I N F S Y S

R E S E A R C H

R E P O R T

INSTITUT FÜR INFORMATIONSSYSTEME

ABTEILUNG WISSENSBASIERTE SYSTEME

CAUSES AND EXPLANATIONS IN THE STRUCTURAL-MODEL APPROACH: TRACTABLE CASES

Thomas Eiter and Thomas Lukasiewicz

INFSYS RESEARCH REPORT 1843-02-03

MARCH 2002; SEPTEMBER 2005

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

TU

TECHNISCHE UNIVERSITÄT WIEN

INFSYS RESEARCH REPORT

INFSYS RESEARCH REPORT 1843-02-03, MARCH 2002; SEPTEMBER 2005

CAUSES AND EXPLANATIONS IN THE STRUCTURAL-MODEL APPROACH: TRACTABLE CASES

30 SEPTEMBER 2005

Thomas Eiter [1] and Thomas Lukasiewicz [2]

Abstract. This paper continues the research on the computational aspects of Halpern and Pearl's causes and explanations in the structural-model approach. To this end, we first explore how an instance of deciding weak cause can be reduced to an equivalent instance in which irrelevant variables in the (potential) weak cause and the causal model are removed, which extends previous work by Hopkins. We then present a new characterization of weak cause for a certain class of causal models in which the causal graph over the endogenous variables has the form of a directed chain of causal subgraphs, called *decomposable causal graph*. Furthermore, we also identify two important subclasses in which the causal graph over the endogenous variables forms a directed tree and more generally a directed chain of layers, called *causal tree* and *layered causal graph*, respectively. By combining the removal of irrelevant variables with this new characterization of weak cause, we then obtain techniques for deciding and computing causes and explanations in the structural-model approach, which can be done in polynomial time under suitable restrictions. This way, we obtain several tractability results for causes and explanations in the structural-model approach. To our knowledge, these are the first explicit ones. They are especially useful for dealing with structure-based causes and explanations in first-order reasoning about actions, which produces large causal models that are naturally layered through the time line, and thus have the structure of layered causal graphs. Another important feature of the tractable cases for causal trees and layered causal graphs is that they can be recognized efficiently, namely in linear time. Finally, by extending the new characterization of weak cause, we obtain similar techniques for computing the degrees of responsibility and blame, and hence also novel tractability results for structure-based responsibility and blame.

[1]Institut für Informationssysteme, Technische Universität Wien, Favoritenstraße 9-11, 1040 Vienna, Austria; e-mail: eiter@kr.tuwien.ac.at.

[2]Institut für Informationssysteme, Technische Universität Wien, Favoritenstraße 9-11, 1040 Vienna, Austria; e-mail: lukasiewicz@kr.tuwien.ac.at. Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Via Salaria 113, 00198 Rome, Italy; e-mail: lukasiewicz@dis.uniroma1.it.

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

Results of this paper appear in: *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI-2002)*, August 1–4, 2002, Edmonton, Alberta, Canada, pp. 146–153, Morgan Kaufmann, 2002.

Contents

1 Introduction

Dealing with causality is an important issue which emerges in many applications of AI. The existing approaches to causality in AI can be roughly divided into those that have been developed as modal nonmonotonic logics (especially in logic programming) and those that evolved from the area of Bayesian networks. A representative of the former is Geffner's modal nonmonotonic logic for handling causal knowledge [12, 13], which is inspired by default reasoning from conditional knowledge bases. Other modal-logic based formalisms play an important role in dealing with causal knowledge about actions and change; see especially the work by Turner [36] and the references therein for an overview. A representative of the latter is Pearl's approach to modeling causality by structural equations [1, 10, 30, 31], which is central to a number of recent research efforts. In particular, the evaluation of deterministic and probabilistic counterfactuals has been explored, which is at the core of problems in fault diagnosis, planning, decision making, and determination of liability [1]. It has been shown that the structural-model approach allows a precise modeling of many important causal relationships, which can especially be used in natural language processing [10]. An axiomatization of reasoning about causal formulas in the structural-model approach has been given by Halpern [14].

Causality also plays an important role in the generation of explanations, which are of crucial importance in areas like planning, diagnosis, natural language processing, and probabilistic inference. Different notions of explanations have been studied quite extensively, see especially [19, 11, 34] for philosophical work, and [29, 35, 20] for work in AI related to Bayesian networks. A critical examination of such approaches from the viewpoint of explanations in probabilistic systems is given in [2].

In [15], Halpern and Pearl formalized causality using a model-based definition, which allows for a precise modeling of many important causal relationships. Based on a notion of weak causality, they offer appealing definitions of actual causality [16] and causal explanations [18]. As they show, their notions of actual cause and causal explanation, which is very different from the concept of causal explanation in [26, 27, 12], models well many problematic examples in the literature.

The following example from [3] illustrates the structural-model approach. Roughly, structural causal models consist of a set of 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. For more details on structural causal models, we refer to Section 2 and especially to [1, 10, 30, 31, 14].

Example 1.1 *(rock throwing)* Suppose that Suzy and Billy pick up rocks and throw them at a bottle. Suzy's rock gets there first, shattering the bottle. Since both throws are fully accurate, Billy's rock would have shattered the bottle, if Suzy had not thrown. We may model such a scenario in the structural-model approach as follows. We assume two binary background variables U_S and U_B, which determine the motivation and the state of mind of Suzy and Billy, where U_S (resp., U_B) is 1 iff Suzy (resp., Billy) intends to throw a rock. We then have five binary variables ST, BT, SH, BH, and BS, which describe the observable situation, where ST (resp., BT) is 1 iff Suzy (resp., Billy) throws a rock, SH (resp., BH) is 1 iff Suzy's (resp., Billy's) rock hits the bottle, and BS is 1 iff the bottle shatters. The causal dependencies between these variables are expressed by functions, which say that (i) the value of ST (resp., BT) is given by the value of U_S (resp., U_B), (ii) SH is 1 iff ST is 1, (iii) BH is 1 iff BT is 1 and SH is 0, and (iv) BS is 1 iff SH or BH is 1. These dependencies can be graphically represented as in Fig. 1.

Some actual causes and explanations in the structural-model approach are then informally given as

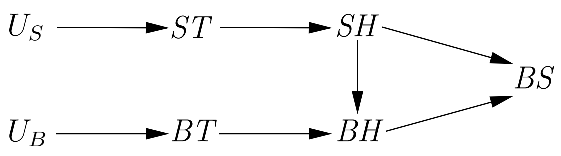

Figure 1: Causal Graph

follows. If both Suzy and Billy intend to throw a rock, then (i) Suzy's throwing a rock is an *actual cause* of the bottle shattering, while (ii) Billy's throwing a rock is not. Furthermore, (iii) if either Suzy or Billy intends to throw a rock, then Suzy's throwing a rock is an *explanation* of the bottle shattering. Here, (i)–(iii) are roughly determined as follows. As for (i), if both Suzy and Billy intend to throw a rock, then Suzy actually throws a rock, and the bottle actually shatters. Moreover, under the *structural contingency* that Billy does not throw a rock, (a) if Suzy does not throw a rock, then the bottle does not shatter, and (b) if Suzy throws a rock, then the bottle shatters, even if any of the other variables would take their actual values. As for (ii), there is no structural contingency under which (a) if Billy does not throw a rock, then the bottle does not shatter, and (b) if Billy throws a rock, then the bottle shatters, even if any of the other variables would take their actual values. Finally, as for (iii), if either Suzy or Billy intends to throw a rock, then the bottle actually shatters, Suzy's throwing a rock is a cause of the bottle shattering whenever she actually throws a rock, and there are some possible contexts in which Suzy throws a rock and some in which she does not. Intuitively, there should be a possible context in which the explanation is false, so that it is not already known, and a possible context in which the explanation is true, so that it is not vacuous. $\Box$

There are a number of recent papers that are based on Halpern and Pearl's definitions of actual causality [16] and causal explanations [18]. In particular, Chockler and Halpern [3] define the notions of responsibility and blame as a refinement of actual causality. Chockler, Halpern, and Kupferman [4] then make use of the notion of responsibility for verifying a system against a formal specification. Along another line of application, Hopkins and Pearl [23] and Finzi and Lukasiewicz [9] generalize structure-based causes and explanations to a first-order framework and make them available in situation-calculus-based reasoning about actions (see Section 8.3). Furthermore, Hopkins and Pearl [24] explore the usage of structure-based causality [16] for commonsense causal reasoning. Finally, inspired by Halpern and Pearl's notions of actual causality [16] and causal explanations [18], Park [28] presents a novel approach allowing for different causal criteria that are influenced by psychological factors not representable in a structural causal model.

The semantic aspects of causes and explanations in the structural-model approach have been thoroughly studied in [15, 16, 17, 18]. Their computational complexity has been analyzed in [6, 7], where it has been shown that associated decision problems are intractable in general. For example, deciding actual causes (as defined in [16]) is complete for the class Σ_2^P (=NP^{NP}) of the Polynomial Hierarchy, while deciding whether an explanation over certain variables exists is complete for Σ_3^P (=$NP^{\Sigma_2^P}$). Thus, these problems are "harder" than the classical propositional satisfiability problem (which is NP-complete), but "easier" than PSPACE-complete problems. Chockler and Halpern [3] and Chockler, Halpern, and Kupferman [4] have shown that computing the degrees of responsibility and blame is complete for polynomial time computation with restricted use of a Σ_2^P oracle (see Section 3.4). As for algorithms, Hopkins [21] explored search-based strategies for computing actual causes in both the general and restricted settings.

However, to our knowledge, no tractable cases for causes and explanations in the structural-model approach were explicitly known so far. In this paper, we aim at filling this gap and provide non-trivial tractability results for the main computational problems on causes and explanations. These tractability results are especially useful for dealing with structure-based causes and explanations in first-order reasoning about actions as recently introduced in [9], where one has to handle binary causal models with a quite large number of variables (see Section 8.3). We make contributions to several issues, which are briefly summarized as follows:

- The first issue concerns focusing of the computation to the relevant part of the causal model. Extending work by Hopkins [21], we explore how an instance of deciding weak cause can be reduced to an equivalent instance in which the (potential) weak cause and the causal model may contain fewer variables. That is, irrelevant variables in weak causes and causal models are identified and removed. We provide two such reductions in this paper, which have different properties, but can be both carried out in polynomial time. These reductions can lead to great simplifications in (potential) weak causes and causal models, and thus speed up considerably computations about causes and explanations. Notice that weak causes are fundamental to the notion of actual cause, to various forms of explanations, as well as to the notions of responsibility and blame.

- The second issue to which we contribute are characterizations of weak causes in the structural-model approach. We present a novel such characterization for a class of causal models in which the causal graph over the endogenous variables has the form of a directed chain of causal subgraphs, which we call a *decomposable causal graph*. We also identify two natural subclasses of decomposable causal graphs, where the causal graph over the endogenous variables forms a directed tree and, more generally, a directed chain of layers, which we call a *causal tree* and a *layered causal graph*, respectively, and provide simplified versions of the characterizations of weak causes.

- By combining the removal of irrelevant variables (in weak causes and causal models) with this new characterization of weak cause in the above causal models, we obtain algorithms for deciding and computing weak causes, actual causes, explanations, partial explanations, and α-partial explanations, as well as for computing the explanatory power of partial explanations, which all run in polynomial time under suitable conditions. This way, we obtain several tractability results for the structural-model approach. To our knowledge, these are the first ones that are explicitly derived for structure-based causes and explanations.

- Furthermore, by slightly extending the new characterization of weak cause in the above causal models, we also obtain algorithms for computing the degrees of responsibility and blame in the structural-model approach, which similarly run in polynomial time under suitable conditions. We thus also obtain new tractability results for the structure-based notions of responsibility and blame. Note that Chockler, Halpern, and Kupferman [4] have recently shown that computing the degree of responsibility in read-once Boolean formulas (which are Boolean formulas in which each variable occurs at most once) is possible in linear time.

- Finally, we show that all the above techniques and results carry over to a refinement of the notion of weak cause and to a generalization of causal models to extended causal models, which have been both recently introduced by Halpern and Pearl in [17]. Furthermore, we describe an application of the results of this paper for dealing with structure-based causes and explanations in first-order reasoning about actions. Here, one has to handle binary causal models with a quite large number of variables, but with a natural layering through the time line. Thus, such causal models often have the structure of layered causal graphs.

An attractive feature of the tractable cases identified for causal trees and layered causal graphs is that the respective problem instances can be recognized efficiently, namely in linear time. For general decomposable causal graphs, however, this is not the case, since this problem is NP-complete in general. Nonetheless, effort spent for the recognition may be more than compensated by the speed up in solving the reasoning problems on weak causes and explanations.

Our results on the computational and semantic properties of weak causes and explanations help, as we believe, to enlarge the understanding of and insight into the structural-model approach by Halpern and Pearl and its properties. Furthermore, they provide the basis for developing efficient algorithms and pave the way for implementations. For example, complexity results on answer set programming [5] have guided the development of efficient solvers such as DLV [25]. The results of this paper are in particular of interest and significant, since a structural decomposition seems natural and applies to a number of examples from the literature.

The rest of this paper is organized as follows. Section 2 contains some preliminaries on structural causal models as well as on causes, explanations, responsibility, and blame in structural causal models. In Section 3, we describe the decision and optimization problems for which we present tractability results in this paper, and we summarize previous complexity results for these problems. In Section 4, we explore the removal of irrelevant variables when deciding weak cause. Section 5 presents tractability results for causal trees. Section 6 then generalizes to decomposable causal graphs, while Section 7 concentrates on layered causal graphs. In Section 8, we generalize the above techniques and results to the refined notion of weak cause and extended causal models, and describe their application in first-order reasoning about actions. Section 9 summarizes our results and gives a conclusion.

To increase readability, all proofs have been moved to Appendices A–E.

2 Preliminaries

In this section, we give some technical preliminaries. We recall Pearl's structural causal models and Halpern and Pearl's notions of weak and actual cause [15, 16] and their notions of explanation, partial explanation, and explanatory power [15, 18].

2.1 Causal Models

We start with recalling structural causal models; for further background and motivation, see especially [1, 10, 30, 31, 14]. Roughly, the main idea behind structural 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 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 all $i \in \{1, \ldots, n\}$; 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. We say that X is *domain-bounded* iff $|D(X_i)| \leq k$ for every $X_i \in X$, where k is some global constant. Lower case letters x, y, z, etc. denote values for the variables or the sets of variables X, Y, Z, etc., respectively. Assignments $X = x$ of values to variables are often abbreviated by the value x. We often identify singletons $\{X_i\}$ with X_i, and their values x with $x(X_i)$.

For $Y \subseteq X$ and $x \in D(X)$, we denote by $x|Y$ the restriction of x to Y. For disjoint sets of variables X, Y and values $x \in D(X)$, $y \in D(Y)$, we denote by xy the union of x and y. For (not necessarily disjoint) sets of variables X, Y and values $x \in D(X)$, $y \in D(Y)$, we denote by $[x\langle y]$ the union of $x|(X \setminus Y)$ and y.

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 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*. We call a causal model $M = (U, V, F)$ *domain-bounded* iff every $X \in V$ is domain-bounded. In particular, M is *binary* iff $|D(X)| = 2$ for all $X \in V$. The parent relationship between the variables of $M = (U, V, F)$ is expressed by the *causal graph* for M, denoted $G(M)$, which is the directed graph (N, E) that has $U \cup V$ as the set of nodes N, and a directed edge from X to Y in E iff X is a parent of Y, for all variables $X, Y \in U \cup V$. We use $G_V(M)$ to denote the subgraph of $G(M)$ induced by V.

We focus here on the principal class of *recursive* causal models $M = (U, V, F)$; as argued in [15], 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 $U = u$ to the exogenous variables determines a unique value y for every set of endogenous variables $Y \subseteq V$, denoted by $Y_M(u)$ (or simply by $Y(u)$, if M is understood). In the following, M is reserved for denoting a recursive causal model.

Example 2.1 *(rock throwing cont'd)* The causal model $M = (U, V, F)$ for Example 1.1 is given by $U = \{U_S, U_B\}$, $V = \{ST, BT, SH, BH, BS\}$, and $F = \{F_{ST}, F_{BT}, F_{SH}, F_{BH}, F_{BS}\}$, where $F_{ST} = U_S$, $F_{BT} = U_B$, $F_{SH} = ST$, $F_{BH} = 1$ iff $BT = 1$ and $SH = 0$, and $F_{BS} = 1$ iff $SH = 1$ or $BH = 1$. Fig. 1 shows the causal graph for M, that is, the parent relationships between the exogenous and endogenous variables in M. Since this graph is acyclic, M is recursive. $\Box$

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 $M_{X=x}$ is given by $(U, V \setminus X, F_{X=x})$, 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 $Y \subseteq V$ and $u \in D(U)$, we write $Y_x(u)$ to abbreviate $Y_{M_x}(u)$. We assume that $X(u) = x$ for all $u \in D(U)$ in the submodel of M where X is set to x.

As for computation, we assume for causal models $M = (U, V, F)$ no particular form of representation of the functions $F_X \colon D(PA_X) \rightarrow D(X)$, $X \in V$, in F (by formulas, circuits, etc.), but that every F_X is evaluable in polynomial time. Furthermore, we assume that the causal graph $G(M)$ for M is part of the input representation of M. Notice that $G(M)$ is computable from M with any common representation of the functions F_X (by formulas, circuits, etc.) in time linear in the size of the representation of M anyway. For any causal model M, we denote by $\|M\|$ the size of its representation.

The following proposition is then 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 Weak and Actual Causes

We now recall weak and actual causes from [15, 16]. We first define events and the truth of events in a causal model $M = (U, V, F)$ under a context $u \in D(U)$.

A *primitive event* is an expression of the form $Y = y$, where Y is an endogenous variable and y is a value for Y. The set of *events* is the closure of the set of primitive events under the Boolean operations $\neg$ and $\wedge$ (that is, every primitive event is an event, and if ϕ and ψ are events, then also $\neg\phi$ and $\phi \wedge \psi$). For any event ϕ, we denote by $V(\phi)$ the set of all variables in ϕ.

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, that is, $\phi \vee \psi$ and $\phi \rightarrow \psi$ stand for $\neg(\neg\phi \wedge \neg\psi)$ and $\neg\phi \vee \psi$, respectively. We write $\phi_M(u)$ (resp., $\phi(u)$ if M is understood) to abbreviate $(M, u) \models \phi$. For $X \subseteq V$ and $x \in D(X)$, we use $\phi_{M_x}(u)$ (resp., $\phi_x(u)$) as an abbreviation of $(M_x, u) \models \phi$. For $X = \{X_1, \ldots, X_k\} \subseteq V$ with $k \geq 1$ and $x_i \in D(X_i)$, $1 \leq i \leq k$, we use $X = x_1 \cdots x_k$ to abbreviate the event $X_1 = x_1 \wedge \ldots \wedge X_k = x_k$. For any event ϕ, we denote by $\|\phi\|$ its size, which is the number of symbols in it.

The following result follows immediately from Proposition 2.1.

Proposition 2.2 *Let $X \subseteq V$ and $x \in D(X)$. Given $u \in D(U)$ and an event ϕ, deciding whether $\phi(u)$ holds (resp., $\phi_x(u)$ holds for given x) is feasible in polynomial time.*

We are now ready to recall the notions of weak and actual cause [15, 16]. Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$, and let ϕ be an event. Then, $X = x$ is a *weak cause* of ϕ under $u \in D(U)$ 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 $\hat{z} = \hat{Z}(u)$.

Loosely speaking, **AC1** says that both $X = x$ and ϕ hold under u, while **AC2** expresses that $X = x$ is a non-trivial reason for ϕ. Here, the dependence of ϕ from $X = x$ is tested under special *structural contingencies*, where some $W \subseteq V \setminus X$ is kept at some value $w \in D(W)$. **AC2**(a) says that ϕ can be false for other values of X under w, while **AC2**(b) essentially ensures that X alone is sufficient for the change from ϕ to $\neg\phi$. Observe that $X = x$ can be a weak cause only if X is nonempty.

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

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

The following characterization of actual causes through weak causes is known.

Theorem 2.3 (see [6]) *Let $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, and $u \in D(U)$. Let ϕ be an event. Then, $X = x$ is an actual cause of ϕ under u iff X is a singleton and $X = x$ is a weak cause of ϕ under u.*

We give an example to illustrate the above notions of weak and actual cause.

Example 2.2 *(rock throwing cont'd)* Consider the context $u_{1,1}=(1,1)$ in which both Suzy and Billy intend to throw a rock. Then, both $ST=1$ and $ST=1 \wedge BT=1$ are weak causes of $BS=1$, while $BT=1$ is not. For instance, let us show that $ST=1$ is a weak cause of $BS=1$ under $u_{1,1}$. As for **AC1**, both ST and BS are 1 under $u_{1,1}$. As for **AC2**, under the contingency that BT is set to 0, we have that (a) if ST is set to 0, then BS has the value 0, and (b) if ST is set to 1, then BS is 1. In fact, by Theorem 2.3, $ST=1$ is an actual cause of $BS=1$ under $u_{1,1}$, while $ST=1 \wedge BT=1$ is not. Furthermore, $ST=1$ (resp., $BT=1$) is the only weak (and by Theorem 2.3 also actual) cause of $BS=1$ under $u_{1,0}=(1,0)$ (resp., $u_{0,1}=(0,1)$) in which only Suzy (resp., Billy) intends to throw a rock. □

2.3 Explanations

We next recall the concept of an explanation from [15, 18]. 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.

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 $\mathcal{C}$ iff 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 context $u \in \mathcal{C}$ exists such that $X'(u) = x|X'$ and $X'=x|X'$ is not a weak cause of ϕ under u.

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

Note that in **EX3**, any counterexample $X' \subset X$ to minimality must be a nonempty set of variables. The following example illustrates the above notion of explanation.

Example 2.3 *(rock throwing cont'd)* Consider the set of contexts $\mathcal{C}=\{u_{1,1},\, u_{1,0},\, u_{0,1}\}$. Then, $ST=1$ is an explanation of $BS=1$ relative to $\mathcal{C}$, since **EX1** $BS(u_{1,1})=BS(u_{1,0})=BS(u_{0,1})=1$, **EX2** $ST=1$ is a weak cause of $BS=1$ under both $u_{1,1}$ and $u_{1,0}$, **EX3** ST is obviously minimal, and **EX4** $ST(u_{1,1})=1$ and $ST(u_{0,1}) \neq 1$. Furthermore, $ST=1 \wedge BT=1$ is not an explanation of $BS=1$ relative to $\mathcal{C}$, since here the minimality condition **EX3** is violated. □

2.4 Partial Explanations and Explanatory Power

We next recall the notions of α-partial / partial explanations and of explanatory power of partial explanations [15, 18]. Roughly, the main idea behind partial explanations is to generalize the notion of explanation of Section 2.3 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 $\mathcal{C}^{\phi}_{X=x}$ to denote the largest subset $\mathcal{C}'$ of $\mathcal{C}$ such that $X=x$ is an explanation of ϕ relative to $\mathcal{C}'$. Note that this set $\mathcal{C}^{\phi}_{X=x}$ is unique. The following proposition from [7] shows that $\mathcal{C}^{\phi}_{X=x}$ is defined, if a subset $\mathcal{C}'$ of $\mathcal{C}$ exists such that $X=x$ is an explanation of ϕ relative to $\mathcal{C}'$; it also characterizes $\mathcal{C}^{\phi}_{X=x}$.

Proposition 2.4 *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 $u \in \mathcal{C}$. 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.*

Let P be a probability function on $\mathcal{C}$ (that is, P is a mapping from $\mathcal{C}$ to the interval $[0,1]$ such that $\sum_{u \in \mathcal{C}} P(u)=1$), and define

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

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 $P(\mathcal{C}^{\phi}_{X=x} \,|\, X=x) \geq \alpha$. 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, $P(\mathcal{C}^{\phi}_{X=x} \,|\, X=x)$ is called its *explanatory power* (or *goodness*).

Example 2.4 *(rock throwing cont'd)* Consider the set of contexts $\mathcal{C}=\{u_{1,1},\, u_{1,0},\, u_{0,1}\}$, and let $P(u_{1,1})=0.2$ and $P(u_{1,0})=P(u_{0,1})=0.4$. Then, $\mathcal{C}^{BS=1}_{ST=1}=\mathcal{C}$, and thus $ST=1$ is a 1-partial explanation of $BS=1$ relative to $(\mathcal{C},P)$. That is, $ST=1$ is a partial explanation of $BS=1$ relative to $(\mathcal{C},P)$ with explanatory power 1. □

As for computation, we assume that the above probability functions P on $\mathcal{C}$ are computable in polynomial time.

2.5 Responsibility and Blame

We finally recall the notions of responsibility and blame from [3]. Intuitively, the notion of responsibility is a refinement of the notion of actual cause, which also measures the minimal number of changes that must be made under a structural contingency to create a counterfactual dependence of ϕ from $X=x$. Whereas the notion of blame then also takes into consideration the belief of an agent about the possible causal models and contexts (before setting the weak cause).

In the sequel, let $M=(U,V,F)$ be a causal model, let $X \subseteq V$, $x \in D(X)$, and $u \in D(U)$, and let ϕ be an event. Let us call the pair (M,u) a *situation*. Then, the *degree of responsibility* of $X=x$ for ϕ in situation (M,u), denoted $\mathrm{dr}((M,u), X{=}x, \phi)$, is 0 if $X=x$ is not an actual cause of ϕ under u in M, and it is $1\,/\,(k{+}1)$ if $X=x$ is an actual cause of ϕ under u in M, and

(i) some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that **AC2**(a) and (b) hold and that k variables in W have different values in w and $W(u)$, and

(ii) no $W' \subseteq V \setminus X$, $\overline{x}' \in D(X)$, and $w' \in D(W')$ exist such that **AC2**(a) and (b) hold and that $k'<k$ variables in W' have different values in w' and $W'(u)$.

Informally, $\mathrm{dr}((M,u), X{=}x, \phi) = 1\,/\,(k{+}1)$, where k is the minimal number of changes that have to be made under u in M to make ϕ counterfactually depend on $X = x$. In particular, if $X = x$ is not an actual cause of ϕ under u in M, then $k = \infty$, and thus $\mathrm{dr}((M,u), X{=}x, \phi) = 0$. Otherwise, $\mathrm{dr}((M,u), X{=}x, \phi)$ is at most 1.

Example 2.5 *(rock throwing cont'd)* Consider again the context $u_{1,1} = (1,1)$ in which both Suzy and Billy intend to throw a rock. As argued in Example 2.2, Suzy's throwing a rock ($ST = 1$) is an actual cause of the bottle shattering ($BS = 1$), witnessed by the contingency that Billy does not throw (and hence does not hit). Here, **AC2** holds also under the contingency that Billy throws a rock, but the rock does not hit the bottle (BT and BH are set to 1 and 0, respectively). Since BT and BH are 1 and 0, respectively, under $u_{1,1}$, the degree of responsibility of Suzy's throwing a rock ($ST = 1$) for the bottle shattering ($BS = 1$) in $(M, u_{1,1})$ is given by 1. □

An *epistemic state* $\mathcal{E} = (\mathcal{K}, P)$ consists of a set of situations $\mathcal{K}$ and a probability distribution P over $\mathcal{K}$. The *degree of blame* of setting X to x for ϕ relative to an epistemic state $(\mathcal{K}, P)$, denoted $\mathrm{db}(\mathcal{K}, P, X \leftarrow x, \phi)$, is defined as

$$\textstyle\sum_{(M,u)\in\mathcal{K}} \mathrm{dr}((M_{X=x}, u), X{=}x, \phi) \cdot P((M,u))\,.$$

Informally, $(\mathcal{K}, P)$ are the situations that an agent considers possible before X is set to x along with their probabilities believed by the agent. Then, $\mathrm{db}(\mathcal{K}, P, X \leftarrow x, \phi)$ is the expected degree of responsibility of $X = x$ for ϕ in $(M_{X=x}, u)$.

Example 2.6 *(rock throwing cont'd)* Suppose that we are computing the degree of blame of Suzy's throwing a rock for the bottle shattering. Assume that Suzy considers possible a modified version of the causal model given in Example 2.1, denoted M', where Billy may also throw extra hard, which is expressed by the additional value 2 of U_B and BT. If Billy throws extra hard, then Billy's rock hits the bottle independently of what Suzy does, which is expressed by additionally assuming that BH is 1 if BT is 2. Assume then that Suzy considers possible the contexts $u_{1,0}$, $u_{1,1}$, and $u_{1,2}$, where Suzy throws a rock, and Billy either does not throw a rock, throws a rock in a normal way, or throws a rock extra hard. Finally, assume that each of the three contexts has the probability $1/3$. It is then not difficult to verify that the degree of responsibility of Suzy's throwing a rock for the bottle shattering is $1/2$ in $(M', u_{1,2})$ and 1 in both $(M', u_{1,0})$ and $(M', u_{1,1})$. Thus, the degree of blame of Suzy's throwing a rock for the bottle shattering is $5/6$. □

3 Problem Statements

We concentrate on the following important computational problems for causes, explanations, responsibility, and blame in the structural-model approach, which comprise both decision problems and problems with concrete output.

3.1 Causes

WEAK/ACTUAL CAUSE: Given $M{=}(U,V,F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, decide if $X = x$ is a weak (resp., an actual) cause of ϕ under u.

WEAK/ACTUAL CAUSE COMPUTATION: Given $M = (U,V,F)$, $X \subseteq V$, $u \in D(U)$, and an event ϕ, compute the set of all $X' = x'$ such that (i) $X' \subseteq X$ and $x' \in D(X')$, and (ii) $X' = x'$ is a weak (resp., an actual) cause of ϕ under u.

3.2 Notions of Explanations

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 COMPUTATION: Given $M = (U, V, F)$, $X \subseteq V$, an event ϕ, and a set of contexts $\mathcal{C} \subseteq D(U)$, compute the set of all $X' = x'$ such that (i) $X' \subseteq X$ and $x' \in D(X')$, and (ii) $X' = x'$ 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 if $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$.

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

PARTIAL EXPLANATION COMPUTATION: 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 $u \in \mathcal{C}$, and a probability function P on $\mathcal{C}$, compute the set of all $X' = x'$ such that (i) $X' \subseteq X$ and $x' \in D(X')$, and (ii) $X' = x'$ 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, a set of contexts $\mathcal{C} \subseteq D(U)$, 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 $X = x$ for ϕ relative to $(\mathcal{C}, P)$.

3.3 Responsibility and Blame

RESPONSIBILITY: Given $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, compute the degree of responsibility of $X = x$ for ϕ in (M, u).

BLAME: Given an epistemic state $\mathcal{E}$, a set of endogenous variables X, $x \in D(X)$, and an event ϕ, compute the degree of blame of setting X to x for ϕ relative to $\mathcal{E}$.

3.4 Previous Results

Several complexity results for the above problems have been established. In particular, as shown in [6], the decision problems WEAK CAUSE and ACTUAL CAUSE are both Σ_2^{P}-complete in the general case, and NP-complete in the case of binary variables. Furthermore, as shown in [7], the decision problems EXPLANATION and PARTIAL/α-PARTIAL EXPLANATION and the optimization problem EXPLANATORY POWER are complete for $\mathrm{D}_2^{\mathrm{P}}$, $\mathrm{P}_{\|}^{\Sigma_2^{\mathrm{P}}}$, and $\mathrm{FP}_{\|}^{\Sigma_2^{\mathrm{P}}}$, respectively, in the general case, and complete for D^{P}, $\mathrm{P}_{\|}^{\mathrm{NP}}$ and $\mathrm{FP}_{\|}^{\mathrm{NP}}$, respectively, in the binary case. Here $\mathrm{D}_2^{\mathrm{P}}$ (resp., D^{P}) is the "logical conjunction" of Σ_2^{P} and Π_2^{P} (resp., NP and co-NP), and $\mathrm{P}_{\|}^{C}$ (resp., $\mathrm{FP}_{\|}^{C}$) is the class of decision problems solvable (resp., functions computable) in polynomial time with access to one round of parallel queries to an oracle in C. Finally, Chockler and Halpern [3] and Chockler, Halpern, and Kupferman [4] have shown that the optimization problems RESPONSIBILITY and BLAME are complete for the classes $\mathrm{FP}^{\Sigma_2^{\mathrm{P}}[\log n]}$ and $\mathrm{FP}_{\|}^{\Sigma_2^{\mathrm{P}}}$, respectively, in the general case, and complete for $\mathrm{FP}^{\mathrm{NP}[\log n]}$ and $\mathrm{FP}_{\|}^{\mathrm{NP}}$, respectively, in the binary case. The class $\mathrm{FP}^{C[\log n]}$ contains the functions computable in polynomial time with $O(\log n)$ many calls to an oracle in C, where n is the size of the problem input.

To our knowledge, there exist no complexity results for the optimization problems WEAK/ACTUAL CAUSE COMPUTATION, EXPLANATION COMPUTATION, and α-PARTIAL/PARTIAL EXPLANATION COMPUTATION so far. But there are complexity results on decision variants of two of the latter problems, which are called EXPLANATION EXISTENCE and α-PARTIAL EXPLANATION EXISTENCE, respectively. They are the decision problems of deciding whether an explanation and an α-partial explanation, respectively, over certain variables exists, which are complete for Σ_3^{P} (resp., Σ_2^{P}) in the general (resp., binary) case; see [7].

To our knowledge, there are no explicit tractability results for the above problems related to causes and explanations so far. As for responsibility and blame, Chockler, Halpern, and Kupferman [4] have shown that computing the degree of responsibility in read-once Boolean formulas can be done in linear time.

4 Irrelevant Variables

In this section, we describe how an instance of deciding weak cause can be reduced with polynomial overhead to an equivalent instance in which the (potential) weak cause and the causal model may contain fewer variables. That is, such reductions identify and remove irrelevant variables in weak causes and also in causal models. This can be regarded as an important preliminary step in the computation of weak and actual causes, which seems to be indispensable in efficient implementations.

We first describe a reduction from [7] and a generalization thereof in which irrelevant variables in weak causes $X = x$ of an event ϕ are characterized and removed. We then generalize these two reductions to two new reductions that identify and remove irrelevant variables in weak causes $X = x$ of ϕ and also in causal models M, producing the *reduced* and the *strongly reduced causal model* of M w.r.t. $X = x$ and an event ϕ. Both new reductions also generalize a reduction due to Hopkins [21] for events of the form $X = x$ and $\phi = \ Y = y$, where X and Y are singletons. The reduced causal model of M w.r.t. $X = x$ and ϕ is in general larger than its strong reduct w.r.t. $X = x$ and ϕ. But the former allows for deciding whether $X' = x'$ is a weak cause of ϕ, for the large class of all $X' \subseteq X$, while the latter generally allows only for deciding whether $X = x$ is a weak cause of ϕ.

In the rest of this section, to illustrate the removal of variables in (potential) weak causes and causal models, we use what is shown in Fig. 2: (i) the causal graph $G_V(M)$ of a causal model $M = (U, V, F)$, (ii) the set of variables $X \subseteq V$ of a (potential) weak cause $X = x$, and (iii) the set of variables $V(\phi)$ in an event ϕ.

4.1 Reducing Weak Causes

The following result (essentially proved in [7]) shows that deciding whether $X = x$ is a weak cause of ϕ under u is reducible to deciding whether $X' = x|X'$ is a weak cause of ϕ under u, where X' is the set of all $X_i \in X$ that are either in ϕ or ancestors of variables in ϕ. That is, in deciding whether $X = x$ is a weak cause of ϕ under u, we can safely ignore all variables in $X = x$ not connected to any variable in ϕ.

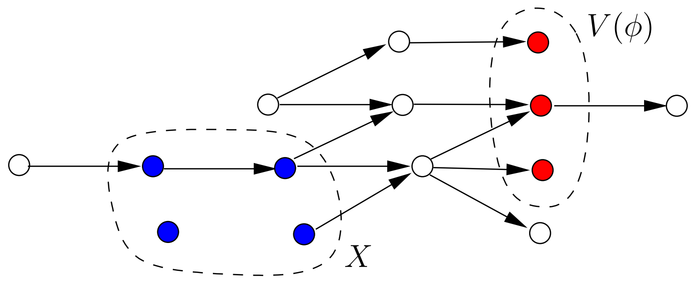

Figure 2: Causal Graph $G_V(M)$ along with X and $V(\phi)$

Theorem 4.1 (essentially [7]) *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' be the set of all variables in X from which a (directed) path exists in $G(M)$ to a variable in ϕ, and let $x' = x|X'$. Then, $X = x$ is a weak cause of ϕ under u iff (i) $(X \backslash X')(u) = x|(X \backslash X')$ and (ii) $X' = x'$ is a weak cause of ϕ under u.*

Example 4.1 Fig. 3 shows X' for a causal model $M = (U, V, F)$ and an event ϕ such that the causal graph $G_V(M)$ and the sets X and $V(\phi)$ are as in Fig. 2. □

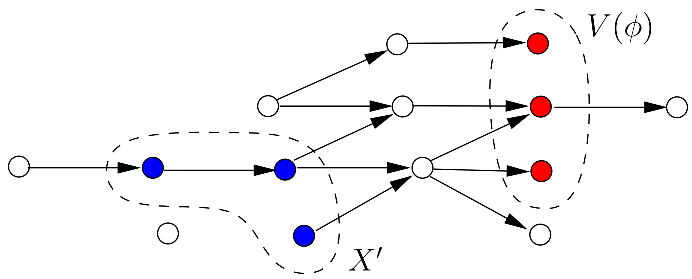

Figure 3: Causal Graph $G_V(M)$ along with X' and $V(\phi)$

The next theorem formulates the more general result that deciding whether $X = x$ is a weak cause of ϕ under u is reducible to deciding whether $X' = x|X'$ is a weak cause of ϕ under u, where X' is the set of all variables in X that occur in ϕ or that are ancestors of variables in ϕ not "blocked" by other variables in X. That is, in deciding whether $X = x$ is a weak cause of ϕ under u, we can even ignore every variable in $X = x$ not connected via variables in $V \backslash X$ to any variable in ϕ.

Theorem 4.2 *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' be the set of all variables $X_i \in X$ from which there exists a path in $G(M)$ to a variable in ϕ that contains no $X_j \in X \backslash \{X_i\}$, and let $x' = x|X'$. Then, $X = x$ is a weak cause of ϕ under u iff (i) $(X \backslash X')(u) = x|(X \backslash X')$ and (ii) $X' = x'$ is a weak cause of ϕ under u.*

Example 4.2 Fig. 4 shows X' for a causal model $M = (U, V, F)$ and an event ϕ such that the causal graph $G_V(M)$ and the sets X and $V(\phi)$ are as in Fig. 2. □

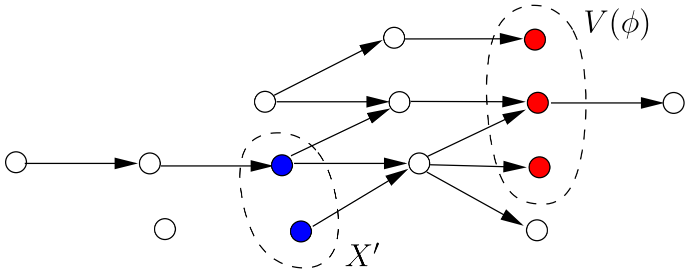

Figure 4: Causal Graph $G_V(M)$ along with X' and $V(\phi)$

The next result shows that computing the set of all variables in a weak cause that are not irrelevant according to Theorems 4.1 and 4.2 can be done in linear time.

Proposition 4.3 *Given a causal model* $M = (U, V, F)$, $X \subseteq V$, *and an event* ϕ,

(a) computing the set X' *of all variables* $X_i \in X$ *from which a path exists to a variable in* ϕ *can be done in time* $O(\|M\|+\|\phi\|)$.

(b) computing the set X' *of all variables* $X_i \in X$ *from which a path exists to a variable in* ϕ *that contains no* $X_j \in X \setminus \{X_i\}$ *can be done in time* $O(\|M\|+\|\phi\|)$.

4.2 Reducing Weak Causes and Causal Models

We now generalize the reduction described in Theorem 4.1 to a reduction which not only removes irrelevant variables from causes, but also removes irrelevant variables in causal models. In the sequel, let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, $x \in D(X)$, and $u \in D(U)$, and let ϕ be an event. We first define irrelevant variables w.r.t. $X = x$ and ϕ, and then the reduced causal model w.r.t. $X = x$ and ϕ, which does not contain these irrelevant variables anymore.

The set of *relevant* endogenous variables of $M = (U, V, F)$ relative to $X = x$ and ϕ, denoted $R^{\phi}_{X=x}(M)$, is the set of all $A \in V$ such that either **R1** or **R2** holds:

R1. A is on a directed path in $G(M)$ from a variable in $X \setminus \{A\}$ to a variable in ϕ.

R2. A does not satisfy **R1**, and either A occurs in ϕ, or A is a parent in $G(M)$ of a variable that satisfies **R1**.

Informally, $R^{\phi}_{X=x}(M)$ is the set of all variables in ϕ, all variables A that connect a different variable in X to one in ϕ, and all the parents of the latter variables. A variable $A \in V$ is *irrelevant* w.r.t. $X = x$ and ϕ iff it is not relevant w.r.t. $X = x$ and ϕ. Note that it does *not* necessarily hold that $X \subseteq R^{\phi}_{X=x}(M)$. The *reduced causal model* of $M = (U, V, F)$ w.r.t. $X = x$ and ϕ, denoted $M^{\phi}_{X=x}$, is the causal model $M' = (U, V', F')$ that is defined by the set of endogenous variables $V' = R^{\phi}_{X=x}(M)$ and the following set of functions $F' = \{F'_A \mid A \in V'\}$:

$$F' = \{F'_A = F_A \mid A \in V' \text{ satisfies } \mathbf{R1}\} \cup \{F'_A = F^{\star}_A \mid A \in V' \text{ satisfies } \mathbf{R2}\},$$

where $F^{\star}_A$ assigns $A_M(u_A)$ to A for every value $u_A \in D(U_A)$ of the set U_A of all ancestors $B \in U$ of A in $G(M)$. We often use $R^{\phi}_X(M)$, $R^Y_X(M)$, M^{ϕ}_X, and M^Y_X to abbreviate $R^{\phi}_{X=x}(M)$, $R^{Y=y}_{X=x}(M)$, $M^{\phi}_{X=x}$, and $M^{Y=y}_{X=x}$, respectively.

Example 4.3 Fig. 5 shows the causal graph $G_V(M^{\phi}_X)$ along with the set of variables $X' = X \cap R^{\phi}_X(M)$ for a causal model $M = (U, V, F)$ and an event ϕ such that the causal graph $G_V(M)$ and the sets X and $V(\phi)$ are as in Fig. 2. □

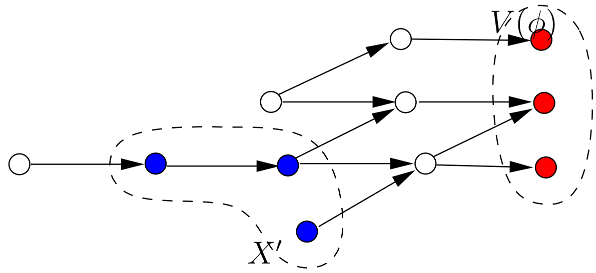

Figure 5: Causal Graph $G_V(M^{\phi}_X)$ along with $X' = X \cap R^{\phi}_X(M)$ and $V(\phi)$

The following result shows that a variable in $X = x$ is irrelevant w.r.t. $X = x$ and ϕ iff it is not connected to a variable in ϕ. Hence, we are heading towards a generalization of the reduction in Theorem 4.1.

Proposition 4.4 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, $x \in D(X)$, and let ϕ be an event. Then, $X \cap R^{\phi}_X(M)$ is the set of all variables $B \in X$ from which there exists a directed path in $G(M)$ to a variable in ϕ.*

The next result shows that deciding whether $X = x$ is a weak cause of ϕ under u in M can be reduced to deciding whether $X' = x'$ is a weak cause of ϕ under u in M^{ϕ}_X, where $X' = X \cap R^{\phi}_X(M)$ and $x' = x|X'$. It generalizes Theorem 4.1. Note that this result and also Theorems 4.7 and 4.10 below do not carry over to responsibility.

Theorem 4.5 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, $x \in D(X)$, and $u \in D(U)$, and let ϕ be an event. Let $X' = X \cap R^{\phi}_X(M)$ and $x' = x|X'$. Then, $X = x$ is a weak cause of ϕ under u in M iff (i) $(X \setminus X')(u) = x|(X \setminus X')$ in M, and (ii) $X' = x'$ is a weak cause of ϕ under u in M^{ϕ}_X.*

Our next result shows that the reduction of a causal model is monotonic in X. Roughly, if $X' \subseteq X$, then the reduced causal model of M w.r.t. $X' = x'$ and ϕ is essentially contained in the reduced causal model of M w.r.t. $X = x$ and ϕ.

Proposition 4.6 *Let $M = (U, V, F)$ be a causal model. Let $X' \subseteq X \subseteq V$, $x' \in D(X')$, $x \in D(X)$, and let ϕ be an event. Then, $M^{\phi}_{X'}$ coincides with $(M^{\phi}_X)^{\phi}_{X'}$.*

We are now ready to formulate the main result of this section. The following theorem shows that deciding whether $X' = x'$, where $X' \subseteq X$, is a weak cause of ϕ under u in M can be reduced to deciding whether its restriction to $R^{\phi}_X(M)$ is a weak cause of ϕ under u in M^{ϕ}_X. It is a generalization of Theorems 4.1 and 4.5, which follows from Theorem 4.5 and Proposition 4.6.

Theorem 4.7 *Let $M = (U, V, F)$ be a causal model. Let $X' \subseteq X \subseteq V$, $x' \in D(X')$, $x \in D(X)$, and $u \in D(U)$, and let ϕ be an event. Let $X'' = X' \cap R^{\phi}_X(M)$ and $x'' = x'|X''$. Then, $X' = x'$ is a weak cause* *of ϕ under u in M iff (i) $(X'\backslash X'')(u) = x'|(X'\backslash X'')$ in M, and (ii) $X'' = x''$ is a weak cause of ϕ under u in M^{ϕ}_{X}.*

The following result shows that the reduced causal model and the restriction of its causal graph to the set of endogenous variables can be computed in quadratic and linear time, respectively. Here, for any set S, we denote by $|S|$ its cardinality.

Proposition 4.8 *Given a causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, and an event ϕ, the directed graph $G_V(M^{\phi}_{X})$ (resp., the causal model M^{ϕ}_{X}) can be computed in time $O(\|M\| + \|\phi\|)$ (resp., in time $O(|V|\|M\| + \|\phi\|)$).*

4.3 Strongly Reducing Weak Causes and Causal Models

In the sequel, let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, $x \in D(X)$, and $u \in D(U)$, and let ϕ be an event. The reduced causal model w.r.t. $X = x$ and ϕ, which generalizes the idea behind Theorem 4.1, still contains some superfluous variables for deciding whether $X = x$ is a weak causes of ϕ under u in M. We now define the strongly reduced causal model w.r.t. $X = x$ and ϕ, which generalizes the idea behind Theorem 4.2, where these superfluous variables are removed. We first define strongly relevant variables w.r.t. $X = x$ and ϕ, and then the strongly reduced causal model w.r.t. $X = x$ and ϕ, which contains only such variables.

The set of *strongly relevant* endogenous variables of $M = (U, V, F)$ relative to $X = x$ and ϕ, denoted $\widehat{R}^{\phi}_{X=x}(M)$, is the set of all $A \in V$ such that either **S1** or **S2** holds:

S1. A is on a directed path P in $G(M)$ from a variable $B \in X \backslash \{A\}$ to a variable in ϕ, where P does not contain any variable from $X \backslash \{B\}$.

S2. A does not satisfy **S1**, and either A occurs in ϕ, or A is a parent in $G(M)$ of a variable that satisfies **S1**.

Note that all variables satisfying **S1** are from $V \backslash X$. Informally, $\widehat{R}^{\phi}_{X=x}(M)$ is the set of all variables in ϕ, all variables A that connect a variable in X to one in ϕ via variables from $V \backslash X$, and all the parents of the latter variables. Observe that $\widehat{R}^{\phi}_{X=x}(M) \subseteq R^{\phi}_{X=x}(M)$. The *strongly reduced causal model* of $M = (U, V, F)$ w.r.t. $X = x$ and ϕ, denoted $\widehat{M}^{\phi}_{X=x}$, is the causal model $M' = (U, V', F')$, where the endogenous variables $V' = \widehat{R}^{\phi}_{X=x}(M)$ and the functions $F' = \{F'_A \mid A \in V'\}$ are defined by:

$$F' = \{F'_A = F_A \mid A \in V' \text{ satisfies } \mathbf{S1}\} \cup \{F'_A = F^{\star}_A \mid A \in V' \text{ satisfies } \mathbf{S2}\},$$

where $F^{\star}_A$ assigns $A_M(u_A)$ to A for every value $u_A \in D(U_A)$ of the set U_A of all ancestors $B \in U$ of A in $G(M)$. We often use $\widehat{R}^{\phi}_{X}(M)$, $\widehat{R}^{Y}_{X}(M)$, $\widehat{M}^{\phi}_{X}$, and $\widehat{M}^{Y}_{X}$ to abbreviate $\widehat{R}^{\phi}_{X=x}(M)$, $\widehat{R}^{Y=y}_{X=x}(M)$, $\widehat{M}^{\phi}_{X=x}$, and $\widehat{M}^{Y=y}_{X=x}$, respectively.

Example 4.4 Fig. 6 shows the causal graph $G_V(\widehat{M}^{\phi}_{X})$ along with the set of variables $X' = X \cap \widehat{R}^{\phi}_{X}(M)$ for a causal model $M = (U, V, F)$ and an event ϕ such that the causal graph $G_V(M)$ and the sets X and $V(\phi)$ are as in Fig. 2. □

The following result shows that a variable in $X = x$ is strongly relevant w.r.t. $X = x$ and ϕ iff it is connected in $G(M)$ via variables in $V \backslash X$ to a variable in ϕ. Thus, we are currently elaborating a generalization of Theorem 4.2.

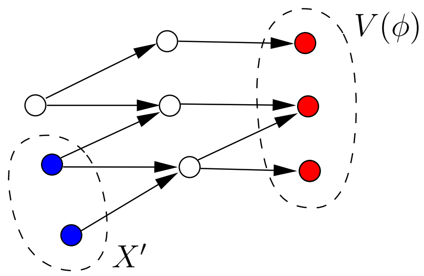

Figure 6: Causal Graph $G_V(\widehat{M}^{\phi}_X)$ along with $X' = X \cap \widehat{R}^{\phi}_X(M)$ and $V(\phi)$

Proposition 4.9 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, $x \in D(X)$, and let ϕ be an event. Then, $X \cap \widehat{R}^{\phi}_X(M)$ is the set of all $X_i \in X$ from which there exists a directed path in $G(M)$ to a variable in ϕ that contains no $X_j \in X \setminus \{X_i\}$.*

It is easy to verify that the monotonicity result of Proposition 4.6 and thus also Theorem 4.7 do not carry over to strongly reduced causal models. Informally, if $X' \subseteq X$, then $\widehat{M}^{\phi}_{X'}$ may contain variables that connect some $X_i \in X'$ to a variable in ϕ *via variables in* $V \setminus X'$, but that do not connect $X_i \in X' \subseteq X$ to a variable in ϕ *via variables in* $V \setminus X \subseteq V \setminus X'$, since some variable from $X \setminus X'$ is needed, and are thus not contained in $\widehat{M}^{\phi}_X$. For example, if the causal graph $G_V(M)$ and the sets X and $V(\phi)$ are as in Fig. 2, and X' consists of the variable in X that is shown upper left in Fig. 2, then this variable even does not occur among the variables of the strongly reduced causal model $\widehat{M}^{\phi}_{X=x}$, since it is pruned away (cf. also Figure 6), and hence $(\widehat{M}^{\phi}_X)^{\phi}_{X'}$ cannot be formed.

However, the weaker result in Theorem 4.5 also holds for strongly reduced causal models. That is, deciding whether $X = x$ is a weak cause of ϕ under u in M can be reduced to deciding whether its restriction to the strongly relevant variables is a weak cause of ϕ under u in $\widehat{M}^{\phi}_X$. This result generalizes Theorem 4.2.

Theorem 4.10 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, $x \in D(X)$, and $u \in D(U)$, and let ϕ be an event. Let $X' = X \cap \widehat{R}^{\phi}_X(M)$ and $x' = x|X'$. Then, $X = x$ is a weak cause of ϕ under u in M iff (i) $(X \setminus X')(u) = x|(X \setminus X')$ in M, and (ii) $X' = x'$ is a weak cause of ϕ under u in $\widehat{M}^{\phi}_X$.*

The following result shows that also the strongly reduced causal model and the restriction of its causal graph to the set of all endogenous variables can be computed in polynomial and linear time, respectively.

Proposition 4.11 *Given a causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, and an event ϕ, the directed graph $G_V(\widehat{M}^{\phi}_X)$ (resp., the causal model $\widehat{M}^{\phi}_X$) is computable in time $O(\|M\| + \|\phi\|)$ (resp., $O(|V|\|M\| + \|\phi\|)$).*

The next result shows that for weak causes of the form $X = x$, where X is a singleton, the reduced causal model coincides with the strongly reduced one. Observe that any such weak cause is an actual cause (by Theorem 2.3). Hence, for deciding actual causes, both reductions coincide. Nonetheless, we need weak causes in particular since they are a basic building block of explanations, not actual causes.

Theorem 4.12 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$, and let ϕ be an event. If X is a singleton, then $M^{\phi}_X = \widehat{M}^{\phi}_X$.*

5 Causal Trees

In this section, we describe our first class of tractable cases of causes and explanations. We show that deciding whether an atom $X = x$ is a weak cause of a primitive event $Y = y$ under a context u in a domain-bounded causal model $M = (U, V, F)$ is tractable, if the reduced causal model $G_V(M_X^Y)$ is a bounded directed tree with root Y, which informally consists of a directed path from X to Y, along with a number of parents for each variable in the path after X bounded by a global constant (see Fig. 7). Under the same conditions, deciding whether $X = x$ is an actual cause, deciding whether $X = x$ is an explanation relative to a set of contexts $\mathcal{C}$, and deciding whether $X = x$ is a partial explanation or an α-partial explanation as well as computing its explanatory power relative to $(\mathcal{C}, P)$ are all tractable.

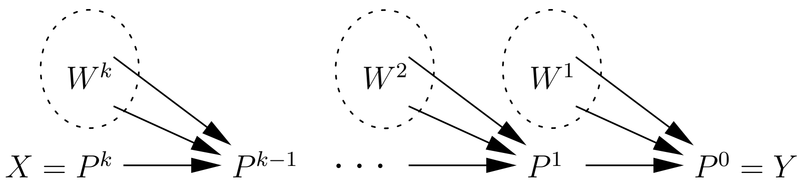

Figure 7: Path from X to Y in a Causal Tree

More precisely, we say that a directed graph $G = (V, E)$, given two nodes $X, Y \in V$, is a *directed tree with root* Y, if it consists of a unique directed path $X \mathrel{\hat{=}} P^k \rightarrow P^{k-1} \rightarrow \cdots \rightarrow P^0 \mathrel{\hat{=}} Y$ from X to Y, and sets W^i of (unconnected) parents $A \neq P^i$ for all P^{i-1} such that $i \in \{1, \ldots, k\}$. Moreover, G is *bounded*, if $|W_i| \leq l$ for each $i \in \{1, \ldots, k\}$, i.e., P_{i-1} has fan-in of variables from V at most $l + 1$, where l is some global constant. If $G = G_V(M)$ for some causal model $M = (U, V, F)$ and $X, Y \in V$, then M is a *(bounded) causal tree* with respect to X and Y.

Example 5.1 An example of a causal tree is the following binary causal model $M = (U, V, F)$ presented in [16] in a discussion of the double prevention problem, where $U = \{U_{BPT}, U_{SPS}\}$ with $D(A) = \{0, 1\}$ for all $A \in U$, $V = \{BPT, LE, LSS, SPS, SST, TD\}$ with $D(A) = \{0, 1\}$ for all $A \in V$. In a World War III scenario, Suzy is piloting a bomber on a mission to blow up an enemy target, and Billy is piloting a fighter as her lone escort. Along comes an enemy plane, piloted by Lucifer. Sharp-eyed Billy spots Lucifer, zooms in, pulls the trigger, and Lucifer's plane goes down in flames. Suzy's mission is undisturbed, and the bombing takes place as planned. The question is whether Billy deserves some credit for the success of the mission. Here, BPT means that Billy pulls the trigger, LE that Lucifer eludes Billy, LSS that Lucifer shoots Suzy, SPS that Suzy plans to shoot the target, SST that Suzy shoots the target, and TD that the target is destroyed.

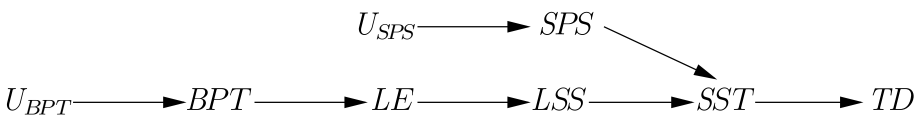

Figure 8: Causal Graph $G(M)$

The set $F = \{F_A \mid A \in V\}$ consists of the functions $F_{BPT} = U_{BPT}$, $F_{SPS} = U_{SPS}$, $F_{LE} = 1 - BPT$, $F_{LSS} = LE$, $F_{SST} = 1$ iff $LSS = 0$ and $SPS = 1$, and $F_{TD} = SST$. The causal graph $G(M)$ is shown in Fig. 8. Let $X = BPT$ and $Y = TD$. Then, $G_V(M)$ is a directed tree with root Y, where the directed path from X to Y is $P^4 = BPT$, $P^3 = LE$, $P^2 = LSS$, $P^1 = SST$, $P^0 = TD$, $W^1 = W^3 = W^4 = \emptyset$ and $W^2 = SPS$. $\square$

As an important property, causal trees can be recognized very efficiently, namely in linear time. The same holds for causal models whose reduced variant with respect to X and Y is a causal tree.

Proposition 5.1 *Given a causal model $M = (U, V, F)$ and variables $X, Y \in V$, deciding whether M resp. M_Y^X is a (unbounded or bounded) causal tree with respect to X and Y is feasible in $O(\|M\|)$ time.*

5.1 Characterizing Weak Causes

We first consider weak causes. In the sequel, let $M = (U, V, F)$ be a causal model, let $X, Y \in V$ such that M is a causal tree with respect to X and Y, and let $x \in D(X)$ and $y \in D(Y)$. We give a new characterization of $X = x$ being a weak cause of $Y = y$ under context $u \in D(U)$, which can be checked in polynomial time under some conditions. We need some preparation by the following definitions. We define $R^0 = \{D(Y) \setminus \{y\}\}$, and for every $i \in \{1, \ldots, k\}$, we define $\hat{p}^i = P^i(u)$ and R^i by:

$$
\begin{aligned}
R^i = \{\boldsymbol{p} \subseteq D(P^i) \mid \exists w \in D(W^i)\, \exists \boldsymbol{p'} \in R^{i-1} :& \\
& \forall p \in D(P^i):\ p \in \boldsymbol{p} \text{ iff } P^{i-1}_{pw}(u) \in \boldsymbol{p'}\,; \\
& P^{i-1}_{\hat{p}^i w}(u) \in D(P^{i-1}) \setminus \boldsymbol{p'}\,\}\,.
\end{aligned}
$$

Intuitively, R^i contains all sets of possible values of P^i in **AC2**(a), for different values of W^i. Here, $P^0 \mathrel{\hat{=}} Y$ must be set to a value different from y, and the possible values of each other P^i depend on the possible values of P^{i-1}. At the same time, the complements of sets in R^i are all sets of possible values of P^i in **AC2**(b). In summary, **AC2**(a) and (b) hold iff some nonempty $\boldsymbol{p} \in R^k$ exists that does not contain x. This result is formally expressed by the following theorem, which can be proved by induction on $i \in \{0, \ldots, k\}$.

Theorem 5.2 *Let $M = (U, V, F)$ be a causal model. Let $X, Y \in V$, $x \in D(X)$, $y \in D(Y)$, and $u \in D(U)$. Suppose that M is a causal tree w.r.t. X and Y, and let R^k be defined as above. Then, $X = x$ is a weak cause of $Y = y$ under u in M iff (α) $X(u){=}x$ and $Y(u){=}y$ in M, and (β) some $\boldsymbol{p} \in R^k$ exists with $\boldsymbol{p} \neq \emptyset$ and $x \notin \boldsymbol{p}$.*

Example 5.2 Consider again the causal tree with respect to $X = BPT$ and $Y = TD$ from Example 5.1. Suppose we want to decide whether $BPT = 1$ is a weak cause of $TD = 1$ under a context $u_{1,1} \in D(U)$, where $u_{1,1}(U_{BPT}) = 1$ and $u_{1,1}(U_{SPS}) = 1$. Here, we obtain the relations $R^0 = \{\{0\}\}$, $R^1 = \{\{0\}\}$, $R^2 = \{\{1\}\}$, $R^3 = \{\{1\}\}$, and $R^4 = \{\{0\}\}$. Observe then that (α) $BPT(u_{1,1})$ and $TD(u_{1,1})$ are both 1, and (β) $\{0\} \in R^4$ and $1 \notin \{0\}$. By Theorem 5.2, it thus follows that $BPT = 1$ is a weak cause of $TD = 1$ under $u_{1,1}$. $\square$

5.2 Deciding Weak and Actual Causes

The following theorem shows that deciding whether an atom $X = x$ is a weak cause of a primitive event $Y = y$ in domain-bounded M is tractable, when M is a bounded causal tree with respect to X and Y. This result follows from Theorem 5.2 and the recursive definition of R^i, which assures that R^k can be computed in polynomial time under the above boundedness assumptions. By Theorem 2.3, the same tractability result holds for actual causes, since the notion of actual cause coincides with the notion of weak cause where X is a singleton.

Theorem 5.3 *Given a domain-bounded causal model $M = (U, V, F)$, variables $X, Y \in V$ such that M is a bounded causal tree with respect to X and Y, and values $x \in D(X)$, $y \in D(Y)$, and $u \in D(U)$, deciding whether $X = x$ is a weak (resp., an actual) cause of $Y = y$ under u in M can be done in polynomial time.*

The next theorem shows that the same tractability result holds when instead of M just the reduced model M^X_Y is required to be a bounded causal tree. The result follows from Theorem 4.5, Proposition 4.8, and Theorem 5.3.

Theorem 5.4 *Given a domain-bounded causal model $M = (U, V, F)$, variables $X, Y \in V$ such that M^Y_X is a bounded causal tree with respect to X and Y, values $x \in D(X)$, $y \in D(Y)$, and $u \in D(U)$, deciding whether $X = x$ is a weak (resp., an actual) cause of $Y = y$ under u in M can be done in polynomial time.*

5.3 Deciding Explanations and Partial Explanations

The following theorem shows that deciding whether $X = x$ is an explanation of $Y = y$ relative to $\mathcal{C}$ in M is tractable under the conditions of the previous subsection. This result follows from Theorem 5.4 and Proposition 2.2.

Theorem 5.5 *Given a domain-bounded causal model $M = (U, V, F)$, variables $X, Y \in V$ such that M^Y_X is a bounded causal tree with respect to X and Y, values $x \in D(X)$ and $y \in D(Y)$, and a set of contexts $\mathcal{C} \subseteq D(U)$, deciding whether $X = x$ is an explanation of $Y = y$ relative to $\mathcal{C}$ in M can be done in polynomial time.*

Similarly, deciding whether $X = x$ is a partial or an α-partial explanation of $Y = y$ relative to $(\mathcal{C}, P)$ in M, as well as computing its explanatory power is tractable under the conditions of the previous subsection. This follows from Theorem 5.4 and Propositions 2.2 and 2.4.

Theorem 5.6 *Let $M = (U, V, F)$ be a domain-bounded causal model, let $X, Y \in V$ be such that M^Y_X is a bounded causal tree with respect to X and Y, and let $x \in D(X)$ and $y \in D(Y)$. Let $\mathcal{C} \subseteq D(U)$ such that $Y(u) = y$ for all $u \in \mathcal{C}$, and let P be a probability function on $\mathcal{C}$. Then,*

(a) deciding whether $X = x$ is a partial explanation of $Y = y$ relative to $(\mathcal{C}, P)$ in M can be done in polynomial time.

(b) deciding whether $X = x$ is an α-partial explanation of $Y = y$ relative to $(\mathcal{C}, P)$ in M, for some given $\alpha \geq 0$ can be done in polynomial time.

(c) given that $X = x$ is a partial explanation of $Y = y$ relative to $(\mathcal{C}, P)$ in M, the explanatory power of $X = x$ is computable in polynomial time.

6 Decomposable Causal Graphs

In this section, we generalize the characterization of weak cause given in Section 5 to more general events and to more general causal graphs. We characterize relationships of the form "$X = x$ is a weak cause of ϕ under u in M", where (i) $X = x$ and ϕ are as in the original definition of weak cause, and thus not restricted to assignments to single variables anymore, and where (ii) $G_V(M)$ is decomposable into a chain of subgraphs (cf. Fig. 9, which is explained in more detail below), and thus not restricted to causal trees anymore. We then use this result to obtain more general tractability results for causes and explanations, and also new tractability results for responsibility and blame.

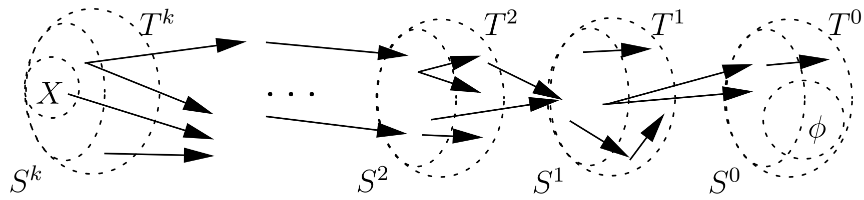

Figure 9: Decomposable Causal Graph

6.1 Characterizing Weak Causes

We first give a new characterization of weak cause. In the sequel, let $M = (U, V, F)$ be a causal model, let $X \subseteq V$, let $x \in D(X)$ and $u \in D(U)$, and let ϕ be an event.

Towards a characterization of "$X = x$ is a weak cause of ϕ under u in M", we define the notion of a decomposition of a causal graph as follows. A *decomposition* of $G_V(M)$ relative to $X = x$ (or simply X) and ϕ is a tuple $((T^0, S^0), \ldots, (T^k, S^k))$, $k \geq 0$, of pairs (T^i, S^i) such that the conditions **D1**–**D6** hold:

D1. $(T^0, \ldots, T^k)$ is an ordered partition of V.

D2. $T^0 \supseteq S^0, \ldots, T^k \supseteq S^k$.

D3. Every $A \in V$ occurring in ϕ belongs to T^0, and $S^k \supseteq X$.

D4. For every $i \in \{0, \ldots, k-1\}$, no two variables $A \in T^0 \cup \cdots \cup T^{i-1} \cup T^i \setminus S^i$ and $B \in T^{i+1} \cup \cdots \cup T^k$ are connected by an arrow in $G_V(M)$.

D5. For every $i \in \{1, \ldots, k\}$, every child of a variable from S^i in $G_V(M)$ belongs to $(T^i \setminus S^i) \cup S^{i-1}$. Every child of a variable from S^0 belongs to $(T^0 \setminus S^0)$.

D6. For every $i \in \{0, \ldots, k-1\}$, every parent of a variable in S^i in $G_V(M)$ belongs to T^{i+1}. There are no parents of any variable $A \in S^k$.

Intuitively, $G_V(M)$ is decomposable into a chain of edge-disjoint subgraphs $G^0, \ldots, G^k$ over some sets of variables $T^0, S^0 \cup T^1, S^1 \cup T^2, \ldots, S^{k-1} \cup T^k$, where $(T^0, \ldots, T^k)$ is an ordered partition of V, such that the sets T_i are connected to each other exactly through some sets $S^i \subseteq T^i$, $i \in \{0, \ldots, k-1\}$, where (i) every arrow that is incident to some $A \in S^i$, $i \in \{1, \ldots, k-1\}$, is either outgoing from A and belongs to G^i, or ingoing into A and belongs to G^{i+1}, and (ii) every variable in ϕ (resp., X) belongs to T^0 (resp., some $S^k \subseteq T^k$); see Fig. 9 for an illustration.

As easily seen, causal trees as in Section 5 are causal models with decomposable causal graphs. For the directed path $X \hat{=} P^k \rightarrow P^{k-1} \rightarrow \cdots \rightarrow P^0 \hat{=} Y$ from X to Y, and the sets W^i, $i \in \{1, \ldots, k\}$, we may define $\mathcal{D} = ((T^0, S^0), \ldots, (T^k, S^k))$ by $S^i = \{P^i\}$, $T^0 = \{P^0\}$, and $T^i = W^i \cup \{P^i\}$, for $i \in \{1, \ldots, k\}$; then, $\mathcal{D}$ is a decomposition of $G_V(M)$ relative to $X = x$ and $Y = y$.

The *width* of a decomposition $\mathcal{D} = ((T^0, S^0), \ldots, (T^k, S^k))$ of $G_V(M)$ relative to X and ϕ is the maximum of all $|T^i|$ such that $i \in \{0, \ldots, k\}$. We say that $\mathcal{D}$ is *width-bounded* iff the width of $\mathcal{D}$ is at most l for some global constant l.

Example 6.1 Fig. 10 shows a decomposition $\mathcal{D} = ((T^0, S^0), (T^1, S^1), (T^2, S^2))$ of a causal graph $G_V(M)$ relative to a set of variables $X \subseteq V$ and an event ϕ. The width of this decomposition $\mathcal{D}$ is given by 6. □

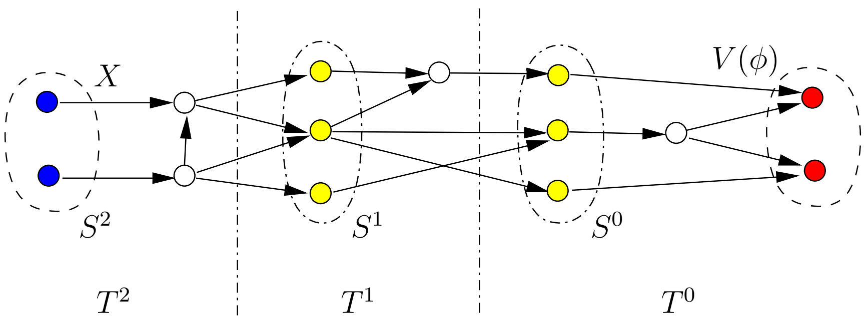

Figure 10: Decomposition $((T^0, S^0), (T^1, S^1), (T^2, S^2))$ of $G_V(M)$ relative to X and ϕ

We use such a decomposition $((T^0, S^0), \ldots, (T^k, S^k))$ of $G_V(M)$ to extend the relations R^i for causal trees from Section 5.1 to decomposable causal graphs. The new relations R^i now contain triples $(\boldsymbol{p}, \boldsymbol{q}, F)$, where $\boldsymbol{p}$ (resp., $\boldsymbol{q}$) specifies a set of possible values of "floating variables" $F \subseteq S^i$ in **AC2**(a) (resp., **AC2**(b)). We define R^0 as follows:

$$
\begin{aligned}
R^0 = \{(\boldsymbol{p}, \boldsymbol{q}, F) \mid\; & F \subseteq S^0,\ \boldsymbol{p}, \boldsymbol{q} \subseteq D(F),\\
& \exists W \subseteq T^0,\ F = S^0 \setminus W,\\
& \exists w \in D(W)\, \forall p, q \in D(F):\\
& p \in \boldsymbol{p} \ \text{ iff } \ \neg\phi_{pw}(u),\\
& q \in \boldsymbol{q} \ \text{ iff } \ \phi_{[q\langle\hat{Z}(u)]w}(u) \ \text{ for all } \hat{Z} \subseteq T^0 \setminus (S^k \cup W)\}\,.
\end{aligned}
$$

For every $i \in \{1, \ldots, k\}$, we then define R^i as follows:

$$
\begin{aligned}
R^i = \{(\boldsymbol{p}, \boldsymbol{q}, F) \mid\, & F \subseteq S^i,\ \boldsymbol{p}, \boldsymbol{q} \subseteq D(F), \\
& \exists W \subseteq T^i,\ F = S^i \setminus W, \\
& \exists w \in D(W)\, \exists(\boldsymbol{p'}, \boldsymbol{q'}, F') \in R^{i-1}\, \forall p, q \in D(F): \\
& p \in \boldsymbol{p} \text{ iff } F'_{pw}(u) \in \boldsymbol{p'}, \\
& q \in \boldsymbol{q} \text{ iff } F'_{[q\langle \hat{Z}(u)]w}(u) \in \boldsymbol{q'} \text{ for all } \hat{Z} \subseteq T^i \setminus (S^k \cup W)\}\,.
\end{aligned}
$$

Intuitively, rather than propagating a set $\boldsymbol{p}$ of possible values of a single variable $P^i \in V$ in **AC2**(a) from $Y = P^0$ back to $X = P^k$ along a path $X \hat{=} P^k \rightarrow P^{k-1} \rightarrow \cdots \rightarrow P^0 \hat{=} Y$ as in Section 5.1, we now propagate triples $(\boldsymbol{p}, \boldsymbol{q}, F)$ consisting of some "floating variables" $F \subseteq S^i \subseteq V$, a set $\boldsymbol{p}$ of possible values of F in **AC2**(a), and a set $\boldsymbol{q}$ of possible values of F in **AC2**(b), from ϕ back to $X \subseteq V$ along the chain of subgraphs $G^0, \ldots, G^k$ over the sets of variables $T^0, S^0 \cup T^1, S^1 \cup T^2, \ldots, S^{k-1} \cup T^k$. Here, R^0 contains all triples $(\boldsymbol{p}, \boldsymbol{q}, F)$ such that $F \subseteq S^0$, $\boldsymbol{p}, \boldsymbol{q} \subseteq D(F)$, $p \in \boldsymbol{p}$ iff $\neg\phi_{pw}(u)$, and $q \in \boldsymbol{q}$ iff $\phi_{[q\langle \hat{Z}(u)]w}(u)$, for all possible $\hat{Z}$ and some appropriate w. Moreover, the triples in R^i depend on the triples in R^{i-1}. In summary, it then follows that **AC2**(a) and (b) hold iff some $(\boldsymbol{p}, \boldsymbol{q}, X) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$.

Note that for a decomposition corresponding to a causal tree as discussed above, for each $(\boldsymbol{p}, \boldsymbol{q}, F)$ in R^0, it holds that $W = \emptyset$ and $F = \{P^0\}$; hence $\boldsymbol{q} = D(F) \backslash \boldsymbol{p}$ is the complement of $\boldsymbol{p}$. Furthermore, for each $(\boldsymbol{p}, \boldsymbol{q}, F)$ in R^i, where $i > 0$, we have $W = W^i$ and $F = \{P^i\}$ and $\boldsymbol{q} = D(P^i) \backslash \boldsymbol{p}$ is the complement of $\boldsymbol{p}$. That is, the sets R^i defined for causal trees correspond to simplified versions of the sets R^i for a decomposed graph, where the redundant components F and $\boldsymbol{q}$ are removed from each triple.

This new characterization of weak cause, which is based on the above concept of a decomposition of $G_V(M)$ and the relations R^i, is expressed by the following theorem, which can be proved by induction on $i \in \{0, \ldots, k\}$.

Theorem 6.1 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, let $x \in D(X)$ and $u \in D(U)$, and let ϕ be an event. Let $((T^0, S^0), \ldots, (T^k, S^k))$ be a decomposition of $G_V(M)$ relative to X and ϕ, and let R^k be defined as above. Then, $X = x$ is a weak cause of ϕ under u in M iff (α) $X(u) = x$ and $\phi(u)$ in M, and (β) some $(\boldsymbol{p}, \boldsymbol{q}, X) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$.*

This result provides a basis for deciding and computing weak and actual causes, and may in particular be fruitfully applied to reduced causal models from which irrelevant variables have been pruned. Often, reduced models have a simple decomposition: Every $\widehat{M}_X^{\phi} = (U, V', F')$ has the trivial decomposition $((V', X))$, and every $M_X^{\phi} = (U, V', F')$ such that no $A \in X$ is on a path from a different variable in X to a variable in ϕ also has the trivial decomposition $((V', X))$.

6.2 Deciding and Computing Weak and Actual Causes

Using the characterization of weak cause given in Section 6.1, we now provide new tractability results for deciding and computing weak and actual causes. The following theorem shows that deciding whether $X = x$ is a weak (resp., an actual) cause of ϕ under u in a domain-bounded M is tractable when $G_V(M)$ has a width-bounded decomposition provided in the input. As for its proof, by Theorem 6.1, deciding whether $X = x$ is a weak cause of ϕ under u in M can be done by recursively computing the R^i's and then deciding whether (α) and (β) of Theorem 6.1 hold. All this can be done in polynomial time under the above boundedness assumptions. By Theorem 2.3, actual causes are weak causes $X = x$ such that X is a singleton. Thus, since deciding whether X is a singleton can be done in constant time, the above tractability result also holds for actual causes.

Theorem 6.2 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, an event ϕ, and a width-bounded decomposition $\mathcal{D}$ of $G_V(M)$ relative to X and ϕ, deciding whether $X = x$ is a weak (resp., an actual) cause of ϕ under u in M is possible in polynomial time.*

The next theorem shows that deciding weak (resp., actual) causes in domain-bounded causal models is also tractable, when $G_V(M^{\phi}_X)$ has a width-bounded decomposition provided in the input. This result essentially combines Theorems 4.7 and 6.2.

Theorem 6.3 *Given a domain-bounded causal model $M = (U, V, F)$, $X' \subseteq X \subseteq V$, $x' \in D(X')$, $u \in D(U)$, an event ϕ, and a width-bounded decomposition $\mathcal{D}$ of the graph $G_V(M^{\phi}_X)$ relative to $X' \cap R^{\phi}_X(M)$ and ϕ, deciding whether $X' = x'$ is a weak (resp., an actual) cause of ϕ under u in M is possible in polynomial time.*

A similar result also holds for strongly reduced causal models. It is expressed by the following theorem, which basically combines Theorems 4.10 and 6.2.

Theorem 6.4 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, an event ϕ, and a width-bounded decomposition $\mathcal{D}$ of the graph $G_V(\widehat{M}^{\phi}_X)$ relative to $X \cap \widehat{R}^{\phi}_X(M)$ and ϕ, deciding whether $X = x$ is a weak (resp., an actual) cause of ϕ under u in M is possible in polynomial time.*

We finally focus on computing weak and actual causes. The following result shows that, given some $X \subseteq V$, computing all weak (resp., actual) causes $X' = x'$, where $X' \subseteq X$ and $x' \in D(X')$, of ϕ under u in domain-bounded M is tractable, when either (a) $G_V(M^{\phi}_X)$ has a width-bounded decomposition provided in the input, or (b) every $G_V(\widehat{M}^{\phi}_{X'})$ with $X' \subseteq X$ has a width-bounded decomposition provided in the input. This result essentially follows from Theorems 6.3 and 6.4. Observe that in Theorems 6.5 to 6.9, each of (a) and (b) implies that $|X|$ is bounded by a constant, and thus also the number of all subsets $X' \subseteq X$ is bounded by a constant. Theorems 6.5 to 6.9 also hold, when the decompositions are relative to $X \cap R^{\phi}_X(M)$ and $X' \cap \widehat{R}^{\phi}_{X'}(M)$ rather than X and X', respectively.

Theorem 6.5 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $u \in D(U)$, an event ϕ, and either (a) a width-bounded decomposition $\mathcal{D}$ of the graph $G_V(M^{\phi}_X)$ relative to X and ϕ, or (b) for every $X' \subseteq X$, a width-bounded decomposition $\mathcal{D}_{X'}$ of $G_V(\widehat{M}^{\phi}_{X'})$ relative to X' and ϕ, computing the set of all $X' = x'$, where $X' \subseteq X$ and $x' \in D(X')$, such that $X' = x'$ is a weak (resp., an actual) cause of ϕ under u in M is possible in polynomial time.*

6.3 Deciding and Computing Explanations and Partial Explanations

We now turn to deciding and computing explanations and partial explanations. The following theorem shows that deciding whether $X = x$ is an explanation of ϕ relative to $\mathcal{C} \subseteq D(U)$ in M can be done in polynomial time, if we assume the same restrictions as in Theorem 6.5. This result follows from Theorems 6.3 and 6.4.

Theorem 6.6 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, an event ϕ, and either (a) a width-bounded decomposition $\mathcal{D}$ of $G_V(M_X^\phi)$ relative to X and ϕ, or (b) for each $X' \subseteq X$, a width-bounded decomposition $\mathcal{D}_{X'}$ of $G_V(\widehat{M}_{X'}^\phi)$ relative to X' and ϕ, deciding whether $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ in M can be done in polynomial time.*

A similar tractability result holds for deciding whether $X = x$ is a partial or an α-partial explanation of ϕ relative to some $(\mathcal{C}, P)$ in M, and for computing the explanatory power of a partial explanation.

Theorem 6.7 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $\mathcal{C} \subseteq D(U)$, an event ϕ, where $\phi(u)$ for all $u \in \mathcal{C}$, a probability function P on $\mathcal{C}$, and either (a) a width-bounded decomposition $\mathcal{D}$ of $G_V(M_X^\phi)$ relative to X and ϕ, or (b) for every $X' \subseteq X$, a width-bounded decomposition $\mathcal{D}_{X'}$ of $G_V(\widehat{M}_{X'}^\phi)$ relative to X' and ϕ,*

(1) deciding whether $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ in M can be done in polynomial time.

(2) deciding whether $X = x$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ in M, for some given $\alpha \geq 0$, can be done in polynomial time.

(3) given that $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ in M, computing the explanatory power of $X = x$ can be done in polynomial time.

Such tractability results also hold for computing explanations and partial explanations. In particular, the next theorem shows that computing all explanations involving variables from a given set of endogenous variables is tractable under the same assumptions as in Theorem 6.5.

Theorem 6.8 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $\mathcal{C} \subseteq D(U)$, an event ϕ, and either (a) a width-bounded decomposition $\mathcal{D}$ of $G_V(M_X^\phi)$ relative to X and ϕ, or (b) for every $X' \subseteq X$, a width-bounded decomposition $\mathcal{D}_{X'}$ of $G_V(\widehat{M}_{X'}^\phi)$ relative to X' and ϕ, computing the set of all $X' = x'$, where $X' \subseteq X$ and $x' \in D(X')$, such that $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$ in M can be done in polynomial time.*

Similarly, also computing all partial and α-partial explanations involving variables from a given set of endogenous variables is tractable under the same restrictions as in Theorem 6.5.

Theorem 6.9 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $\mathcal{C} \subseteq D(U)$, an event ϕ, where $\phi(u)$ for all $u \in \mathcal{C}$, a probability function P on $\mathcal{C}$, and either (a) a width-bounded decomposition $\mathcal{D}$ of $G_V(M_X^\phi)$ relative to X and ϕ, or (b) for every $X' \subseteq X$, a width-bounded decomposition $\mathcal{D}_{X'}$ of $G_V(\widehat{M}_{X'}^\phi)$ relative to X' and ϕ,*

(1) computing the set of all $X' = x'$, where $X' \subseteq X$ and $x' \in D(X')$, such that $X' = x'$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ in M can be done in polynomial time.

(2) computing the set of all $X' = x'$ where $X' \subseteq X$ and $x' \in D(X')$, such that $X' = x'$ is an α-partial explanation of ϕ relative to $(\mathcal{C}, P)$ in M, for some given $\alpha \geq 0$, can be done in polynomial time.

6.4 Computing Degrees of Responsibility and Blame

We now show that the tractability results for deciding and computing causes and explanations of Sections 6.2 and 6.3 can also be extended to computing degrees of responsibility and blame. To this end, we slightly generalize the relations R^i, $i \in \{0, \ldots, k\}$, of Section 6.1.

We use the following notation. For sets of variables X and values $x, x' \in D(X)$, the *difference between x and x'*, denoted $\mathrm{diff}(x, x')$, is the number of all variables $A \in X$ such that $x(A) \neq x'(A)$.

We define R^0 as follows:

$$\begin{aligned} R^0 = \{(\boldsymbol{p}, \boldsymbol{q}, F, l) \mid\ & F \subseteq S^0,\ \boldsymbol{p}, \boldsymbol{q} \subseteq D(F),\ l \in \{0, \ldots, |T^0|\}, \\ & \exists W \subseteq T^0,\ F = S^0 \setminus W, \\ & \exists w \in D(W)\, \forall p, q \in D(F): \\ & l = \mathrm{diff}(w, W(u)),\ \ p \in \boldsymbol{p} \text{ iff } \neg\phi_{pw}(u), \\ & q \in \boldsymbol{q} \text{ iff } \phi_{[q\langle \hat{Z}(u)\rangle]w}(u) \text{ for all } \hat{Z} \subseteq T^0 \setminus (S^k \cup W)\}\,. \end{aligned}$$

For every $i \in \{1, \ldots, k\}$, we then define R^i as follows:

$$\begin{aligned} R^i = \{(\boldsymbol{p}, \boldsymbol{q}, F, l) \mid\ & F \subseteq S^i,\ \boldsymbol{p}, \boldsymbol{q} \subseteq D(F),\ l \in \{0, \ldots, \textstyle\sum_{j=0}^{i} |T^j|\}, \\ & \exists W \subseteq T^i,\ F = S^i \setminus W, \\ & \exists w \in D(W)\, \exists (\boldsymbol{p'}, \boldsymbol{q'}, F', l') \in R^{i-1}\, \forall p, q \in D(F): \\ & l = \mathrm{diff}(w, W(u)) + l',\ \ p \in \boldsymbol{p} \text{ iff } F'_{pw}(u) \in \boldsymbol{p'}, \\ & q \in \boldsymbol{q} \text{ iff } F'_{[q\langle \hat{Z}(u)\rangle]w}(u) \in \boldsymbol{q'} \text{ for all } \hat{Z} \subseteq T^i \setminus (S^k \cup W)\}\,. \end{aligned}$$

Intuitively, rather than triples $(\boldsymbol{p}, \boldsymbol{q}, F)$, the new relations R^i contain tuples $(\boldsymbol{p}, \boldsymbol{q}, F, l)$, where $\boldsymbol{p}$ (resp., $\boldsymbol{q}$) is a set of possible values of $F \subseteq S^i$ in **AC2**(a) (resp., (b)) as in Section 6.1, and l is the sum of all differences between $w \in D(W)$ and $W(u)$ in T^j for all $j \in \{0, \ldots, i\}$. Thus, **AC2** holds with some $W \subseteq V \setminus X$ and $w \in D(W)$ such that $\mathrm{diff}(w, W(u)) = l$ iff some $(\boldsymbol{p}, \boldsymbol{q}, X, l) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$. This result is expressed by the following theorem.

Theorem 6.10 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, let $x \in D(X)$ and $u \in D(U)$, and let ϕ be an event. Let $((T^0, S^0), \ldots, (T^k, S^k))$ be a decomposition of $G_V(M)$ relative to X and ϕ, and let R^k be defined as above. Then,* **AC2** *holds with some $W \subseteq V \setminus X$ and $w \in D(W)$ such that* $\mathrm{diff}(w, W(u)) = l$ *iff some $(\boldsymbol{p}, \boldsymbol{q}, X, l) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$.*

The next theorem shows that the degree of responsibility of $X = x$ for ϕ in a situation (M, u) with domain-bounded M can be computed in polynomial time given that $G_V(M)$ has a width-bounded decomposition provided in the input. It follows from Theorem 6.10 and the fact that recursively computing the R^i's and deciding whether there exists some $(\boldsymbol{p}, \boldsymbol{q}, X, l) \in R^k$ with $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$ can be done in polynomial time under the above boundedness assumptions.

Theorem 6.11 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, an event ϕ, and a width-bounded decomposition $\mathcal{D}$ of $G_V(M)$ relative to X and ϕ, computing the degree of responsibility of $X = x$ for ϕ in (M, u) is possible in polynomial time.*

Similarly, computing the degree of blame relative to an epistemic state $(\mathcal{K}, P)$ is tractable, when every causal model in $\mathcal{K}$ satisfies the same boundedness assumptions as in Theorem 6.11. This is expressed by the following theorem.

Theorem 6.12 *Given an epistemic state $(\mathcal{K}, P)$, where for every $(M,u) \in \mathcal{K}$, M is domain-bounded, a set of endogenous variables X, a value $x \in D(X)$, an event ϕ, and for every $(M,u) = ((U,V,F),u) \in \mathcal{K}$ a width-bounded decomposition of $G_V(M)$ relative to X and ϕ, computing the degree of blame of setting X to x for ϕ relative to $(\mathcal{K}, P)$ is possible in polynomial time.*

6.5 Computing Decompositions

The tractability results of Sections 6.2 to 6.4 are all based on the assumption that some decomposition of $G_V(M)$ is provided in the input. It thus remains to decide whether such decompositions exist at all, and if so, then to compute one, especially one of minimal width. The problem of deciding whether there exists some decomposition of width below a given integer $l > 0$ is formally expressed as follows.

LAYERWIDTH WITH CONSTRAINTS: Given $G_V(M)$ for $M = (U,V,F)$, $X \subseteq V$, an event ϕ, and an integer $l > 0$, decide whether there exists a decomposition $((T^0, S^0), \ldots, (T^k, S^k))$ of $G_V(M)$ relative to X and ϕ of width at most l.

As shown by Hopkins [22], LAYERWIDTH WITH CONSTRAINTS is NP-complete. Hopkins [22] also presents an algorithm for computing a layer decomposition of lowest width, where a *layer decomposition* satisfies every condition among **D1** to **D6** except for **D3**. It is an any-time depth-first branch-and-bound algorithm, which searches through a binary search tree that represents the set of all possible layer decompositions. This algorithm can also be used to compute the set of all decompositions of $G_V(M)$ relative to X and ϕ of lowest width.

The intractability of computing a decomposition of lowest width, which is a consequence of the NP-completeness of LAYERWIDTH WITH CONSTRAINTS, is not such a negative result as it might appear at first glance. It means that decompositions are an expressive concept, for which sophisticated algorithms like Hopkin's are needed to obtain good performance. However, the effort for decomposition pays off by subsequent polynomial-time solvability of a number of reasoning tasks given that the ramifying conditions are met, such that overall, the effort is polynomial time modulo calls to an NP-oracle. This complexity is lower than the complexity of weak and actual causes, as well as the complexity of explanations in the general case, which are located at the second and the third level of the Polynomial Hierarchy, respectively [6, 7] (see also Section 3.4). On the other hand, the lower complexity means that suitable decompositions will not always exist. However, the worst-case complexity results in [6, 7] use quite artificial constructions, and the causal models involved will hardly occur in practice. In fact, many of the examples in the literature seem to have decomposable causal graphs; it remains to be seen whether this holds for a growing stock of applications.

7 Layered Causal Graphs

In general, as described in Section 6.5, causal graphs $G_V(M)$ with width-bounded decompositions cannot be efficiently recognized, and such decompositions also cannot be efficiently computed. But, from Section 5, we already know width-bounded causal trees as a large class of causal graphs, which have width-bounded decompositions that can be computed in linear time. In this section, we discuss an even larger class of causal graphs, called *layered causal graphs*, which also have natural nontrivial decompositions that can be computed in linear time. Intuitively, in layered causal graphs $G_V(M)$, the set of endogenous variables V can be partitioned into *layers* $S^0, \ldots, S^k$ such that every arrow in $G_V(M)$ goes from a variable in some layer S^i to a variable in S^{i-1} (see Fig. 11).

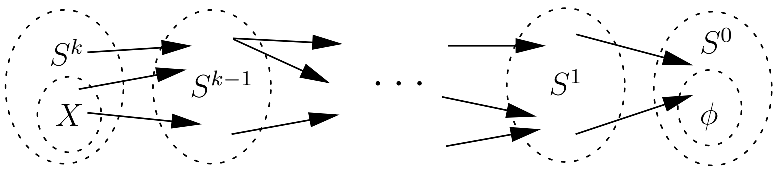

Figure 11: Path from X to ϕ in a Layered Causal Graph

We now first define layered causal graphs. We then prove that they are a special case of decomposable causal graphs, and that recognizing them and computing their layers can be done in linear time. In the sequel, let $M = (U, V, F)$ be a causal model, let $X \subseteq V$, let $x \in D(X)$ and $u \in D(U)$, and let ϕ be an event.

More formally, a *layering* of $G_V(M)$ relative to X and ϕ is an ordered partition $(S^0, \ldots, S^k)$ of V that satisfies the following conditions **L1** and **L2**:

L1. For every arrow $A \to B$ in $G_V(M)$, there exists some $i \in \{1, \ldots, k\}$ such that $A \in S^i$ and $B \in S^{i-1}$.

L2. Every $A \in V$ occurring in ϕ belongs to S^0, and $S^k \supseteq X$.

We say that $G_V(M)$ is *layered* relative to X and ϕ iff it has a layering $(S^0, \ldots, S^k)$ relative to X and ϕ. The *width* of such a layering $(S^0, \ldots, S^k)$ is the maximum of all $|S^i|$ such that $i \in \{0, \ldots, k\}$. A layered causal graph $G_V(M)$ relative to X and ϕ is *width-bounded* for an integer $l \geq 0$ iff there exists a layering $(S^0, \ldots, S^k)$ of $G_V(M)$ relative to X and ϕ of a width of at most l.

Example 7.1 Fig. 12 provides a layering $\mathcal{L} = (S^0, S^1, S^2)$ of the causal graph $G_V(\widehat{M}^\phi_X)$ in Fig. 6 relative to $X' = X \cap \widehat{R}^\phi_X(M)$ and ϕ, where $M = (U, V, F)$ is a causal model and ϕ is an event such that the causal graph $G_V(M)$ and the sets X and $V(\phi)$ are as in Fig. 2. The width of this layering $\mathcal{L}$ is given by 3. $\Box$

The following result shows that layered causal graphs $G_V(M)$ relative to X and ϕ have a natural nontrivial decomposition relative to X and ϕ.

Proposition 7.1 *Let $M = (U, V, F)$ be a causal model, let $X \subseteq V$, and let ϕ be an event. Let $(S^0, \ldots, S^k)$ be a layering of $G_V(M)$ relative to X and ϕ. Then, $((S^0, S^0), \ldots, (S^k, S^k))$ is a decomposition of $G_V(M)$ relative to X and ϕ.*

We next give a condition under which a layered causal graph $G_V(M)$ has a unique layering. Two variables $A, B \in V$ are *connected* in $G_V(M)$ iff they are connected via a path in the undirected graph $(V, \{\{A, B\} \mid A \to B \text{ in } G_V(M)\})$. A causal graph $G_V(M)$ is *connected* relative to X and ϕ iff (i) $X \neq \emptyset$, (ii) there exists a variable in X that is connected to a variable in ϕ, and (iii) every variable in $V \setminus (X \cup V(\phi))$ is connected to a variable in $X \cup V(\phi)$. Notice that if $X = x$ is a weak cause of ϕ under $u \in D(U)$, then (i) and (ii) hold. Furthermore, in (iii), each variable $A \in V \setminus (X \cup V(\phi))$ which is not connected to a variable in $X \cup V(\phi)$ is irrelevant to "$X = x$ is a weak cause of ϕ under u".

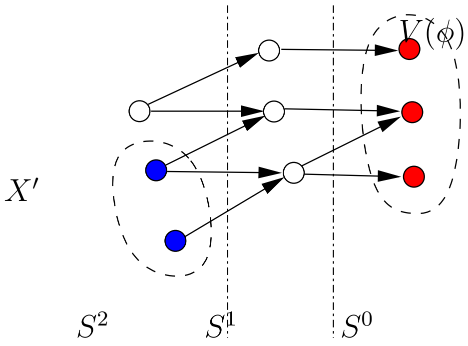

Figure 12: Layering (S^0, S^1, S^2) of $G_V(\widehat{M}^{\phi}_X)$ relative to $X' = X \cap \widehat{R}^{\phi}_X(M)$ and ϕ

The next result shows that when layered causal graphs $G_V(M)$ relative to X and ϕ are *connected* relative to X and ϕ, then the layering is unique. For this result, observe that every event ϕ contains some variables $A \in V$, which are all placed in the layer S_0. By conditions (i) and (ii), also X contains some variables, which are all placed in some layer S^k. By condition (iii), any other variable belongs to at most one layer S^i, and thus to exactly one layer S^i, since $G_V(M)$ is layered.

Proposition 7.2 *Let $M = (U, V, F)$ be a causal model, let $X \subseteq V$, and let ϕ be an event. If $G_V(M)$ is layered and connected relative to X and ϕ, then $G_V(M)$ has a unique layering relative to X and ϕ.*

We now provide an algorithm for deciding if a connected causal graph $G_V(M)$ relative to X and ϕ is layered and, if so, for computing its unique layering: Algorithm LAYERING (see Fig. 13) computes the unique layering $\mathcal{L} = (S^0, \ldots, S^k)$ of a connected causal graph $G_V(M)$ relative to X and ϕ, if it exists. The layering $\mathcal{L}$ is represented by the mapping $\lambda\colon V \to \{0, \ldots, k\}$, defined by $\lambda(A) = j$ for all $A \in S^j$ and all $j \in \{0, \ldots, k\}$. The following proposition states the correctness of LAYERING.

Proposition 7.3 *Let $M = (U, V, F)$ be a causal model, let $X \subseteq V$, and let ϕ be an event, where $G_V(M)$ is connected relative to X and ϕ. Then,* LAYERING *returns the unique layering of $G_V(M)$ relative to X and ϕ, if it exists, and Nil, otherwise.*

The next result shows that recognizing layered and width-bounded causal graphs $G_V(M)$ and computing their unique layerings can be done in linear time. Note that deciding whether $G_V(M)$ is connected w.r.t. X and ϕ is also possible in linear time.

Proposition 7.4 *Given a causal model $M = (U, V, F)$, $X \subseteq V$, and an event ϕ, where $G_V(M)$ is connected w.r.t. X and ϕ, deciding whether $G_V(M)$ is layered w.r.t. X and ϕ as well as width-bounded for some integer $l \geq 0$, and computing the unique layering of $G_V(M)$ w.r.t. X and ϕ can be done in $O(\|G_V(M)\| + \|\phi\|)$ time.*

By Proposition 7.1, all results of Sections 6.1–6.4 on causes, explanations, responsibility, and blame in decomposable causal graphs also apply to layered causal graphs as a special case. In particular, the relations R^i of Section 6.1 can be simplified to the following ones for layered causal graphs. The relation R^0 is

Algorithm LAYERING

Input: causal model $M = (U, V, F)$, $X \subseteq V$, and an event ϕ,
where $G_V(M) = (V, E)$ is connected relative to X and ϕ.

Output: unique layering $\mathcal{L} = (S^0, \ldots, S^k)$ of $G_V(M)$ relative to X and ϕ,
if it exists; *Nil*, otherwise.

Notation: $\mathcal{L}$ is represented by the mapping $\lambda\colon V \to \{0, \ldots, k\}$, which is
defined by $\lambda(A) = j$ for all $A \in S^j$ and all $j \in \{0, \ldots, k\}$.

for each $A \in V \setminus V(\phi)$ **do** $\lambda(A) := \bot$ (i.e., *undefined*);
for each $A \in V \cap V(\phi)$ **do** $\lambda(A) := 0$;
if $X \cap V(\phi) \neq \emptyset$ **then for each** $A \in X$ **do** $\lambda(A) := 0$;
while $E \neq \emptyset$ **do begin**
  select some $A \to B$ in E such that $\lambda(A) \neq \bot$ or $\lambda(B) \neq \bot$;
  if $B \in X \vee \lambda(A) = 0$ **then return** *Nil*;
  if $\lambda(A) \neq \bot \wedge \lambda(B) = \bot$ **then** $\lambda(B) := \lambda(A) - 1$
  else if $\lambda(A) = \bot \wedge \lambda(B) \neq \bot$ **then begin** $\lambda(A) := \lambda(B) + 1$;
    if $A \in X$ **then for each** $A' \in X \setminus \{A\}$ **do** $\lambda(A') := \lambda(A)$
  end
  else /* $\lambda(A), \lambda(b) \neq \bot$ */ **if** $\lambda(A) \neq \lambda(B) + 1$ **then return** *Nil*;
  $E := E \setminus \{A \to B\}$
end;
if $X \subseteq \{A \in V \mid \lambda(A) = k\}$, where $k = \max\{\lambda(A) \mid A \in V\}$ **then return** λ
else return *Nil*.

Figure 13: Algorithm LAYERING

given by:

$$
\begin{aligned}
R^0 = \{(\boldsymbol{p}, \boldsymbol{q}, F) \mid\ & F \subseteq S^0,\ \boldsymbol{p}, \boldsymbol{q} \subseteq D(F), \\
& \exists w \in D(S^0 \setminus F)\ \forall p, q \in D(F): \\
& p \in \boldsymbol{p} \ \text{ iff } \ \neg\phi_{pw}(u), \\
& q \in \boldsymbol{q} \ \text{ iff } \ \phi_{[q\langle \hat{Z}(u)]w}(u) \ \text{ for all } \hat{Z} \subseteq F \setminus S^k\}\,,
\end{aligned}
$$

For each $i \in \{1, \ldots, k\}$, the relation R^i is given by:

$$
\begin{aligned}
R^i = \{(\boldsymbol{p}, \boldsymbol{q}, F) \mid\ & F \subseteq S^i,\ \boldsymbol{p}, \boldsymbol{q} \subseteq D(F), \\
& \exists w \in D(S^i \setminus F)\ \exists(\boldsymbol{p'}, \boldsymbol{q'}, F') \in R^{i-1}\ \forall p, q \in D(F): \\
& p \in \boldsymbol{p} \ \text{ iff } \ F'_{pw}(u) \in \boldsymbol{p'}, \\
& q \in \boldsymbol{q} \ \text{ iff } \ F'_{[q\langle \hat{Z}(u)]w}(u) \in \boldsymbol{q'} \text{ for all } \hat{Z} \subseteq F \setminus S^k\}\,.
\end{aligned}
$$

The following theorem is immediate by Theorem 6.1 and Proposition 7.1.

Theorem 7.5 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$, let $x \in D(X)$ and $u \in D(U)$, and let ϕ be an event. Let $(S^0, \ldots, S^k)$ be a layering of $G_V(M)$ relative to X and ϕ, and let R^k be defined* *as above. Then, $X = x$ is a weak cause of ϕ under u in M iff (α) $X(u) = x$ and $\phi(u)$ in M, and (β) some $(\boldsymbol{p}, \boldsymbol{q}, X) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$.*

The next theorem shows that deciding whether $X = x$ is a weak respectively actual cause of ϕ under u in domain-bounded M is tractable, when $G_V(M)$ is layered and width-bounded. This is immediate by Theorem 6.2 and Proposition 7.1.

Theorem 7.6 *Given a domain-bounded causal model $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, where $G_V(M)$ is layered (relative to X and ϕ) and width-bounded for a constant $l \geq 0$, deciding whether $X = x$ is a weak (resp., an actual) cause of ϕ under u in M is possible in polynomial time.*

Similarly, by Proposition 7.1, all the tractability results of Theorems 6.3–6.9 and Theorems 6.11 and 6.12 also hold for width-bounded layered causal graphs.

8 Refinements and Model Application

In this section, we show that with some slight technical adaptations, all the above techniques and results carry over to a recent refinement of the notion of weak cause and to an extension of causal models due to Halpern and Pearl [17]. This shows that the results are robust at their core. Furthermore, we describe an application of our results for dealing with structure-based causes and explanations in first-order reasoning about actions.

8.1 Refined Weak Causes

We first consider the refined notion of weak cause that has been recently introduced by Halpern and Pearl in [17]. 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 *(refined) weak cause* of ϕ under $u \in D(U)$ in M iff **AC1** and the following condition **AC2′** hold:

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 $W' \subseteq W$, $\hat{Z} \subseteq V \setminus (X \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$.

Nearly all the results of this paper carry over to this refined notion of weak cause. The following theorem shows that this applies directly to Theorems 4.1 and 4.2.

Theorem 8.1 *Theorems 4.1 and 4.2 hold also for the refined notion of weak cause.*

For the results of Sections 4.2 and 4.3 to carry over to the refined notion of weak cause, we slightly adapt the definitions there as follows. The set of *relevant* (resp., *strongly relevant*) endogenous variables of $M = (U, V, F)$ w.r.t. $X = x$ and ϕ, denoted $R^{\phi}_{X=x}(M)$ (resp., $\widehat{R}^{\phi}_{X=x}(M)$), is redefined as the set of all $A \in V$ such that either **R1** (resp., **S1**), or **R2** (resp., **S2**), or the following condition **R3** (resp., **S3**) holds:

R3. A satisfies neither **R1** nor **R2**, and A is an ancestor in $G(M)$ of a variable $B \in V$ that satisfies **R2**.

S3. A satisfies neither **S1** nor **S2**, and A is an ancestor in $G(M)$ of a variable $B \in V \setminus X$ that satisfies **S2**.

Notice that nodes behind parents of nodes that are on directed paths from a variable in X to a variable in ϕ cannot be simply pruned, since by the refined condition **AC2′** subtle interactions between the variables in $W \cup \hat{Z}$ are possible. The *reduced* (resp., *strongly reduced*) *causal model* of $M=(U,V,F)$, where $F = \{F_A \mid A \in V\}$, w.r.t. $X=x$ and ϕ, denoted $M^{\phi}_{X=x}$ (resp., $\widehat{M}^{\phi}_{X=x}$), is redefined as the causal model $M'=(U,V',F')$, where $V'=R^{\phi}_{X=x}(M)$ (resp., $V'=\widehat{R}^{\phi}_{X=x}(M)$) and $F'=\{F'_A \mid A \in V'\}$ with $F'_A=F_A$ for all $A \in V'$ (resp., $F'_A=F^{\star}_A$ (defined as in Section 4.3) for all $A \in V' \cap X$ and $F'_A=F_A$ for all $A \in V' \setminus X$).

It is then not difficult to verify that all the results of Sections 4.2 and 4.3, except for Theorem 4.12, also hold for the refined notion of weak cause, using the above reduced and strongly reduced causal models. In particular, the following theorem shows that Theorems 4.7 and 4.10 carry over to the refined notion of weak cause.

Theorem 8.2 *Let $M=(U,V,F)$ be a causal model. Let $X' \subseteq X \subseteq V$ (resp., $X' = X \subseteq V$), $x' \in D(X')$, $x \in D(X)$, and $u \in D(U)$, and let ϕ be an event. Let $X'' = X' \cap R^{\phi}_{X}(M)$ (resp., $X''=X' \cap \widehat{R}^{\phi}_{X}(M)$) and $x''=x'|X''$. Then, $X'=x'$ is a (refined) weak cause of ϕ under u in M iff (i) $(X' \setminus X'')(u)=x'|(X' \setminus X'')$ in M, and (ii) $X''=x''$ is a (refined) weak cause of ϕ under u in M^{ϕ}_{X} (resp., $\widehat{M}^{\phi}_{X}$).*

For the results of Sections 6.1 to 6.3 to carry over to the refined notion of weak cause, we slightly adapt the relations R^i, $i \in \{0,\dots,k\}$, in Section 6.1 by replacing "$\phi_{[q\langle\hat{Z}(u)\rangle]w}(u)$" and "$F'_{[q\langle\hat{Z}(u)\rangle]w}(u) \in \boldsymbol{q'}$" by "$\phi_{[q\langle\hat{Z}(u)\rangle]w'}(u)$ for all $W' \subseteq W$ and $w'=w|W'$" and "$F'_{[q\langle\hat{Z}(u)\rangle]w'}(u) \in \boldsymbol{q'}$ for all $W' \subseteq W$ and $w'=w|W'$", respectively.

Using these new R^i's, all the results of Sections 6.1 to 6.3 (and thus all the results of Sections 5 and 7) hold also for the refined notion of weak cause. In particular, the following theorem is an extension of Theorem 6.1 to the refined notion of weak cause. Note that the results of Section 6.4 can be similarly extended.

Theorem 8.3 *Let $M=(U,V,F)$ be a causal model. Let $X \subseteq V$, let $x \in D(X)$ and $u \in D(U)$, and let ϕ be an event. Let $((T^0,S^0),\dots,(T^k,S^k))$ be a decomposition of $G_V(M)$ relative to X and ϕ, and let R^k be defined as above. Then, $X=x$ is a (refined) weak cause of ϕ under u in M iff (α) $X(u)=x$ and $\phi(u)$ in M, and (β) some $(\boldsymbol{p},\boldsymbol{q},X) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$.*

8.2 Refined Weak Causes in Extended Causal Models

We next consider the recent generalization of causal models to extended causal models [17]. An *extended causal model* $M=(U,V,F,E)$ consists of a standard causal model (U,V,F) as in Section 2.1 and a set $E \subseteq D(V)$ of *allowable settings* for V. For any $Y \subseteq V$, a value $y \in D(Y)$ is an *allowable setting* for Y iff $y=v|Y$ for some $v \in E$. Informally, y can be extended to an allowable setting for V. In the sequel, we assume E is represented in such a way that deciding whether a given $y \in D(Y)$, $Y \subseteq V$, is an allowable setting for Y is possible in polynomial time.

The notion of *(refined) weak cause* in extended causal models $M=(U,V,F,E)$ is then defined by slightly modifying the conditions **AC2′(a)** and **AC2′(b)** in the definition of (refined) weak causality to restrict to allowable settings.

To extend the results in Section 4.1 to the refined notion of weak cause in extended causal models, we introduce a natural closure property as follows. We say $M=(U,V,F,E)$ is *closed* (resp., *closed relative to* $X \subseteq V$) iff $y \cup (V \setminus Y)_y(u)$ is an allowable setting for V for all allowable settings y for Y, all $Y \subseteq V$ (resp., all $Y \subseteq V$ with $X \subseteq Y$), and all $u \in D(U)$. Informally, if y is an allowable setting for Y, then extending y

by the values of all other endogenous variables in M_y under any $u \in D(U)$ is an allowable setting for V. Notice that M is closed relative to all $X \subseteq V$, if M is closed. The following result says that Theorems 4.1 and 4.2 carry over to the refined notion of weak cause in closed extended causal models.

Theorem 8.4 *Theorems 4.1 and 4.2 hold also for the refined notion of weak cause in extended causal models* $M = (U, V, F, E)$ *that are closed relative to* X'.

For the results of Sections 4.2 and 4.3, we generalize the notions of a reduction and a strong reduction to extended causal models as follows. The *reduced* (resp., *strongly reduced*) *extended causal model* of $M = (U, V, F, E)$ w.r.t. $X = x$ and ϕ, denoted $M^{\phi}_{X=x}$ (resp., $\widehat{M}^{\phi}_{X=x}$), is defined as the extended causal model $M' = (U, V', F', E')$, where (U, V', F') is the reduced (resp., strongly reduced) causal model of (U, V, F) w.r.t. $X = x$ and ϕ, and $E' = \{v|V' \mid v \in E\}$. Notice here that, since E' is the restriction of E to V', any procedure for deciding allowability relative to E is immediately a procedure for deciding allowability relative to E'. The following result says that reductions and strong reductions keep the closure property.

Theorem 8.5 *Let* $M = (U, V, F, E)$ *be an extended causal model. Let* $X \subseteq V$ *and* $x \in D(X)$*, let* $X' = X \cap \widehat{R}^{\phi}_X(M)$*, and let* ϕ *be an event. Then: (a) If* M *is closed, then also* M^{ϕ}_X *is closed. (b) If* M *is closed relative to* X'*, then also* $\widehat{M}^{\phi}_X$ *is closed relative to* X'.

Using these notations, all the results of Sections 4.2 and 4.3, except for Theorem 4.12, hold also for the refined notion of weak cause in closed extended causal models. In particular, the following theorem generalizes Theorems 4.7 and 4.10.

Theorem 8.6 *Let* $M = (U, V, F, E)$ *be an extended causal model. Let* $X' \subseteq X \subseteq V$ *(resp.,* $X' = X \subseteq V$*), let* $x' \in D(X')$*,* $x \in D(X)$*, and* $u \in D(U)$*, and let* ϕ *be an event. Let* $X'' = X' \cap R^{\phi}_X(M)$ *(resp.,* $X'' = X' \cap \widehat{R}^{\phi}_X(M)$*) and* $x'' = x'|X''$*. Suppose that* M *is closed relative to* X''*. Then,* $X' = x'$ *is a (refined) weak cause of* ϕ *under* u *in* M *iff (i)* $(X' \backslash X'')(u) = x'|(X' \backslash X'')$ *in* M*, and (ii)* $X'' = x''$ *is a (refined) weak cause of* ϕ *under* u *in* M^{ϕ}_X *(resp.,* $\widehat{M}^{\phi}_X$*).*

For the results of Sections 6.1 to 6.3, we generalize the notion of a decomposition of $G_V(M)$ in Section 6.1 and the relations R^i, $i \in \{0, \dots, k\}$, in Section 8.1 to extended causal models as follows. A *decomposition* of $G_V(M)$ relative to $X = x$ (or simply X) and ϕ is a tuple $((T^0, S^0), \dots, (T^k, S^k))$, $k \geq 0$, of pairs (T^i, S^i) such that **D1–D6** in Section 6.1 and the following condition **D7** hold:

D7. Every y^i with $i \in \{0, \dots, k\}$ is an allowable setting of $Y^i \subseteq T^i$
iff $y^0 \cup \dots \cup y^k$ is an allowable setting of $Y^0 \cup \dots \cup Y^k \subseteq V$.

We then finally adapt the relations R^i, $i \in \{0, \dots, k\}$, in Section 8.1 by replacing "$\neg\phi_{pw}(u)$" and "${F'}_{pw}(u) \in \boldsymbol{p'}$" with "$\neg\phi_{pw}(u)$ and $pw|(X \cup W)$ is allowable" and "${F'}_{pw}(u) \in \boldsymbol{p'}$ and $pw|(X \cup W)$ is allowable", respectively.

Using the above notations, all the results of Sections 6.1 to 6.3 (and thus all the results of Sections 5 and 7) hold also for the refined notion of weak cause in closed extended causal models. In particular, the following theorem is a generalization of Theorem 6.1 to the refined notion of weak cause in closed extended causal models. Note that the results of Section 6.4 can be similarly generalized.

Theorem 8.7 *Let* $M = (U, V, F, E)$ *be an extended causal model. Let* $X \subseteq V$*, let* $x \in D(X)$ *and* $u \in D(U)$*, and let* ϕ *be an event. Let* $((T^0, S^0), \dots, (T^k, S^k))$ *be a decomposition of* $G_V(M)$ *relative to* X *and* ϕ*, and* *let R^k be defined as above. Suppose that M is closed relative to X. Then, $X = x$ is a (refined) weak cause of ϕ under u in M iff (α) $X(u) = x$ and $\phi(u)$ in M, and (β) some $(\boldsymbol{p}, \boldsymbol{q}, X) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$.*

8.3 Causes and Explanations in First-Order Reasoning about Actions

The work [9] presents a combination of the structural-model approach with first-order reasoning about actions in Poole's independent choice logic (ICL) [32, 33]. It shows how the ICL can be extended by structure-based causes and explanations, and thus how structure-based concepts can be made available in first-order reasoning about actions. From another perspective, it also shows how first-order modeling capabilities and explicit actions can be added to the structural-model approach.

From a technical point of view, this combination is based on a mapping of first-order theories in the ICL to binary causal models via some grounding step. The generated causal models have a subset of the Herbrand base as a set of endogenous variables, and thus they generally have a quite large number of variables. But they also have a natural layering through the time line, and thus they often have the structure of layered causal graphs as described in Section 7.

Roughly, ICL-theories are defined as follows. A *choice space* C is a set of pairwise disjoint and nonempty subsets of the Herbrand base, called the *alternatives* of C. The elements of the alternatives of C are called the *atomic choices* of C. A *total choice* of C is a set of atomic choices B such that $|B \cap A| = 1$ for all alternatives A of C. An *independent choice logic theory* (or *ICL-theory*) $T = (C, L)$ consists of a choice space C and an acyclic logic program L such that no atomic choice in C coincides with the head of any clause in the grounding of L. Semantically, every total choice of C along with the acyclic logic program L produces a first-order model [9]. Hence, $T = (C, L)$ encodes the set of all such models. Every total choice and thus every first-order model is often also associated with a probability value.

It is not difficult to see that there is a natural relationship between binary structure-based causal models $M = (U, V, F)$ and ICL-theories $T = (C, L)$: (i) The exogenous variables in U along with their domains correspond to the alternatives of C along with their atomic choices, (ii) the endogenous variables in V along with their binary domains correspond to the ground atoms of T that do not act as atomic choices, along with their binary truth values, (iii) the functions in F correspond to collections of clauses with the same head in the grounding of L, and (iv) a probability function on the contexts in $D(U)$ corresponds to a probability function on the atomic choices of C. This natural relationship nicely supports the definition of structure-based causes and explanations in the ICL. The following example illustrates ICL-theories and structure-based causes in ICL-theories.

Example 8.1 *(mobile robot)* Consider a mobile robot, which can navigate in an environment and pick up objects. We assume the constants r_1 (robot), o_1 and o_2 (two objects), p_1 and p_2 (two positions), and $0, 1, \ldots, h$ (time points within a *horizon* $h \geq 0$). The domain is described by the fluents $carrying(O, T)$ and $at(X, Pos, T)$, where $O \in \{o_1, o_2\}$, $T \in \{0, 1, \ldots, h\}$, $X \in \{r_1, o_1, o_2\}$, and $Pos \in \{p_1, p_2\}$, which encode that the robot r_1 is carrying the object O at time T (where we assume that at any time T the robot can hold at most one object), and that the robot or object X is at the position Pos at time T, respectively. The robot is endowed with the actions $moveTo(Pos)$, $pickUp(O)$, and $putDown(O)$, where $Pos \in \{p_1, p_2\}$ and $O \in \{o_1, o_2\}$, which represent the actions "move to the position P," "pick up the object O," and "put down the object O," respectively. The action $pickUp(O)$ is stochastic: It is not reliable, and thus can fail. Furthermore, we have the predicates $do(A, T)$, which represents the execution of an action A at time T, and $fa(A, T)$ (resp., $su(A, T)$), which represents the failure (resp., success) of an action A executed at time T.

An ICL-theory (C, L) is then given as follows. The choice space C encodes that picking up an object $o_i \in \{o_1, o_2\}$ at time $t \in \{0, 1, \ldots, h\}$ may fail ($fa(pickUp(o_i), t)$) or succeed ($su(pickUp(o_i), t)$):

$$C \;=\; \{\{fa(pickUp(o_i), t),\; su(pickUp(o_i), t)\} \mid i \in \{1, 2\},\; t \in \{0, 1, \ldots, h\}\}\,.$$

The acyclic logic program L consists of the clauses below, which encode the following knowledge:

- The robot is carrying the object O at time $T{+}1$, if either (i) the robot and the object O were both at Pos at time T, and the robot was not carrying any object and successfully picking up the object O at time T, or (ii) the robot was carrying the object O and not putting it down at time T.

  (1) $carrying(O, T{+}1) \Leftarrow (\neg carrying(o_1, T) \wedge \neg carrying(o_2, T) \wedge at(r_1, Pos, T) \wedge at(O, Pos, T) \wedge do(pickUp(O), T) \wedge su(pickUp(O), T)) \vee (carrying(O, T) \wedge \neg do(putDown(O), T))$.

- The robot is at Pos at time $T{+}1$, if either (i) it moved to Pos at time T, or (ii) it was at Pos and did not move away at time T.

  (2) $at(r_1, Pos, T{+}1) \Leftarrow do(moveTo(Pos), T) \vee (at(r_1, Pos, T) \wedge \neg do(moveTo(Pos'), T) \wedge Pos \neq Pos')$.

- The object O is at Pos at time $T{+}1$, if either (i) the object was at Pos and not carried by the robot at time T, or (ii) the robot was carrying the object O and moved to Pos at time T, or (iii) the object was at Pos and carried by the robot, who did not move away at time T.

  (3) $at(O, Pos, T{+}1) \Leftarrow (\neg carrying(O, T) \wedge at(O, Pos, T)) \vee (carrying(O, T) \wedge do(moveTo(Pos), T)) \vee (carrying(O, T) \wedge at(O, Pos, T) \wedge \neg do(moveTo(Pos'), T) \wedge Pos \neq Pos')$.

- The object o_1 is at the position p_2 at time 0.

  (4) $at(o_1, p_2, 0) \Leftarrow \top$.

- The robot is at the position p_2 at time 0.

  (5) $at(r_1, p_2, 0) \Leftarrow \top$.

Consider the horizon $h = 3$ and suppose that picking up an object succeeds at every time $t \in \{0, 1, 2, 3\}$, which is encoded by the total choice

$$B = \{su(pickUp(o_i), t) \mid i \in \{1, 2\}, t \in \{0, 1, 2, 3\}\}\,.$$

Suppose that the robot executes a pick up of o_1 at time 0, a move to p_1 at time 1, and a pick up of o_2 at time 2, which is represented by the additional facts

$$E = \{do(pickUp(o_1), 0),\; do(moveTo(p_1), 1),\; do(pickUp(o_2), 2)\}\,.$$

The structural-model approach now allows to give a semantics to causal statements in the ICL such as e.g. “the object o_1 being at position p_2 at time 0 is an actual cause of the robot not carrying the object o_2 at time 3 under the above B in $T \cup E$”. Intuitively, the robot and the object o_1 are both at position p_2 at time 0. Hence, picking up o_1 succeeds at time 0, the robot moves with o_1 to position p_1 at time 1, there its picking up o_2 fails at time 2, and this is why the robot is not carrying o_2 at time 3. However, if o_1 was not in position p_2 at time 0, and o_2 was always at position p_1, then the robot would hold no object at time 2, and its picking up o_2 at time 2 would succeed, and thus the robot would then be carrying o_2 at time 3.

Notice that the grounding step produces a causal model that has, even in this simple example, more than 90 variables (for a horizon $h \geq 0$, we have $24 \cdot (h+1)$ variables), which largely increases if we have more than only two positions and two objects different from the robot. Furthermore, the causal graph of this model is naturally layered through the time line, such that the results of Section 7 can be fruitfully applied to it. □

9 Conclusion

Defining causality between events is an issue which beyond the philosophical literature has also been considered in AI for a long time. Because of its key role for hypothesis and explanation forming, it is an important problem for which a number of different approaches have been proposed. In the approach by Halpern and Pearl [15, 16], causality is modeled using structural equations, distinguishing between weak and actual causes of events which are modeled by Boolean combinations of atomic value statements. Based on weak causes, a notion of causal explanation and probabilistic variants thereof have been defined in [18], while a refinement of actual causality in terms of responsibility and blame has been recently given in [3]. As has been argued and demonstrated, the structural-model approach by Halpern and Pearl deals well with difficulties of other approaches, including recent ones in the literature (see [16, 18]).

In order to bring the theoretical approach by Halpern and Pearl to practice, an understanding of the computational properties and (efficient) algorithms are required. In this direction, the computational complexity of major decision and computation problems for the approach has been studied in [3, 6, 7], and algorithms for computing causes proposed in [21]. Since arbitrary Boolean functions are used to model structural equations, determining causes and explanations is unsurprisingly intractable in general. Hence, the important issue of tractable cases arose, as well as how unnecessary complexity in computations can be avoided.

Investigating these issues, we have first explored, extending work by Hopkins [21], how to focus the computation of (potential) weak causes and causal models, by efficient removal of irrelevant variables. We have then presented a new characterization of weak cause for a certain class of causal models in which the causal graph over the endogenous variables is benignly decomposable. Two natural and important subclasses of it are causal trees and layered causal graphs, which can be be efficiently recognized, namely in linear time. By combining the removal of irrelevant variables with this new characterization of weak cause, we have then obtained techniques for deciding and computing causes and explanations in the structural-model approach, which show that these problems are tractable under suitable conditions. To our knowledge, these are the first explicit tractability results for causes and explanations in the structural-model approach. Moreover, by slightly extending the new characterization of weak cause, we have obtained similar techniques for computing the degrees of responsibility and blame, and thus also new tractability results for structure-based responsibility and blame. Finally, we have shown that all the above techniques and results carry over to recent refinements of the notion of weak cause and causal models due to Halpern and Pearl [17], and we have also described an application of our results and techniques for dealing with structure-based causes and explanations in first-order reasoning about actions in Poole’s ICL.

We have thus identified tractable special cases for decision and optimization problems of relatively high complexity, which is to some extend remarkable. These tractability results are a nice computational property of causes, explanations, responsibility, and blame in the structural-model approach.

An interesting topic of further studies is to explore whether there are other important classes of causal graphs different from causal trees and layered causal graphs in which the tractable cases can be recognized efficiently (that is, in which width-bounded decompositions can be recognized and computed efficiently). Another direction is analyzing how the techniques and results of this paper can be further developed for reasoning about actions [9] and commonsense causal reasoning [24]. Finally, implementation and further optimization remains for future work.

A Appendix: Proofs for Section 4

Proof of Theorem 4.1. Let $X_0 \in X$ be such that in $G(M)$, there is no directed path from X_0 to any variable in ϕ. Let $X'' = X \setminus \{X_0\}$ and $x'' = x|X''$. Then, as shown in [7], $X = x$ 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. By iteratively applying this result to every $X_0 \in X \backslash X'$, we thus obtain that $X = x$ is a weak cause of ϕ under u iff (i) $(X \backslash X')(u) = x|(X \backslash X')$ and (ii) $X' = x'$ is a weak cause of ϕ under u. □

Proof of Theorem 4.2. Let $X_0 \in X$ be such that in $G(M)$, each directed path from X_0 to a variable in ϕ contains some $X_i \in X'' = X \backslash \{X_0\}$, and let $x'' = x|X''$. As in the proof of Theorem 4.1, it is sufficient to show that $X = x$ 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.

($\Rightarrow$) Assume that $X = x$ is a weak cause of ϕ under u. That is, **AC1** $X(u) = x$ and $\phi(u)$, and **AC2** some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, 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 $\hat{z} = \hat{Z}(u)$. In particular, (i) $X_0(u) = x(X_0)$, and also **AC1** $X''(u) = x''$ and $\phi(u)$. Furthermore, as every directed path in $G(M)$ from X_0 to a variable in ϕ contains some $X_i \in X''$, it follows that **AC2**(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 $W' = W \cup \{X_0\}$, $\overline{x}'' = \overline{x}|X''$, $w' = wx_0 \in D(W')$, and $x_0 = x(X_0)$. Hence, (ii) $X'' = x''$ is a weak cause of ϕ under u.

($\Leftarrow$) Assume that (i) $X_0(u) = x(X_0)$ and (ii) $X'' = x''$ is a weak cause of ϕ under u. That is, **AC1** $X''(u) = x''$ and $\phi(u)$ hold, and **AC2** there exist some $W \subseteq V \backslash X''$, $\overline{x}'' \in D(X'')$, $w \in D(W)$ such that (a) $\neg\phi_{\overline{x}''w}(u)$, and (b) $\phi_{x''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \backslash (X'' \cup W)$ and $\hat{z} = \hat{Z}(u)$. By (i), it holds that **AC1** $X(u) = x$ and $\phi(u)$. Furthermore, since every directed path in $G(M)$ from X_0 to a variable in ϕ contains some $X_i \in X''$, it follows that **AC2**(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 $W' = W \setminus \{X_0\}$, $w' = w|W' \in D(W')$, and $x_0 = x(X_0)$. It thus follows that $X = x$ is a weak cause of ϕ under u. □

Proof of Proposition 4.3. (a) We first compute the set A_ϕ of all variables in ϕ and their ancestors in $G_V(M)$, and then the set $X' = A_\phi \cap X$. Using standard methods and data structures, the first step can be done in time $O(\|\phi\| + |E|)$ where $G_V(M) = (V, E)$, and the second step in time $O(|V|)$. In summary, X' is computable in time $O(\|G_V(M)\| + \|\phi\|)$, and hence in time $O(\|M\| + \|\phi\|)$.

(b) We first compute the directed graph G' obtained from $G_V(M) = (V, E)$ by removing every arrow $X_k \rightarrow X_l$ with $X_l \in X$. We then compute the set A'_ϕ of all ancestors in G' of variables in ϕ. We finally compute $X' = A'_\phi \cap X$. Using standard methods and data structures, the first step can be done in time $O(|V| + |E|)$, the second step in time $O(|V| + |E| + \|\phi\|)$, and the third step in time $O(|V|)$. In summary, X' is computable in time $O(|V| + |E| + \|\phi\|)$, hence in time $O(\|M\| + \|\phi\|)$. □

Proof of Proposition 4.4. Consider any $B \in X \cap R^{\phi}_X(M)$. That is, $B \in X$ and $B \in R^{\phi}_X(M)$. If B is included into $R^{\phi}_X(M)$ by **R1**, then there exists a directed path in $G(M)$ from B to a variable in ϕ. If B is included into $R^{\phi}_X(M)$ by **R2**, then either B occurs in ϕ or B is a parent of a variable that satisfies **R1**. Thus, there also exists a directed path in $G(M)$ from B to a variable in ϕ.

Conversely, suppose that $B \in X$ and that there exists a directed path in $G(M)$ from B to a variable in ϕ. If B occurs in ϕ, then $B \in R^{\phi}_X(M)$ by either **R1** or **R2**. Otherwise, there exists a child A of B in $G(M)$ such that A is on a directed path in $G(M)$ from $B \in X \backslash \{A\}$ to a variable in ϕ. Hence, $A \in R^{\phi}_X(M)$ by **R1**. It thus follows that $B \in R^{\phi}_X(M)$ by either **R1** or **R2**. □

Proof of Theorem 4.5. Let $M^{\phi}_X = (U, V', F')$. Let $X' = X \cap V'$ and $x' = x|X'$. We have to show that $X = x$ is a weak cause of ϕ under u in M iff (i) $(X \backslash X')(u) = x|(X \backslash X')$ in M, and (ii) $X' = x'$ is a weak cause of ϕ under u in M^{ϕ}_X. Let V'_1 (resp., V'_2) denote the set of all $A \in V'$ that satisfy **R1** (resp., **R2**).

Roughly, the main idea behind the proof is to move all the variables in $V'_2 \setminus X'$ into W in **AC2**. Then, setting the variables in X' and W in **AC2** makes the truth of ϕ under u independent from the values of the variables in $V \setminus V'$. Thus, the variables in $V \setminus V'$ can be simply ignored in M and added to M^{ϕ}_X, respectively.

Fact A. $V'_M(u) = V'_{M^{\phi}_X}(u)$ and $\phi_M(u) = \phi_{M^{\phi}_X}(u)$.

($\Rightarrow$) Assume that $X = x$ is a weak cause of ϕ under u in M. That is, **AC1** $X(u) = x$ and $\phi(u)$ in M, and **AC2** some $W \subseteq V \backslash X$, $\overline{x} \in D(X)$, $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}w}(u)$ and (b) $\phi_{xw\hat{z}}(u)$ in M for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$. This already shows that (i) $(X \backslash X')(u) = x|(X \backslash X')$ in M. We next show that also (ii) holds. From Fact A, it follows that **AC1** $X'(u) = x'$ and $\phi(u)$ in M^{ϕ}_X. Since for any value $\hat{x}'$ of X' it holds that $(V'_2 \backslash X')_{\hat{x}'w} = (V'_2 \backslash X')_w$ in M, it follows in particular that $(V'_2 \backslash X')_{\overline{x}'w}(u) = (V'_2 \backslash X')_w(u) = (V'_2 \backslash X')_{x'w}(u)$ in M, where $\overline{x}' = \overline{x}|X'$. It then follows that (a) $\neg\phi_{\overline{x}'w'}(u)$ and (b) $\phi_{x'w'\hat{z}}(u)$ in M for all $\hat{Z} \subseteq V \backslash (X' \cup W')$ and $\hat{z} = \hat{Z}(u)$, where $W' = (W \cap V') \cup (V'_2 \backslash X')$, and $w' = W'_w(u)$ in M. Hence, **AC2**(a) $\neg\phi_{\overline{x}'w'}(u)$ in M^{ϕ}_X and (b) $\phi_{x'w'\hat{z}}(u)$ in M^{ϕ}_X for all $\hat{Z} \subseteq V' \setminus (X' \cup W')$ and $\hat{z} = \hat{Z}(u)$. In summary, (ii) $X' = x'$ is a weak cause of ϕ under u in M^{ϕ}_X.

($\Leftarrow$) Assume that (i) $(X \backslash X')(u) = x|(X \backslash X')$ in M and (ii) $X' = x'$ is a weak cause of ϕ under u in M^{ϕ}_X. Thus, **AC1** $X'(u) = x'$ and $\phi(u)$ in M^{ϕ}_X, and **AC2** some $W \subseteq V' \backslash X'$, $\overline{x}' \in D(X)$, $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}'w}(u)$ and (b) $\phi_{x'w\hat{z}}(u)$ in M^{ϕ}_X for all $\hat{Z} \subseteq V' \setminus (X' \cup W)$ and $\hat{z} = \hat{Z}(u)$. By Fact A, we have **AC1** $X(u) = x$ and $\phi(u)$ in M. Since in M^{ϕ}_X, the variables in $V'_2 \backslash X'$ do not depend on any variable in X', it holds that $(V'_2 \backslash X')_{\overline{x}'w}(u) = (V'_2 \backslash X')_w(u) = (V'_2 \backslash X')_{x'w}(u)$ in M^{ϕ}_X. It then follows that (a) $\neg\phi_{\overline{x}'w'}(u)$ and (b) $\phi_{x'w'\hat{z}}(u)$ in M^{ϕ}_X for all $\hat{Z} \subseteq V' \setminus (X' \cup W')$ and $\hat{z} = \hat{Z}(u)$, where $W' = W \cup (V'_2 \backslash X')$, and w' is such that $w'|W = w$ and $w'|(V'_2 \backslash (X' \cup W)) = ((V'_2 \backslash (X' \cup W))_w(u)$ in M^{ϕ}_X. Since no variable from $V \backslash V'$ can influence any variable in ϕ if all variables in X and W' have a value assigned in M, it follows that **AC2**(a) $\neg\phi_{\overline{x}w'}(u)$ in M and (b) $\phi_{xw'\hat{z}}(u)$ in M for all $\hat{Z} \subseteq V \backslash (X \cup W')$ and $\hat{z} = \hat{Z}(u)$, where $\overline{x}|X' = \overline{x}'$ and $\overline{x}|(X \backslash X') = (X \backslash X')(u)$. In summary, $X = x$ is a weak cause of ϕ under u in M. □

Proof of Proposition 4.6. Let $M^{\phi}_X = (U, V', F')$, $M^{\phi}_{X'} = (U, V'', F'')$, and $(M^{\phi}_X)^{\phi}_{X'} = (U, V''', F''')$. We first show that $V'' = V'''$ and then that $F'' = F'''$, which proves that $M^{\phi}_{X'}$ coincides with $(M^{\phi}_X)^{\phi}_{X'}$. We note the following easy fact.

Fact B. **R1** and $R^{\phi}_X(M)$ are monotonic in X, i.e., if A satisfies **R1** for X, then also for each superset of X, and $X \subseteq X'$ implies $R^{\phi}_X(M) \subseteq R^{\phi}_{X'}(M)$.

Let V_1'' (resp., V_1''') denote the set of all variables included into V'' (resp., V''') by **R1**. We now first show that $V_1'' = V_1'''$. Consider any $A \in V_1'''$. Then, A is on a directed path in $G(M_X^\phi)$ from a variable in $X' \setminus \{A\}$ to a variable in ϕ. Since $G(M_X^\phi)$ is a subgraph of $G(M)$, it follows that A is also on a directed path in $G(M)$ from a variable in $X' \setminus \{A\}$ to a variable in ϕ, and thus $A \in V_1''$. Conversely, suppose that $A \in V_1''$. Then, A is on a directed path in $G(M)$ from a variable in $X' \setminus \{A\}$ to a variable in ϕ. Since $X' \subseteq X$, this path also exists in $G(M_X^\phi)$, and thus $A \in V_1'''$. This shows that $V_1'' = V_1'''$. Observe then that $V(\phi) \subseteq V''$ and $V(\phi) \subseteq V'''$. Consider finally any parent $A \in V$ of a variable $B \in V_1''$ in $G(M)$. By Fact B, $A \in V'$ and since $V_1'' = V_1'''$, A is also parent of $B \in V_1'''$ in $G(M_X^\phi)$. Conversely, if $A \in V$ is parent of $B \in V_1'''$ in $G(M_X^\phi)$, then $A \in V$ is also a parent of $B \in V_1''$ in $G(M)$. In summary, this shows that $V'' = V'''$.

We finally show that $F'' = F'''$. As shown above, $V_1'' = V_1'''$ and $V'' = V'''$. By Fact B, for each $A \in V_1''$ we have $F''_A = F_A$ and $F'''_A = F'_A = F_A$. For each $A \in V'' \setminus V_1''$, we have $F''_A = F^\star_A$ and $F'''_A = (F'_A)^\star = F^\star_A$. Hence, $F'' = F'''$. □

Proof of Theorem 4.7. Let $X^\star$ be the set of all variables in X that are not connected by a directed path in $G(M)$ to a variable in ϕ. By Proposition 4.4, $X'' = X' \setminus X^\star$. By Theorem 4.5, $X' = x'$ is a weak cause of ϕ under u in M iff (i) $(X' \setminus X'')(u) = x' | (X' \setminus X'')$ in M, and (ii) $X'' = x''$ is a weak cause of ϕ under u in $M^\phi_{X'}$. Moreover, again by Theorem 4.5 (invoked for X there equal to X'', which means $X' = X'' \cap R^\phi_{X''}(M^\phi_X) = X''$), $X'' = x''$ is a weak cause of ϕ under u in M^ϕ_X iff $X'' = x''$ is a weak cause of ϕ under u in $(M^\phi_X)^\phi_{X''} = (M^\phi_X)^\phi_{X'}$. By Proposition 4.6, $M^\phi_{X'} = (M^\phi_X)^\phi_{X'}$, which proves the result. □

Proof of Proposition 4.8. We first show that the directed graph $G_V(M^\phi_X)$ is computable in linear time. Its set of nodes $V' = R^\phi_X(M)$ is the set of all variables $A \in V$ that satisfy **R1** or **R2**. The set of all variables $A \in V$ that satisfy **R1** is given by $D_X \cap A_\phi$, where D_X denotes the set of all proper descendents of variables in X, and A_ϕ denotes the set of all variables in ϕ and of all ancestors of variables in ϕ. Thus, the part of V' satisfying **R1** can be computed in time $O(\|G(M)\| + \|\phi\|)$, since D_X is computable in time $O(\|G(M)\|) = O(|U| + |V| + |E|)$ where $G(M) = (U \cup V, E)$, and A_ϕ and $D_X \cap A_\phi$ are computable in time $O(\|G(M)\| + \|\phi\|)$ using standard methods and data structures. The set of all variables $A \in V$ that satisfy **R2** is given by $(V_\phi \cup PA(D_X \cap A_\phi)) \setminus (D_X \cap A_\phi)$. As already noted, $D_X \cap A_\phi$ can be computed in time $O(\|G(M)\| + \|\phi\|)$. Furthermore, V_ϕ and $PA(D_X \cap A_\phi)$ given $D_X \cap A_\phi$ can be computed in time $O(\|\phi\|)$ and $O(\|G(M)\|)$, respectively. Since all set operations are feasible in linear time using standard methods and data structures, it thus follows that the part of V' satisfying **R2** can be computed in time $O(\|G(M)\| + \|\phi\|)$. In summary, V' is computable in time $O(\|G(M)\| + \|\phi\|)$, hence in time $O(\|M\| + \|\phi\|)$. This already shows that $G_V(M^\phi_X)$ can be computed in time linear in the size of M and ϕ, since it is the restriction of $G(M)$ to V'.

We next show that $M^\phi_X = (U, V', F')$ can be computed in polynomial time. As argued above, V' and its partition into variables that satisfy **R1** and those that satisfy **R2** is computable in time $O(\|G(M)\| + \|\phi\|)$. We next show that a representation of every function $F^\star_A$, where A satisfies **R2**, is computable in time $O(\|M\|)$. Every $F^\star_A(U_A)$ is given as follows. The set of arguments U_A is the set of all ancestors $B \in U$ of A in $G(M)$. The function $F^\star_A$ itself can be represented by the restriction M_A of $M = (U, V, F)$ to V and all ancestors $B \in U$ of A in $G(M)$. Then, $F^\star_A(u_A)$ for $u_A \in D(U_A)$ is given by $A(u_A)$ in M_A. Observe that by Proposition 2.1, every $F^\star_A(u_A)$ is computable in polynomial time. Clearly, U_A and M_A can be computed in linear time. Thus, the set of all functions $F^\star_A$, where A satisfies **R2**, can be computed in $O(|V| \|M\|)$ time. In summary, $M^\phi_X = (U, V', F')$ can be computed in $O(|V| \|M\| + \|\phi\|)$ time. □

Proof of Proposition 4.9. Consider any $B \in X \cap \widehat{R}^\phi_X(M)$. That is, $B \in X$ and $B \in \widehat{R}^\phi_X(M)$. Then, B

is not included into $\widehat{R}^{\phi}_{X}(M)$ by **S1**, as otherwise $B \notin X$. Thus, B is included into $\widehat{R}^{\phi}_{X}(M)$ by **S2**. Hence, either B occurs in ϕ, or B is a parent of a variable that satisfies **S1**. Thus, there exists a directed path in $G(M)$ from B to a variable in ϕ that contains no $X_j \in X \backslash \{B\}$.

Conversely, suppose that $B \in X$ and that there exists a directed path in $G(M)$ from B to a variable in ϕ that contains no $X_j \in X \backslash \{B\}$. If B occurs in ϕ, then $B \in \widehat{R}^{\phi}_{X}(M)$ by **S2** (note that B does not satisfy **S1**). Otherwise, there exists a child A of B in $G(M)$ such that A is on a directed path P in $G(M)$ from a variable in $X \backslash \{A\}$ (=X) to a variable in ϕ, where P does not contain any variable from $X \backslash \{B\}$. It follows that $A \in \widehat{R}^{\phi}_{X}(M)$ by **S1**, and thus $B \in \widehat{R}^{\phi}_{X}(M)$ by **S2**. □

Proof of Theorem 4.10. The proof is similar to the one of Theorem 4.5, using now $\widehat{M}^{\phi}_{X}$ instead of M^{ϕ}_{X} and **S1** (resp., **S2**) instead of **R1** (resp., **R2**). □

Proof of Proposition 4.11. The proof is similar to the proof of Proposition 4.8. The main difference is that we now define D_X as the set of all proper descendents in G' of variables in X, and A_ϕ as the set of all variables in ϕ and of all ancestors in G' of variables in ϕ, where the directed graph G' is obtained from $G_V(M)$ by removing every arrow $X_k \rightarrow X_l$ with $X_l \in X$. The result then follows from the observation that the new D_X and A_ϕ can also be both computed in time $O(\|M\| + \|\phi\|)$, since G' is computable in time $O(\|G_V(M)\|)$, and D_X and A_ϕ are both computable in time $O(\|G'\| + \|\phi\|)$ and $O(\|G'\|) = O(\|M\|)$. □

Proof of Theorem 4.12. If X is a singleton, then **R1** (resp., **R2**) in the definition of $R^{\phi}_{X}(M)$ coincides with **S1** (resp., **S2**) in the definition of $\widehat{R}^{\phi}_{X}(M)$, since $X \backslash \{B\} = \emptyset$. This shows that $R^{\phi}_{X}(M) = \widehat{R}^{\phi}_{X}(M)$ and $M^{\phi}_{X} = \widehat{M}^{\phi}_{X}$. □

B Appendix: Proofs for Section 5

Proof of Proposition 5.1. Using standard methods and data structures, deciding whether there exists exactly one directed path in $G_V(M) = (V, E)$ from every variable $A \in V \setminus \{Y\}$ to Y can be done in $O(|V| + |E|)$ time. Moreover, deciding whether every $A \in V \backslash \{X\}$ has a bounded number of parents can also be done in $O(|V| + |E|)$ time. In summary, deciding whether M is a causal tree with respect to X and Y is feasible in $O(|V| + |E|) = O(\|M\|)$ time. By Proposition 4.8, the directed graph $G_V(M^{Y}_{X})$ can be computed in $O(\|M\|)$ time from M and X, Y. Thus, deciding if M^{Y}_{X} is a (bounded) causal tree can also be done in time $O(\|M\|)$. □

Proof of Theorem 5.2. Clearly, (α) coincides with **AC1**. Assume that (α) holds. We now show that (β) is equivalent to **AC2**:

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

(a) $Y_{\overline{x}w}(u) \neq y$,

(b) $Y_{xw\hat{Z}(u)}(u) = y$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$.

Clearly, we can assume that $P^i \notin W$ for all $i \in \{0, \ldots, k-1\}$, since otherwise $Y_{\overline{x}w}(u) = Y_{xw}(u)$. This shows that $W \subseteq W^1 \cup \cdots \cup W^k$. Observe then that we can enlarge every $w \in D(W)$ to some $w' \in D(W')$, where $W' = W^1 \cup \cdots \cup W^k$, by defining $w'|W = w$ and $w'|(W' \backslash W) = (W' \backslash W)(u)$. Hence, we can assume that $\hat{Z} \subseteq \{P^0, \ldots, P^{k-1}\}$ and thus also, by the path structure of the causal tree, that $\hat{Z} = \{P^i\}$ with $i \in \{1, \ldots, k-1\}$. Hence, it is sufficient to prove that (β) is equivalent to the following condition **AC2′**:

AC2′. Some values $\overline{x} \in D(X)$ and $\overline{w} \in D(W^1 \cup \cdots \cup W^k)$ exist such that:

(a) $Y_{\overline{x}\,\overline{w}}(u) \neq y$,

(b) $Y_{\hat{p}^j \overline{w}}(u) = y$ for all $j \in \{1, \ldots, k\}$.

We now show that $(\star)$ for every $i \in \{1, \ldots, k\}$, it holds that $\boldsymbol{p} \in R^i$ iff there exists some $\overline{w} \in D(W^1 \cup \cdots \cup W^i)$ such that:

(i) $p \in \boldsymbol{p}$ iff $Y_{p\overline{w}}(u) \neq y$, for all $p \in D(P^i)$,

(ii) $Y_{\hat{p}^j \overline{w}}(u) = y$ for all $j \in \{1, \ldots, i\}$.

This then shows that (β) is equivalent to **AC2′**: $(\Rightarrow)$ Assume some $\boldsymbol{p} \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \notin \boldsymbol{p}$. Then, some $\overline{w} \in D(W^1 \cup \cdots \cup W^k)$ and $p \in \boldsymbol{p}$ exist such that $Y_{p\overline{w}}(u) \neq y$ and $Y_{\hat{p}^j \overline{w}}(u) = y$ for all $j \in \{1, \ldots, k\}$. That is, **AC2′** holds. $(\Leftarrow)$ Conversely, suppose that **AC2′**(a) and (b) hold for some $\overline{x} \in D(X)$ and $\overline{w} \in D(W^1 \cup \cdots \cup W^k)$. Let $\boldsymbol{p} = \{p \in D(P^k) \mid Y_{p\overline{w}}(u) \neq y\}$. Then, $\boldsymbol{p} \in R^k$, $\boldsymbol{p} \neq \emptyset$, and $x \notin \boldsymbol{p}$. That is, (β) holds.

We prove $(\star)$ by induction on $i \in \{1, \ldots, k\}$:

Basis: Since $R^0 = \{D(Y) \setminus \{y\}\}$, it holds that $\boldsymbol{p} \in R^1$ iff some $w \in D(W^1)$ exists such that:

(i) $p \in \boldsymbol{p}$ iff $Y_{pw}(u) \in D(Y) \setminus \{y\}$ iff $Y_{pw}(u) \neq y$, for all $p \in D(P^1)$,

(ii) $P^0_{\hat{p}w}(u) \in \{y\}$, i.e., $Y_{\hat{p}^1 w}(u) = y$.

Induction: Observe that $\boldsymbol{p} \in R^i$ iff some $w \in D(W^i)$ and $\boldsymbol{p'} \in R^{i-1}$ exist such that:

(i′) $p \in \boldsymbol{p}$ iff $P^{i-1}_{pw}(u) \in \boldsymbol{p'}$, for all $p \in D(P^i)$,

(ii′) $P^{i-1}_{\hat{p}^i w}(u) \in D(P^{i-1}) \setminus \boldsymbol{p'}$.

By the induction hypothesis, $\boldsymbol{p'} \in R^{i-1}$ iff some $\overline{w}' \in D(W^1 \cup \cdots \cup W^{i-1})$ exists such that:

(i″) $p' \in \boldsymbol{p'}$ iff $Y_{p'\overline{w}'}(u) \neq y$, for all $p' \in D(P^{i-1})$,

(ii″) $Y_{\hat{p}^j \overline{w}'}(u) = y$ for all $j \in \{1, \ldots, i-1\}$.

Thus, $\boldsymbol{p} \in R^i$ iff some $w \in D(W^i)$ and $\overline{w}' \in D(W^1 \cup \cdots \cup W^{i-1})$ exist such that:

(i) $p \in \boldsymbol{p}$ iff $P^{i-1}_{pw}(u) \in \boldsymbol{p'}$ iff $Y_{pw\overline{w}'}(u) \neq y$, for all $p \in D(P^i)$, by (i′) and (i″),

(ii) $P^{i-1}_{\hat{p}^i w}(u) = p'$ and $Y_{p'\overline{w}'}(u) = y$, by (ii′) and (i″), as well as $Y_{\hat{p}^j w\overline{w}'}(u) = y$ for all $j \in \{1, \ldots, i-1\}$ by (ii″).

That is, $\boldsymbol{p} \in R^i$ iff some $\overline{w} \in D(W^1 \cup \cdots \cup W^i)$ exists such that:

(i) $p \in \boldsymbol{p}$ iff $Y_{p\overline{w}}(u) \neq y$, for all $p \in D(P^i)$,

(ii) $Y_{\hat{p}^j \overline{w}}(u) = y$ for all $j \in \{1, \ldots, i\}$ (note that $Y_{p'\overline{w}}(u) = Y_{\hat{p}^i w\overline{w}})$. $\Box$

Proof of Theorem 5.3. By Theorem 5.2, $X = x$ is a weak cause of $Y = y$ under u in M iff (α) $X(u) = x$ and $Y(u) = y$ in M, and (β) some $\boldsymbol{p} \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \notin \boldsymbol{p}$. It is thus sufficient to show that deciding whether (α) and (β) hold can be done in polynomial time. By Proposition 2.1, deciding whether (α) holds can be done in polynomial time. Next, we observe that $P^0, \ldots, P^k$ and $W^1, \ldots, W^k$ can be computed in time $O(\|M\|)$. By Proposition 2.1, every $\hat{p}^i$ with $i \in \{1, \ldots, k\}$ can be computed in polynomial time. We then iteratively compute every R^i with $i \in \{0, \ldots, k\}$. Clearly, R^0 can be computed in constant time, since V is domain-bounded. Observe then that the cardinality of each $D(W^i)$ is bounded by a constant, since V is domain-bounded and $G_V(M)$ is bounded. Furthermore, the size of each R^{i-1} and the cardinality of each $D(P^i)$ are both bounded by a constant, since V is domain-bounded. By Proposition 2.1, the values $P^{i-1}_{\hat{p}^i w}(u)$ and $P^{i-1}_{pw}(u)$ can be computed in polynomial time. Hence, every R^i can be computed by a constant number of polynomial computations, and thus in polynomial time. Hence, R^k can be computed in polynomial time. Given R^k, deciding whether (β) holds can be done in constant time. In summary, computing R^k and deciding whether (β) holds, and thus deciding whether (α) and (β) hold, can be done in polynomial time.

By Theorem 2.3, $X = x$ is an actual cause of $Y = y$ under u in M iff X is a singleton and $X = x$ is a weak cause of $Y = y$ under u in M. Thus, deciding whether $X = x$ is an actual cause of $Y = y$ under u in M can also be done in polynomial time. $\Box$

Proof of Theorem 5.4. By Theorem 4.5, $X = x$ is a weak cause of $Y = y$ under u in M iff $X = x$ is a weak cause of $Y = y$ under u in M^Y_X. By Proposition 4.8, M^Y_X is computable in polynomial time. By Theorem 5.3, given M^Y_X, deciding whether $X = x$ is a weak cause of $Y = y$ under u in M^Y_X can be done in polynomial time. In summary, deciding whether $X = x$ is a weak cause of $Y = y$ under u in M^Y_X and thus in M is possible in polynomial time. $\Box$

Proof of Theorem 5.5. Recall that $X = x$ is an explanation of $Y = y$ relative to $\mathcal{C}$ iff **EX1** $Y(u) = y$ for every $u \in \mathcal{C}$, **EX2** $X = x$ is a weak cause of $Y = y$ under every $u \in \mathcal{C}$ such that $X(u) = x$, **EX3** X is minimal, and **EX4** $X(u) = x$ and $X(u') \neq x$ for some $u, u' \in \mathcal{C}$. By Proposition 2.1, checking whether **EX1** and **EX4** hold can be done in polynomial time. Clearly, **EX3** always holds, since X is a singleton. By Theorem 5.4, deciding whether $X = x$ is a weak cause of $Y = y$ under some $u \in \mathcal{C}$ in M such that $X(u) = x$ can be done in polynomial time. Thus, by Proposition 2.1, deciding whether **EX2** holds can be done in polynomial time. In summary, deciding whether **EX1**–**EX4** hold can be done in polynomial time. $\Box$

Proof of Theorem 5.6. We first compute the set $\mathcal{C}^\star$ of all $u \in \mathcal{C}$ such that either (i) $X(u) \neq x$ in M, or (ii) $X(u) = x$ and $X = x$ is a weak cause of $Y = y$ under u in M. By Proposition 2.1 and Theorem 5.4, this can be done in polynomial time. If $X = x$ is a partial explanation of $Y = y$ relative to $(\mathcal{C}, P)$ in M, then $\mathcal{C}^{Y=y}_{X=x}$ is defined, and $\mathcal{C}^{Y=y}_{X=x} = \mathcal{C}^\star$ by Proposition 2.4. Given $\mathcal{C}^{Y=y}_{X=x}$, the explanatory power $P(\mathcal{C}^{Y=y}_{X=x} \mid X = x)$ is computable in polynomial time by Proposition 2.1, if we assume as usual that P is computable in polynomial time. In summary, this shows (c).

To check partial (resp., α-partial) explanations in (a) (resp., (b)), we compute $\mathcal{C}^\star$ as above. We then check whether $\mathcal{C}^{Y=y}_{X=x}$ is defined. That is, by Proposition 2.4, we check that $X = x$ is an explanation of $Y = y$ relative to $\mathcal{C}^\star$ in M, which is possible in polynomial time by Theorem 5.5. Then, $\mathcal{C}^{Y=y}_{X=x} = \mathcal{C}^\star$ by Proposition 2.4. We finally compute $P(\mathcal{C}^{Y=y}_{X=x} \mid X = x)$ as above and check that it is positive (resp., at least α), which can clearly be done in polynomial time. In summary, this proves (a) (resp., (b)). $\Box$

C Appendix: Proofs for Section 6

Proof of Theorem 6.1. Obviously, (α) coincides with **AC1**. We now prove that (β) is equivalent to **AC2**:

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

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

(b) $\phi_{xw\hat{Z}(u)}(u)$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$.

By **D6**, the variables in S^k have no parents in $G_V(M)$. Hence, every variable in S^k only depends on the variables in U, and thus we can move any $A \in S^k \setminus (W \cup X)$ into W by setting $w(A) = A(u)$. We can thus assume that $X = S^k \setminus W$ holds. Since (i) $W \subseteq V$ and $X = S^k \setminus W$ implies $W \subseteq V \setminus X$, and (ii) $X = S^k \setminus W$ implies $S^k \cup W = X \cup W$, it is thus sufficient to show that (β) is equivalent to **AC2***:

AC2***.** Some $W \subseteq V$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that $X = S^k \setminus W$ and

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

(b) $\phi_{xw\hat{Z}(u)}(u)$ for all $\hat{Z} \subseteq V \setminus (S^k \cup W)$.

We now prove that $(\star)$ for all $i \in \{0, \ldots, k\}$, it holds that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^i$ iff some $\overline{W} \subseteq T^0 \cup \cdots \cup T^i$ and $\overline{w} \in D(\overline{W})$ exist such that $F = S^i \setminus \overline{W}$ and

(i) for every $p, q \in D(F)$:

(i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{p\overline{w}}(u)$,

(i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle\hat{Z}(u)]\overline{w}}(u)$ for all $\hat{Z} \subseteq (T^0 \cup \cdots \cup T^i) \setminus (S^k \cup \overline{W})$.

In particular, this then implies that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^k$ iff some $\overline{W} \subseteq T^0 \cup \cdots \cup T^k = V$ and $\overline{w} \in D(\overline{W})$ exist such that $F = S^k \setminus \overline{W}$ and

(i) for every $p, q \in D(F)$:

(i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{p\overline{w}}(u)$,

(i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle\hat{Z}(u)]\overline{w}}(u)$
for all $\hat{Z} \subseteq (T^0 \cup \cdots \cup T^k) \setminus (S^k \cup \overline{W}) = V \setminus (S^k \cup \overline{W})$.

This then shows that **AC2*** is equivalent to (β) some $(\boldsymbol{p}, \boldsymbol{q}, X) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$: $(\Leftarrow)$ Suppose first that (β) holds. Hence, some $\overline{W} \subseteq V$ and some $\overline{w} \in D(\overline{W})$ exist such that $X = S^k \setminus \overline{W}$ and (a) $\neg\phi_{p\overline{w}}(u)$ for some $p \in \boldsymbol{p} \neq \emptyset$, and (b) $\phi_{[q\langle\hat{Z}(u)]\overline{w}}(u)$ for $q = x \in \boldsymbol{q}$ and all $\hat{Z} \subseteq V \setminus (S^k \cup \overline{W})$. That is, **AC2*** holds. $(\Rightarrow)$ Conversely, suppose now that **AC2*** holds. Let $(\boldsymbol{p}, \boldsymbol{q}, X)$ be defined by (i), using $W \subseteq V$ and $w \in D(W)$ from **AC2*** as $\overline{W} \subseteq V$ and $\overline{w} \in D(\overline{W})$, respectively. Then, $(\boldsymbol{p}, \boldsymbol{q}, X) \in R^k$, $\boldsymbol{p} \neq \emptyset$, and $x \in \boldsymbol{q}$. That is, (β) holds.

We give a proof of $(\star)$ by induction on $i \in \{0, \ldots, k\}$:

Basis: Recall that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^0$ iff some $W \subseteq T^0$ and $w \in D(W)$ exist such that $F = S^0 \setminus W$ and

(i) for every $p, q \in D(F)$:

(i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{pw}(u)$, (i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle \hat{Z}(u)]w}(u)$ for all $\hat{Z} \subseteq T^0 \setminus (S^k \cup W)$.

Induction: Recall that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^i$ iff some $W \subseteq T^i$, $w \in D(W)$, and $(\boldsymbol{p'}, \boldsymbol{q'}, F') \in R^{i-1}$ exist such that $F = S^i \setminus W$ and

(i′) for every $p, q \in D(F)$:

(i.1′) $p \in \boldsymbol{p}$ iff $F'_{pw}(u) \in \boldsymbol{p'}$,

(i.2′) $q \in \boldsymbol{q}$ iff $F'_{[q\langle \hat{Z}(u)]w}(u) \in \boldsymbol{q'}$ for all $\hat{Z} \subseteq T^i \setminus (S^k \cup W)$.

The induction hypothesis says that $(\boldsymbol{p'}, \boldsymbol{q'}, F') \in R^{i-1}$ iff some $\overline{W}' \subseteq T^0 \cup \dots \cup T^{i-1}$ and $\overline{w}' \in D(\overline{W}')$ exist such that $F' = S^{i-1} \setminus \overline{W}'$ and

(i″) for every $p', q' \in D(F')$:

(i.1″) $p' \in \boldsymbol{p'}$ iff $\neg\phi_{p'\overline{w}'}(u)$,

(i.2″) $q' \in \boldsymbol{q'}$ iff $\phi_{[q'\langle \hat{Z}'(u)]\overline{w}'}(u)$ for all $\hat{Z}' \subseteq (T^0 \cup \cdots \cup T^{i-1}) \setminus (S^k \cup \overline{W}')$.

It thus follows that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^i$ iff some $\overline{W}' \subseteq T^0 \cup \cdots \cup T^{i-1}$, $W \subseteq T^i$, $\overline{w}' \in D(\overline{W}')$, and $w \in D(W)$ exist such that $F = S^i \setminus W$ and

(i‴) for $F' = S^{i-1} \setminus \overline{W}'$ and every $p, q \in D(F)$:

(i.1‴) $p \in \boldsymbol{p}$ iff $\neg\phi_{p'\overline{w}'}(u)$, where $p' = F'_{pw}(u)$, by (i.1′) and (i.1″),

(i.2‴) $q \in \boldsymbol{q}$ iff $\phi_{[q'\langle \hat{Z}'(u)]\overline{w}'}(u)$, where $q' = F'_{[q\langle \hat{Z}(u)]w}(u)$, for all $\hat{Z}' \subseteq (T^0 \cup \dots \cup T^{i-1}) \setminus (S^k \cup \overline{W}')$ and all $\hat{Z} \subseteq T^i \setminus (S^k \cup W)$, by (i.2′) and (i.2″).

By **D4–D6** in the definition of a decomposition, setting some of the T^i-variables as W- or $\hat{Z}$-variables in (i.1‴) and (i.2‴) does not influence the values of the variables in $S^i \setminus (W \cup \hat{Z})$. Thus, $F'_{pw}(u) = F'_{pw\overline{w}'}(u)$, and so $\neg\phi_{p'\overline{w}'}(u) = \neg\phi_{pw\overline{w}'}(u)$. Furthermore, $A_{[q\langle \hat{Z}(u)]w}(u) = A_{[q\langle \hat{Z}(u)]w\hat{Z}'(u)\overline{w}'}(u)$ for all $A \in F' \setminus \hat{Z}'$, and thus $\phi_{[q'\langle \hat{Z}'(u)]\overline{w}'}(u)$, where $q' = F'_{[q\langle \hat{Z}(u)]w}(u)$, is equivalent to $\phi_{[q\langle \hat{Z}(u)]w\hat{Z}'(u)\overline{w}'}(u) = \phi_{[q\langle \hat{Z}(u)\hat{Z}'(u)]w\overline{w}'}(u)$. Hence, it follows that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^i$ iff some $\overline{W}' \subseteq T^0 \cup \dots \cup T^{i-1}$, $W \subseteq T^i$, $\overline{w}' \in D(\overline{W}')$, and $w \in D(W)$ exist such that $F = S^i \setminus W$ and

(i) for every $p, q \in D(F)$:

(i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{pw\overline{w}'}(u)$,

(i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle \hat{Z}(u)]w\overline{w}'}(u)$ for all $\hat{Z} \subseteq (T^0 \cup \cdots \cup T^i) \setminus (S^k \cup W \cup \overline{W}')$.

That is, it holds that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^i$ iff some $\overline{W} \subseteq T^0 \cup \dots \cup T^i$ and $\overline{w} \in D(\overline{W})$ exist such that $F = S^i \setminus \overline{W}$ and

(i) for every $p, q \in D(F)$:

(i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{p\overline{w}}(u)$,

(i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle \hat{Z}(u)]\overline{w}}(u)$ for all $\hat{Z} \subseteq (T^0 \cup \cdots \cup T^i) \setminus (S^k \cup \overline{W})$. □

Proof of Theorem 6.2. By Theorem 2.3, $X = x$ is an actual cause of ϕ under u in M iff (i) $X = x$ is a weak cause of ϕ under u in M and (ii) X is a singleton. Since deciding whether X is a singleton can clearly be done in constant time, it is sufficient to prove the statement of the theorem for the notion of weak cause. Let $\mathcal{D} = ((T^0, S^0), \ldots, (T^k, S^k))$. By Theorem 6.1, $X = x$ is a weak cause of ϕ under u in M iff (α) $X(u) = x$ and $\phi(u)$ in M, and (β) some $(\boldsymbol{p}, \boldsymbol{q}, X) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$, where R^k is computed using the decomposition $\mathcal{D}$ of $G_V(M)$ relative to X and ϕ. By Proposition 2.2, deciding whether (α) holds can be done in polynomial time. Since V is domain-bounded and $\mathcal{D}$ is width-bounded, it follows that R^0 can be computed in polynomial time, and that each R^i, $i \in \{1, \ldots, k\}$, can be computed in polynomial time from R^{i-1}. Hence, R^k can be computed in polynomial time. Since V is domain-bounded and $\mathcal{D}$ is width-bounded, it then follows that, given R^k, checking (β) can be done in constant time. In summary, deciding whether (β) holds can also be done in polynomial time. $\Box$

Proof of Theorem 6.3. By Theorem 2.3, it is sufficient to prove the statement of the theorem for the notion of weak cause. Let $X'' = X' \cap R^{\phi}_X(M)$ and $x'' = x'|X''$. By Theorem 4.7, $X' = x'$ is a weak cause of ϕ under u in M iff (i) $(X' \setminus X'')(u) = x'|(X' \setminus X'')$ in M, and (ii) $X'' = x''$ is a weak cause of ϕ under u in M^{ϕ}_X. By Proposition 4.8, $R^{\phi}_X(M)$ can be computed in linear time, and thus $X' \setminus X'' = X' \setminus R^{\phi}_X(M)$ can be computed in linear time. By Proposition 2.1, given $X' \setminus X''$, checking (i) can be done in polynomial time. In summary, deciding whether (i) holds can be done in polynomial time. By Proposition 4.8, M^{ϕ}_X can be computed in polynomial time. By Theorem 6.2, given M^{ϕ}_X, checking (ii) can be done in polynomial time. In summary, deciding whether (ii) holds can be done in polynomial time. $\Box$

Proof of Theorem 6.4. By Theorem 2.3, it is sufficient to prove the statement of the theorem for the notion of weak cause. Let $X' = X \cap \widehat{R}^{\phi}_X(M)$ and $x' = x|X'$. By Theorem 4.10, $X = x$ is a weak cause of ϕ under u in M iff (i) $(X \setminus X')(u) = x|(X \setminus X')$ in M, and (ii) $X' = x'$ is a weak cause of ϕ under u in $\widehat{M}^{\phi}_X$. By Proposition 4.11, $\widehat{R}^{\phi}_X(M)$ can be computed in linear time, and thus $X \setminus X' = X \setminus \widehat{R}^{\phi}_X(M)$ can be computed in linear time. By Proposition 2.1, given $X \setminus X'$, checking (i) can be done in polynomial time. In summary, deciding whether (i) holds can be done in polynomial time. By Proposition 4.11, $\widehat{M}^{\phi}_X$ can be computed in polynomial time. By Theorem 6.2, given $\widehat{M}^{\phi}_X$, checking (ii) can be done in polynomial time. In summary, deciding whether (ii) holds can be done in polynomial time. $\Box$

Proof of Theorem 6.5. By Theorem 2.3, it is sufficient to prove the statement of the theorem for weak causes. Since $\mathcal{D}$ (resp., $\mathcal{D}_X$) for (a) (resp., (b)) is width-bounded, it follows that $|X|$ is bounded by a constant. Moreover, if $X' = x'$ is a weak cause of ϕ under u in M, then $X'(u) = x'$ in M. Thus, it is sufficient to show that for every $X' \subseteq X$ and $x' \in D(X)$, where $x' = X'(u)$ in M, deciding whether $X' = x'$ is a weak cause of ϕ under u in M can be done in polynomial time. Observe then for (a) that $\mathcal{D}$ is also a decomposition of $G_V(M^{\phi}_X)$ relative to $X' \cap R^{\phi}_X(M)$ and ϕ. By Theorem 6.3 (resp., 6.4) for (a) (resp., (b)), it then follows that deciding whether $X' = x'$ is a weak cause of ϕ under u in M can be done in polynomial time.

In case (a), by exploiting the monotonicity of $R^{\phi}_X(M)$ w.r.t. X, we can proceed as follows, avoiding multiple computations of the set R^k. First, check that $\phi(u)$ holds and compute R^k for $\mathcal{D}$ and $X'' = X \cap R^{\phi}_X(M)$. Then, for each subset $X' \subseteq X''$ such that some triple $(\boldsymbol{p}, \boldsymbol{q}, X')$ exists in R^k such that $\boldsymbol{p} \neq \emptyset$ and $x' \notin \boldsymbol{q}$, where $x' = X'(u)$ in M, we have that $X' = x'$ is a weak cause of ϕ under u in M. Extending each such X' by an arbitrary subset Z of variables from $X \setminus X''$, we obtain that $X'Z = x'z$, where $z = Z(u)$ in M, is also a weak cause of ϕ under u. In this way, all weak causes $X' = x'$ for ϕ under u in M where $X' \subseteq X$ can be computed.

For computing all actual causes in case (a), by Theorem 2.3, one can similarly check that $\phi(u)$ holds, compute R^k for $\mathcal{D}$ and $X'' = X \cap R^{\phi}_X(M)$, and then output $X' = x'$ for each tuple $(\boldsymbol{p}, \boldsymbol{q}, X')$ in R^k such that $X' \subseteq X$ is a singleton and $x' = X'(u)$ in M. No extension of X' by variables Z from $X \setminus X''$ needs to be considered. □

Proof of Theorem 6.6. Recall that $X = x$ is an explanation of ϕ relative to $\mathcal{C}$ iff **EX1** $\phi(u)$ for every $u \in \mathcal{C}$, **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 (1) $X'(u) = x|X'$ and (2) $X' = x|X'$ is not a weak cause of ϕ under u, and **EX4** $X(u) = x$ and $X(u') \neq x$ for some $u, u' \in \mathcal{C}$. By Proposition 2.2, checking whether **EX1** and **EX4** hold can be done in polynomial time. By Theorem 6.3 (resp., 6.4) for (a) (resp., (b)), deciding whether $X = x$ is a weak cause of ϕ under some $u \in \mathcal{C}$ such that $X(u) = x$ can be done in polynomial time. Thus, by Proposition 2.1, deciding whether **EX2** holds can be done in polynomial time. We finally show that checking **EX3** is possible in polynomial time. For (a), notice that $\mathcal{D}$ is also a decomposition of $G_V(M^{\phi}_X)$ relative to $X' \cap R^{\phi}_X(M)$ and ϕ, for each $X' \subset X$. Since $\mathcal{D}$ (resp., $\mathcal{D}_X$) for (a) (resp., (b)) is width-bounded, it follows that $|X|$ is bounded by a constant. By Proposition 2.1 and Theorem 6.3 (resp., 6.4) for (a) (resp., (b)), deciding whether (1) $X'(u) = x|X'$ and (2) $X' = x|X'$ is not a weak cause of ϕ under some $u \in \mathcal{C}$ can be done in polynomial time, for every $X' \subset X$. Hence, deciding whether **EX3** holds can be done in polynomial time. In summary, deciding whether **EX1–EX4** hold can be done in polynomial time. □

Proof of Theorem 6.7. We first compute the set $\mathcal{C}^{\star}$ of all $u \in \mathcal{C}$ such that either (i) $X(u) \neq x$ in M, or (ii) $X(u) = x$ and $X = x$ is a weak cause of ϕ under u in M. By Proposition 2.1 and Theorem 6.3 (resp., 6.4) for (a) (resp., (b)), this can be done in polynomial time. If $X = x$ is a partial explanation of ϕ relative to $(\mathcal{C}, P)$ in M, then $\mathcal{C}^{\phi}_{X=x}$ is defined, and $\mathcal{C}^{\phi}_{X=x} = \mathcal{C}^{\star}$ by Proposition 2.4. Given $\mathcal{C}^{\phi}_{X=x}$, the explanatory power $P(\mathcal{C}^{\phi}_{X=x} \mid X = x)$ is computable in polynomial time by Proposition 2.1, if P is computable in polynomial time, as usual. In summary, this shows (3).

To check partial (resp., α-partial) explanations in (1) (resp., (2)), we compute $\mathcal{C}^{\star}$ as above. We then check that $\mathcal{C}^{\phi}_{X=x}$ is defined. That is, by Proposition 2.4, we check that $X = x$ is an explanation of ϕ relative to $\mathcal{C}^{\star}$ in M, which is possible in polynomial time by Theorem 6.6. Then, $\mathcal{C}^{\phi}_{X=x} = \mathcal{C}^{\star}$ by Proposition 2.4. We finally compute $P(\mathcal{C}^{\phi}_{X=x} \mid X = x)$ as above and check that it is positive (resp., at least α), which can be done in polynomial time. In summary, this proves (1) (resp., (2)). □

Proof of Theorem 6.8. Observe that the set of all $X' = x'$ such that $X' \subseteq X$ and $x' \in D(X')$ is bounded by a constant, since V is domain-bounded, and $\mathcal{D}$ (resp., $\mathcal{D}_X$) for (a) (resp., (b)) is width-bounded, and thus $|X|$ is bounded by a constant. Hence, it is sufficient to show that for every $X' \subseteq X$ and $x' \in D(X')$, deciding whether $X' = x'$ is an explanation of ϕ relative to $\mathcal{C}$ in M is possible in polynomial time. This can be done in a similar way as the proof of Theorem 6.6. □

Proof of Theorem 6.9. As argued in the proof of Theorem 6.8, the set of all $X' = x'$ such that $X' \subseteq X$ and $x' \in D(X')$ is bounded by a constant. Hence, it is sufficient to show that for every $X' \subseteq X$ and $x' \in D(X')$, deciding whether $X' = x'$ is a partial (resp., an α-partial) explanation of ϕ relative to $(\mathcal{C}, P)$ in M is possible in polynomial time. This can be done in the same way as the proof of Theorem 6.7 (1) (resp., (2)), using only Theorem 6.8 instead of Theorem 6.6. □

Proof of Theorem 6.10. We generalize the proof of Theorem 6.1. We show that some $(\boldsymbol{p}, \boldsymbol{q}, X, l) \in R^k$ exists with $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$ iff **AC2l** holds:

AC2l. Some $W \subseteq V \setminus X$ and some $\overline{x} \in D(X)$ and $w \in D(W)$ exist such that:

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

(b) $\phi_{xw\hat{Z}(u)}(u)$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$,

(c) $\mathrm{diff}(w, W(u)) = l$.

As in the proof of Theorem 6.1, by moving any $A \in S^k \setminus (W \cup X)$ into W by setting $w(A) = A(u)$ (which does not influence $\mathrm{diff}(w, W(u))$), it is sufficient to show that some $(\boldsymbol{p}, \boldsymbol{q}, X, l) \in R^k$ exists with $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$ iff **AC2l′** holds:

AC2l′. Some $W \subseteq V$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that $X = S^k \setminus W$ and

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

(b) $\phi_{xw\hat{Z}(u)}(u)$ for all $\hat{Z} \subseteq V \setminus (S^k \cup W)$,

(c) $\mathrm{diff}(w, W(u)) = l$.

This can be done in a similar way as showing that (β) is equivalent to **AC2′** in the proof of Theorem 6.1, where we use the following result $(\star\star)$ instead of $(\star)$, which can be proved by induction on $i \in \{0, \ldots, k\}$ (in a similar way as $(\star)$): $(\star\star)$ For all $i \in \{0, \ldots, k\}$, it holds that $(\boldsymbol{p}, \boldsymbol{q}, F, l) \in R^i$ iff some $\overline{W} \subseteq T^0 \cup \cdots \cup T^i$ and $\overline{w} \in D(\overline{W})$ exist such that $F = S^i \setminus \overline{W}$, $\mathrm{diff}(\overline{w}, \overline{W}(u)) = l$, and

(i) for every $p, q \in D(F)$:

(i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{p\overline{w}}(u)$,

(i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle\hat{Z}(u)]\overline{w}}(u)$ for all $\hat{Z} \subseteq (T^0 \cup \cdots \cup T^i) \setminus (S^k \cup \overline{W})$. □

Proof of Theorem 6.11. We first decide if $(\star)$ $X = x$ is an actual cause of ϕ under u in M, which can be done in polynomial time by Theorem 6.2. If $(\star)$ does not hold, then $\mathrm{dr}((M, u), X{=}x, \phi) = 0$. Otherwise, $\mathrm{dr}((M, u), X{=}x, \phi) = 1/(l^\star{+}1)$, where $l^\star$ is the minimal l for which some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that **AC2**(a) and (b) hold and $\mathrm{diff}(w, W(u)) = l$. By Theorem 6.10, $l^\star$ is the minimal l for which some $(\boldsymbol{p}, \boldsymbol{q}, X, l) \in R^k$ exists such that $\boldsymbol{p} \neq \emptyset$ and $x \in \boldsymbol{q}$. Since V is domain-bounded and $\mathcal{D}$ is width-bounded, R^0 can be computed in polynomial time, and each R^i, $i \in \{1, \ldots, k\}$, can be computed in polynomial time from R^{i-1}. Thus, R^k can be computed in polynomial time. Since V is domain-bounded and $\mathcal{D}$ is width-bounded, $l^\star$ can be computed in polynomial time from R^k. In summary, $l^\star$ and thus $\mathrm{dr}((M, u), X{=}x, \phi) = 1/(l^\star{+}1)$ can be computed in polynomial time. □

Proof of Theorem 6.12. By Theorem 6.11, every $\mathrm{dr}((M, u), X{=}x, \phi)$, $(M, u) \in \mathcal{K}$, can be computed in polynomial time. Assuming that P can be computed in polynomial time, also $\mathrm{db}(\mathcal{K}, P, X \leftarrow x, \phi)$ can be computed in polynomial time. □

D Appendix: Proofs for Section 7

Proof of Proposition 7.1. Let $(S^0, \ldots, S^k)$ be an arbitrary layering of $G_V(M)$ w.r.t. X and ϕ. We now show that $((T^0, S^0), \ldots, (T^k, S^k))$, where $T^0 = S^0, \ldots, T^k = S^k$, is a decomposition of $G_V(M)$ w.r.t. X

and ϕ, that is, that **D1**–**D6** hold. Trivially, **D1** and **D2** hold. Moreover, **L2** implies **D3**, and **L1** implies **D4**–**D6**. □

Proof of Proposition 7.2. Assume that $\mathcal{L} = (S^0, \ldots, S^k)$ is an arbitrary layering of $G_V(M)$ relative to X and ϕ. By **L2**, every $A \in V(\phi) \cap V$ belongs to S^0, and at least one such variable exists. By **L2** and since $G_V(M)$ is connected relative to X and ϕ, every variable $A \in X$ belongs to S^k, and at least on such variable exists, where k is given via a path P from a variable $B \in V(\phi)$ to a variable in X (in the undirected graph for $G_V(M)$) as the number of arrows in $G_V(M)$ that go against the direction of P minus the number of arrows in $G_V(M)$ that go in the same direction as P. Indeed, if we move from B to A (against the direction of P), any step backwards toward S^i must be compensated later with a step forward. By **L1** and since $G_V(M)$ is connected relative to X and ϕ, for every $i \in \{0, \ldots, k\}$, the set S^i is the set of all $A \in V$ that are reachable from some $B \in X \cup V(\phi)$ on a path P (in the undirected graph for $G_V(M)$) such that i is the number of arrows in $G_V(M)$ that go against the direction of P minus the number of arrows in $G_V(M)$ that go in the same direction as P plus j with $B \in S^j$. That is, the layering $\mathcal{L}$ is unique. □

Proof of Proposition 7.3. In Step (1), we initialize $\lambda(A)$ to undefined for all $A \in V \backslash V(\phi)$. In Step (2), every variable occurring in ϕ is put into S^0, in order to satisfy one part of **L2**. In Steps (3)–(13), since $G_V(M)$ is connected, all the other variables are put into some S^j such that **L1** is satisfied. Step (3) takes care of the special case in which variables from ϕ belong to X, where then only a trivial layered decomposition is possible. Steps (6) and (11) catch cases in which no layering mapping as desired exists, and then *Nil* is returned. Notice that the for-loop in Step (9) is executed at most once. Finally, we check in Steps (14) and (15) that $X \subseteq S^k$, where k is the maximal index j of some S^j, and thus whether the other part of **L2** is also satisfied. If so, then we return the computed layering λ; otherwise, we return *Nil*. □

Proof of Proposition 7.4. By Proposition 7.2, if a layering of $G_V(M)$ relative to X and ϕ exists, then it is unique. By Proposition 7.3, Algorithm LAYERING returns the unique layering $\mathcal{L}$ of $G_V(M)$ relative to X and ϕ, if it exists, and *Nil*, otherwise. Observe then that Steps (1)-(3) of LAYERING take $O(|V| + |V(\phi)|)$ time, Steps (4)-(13) take $O(|E| + |X|)$ time, and Step (14) is feasible in $O(|V|)$ time (using an auxiliary variable for the maximum of λ, even in constant time). Hence, LAYERING can be implemented to run in $O(|V| + |V(\phi)| + |E|)$ time, i.e., in $O(\|G_V(M)\| + |V(\phi)|)$ time. Given that $G_V(M)$ is layered, deciding whether $\mathcal{L}$ is width-bounded by some integer $l \geq 0$ can be done in time in $O(|V|)$. □

E Appendix: Proofs for Section 8

Proof of Theorem 8.1. We generalize the proof of Theorem 4.1 (resp., 4.2) to the refined notion of weak cause. Let $X_0 \in X$ be such that (α) there is no directed path in $G(M)$ from X_0 to a variable in ϕ (resp., (β) each directed path in $G(M)$ from X_0 to a variable in ϕ contains some $X_i \in X \backslash \{X_0\}$). Let $X'' = X \backslash \{X_0\}$ and $x'' = x|X''$. It is now sufficient to show that $X = x$ is a (refined) weak cause of ϕ under u iff (i) $X_0(u) = x(X_0)$ and (ii) $X'' = x''$ is a (refined) weak cause of ϕ under u.

($\Rightarrow$) Suppose $X = x$ is a (refined) weak cause of ϕ under u. That is, **AC1** $X(u) = x$ and $\phi(u)$, and **AC2′** some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}w}(u)$ and (b) $\phi_{xw'\hat{z}}(u)$ for all $W' \subseteq W$, $\hat{Z} \subseteq V \setminus (X \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$. In particular, (i) $X_0(u) = x(X_0)$, and also **AC1** $X''(u) = x''$ and $\phi(u)$. By (α) (resp., (β)), it then follows that **AC2′**(a) $\neg\phi_{\overline{x}''w}(u)$ and (b) $\phi_{x''w'\hat{z}}(u)$ hold for all $W' \subseteq W$, $\hat{Z} \subseteq V \setminus (X'' \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$, where $\overline{x}'' = \overline{x}|X''$. This shows that (ii) $X'' = x''$ is a (refined) weak cause of ϕ under u.

($\Leftarrow$) Suppose (i) $X_0(u) = x(X_0)$ and (ii) $X'' = x''$ is a (refined) weak cause of ϕ under u. That is, **AC1** $X''(u) = x''$ and $\phi(u)$, and **AC2′** some $W \subseteq V \setminus X''$, $\overline{x}'' \in D(X'')$, and $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}''w}(u)$, and (b) $\phi_{x''w'\hat{z}}(u)$ for all $W' \subseteq W$, $\hat{Z} \subseteq V \setminus (X'' \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$. By (i), we thus obtain **AC1** $X(u) = x$ and $\phi(u)$. By (α) (resp., (β)), it follows that **AC2′**(a) $\neg\phi_{\overline{x}''\overline{x}_0 w'}(u)$ and (b) $\phi_{x'' x_0 w'' \hat{z}}(u)$ for all $W'' \subseteq W'$, $\hat{Z} \subseteq V \setminus (X \cup W')$, $w'' = w'|W''$, and $\hat{z} = \hat{Z}(u)$, where $W' = W \setminus \{X_0\}$, $w' = w|W'$, $\overline{x}_0 = (X_0)_{\overline{x}''w}(u)$, and $x_0 = x(X_0)$. This shows that $X = x$ is a (refined) weak cause of ϕ under u. $\Box$

Proof of Theorem 8.2. Let $M' = M^{\phi}_X$ (resp., $M' = \widehat{M}^{\phi}_X$). We prove the statement of the theorem for the case $X' = X$ and $M' = M^{\phi}_X$. The proof for $X' = X$ and $M' = \widehat{M}^{\phi}_X$ can be done in a similar way, using $\widehat{M}^{\phi}_X$ instead of M^{ϕ}_X. The proof for $X' \subset X$ and $M' = M^{\phi}_X$ is similar to the proof of Theorem 4.7.

Let $X' = X$ and $M' = M^{\phi}_X = (U, V', F')$. We extend the proof of Theorem 4.5 to the refined notion of weak cause. Let $X'' = X' \cap V'$ and $x'' = x'|X''$. We have to show that $X' = x'$ is a (refined) weak cause of ϕ under u in M iff (i) $(X' \setminus X'')(u) = x'|(X' \setminus X'')$ in M, and (ii) $X'' = x''$ is a (refined) weak cause of ϕ under u in M^{ϕ}_X.

Fact A. $V'_M(u) = V'_{M^{\phi}_X}(u)$ and $\phi_M(u) = \phi_{M^{\phi}_X}(u)$.

($\Rightarrow$) Suppose $X' = x'$ is a (refined) weak cause of ϕ under u in M. That is, **AC1** $X'(u) = x'$ and $\phi(u)$ in M, and **AC2′** some $W \subseteq V \setminus X'$, $\overline{x}' \in D(X')$, $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}'w}(u)$ in M and (b) $\phi_{x'w'\hat{z}}(u)$ in M for all $W' \subseteq W$, $\hat{Z} \subseteq V \setminus (X' \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$ in M. This shows that (i) $(X' \setminus X'')(u) = x'|(X' \setminus X'')$ in M. We next show that also (ii) holds. By Fact A, **AC1** $X''(u) = x''$ and $\phi(u)$ in M^{ϕ}_X. Notice then that (a) $\neg\phi_{\overline{x}''\overline{w}}(u)$ in M and (b) $\phi_{x''\overline{w}'\hat{z}'}(u)$ in M, where $\overline{x}'' = \overline{x}'|X''$, $\overline{W} = W \cap V'$, $\overline{w} = w|\overline{W}$, $\overline{W}' = W' \cap V'$, $\overline{w}' = w'|\overline{W}' = \overline{w}|\overline{W}'$, $\hat{Z}' = \hat{Z} \cap V'$, and $\hat{z}' = \hat{z}|\hat{Z}'$. Since each among $\neg\phi_{\overline{x}''\overline{w}}(u)$, $\phi_{x''\overline{w}'\hat{z}'}(u)$, and $\hat{Z}'(u)$ has the same values in M and M^{ϕ}_X, this shows that **AC2′** (a) $\neg\phi_{\overline{x}''\overline{w}}(u)$ in M^{ϕ}_X and (b) $\phi_{x''\overline{w}'\hat{z}'}(u)$ in M^{ϕ}_X for all $\hat{Z}' \subseteq V' \setminus (X'' \cup \overline{W})$, $\overline{W}' \subseteq \overline{W}$, $\overline{w}' = \overline{w}|\overline{W}'$, and $\hat{z}' = \hat{Z}'(u)$ in M^{ϕ}_X. In summary, (ii) $X'' = x''$ is a (refined) weak cause of ϕ under u in M^{ϕ}_X.

($\Leftarrow$) Suppose (i) $(X' \setminus X'')(u) = x'|(X' \setminus X'')$ in M and (ii) $X'' = x''$ is a (refined) weak cause of ϕ under u in M^{ϕ}_X. Thus, **AC1** $X''(u) = x''$ and $\phi(u)$ in M^{ϕ}_X, and **AC2′** some $W \subseteq V' \setminus X''$, $\overline{x}'' \in D(X'')$, $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}''w}(u)$ in M^{ϕ}_X and (b) $\phi_{x''w'\hat{z}}(u)$ in M^{ϕ}_X for all $W' \subseteq W$, $\hat{Z} \subseteq V' \setminus (X'' \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$ in M^{ϕ}_X. By Fact A, **AC1** $X'(u) = x'$ and $\phi(u)$ in M. Since each among $\neg\phi_{\overline{x}''w}(u)$, $\phi_{x''w'\hat{z}}(u)$, and $\hat{Z}(u)$ has the same values in M and M^{ϕ}_X, this shows that (a) $\neg\phi_{\overline{x}''w}(u)$ in M and (b) $\phi_{x''w'\hat{z}}(u)$ in M for all $W' \subseteq W$, $\hat{Z} \subseteq V' \setminus (X'' \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$ in M. It then follows that **AC2′** (a) $\neg\phi_{\overline{x}'w}(u)$ in M and (b) $\phi_{x'w'\hat{z}}(u)$ in M for all $W' \subseteq W$, $\hat{Z} \subseteq V \setminus (X' \cup W)$, $w' = w|W'$, and $\hat{z} = \hat{Z}(u)$ in M, where $\overline{x}'|X'' = \overline{x}''$ and $\overline{x}'|(X' \setminus X'') = (X' \setminus X'')_{\overline{x}''w}(u)$ in M. In summary, this shows that $X' = x'$ is a (refined) weak cause of ϕ under u in M. $\Box$

Proof of Theorem 8.3. The proof is nearly identical to the proof of Theorem 6.1, except that **AC2** is now replaced by **AC2′** (for the refined notion of weak cause), the relations R^i are replaced by the relations R^i for the refined notion of weak cause, and ($\star$) is replaced by the following statement ($\star'$): for all $i \in \{0, \ldots, k\}$, it holds that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^i$ iff some $\overline{W} \subseteq T^0 \cup \cdots \cup T^i$ and $\overline{w} \in D(\overline{W})$ exist such that $F = S^i \setminus \overline{W}$ and (i) for every $p, q \in D(F)$: (i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{p\overline{w}}(u)$, and (i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle\hat{Z}(u)]\overline{w}'}(u)$ for all $\hat{Z} \subseteq (T^0 \cup \cdots \cup T^i) \setminus (S^k \cup \overline{W})$, $\overline{W}' \subseteq \overline{W}$, and $\overline{w}' = \overline{w}|\overline{W}'$. $\Box$

Proof of Theorem 8.4. The proof is nearly identical to the proof of Theorem 8.1, except that **AC2′** is now replaced by **AC2″** (for the refined notion of weak cause in extended causal models). In the "$\Rightarrow$"-part, we use that $\overline{x}''w$ is allowable if $\overline{x}w$ is allowable, while in the "$\Leftarrow$"-part, we use that $\overline{x}''\overline{x}_0 w'$ is allowable if $\overline{x}''w$ is allowable, which follows from the assumption that M is closed relative to X''. $\Box$

Proof of Theorem 8.5. (a) Let $V' = R^{\phi}_{X}(M)$ and $M' = M^{\phi}_{X}$. Assume M is closed. Let $Y \subseteq V'$, let y be an allowable setting for Y in M', and let $u \in D(U)$. Then, y is an allowable setting for Y in M, and $(V' \backslash Y)_y(u)$ has the same value in M and M'. Since M is closed, $y \cup (V \backslash Y)_y(u)$ is an allowable setting for Y in M, and thus $y \cup (V' \backslash Y)_y(u)$ is an allowable setting for Y in M'. Hence, M' is closed.

(b) Let $V' = \widehat{R}^{\phi}_{X}(M)$ and $M' = \widehat{M}^{\phi}_{X}$. Suppose M is closed relative to X'. Let $Y \subseteq V'$ with $X' \subseteq Y$, let y be an allowable setting for Y in M', and let $u \in D(U)$. Then, y is an allowable setting for Y in M, and $(V' \backslash Y)_y(u)$ has the same value in M and M'. Since M is closed relative to X', it follows that $y \cup (V \backslash Y)_y(u)$ is an allowable setting for Y in M, and thus $y \cup (V' \backslash Y)_y(u)$ is an allowable setting for Y in M'. This shows that M' is closed relative to X'. $\Box$

Proof of Theorem 8.6. The proof is nearly identical to the proof of Theorem 8.2, except that **AC2′** is now replaced by **AC2″** (for the refined notion of weak cause in extended causal models). In the "$\Rightarrow$"-part, we use that $\overline{x}''\overline{w}$ is allowable in M^{ϕ}_{X} if $\overline{x}'w$ is allowable in M, while in the "$\Leftarrow$"-part, we use that $\overline{x}'w$ is allowable in M if $\overline{x}''w$ is allowable in M^{ϕ}_{X}, which follows from M being closed relative to X''.

Proof of Theorem 8.7. The proof is nearly identical to the proof of Theorem 8.3, except that **AC2′** is now replaced by **AC2″** (for the refined notion of weak cause in extended causal models), the relations R^i for the refined notion of weak cause are replaced by the relations R^i for the refined notion of weak cause in extended causal models, and $(\star')$ is replaced by the following statement $(\star'')$: for all $i \in \{0, \ldots, k\}$, it holds that $(\boldsymbol{p}, \boldsymbol{q}, F) \in R^i$ iff some $\overline{W} \subseteq T^0 \cup \cdots \cup T^i$ and $\overline{w} \in D(\overline{W})$ exist such that $F = S^i \setminus \overline{W}$ and (i) for every $p, q \in D(F)$: (i.1) $p \in \boldsymbol{p}$ iff $\neg\phi_{p\overline{w}}(u)$ and $p\overline{w}|(X \cup W)$ is allowable, and (i.2) $q \in \boldsymbol{q}$ iff $\phi_{[q\langle\hat{Z}(u)]\overline{w}'}(u)$ for all $\hat{Z} \subseteq (T^0 \cup \cdots \cup T^i) \setminus (S^k \cup \overline{W})$, $\overline{W}' \subseteq \overline{W}$, and $\overline{w}' = \overline{w}|\overline{W}'$. Observe that in the step from **AC2″** to **(AC2″)***, we then use the assumption that M is closed relative to X. Moreover, in the induction step, we use the property **D7** of decompositions in extended causal models. $\Box$

References

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

[2] U. Chajewska and J. Y. Halpern. Defining explanation in probabilistic systems. In *Proceedings UAI-1997*, pp. 62–71. Morgan Kaufmann, 1997.

[3] H. Chockler and J. Y. Halpern. Responsibility and blame: A structural-model approach. In *Proceedings IJCAI-2003*, pp. 147–153. Morgan Kaufmann, 2003. Extended version in *J. Artif. Intell. Res.*, 22:93–115, 2004.

[4] H. Chockler, J. Y. Halpern, and O. Kupferman. What causes a system to satisfy a specification? CoRR, 2003.

[5] E. Dantsin, T. Eiter, G. Gottlob, and A. Voronkov. Complexity and expressive power of logic programming. *ACM Comput. Surv.* 33(3):374–425, 2001.

[6] T. Eiter and T. Lukasiewicz. Complexity results for structure-based causality. In *Proceedings IJCAI-2001*, pp. 35–40. Morgan Kaufmann, 2001. Extended version in *Artif. Intell.*, 142(1):53–89, 2002.

[7] T. Eiter and T. Lukasiewicz. Complexity results for explanations in the structural-model approach. In *Proceedings KR-2002*, pp. 49–60. Morgan Kaufmann, 2002. Extended version in *Artif. Intell.*, 154(1–2):145–198, 2004.

[8] T. Eiter and T. Lukasiewicz. Causes and explanations in the structural-model approach: Tractable cases. In *Proceedings UAI-2002*, pp. 146–153. Morgan Kaufmann, 2002.

[9] A. Finzi and T. Lukasiewicz. Structure-based causes and explanations in the independent choice logic. In *Proceedings UAI-2003*, pp. 225–232. Morgan Kaufmann, 2003.

[10] D. Galles and J. Pearl. Axioms of causal relevance. *Artif. Intell.*, 97(1–2):9–43, 1997.

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

[12] H. Geffner. Causal theories for nonmonotonic reasoning. In *Proceedings AAAI-1990*, pp. 524–530. AAAI Press / MIT Press, 1990.

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

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

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

[16] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach – Part I: Causes. In *Proceedings UAI-2001*, pp. 194–202. Morgan Kaufmann, 2001.

[17] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach – Part I: Causes. *British Journal for the Philosophy of Science*, 2005.

[18] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach – Part II: Explanations. In *Proceedings IJCAI-2001*, pp. 27–34. Morgan Kaufmann, 2001. Extended version in *British Journal for the Philosophy of Science*, 2005.

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

[20] 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.

[21] M. Hopkins. Strategies for determining causes of reported events. In *Proceedings AAAI-2002*, pp. 546–552. AAAI Press, 2002.

[22] M. Hopkins. LAYERWIDTH: Analysis of a new metric for directed acyclic graphs. In *Proceedings UAI-2003*, pp. 321–328. Morgan Kaufmann, 2003.

[23] M. Hopkins and J. Pearl. Causality and counterfactuals in the situation calculus. Technical Report R-301, UCLA Cognitive Systems Laboratory, 2002.

[24] M. Hopkins and J. Pearl. Clarifying the usage of structural models for commonsense causal reasoning. In *Proceedings AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*. AAAI Press, 2003.

[25] N. Leone, G. Pfeifer, W. Faber, T. Eiter, G. Gottlob, S. Perri, and F. Scarcello. The DLV System for knowledge representation and reasoning. *ACM Transactions on Computational Logic (TOCL)*, forthcoming.

[26] V. Lifschitz. On the logic of causal explanation. *Artif. Intell.*, 96(2):451–465, 1997.

[27] N. McCain and H. Turner. Causal theories of action and change. In *Proceedings AAAI-1997*, pp. 460–465. AAAI Press / MIT Press, 1997.

[28] J. D. Park. Causes and explanations revisited. In *Proceedings IJCAI-2003*, pp. 154–162. Morgan Kaufmann, 2003.

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

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

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

[32] D. Poole. The independent choice logic for modelling multiple agents under uncertainty. *Artif. Intell.*, 94(1–2):7–56, 1997.

[33] D. Poole. Decision theory, the situation calculus and conditional plans. *Electronic Transactions on Artificial Intelligence*, 2(1–2):105–158, 1998.

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

[35] S. E. Shimony. Explanation, irrelevance, and statistical independence. In *Proceedings AAAI-1991*, pp. 482–487. AAAI Press / MIT Press, 1991.

[36] H. Turner. A logic of universal causation. *Artif. Intell.*, 113(1–2):87–123, 1999.