')$.*

Proof. ($\Rightarrow$) Obviously, any explanation is a pseudo-explanation.

($\Leftarrow$) Let $(\psi', X' = x')$ be a pseudo-explanation of ϕ relative to $\mathcal{S}$ such that $\psi \models_{\mathcal{S}} \psi'$. We show that there exists some explanation $(\psi'', X' = x')$ of ϕ relative to $\mathcal{S}$ such that $\psi' \models_{\mathcal{S}} \psi''$. Let ψ^* be a weakest formula ψ'' such that $\psi' \models_{\mathcal{S}} \psi''$ and ES2 holds for $(\psi'', X' = x')$. We claim that $(\psi^*, X' = x')$ is an explanation of ϕ relative to $\mathcal{S}$. Since $\psi \models_{\mathcal{S}} \psi^*$, this will prove the result.

Towards a contradiction, suppose $(\psi'', X'' = x'')$, where $\psi^* \models_{\mathcal{S}} \psi''$, is such that it satisfies ES2 and either (i) $X'' \subset X'$ and $x'' = x'|X''$, or (ii) $\psi^* \not\equiv_{\mathcal{S}} \psi''$. In case (i), each causal model M selected by ψ' is also selected by ψ^*, and thus by ψ''; furthermore, $X'' = x''$ is a weak cause of ϕ in M under u for each $(M, u) \in \mathcal{S}$ such that $M \models \psi''$ and $(M, u) \models X'' = x''$. This contradicts that $(\psi', X' = x')$ is a pseudo-explanation of ϕ relative to $\mathcal{S}$. Thus, $X'' = X'$ must hold, and case (ii) must apply. However, this means that ψ^* is not a weakest formula ψ'' such that $\psi' \models_{\mathcal{S}} \psi''$ and $(\psi'', X' = x')$ satisfies ES2, which is a contradiction. This proves that $(\psi^*, X' = x')$ is an explanation of ϕ relative to $\mathcal{S}$. □

Theorem 6.9 *Problem Explanation Existence for situations is Σ_3^P-complete.*

Proof. We first prove membership in Σ_3^P. By Lemmas 6.6 and 6.8, it is sufficient to guess some $X' \subseteq X$, $x' \in D(X')$, and $\mathcal{M}' = \{M_1, \ldots, M_n\} \subseteq \mathcal{M}$ such that (i) $\{M \in \mathcal{M} \mid M \models \psi\} \subseteq \mathcal{M}'$, (ii) $M_1, \ldots, M_n \not\leq M$ for all $M \in \mathcal{M} \setminus \mathcal{M}'$, and ES1, ES2 where "$M \in \mathcal{M}'$" replaces "$M \models \psi$", ES3′, and ES4 hold. Task (i) can be done in polynomial time with an NP-oracle, while task (ii) can be done, by Theorem 6.4, in nondeterministic polynomial time with a Π_2^P-oracle. Checking ES1 and ES4 is possible in polynomial time, while ES2 can be checked, by Proposition 6.1 and Theorem 3.3, in polynomial time with a Σ_2^P oracle. Finally, checking ES3′ is in Π_2^P, since deciding the existence of a counterexample to minimality is in Σ_2^P. In summary, the whole procedure runs in nondeterministic polynomial time using a Π_2^P oracle. Hence, Explanation Existence is in Σ_3^P.

For the case of unrestricted models, Σ_3^P-hardness is inherited from the Σ_3^P-completeness of Explanation Existence for context explanations, which occurs as a special case of Explanation Existence for situations. We show Σ_3^P-hardness for the binary case by a reduction from deciding non-subsumption between causal models, which is Σ_3^P-complete by Theorem 6.4: Given two causal models $M_1 = (U_1, V, F_1)$ and $M_2 = (U_2, V, F_2)$, decide whether $M_1 \not\leq M_2$. Without loss of generality, we assume that $U_1 = U_2 = U$.

The reduction is similar in spirit to the one in the proof of Theorem 6.7, yet different. We construct a causal formula ψ, a set of endogenous variables X, an event ϕ, and a set of situations $\mathcal{S}$, such that some explanation $(\psi', X' = x')$ of ϕ relative to $\mathcal{S}$ exists such that $\psi \models_{\mathcal{S}} \psi'$, $X' \subseteq X$, and $x' \in D(X')$ iff $M_1 \not\leq M_2$.

The set of situations is defined by $\mathcal{S} = \{S_i = (M_i, u_i) \mid 3 \leq i \leq 6\}$, where the causal models $M_i = (U_i, V_i, F_i)$ and the contexts u_i are given as follows. For $i \in \{3, \ldots, 6\}$, the sets of exogenous and endogenous variables are defined by $U_i = U \cup \{U_0\}$ and $V_i = V \cup \{X_0, X_1, Y, T\}$, respectively, where $D(X) = \{0, 1\}$ for all $X \in \{U_0, X_0, X_1, Y, T\}$. For $i \in \{3, \ldots, 6\}$, the functions $F_i = \{F_X^i \mid X \in V_i\}$ are defined as follows:

- $F_3 = \{F_{X_0}^3 = 0,\ F_{X_1}^3 = 0,\ F_Y^3 = (U_0 = 0) \wedge (X_0 = 1),\ F_T^3 = 1\} \cup F_1$;

- $F_4 = \{F^4_{X_0} = 0,\ F^4_{X_1} = 0,\ F^4_Y = (U_0 = 0) \wedge (X_0 = 1),\ F^4_T = 1\} \cup F_2$;
- $F_5 = \{F^5_{X_0} = 0,\ F^5_{X_1} = 0,\ F^5_Y = X_1,\ F^5_T = 0\} \cup \{F^5_X = 0 \mid X \in V\}$;
- $F_6 = \{F^6_{X_0} = 1,\ F^6_{X_1} = 0,\ F^6_Y = (T = 1) \wedge (X_0 = 1 \vee X_1 = 1),\ F^6_T = 0\} \cup \{F^6_X = 0 \mid X \in V\}$.

The contexts $u_3, \ldots, u_6$ are arbitrary such that $u_3(U_0) = 0$ and $u_4(U_0) = 1$.

Observe now that the situations S_i have the following weak causes of $Y = 0$ involving only variables in $X = \{X_0, X_1\}$: S_3 has $X_0 = 0$ and $X_0 = 0 \wedge X_1 = 0$, S_4 has no weak cause, S_5 has $X_1 = 0$ and $X_0 = 0 \wedge X_1 = 0$, and S_6 has $X_1 = 0$.

Define now $\psi = T = 0$, $\phi = Y = 0$, and $X = \{X_0, X_1\}$. Note that ψ selects the models M_5 and M_6.

Intuitively, S_5 and S_6 create a single candidate event, $X_0 = 0 \wedge X_1 = 0$, for an explanation $(\psi', X' = x')$ of ϕ as desired. This candidate is good if S_3 but not S_4 can be respected in the explanation, i.e., ψ' selects M_3 but not M_4, which is equivalent to $M_1 \not\leq M_2$.

Formally, in any explanation $(\psi', X' = x')$ for $Y = 0$ relative to $\mathcal{S}$ such that $\psi \models_{\mathcal{S}} \psi'$, the set X' must be different from X_0; otherwise, $X' = x'$ is not a weak cause under u_5 in M_5 and under u_6 in M_6, which means that ES2 is violated. Thus, X' must include X_1. On the other hand, $X \neq \{X_1\}$ and $x'|X_1 = 0$ must hold, since otherwise ES4 is violated. Since $X_0 = 1 \wedge X_1 = 0$ is not weak cause of $Y = 0$ in M_6 under u_6, we have that $X' = x'$ must be of form $X_0 = 0 \wedge X_1 = 0$.

We claim that some $(\psi', X_0 = 0 \wedge X_1 = 0)$ with $\psi \models_{\mathcal{S}} \psi'$ is an explanation of ϕ relative to $\mathcal{S}$ iff $M_1 \not\leq M_2$ holds.

Suppose that $(\psi', X_0 = 0 \wedge X_1 = 0)$ is an explanation of ϕ relative to $\mathcal{S}$. We must have $\psi' \not\equiv_{\mathcal{S}} \psi$: indeed, $(\psi, X_0 = 0 \wedge X_1 = 0)$ is not a pseudo-explanation of ϕ, since ES3′ fails, which is witnessed by $(\psi, X_1 = 0)$ satisfying ES2. Therefore, ψ' must select either M_3 or M_4. Since X allows no weak cause of $Y = 0$ in M_4 under u_4, ψ' must not select M_4. This implies $M_3 \not\leq M_4$, which in turn implies that $M_1 \not\leq M_2$.

Conversely, suppose that $M_1 \not\leq M_2$. Then, $M_3 \not\leq M_4$, and the set $\{M_3, M_5, M_6\}$ is definable by a formula ψ' such that $\psi \models_{\mathcal{S}} \psi'$. Consider $(\psi', X_0 = 0 \wedge X_1 = 0)$. Clearly, ES1 holds for $\phi = Y = 0$ and ES4 holds for $X_0 = 0 \wedge X_1 = 0$. Also ES2 holds, since $X_0 = 0 \wedge X_1 = 0$ is a weak cause of $Y = 0$ in M_3 under u_3 and in M_5 under u_5. Furthermore, neither for $(\psi', X_0 = 0)$ nor for $(\psi', X_1 = 0)$ is ES2 satisfied, since $X_0 = 0$ is not a weak cause of $Y = 0$ in M_5 under u_5 and $X_1 = 0$ is not a weak cause of $Y = 0$ in M_3 under u_3. Thus, $(\psi', X_0 = 0 \wedge X_1 = 0)$ is a pseudo-explanation of ϕ relative to $\mathcal{S}$. From Lemma 6.8, it follows that some explanation $(\psi'', X_0 = 0 \wedge X_1 = 0)$ of $Y = 0$ relative to $\mathcal{S}$ exists (in fact, $\psi'' \equiv_{\mathcal{S}} \psi'$ must hold).

As the above reduction is polynomial, this shows Σ_3^P-hardness. □

We remark that the existence of specific explanations may have higher complexity. For example, deciding the existence of an explanation $(\psi', X' = x')$ where $\psi' = \psi$, is both Σ_3^P-hard and Π_3^P-hard; the latter is implicit in the proof of Theorem 6.7.

6.3 Causal Formulas with Exogenous Variables

We now give some remarks on the impact of the language of events that is considered in defining explanations and situations. In this paper, like in [25,26], primitive events involve only endogenous variables. The setting stated in [17] is slightly more liberal and also admits exogenous variables to occur in primitive events. While such enhanced expressiveness does not increase the complexity results for explanations in Sections 3–5, it allows to simplify some of the technical hardness proofs. On the other hand, the higher expressiveness of causal formulas which may also involve exogenous variables via primitive events implies a refinement of the subsumption and indiscernibility relation, which is also easier to test: The characterization of $M_1, \ldots, M_n \trianglelefteq M$ in Theorem 6.3, where $M = (U, V, F)$ and $M_i = (U, V, F_i)$ for all $i \in \{1, \ldots, n\}$, can be replaced by the following simpler condition:

$(**)$ There exists some $u \in D(U)$ such that for every $i \in \{1, \ldots, n\}$, there exists a causal formula $[Y \leftarrow y]\, X = x$, where Y is a (possibly empty) set of endogenous variables and X is a single variable, such that (i) $(M, u) \not\models [Y \leftarrow y]\, X = x$, and (ii) $(M_i, u) \models [Y \leftarrow y]\, X = x$.

The check of this condition is easily seen to be NP-complete. Therefore, the subsumption test $M_1 \leq M_2$ (resp., equivalence test $M_1 \equiv M_2$) is $\mathrm{co\text{-}NP}$-complete rather than Π_3^P-complete, and thus two levels lower in the polynomial hierarchy. Consequently, it does not dominate the complexity of the conditions ES1–ES4; the same algorithm for checking an explanation, performed in this setting, yields then a $\mathrm{D}_2^{\mathrm{P}}$ (resp., D^{P}) upper bound in the general (resp., binary) case. A matching lower bound is inherited from the complexity of Explanation in the basic setting, as it is a special case of situations, and thus the problem is complete for $\mathrm{D}_2^{\mathrm{P}}$ (resp., D^{P}). Similarly, for the problem Explanation Existence, we obtain completeness for Σ_3^P and Σ_2^P in the general and the binary case, respectively.

7 Related Work

In this section, we give a comparison of our work to related work on complexity of explanations in the areas of abduction and of Bayesian networks.

Abduction has been recognized as an important principle of common-sense reasoning, and plays an important role in many AI problems including diagnosis, planning, or natural language processing to mention but a few. One of the uses of abduction is to obtain explanations for observations, which loosely speaking is accomplished by a kind of reversed modus ponens. There is quite some work on algorithms and complexity of finding abductive explanations (e.g. [4,8,9,11,44,47]).

Roughly, in a logic-based setting, abductive explanations are defined as follows (cf. [35,47]). Given some background knowledge T, which is a theory, i.e., a set of sentences in some logic, and a set of observations O, which are typically facts, a set of sentences E from a set of hypotheses H is an *explanation* of O from T, iff

(1) $T \cup E$ is satisfiable, i.e., not contradictory, and
(2) $T \cup E \models O$, i.e., the observations are logically entailed from the background knowledge and the explanation, under a notion of logical entailment $\models$.

Usually, further conditions are imposed on E in order to single out most plausible explanations. A standard such condition is the application of Occam's razor, i.e., minimality in terms of set inclusion.

While causal and abductive explanations, in a standard logical setting such as above, are apparently different concepts, they have similar complexity. In particular, deciding the existence of an abductive explanation in the propositional context (i.e., T, O, and H are in classical propositional logic) is Σ_2^P-complete, as shown in [11]. This matches our respective result on causal explanations for binary causal models. In fact, computing causal explanations can be polynomially transformed into computing abductive explanations in this case, and vice versa.

In the case of causal models with non-binary domains, explanations are one level higher up in the Polynomial Hierarchy, and deciding the existence of a causal explanation is Σ_3^P-complete. This matches, interestingly, the complexity of abductive explanations from disjunctive logic programs under the stable resp. answer set semantics. In this setting, the background theory T is a propositional disjunctive logic programs, H and E are set of atoms, and $\models$ is standard cautious inference, i.e., truth in all stable models resp. answer sets of a program. As was shown in [12], deciding the existence of an abductive explanation is Σ_3^P-complete in this scenario. Thus, deciding the existence of causal and of abductive LP explanations is polynomially intertranslatable, which extends easily to computing some causal resp. abductive LP explanation, and computational engines could be mutually exploited.

The issue of efficient transformations of causal into abductive explanations, as well as into related reasoning tasks of nonmonotonic formalisms, is an interesting subject for further work, which may also be exploited for obtaining rapid prototype

implementations. E.g., by mapping binary causal explanations to abductive explanations, (extended) variants of the Truth Maintenance System (cf. [44]) could be utilized for this purpose, or the diagnostic frontend of the DLV system [10]. Another possibility would be an encoding of causal explanations in Answer Set Programming, and using the DLV engine to compute solutions. For the case of general causal explanations, reductions to QBF solvers such as [5,42,19] could be used.

7.2 Bayesian Networks

After Cooper's well-known intractability result [7] for probabilistic inference in Bayesian networks, a number of papers in this area have investigated complexity issues for reasoning and in particular for explanation finding.

A dominating notion of explanation in the probabilistic AI literature is the *maximum a posteriori explanation* (MAP, alias *most probable explanation* [33,34]), which is an assignment to all variables given a partial assignment to the variables in a Bayesian network, such that its probability is maximum. Some complexity results for MAPs have been derived, which however are only weakly related to our results for causal explanations. In particular, computing a MAP in a Bayesian network is NP-hard [49], and the same applies to computing a MAP approximation [1]; on the other hand, this is feasible in polynomial time with an NP oracle.

This result on computing a MAP is quite different from our results on α-partial explanations, for two reasons: firstly, MAPs are computed from the set of all contexts, which is not part of the input. In this setting, α-partial explanations have higher complexity. Secondly, MAPs are *single contexts* which maximize probability for a given evidence, while α-partial explanations single out *subsets of contexts* which sensibly respect relevant information [27].

From the computational side, it is more suitable to compare deciding $P(X = x) > 0$ in a Bayesian network with our problem Partial Explanation under succinct context sets, where $\mathcal{C}$ contains all possible contexts and P emerges from independent exogenous variables. However, the former problem is NP-complete [7], while the latter is, by our results, Π_4^P-complete and thus much harder. We may thus expect a similar relationship between computing the explanatory power and the probability $P(X = x)$ in a Bayesian network, which can be done in polynomial time with the help of a #P oracle [43].

8 Conclusion

In this paper, we have considered explanations in Halpern and Pearl's structural-model approach to causality from a computational perspective, and we have obtained a number of complexity results which precisely characterize the intrinsic difficulty of major computational tasks on explanations.

Our results give a clear picture of the complexity of explanations in the case of general structural models, as well as under the restriction to the case where all variables are binary. As we have shown, causal explanations reside at the third level of the Polynomial Hierarchy (PH) in the basic setting, and thus are, computationally speaking, harder to compute than, for example, abductive explanations in the standard logic-based setting, which are at the second level of PH. Intuitively, causal explanations harbor three intermingled sources of complexity, which make the concept difficult: (1) the, in general, exponential set of candidates $X = x$ for an explanation formed from variables X in a given set X' of variables; (2) condition AC2(b), which informally is a kind of validity test ensuring that X alone is sufficient to bring about the change of the event ϕ to $\neg\phi$, and thus impacts on ϕ; and (3) minimality of explanations, which implies an exponential set of candidates in condition EX3 / ES3 for spoiling a candidate explanation. The complexity of causal explanations further increases, as demonstrated for the recognition problem, under a natural concise form of model representation by two levels in PH. In particular, the recognition problems was proved to be Π_4^P-complete, and thus is, compared to validity checking in classical propositional logic, a rather complex problem.

Some of our hardness results remain valid under further restrictions, such as a boundedness condition on the causal model [14,18]. In particular, all hardness results from Tables 1–3 in Sections 3–5 hold for primitive events ϕ. Thus, complex events are not a source of complexity. However, to avoid a proliferation of results, we did not further consider such restrictions here.

For "efficient" algorithms to generate explanations or "best" α-partial explanations, we can conclude the following. Both must solve an inherent Σ_3^P-hard problem; thus, simple backtracking is infeasible, as well as polynomial reductions to a SAT solver or a computational logic system which can handle problems with complexity up to Σ_2^P, such as DLV [13,36]. However, an explanation may be computed using nested backtracking, or flat backtracking calling a subroutine for Σ_2^P tasks (e.g., calls to DLV). A further possible perspective are translations to QBF-solvers, which proved valuable in other applications [42]. We can compute an α-partial explanation similarly. Computing a best one amounts to an optimization problem, which can be solved by binary search over the range [0,1] of α, and thus in polynomial time with a Σ_3^P-oracle. A substantially faster algorithm seems unlikely to exist.

Once the basic results about the complexity of a framework are known, and in-

tractability of some tasks has been evidenced, a natural next step of research is to identify cases of lower complexity, and in particular to find islands of tractability. For that, meaningful restrictions must be found which eliminate the various sources of complexity, which is not straightforward.

While the complexity results for explanations established in this paper may look discouraging, and leave us with little hope for tractable cases, it turned out that there are meaningful restrictions of causal models for which explanations have polynomial complexity. In a companion paper [15,16] to [14,18] and the present paper, we describe nontrivial syntactic restrictions on causal models under which the notions of weak causes and explanations are tractable. In particular, we have identified a hierarchy of tractable classes, starting with simple causal trees, i.e., the causal graphs are trees, over layered causal graphs, i.e., the causal graphs can be layered so as to permit a step by step propagation of effects, to a general class of decomposable causal graphs. On such causal models, small weak causes under explanations can be computed efficiently under further assumptions which are needed to gain tractability. However, the technical definitions and the characterizations are far too involved to be discussed here; we refer the interested reader to [15,16] for details.

Hence, there are some positive results on the computation of causal explanations for certain instances already. It remains, however, to find other classes of instances that have lower complexity, and in particular that guarantee tractability; delineating the tractability frontier is a challenging task for future work. Likewise, the development of suitable algorithms, continuing and extending the work of Hopkins [30], is indispensable for making the structural-model approach amenable to efficient implementation and use in practice.

Acknowledgments

This work has been partially supported by the Austrian Science Fund under project Z29-N04, by a DFG grant, and by a Marie Curie Individual Fellowship of the European Community programme "Human Potential" under contract number HPMF-CT-2001-001286 (Disclaimer: The authors are solely responsible for information communicated and the European Commission is not responsible for any views or results expressed). We are thankful to the reviewers of this paper and its KR 2002 abstract, whose constructive comments helped to improve our work.

A Appendix: Proofs for Section 3

Proof of Theorem 3.3 (continued). Hardness for Σ_3^P is shown by a reduction from deciding whether a given QBF $\Phi = \exists B\, \forall C\, \exists D\, \gamma$ is valid, where γ is a propositional

formula on the variables $B = \{B_1, \ldots, B_l\}$, $C = \{C_1, \ldots, C_m\}$, and $D = \{D_1, \ldots, D_n\}$. We construct $M = (U, V, F)$, $X \subseteq V$, $\mathcal{C} \subseteq D(U)$, and ϕ as in the statement of the theorem such that Φ is valid iff some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$.

We define $U = \{I, U_0, {U_0}', \ldots, U_l, {U_l}'\}$, where $D(I) = \{0, \ldots, l+1\}$ and $D(S) = \{0, 1\}$ for all $S \in U \setminus \{I\}$. Let $\mathcal{C} = \{u_0, {u_0}', \ldots, u_l, {u_l}', u_{l+1}\}$, where u_i (resp., ${u_i}'$) is the unique $u \in D(U)$ such that $\varepsilon_i(u)$ (resp., ${\varepsilon_i}'(u)$) holds, and ε_i (resp., ${\varepsilon_i}'$) for every $i \in \{0, \ldots, l+1\}$ (resp., $i \in \{0, \ldots, l\}$) is defined by:

$$\varepsilon_i \;=\; I = i \wedge U_0 = 0 \wedge {U_0}' = 1 \wedge \bigwedge_{i=1}^{l} (U_i = 0 \wedge {U_i}' = 0)\,,$$

$${\varepsilon_i}' \;=\; I = i \wedge U_0 = 0 \wedge {U_0}' = 0 \wedge \bigwedge_{i=1}^{l} (U_i = 0 \wedge {U_i}' = 0)\,.$$

We define $M = (U, V, F)$ as follows. Let $V = B \cup B' \cup C \cup D \cup \{X_0, {X_0}', E, E', Y\}$, where $B' = \{{B_1}', \ldots, {B_l}'\}$, $D(S) = \{0, 1, 2\}$ for all $S \in D$, and $D(S) = \{0, 1\}$ for all $S \in V \setminus D$. Let

$$\alpha \;=\; (\neg\gamma' \wedge \bigwedge_{S \in D} S \neq 2) \vee (E = 0) \vee (X_0 = 1 \wedge E = 1 \wedge \bigvee_{S \in D} S \neq 2)\,,$$

$$\phi_1' \;=\; (\varepsilon_0 \vee {\varepsilon_0}' \rightarrow (X_0 = 0 \wedge \bigwedge_{i=1}^{l} B_i \neq {B_i}') \vee (\bigvee_{i=1}^{l} (B_i = 1 \wedge {B_i}' = 1)) \vee E' = 0)\,,$$

$$\phi_2' \;=\; \bigwedge_{i=1}^{l} (\varepsilon_i \vee {\varepsilon_i}' \rightarrow B_i = 0 \vee {B_i}' = 0)\,,$$

$$\phi_3' \;=\; (\varepsilon_{l+1} \rightarrow (\alpha \wedge \bigwedge_{i=1}^{l} B_i \neq {B_i}') \vee (\bigvee_{i=1}^{l} (B_i = 1 \wedge {B_i}' = 1)) \vee E' = 0)\,,$$

where γ' is obtained from γ by replacing each $S \in B \cup C \cup D$ by "$S = 1$". We are now ready to define the functions $F = \{F_S \mid S \in V\}$ as follows:

- $F_{B_i} = U_i$ and $F_{{B_i}'} = {U_i}'$ for all $i \in \{1, \ldots, l\}$,
- $F_{X_0} = U_0$ and $F_{{X_0}'} = {U_0}'$,
- $F_S = 0$ for all $S \in C \cup \{E, E'\}$,
- $F_S = X_0 + E$ for all $S \in D$,
- $F_Y = 1$ iff $\phi_1' \vee \phi_2' \vee \phi_3'$ is true.

Let $X = B \cup B' \cup \{X_0, {X_0}'\}$. Let ϕ be $Y = 1$. Notice that ϕ is primitive.

For every truth assignment τ to the variables in B, denote by $[B/\tau(B)]$ the substitution $[B_1/\tau(B_1), \ldots, B_l/\tau(B_l)]$, and we define $\alpha^\tau = \alpha\,[B/\tau(B)]$. Let $x_0 = 0$, and let $u \in D(U)$ with $X_0(u) = x_0$. Then, $X_0 = x_0$ is a weak cause of α^τ under u iff $\exists C\, \forall D\, \neg\gamma\,[B/\tau(B)]$ is valid [14,18]. That is, $X_0 = x_0$ is not a weak cause of α^τ under u iff $\forall C\, \exists D\, \gamma\,[B/\tau(B)]$ is valid. Thus, Proposition 2.5 implies the following:

(∗) For every $X' \subseteq B \cup B' \cup \{X_0, {X_0}'\}$ with $X_0 \in X'$, it holds that $X' = X'(u)$ is not a weak cause of α^τ under u iff $\forall C\, \exists D\, \gamma\,[B/\tau(B)]$ is valid.

We now show that Φ is valid iff some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$.

($\Leftarrow$) Assume that some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$. Then,

- $x'(S) = 0$ for all $S \in X' \cap (B \cup B' \cup \{X_0\})$,

as otherwise $X'(u) \neq x'$ for all $u \in \mathcal{C}$, and thus EX4 is violated. For every $i \in \{0, \ldots, l\}$, it holds either $X'(u_i) = x'$ or $X'({u_i}') = x'$. Thus, $X' \cap \{B_i, {B_i}'\} \neq \emptyset$ for all $i \in \{1, \ldots, l\}$, as otherwise $X' = x'$ is not a weak cause of ϕ under any $u \in \{u_i, {u_i}'\}$, and thus EX2 is violated. It follows that

- $X_0 \in X'$ and
- $|X' \cap \{B_i, {B_i}'\}| = 1$

for all $i \in \{1, \ldots, l\}$, as otherwise $X' = x'$ is not a weak cause of ϕ under any $u \in \{u_0, {u_0}'\}$, and thus EX2 is violated. It holds

- ${X_0}' \in X'$,

as otherwise $X'(u) = x'$ for all $u \in \mathcal{C}$, and thus EX4 is violated. We have

- $x'({X_0}') = 0$,

as otherwise, by Proposition 2.5, $X'' = x''$ is a weak cause of ϕ under every $u \in \mathcal{C}$, where $X'' = X' \setminus \{{X_0}'\}$ and $x'' = x'|X''$, and thus EX3 is violated. Observe now that $X'' = x''$ is not a weak cause of ϕ under $u = u_{l+1}$, where $X'' = X' \setminus \{{X_0}'\}$ and $x'' = x'|X''$, as otherwise EX3 is violated. Let the truth assignment τ to the variables in B be defined by $\tau(S) = 0$ iff $S \in X'$ for all $S \in B$. We now show that $X'' = x''$ is not a weak cause of α^τ under u. Towards a contradiction, assume the contrary. Thus, there exists some $W \subseteq V \setminus X''$, $\overline{x}'' \in D(X'')$, and $w \in D(W)$ such that $\neg\alpha^\tau_{\overline{x}''w}(u)$ and $\alpha^\tau_{x''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X'' \cup W)$ and $\hat{z} = \hat{Z}(u)$. Here, we can assume that $\overline{x}''(X_0) = 1, \overline{x}''(S) = 0$ for all $S \in X'' \setminus \{X_0\}$, $\{E'\} \cup ((B \cup B') \setminus X'') \subseteq W$, and $w(S) = 1$ for all $S \in \{E'\} \cup ((B \cup B') \setminus X'')$. Hence, it holds $\neg\alpha_{\overline{x}''w}(u)$ and $\alpha_{x''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X'' \cup W)$ and $\hat{z} = \hat{Z}(u)$. Thus, $\neg\phi_{\overline{x}''w}(u)$ and $\phi_{x''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X'' \cup W)$ and $\hat{z} = \hat{Z}(u)$. As $X''(u) = x''$ and $\phi(u)$, it follows that $X'' = x''$ is a weak cause of ϕ under u, which is a contradiction. Hence, $X'' = x''$ is not a weak cause of α^τ under u. By (∗), it follows that $\forall C\, \exists D\, \gamma\,[B/\tau(B)]$ is valid. That is, Φ is valid.

($\Rightarrow$) Assume that Φ is valid. That is, there exists a truth assignment τ to the variables in B such that $\forall C\, \exists D\, \gamma\,[B/\tau(B)]$ is valid. Define $X' = \{X_0, {X_0}'\} \cup \{S \in B \mid$

$\tau(B)=0\} \cup \{S' \in B' \mid \tau(B)=1\}$ and $x'(S)=0$ for all $S \in X'$. We now show that $X'=x'$ is an explanation of ϕ relative to $\mathcal{C}$. EX1 holds, as $\phi(u)$ for all $u \in \mathcal{C}$. EX2 holds, as $X'=x'$ is weak cause of ϕ under every $u_i{}'$ with $i \in \{0,\ldots,l\}$. EX4 holds, as $X'(u_0) \neq x'$ and $X'(u_0{}')=x'$. We next show that EX3 holds. Towards a contradiction, assume the contrary. That is, there exists some $X'' \subset X'$ such that $X''=x''$ is a weak cause of ϕ under every $u \in \mathcal{C}$ with $X''(u)=x''$, where $x''=x'|X''$. It holds $X'' \cap \{B_i, B_i{}'\} \neq \emptyset$ for all $i \in \{1,\ldots,l\}$, as otherwise $X''=x''$ is not a weak cause of ϕ under any $u \in \{u_i, u_i{}'\}$. It follows that $X_0 \in X''$, as otherwise $X''=x''$ is not a weak cause of ϕ under any $u \in \{u_0, u_0{}'\}$. It thus follows $X''=X' \setminus \{X_0{}'\}$. Hence, $X''=x''$ is a weak cause of ϕ under $u=u_{l+1}$. That is, some $W \subseteq V \setminus X''$, $\overline{x}'' \in D(X'')$, and $w \in D(W)$ exist such that $\neg\phi_{\overline{x}''w}(u)$ and $\phi_{x''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X'' \cup W)$ and $\hat{z}=\hat{Z}(u)$. As $\neg\phi_{\overline{x}''w}(u)$, it follows that $E' \in W$ and $w(E')=1$. As $\phi_{x''w}(u)$, for every $S \in (B \cup B') \setminus X''$, it holds either $S(u)=1$ or $S \in W$ and $w(S)=1$. As $\neg\phi_{\overline{x}''w}(u)$, it thus follows that $\overline{x}''(S)=0$ for all $S \in X'' \setminus \{X_0\}$. Hence, $\neg\alpha_{\overline{x}''w}(u)$ and $\alpha_{x''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X'' \cup W)$ and $\hat{z}=\hat{Z}(u)$. That is, $\neg\alpha^{\tau}_{\overline{x}''w}(u)$ and $\alpha^{\tau}_{x''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X'' \cup W)$ and $\hat{z}=\hat{Z}(u)$. As $X''(u)=x''$ and $\alpha^{\tau}(u)$, this shows that $X''=x''$ is a weak cause of α^{τ} under u. By $(*)$, it follows that $\forall C\, \exists D\, \gamma\,[B/\tau(B)]$ is not valid, which is a contradiction. This shows that EX3 holds. $\square$

Proof of Theorem 3.5 (continued). We define $U=\{I, U_0, U_0{}', \ldots, U_k, U_k{}'\}$, where $D(I)=\{0,\ldots,l+1\}$ and $D(S)=\{0,1\}$ for all $S \in U \setminus \{I\}$. Let $\mathcal{C}=\{u_0, u_0{}', \ldots, u_l, u_l{}', u_{l+1}\}$, where u_i (resp., $u_i{}'$) is the unique $u \in D(U)$ such that $\varepsilon_i(u)$ (resp., $\varepsilon_i{}'(u)$) holds, and ε_i (resp., $\varepsilon_i{}'$) for every $i \in \{0,\ldots,l+1\}$ (resp., $i \in \{0,\ldots,l\}$) is defined as follows:

$$\varepsilon_i \;=\; I=i \wedge U_0=0 \wedge U_0{}'=1 \wedge \bigwedge_{i=1}^{l}(U_i=0 \wedge U_i{}'=0)\,,$$

$$\varepsilon_i{}' \;=\; I=i \wedge U_0=0 \wedge U_0{}'=0 \wedge \bigwedge_{i=1}^{l}(U_i=0 \wedge U_i{}'=0)\,.$$

We define $M=(U,V,F)$ as follows. The endogenous variables are given by $V = B \cup B' \cup C \cup \{X_0, X_0{}', E', Y\}$, where $B'=\{B_1{}', \ldots, B_l{}'\}$ and $D(S)=\{0,1\}$ for all $S \in V$. Let

$$\alpha \;=\; X_0=0 \vee \gamma'\,,$$

$$\phi_1' \;=\; (\varepsilon_0 \vee \varepsilon_0{}' \rightarrow (X_0=0 \wedge \bigwedge_{i=1}^{l} B_i \neq B_i{}') \vee (\bigvee_{i=1}^{l}(B_i=1 \wedge B_i{}'=1)) \vee E'=0)\,,$$

$$\phi_2' \;=\; \bigwedge_{i=1}^{l}(\varepsilon_i \vee \varepsilon_i{}' \rightarrow B_i=0 \vee B_i{}'=0)\,,$$

$$\phi_3' \;=\; (\varepsilon_{l+1} \rightarrow (\alpha \wedge \bigwedge_{i=1}^{l} B_i \neq B_i{}') \vee (\bigvee_{i=1}^{l}(B_i=1 \wedge B_i{}'=1)) \vee E'=0)\,,$$

where γ' is obtained from γ by replacing each $S \in B \cup C$ by "$S=1$". We are now

ready to define the functions $F = \{F_S \mid S \in V\}$ as follows:

- $F_{B_i} = U_i$ and $F_{B_i'} = U_i'$ for all $i \in \{1, \ldots, l\}$,
- $F_{X_0} = U_0$ and $F_{X_0'} = U_0'$,
- $F_S = 0$ for all $S \in C \cup \{E'\}$,
- $F_Y = 1$ iff $\phi_1' \vee \phi_2' \vee \phi_3'$ is true.

Let $X = B \cup B' \cup \{X_0, X_0'\}$. Let ϕ be $Y = 1$. Notice that ϕ is primitive. For every truth assignment τ to the variables in B, we denote by $[B/\tau(B)]$ the substitution $[B_1/\tau(B_1), \ldots, B_l/\tau(B_l)]$, and we define $\alpha^\tau = \alpha\,[B/\tau(B)]$. Let $x_0 = 0$, and let $u \in D(U)$ with $X_0(u) = x_0$. Then, $X_0 = x_0$ is a weak cause of α^τ under u iff $\exists C\, \neg\gamma\,[B/\tau(B)]$ is valid. That is, $X_0 = x_0$ is not a weak cause of α^τ under u iff $\forall C\, \gamma\,[B/\tau(B)]$ is valid. Thus, Proposition 2.5 implies the following fact:

($*$) For every $X' \subseteq B \cup B' \cup \{X_0, X_0'\}$ with $X_0 \in X'$, it holds that $X' = X'(u)$ is not a weak cause of α^τ under u iff $\forall C\, \gamma\,[B/\tau(B)]$ is valid.

Using ($*$), by a similar line of argumentation as in the proof of Theorem 3.3, it follows that Φ is valid iff some $X' \subseteq X$ and $x' \in D(X')$ exist such that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$. □

B Appendix: Proofs for Section 4

Proof of Theorem 4.3 (continued). We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, ϕ, $\mathcal{C} \subseteq D(U)$, P, and α as in the statement of the theorem, such that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff the number of valid formulas among $\Phi_1, \ldots, \Phi_k$ is even.

For $i \in \{1, \ldots, k\}$, define the causal models $M_i = (U_i, V_i, F_i)$ as follows. The exogenous and endogenous variables are defined by $U_i = \{E_i\}$ and $V_i = A_i \cup B_i \cup \{C_i, G_i\}$, respectively. Define $D(S) = \{0, 1, 2\}$ for all $S \in B_i$, and $D(S) = \{0, 1\}$ for all $S \in U_i \cup V_i \setminus B_i$. We define

$$\phi_i = (\gamma_i' \wedge \bigwedge_{S \in B_i} S \neq 2) \vee (C_i = 0) \vee (G_i = 1 \wedge C_i = 1 \wedge \bigvee_{S \in B_i} S \neq 2),$$

where γ_i' is obtained from γ_i by replacing each $S \in A_i \cup B_i$ by "$S = 1$". The functions in $F_i = \{F_S^i \mid S \in V_i\}$ are defined as follows:

- $F_{G_i}^i = E_i$,
- $F_S^i = 0$ for all $S \in \{C_i\} \cup A_i$,
- $F_S^i = G_i + C_i$ for all $S \in B_i$.

For each $i \in \{1, \ldots, k\}$, let $X_i = \{G_i\}$, and define $x_i \in D(X_i)$ and $u_i \in D(U_i)$ by $x_i(G_i) = 0$ and $u_i(E_i) = 0$. Then, for every $i \in \{1, \ldots, k\}$, $X_i = x_i$ is a weak cause of ϕ_i under u_i in M_i iff Φ_i is valid (the construction is similar as in the proof of Theorem 3.2, the only difference is that we have $F^i_{G_i} = E_i$ here, instead of $F^i_{G_i} = 0$). Observe also that $\phi_i(u)$ holds for all $u \in D(U_i)$.

Define the causal model $M = (U, V, F)$ by $U = U_1 \cup \cdots \cup U_k \cup \{E\}$, where $D(E) = \{0, \ldots, k\}$, $V = V_1 \cup \cdots \cup V_k \cup \{H\}$, and $F = F_1 \cup \cdots \cup F_k \cup \{F_H\}$, where

$$F_H = 1 \text{ iff } \Big(\bigwedge_{i \in \{1,\ldots,k\}} \varepsilon_i \rightarrow \phi_i \Big) \wedge \Big(\bigwedge_{\substack{i \in \{1,\ldots,k\}, \\ i \text{ even}}} \varepsilon'_i \rightarrow \phi_{i-1} \Big) \wedge \Big(\bigwedge_{\substack{i \in \{1,\ldots,k\}, \\ i \text{ odd}}} \varepsilon'_i \rightarrow \top \Big) \text{ is true,}$$

and ε_i and ε'_i are defined as follows for every $i \in \{1, \ldots, k\}$:

$$\varepsilon_i = (E = i) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\}} (E_j = 0) \Big),$$

$$\varepsilon'_i = (E = 0) \wedge (E_i = 1) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\} \setminus \{i\}} (E_j = 0) \Big).$$

For every $i \in \{1, \ldots, k\}$, let u_i (resp., u'_i) be the unique $u \in D(U)$ such that $\varepsilon_i(u)$ (resp., $\varepsilon'_i(u)$). Let $Y = \{H\}$, and let ϕ be $Y = 1$. Let $\mathcal{C} = \{u_1, \ldots, u_k, u'_1, \ldots, u'_k\}$, $P(u) = 1 \,/\, 2k$ for all $u \in \mathcal{C}$, and $\alpha = 1 \,/\, 2k$. Define $X = \{G_1, \ldots, G_k\}$ and $x = x_1 \cdots x_k$ $(= 0 \cdots 0)$.

Observe that ϕ is primitive, P is the uniform distribution over $\mathcal{C}$, and $\phi(u)$ for all $u \in \mathcal{C}$. By Proposition 2.5, the following holds for all $i \in \{1, \ldots, k\}$, all $X' \subseteq X$, and $x' = x|X'$:

(i) If $X_i \subseteq X'$, then $X' = x'$ is a weak cause of ϕ under u_i iff Φ_i is valid.

(ii) If i is even and $X_{i-1} \subseteq X'$, then $X' = x'$ is a weak cause of ϕ under u'_i iff Φ_{i-1} is valid.

(iii) If i is odd, then $X' = x'$ is not a weak cause of ϕ under u'_i.

(iv) If $X_i \not\subseteq X'$, then $X' = x'$ is not a weak cause of ϕ under u_i.

By Proposition 4.1, $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (a) $X(u) \neq x$, or (b) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. By (i), it thus follows $\mathcal{C}^{\phi}_{X=x} = \{u'_1, \ldots, u'_k\} \cup \{u_i \mid i \in \{1, \ldots, k\},\ \Phi_i \text{ is valid}\}$.

We now show that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff the number of valid formulas among $\Phi_1, \ldots, \Phi_k$ is even.

($\Rightarrow$) Assume that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$. In particular, $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$. Towards a contradiction, assume that the number of valid formulas among $\Phi_1, \ldots, \Phi_k$ is odd. Let $j \in \{1, \ldots, k\}$ be the smallest index such that Φ_j is not valid. Notice that j is even. We define

$X' = X \setminus X_j$ and $x' = x|X'$. The set of all $u \in \mathcal{C}^{\phi}_{X=x}$ such that $X'(u) = x'$ is given by $\mathcal{C}' = \{u'_j\} \cup \{u_1, \ldots, u_{j-1}\}$ (as Φ_j implies Φ_{j-1}, for every $j \in \{2, \ldots, k\}$). By (i) and (ii), $X' = x'$ is a weak cause of ϕ under every $u \in \mathcal{C}'$. That is, $X' = x'$ is a weak cause of ϕ under every $u \in \mathcal{C}^{\phi}_{X=x}$ with $X'(u) = x'$, which violates EX3, and thus contradicts $X = x$ being an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$.

($\Leftarrow$) Assume that the number of valid formulas among $\Phi_1, \ldots, \Phi_k$ is even. We now show that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$. By Proposition 4.1, it is sufficient to show that (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$, and (b) $P(\mathcal{C}^{\phi}_{X=x} \mid X = x) \geq \alpha$. We first prove (a) by showing that EX1–EX4 hold. Clearly, EX1 and EX2 hold. Observe that ${u_1}' \in \mathcal{C}^{\phi}_{X=x}$. As Φ_1 is valid, we also have $u_1 \in \mathcal{C}^{\phi}_{X=x}$. Hence, as $X(u'_1) \neq x$ and $X(u_1) = x$, also EX4 holds. We next show that EX3 holds. Towards a contradiction, assume that some $X' \subset X$ exists such that $X' = x'$ is a weak cause of ϕ under all $u \in \mathcal{C}^{\phi}_{X=x}$ with $X'(u) = x'$, where $x' = x|X'$. Let $X_j \in X \setminus X'$ such that $j \in \{1, \ldots, k\}$ is minimal. As $u'_j \in \mathcal{C}^{\phi}_{X=x}$ and $X'(u'_j) = x'$, it follows that $X' = x'$ is a weak cause of ϕ under u'_j. By (iii), j is even. By (ii), Φ_{j-1} is valid. By (iv), u_j does not belong to $\mathcal{C}^{\phi}_{X=x}$. That is, Φ_j is not valid. But this contradicts the number of valid formulas among $\Phi_1, \ldots, \Phi_k$ being even. Thus, also EX3 holds. Clearly, (b) follows from EX4 and P being the uniform distribution over $\mathcal{C}$. $\square$

Proof of Theorem 4.6 (continued). We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, ϕ, $\mathcal{C} \subseteq D(U)$, and P as required, such that $(v_1, \ldots, v_k)$ is the bit-vector representation of the explanatory power of $X = x$ multiplied with $\sum_{j=0}^{k-1} 2^j$.

For every $i \in \{1, \ldots, k\}$, define $M_i = (U_i, V_i, F_i)$ and $X_i \subseteq V_i$ as in the proof of Theorem 4.3. We define $M = (U, V, F)$ by $U = U_1 \cup \cdots \cup U_k \cup \{E\}$, where $D(E) = \{0, \ldots, k\}$, $V = V_1 \cup \cdots \cup V_k \cup \{H\}$, and $F = F_1 \cup \cdots \cup F_k \cup \{F_H\}$, where

$$F_H = 1 \text{ iff } \Big(\bigwedge_{i \in \{1,\ldots,k\}} \varepsilon_i \rightarrow \phi_i \Big) \wedge \Big(\bigwedge_{i \in \{1,\ldots,k\}} \varepsilon'_i \rightarrow \top \Big) \text{ is true,}$$

and ε_i and ε'_i are defined as follows for every $i \in \{1, \ldots, k\}$:

$$\varepsilon_i = (E = i) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\}} (E_j = 0) \Big),$$

$$\varepsilon'_i = (E = 0) \wedge (E_i = 1) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\} \setminus \{i\}} (E_j = 0) \Big).$$

For every $i \in \{1, \ldots, k\}$, let u_i (resp., u'_i) be the unique $u \in D(U)$ such that $\varepsilon_i(u)$ (resp., $\varepsilon'_i(u)$). Let $Y = \{H\}$, and let ϕ be $Y = 1$. We define $\mathcal{C} = \{u_1, \ldots, u_k, u'_1, \ldots, u'_k\}$, $P(u'_i) = 0$ for all $i \in \{1, \ldots, k\}$, and $P(u_i) = 2^{i-1} \,/\, \sum_{j=0}^{k-1} 2^j$ for all $i \in \{1, \ldots, k\}$. We define $X = \{G_1, \ldots, G_k\}$ and $x = x_1 \cdots x_k$.

Observe that ϕ is primitive. Moreover, $\phi(u)$ for all $u \in \mathcal{C}$, and for all $i \in \{1, \ldots, k\}$:

(i) $X = x$ is a weak cause of ϕ under u_i iff Φ_i is valid.

(ii) If $X' \subseteq X$ and $x' = x|X'$, then $X' = x'$ is not a weak cause of ϕ under u_i'.

By Proposition 4.1, $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (a) $X(u) \neq x$, or (b) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. By (i), it thus follows $\mathcal{C}^{\phi}_{X=x} = \{u_1', \ldots, u_k'\} \cup \{u_i \mid i \in \{1, \ldots, k\},\ \Phi_i \text{ is valid}\}$. By (ii), $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$. Thus, $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$. Its explanatory power is the sum of all $P(u_i) = 2^{i-1} \,/\, \sum_{j=0}^{k-1} 2^j$ with $i \in \{1, \ldots, k\}$ such that Φ_i is valid. □

Theorem B.1 *α-Partial Explanation is $\mathrm{P}^{\mathrm{NP}}_{\|}$-complete in the binary case.*

Proof. As for membership, recall that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$, and (b) $P(\mathcal{C}^{\phi}_{X=x} \mid X = x) \geq \alpha$. By Proposition 4.1, $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (i) $X(u) \neq x$, or (ii) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. Deciding (i) is polynomial, and, by Theorem 2.6, deciding (ii) is in NP in the binary case. Thus, computing $\mathcal{C}^{\phi}_{X=x}$ is in $\mathrm{FP}^{\mathrm{NP}}_{\|}$ in the binary case. Once $\mathcal{C}^{\phi}_{X=x}$ is given, deciding (a) is possible with two NP-oracle calls, by Theorem 3.4, and deciding (b) is polynomial. As two rounds of parallel NP-oracle queries in a polynomial-time computation can be replaced by a single one [3], the problem is in $\mathrm{P}^{\mathrm{NP}}_{\|}$.

Hardness for $\mathrm{P}^{\mathrm{NP}}_{\|}$ is shown by a reduction from the following $\mathrm{P}^{\mathrm{NP}}_{\|}$-complete problem [51]. Given k propositional formulas γ_i, $i \in \{1, \ldots, k\}$, where each γ_i is defined on the variables $A_i = \{A_{i,1}, \ldots, A_{i,m_i}\}$, decide whether the number of tautologies among $\gamma_1, \ldots, \gamma_k$ is even. Without loss of generality, k is even, the A_i's are pairwise disjoint, γ_1 is not a tautology, and for every $j \in \{1, \ldots, k-1\}$, if γ_j is a tautology, then also γ_{j+1} [51]. We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\phi, \mathcal{C} \subseteq D(U)$, P, and α as required, such that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff the number of tautologies among $\gamma_1, \ldots, \gamma_k$ is even. The construction is similar to the one in the proof of Theorem 4.3. Roughly, we replace the part for Σ^P_2-hardness of deciding general weak cause by a new part for NP-hardness of deciding binary weak cause.

For $i \in \{1, \ldots, k\}$, define the causal models $M_i = (U_i, V_i, F_i)$ as follows. The exogenous and endogenous variables are defined by $U_i = \{E_i\}$ and $V_i = A_i \cup \{G_i\}$, respectively, where $D(S) = \{0, 1\}$ for all $S \in U_i \cup V_i$. We define the functions in $F_i = \{F^i_S \mid S \in V_i\}$ as follows:

- $F^i_{G_i} = E_i$,
- $F^i_S = 0$ for all $S \in A_i$.

We then define $\phi_i = G_i = 0 \vee \gamma_i{}'$, where $\gamma_i{}'$ is obtained from γ_i by replacing each $S \in A_i$ by "$S = 1$". For each $i \in \{1, \ldots, k\}$, let $X_i = \{G_i\}$, and define $x_i \in D(X_i)$

and $u_i \in D(U_i)$ by $x_i(G_i) = 0$ and $u_i(E_i) = 0$. Then, for every $i \in \{1, \ldots, k\}$, $X_i = x_i$ is a weak cause of ϕ_i under u_i in M_i iff γ_i is not a tautology.

We define the causal model $M = (U, V, F)$ as follows. The exogenous and endogenous variables are given by $U = U_1 \cup \cdots \cup U_k \cup \{E\}$ and $V = V_1 \cup \cdots \cup V_k \cup \{H\}$, respectively, where $D(E) = \{0, \ldots, k\}$ and $D(H) = \{0, 1\}$. The functions are given by $F = F_1 \cup \cdots \cup F_k \cup \{F_H\}$, where

$$F_H = 1 \text{ iff } \Big(\bigwedge_{i \in \{1,\ldots,k\}} \varepsilon_i \rightarrow \phi_i \Big) \wedge \Big(\bigwedge_{\substack{i \in \{1,\ldots,k\}, \\ i \text{ even}}} \varepsilon'_i \rightarrow \phi_{i-1} \Big) \wedge \Big(\bigwedge_{\substack{i \in \{1,\ldots,k\}, \\ i \text{ odd}}} \varepsilon'_i \rightarrow \top \Big) \text{ is true,}$$

and ε_i and ε'_i are defined as follows for every $i \in \{1, \ldots, k\}$:

$$\varepsilon_i = (E = i) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\}} (E_j = 0) \Big),$$

$$\varepsilon'_i = (E = 0) \wedge (E_i = 1) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\} \setminus \{i\}} (E_j = 0) \Big).$$

For every $i \in \{1, \ldots, k\}$, let u_i (resp., u'_i) be the unique $u \in D(U)$ such that $\varepsilon_i(u)$ (resp., $\varepsilon'_i(u)$). Let $Y = \{H\}$, and let ϕ be $Y = 1$. Let $\mathcal{C} = \{u_1, \ldots, u_k, u'_1, \ldots, u'_k\}$, $P(u) = 1 \,/\, 2k$ for all $u \in \mathcal{C}$, and $\alpha = 1 \,/\, 2k$. Define $X = \{G_1, \ldots, G_k\}$ and $x = x_1 \ldots x_k$. Observe that ϕ is primitive, that $\phi(u)$ for all $u \in \mathcal{C}$, and that P is the uniform distribution over $\mathcal{C}$. By Proposition 4.1, $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (a) $X(u) \neq x$, or (b) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. By Proposition 2.5, it thus follows $\mathcal{C}^{\phi}_{X=x} = \{u'_1, \ldots, u'_k\} \cup \{u_i \mid i \in \{1, \ldots, k\}, \gamma_i \text{ is not a tautology}\}$.

By a line of argumentation similar to the one in the proof of Theorem 4.3, it follows that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff the number of non-tautologies among $\gamma_1, \ldots, \gamma_k$ is even, that is, as k is even, iff the number of tautologies among $\gamma_1, \ldots, \gamma_k$ is even. □

Theorem B.2 *α-Partial Explanation Existence is Σ^P_2-complete in the binary case.*

Proof. As for membership in Σ^P_2, by Theorem B.1, deciding whether $X' = x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ is in $\mathrm{P}^{\mathrm{NP}}_{\|}$ in the binary case. Thus, guessing some $X' \subseteq X$ and $x' \in D(X')$, and deciding whether $X' = x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ is in Σ^P_2 in the binary case.

Hardness for Σ^P_2 is shown by a reduction from Explanation Existence in the binary case (see Theorem 3.5). Given an instance of it, let P be the uniform distribution on $\mathcal{C}$, and let $\alpha = 1$. Then, $X' = x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$. □

Theorem B.3 *Partial Explanation is $\mathrm{P}^{\mathrm{NP}}_{\|}$-complete in the binary case.*

Proof. As for membership in $\mathrm{P}_{\|}^{\mathrm{NP}}$, recall that $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$, and (b) $\mathcal{C}^{\phi}_{X=x}$ contains some u such that $X(u) = x$ and $P(u) > 0$. By the proof of Theorem B.1, computing $\mathcal{C}^{\phi}_{X=x}$ is in $\mathrm{FP}_{\|}^{\mathrm{NP}}$ in the binary case. Once $\mathcal{C}^{\phi}_{X=x}$ is given, checking (a) is in D^{P} in the binary case, by Theorem 3.4, and checking (b) is polynomial. As two rounds of parallel NP-oracle queries in a polynomial-time computation can be replaced by a single one [3], Partial Explanation is in $\mathrm{P}_{\|}^{\mathrm{NP}}$ in the binary case.

We next show $\mathrm{P}_{\|}^{\mathrm{NP}}$-hardness. If P is the uniform distribution over $\mathcal{C}$, then $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff $X = x$ is a $\frac{1}{|\mathcal{C}|}$-partial explanation of ϕ relative to $(\mathcal{C}, P)$. By the proof of Theorem B.1, deciding the latter is complete for $\mathrm{P}_{\|}^{\mathrm{NP}}$. Thus, deciding whether $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ is $\mathrm{P}_{\|}^{\mathrm{NP}}$-hard, and hardness holds even if P is the uniform distribution over $\mathcal{C}$. □

Theorem B.4 *Explanatory Power is* $\mathrm{FP}_{\|}^{\mathrm{NP}}$*-complete in the binary case.*

Proof. We compute $\mathcal{C}^{\phi}_{X=x}$ and $P(\mathcal{C}^{\phi}_{X=x} \mid X = x)$. By the proof of Theorem B.1, the former is in $\mathrm{FP}_{\|}^{\mathrm{NP}}$ in the binary case, while the latter is polynomial. Thus, Explanatory Power is in $\mathrm{FP}_{\|}^{\mathrm{NP}}$ in the binary case.

Hardness for $\mathrm{FP}_{\|}^{\mathrm{NP}}$ is shown by a reduction from the following $\mathrm{FP}_{\|}^{\mathrm{NP}}$-complete problem. Given k propositional formulas γ_i, $i \in \{1, \ldots, k\}$, where each γ_i is defined on the variables $A_i = \{A_{i,1}, \ldots, A_{i,m_i}\}$, compute the vector $(v_1, \ldots, v_k) \in \{0,1\}^k$ such that $v_i = 1$ iff γ_i is not a tautology, for all $i \in \{1, \ldots, k\}$. Without loss of generality, the A_i's are pairwise disjoint, and γ_1 is not a tautology.

We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, ϕ, $\mathcal{C} \subseteq D(U)$, and P as required, such that $(v_1, \ldots, v_k)$ is the bit-vector representation of the explanatory power of $X = x$ multiplied with $\sum_{j=0}^{k-1} 2^j$. For every $i \in \{1, \ldots, k\}$, let $M_i{=}(U_i, V_i, F_i)$ and $X_i \subseteq V_i$ be defined as in the proof of Theorem B.1. The rest of the construction is similar as in the proof of Theorem 4.6. We define the causal model $M = (U, V, F)$ as follows. The exogenous and endogenous variables are given by $U = U_1 \cup \cdots \cup U_k \cup \{E\}$ and $V = V_1 \cup \cdots \cup V_k \cup \{H\}$, respectively, where $D(E) = \{0, \ldots, k\}$ and $D(H) = \{0, 1\}$. The functions are given by $F = F_1 \cup \cdots \cup F_k \cup \{F_H\}$, where

$$F_H = 1 \text{ iff } \Big(\bigwedge_{i \in \{1,\ldots,k\}} \varepsilon_i \rightarrow \phi_i \Big) \wedge \Big(\bigwedge_{i \in \{1,\ldots,k\}} \varepsilon'_i \rightarrow \top \Big) \text{ is true,}$$

and ε_i and ε'_i are defined as follows for every $i \in \{1, \ldots, k\}$:

$$\varepsilon_i = (E = i) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\}} (E_j = 0) \Big),$$

$$\varepsilon'_i = (E = 0) \wedge (E_i = 1) \wedge \Big(\bigwedge_{j \in \{1,\ldots,k\} \setminus \{i\}} (E_j = 0) \Big).$$

For every $i \in \{1, \ldots, k\}$, let u_i (resp., u_i') be the unique $u \in D(U)$ such that $\varepsilon_i(u)$ (resp., $\varepsilon_i'(u)$). Let $Y = \{H\}$, and let ϕ be $Y = 1$. We define $\mathcal{C} = \{u_1, \ldots, u_k, u_1', \ldots, u_k'\}$, $P(u_i') = 0$ for all $i \in \{1, \ldots, k\}$, and $P(u_i) = 2^{i-1} \,/\, \sum_{j=0}^{k-1} 2^j$ for all $i \in \{1, \ldots, k\}$. We define $X = \{G_1, \ldots, G_k\}$ and $x = x_1 \cdots x_k$.

Observe that ϕ is primitive. Moreover, $\phi(u)$ for all $u \in \mathcal{C}$, and for all $i \in \{1, \ldots, k\}$:

(i) $X = x$ is a weak cause of ϕ under u_i iff γ_i is not a tautology.

(ii) If $X' \subseteq X$ and $x' = x|X'$, then $X' = x'$ is not a weak cause of ϕ under u_i'.

By Proposition 4.1, $\mathcal{C}^{\phi}_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (a) $X(u) \neq x$, or (b) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. By (i), it thus follows $\mathcal{C}^{\phi}_{X=x} = \{u_1', \ldots, u_k'\} \cup \{u_i \mid i \in \{1, \ldots, k\},\ \gamma_i \text{ is not a tautology}\}$. By (ii), $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\phi}_{X=x}$. Hence, $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$. The explanatory power of $X = x$ is the sum of all $P(u_i) = 2^{i-1} \,/\, \sum_{j=0}^{k-1} 2^j$ with $i \in \{1, \ldots, k\}$ such that γ_i is not a tautology. □

C Appendix: Proofs for Section 5

Proof of Theorem 5.1 (continued). Hardness for Π_4^P is shown by a reduction from the Π_4^P-complete problem of deciding whether a given QBF $\Phi = \forall A\, \exists B\, \forall C\, \exists D\, \gamma$ is valid, where γ is a propositional formula on the variables $A = \{A_1, \ldots, A_k\}$, $B = \{B_1, \ldots, B_l\}$, $C = \{C_1, \ldots, C_m\}$, and $D = \{D_1, \ldots, D_n\}$. We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, and ϕ as in the statement of the theorem such that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff Φ is valid.

We define the exogenous variables by $U = B \cup \{U_0, U_1, {U_1}', \ldots, U_k, {U_k}'\}$, where $D(S) = \{0, 1\}$ for all $S \in U$. We define the set of contexts by $\mathcal{C} = \{u \in D(U) \mid (\varepsilon_0 \vee \varepsilon_1 \vee \varepsilon_2)(u)\}$, where:

$$\varepsilon_0 \;=\; U_0 = 0 \wedge \bigwedge_{i=1}^{k} (U_i = 0 \wedge {U_i}' = 0)\,,$$

$$\varepsilon_1 \;=\; U_0 = 0 \wedge \bigvee_{i=1}^{k} \Big(((U_i = 1 \wedge {U_i}' = 0) \vee (U_i = 0 \wedge {U_i}' = 1)) \wedge \bigwedge_{j \in \{1,\ldots,k\} - \{i\}} (U_j = 0 \wedge {U_j}' = 0) \Big)\,,$$

$$\varepsilon_2 \;=\; U_0 = 1 \vee \bigvee_{i=1}^{k} (U_i = 1 \wedge {U_i}' = 1)\,.$$

We define $M = (U, V, F)$ as follows. We define $V = A \cup A' \cup C \cup D \cup \{X_0, E, E', Y\}$, where $A' = \{{A_1}', \ldots, {A_k}'\}$, $D(S) = \{0, 1, 2\}$ for all $S \in D$, and $D(S) = \{0, 1\}$ for

all $S \in V \setminus D$. Let

$$\alpha \;=\; (\neg\gamma' \wedge \bigwedge_{S\in D} S \neq 2) \vee (E=0) \vee (X_0 = 1 \wedge E = 1 \wedge \bigvee_{S\in D} S \neq 2)\,,$$

$$\phi' \;=\; (\varepsilon_0 \rightarrow X_0 = 0) \wedge (\varepsilon_2 \rightarrow \top)$$
$$\wedge\; (\varepsilon_1 \rightarrow (\alpha \wedge \bigwedge_{i=1}^{k} A_i \neq A_i{}') \vee (\bigvee_{i=1}^{k} (A_i = 1 \wedge A_i{}' = 1)) \vee E' = 0)\,,$$

where γ' is obtained from γ by replacing each $S \in A \cup B \cup C \cup D$ by "$S = 1$". We are now ready to define the functions $F = \{F_S \mid S \in V\}$ as follows:

- $F_{A_i} = U_i$ and $F_{A_i{}'} = U_i{}'$ for all $i \in \{1,\ldots,k\}$,
- $F_{X_0} = U_0$, and $F_S = 0$ for all $S \in C \cup \{E, E'\}$,
- $F_S = X_0 + E$ for all $S \in D$,
- $F_Y = 1$ iff ϕ' is true.

Let $X = A \cup A' \cup \{X_0\}$, and let $x \in D(X)$ be given by $x(S) = 0$ for all $S \in X$. Let ϕ be $Y = 1$. Notice that ϕ is primitive. We now show that Φ is valid iff $X = x$ is an explanation of ϕ relative to $\mathcal{C}$.

We first show that EX1, EX2, and EX4 always hold. As $\phi(u)$ for all $u \in \mathcal{C}$, EX1 always holds. For every $u \in \mathcal{C}$ with $X(u) = x$, it holds $\varepsilon_0(u)$. Hence, $X = x$ is a weak cause of ϕ under every $u \in \mathcal{C}$ with $X(u) = x$. That is, also EX2 always holds. As some $u, u' \in \mathcal{C}$ exist such that $X(u) = x$ and $X(u') \neq x$, also EX4 always holds. It thus remains to show that Φ is valid iff EX3 holds. Recall that EX3 says that for every $X' \subset X$, some $u \in \mathcal{C}$ exists such that (i) $X'(u) = x|X'$ and (ii) $X = x|X'$ is not a weak cause of ϕ under u. If $X_0 \notin X'$ or $X' \cap \{A_i, A_i{}'\} = \emptyset$ for some $i \in \{1,\ldots,k\}$, then (i) and (ii) hold for some $u \in \mathcal{C}$ with $\varepsilon_2(u)$. If $X_0 \in X'$, $X' \cap \{A_i, A_i{}'\} \neq \emptyset$ for all $i \in \{1,\ldots,k\}$, and $A_i, A_i{}' \in X'$ for some $i \in \{1,\ldots,k\}$, then (i) and (ii) hold for some $u \in \mathcal{C}$ with $\varepsilon_1(u)$.

It thus remains to show that Φ is valid iff for every $X' \subset X$ such that (a) $X_0 \in X'$ and (b) $|X' \cap \{A_i, A_i{}'\}| = 1$ for all $i \in \{1,\ldots,k\}$, some $u \in \mathcal{C}$ exists such that (i) $X'(u) = x|X'$ and (ii) $X' = x|X'$ is not a weak cause of ϕ under u.

For all truth assignments σ and τ to the variables in A and B, respectively, denote by $[A/\sigma(A), B/\tau(B)]$ the substitution $[A_1/\sigma(A_1),\ldots,A_k/\sigma(A_k), B_1/\tau(B_1),\ldots, B_l/\tau(B_l)]$, and we define $\alpha^{\sigma,\tau} = \alpha\,[A/\sigma(A),\ B/\tau(B)]$. Let $x_0 = 0$, and let $u \in D(U)$ such that $X_0(u) = x_0$. Then, $X_0 = x_0$ is a weak cause of $\alpha^{\sigma,\tau}$ under u iff $\exists C\,\forall D\,\neg\gamma\,[A/\sigma(A), B/\tau(B)]$ is valid [14,18]. That is, $X_0 = x_0$ is not a weak cause of $\alpha^{\sigma,\tau}$ under u iff $\forall C\,\exists D\,\gamma\,[A/\sigma(A), B/\tau(B)]$ is valid. Thus, Proposition 2.5 implies the following fact:

$(*)$ For every $X' \subseteq A \cup A' \cup \{X_0\}$ with $X_0 \in X'$, it holds that $X' = X'(u)$ is not

a weak cause of $\alpha^{\sigma,\tau}$ under u iff $\forall C\, \exists D\, \gamma\,[A/\sigma(A), B/\tau(B)]$ is valid.

($\Rightarrow$) Assume that Φ is valid. Let $X' \subset X$ such that (a) and (b) holds. Define the truth assignment σ to the variables in A by $\sigma(A_i) = 0$ iff $A_i \in X'$ for all $i \in \{1, \ldots, k\}$. As Φ is valid, there exists a truth assignment τ to the variables in B such that $\forall C\, \exists D\, \gamma\,[A/\sigma(A), B/\tau(B)]$ is valid. Let $x' = x|X'$, and let $u \in D(U)$ be arbitrary such that $X'(u) = x'$, $\varepsilon_1(u)$, and $u(B_i) = \tau(B_i)$ for all $i \in \{1, \ldots, l\}$. By ($*$), $X' = x'$ is not a weak cause of $\alpha^{\sigma,\tau}$ under u. We now show that $X' = x'$ is also not a weak cause of ϕ under u. Towards a contradiction, assume the contrary. Thus, some $W \subseteq V \setminus X'$, $\overline{x}' \in D(X')$, and $w \in D(W)$ exist such that $\neg\phi_{\overline{x}'w}(u)$ and $\phi_{x'w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. As $\neg\phi_{\overline{x}'w}(u)$, it follows that $F \in W$ and $w(F) = 1$. As $\phi_{x'w}(u)$, for every $S \in (A \cup A') \setminus X'$, it holds either $S(u) = 1$ or $S \in W$ and $w(S) = 1$. As $\neg\phi_{\overline{x}'w}(u)$, it thus follows that $\overline{x}'(S) = 0$ for all $S \in X' \setminus \{X_0\}$. Hence, $\neg\alpha_{\overline{x}'w}(u)$ and $\alpha_{x'w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. That is, $\neg\alpha^{\sigma,\tau}_{\overline{x}'w}(u)$ and $\alpha^{\sigma,\tau}_{x'w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. As $X'(u) = x'$ and $\alpha^{\sigma,\tau}(u)$, this shows that $X' = x'$ is a weak cause of $\alpha^{\sigma,\tau}$ under u. Equivalently, by ($*$), $\forall C\, \exists D\, \gamma\,[A/\sigma(A), B/\tau(B)]$ is not valid, which is a contradiction. This shows that $X' = x'$ is not a weak cause of ϕ under u.

($\Leftarrow$) Assume that Φ is not valid. That is, there is a truth assignment σ to the variables in A such that for every truth assignment τ to the variables in B, it holds that $\forall C\, \exists D\, \gamma\,[A/\sigma(A), B/\tau(B)]$ is not valid. Let $X' = \{X_0\} \cup \{S \in A \mid \sigma(S) = 0\} \cup \{S' \in A' \mid \sigma(S) = 1\}$, and let $x' = x|X'$. Let $u \in \mathcal{C}$ be any context such that $X'(u) = x'$. We now show that $X' = x'$ is a weak cause of ϕ under u. If $\varepsilon_0(u)$, then $X' = x'$ is trivially a weak cause of ϕ under u. Assume now $\varepsilon_1(u)$. Let τ be the truth assignment to the variables in B with $u(B_i) = \tau(B_i)$ for all $i \in \{1, \ldots, l\}$. As $\forall C\, \exists D\, \gamma\,[A/\sigma(A), B/\tau(B)]$ is not valid, by ($*$), $X' = x'$ is a weak cause of $\alpha^{\sigma,\tau}$ under u. Thus, some $W \subseteq V \setminus X'$, $\overline{x}' \in D(X')$, and $w \in D(W)$ exist such that $\neg\alpha^{\sigma,\tau}_{\overline{x}'w}(u)$ and $\alpha^{\sigma,\tau}_{x'w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. Here, we can assume that $\overline{x}'(X_0) = 1$, $\overline{x}'(S) = 0$ for all $S \in X' \setminus \{X_0\}$, $\{F\} \cup ((A \cup A') \setminus X') \subseteq W$, and $w(S) = 1$ for all $S \in \{F\} \cup ((A \cup A') \setminus X')$. Hence, $\neg\alpha_{\overline{x}'w}(u)$ and $\alpha_{x'w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. Thus, $\neg\phi_{\overline{x}'w}(u)$ and $\phi_{x'w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. As $X'(u) = x'$ and $\phi(u)$, it follows that $X' = x'$ is a weak cause of ϕ under u. □

Theorem C.1 *Explanation is Π^P_3-complete for succinct context sets and binary causal models.*

Proof. As for membership in Π^P_3, recall that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff EX1–EX4 hold. As argued in the proof of Theorem 5.1, deciding whether EX1 and EX4 hold is in co-NP and NP, respectively, for succinct context sets. By Theorem 2.6, deciding whether $X = x$ is a weak cause of ϕ under some $u \in D(U)$ is in NP in the binary case. Thus, in EX2, deciding whether $X = x$ is a weak cause of ϕ under every $u \in \mathcal{C}$ with $X(u) = x$ is in Π^P_2 for succinct context sets and binary

causal models. Hence, deciding whether some $X' \subset X$ exists such $X' = x|X'$ is a weak cause of ϕ under every $u \in \mathcal{C}$ with $X'(u) = x|X'$ is in Σ_3^P for succinct context sets and binary causal models. Thus, deciding whether EX3 holds is in Π_3^P. In summary, deciding whether EX1–EX4 hold is in Π_3^P for succinct context sets and binary causal models.

Hardness for Π_3^P is shown by a reduction from the Π_3^P-complete problem of deciding whether a given QBF $\Phi = \forall A\, \exists B\, \forall C\, \gamma$ is valid, where γ is a propositional formula on the variables $A = \{A_1, \ldots, A_k\}$, $B = \{B_1, \ldots, B_l\}$, and $C = \{C_1, \ldots, C_m\}$. We construct $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, and ϕ as required such that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff Φ is valid. The construction is similar to the one in the proof of Theorem 5.1. Roughly, we replace the part for Σ_2^P-hardness of deciding general weak cause by a new part for NP-hardness of deciding binary weak cause.

We define the exogenous variables by $U = B \cup \{U_0, U_1, {U_1}', \ldots, U_k, {U_k}'\}$, where $D(S) = \{0, 1\}$ for all $S \in U$. We define the set of contexts by $\mathcal{C} = \{u \in D(U) \mid (\varepsilon_0 \vee \varepsilon_1 \vee \varepsilon_2)(u)\}$, where:

$$\begin{aligned}
\varepsilon_0 &= U_0 = 0 \wedge \bigwedge_{i=1}^{k} (U_i = 0 \wedge {U_i}' = 0)\,, \\
\varepsilon_1 &= U_0 = 0 \wedge \bigvee_{i=1}^{k} \Big(((U_i = 1 \wedge {U_i}' = 0) \vee (U_i = 0 \wedge {U_i}' = 1)) \\
&\qquad \wedge \bigwedge_{j \in \{1,\ldots,k\} - \{i\}} (U_j = 0 \wedge {U_j}' = 0) \Big)\,, \\
\varepsilon_2 &= U_0 = 1 \vee \bigvee_{i=1}^{k} (U_i = 1 \wedge {U_i}' = 1)\,.
\end{aligned}$$

We define the causal model $M = (U, V, F)$ as follows. The exogenous variables are given by $V = A \cup A' \cup C \cup \{X_0, E', Y\}$, where $A' = \{{A_1}', \ldots, {A_k}'\}$ and $D(S) = \{0, 1\}$ for all $S \in V$. Let

$$\alpha = X_0 = 0 \vee \gamma'\,,$$

$$\begin{aligned}
\phi' = {} & (\varepsilon_0 \rightarrow X_0 = 0) \wedge (\varepsilon_2 \rightarrow \top) \\
& \wedge\ (\varepsilon_1 \rightarrow (\alpha \wedge \bigwedge_{i=1}^{k} A_i \neq {A_i}') \vee (\bigvee_{i=1}^{k} (A_i = 1 \wedge {A_i}' = 1)) \vee E' = 0)\,,
\end{aligned}$$

where γ' is obtained from γ by replacing each $S \in A \cup B \cup C$ by "$S = 1$". We are now ready to define the functions $F = \{F_S \mid S \in V\}$ as follows:

- $F_{A_i} = U_i$ and $F_{{A_i}'} = {U_i}'$ for all $i \in \{1, \ldots, k\}$,
- $F_{X_0} = U_0$, and $F_S = 0$ for all $S \in C \cup \{E'\}$,

- $F_Y = 1$ iff ϕ' is true.

Let $X = A \cup A' \cup \{X_0\}$, and let $x \in D(X)$ be given by $x(S) = 0$ for all $S \in X$. Let ϕ be $Y = 1$. Notice that ϕ is primitive. For all truth assignments σ and τ to the variables in A and B, respectively, we denote by $[A/\sigma(A), B/\tau(B)]$ the substitution $[A_1/\sigma(A_1), \ldots, A_k/\sigma(A_k), B_1/\tau(B_1), \ldots, B_l/\tau(B_l)]$, and we define $\alpha^{\sigma,\tau} = \alpha\,[A/\sigma(A),\, B/\tau(B)]$. Let $x_0 = 0$, and let $u \in D(U)$ such that $X_0(u) = x_0$. Then, $X_0 = x_0$ is a weak cause of $\alpha^{\sigma,\tau}$ under u iff $\exists C\, \neg\gamma\,[A/\sigma(A), B/\tau(B)]$ is valid. That is, $X_0 = x_0$ is not a weak cause of $\alpha^{\sigma,\tau}$ under u iff $\forall C\, \gamma\,[A/\sigma(A),$ $B/\tau(B)]$ is valid. Thus, Proposition 2.5 implies the following fact:

($*$) For every $X' \subseteq A \cup A' \cup \{X_0\}$ with $X_0 \in X'$, it holds that $X' = X'(u)$ is not a weak cause of $\alpha^{\sigma,\tau}$ under u iff $\forall C\, \gamma\,[A/\sigma(A), B/\tau(B)]$ is valid.

Using ($*$), by a line of argumentation similar to the one in the proof of Theorem 5.1, it follows that Φ is valid iff $X = x$ is an explanation of ϕ relative to $\mathcal{C}$. □

Theorem C.2 *Partial Explanation is Π_3^P-complete for succinct context sets and binary causal models.*

Proof. As for membership in Π_3^P, recall that $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}^\phi_{X=x}$, and (b) $X(u) = x$ and $P(u) > 0$ for some $u \in \mathcal{C}^\phi_{X=x}$. By Proposition 4.1, $\mathcal{C}^\phi_{X=x}$ is the set of all $u \in \mathcal{C}$ such that either (i) $X(u) \neq x$, or (ii) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. To check that (a) holds, we check that EX1–EX4 hold. Clearly, EX1 and EX2 always hold. The complement of EX3 says that some $X' \subset X$ exists such that for every $u \in \mathcal{C}$, it holds that $X'(u) = x|X'$ and $u \in \mathcal{C}^\phi_{X=x}$ implies that $X' = x|X'$ is a weak cause of ϕ under u. That is, some $X' \subset X$ exists such that for every $u \in \mathcal{C}$, it holds either (a) $X'(u) \neq x|X'$, or (b) $X(u) = x$ and $X = x$ is not a weak cause of ϕ under u, or (c) $X' = x|X'$ is a weak cause of ϕ under u. By Theorem 2.6, deciding weak cause is in NP in the binary case. Thus, deciding whether EX3 does not hold is in Σ_3^P for succinct context sets and binary causal models. Hence, deciding whether EX3 holds is in Π_3^P. EX4 says that some $u, u' \in \mathcal{C}^\phi_{X=x}$ exist such that $X(u) \neq x$ and $X(u') = x$. Equivalently, some $u, u' \in \mathcal{C}$ exist such that $X(u) \neq x$, and $X(u') = x$ and $X = x$ is a weak cause of ϕ under u'. Thus, deciding whether EX4 holds is in NP in the binary case. In summary, deciding whether (a) holds is in Π_3^P for succinct context sets and binary causal models. Finally, (b) says that some $u \in \mathcal{C}$ exists such that $X(u) = x$, $P(u) > 0$, and $X = x$ is a weak cause of ϕ under u. Thus, checking (b) is in NP in the binary case. In summary, deciding whether (a) and (b) holds is in Π_3^P for succinct context sets and binary causal models.

Hardness for Π_3^P is shown a reduction from the Π_3^P-complete problem of deciding whether a given QBF $\Phi = \forall A\, \exists B\, \forall C\, \gamma$ is valid, where γ is a propositional formula on the variables $A = \{A_1, \ldots, A_k\}$, $B = \{B_1, \ldots, B_l\}$, and $C = \{C_1, \ldots, C_m\}$. We define $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, ϕ, and $\mathcal{C} \subseteq D(U)$ as in the proof of The-

orem C.1, and let P be the uniform distribution over $\mathcal{C}$. Observe that ϕ is primitive and that $\phi(u)$ holds for all $u \in \mathcal{C}$. For every $u \in \mathcal{C}$, either (i) $X(u) \neq x$, or (ii) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u. By Proposition 4.1, $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff (a) $X = x$ is an explanation of ϕ relative to $\mathcal{C}$, and (b) $\mathcal{C}$ contains some u such that $X(u) = x$ and $P(u) > 0$. Here, (a) implies (b). By the proof of Theorem C.1, $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff Φ is valid. In summary, $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff Φ is valid. □

D Appendix: Proofs for Section 6

Proof of Theorem 6.4 (continued). It remains to prove that Φ is not valid iff $(*)$ some $u \in D(U)$ exists such that for every $u_1 \in D(U_1)$, there exists a causal formula $[Y \leftarrow y]\, X = x$, where $Y \subseteq V$ and $X \in V$, such that (i) $(M, u) \not\models [Y \leftarrow y]\, X = x$ and (ii) $(M_1, u_1) \models [Y \leftarrow y]\, X = x$.

$(\Rightarrow)$ Suppose that Φ is not valid. Let then τ be any truth assignment to B such that $\exists C \forall D \gamma(B/\tau(B), C, D)$ is not valid, that is, $\forall C \exists D \neg\gamma(B/\tau(B), C, D)$ is valid. Let u be any context from $D(U)$ such that $u(B_i) = \tau(B_i)$ for all $i \in \{1, \ldots, l\}$. Consider now any context $u_1 \in D(U_1)$. We then distinguish two cases as follows. (a) If $u_1(B_i) \neq \tau(B_i)$ for some $i \in \{1, \ldots, l\}$, then $(M_1, u_1) \models [W \leftarrow \tau(B)]\, Z = 1$, while $(M, u) \not\models [W \leftarrow \tau(B)]\, Z = 1$, where $W \leftarrow \tau(B)$ abbreviates $W_1 \leftarrow \tau(B_1)$, $\ldots, W_l \leftarrow \tau(B_l)$. (b) If $u_1(B_i) = \tau(B_i)$ for all $i \in \{1, \ldots, l\}$, then some truth assignment τ'' to D exists such that $\gamma(B/\tau(B), C/\tau'(C), D/\tau''(D))$ is false, where τ' is the truth assignment to C defined by $\tau'(C_i) = u_1(C_i)$ for all $i \in \{1, \ldots, m\}$. Hence, $(M_1, u_1) \models [W \leftarrow \tau(B), D \leftarrow \tau''(D)]\, Z = 1$, while $(M, u) \not\models [W \leftarrow \tau(B), D \leftarrow \tau''(D)]\, Z = 1$. In summary, if Φ is not valid, then $(*)$ holds.

$(\Leftarrow)$ Suppose that $(*)$ holds. That is, some $u \in D(U)$ exists such that for every $u_1 \in D(U_1)$, there exists a causal formula $[Y \leftarrow y]\, X = x$, where $Y \subseteq V$ and $X \in V$, such that (i) $(M, u) \not\models [Y \leftarrow y]\, X = x$ and (ii) $(M_1, u_1) \models [Y \leftarrow y]\, X = x$. In particular, some $u \in D(U)$ exists such that for every $u_1 \in D(U_1)$ with $u|B = u_1|B$, there exists a causal formula $[Y \leftarrow y]\, X = x$ as above with (i) and (ii). Trivially, (i) and (ii) implies $X \notin Y$ for all such $u_1 \in D(U_1)$. Moreover, as $F_X = F_X^1 = 0$ for all $X \in V \setminus \{Z\}$, it follows that $X = Z$ must hold for all such $u_1 \in D(U_1)$. It then follows that $(M, u) \models [Y \leftarrow y]\, W_i = u(B_i)$ for all $i \in \{1, \ldots, l\}$, since otherwise $(M, u) \models [Y \leftarrow y]\, Z = 1$ and $(M_1, u_1) \models [Y \leftarrow y]\, Z = 1$, for all $u_1 \in D(U_1)$ with $u|B = u_1|B$. This also shows that we have $(M, u) \not\models [Y \leftarrow y]\, Z = 1$ and that $(M_1, u_1) \models [Y \leftarrow y]\ Z = 1$, and, moreover, that $x = 1$ must hold, for all $u_1 \in D(U_1)$ with $u|B = u_1|B$. It then follows that for every truth assignment τ' to C defined by $\tau'(C_i) = u_1(C_i)$ for all $i \in \{1, \ldots, m\}$, there exists a truth assignment τ'' to D which is defined by $(M_1, u_1) \models [Y \leftarrow y]\, D_i = \tau''(D_i)$ for all $i \in \{1, \ldots, n\}$, such that $\gamma(B/\tau(B), C/\tau'(C),\ D/\tau''(D))$ is false, where the truth assignment τ to B

is defined by $\tau(B_i) = u_1(B_i)$ for all $i \in \{1, \ldots, l\}$. This shows that $\exists B \forall C \exists D \neg\gamma$ is valid. That is, $\Phi = \forall B \exists C \forall D \gamma$ is not valid. □

References

[1] A. M. Abdelbar, S. T. Hedetniemi, and S. M. Hedetniemi. The complexity of approximating MAPs for belief networks with bounded probabilities. *Artificial Intelligence*, 124:283–288, 2000.

[2] A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94), Seattle, WA, USA, July 31 - August 4, 1994*, pages 230–237, 1994.

[3] S. Buss and L. Hay. On truth-table reducibility to SAT. *Information and Computation*, 91:86–102, 1991.

[4] T. Bylander, D. Allemang, M. C. Tanner, and J. R. Josephson. The computational complexity of abduction. *Artificial Intelligence*, 49:25–60, 1991.

[5] M. Cadoli, M. Schaerf, A. Giovanardi, and M. Giovanardi. An algorithm to evaluate quantified Boolean formulae and its experimental evaluation. *Journal of Automated Reasoning*, 28:101–142, 2002. Preliminary abstract: An algorithm to evaluate quantified Boolean formulae, Proceedings AAAI/IAAI-98, pages 262–267, 1998.

[6] U. Chajewska and J. Y. Halpern. Defining explanation in probabilistic systems. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI-97), August 1-3, 1997, Brown University, Providence, Rhode Island, USA*, pages 62–71, 1997.

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

[8] A. del Val. On some tractable classes in deduction and abduction. *Artificial Intelligence*, 116(1–2):297–313, 2000.

[9] A. del Val. The complexity of restricted consequence finding and abduction. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI-00), July 30 - August 3, 2000, Austin, Texas, USA*, pages 337–342, 2001.

[10] T. Eiter, W. Faber, N. Leone, and G. Pfeifer. The Diagnosis Frontend of the `dlv` System. *The European Journal on Artificial Intelligence (AI Communications)*, 12(1–2):99–111, 1999.

[11] T. Eiter and G. Gottlob. The complexity of logic-based abduction. *Journal of the ACM*, 42(1):3–42, January 1995.

[12] T. Eiter, G. Gottlob, and N. Leone. Abduction from logic programs: Semantics and complexity. *Theoretical Computer Science*, 189(1-2):129–177, December 1997.

[13] T. Eiter, N. Leone, C. Mateis, G. Pfeifer, and F. Scarcello. The KR System dlv: Progress report, comparisons, and benchmarks. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98), Trento, Italy, June 2-5, 1998*, pages 406–417, 1998.

[14] T. Eiter and T. Lukasiewicz. Complexity results for structure-based causality. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI 2001), Seattle, Washington, USA, August 4-10, 2001*, pages 35–40, 2001.

[15] T. Eiter and T. Lukasiewicz. Causes and explanations in the structural-model approach: Tractable cases. In A. Darwiche and N. Friedman, editors, *Proceedings Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-2002), August 1 - 4, 2002, Edmonton, Alberta, Canada*, pages 146–153, 2002.

[16] T. Eiter and T. Lukasiewicz. Causes and explanations in the structural-model approach: Tractable cases. Technical Report INFSYS RR-1843-02-03, Institut für Informationssysteme, Technische Universität Wien, 2002.

[17] T. Eiter and T. Lukasiewicz. Complexity results for explanations in the structural-model approach. In *Proceedings Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR 2002), Toulouse, France*, pages 49–60. Morgan Kaufmann, 2002.

[18] T. Eiter and T. Lukasiewicz. Complexity results for structure-based causality. *Artificial Intelligence*, 142(1):53–89, 2002.

[19] R. Feldmann, B. Monien, and S. Schamberger. A distributed algorithm to evaluate quantified Boolean formulae. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI-00), July 30 - August 3, 2000, Austin, Texas, USA*, pages 285–290, 2000.

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

[21] P. Gärdenfors. *Knowledge in Flux*. MIT Press, 1988.

[22] H. Geffner. Causal theories for nonmonotonic reasoning. In *Proceedings of the 8th National Conference on Artificial Intelligence (AAAI-90), Boston, Massachusetts, July 29 - August 3, 1990*, pages 524–530, 1990.

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

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

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

[26] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach – Part I: Causes. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence (UAI-01), University of Washington, Seattle, Washington, USA, August 2-5, 2001*, pages 194–202, 2001.

[27] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach – Part II: Explanations. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI 2001), Seattle, Washington, USA, August 4-10, 2001*, pages 27–34, 2001.

[28] C. G. Hempel. *Aspects of Scientific Explanation*. Free Press, 1965.

[29] M. Henrion and M. J. Druzdzel. Qualitative propagation and scenario-based approaches to explanation of probabilistic reasoning. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, July 27-29, 1990, MIT, Cambridge, MA, USA*, pages 17–32. Elsevier Science, 1990.

[30] M. Hopkins. Strategies for determining causes of reported events. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI-02), July 28 - August 1, 2002, Edmonton, Alberta, Canada*, pages 546–552, 2002.

[31] B. Jenner and J. Toran. Computing functions with parallel queries to NP. *Theoretical Computer Science*, 141:175–193, 1995.

[32] D. S. Johnson. A catalog of complexity classes. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume A, chapter 2. Elsevier Science, 1990.

[33] J.Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[34] K. Kask and R. Dechter. A general scheme for automatic generation of search heuristics from specification dependencies. *Artificial Intelligence*, 129:91–131, 2001.

[35] K. Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53:255–272, 1992.

[36] N. Leone, G. Pfeifer, W. Faber, T. Eiter, G. Gottlob, C. Koch, C. Mateis, S. Perri, and F. Scarcello. The DLV System for Knowledge Representation and Reasoning. Technical Report INFSYS RR-1843-02-14, Institut für Informationssysteme, Technische Universität Wien, A-1040 Vienna, Austria, October 2002.

[37] V. Lifschitz. On the logic of causal explanation. *Artificial Intelligence*, 96:451–465, 1997.

[38] N. McCain and H. Turner. Causal theories of action and change. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference, (AAAI/IAAI-97), July 27-31, 1997, Providence, Rhode Island*, pages 460–465, 1997.

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

[40] J. Pearl. Reasoning with cause and effect. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI 99), Stockholm, Sweden, July 31 - August 6, 1999*, pages 1437–1449, 1999.

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

[42] J. Rintanen. Improvements to the evaluation of quantified boolean formulae. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI 99), Stockholm, Sweden, July 31 - August 6, 1999*, pages 1192–1197, 1999.

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

[44] V. Rutenburg. Propositional truth maintenance systems: Classification and complexity analysis. *Annals of Mathematics and Artificial Intelligence*, 10:207–231, 1994.

[45] W. C. Salmon. *Four Decades of Scientific Explanation*. University of Minnesota Press, 1989.

[46] A. Selman. A taxonomy of complexity classes of functions. *Journal of Computer and Systems Sciences*, 48:357–381, 1994.

[47] B. Selman and H. J. Levesque. Support set selection for abductive and default reasoning. *Artificial Intelligence*, 82:259–272, 1996.

[48] S. E. Shimony. Explanation, irrelevance, and statistical independence. In *Proceedings of the 9th National Conference on Artificial Intelligence (AAAI-91), July 14-19, 1991*, pages 482–487, 1991.

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

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

[51] K. Wagner. Bounded query classes. *SIAM Journal of Computing*, 19(5):833–846, 1990.