In: Proc. Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR 2002), Toulouse, France. Morgan Kaufman, 2002.

Complexity Results for Explanations in the Structural-Model Approach

Thomas Eiter
Institut für Informationssysteme,
Technische Universität Wien
Favoritenstraße 9-11, 1040 Wien, Austria
eiter@kr.tuwien.ac.at

Thomas Lukasiewicz*
Dipartimento di Informatica e Sistemistica,
Università di Roma "La Sapienza"
Via Salaria 113, 00198 Roma, Italy
lukasiewicz@dis.uniroma1.it

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 causality. 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, and the complexity of deciding explanations in the general case of situations. 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.

1 INTRODUCTION

The automatic generation of explanations for humans is of crucial importance in areas like planning, diagnosis, natural language processing, and probabilistic inference. Notions of explanations have been studied quite extensively in the literature, see especially [21, 14, 36] for philosophical work, and [25, 38, 22] 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 [4].

*Alternate address: Institut für Informationssysteme, Technische Universität Wien, Favoritenstraße 9-11, 1040 Wien, Austria. E-mail: lukasiewicz@kr.tuwien.ac.at.

In a recent paper [18, 20], 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 causality [18, 19]. They showed that this notion of causal explanation models well many problematic examples in the literature. Note that Halpern and Pearl's causal explanation is very different from the concept of causal explanation in [28, 29, 15].

The following example from [18, 19, 20] illustrates the structural-model approach. See especially [1, 13, 31, 32, 17] for more details on structural causal models.

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.

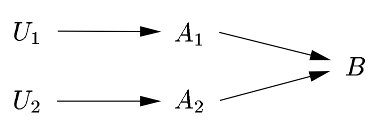

Figure 1: Causal Graph

While the semantic aspects of explanations in the structural-model approach have been thoroughly studied in [18, 20], an analysis of their computational properties is missing so far. In this paper, we fill this gap. Our main contributions are summarized as follows:

- We analyze the computational complexity of Halpern and Pearl's explanations in the structural-model approach [18, 20]. In particular, we draw a precise picture of the complexity of deciding explanations, α-partial explanations, and partial explanations, and of computing the explanatory power of partial explanations.

- We also analyze the complexity of the natural problem of deciding whether an explanation or an α-partial explanation over certain variables exists.

- We show that deciding explanations and partial explanations has a higher complexity in the case of succinctly represented context sets. Generalizing from contexts to situations, in contrast, does not increase the complexity of deciding explanations.

- We also show that all analyzed problems have a lower complexity in the binary case.

Note that detailed proofs of all results are given in the extended paper [10].

2 PRELIMINARIES

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

2.1 CAUSAL MODELS

A *causal model* $M = (U, V, F)$ consists of two disjoint finite sets U and V of *exogenous* and *endogenous* variables, respectively, and a set $F = \{F_X \mid X \in V\}$ of functions $F_X\colon D(PA_X) \to D(X)$ that assign a value of X to each value of the *parents* $PA_X \subseteq U \cup V - \{X\}$ of X.

We focus here on the principal class [18] of *recursive* causal models $M = (U, V, F)$ in which a total ordering $\prec$ on V exists such that $Y \in PA_X$ implies $Y \prec X$, for all $X, Y \in V$. In such models, every assignment to the exogenous variables $U = u$ determines a unique value y for every set of endogenous variables $Y \subseteq V$, denoted $Y_M(u)$ (or simply $Y(u)$). In the sequel, M is reserved for denoting a recursive causal model. For any causal model $M = (U, V, F)$, set of variables $X \subseteq V$, and $x \in D(X)$, the causal model $M_x = (U, V, F_x)$, where $F_x = \{F_Y \mid Y \in V - X\} \cup \{F_{X'} = x(X') \mid X' \in X\}$, is a *submodel* of M. For $Y \subseteq V$, we abbreviate $Y_{M_x}(u)$ by $Y_x(u)$. We say $M = (U, V, F)$ is *binary* iff $|D(X)| = 2$ for all $X \in V$.

Example 2.1 *(arsonists continued)* $M = (U, V, F)$ for Example 1.1 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 $A_2 = 1$ (Fig. 1 shows the parent relationship between the variables).

As for computation, we assume that in $M = (U, V, F)$, every function $F_X\colon D(PA_X) \to D(X)$, $X \in V$, is computable in polynomial time. The following is immediate.

Proposition 2.1 *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 CAUSALITY

We now recall weak causes from [18, 19]. A *primitive event* is an expression of the form $Y = y$, where Y is a variable and y is a value for Y. The set of *events* is the closure of the set of primitive events under the Boolean operations $\neg$ and $\wedge$. The *truth* of an event ϕ in $M = (U, V, F)$ under $u \in D(U)$, denoted $(M, u) \models \phi$, is inductively defined by:

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

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

Proposition 2.2 *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.*

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 set of variables $W \subseteq V - X$ and some values $\overline{x} \in D(X)$, $w \in D(W)$ exist with:

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

Example 2.2 *(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 $B = 1$. 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 lemma characterizes irrelevant variables in weak causes.

Lemma 2.3 *Let $M = (U,V,F)$. Let $X \subseteq V$ and $x \in D(X)$, let ϕ be an event, and let $u \in D(U)$. Let $X_0 \in V$ such that in the causal network for M, it holds that X_0 is not a predecessor of any variable in ϕ, and $X_0(u) = x(X_0)$. Let $X' = X - \{X_0\}$ and $x' = x|X'$. Then, $X = x$ is a weak cause of ϕ under u iff $X' = x'$ is a weak cause of ϕ under u.*

We recall a result from [11, 12], 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 specified by $n_X \geq 1$.

Theorem 2.4 (see [11, 12]) *Given $M = (U,V,F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, deciding whether $X = x$ is a weak cause of ϕ under u is complete for Σ_2^{P} (resp.,* NP*) in the general (resp., binary) case.*

2.3 EXPLANATION

We now recall the concept of explanation from [18, 20]. Let $M = (U,V,F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$, ϕ be an event, and $\mathcal{C} \subseteq D(U)$ be a set of contexts. Then, $X = x$ is an *explanation* of ϕ relative to $\mathcal{C}$ iff the following conditions hold:

EX1. $\phi(u)$ for each 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 $u' \in \mathcal{C}$.

Example 2.3 *(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 $\mathcal{C}$, while $A_1 = 1 \wedge A_2 = 1$ is not, as here, the minimality condition EX3 is violated.

2.4 PARTIAL EXPLANATION AND EXPLANATORY POWER

We finally recall the notions of partial and α-partial explanations and of explanatory power [18, 20]. Let $M = (U,V,F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$, let ϕ be an event, let $\mathcal{C} \subseteq D(U)$ 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 $\mathcal{C}'$; it is easy to see that if such a set $\mathcal{C}'$ exists, then $\mathcal{C}^{\phi}_{X=x}$ is defined. Let P be a probability function on $\mathcal{C}$, and define

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

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* of ϕ relative to $(\mathcal{C}, P)$ for some $\alpha > 0$; furthermore, $P(\mathcal{C}^{\phi}_{X=x} \mid X = x)$ is called its *explanatory power* (or *goodness*).

Example 2.4 *(arsonists continued)* Let $\mathcal{C} = \{u_{1,1}, u_{1,0}, u_{0,1}\}$, 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.

2.5 COMPLEXITY CLASSES

The complexity classes that we encounter are shown in Fig. 2. They are well-known classes from the Polynomial Hierarchy (PH), or derived from them. We recall that $\mathrm{NP} = \Sigma_1^{\mathrm{P}}$, co-NP $= \Pi_1^{\mathrm{P}}$, $\Sigma_{k+1}^{\mathrm{P}} = \mathrm{NP}^{\Sigma_k^{\mathrm{P}}}$, and $\Pi_k^{\mathrm{P}} =$ co-Σ_k^{P}, $k \geq 1$, are classes in PH. The class $\mathrm{D}_k^{\mathrm{P}} = \{L \times L' \mid L \in \Sigma_k^{\mathrm{P}}, L' \in \Pi_k^{\mathrm{P}}\}$, $k \geq 1$, 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 class $\mathrm{P}_{\parallel}^{\Sigma_k^{\mathrm{P}}}$, $k \geq 1$, contains the decision problems which can be solved in polynomial time with parallel calls to a Σ_k^{P} oracle; $\mathrm{FP}_{\parallel}^{\Sigma_k^{\mathrm{P}}}$ is the analog for function computations. Note that $\mathrm{P}_{\parallel}^{\mathrm{NP}} = \mathrm{P}_{\parallel}^{\Sigma_1^{\mathrm{P}}}$ and $\mathrm{FP}_{\parallel}^{\mathrm{NP}} = \mathrm{FP}_{\parallel}^{\Sigma_1^{\mathrm{P}}}$. For further background on the complexity classes, see e.g. [23, 24, 30, 41].

3 OVERVIEW OF RESULTS

In this section, we give an overview on the complexity results that we derive, and discuss possible implications.

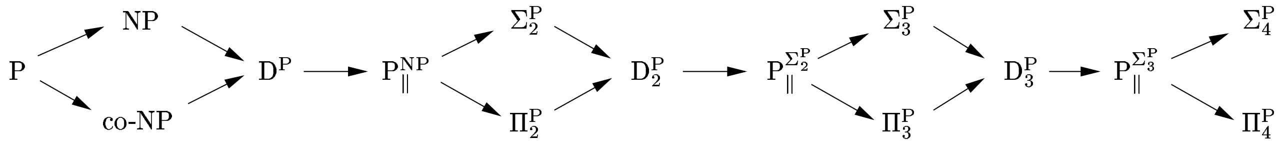

Figure 2: Containment between Complexity Classes

3.1 PROBLEM STATEMENTS

In our 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 $\mathcal{C}$.

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 $X=x$.

In all problems, the probability function P is assumed to be polynomially computable.

The first problem, Explanation, is the recognition of an explanation. It emerges directly from the definition of explanation in Section 2.3 and captures its intrinsic complexity. The problem Explanation Existence is associated with the important task of finding an explanation for an event ϕ. Similar as in other frameworks for explanations

Table 1: Complexity of Explanations

Problem	general case	binary case
Explanation	$\mathrm{D_2^P}$-complete	$\mathrm{D^P}$-complete
Explanation Existence	Σ_3^{P}-complete	Σ_2^{P}-complete
α-Partial Explanation	$\mathrm{P_{\|}^{\Sigma_2^P}}$-complete	$\mathrm{P_{\|}^{NP}}$-complete
α-Partial Explanation Existence	Σ_3^{P}-complete	Σ_2^{P}-complete
Partial Explanation	$\mathrm{P_{\|}^{\Sigma_2^P}}$-complete	$\mathrm{P_{\|}^{NP}}$-complete
Explanatory Power	$\mathrm{FP_{\|}^{\Sigma_2^P}}$-complete	$\mathrm{FP_{\|}^{NP}}$-complete

Table 2: Complexity of Explanations: Succinct Contexts

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

Table 3: Complexity of Explanations: Situations

Problem	general case	binary case
Explanation	$\mathrm{D_2^P}$-complete	$\mathrm{D^P}$-complete

(e.g. [27, 37]), 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. 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.

3.2 MAIN RESULTS

Our main complexity results are compactly summarized in Tables 1–3. Besides the general case, they include results for binary causal models, and also address succinct context representation (see Table 2) and a generalization from contexts to situations in [20] (see Table 3).

All results in Tables 1–3 show completeness under standard polynomial-time transformations [24, 30], and thus sharply characterize the complexity of the problems. From the results in Table 1, it appears that finding explanations and α-partial explanations is at the third level of PH. Thus, explanations are harder to compute than weak causes, which lie at the second level of PH [11]. On the other hand, recognizing explanations and α-partial explanations is only mildly harder than recognizing weak causes, which is Σ_2^{P}-complete. The reason is that by the latter, 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}. In the case of an α-partial explanation, we need to know the context $\mathcal{C}^{\phi}_{X=x}$; by exploiting a basic characterization result (Lemma 4.3), it can be computed efficiently with parallel calls to a Σ_2^{P} oracle. Once $\mathcal{C}^{\phi}_{X=x}$ is known, we need to check whether $X=x$ is an explanation relative to it, the rest is easy. Thus, the complexity of this problem, as well as of Explanatory Power, lies here in the computation of $\mathcal{C}^{\phi}_{X=x}$. The Σ_3^{P} upper bound for Explanation Existence and α-Partial Explanation Existence is then straightforward by 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. The Σ_3^{P}-hardness of α-Partial Explanation Existence is inherited from the hardness of Explanation Existence.

3.3 SUCCINCT CONTEXTS

Table 2 shows results for some of the problems in a setting where contexts are succinctly represented. In fact, Table 1 assumes that the set of contexts $\mathcal{C}$ is enumerated in the input. However, $\mathcal{C}$ may contain exponentially many contexts; 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 $\mathcal{C}$. It turns out that succinct representation increases the complexity of Explanation and α-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 explanations, we have similar effects. Worse, we need to calculate sums of probabilities over succinctly represented context sets. This leads us outside PH: It requires to solve problems which are at least as hard 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. For Partial Explanation, we obtain the entry shown in the table. A complexity increase for Explanation Existence under succinct context sets to Σ_5^{P} is plausible, though we have not analyzed it yet; note that already the Π_4^{P}-hardness proof for Explanation is rather involved.

3.4 SITUATIONS

Table 3 shows results for the generalization of explanations from contexts to situations discussed in [20]. Without going into the details here, in this scenario the epistemic state consists of a set $\mathcal{S}$ of pairs (M, u) called *situations*, where M is a causal model and u is a context, rather than a set of contexts. Informally, a situation (M, u) encodes some causal knowledge in M and some known facts in u. General explanations are then defined as pairs $(\psi, X=x)$ where ψ is an arbitrary causal formula that restricts the causal models to be considered, and $X=x$ is a conjunction of primitive events. The definition is similar to the one of explanations, and is too involved to be presented here (see [20]); it covers basic explanations as a special case. Interestingly, general explanations are not more difficult to recognize than basic explanations.

3.5 RESTRICTED CASES

This concludes our exposition of the complexity results in the general case. Tables 1-3 also show results for the restriction to *binary causal models*, where each endogenous variable may take only two values. In this 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 [11]. The membership parts can be derived analogously 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.

Some of our hardness results remain valid under further restrictions, such as a boundedness condition on the causal model [11, 12]. In particular, all hardness results from Tables 1-3 hold for primitive events ϕ; thus, complex events

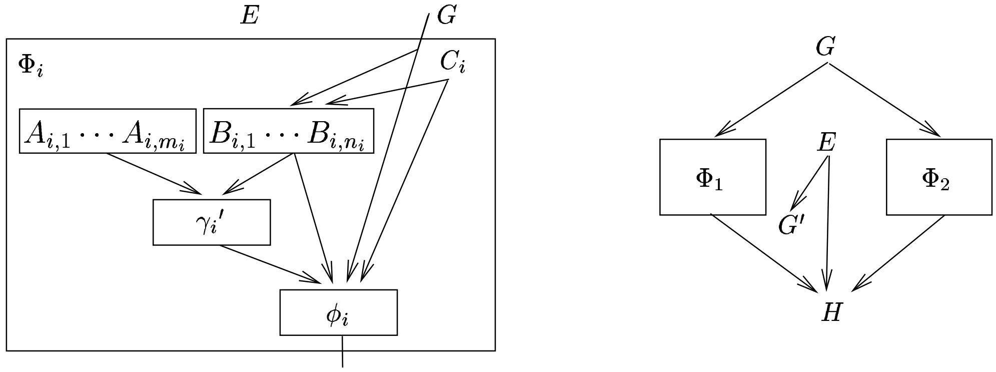

Figure 3: Schematic Construction for Evaluating two QBFs Φ_1 and Φ_2

are not a source of complexity. To avoid a proliferation of results, we do not further consider restrictions here.

3.6 IMPLICATIONS FOR COMPUTATION

For "efficient" algorithms that generate explanations or "best" α-partial explanations, we can draw the following conclusions. Both must solve an inherent Σ_3^{P}-hard problem, i.e., a problem at the third level of PH; such problems are rather hard to solve. Informally, the problem is "triple NP-hard:" even if we could use a subroutine for solving Σ_2^{P}-complete problems for free, the problem would be intractable (NP-hard). Similarly, Σ_2^{P}-complete problems are intractable even if we could use a subroutine for solving NP-complete problems for free. Thus, computationally speaking, generating explanations is rather difficult. In particular, a simple NP-style backtracking strategy that explores, similar as a simple Davis-Putnam style SAT-solver, a polynomial-depth search tree is infeasible. By similar arguments, polynomial-time reductions to a SAT-solver or a computational logic system which can handle problems with complexity up to Σ_2^{P}, such as DLV [9] are infeasible.

On the other hand, an explanation can be computed using a nested backtracking procedure (modeling nested subroutine calls), or using 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 [33]. 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.

4 DERIVATION OF RESULTS

We now sketch how some of our complexity results can be formally derived. More detailed proofs are given in Appendix A. Detailed proofs of all results are given in [10]. Many of these proofs are technically quite involved.

4.1 EXPLANATION

Theorem 4.1 *Explanation is* $\mathrm{D}_2^{\mathrm{P}}$*-complete.*

Proof (sketch). 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 2.4 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 build $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, 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, 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) $V_1 \cap V_2 = \{G\}$, 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 additional endogenous variables (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.

Theorem 4.2 *Explanation Existence is Σ_3^{P}-complete.*

Proof (sketch). We guess some $X' \subseteq X$ and $x' \in D(X')$, and verify that $X'=x'$ is an explanation of ϕ relative to $\mathcal{C}$. By Theorem 4.1, this can be done in polynomial time with two calls to a Σ_2^{P}-oracle. Thus, the problem is in Σ_3^{P}.

Σ_3^{P}-hardness 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 quantor "$\exists B$" 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. □

4.2 PARTIAL EXPLANATION AND EXPLANATORY POWER

We now focus on the complexity of deciding partial and α-partial explanations. The following lemma gives a useful characterization of the set $\mathcal{C}^{\phi}_{X=x}$, which is used below.

Lemma 4.3 *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 $\mathcal{C} \subseteq D(U)$ such that $\phi(u)$ for all $u \in \mathcal{C}$. 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.*

Theorem 4.4 *α-Partial Explanation is $\mathrm{P}_{\|}^{\Sigma_2^{\mathrm{P}}}$-complete.*

Proof (sketch). We first prove membership in $\mathrm{P}_{\|}^{\Sigma_2^{\mathrm{P}}}$. 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) $P(\mathcal{C}^{\phi}_{X=x} \,|\, X=x) \geq \alpha$. By Lemma 4.3, $\mathcal{C}^{\phi}_{X=x}$ 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 2.4, computing $\mathcal{C}^{\phi}_{X=x}$ is in $\mathrm{FP}_{\|}^{\Sigma_2^{\mathrm{P}}}$. Once $\mathcal{C}^{\phi}_{X=x}$ is given, deciding (a) is possible with two Σ_2^{P}-oracle calls, by Theorem 4.1, 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 [2]. Hence, the problem is in $\mathrm{P}_{\|}^{\Sigma_2^{\mathrm{P}}}$.

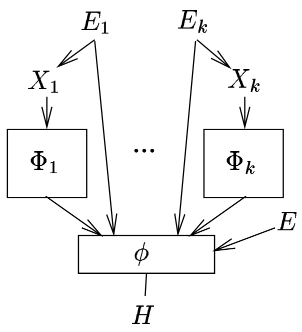

Figure 4: Schematic Construction for Evaluating k QBFs $\Phi_1,\ldots,\Phi_k$

Hardness for $\mathrm{P}_{\|}^{\Sigma_2^{\mathrm{P}}}$ 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. W.l.o.g., $A_1 \cup B_1,\ldots,A_k \cup B_k$ 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} [41]. We create $M=(U,V,F)$, $X \subseteq V$, $x \in D(X)$, 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. 4), 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. □

Theorem 4.5 *α-Partial Explanation Existence is Σ_3^{P}-complete.*

Proof. As for membership in Σ_3^{P}, by Theorem 4.4, deciding whether $X'=x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C},P)$ is in $\mathrm{P}_{\|}^{\Sigma_2^{\mathrm{P}}}$. 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_3^{P}.

Σ_3^{P}-hardness is shown by a reduction from Explanation Existence. 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 4.6 *Partial Explanation is* $\mathrm{P}_{\|}^{\Sigma_2^{\mathrm{P}}}$*-complete.*

Proof (sketch). The membership part can be proved similarly as in the proof of Theorem 4.4. The hardness part follows easily from the hardness result in Theorem 4.4. □

Theorem 4.7 *Explanatory Power is* $\mathrm{FP}_{\|}^{\Sigma_2^{\mathrm{P}}}$*-complete.*

Proof (sketch). We compute first $\mathcal{C}^{\phi}_{X=x}$ and then $P(\mathcal{C}^{\phi}_{X=x} \mid X=x)$. By the proof of Theorem 4.4, the former is in $\mathrm{FP}_{\|}^{\Sigma_2^{\mathrm{P}}}$, while the latter can clearly be done in polynomial time. In summary, computing the explanatory power is in $\mathrm{FP}_{\|}^{\Sigma_2^{\mathrm{P}}}$.

$\mathrm{FP}_{\|}^{\Sigma_2^{\mathrm{P}}}$-hardness 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 $i \in \{1,\ldots,k\}$. W.l.o.g., $A_1 \cup B_1, \ldots, A_k \cup B_k$ are pairwise disjoint, and Φ_1 is valid. Roughly, the main idea is to construct a problem instance such that $(v_1,\ldots,v_k)$ is the bitvector representation of the explanatory power of $X = x$. 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 $u_i \in \mathcal{C}^{\phi}_{X=x}$ iff Φ_i is valid. Defining $P(u_i) = 2^{i-1}$ for all $i \in \{1,\ldots,k\}$ completes the reduction. □

4.3 SUCCINCT REPRESENTATION

Theorem 4.8 *Explanation is* Π_4^{P}*-complete in the case of 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 co-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 2.4, 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 $A \cup B \cup C \cup D$. Roughly speaking, the main idea is to encode Φ in EX3, where the quantor "$\forall A$" is represented by considering all $X' \subset X$, the quantor "$\exists B$" 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. □

Theorem 4.9 *Partial Explanation is* Π_4^{P}*-complete in the case of succinct context sets.*

Proof (sketch). Membership in Π_4^{P} follows from Lemma 4.3 and Theorem 2.4. Hardness for Π_4^{P} can be proved similarly as in the proof of Theorem 4.8. □

5 RELATED WORK AND CONCLUSION

There is quite some work on algorithms and complexity of finding abductive explanations (e.g. [3, 6, 7, 8, 35, 37]) which play an important role in many AI problems including diagnosis, planning, or natural language processing. Roughly, a set of facts E is an abductive explanation of an observation O on some background theory T, if E is compatible with T and entails O; further minimality conditions are usually imposed on E. While causal and abductive explanations (in a standard logical setting [27, 37]) are apparently different concepts, they have similar complexity. In particular, deciding the existence of an abductive explanation is Σ_2^{P}-complete in the propositional context [8]; this matches our respective result on causal explanations for binary causal models. Computing causal and abductive explanations is polynomially intertranslatable in this case, while causal explanations from general causal models are harder to compute. Efficient transformations of causal into abductive explanations, and vice versa, is an interesting subject for further work.

Rather weakly related to ours are complexity results on maximum a posteriori explanations (MAPs, alias most probable explanations [25, 26]), which are a dominating notion of explanation in the probabilistic AI literature. Computing a MAP in a Bayesian belief network, i.e., an assignment to all variables given a partial assignment such that its probability is maximum, is NP-hard [39] but is feasible in polynomial time with an NP oracle. This result 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 [20]. Computationally, it is more suitable to compare deciding $P(X = x) > 0$ in a belief 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 [5], while the latter is, by our results, Π_4^{P}-complete and thus much harder. We may expect a similar relationship for computing the explanatory power vs the probability $P(X = x)$ in a belief network, which can be done in polynomial time with a #P oracle [34].

Our work on causal explanations continues and extends [11], and contributes in paving the way for efficient algorithms and implementations of the structural-model approach by Halpern and Pearl. Our results give a picture of the complexity of explanations in the general and the binary case. However, it remains to identify cases of lower complexity, and in particular islands of tractability. Meaningful restrictions must be found that eliminate several sources of complexity, which is not straightforward. This and refining the complexity picture is part of our ongoing work.

A APPENDIX: SELECTED PROOFS AND PROOF SKETCHES

A.1 PROOFS FOR SECTION 2

Proof of Lemma 2.3. ($\Rightarrow$) Assume that $X = x$ is a weak cause of ϕ under u. That is, (AC1) $X(u) = x$ and $\phi(u)$, and (AC2) some $W \subseteq V - X$, $\overline{x} \in D(X)$, $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}w}(u)$ and (b) $\phi_{xw\hat{z}}(u)$ for all $\hat{Z} \subseteq V - (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. In particular, $X'(u) = x'$ and $\phi(u)$. Moreover, as X_0 is no predecessor of any variable in ϕ, it follows that (a) $\neg\phi_{\overline{x}'w'}(u)$ and (b) $\phi_{x'w'\hat{z}}(u)$ for all $\hat{Z} \subseteq V - (X \cup W)$ 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.

($\Leftarrow$) Assume that $X' = x'$ is a weak cause of ϕ under u. That is, (AC1) $X'(u) = x'$ and $\phi(u)$, and (AC2) some $W \subseteq V - X'$, $\overline{x}' \in D(X')$, $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}'w}(u)$, and (b) $\phi_{x'w\hat{z}}(u)$ for all $\hat{Z} \subseteq V - (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. As $X_0(u) = x(X_0)$, it holds $X(u) = x$ and $\phi(u)$. Moreover, 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 $\hat{Z} \subseteq V - (X \cup W)$ and $\hat{z} = \hat{Z}(u)$, where $w' = w|(W - \{X_0\})$, and $x_0 = x(X_0)$. Hence, $X = x$ is a weak cause of ϕ under u. □

A.2 SELECTED PROOFS FOR SECTION 4

Proof of Theorem 4.1 (continued). 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 $S \in U_i \cup V_i - B_i$. 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 $F_i = \{F^i_S \mid S \in V_i\}$ are defined as follows:

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

As shown in [11, 12], 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 $V = V_1 \cup V_2 \cup \{G', H\}$ 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 Lemma 2.3, 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. Hence, $G = 0$ 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. □

Proof of Theorem 4.2 (continued). Hardness for Σ^{P}_3 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 build $M=(U,V,F)$, $X\subseteq V$, $\mathcal{C}\subseteq D(U)$, and ϕ as required 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 $U=\{I,U_0,U_0',\ldots,U_k,U_k'\}$, where $D(I)=\{0,\ldots,l+1\}$ and $D(S)=\{0,1\}$ for all $S\in U-\{I\}$. 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., ε_i') for every $i\in\{0,\ldots,l+1\}$ (resp., $i\in\{0,\ldots,l\}$) is defined by:

$$\begin{aligned}\varepsilon_i &= I=i\wedge U_0=0\wedge U_0'=1\\ &\qquad\wedge\bigwedge_{i=1}^{l}(U_i=0\wedge U_i'=0)\,,\\ \varepsilon_i' &= I=i\wedge U_0=0\wedge U_0'=0\\ &\qquad\wedge\bigwedge_{i=1}^{l}(U_i=0\wedge U_i'=0)\,.\end{aligned}$$

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 $S\in V-D$. Let

$$\begin{aligned}\alpha &= (\neg\gamma'\wedge\bigwedge_{S\in D}S\neq 2)\vee(E=0)\\ &\qquad\vee\ (X_0=1\wedge E=1\wedge\bigvee_{S\in D}S\neq 2)\,,\\ \phi_1' &= (\varepsilon_0\vee\varepsilon_0'\rightarrow(X_0=0\wedge\bigwedge_{i=1}^{l}B_i\neq B_i')\\ &\qquad\vee\ (\bigvee_{i=1}^{l}(B_i=1\wedge B_i'=1))\vee E'=0)\,,\\ \phi_2' &= \bigwedge_{i=1}^{l}(\varepsilon_i\vee\varepsilon_i'\rightarrow B_i=0\vee B_i'=0)\,,\\ \phi_3' &= (\varepsilon_{l+1}\rightarrow(\alpha\wedge\bigwedge_{i=1}^{l}B_i\neq B_i')\\ &\qquad\vee\ (\bigvee_{i=1}^{l}(B_i=1\wedge B_i'=1))\vee E'=0)\,,\end{aligned}$$

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.

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

For every truth assignment τ to the variables in B, we use $[B/\tau(B)]$ to denote the substitution $[B_1/\tau(B_1),\ldots,B_l/\tau(B_l)]$, and we define $\alpha^\tau=\alpha\,[B/\tau(B)]$. Define $x_0=0$, 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 [11, 12]. 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, Lemma 2.3 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 $\forall C\,\exists D\,\gamma\,[B/\tau(B)]$ is valid.

Using this result, it can be shown 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 $\mathcal{C}$ (see [10] for details). □

Proof of Lemma 4.3. 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). □

Proof of Theorem 4.4 (continued). We construct $M=(U,V,F)$, $X\subseteq V$, $x\in D(X)$, ϕ, $\mathcal{C}\subseteq D(U)$, 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 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 $S\in U_i\cup V_i-B_i$. We define

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

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

- $F^i_{G_i}=E_i$,

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

For each $i \in \{1,\dots,k\}$, 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 $i \in \{1,\dots,k\}$, $X_i = x_i$ 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 4.1, 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 $U = U_1 \cup \dots \cup U_k \cup \{E\}$, where $D(E) = \{0,\dots,k\}$, $V = V_1 \cup \dots \cup V_k \cup \{H\}$, and $F = F_1 \cup \dots \cup F_k \cup \{F_H\}$, where $F_H = 1$ iff

$$(\bigwedge_{i \in \{1,\dots,k\}} \varepsilon_i \rightarrow \phi_i) \wedge (\bigwedge_{i \in \{1,\dots,k\},\, i \text{ even}} \varepsilon'_i \rightarrow \phi_{i-1}) \wedge (\bigwedge_{i \in \{1,\dots,k\},\, i \text{ odd}} \varepsilon'_i \rightarrow \top)$$

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

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

$$\varepsilon'_i \;=\; (E = 0) \wedge (E_i = 1) \wedge (\bigwedge_{j \in \{1,\dots,k\} - \{i\}} (E_j = 0)).$$

For every $i \in \{1,\dots,k\}$, 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 $\mathcal{C} = \{u_1,\dots,u_k, u'_1,\dots,u'_k\}$, $P(u) = 1/2k$ for all $u \in \mathcal{C}$, and $\alpha = 1/2k$. Define $X = \{G_1,\dots,G_k\}$ 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 Lemma 2.3, the following holds for all $i \in \{1,\dots,k\}$, 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 Lemma 4.3, $\mathcal{C}^{\phi}_{X=x}$ 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), $\mathcal{C}^{\phi}_{X=x} = \{u'_1,\dots,u'_k\} \cup \{u_i \mid i \in \{1,\dots,k\},\ \Phi_i \text{ is valid}\}$. It can now be shown that $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ iff the number of valid formulas among $\Phi_1,\dots,\Phi_k$ is even (see [10] for details). □

Proof of Theorem 4.8 (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 $A = \{A_1,\dots,A_k\}$, $B = \{B_1,\dots,B_l\}$, $C = \{C_1,\dots,C_m\}$, and $D = \{D_1,\dots,D_n\}$. We build $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, 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 $U = B \cup \{U_0, U_1, {U_1}',\dots,U_k,{U_k}'\}$, where $D(S) = \{0,1\}$ for all $S \in U$. Define the set of contexts by $\mathcal{C} = \{u \in D(U) \mid (\varepsilon_0 \vee \varepsilon_1 \vee \varepsilon_2)(u)\}$, where:

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

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

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

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

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

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

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 $i \in \{1,\dots,k\}$,
- $F_{X_0} = U_0$, and $F_S = 0$ for all $S \in C \cup \{E, F\}$,
- $F_S = X_0 + E$ 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. It can now be shown that Φ is valid iff $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ (see [10] for details). $\square$

Acknowledgments

This work has been partially supported by the Austrian Science Fund Project Z29-INF, a DFG grant, and a Marie Curie Individual Fellowship of the European Community. We are grateful to the referees for their useful comments.

References

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

[2] S. Buss and L. Hay. On truth-table reducibility to SAT. *Inf. Comput.*, 91:86–102, 1991.

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

[4] U. Chajewska and J. Y. Halpern. Defining explanation in probabilistic systems. In *Proc. UAI-97*, pp. 62–71, 1997.

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

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

[7] A. del Val. The complexity of restricted consequence finding and abduction. In *Proc. AAAI/IAAI-00*, pp. 337–342, 2001.

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

[9] T. Eiter, N. Leone, C. Mateis, G. Pfeifer, and F. Scarcello. The KR System dlv: Progress report, comparisons, and benchmarks. In *Proc. KR-98*, pp. 406–417, 1998.

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

[11] T. Eiter and T. Lukasiewicz. Complexity results for structure-based causality. In *Proc. IJCAI-01*, pp. 35–40, 2001.

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

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

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

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

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

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

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

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

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

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

[22] M. Henrion and M. J. Druzdzel. Qualitative propagation and scenario-based approaches to explanation of probabilistic reasoning. In *Uncertainty in Artificial Intelligence 6*, pp. 17–32. Elsevier Science, 1990.

[23] B. Jenner and J. Toran. Computing functions with parallel queries to NP. *Theor. Comput. Sci.*, 141:175–193, 1995.

[24] 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.

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

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

[27] K. Konolige. Abduction versus closure in causal theories. *Artif. Intell.*, 53:255–272, 1992.

[28] V. Lifschitz. On the logic of causal explanation. *Artif. Intell.*, 96:451–465, 1997.

[29] N. McCain and H. Turner. Causal theories of action and change. In *Proc. AAAI-97*, pp. 460–465, 1997.

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

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

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

[33] J. Rintanen. Constructing conditional plans by a theorem-prover. *J. Artif. Intell. Res.*, 10:323–352, 1999.

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

[35] V. Rutenburg. Propositional truth maintenance systems: Classification and complexity analysis. *Ann. Math. Artif. Intell.*, 10:207–231, 1994.

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

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

[38] S. E. Shimony. Explanation, irrelevance, and statistical independence. In *Proc. AAAI-91*, pp. 482–487, 1991.

[39] S. E. Shimony. Finding MAPs for belief networks is NP-hard. *Artif. Intell.*, 68:399–410, 1994.

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

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