Artificial Intelligence, 57(2-3):227–270, 1992.

On the Complexity of Propositional Knowledge Base Revision, Updates, and Counterfactuals*

Thomas Eiter and Georg Gottlob

Christian Doppler Laboratory for Expert Systems
Information Systems Department†
Technical University of Vienna

Abstract

We study the complexity of several recently proposed methods for updating or revising propositional knowledge bases under the principle of minimal change. In particular, we derive complexity results for the following problem: given a knowledge base T, an update p, and a formula q, decide whether q is derivable from $T \circ p$, the updated (or revised) knowledge base. Note that this problem includes the evaluation of the counterfactual $p > q$ over T, that is a conditional statement "if p, then q" where p is known or expected to be false. We consider the general case where T is an arbitrary propositional formula (or theory) as well as restricted versions of this problem, in particular where T is a conjunction of Horn clauses, or where the size of the update p is bounded by a constant.

1 Introduction

In many areas of Artificial Intelligence, one has to deal with *knowledge in flux*, i.e. with knowledge bases that are subject to frequent updates or revisions in the light of new information [15, 58]. Different approaches for changing knowledge bases have been proposed recently by several researchers. All of these methods adhere to the principle of *Minimality of Change* which states that the knowledge base should change as little as possible if new information is incorporated.

In this paper we deal with the following methods of updating knowledge bases: Ginsberg's approach [19], also considered by Fagin, Ullman, and Vardi [13], and by Nebel [44, 45],

*An extended abstract of this paper appeared in the proceedings of the Eleventh ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, San Diego, CA, June 2-4, 1992.

†Mailing address: Paniglgasse 16, A-1040 Wien, Austria.
Internet e-mail: {eiter,gottlob}@vexpert.dbai.tuwien.ac.at

the WIDTIO approach [19, 21, 58], Dalal's revision [7], the approach by Forbus [14], Borgida's method [3], Winslett's "Possible Models Approach" [57], Satoh's method [53], and Weber's approach [56].

Each of these methods is suited for particular application domains; most researchers agree that none of them can be considered as a general purpose means of changing knowledge bases. Moreover, it is now widely believed that there is no general purpose, domain independent means of updating knowledge bases that will "do the right thing" under all circumstances [58].

As pointed out by Katsuno and Mendelzon [30], as well as by various other authors [58, 45, 23], an important problem to explore is the computational complexity of the different approaches. Note that – though all methods are clearly intractable in the general case – for almost all of the cited methods a precise determination of the complexity was left as an open issue. Furthermore, it was not clear under which restrictions the single methods would become tractable.

In this paper we shed light on these complexity problems. In particular, we consider the following "implication problem": given a knowledge base T, an update p, and a formula q, decide whether q is derivable from $T \circ p$, the updated (or revised) knowledge base. For each of the different update or revision methods, we consider the general case where T is an arbitrary propositional formula as well as restricted versions of this problem where T is a conjunction of Horn clauses (or a set of Horn conjunctions), or where the size of the knowledge change p is bounded by a constant. The former restriction is an important special case to consider because many rule-based systems rely on Horn clauses, and the latter restriction is, according to Winslett [58, p.33], reasonable for consideration, since this assumption will certainly be met by human-generated updates. In each case (and for all combinations of cases) we are able to locate the complexity of this decision problem at the appropriate level of the polynomial hierarchy. In particular, we succeed in elucidating the tractability/intractability frontier for many approaches.

Note that the implication problem we consider exactly corresponds to evaluating a counterfactual according to the given revision or update semantics. Counterfactuals are conditional statements of the form "if p were true, then q would hold", where p is assumed to be false in the actual world. According to the *Ramsey Test* [15], evaluating such a counterfactual in a given knowledge base T is equivalent to test whether q is a logical consequence of $T \circ p$.

By our results it turns out that for all approaches (except for Dalal's) the implication problem is at the second level of the polynomial hierarchy, thus harder than classical implication in propositional logic. In the Horn case, the complexity level of most approaches is lowered by one, but the problem remains intractable in each case. The same happens if the size of the knowledge change (formula p) is bounded by a constant. Tractability is obtained only if we consider both restrictions simultaneously.

Our results extend and complement very recent results by Nebel [45] and Winslett [58]. Nebel has shown that the implication problem under Ginsberg's approach is $\mathbf{\Pi}_2^P$-complete (thus at the second level of the polynomial hierarchy). His proof, however, does not

carry over to most other approaches that we discuss. Nebel also shows that evaluating a counterfactual in Dalal's setting is both **NP**-hard and co-**NP**-hard, but in Δ_2^P, hence solvable with a polynomial number of calls to an **NP** oracle (i.e. to a subroutine solving an **NP**-complete problem). We improve this result by showing that a logarithmic number of such oracle calls suffices for this task and that the problem is, moreover, also complete for the class $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$, which are the decision problems solvable in polynomial time by a logarithmic number of calls to an **NP** oracle.

Winslett [58] provides update algorithms for a number of approaches and derives from these algorithms bounds for computing the updated knowledge base $T \circ p$ from T and p. Note that this problem differs from the implication problem we consider. We will refer to Winslett's results in Section 8. Winslett poses the following issue as an open problem [58, p.33]: Is it $\mathbf{\Sigma}_2^P$-complete to decide if a change under the WIDTIO semantics removes a particular formula f from the knowledge base (the length of the update formula p is bounded by a constant) ? We show that the answer is yes, even under the restriction that f and p are literals (Section 8, discussion of Theorem 8.2).

The rest of this paper is organized as follows. In Section 2 we introduce some notation and introduce the relevant concepts of complexity theory. In Section 3 we report about basic rationality postulates for change operators, review the change operators under consideration, and relate change operators to counterfactuals. Section 4 presents an overview and a thorough discussion of our results. Complexity classes or bounds for each update operator under all possible restrictions are shown in a table. Note that we are able to determine the exact complexity of most change operators by a completeness result. In Section 5, we point out consequences of the complexity findings and consider related results. Sections 6-8 constitute the technical part of this paper, and state the complexity theorems and their proofs. The proofs are succinct; often one single transformation covers a number of different change semantics simultaneously. In particular, Section 6 contains the complexity analysis for the general case, while Sections 7 and 8 treat the Horn case and the case of bounded knowledge change, respectively. Section 9 concludes the paper.

2 Preliminaries

Throughout this paper, the language L of propositional logic with finite variable set V is considered. A knowledge base is a finite set of propositional formulae in L. Unless stated otherwise, T always denotes an arbitrary knowledge base, and p, q are a propositional formulae. As usual, we will identify T with the conjunction of its formulae. If $S = \{T_1, \ldots, T_n\}$ is a family of knowledge bases $T_1, \ldots, T_n$, we identify S with the formula $T_1 \vee \cdots \vee T_n$.

$V(T)$ denotes the variables of V that occur in T. Any truth value assignment to V is a model, which we identify with the set of variables that are assigned true by this assignment. Accordingly, set operations on models M, M' (e.g., $M - M'$) refer to the respective sets of true variables of M, M'.

A model M is a model of a formula p iff p is satisfied by M (written $M \models p$). The set of all models of p is denoted by $Mod(p)$. As usual, we write $p \models q$ if satisfaction of p entails satisfaction of q, i.e. $Mod(p) \subseteq Mod(q)$ holds. We say that p is valid in the knowledge base T iff $T \models p$ holds.

T is *complete* if for every formula p in L it holds that $T \models p$ or $T \models \neg p$. Every $x \in V$ is an *atom*, and T is called *atomic* if every formula in T is an atom. Any atom or negated atom is a *literal*, and T is a *literal base* if it contains only literals. A *clause* is a disjunction $l_1 \vee \cdots \vee l_n$ of literals l_i.

Given a knowledge base T and a formula p, $T \circ p$ denotes the change of T by incorporating p into the knowledge base applying the change operator $\circ$. The size, i.e. the length of a formula p is denoted by $\|p\|$.

We assume that the reader is familiar with the basic concepts of the theory of **NP**-completeness, as excellently presented in [16]. Many of the problems that we will discuss are most probably not in **NP** or co-**NP**, but reside in low levels of the *polynomial hierarchy* (PH). The PH is a hierarchy of complexity classes that account for computations that may query a certain oracle (rf. [16]). An oracle for a class of decision problems **C** enables to decide any problem in **C** in unit time; one can think of an oracle as a "subroutine without cost". $\mathbf{P}^{\mathbf{C}}$ ($\mathbf{NP}^{\mathbf{C}}$) are the decision problems solvable by some deterministic (nondeterministic) Turing machine in polynomial time with an oracle for the problems in **C**. The classes Δ_k^P, $\mathbf{\Sigma}_k^P$, and $\mathbf{\Pi}_k^P$ of PH are defined as follows:

$$\Delta_0^P = \mathbf{\Sigma}_0^P = \mathbf{\Pi}_0^P = \mathbf{P}$$

and for all $k \geq 0$,

$$\Delta_{k+1}^P = \mathbf{P}^{\Sigma_k^P}, \quad \mathbf{\Sigma}_{k+1}^P = \mathbf{NP}^{\Sigma_k^P}, \quad \mathbf{\Pi}_{k+1}^P = \text{co-}\mathbf{\Sigma}_{k+1}^P.$$

Note that $\mathbf{NP} = \mathbf{\Sigma}_1^P$, co-$\mathbf{NP} = \mathbf{\Pi}_1^P$, and $\Delta_2^P = \mathbf{P}^{\mathbf{NP}}$. Moreover, $\mathbf{\Sigma}_2^P = \mathbf{NP}^{\mathbf{NP}}$ and $\mathbf{\Pi}_2^P =$ co-$\mathbf{\Sigma}_2^P$. The classes co-**NP** and $\mathbf{\Pi}_2^P$ are of main interest in this paper. The "classical" co-**NP**-complete problem is to decide if a set of propositional clauses is not simultaneously satisfiable (UNSAT); the most prominent $\mathbf{\Pi}_2^P$-complete problem is to decide validity of a formula from $\text{QBF}_{2,\forall}$, the set of quantified Boolean formulas of the form $\forall \underline{x} \exists \underline{y} E(\underline{x}, \underline{y})$, where $\underline{x} =< x_1, \ldots, x_m >, \underline{y} =< y_1, \ldots, y_n >$ are tuples of propositional variables and $E(\underline{x}, \underline{y})$ is a propositional formula on $\underline{x}, \underline{y}$.[1] We say that a problem is at the k-th level of the PH if it is in Δ_{k+1}^P and is $\mathbf{\Sigma}_k^P$-hard or $\mathbf{\Pi}_2^P$-hard (hence, complete for Δ_{k+1}^P under Turing reductions). According to this, UNSAT is at the first and $\text{QBF}_{\forall,2}$ at the second level of the PH.

More recently, $\mathbf{P}^{\mathbf{NP}}$ has been refined to take into account how many **NP** oracle calls, i.e. subroutine calls to an **NP**-complete problem, are needed to solve a problem [46, 33, 27, 55]. This grasps in a certain sense "how much **NP**-hardness" is in a problem [33]. $\mathbf{P}^{\mathbf{NP}[f(n)]}$ is the set of problems decidable in polynomial time with no more than $f(n)$ calls to

[1] In fact, Quantified Propositional Formula (QPF) rather than QBF would be correct. In abuse of terminology, we do not distinguish between the isomorphic concepts of QPF and QBF.

an **NP** oracle, where n is the input size [26]. For example, for any integer k the class $\mathbf{P}^{\mathbf{NP}[k]}$ are those problems polynomially decidable with at most k **NP** oracle calls, and $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$ is the class of problems polynomially decidable with at most $O(\log n)$ **NP** oracle calls. Classes with bounded number $f(n)$ of calls to an arbitrary oracle may be defined analogous to $\mathbf{P}^{\mathbf{NP}[f(n)]}$. For example, $\mathbf{P}^{\Sigma_2^P[O(\log n)]}$ contains all problems decidable in polynomial time with at most $O(\log n)$ calls to an $\mathbf{\Sigma}_2^P$ oracle.

Intuitively, a problem solvable with one **NP** oracle call is "easier" than a problem requiring a polynomial number of **NP** oracle queries; indeed, it is known that $\mathbf{P}^{\mathbf{NP}[k]} = \mathbf{P}^{\mathbf{NP}[k+1]}$ for any $k \geq 1$ implies the collapse of the PH, which is not expected by the experts [25]. Further, any algorithm for a $\mathbf{P}^{\mathbf{NP}[f(n)]}$-hard problem that is polynomial upon subroutine calls to an **NP**-complete problem is expected to make at least $f(n)$ subroutine calls on input of length n.

3 Knowledge change and counterfactuals

We start with a review of operators for knowledge base change proposed in the literature. The proposed methods differ noticeably from one another, if we compare them according to the following three orthogonal classification criteria:

- **Irrelevance of Syntax.** In some approaches, the change consists in adding a new formula (the update) and retracting some formulas from the knowledge base T if this is necessary for preserving consistency. In this way, a formula has to be either retracted as a whole or left in the knowledge base. These approaches, referred to as *formula-based changes*, heavily depend on the syntax of T. On the other hand, *model-based changes* refer only to the extensions, i.e. models, of T. These methods respect Dalal's principle of *Irrelevance of Syntax* [7].

- **Revision versus Update.** Different change operators are appropriate according to whether the theory-change consists of a *revision*, i.e. a correction to the actual state of beliefs due to new, more precise, or more reliable information for instance, or of an *update* reflecting a change in the world described by the knowledge base. This fundamental distinction is extensively discussed in [32, 57, 58, 30]. Katsuno and Mendelzon give a nice characterization of revision versus update for model-based approaches in [30], where they set forth a set of update postulates different from revision postulates [1]. They point out that revision operators operate globally on sets of models, that is, each model of the revised knowledge base must be as close as possible to the set of all models of the initial knowledge base. One particular measure of distance is to define the "distance" between a model M and a set S of models as the minimum over the "distances" between M and any element of S; for a general definition, see [31]. On the other hand, update semantics is always "pointwise", that is, each model M of the initially given knowledge base is changed separately with minimal change, and the result of the update is the union of the updates of all those models M.

- **Measure of Closeness.** Both the formula-based approaches and the model-based approaches need a notion of "closeness". In the formula-based approaches, one compares the closeness between theories, i.e. between sets of formulas. In the model-based approaches one compares the closeness between models, i.e. between sets of atoms. Two well-known measures of closeness are the cardinality-based and the inclusion-based measure. The *cardinality-based* measure expresses the distance between two sets by the number of items in which they differ. The *inclusion-based* measure, instead, states that a set S_1 is closer to a set S than a set S_2 if the elements on which S_1 and S differ are a proper subset of the elements on which S_2 and S differ, i.e. if $S_1 \triangle S \subset S_2 \triangle S$ holds ($A \triangle B$ is the symmetric set difference $(A \cup B) - (A \cap B)$). Note that other measures of closeness are conceivable; Katsuno and Mendelzon developed a very general notion of closeness in terms of pre-orders on the sets of models [28, 31].

3.1 Formula-Based Change

If a knowledge base T is inconsistent with new knowledge p, a straightforward method to gain consistency with p is to remove formulas from the knowledge base. This approach apparently injures the principle of Irrelevance of Syntax in knowledge base change, and is thus rejected by many researchers; but it also has its defenders, cf. [19, 44], and there are scenarios which illustrate the need to differentiate in some epistemic contexts between syntactically deviating descriptions of the world [31]. Nebel points out an important relation between formula-based and model-based change, however, showing that formula-based change operators can be reconstructed as change operators acting on the deductive closure (i.e., as model-based change operators) [44].

Ginsberg's Approach:

A straight solution for knowledge base change that obeys the principle of Minimal Change is the Set-Of-Theories approach in [13, 12], which is the basis for counterfactuals in [19, 44, 45]. We refer to this approach as "Ginsberg's approach" throughout this paper. Let T be satisfiable, and let

$$W(p,T) = \{T' \subseteq T : T' \not\models \neg p,\ \ T' \subset U \subseteq T \Rightarrow U \models \neg p\}.$$

($W(p,T)$ are the "possible worlds" for p from belief in T [19].[2]) Then, define

$$T \circ_G p = \{T' \cup \{p\} : T' \in W(p,T)\}.$$

That is, $T \circ_G p$ is the family of maximal subsets $T' \subseteq T$ that are consistent with p, each of them augmented by p.

[2]Ginsberg's term "possible worlds" may be misleading since $W(p,T)$ does *not* refer to possible worlds (in the technical sense) at all but to subsets of a set of formulas.

Example 3.1 *Let* $T = \{a, b, a \wedge b \Rightarrow c\}$ *and* $p = \neg c$. *Then,* $W(p,T) = \{\{a \wedge b \Rightarrow c, b\}, \{a \wedge b \Rightarrow c, a\}, \{a, b\}\}$, *hence* $T \circ_G p = \{\{a \wedge b \Rightarrow c, b, \neg c\}, \{a \wedge b \Rightarrow c, b, \neg c\}, \{a, b, \neg c\}\}$.

A formula q is true after a change iff it is valid in all knowledge bases of $T \circ_G p$, which are considered to be independent. Thus after a single change, a formula is true iff it is valid in the logical disjunction of the knowledge bases in $T \circ_G p$.[3] For example, the formula $a \vee b$ is valid in $T \circ_G p$ from Example 3.1. Note that maximal consistent subsets of theories were used in a similar way (but in a different context) by Rescher [50, 51]. The disadvantage that incorporating p into T may result in multiple knowledge bases gave rise to the Cross-Product approach and the WIDTIO approach, which combine all resulting knowledge bases into a single one.

The Cross-Product Approach:

This approach [13] constructs a knowledge base whose formulas are, roughly speaking, all disjunctions with one formula from each knowledge base in $T \circ_G p$. Precisely, if $T \circ_G p = \{T_1, \ldots, T_m\}$, then

$$T \circ_C p = \{p_1 \vee \cdots \vee p_m : p_i \in T_i, 1 \leq i \leq m\}.$$

The difficulty with this change operator is to construct an efficient implementation (the number of formulas may grow exponentially in the size of $T \circ_G p$). It is known that $T \circ_G p \equiv T \circ_C p$; therefore we will not explicitily deal with $\circ_C$; all results derived for $\circ_G$ are also valid for $\circ_C$.

The WIDTIO Approach:

The WIDTIO (When In Doubt Throw It Out [58]) approach [19, 21] solves the multiple knowledge bases problem by the following definition:

$$T \circ_{Wid} p = \bigcap_{W \in W(p,T)} W \cup \{p\},$$

i.e. only the formulae that appear in all possible worlds are retained. This method is easy to implement in terms of the size of the revised knowledge base, but has the disadvantage that in the worst case all knowledge in T has to be retracted.

Example 3.2 *Consider* $T = \{a, b, a \wedge b \Rightarrow c\}$, $p = \neg c$ *from Example 3.1 again. Then* $T \circ_{Wid} p = \{\neg c\}$. *Note that* $a \vee b$ *is not implied by* $T \circ_{Wid} p$.

[3] Nebel [45] gives an elegant characterization of $T \circ_G p$ using deductive closure of a set of formulas.

3.2 Model-Based Change

Model-based change operators refer only to the extensions, i.e. the models of a knowledge base to gain irrelevance of syntactical knowledge representation. Model theoretic characterizations of minimal change are given in [28, 58, 30].

The considered model-based change operators rely heavily on the notion of model closeness defined by $M \triangle M$, that is the set of propositional variables whose value in M is different than in M'. For quick reference, we provide at the begin of this section a compact definition of all model-based change operators that we consider (Tables 1,2).

(Tables 1,2 go here)

More details and examples are provided in the rest of this section.

3.2.1 Revision

Alchourrón, Gärdenfors, and Makinson [1] proposed eight postulates G*1~G*8 that a change operator $\circ$, in their terms *revision* operator, should satisfy. Katsuno and Mendelzon gave the following equivalent formulation for the specialization to propositional logic and finite knowledge bases (see [28, 31] for a full discussion):

(R1) $T \circ p$ implies p.

(R2) If $T \wedge p$ is satisfiable, then $T \circ p \equiv T \wedge p$.

(R3) If p is satisfiable then $T \circ p$ is also satisfiable.

(R4) If $T_1 \equiv T_2$ and $p_1 \equiv p_2$, then $T_1 \circ p_1 \equiv T_2 \circ p_2$.

(R5) $(T \circ p_1) \wedge p_2$ implies $T \circ (p_1 \wedge p_2)$.

(R6) If $(T \circ p_1) \wedge p_2$ is satisfiable, then $T \circ (p_1 \wedge p_2)$ implies $(T \circ p_1) \wedge p_2$.

Dalal's Approach:

Dalal [6] proposes to interpret minimality of change in terms of minimal cardinality of model change. In his approach, $T \circ p$ has those models of p which have as few variables different from any model of T as possible. Define the *distance* between formulas p and q as

$$|\triangle|^{min}(p, q) = \min \{|M \triangle M'| : M \in Mod(p), M' \in Mod(q)\},$$

that is $|\triangle|^{min}(p, q)$ is the minimum number of propositional symbols in which models of p and q diverge. Dalal's revision operator $\circ_D$ can be defined as follows: If T is satisfiable, then

$$Mod(T \circ_D p) = \left\{M \in Mod(p) : \exists M' \in Mod(T) \text{ s.t. } |M \triangle M'| = |\triangle|^{min}(T, p)\right\}$$

$M \ldots$ model, $\quad p, q \ldots$ propositional formulas
$\triangle^{min}(p,q) = \min_{\subseteq}(\{M \triangle M' : M \in Mod(p), M' \in Mod(q)\})$
$\triangle_M^{min}(p) = \min_{\subseteq}(\{M \triangle M' : M' \in Mod(p)\})$
$\lvert\triangle\rvert^{min}(p,q) = \min\{\lvert M \triangle M'\rvert : M \in Mod(p), M' \in Mod(q)\}$
$\lvert\triangle\rvert_M^{min}(p) = \min\{\lvert M \triangle M'\rvert : M' \in Mod(p)\}$
$\Omega(p,q) = \left\{x : \exists D \in \triangle^{min}(p,q) \text{ s.t. } x \in D\right\}$

$\min_{\subseteq}$ selects the minimal elements under set inclusion

Table 1: Summary of functions for closeness measures

operator $\circ$	$Mod(T \circ p)^*$	
	$T \wedge p$ unsatisfiable	$T \wedge p$ satisfiable
Dalal $\circ_D$	$\left\{M \in Mod(p) : \exists M' \in Mod(T) \text{ s.t. } \lvert M \triangle M'\rvert = \lvert\triangle\rvert^{min}(T,p)\right\}$	$Mod(T \wedge p)$
Forbus $\circ_F$	$\bigcup_{M \in Mod(T)} \left\{M' \in Mod(p) : \lvert M' \triangle M\rvert = \lvert\triangle\rvert_M^{min}(p)\right\}$	
Borgida $\circ_B$	$\bigcup_{M \in Mod(T)} \left\{M' \in Mod(p) : M \triangle M' \in \triangle_M^{min}(p)\right\}$	$Mod(T \wedge p)$
Winslett $\circ_{Win}$	$\bigcup_{M \in Mod(T)} \left\{M' \in Mod(p) : M \triangle M' \in \triangle_M^{min}(p)\right\}$	
Satoh $\circ_S$	$\left\{M \in Mod(p) : \exists M' \in Mod(T) \text{ s.t. } M \triangle M' \in \triangle^{min}(p,T)\right\}$	$Mod(T \wedge p)$
Weber $\circ_{Web}$	$\{M \in Mod(p) : \exists M' \in Mod(T) \text{ s.t. } M - \Omega(T,p) = M' - \Omega(T,p)\}$	$Mod(T \wedge p)$

* $Mod(T \circ_D p) = Mod(p), Mod(T \circ_{Web} p) = Mod(T)$ if T or p is unsatisfiable

Table 2: Model-based revision operators

and if T is unsatisfiable, then $Mod(T \circ_D p) = Mod(p)$. It is well-known that $\circ_D$ satisfies (R1)$\sim$(R6) [28] (Dalal himself already noticed that $\circ_D$ satisfies the Gärdenfors postulates [6]).

Example 3.3 *Let $T = \{a, b, c\}$ and $p = (\neg a \wedge d) \vee (a \wedge \neg b \wedge \neg c \wedge \neg d)$. Note that $|\triangle|^{min}(T, p) = 1$: Since $T \wedge p$ is unsatisfiable, $|\triangle|^{min}(T, p) > 0$, but $\{a, b, c, d\} \in Mod(T)$, $\{b, c, d\} \in Mod(p)$ and $|\{a, b, c, d\} \triangle \{b, c, d\}| = 1$. It is easily verified that $T \circ_D p \equiv \neg a \wedge b \wedge c \wedge d$.*

Satoh's Approach:

According to Satoh's approach [53], the change operator selects, similar to Dalal's approach, only the models of p that are closest to T; the notion of closeness is set-inclusion instead of cardinality of model difference, however. Satoh defined his operator for the first-order case using circumscription. Let for any formulae p, q

$$\triangle^{min}(p, q) = \min_{\subseteq} (\{M \triangle M' : M \in Mod(p), M' \in Mod(q)\}),$$

that is $\triangle^{min}(p, q)$ is the family of $\subseteq$-minimal sets of differences between models of p and q. Satoh's change operator $\circ_S$ can be defined in the propositional context by

$$Mod(T \circ_S p) = \left\{M \in Mod(p) : \exists M' \in Mod(T) \text{ s.t. } M \triangle M' \in \triangle^{min}(T, p)\right\}.$$

It is well-known that if T is satisfiable, $\circ_S$ satisfies (R1)$\sim$(R5) but not (R6) [28].

Example 3.4 *Let $T = \{a, b\}$ and $p = (\neg a \wedge c) \vee (\neg a \wedge \neg b \wedge \neg c)$. Note that $T \wedge p$ is unsatisfiable. It is easily checked that $\triangle^{min}(T, p) = \{\{a\}\}$ and that $\{b, c\}$ is the only model of p selected by $\circ_S$. Therefore, $T \circ_S p \equiv \neg a \wedge b \wedge c$.*

Borgida's Approach:

The revision operator $\circ_B$ proposed by Borgida [3] uses like $\circ_S$ set inclusion instead of cardinality to interpret minimality of change. Let for a model M and a formula p be

$$\triangle_M^{min}(p) = \min_{\subseteq} (\{M \triangle M' : M' \in Mod(p)\}),$$

that is, $\triangle_m^{min}(p)$ is the family of $\subseteq$-minimal difference sets between any model of p and M. Borgida's operator can be defined as follows. If $T \wedge p$ is satisfiable, then $T \circ_B p \equiv T \wedge p$; otherwise, for every $M \in Mod(T)$ all models of p that are closest to M with respect to set inclusion are selected, that is

$$Mod(T \circ_B p) = \bigcup_{M \in Mod(T)} \left\{M' \in Mod(p) : M \triangle M' \in \triangle_M^{min}(p)\right\}.$$

If T is satisfiable, $\circ_B$ satisfies (R1)$\sim$(R5) but not (R6) [28]. Note that $\circ_B$ operates globally if $T \wedge p$ is satisfiable, but pointwise if $T \wedge p$ is not satisfiable.

Example 3.5 *Consider* $T = \{a, b\}$ *and* $p = (\neg a \wedge c) \vee (\neg a \wedge \neg b \wedge \neg c)$ *from Example 3.4 again.* T *has two models* M_1, M_2 *on variables* a, b, c*:* $M_1 = \{a, b\}$ *and* $M_2 = \{a, b, c\}$*. Then,* $\triangle_{M_1}^{min}(p) = \{\{a, c\}, \{a, b\}\}$*, and changing* M_1 *yields the two models* $\{\}$ *and* $\{b, c\}$*;* $\triangle_{M_2}^{min}(p) = \{\{a\}\}$*, and changing* M_2 *yields the model* $\{b, c\}$*. Thus* $T \circ_B p \equiv (\neg a \wedge \neg b \wedge \neg c) \vee (\neg a \wedge b \wedge c)$*. Note that* $T \circ_B p \not\equiv T \circ_S p$*.*

Weber's Approach:

Weber [56] proposes a change operation that eliminates first all variables from T that occur in any minimal model change under Satoh's semantics and then joins p to the resulting knowledge base.

Define $res_x(T) = T_x^+ \vee T_x^-$ for variable x, where T_x^+ (T_x^-) denotes the formula obtained by replacing all occurrences of x in T by *true* (*false*). For any set $U = \{u_1, \ldots, u_n\}$ of variables, define $res_U(T) = res_{u_1}(res_{u_2}(\cdots res_{u_n}(T)\cdots))$. Let for formulas p, q be $\Omega(p, q)$ the variables in $\triangle^{min}(p, q)$, i.e.

$$\Omega(p, q) \quad = \quad \Big\{x : \exists D \in \triangle^{min}(p, q) \text{ s.t. } x \in D\Big\}.$$

If T or p is unsatisfiable, then $T \circ_{Web} p$ is defined as T; otherwise, $T \circ_{Web} p = res_\Omega(T) \wedge p$, $\Omega = \Omega(T, p)$. In terms of models, $\circ_{Web}$ can be defined for satisfiable p by

$$Mod(T \circ_{Web} p) = \{M \in Mod(p) : \exists M' \in Mod(T) \text{ s.t. } M - \Omega(T, p) = M' - \Omega(T, p)\}.$$

If T and p are satisfiable, then $\circ_{Web}$ satisfies (R1)$\sim$(R4), but neither (R5) nor (R6) [31].

Example 3.6 *Consider* $T = \{a, b\}$ *and* $p = (\neg a \wedge c) \vee (\neg a \wedge \neg b \wedge \neg c)$ *from Example 3.4 for the third time. Since* $\triangle^{min}(T, p) = \{\{a\}\}$*, we get* $\Omega(T, p) = \{a\}$ *and* $res_{\{a\}}(T) = (true \wedge b) \vee (false \wedge b)$*. Thus* $T \circ_{Web} p$*, which is* $res_\Omega(T) \wedge p$*, is equivalent to* $\neg a \wedge b \wedge c$*.*

3.2.2 Update

A clear distinction between the notions of revision and update in theory and knowledge base change, as provided for databases in [32], emerged more recently [30]. *Revision* describes the incorporation of new knowledge about the unchanged world into the knowledge base, whereas *updating* means to bring the knowledge base up to date when the described world changes. Katsuno and Mendelzon set up the following postulates that any update operator $\diamond$ should obey (see [31, 30] for a full discussion):

(U1) $T \diamond p$ implies p.

(U2) If T implies p then $T \diamond p$ is equivalent to T.

(U3) If both T and p are satisfiable, then $T \diamond p$ is also satisfiable.

(U4) If $T_1 \equiv T_2$ and $p_1 \equiv p_2$, then $T_1 \diamond p_1 \equiv T_2 \diamond p_2$.

(U5) $(T \diamond p_1) \wedge p_2$ implies $T \diamond (p_1 \wedge p_2)$.

(U6) If $T \diamond p_1$ implies p_2 and $T \diamond p_2$ implies p_1, then $T \diamond p_1 \equiv T \diamond p_2$.

(U7) If T is complete, then $(T \diamond p_1) \wedge (T \diamond p_2)$ implies $T \diamond (p_1 \vee p_2)$.

(U8) $(T_1 \vee T_2) \diamond p \equiv (T_1 \diamond p) \vee (T_2 \diamond p)$.

The most important differences between update and revision are that updating means changing every model of the knowledge base separately (U8) and that an unsatisfiable knowledge base remains unsatisfiable when updated (U2). For convenience, we use $\circ$ for both kinds of knowledge base change.

Forbus's Approach:

Forbus developed a method for updating in the context of qualitative simulation [14]. As Dalal, Forbus bases change semantics on cardinality of model differences, but decides to change each model of T independently (cf. [58, p.24]). Define the *distance* $|\triangle|_M^{min}(p)$ between a model M and a formula p as

$$|\triangle|_M^{min}(p) = \min\{|M \triangle M'| : M' \in Mod(p)\},$$

i.e. as the minimum number of variables in which M and a model of p diverge. In our framework, Forbus's change operator $\circ_F$ can be defined by

$$Mod(T \circ_F p) = \bigcup_{M \in Mod(T)} \left\{ M' \in Mod(p) : \ |M' \triangle M| = |\triangle|_M^{min}(p) \right\}.$$

It is easily verified from Theorem 3.1 in [30] that $\circ_F$ satisfies (U1)$\sim$(U8).[4]

Example 3.7 *Consider $T = \{a, b, c\}$ and $p = (\neg a \wedge d) \vee (a \wedge \neg b \wedge \neg c \wedge \neg d)$ from Example 3.3 again. T has two models $M_1 = \{a, b, c\}$ and $M_2 = \{a, b, c, d\}$ on variables a, b, c, d. Then, $|\triangle|_{M_1}^{min}(p) = 2$ and changing M_1 yields the models $\{a\}$, $\{b, c, d\}$, and $|\triangle|_{M_2}^{min}(p) = 1$ and changing M_2 yields the model $\{b, c, d\}$. Therefore, $T \circ_F p \equiv (a \wedge \neg b \wedge \neg c \wedge \neg d) \vee (\neg a \wedge b \wedge c \wedge d)$. Note that $T \circ_D p \not\equiv T \circ_F p$.*

Winslett's Approach:

Winslett introduced her "Possible Models Approach" for reasoning about action in first-order calculus [57, 58]. Let for a model M and a formula p denote

$$\triangle_M^{min}(p) = \min_{\subseteq} \left(\{M \triangle M' : M' \in Mod(p)\} \right),$$

[4] [31, Example 6.2] considers an operator that operates as $\circ_F$ if $T \wedge p$ is unsatisfiable and yields $T \wedge p$ if $T \wedge p$ is satisfiable. Our complexity results for $\circ_F$ carry over to this operator analogously as the results for $\circ_{Win}$ carry over to $\circ_B$.

i.e. $\triangle_M^{min}(p)$ is the family of minimal differences between a M and modeled of p. Winslett's change operator can be defined for the propositional case by

$$Mod(T \circ_{Wid} p) = \bigcup_{M \in Mod(T)} \left\{ M' \in Mod(p) : M \triangle M' \in \triangle_M^{min}(p) \right\}.$$

The operator $\circ_{Win}$ satisfies (U1)$\sim$(U8) [30]. Note that the only case in which $\circ_{Win}$ may diverge from $\circ_B$ is if $T \wedge p$ is satisfiable: here $\circ_{Win}$ operates pointwise while $\circ_B$ operates global (hence, (U8) is not satisfied by $\circ_B$). This subtle difference, however, proved to be favorable for appropriate change semantics for reasoning about action.

Example 3.8 *Consider the classic example $T = \{a \equiv \neg b\}$ and $p = a$ [57]. Note that $T \wedge p$ is satisfiable; hence, $T \circ_B p \equiv a \wedge \neg b$. Note that T has the two models $M_1 = \{a\}$, $M_2 = \{b\}$ on variables a, b; changing M_1 yields the model $\{a\}$ ($\triangle_{M_1}^{min}(p) = \{\}$), and changing M_2 yields the model $\{a, b\}$ ($\triangle_{M_2}^{min}(p) = \{\{a\}\}$), thus $T \circ_{Win} p \equiv a$. If a means "a book is on the floor" and b "a magazine is on the floor", T states that either the book or the magazine is on the floor. If a robot is ordered to put the book on the floor, it is reasonable to accept only that a holds after this action; why should we accept that $a \wedge \neg b$ is true afterwards, i.e. the magazine is not on the floor, as suggested by revision operators satisfying* (R2)*? This indicates that Winslett's approach is more appropriate for describing change actions.*

3.3 Counterfactuals

A counterfactual is a conditional statement such as "If p, then q", where it is known or expected that p is false [34, 19].[5] The epistemic semantics for counterfactuals of our concern is based on the *Ramsey Test* [15, p. 147]:

> Accept a sentence of the form 'If A, then C' in a state of belief K if and only if the minimal change of K needed to accept A also requires accepting C.

For extensive studies of such counterfactuals refer to [19, 15]. Ginsberg [19] points out the importance of counterfactuals for a large number of areas within AI.

Technically, counterfactuals are definable under all change operators from above. Given T and change operator $\circ$, define that "If p, then q" (denoted by $p > q$) is true over T iff $T \circ p \models q$ holds, and that that $p > q$ is false over T iff $T \circ p \not\models q$ holds.

Gärdenfors tried to define a counterfactual connective based for the object language based upon a change operator subject to revision semantics, but found that this is incompatible with the Ramsey Test in all but in a certain sense "trivial" logics (Gärdenfors's Triviality Theorem, cf. [15, p.158,Theorem 7.10]). Recent investigations showed that update semantics is not subject to the Triviality Theorem and may be a more fruitful way of

[5] In a somewhat weaker notion it is sufficient that p may be false, e.g. [23]. Under the Ramsey principle, this notion matches inference from a knowledge base which has to be updated or revised under minimal change.

modeling counterfactuals [23, 29]. (Lewis's logic of conditionals [34], in which a *system of spheres* plays an important role in defining "most similar worlds", also has an update-semantics flavor.) Note that we consider counterfactuals at the meta-level, to which the triviality result does not apply.

The problem of our study is the following basic issue in counterfactual reasoning: Given a knowledge base T and formulae p, q, is $p > q$ true over T ? We name this problem COUNTERFACTUAL (CF); an index will indicate which change semantics is applied.

Example 3.9 *Consider the following example. Let r denote "It is raining", c, "It is cold" h, "I stay at home", and g, "Gorbachev is president of the United States". Assume that the knowledge base T is given by $T = \{r, \neg c, \neg g, r \wedge c \Rightarrow h\}$. Consider the counterfactual $c > h$, which means, "If it would be cold, I would stay at home". We get $W(T, c) = \{\{r, \neg g, r \wedge c \Rightarrow h\}\}$, and hence $T \circ_G c = \{\{r, \neg g, r \wedge c \Rightarrow h, c\}\}$. Clearly $T \circ_G c \models h$ holds, and therefore $c >_G h$ is true over T. On the other hand, the counterfactual "If it would be cold, Gorbachev would be president of the United States" is false over T, because $T \circ_G c \not\models g$ holds. Note that the "classical" implications $c \Rightarrow h$ and $c \Rightarrow g$ are both valid in the knowledge base T.*

Example 3.10 *Consider Winslett's example $T = \{a \equiv \neg b\}$, $p = a$ again. The counterfactual $a >_{Win} \neg b$, which means "If the robot would put the book on the floor (assuming it's not already there), then the magazine would not be on the floor", is not true over T since $T \circ_{Win} p \equiv a$. Not that $a > b$ is true over T under all semantics but $>_{Win}$ and $>_F$.*

Example 3.11 *Assume a court has to hear three independent witnesses W_1, W_2, and W_3 in a trial. The belief that W_i tells the truth is denoted by the atom t_i. The initial assumption of the court is that all witnesses are sincere, i.e. $T = \{t_1, t_2, t_3\}$. The testimonies of the witnesses are contradictory, however, such that if W_1 is honest, then W_2 and W_3 are liars, but if one of W_2, W_3 is honest, then W_1 is a liar. This amounts to the following formula p:*

$$p = (t_1 \Rightarrow \neg t_2 \wedge \neg t_3) \wedge (t_2 \vee t_3 \Rightarrow \neg t_1).$$

The court usually tends to assume that a maximum number of independent witnesses are honest, and therefore concludes that W_1 is a liar (for otherwise both W_2 and W_3 would lie). This policy is appropriately represented by a cardinality-based revision operator like Dalal's. Indeed, the counterfactual $p >_D \neg t_1$, which means "If honesty of W_1 would entail that W_2 and W_3 lie, and if honesty of W_2 or W_3 would mean that W_1 lies, then the court would belief that W_1 lies", is true over $T = \{t_1, t_2, t_3\}$. Note that $p > \neg t_1$ evaluates to false over T under all considered semantics but $>_D$ and $>_F$.

The examples show that different change semantics are appropriate in different situations, and that choosing a particular approach over another is domain dependent. It is currently unclear for which domains the consider change semantics are favorable, and further research will be necessary to clarify this issue; our concern is with their computational properties, however.

We will pay special attention to the following restrictions of CF: the case in which T is a set of Horn formulas and p, q are Horn formulas, and the case in which the change knowledge described by p involves only a few facts. Moreover, we will consider most time counterfactuals on knowledge bases that contain only atoms or negated atoms, as our hardness results will hold even under this restriction.

4 Overview and discussion of results

In this section we summarize and discuss our results and indicate possible extensions that we are currently investigating.

The main results of this paper are compactly presented in Table 3. Different methods according to the different change semantics for evaluating a counterfactual $p > q$ over a propositional knowledge base correspond to different rows in this table. Recall that a counterfactual $p > q$ over a knowledge base T is true iff $T \circ p \models q$ where "$\circ$" is the chosen change operator; in other terms, evaluating a counterfactual $p > q$ is equivalent to testing whether q could be inferred after a hypothetical knowledge base change p.

The table contains four columns. The first two columns account for the case where T is a general propositional knowledge base, while the last two columns reflect the Horn case. In the Horn case we assume that p,q and all formulae in T are conjunctions of Horn clauses. Columns one and three describe those cases in which no bound is imposed on the size $\|p\|$ of the knowledge base change (or antecedent) p, columns two and four account for those cases where the size of p is bounded by a constant.

Let us discuss the results. The first thing to notice is that in the general case with no restriction on p, for all approaches except Dalal's, the complexity of testing $p > q$ is at the second level of the polynomial hierarchy. More precisely, for most change operators the problem is $\mathbf{\Pi}_2^P$-complete and for two (WIDTIO and Weber's) it is very close to $\mathbf{\Pi}_2^P$. Under the assumption that the polynomial hierarchy does not collapse,[6] this means that counterfactual inference is harder than classical inference. On the other hand, the second level of the polynomial hierarchy is still a very low level, and, in a certain sense, much closer to the first level than to **PSPACE**.

In more practical terms, the $\mathbf{\Pi}_2^P$-hardness results state that with current methods, if we have at our disposal any subroutine for solving **NP**-complete problems, we would need (in the worst case) an exponential number of calls to this subroutine for deciding whether the counterfactual $p > q$ holds.

Loosely speaking, the $\mathbf{\Pi}_2^P$-hardness is mostly due to the interaction of two different sources of complexity. The first source is the complexity of inference in propositional logic (co-**NP**-hard) and the second source is the large number of choices that we have in general to make a knowledge base (or a model) consistent with a proposition p performing a minimal change.

[6] We will make this generally accepted assumption [25] for the rest of the discussion in this section.

COUNTERFACTUAL	GENERAL CASE		HORN FORMULAS	
$p > q$ true in T?	arbitrary p	$\|\|p\|\| \leq k$	arbitrary p	$\|\|p\|\| \leq k$
Ginsberg: $p >_G q$	$\mathbf{\Pi}_2^P$-complete		co-$\mathbf{NP}$-complete	
WIDTIO: $p >_{Wid} q$	$\mathbf{\Pi}_2^P$-hard, in $\mathbf{P}^{\mathbf{\Sigma}_2^P[O(\log n)]}$		co-$\mathbf{NP}$-hard, in $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$	
Dalal: $p >_D q$	$\mathbf{P}^{\mathbf{NP}[O(\log n)]}$-complete	$\mathbf{NP}$- and co-$\mathbf{NP}$-hard, in $\mathbf{P}^{\mathbf{NP}[O(1)]}$	$\mathbf{P}^{\mathbf{NP}[O(\log n)]}$-complete	$O(\|\|T\|\| \cdot \|\|q\|\|)$
Forbus: $p >_F q$	$\mathbf{\Pi}_2^P$-complete	co-$\mathbf{NP}$-complete	$\mathbf{\Pi}_2^P$-complete	$O(\|\|T\|\| \cdot \|\|q\|\|)$
Borgida: $p >_B q$	$\mathbf{\Pi}_2^P$-complete	$\mathbf{NP}$- and co-$\mathbf{NP}$-hard, in $\mathbf{P}^{\mathbf{NP}[2]}$	co-$\mathbf{NP}$-complete	$O(\|\|T\|\| \cdot \|\|q\|\|)$
Winslett: $p >_{Win} q$	$\mathbf{\Pi}_2^P$-complete	co-$\mathbf{NP}$-complete	co-$\mathbf{NP}$-complete	$O(\|\|T\|\| \cdot \|\|q\|\|)$
Satoh: $p >_S q$	$\mathbf{\Pi}_2^P$-complete	$\mathbf{NP}$- and co-$\mathbf{NP}$-hard, in $\mathbf{P}^{\mathbf{NP}[O(1)]}$	co-$\mathbf{NP}$-complete	$O(\|\|T\|\| \cdot \|\|q\|\|)$
Weber: $p >_{Web} q$	$\mathbf{\Pi}_2^P$-hard, in $\mathbf{P}^{\mathbf{\Sigma}_2^P[O(\log n)]}$	$\mathbf{NP}$- and co-$\mathbf{NP}$-hard, in $\mathbf{P}^{\mathbf{NP}[O(1)]}$	co-$\mathbf{NP}$-hard, in $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$	$O(\|\|T\|\| \cdot \|\|q\|\|)$

Table 3: Complexity results for propositional counterfactual evaluation

The only exception in the first column of our table is Dalal's approach, which is complete for the class $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$. This means that the we can decide $p > q$ with $O(\log n)$ calls to an **NP** oracle, where n stands for the size of T plus p plus q. As will become clear from the proof in Section 6, the deep reason for this relative easiness is that in Dalal's approach the two sources of complexity are not intermingled and can be separated from one another.

Interestingly, the syntax-based approaches remain $\mathbf{\Pi}_2^P$-hard even if the size of p is bounded by a constant k. This insensitivity towards the size of the knowledge change p can be intuitively explained as follows: even a small change may be have drastic effects by forcing some originally unessential part of a formula in T to become true. For instance, if T contains two formulas $f_1 = \neg u \vee (x_1 \wedge x_2 \wedge \ldots \wedge x_n)$, $f_2 = \neg u$ and if p is equal to u, then p forces the originally completely irrelevant conjunction $(x_1 \wedge x_2 \wedge \ldots \wedge x_n)$ to become true.

On the other hand, the complexity of the model-based approaches decreases when the size of p is bounded by a constant. The intuitive reason for this is that a small change p can affect only few literals; since in this case there are only few alternatives for changing a given model, the second source of complexity disappears.

In the Horn case, the complexity of all approaches except Forbus's is at the first level of the polynomial hierarchy, more precisely, co-**NP**-equivalent. Intuitively, this lowering of complexity is due to the fact that with Horn clauses the inference problem becomes polynomial, and hence the first of the aforementioned two sources of complexity is eliminated. The cardinality based approaches of Dalal and Forbus are the only ones whose complexity does not decrease under the Horn restriction. This is because their cardinality-based measure of closeness between models has a detrimental effect on the benign properties of Horn clauses.

Let us finally consider the case where both restrictions, Horn formulas and bounded p, are combined. The complexity results corresponding to this case are shown in the fourth column of our table. As in the general case, bounding the size of p does not improve the complexity of the formula-based methods. The model-based methods, however, all become polynomial under these restrictions and allow counterfactual evaluation even in $O(\|T\| \cdot \|q\|)$ time in this case.

Note that the theorems presented in the following sections often state yet more acute results than the ones shown in Table 3. For instance, we show that most hardness and completeness results hold even in case T is literal base, i.e. T consists of a set (or conjunction) of atoms or negated atoms.

All results presented in Table 3 are new except the $\mathbf{\Pi}_2^P$-completeness of Ginsberg's approach for the general case. The latter has recently been shown by Nebel in [45]. Note that in Nebel's proof it is assumed that the given knowledge base T is unsatisfiable. This assumption is not reasonable in the context of counterfactuals [19]. Our proof is completely different and does not make this unsatisfiability assumption. We note, however, that Nebel's proof could be easily transformed using Corollary 1 in [45] into a proof that works also for a satisfiable T. Nebel's proof, on the other hand, does not generalize to most other approaches (in particular, the model based approaches require a different proof

technique).

An important extension that we do not deal with explicitly is the possibility of *protecting* formulas in T. The set $I \subseteq T$ of protected formulas of T can be viewed as integrity constraints that must not be violated by any change. It is thus required that the changed knowledge base $T \circ p$ satisfies I, i.e. $T \circ p \models I$. Note that in the case where p is not restricted, this can always be enforced by "adding" I to the change p, i.e. computing $T \circ (p \wedge I)$. Hence in such cases, the possibility of protecting formulas does not alter our complexity results. It is also quite easy to see that even in the case where $\|p\|$ is bounded by a constant the worst case complexity of CF is not altered if protection of formulas is allowed. For these reasons, we do not treat this feature as a special case in this paper.

Nebel has also studied in [45] a restriction for the syntax based methods that we do not consider here: disambiguating the knowledge base change by attaching different priorities to the sentences in T. Clearly, such a policy eliminates the second source of complexity, and therefore the complexity level will be lowered by one in general. Note that this prioritization of sentences cannot be applied in the model-based cases. Here we could use a total order on the propositional variables (similar in some sense to the partial order used in prioritized circumscription [35, 38]) indicating which variable should change its value first if an inconsistency has to be removed. We leave a detailed study of such restrictions for the future.

Another very interesting issue are compound counterfactuals and sequences of knowledge change. A counterfactual of the form $p > (q > r)$ corresponds to the question "if p would hold, would q counterfactually imply r ?". Some recent preliminary results of the authors indicate that as long as we use nested sequences of counterfactuals, the problem remains $\mathbf{\Pi}_2^P$-complete for most approaches. However, if we mix counterfactual implication ($>$) with classical connectives (such as negation) then the problem becomes **PSPACE**-complete. Independently, Grahne and Mendelzon obtained a closely related **PSPACE**-completeness result in [24], where, loosely speaking, the complexity of evaluating formulas built with classical connectives and a counterfactual connective that adheres to the $\circ_{Win}$ semantics on a set of models is considered. Their results entail also that, in our terms, evaluating in this setting a nested counterfactual mixed with classical connectives over a knowledge base is **PSPACE**-complete.

5 Related results and further consequences

Counterfactuals are just one form of nonmonotonic reasoning, and their relationship to other well-known nonmonotonic reasoning mechanisms auch as circumscription, default logic, and truth maintenance systems has already been investigated by Ginsberg [19]. It is interesting to observe that similar complexity results have recently been derived for these approaches, which locate reasoning problems in all these formalism at the second level of the polynomial hierarchy. In particular, the following results have been found.

The nonmonotonic inference techniques of *closed world reasoning* are widely used in Artifi-

cial Intelligence, Database Theory, and Logic Programming [2]. Starting with the (naive) Closed World Assumption (*CWA*) introduced by Reiter [48], several formalizations of closed world reasoning have been developed: the Generalized Closed World Assumption (*GCWA*) by Minker [41], the Extended Generalized Closed World Assumption (*EGCWA*) by Yahya and Henschen [60], the Careful Closed World Assumption (*CCWA*) by Gelfond and Przymusinska [18], and the Extended Closed World Assumption (*ECWA*) by Gelfond, Przymusinska, and Przymusinski [17]. Circumscription was introduced by McCarthy in [37]. It is known that circumscription coincides with the *ECWA* in the case of propositional logic [17].

[10] deals with the complexity of closed world reasoning applied to general propositional theories and to unrestricted clause-sets. In particular, the deduction problem is studied, i.e., given formulae F and G, does G follow from F under a certain closed-world inference rule. It is shown that the deduction problems under the *GCWA*, the *EGCWA*, the *CCWA*, the *ECWA* are all at the second level of PH. In particular, the deduction problem both under the *EGCWA* and the *ECWA* (= propositional circumscription) is shown $\mathbf{\Pi}_2^P$-complete, and $\mathbf{\Pi}_2^P$-hardness is demonstrated for the restriction that F is in clause form with at most three literals per clause and G is a literal. This is proved to be a limit case for $\mathbf{\Pi}_2^P$-completeness, since if F has at most 2 literals per clause, the deduction problem is co-**NP**-complete for both *EGCWA* and *ECWA* (co-**NP**-hardness is shown in [5]).

For the *GCWA* and the *CCWA*, upper bounds are provided by showing that the inference problem can be solved with $O(\log m)$ calls to a $\mathbf{\Sigma}_2^P$-oracle, where m is the number of propositional variables in the formula F. It is also shown testing whether the closure under *CWA* of a formula F is consistent (a co-**NP**-hard problem, but most likely not in co-**NP**) can be done with $O(\log m)$ calls to an **NP**-oracle, where m is the number of propositional variables in F.

The results in [10] can be interpreted as follows: If the polynomial hierarchy does not collapse, then, for all considered closed world reasoning principles, the deduction problem is strictly harder than the deduction problem in classical propositional logic. In particular, depending on the chosen closed world principle except for the naive *CWA*, the deduction problem is either complete for $\mathbf{\Pi}_2^P$ or only "mildly" harder than $\mathbf{\Pi}_2^P$. The deduction problem under the *CWA* is complete for Δ_2^P under Turing reductions, and thus not much harder than the deduction problem in the classical propositional logic. It is harder, however, since it is shown that this problem is neither in **NP** nor in co-**NP** unless the polynomial hierarchy collapses.

The complexity of several reasoning tasks in *nonmonotonic propositional logics* is investigated in [22]. The systems of logics considered are Reiter's default logic [49], McDermott's and Doyle's nonmonotonic logic [40, 39], Moore's autoepistemic logic [43], and Marek's and Truszczyński's nonmonotonic logic N [36]. All these logics have in common that the semantics of an initially given set of premises is explained through a corresponding set of fixed points (also called extensions or expansions). Given a set of premises, the main reasoning tasks in these logics are (*i*) testing for existence of a fixed point; (*ii*) deciding whether a given formula belongs to at least one fixed point (brave reasoning); and (*iii*) deciding whether a given formula belongs to all fixed points (cautious reasoning). It is

shown that for all mentioned logics, problems (i), (ii) are complete for $\mathbf{\Sigma}_2^P$, while problem (iii) is complete for $\mathbf{\Pi}_2^P$.

Similar complexity results on abductive reasoning are given in [11], while in [52] it is shown that some special reasoning task in a generalization of De Kleer's assumption-based truth management system (ATMS) [8] is complete for $\mathbf{\Sigma}_2^P$.

In summary, it appears that the complexity of almost all approaches to nonmonotonic propositional reasoning is located at the second level of the polynomial hierarchy. As a practical consequence, most formalizations of nonmonotonic reasoning can be polynomially translated into one another, and existing algorithms or proof procedures for one approach can be fruitfully applied to the other.

For instance, $\mathbf{\Pi}_2^P$-completeness of Winslett's method and cautious default reasoning guarantees that the problem of evaluating a counterfactual $p >_{Win} q$ can be polynomially transformed into an equivalent problem of brave reasoning in default logic. Thus, one can take profit of existing sophisticated procedures for default reasoning and of implemented default theorem-provers to solve counterfactuals. Similarly, by the established complexity results, one can polynomially translate most decision problems related to theory update or revision under various approaches to the problem of inferencing under (propositional) circumscription. Practical methods for inferencing under circumscription have been developed by [4, 20, 47]. Note that in the more general context of first-order logic, Winslett already pointed out the translatability of theory-update problems to circumscription [59].

6 Computational complexity: the general case

We consider CF in the most general case first, where p, q, and all formulas in T are arbitrary propositional formulas. We will see, however, that the worst case complexity of CF still occurs even if the knowledge base T contains only facts, that is, atomic chunks of knowledge.

Prior to complexity considerations, we make several observations on changing literal bases under various change semantics. If T is a literal base, then all knowledge in the knowledge base is represented by facts and negated facts, and if all facts of interest are considered (i.e. if the literal base is complete), the knowledge base is a unique description of the world. Certain change operators from above behave identical when applied to a literal base; several noticeable relationships in this context are captured by the following Lemma.

Lemma 6.1 *Let T be a satisfiable literal base, and let p, q be propositional formulas.*

(1) *$T \circ_G p \equiv T \circ_S p$ and $p >_G q$ is true over T iff $p >_S q$ is true over T.*

(2) *If $V(p) \subseteq V(T)$, then $T \circ_S p \equiv T \circ_B p \equiv T \circ_{Win} p$, and the validity of $p >_S q$, $p >_B q$, and $p >_{Win} q$ over T is equivalent.*

(3) *If* $V(p) \subseteq V(T)$, *then* $T \circ_D p \equiv T \circ_F p$, *and* $p >_D q$ *is true over* T *iff* $p >_F q$ *is true over* T.

(4) *If* p *is satisfiable, then* $T \circ_{Wid} p \equiv T \circ_{Web} p$, *and* $p >_{Wid} q$ *is true over* T *iff* $p >_{Web} q$ *is true over* T.

(5) *Let* $q \in T$, *and let* p *be satisfiable. Then the validity of* $p >_G q$, $p >_{Wid} q$, $p >_S q$, *and* $p >_{Web} q$ *over* T *is equivalent, and* $q \in T \circ_{Wid} p$ *iff* $p >_G q$ *is true over* T.

Proof: The equivalence of counterfactual validity under different change semantics is trivial if the knowledge bases are logically equivalent after change; thus for (1)–(4) it suffices to consider the equivalence of the change operators.

The definition of $\circ_S$ entails that each $D \in \triangle^{min}(T, p)$ satisfies $D \subseteq V(T) \cap V(p)$, i.e. the models of T change merely on the common variables of T and p. It is obvious that either $W \wedge p \models f$ or $W \wedge p \models \neg f$ holds, for every $W \in W(p, T)$ and $f \in T$. Therefore, if T is a literal base, maximality of W in Ginsberg's approach is equivalent to Satoh's minimality of truth value change of the variables in $V(T) \cap V(p)$. Thus (1), i.e. $Mod(T \circ_G p) = Mod(T \circ_S p)$, holds.

Next we verify (2) and (3). Under change operation $\circ \in \{\circ_B, \circ_{Win}, \circ_F\}$, the models of T change only on the variables $V(p)$ to become models of $T \circ p$, but for $\circ \in \{\circ_S, \circ_D\}$, changes restrict to $V(p) \cap V(T)$. Since $V(p) \subseteq V(T)$ and T has only one model on $V(T)$, however, all stated equivalences must obviously hold (note: if in this case $T \wedge p$ is satisfiable, $Mod(T \circ_B p) = Mod(T) = Mod(T \circ_{Win} p)$).

(4) and (5) are verified as follows. Consider any formula f in the literal base T. From the definition of $\circ_G$ follows that $p >_G f$ is true over T iff $f \in W$, for all $W \in W(p, T)$, which holds by the definition of $\circ_{Wid}$ iff $f \in T \circ_{Wid} p$; an easy consideration is that the latter is true iff $p >_{Wid} f$ is true over T. Thus $p >_G f$ is equivalent to $p >_{Wid} f$ over T; applying (1) we conclude that $f \in T \circ_{Wid} p$ iff $p >_S f$ is true over T. As T is a literal base, it is straightforward that the variable of the literal f is in $\Omega(T, p)$ exactly if $p >_S f$ is false over T. As f is arbitrary from T, it follows that $\Omega(T, p) = T - (T \circ_{Wid} p)$, and by the model characterization of $T \circ_{Web} p$ (T and p are assumed satisfiable) we get easily $Mod(T \circ_{Web} p) = Mod(T \circ_{Wid} p)$; thus (4) is true, and hence (5) is clearly also true. □

This interrelation of various change semantics will be heavily exploited to derive several complexity results with few proofs. For example, Lemma 6.1 is useful for one single proof of $\mathbf{\Pi}_2^P$-hardness of CF under all change operators in Section 3 but Forbus's and Dalal's. In fact, we show the following.

Lemma 6.2 *Let* T *be an atomic knowledge base. To decide if* $p > q$ *is true over* T *is* $\mathbf{\Pi}_2^P$*-hard for* $> \in \{>_G, >_{Wid}, >_B, >_{Win}, >_S, >_{Web}\}$, *even if* q *is an atom with* $q \in T$.

Proof: Given an instance $F = \forall \underline{x} \exists \underline{y} E(\underline{x}, \underline{y})$ of $\mathrm{QBF}_{2,\forall}$, we construct in polynomial time an atomic knowledge base T and formulae p, q such that q is an atom from T and $p >_G q$

is true over T iff F is valid; since $V(p) \subseteq V(T)$ holds, the claim will be easy to prove with Lemma 6.1.

Let $\underline{x} = <x_1, \ldots, x_m>$ and $\underline{y} = <y_1, \ldots, y_n>$. Let $z_1, \ldots, z_m$ and r be additional propositional variables. We define T, p, and q as follows:

$$\begin{aligned} T &= \{x_1, \ldots, x_m, z_1, \ldots, z_m, y_1, \ldots y_n, r\} \qquad (1) \\ p &= [(x_1 \equiv \neg z_1) \wedge \ldots \wedge (x_m \equiv \neg z_m) \ \wedge \ (r \Rightarrow E(\underline{x}, \underline{y})) \ \wedge \ ((y_1 \vee \ldots \vee y_n) \Rightarrow r)] \\ q &= r \end{aligned}$$

Note that T and p are satisfiable, but $T \wedge p$ is unsatisfiable. The construction of T, p, q is obviously feasible in time polynomial in the size of F.[7]

Before proving that F is valid iff $p >_G q$ is true over T, let us say a few words about the spirit behind our transformation.

Recall that each "possible world" is a maximal subset of $T' \subseteq T$ such that $T' \wedge p$ is satisfiable. It is easy to see that the counterfactual $p >_G q$ is true over T iff r belongs to each "possible world" in $W(p, T)$. The conjuncts of p serve for the following purposes:

- The conjuncts $x_i \equiv \neg z_i$ make sure that each $W \in W(p, T)$ contains at most one atom from $\{x_i, z_i\}$. Moreover, by the maximality requirement, W contains exactly one atom from $\{x_i, z_i\}$. Thus, $W_1, W_2 \in W(p, T)$ differing in their x_i elements are incomparable with respect to set inclusion.
- The conjunct $r \Rightarrow E(\underline{x}, \underline{y})$ makes sure that for each $W \in W(p, T)$ such that $r \in W$, $W \wedge p \models E(\underline{x}, \underline{y})$ holds.
- The conjunct $(y_1 \vee \cdots \vee y_n) \Rightarrow r$ enforces that for each $W \in W(p, T)$ with $r \notin W$, $W \subseteq \{x_1, z_1, \ldots, x_m, z_m\}$ holds.

Let us now show that F is valid iff $p >_G q$ is true over T.

First observe that for each $W \in W(p, T)$, either $x_i \in W$ or $z_i \in W$, $1 \leq i \leq m$. Indeed, assume that $\{x_i, z_i\} \cap W = \emptyset$ for some i. Then, since $W \wedge p$ is satisfiable, $M \in \mathit{Mod}(W \wedge p)$ exists such that $M \models x_i$ or $M \models z_i$. Thus $W \wedge x_i \wedge p$ or $W \wedge z_i \wedge p$ is satisfiable, and hence $W \notin W(p, T)$, contradiction. Note that clearly at most one of x_i, z_i is contained in W.

It thus holds that every $W \in W(p, T)$ "fixes" together with p a unique truth assignment ϕ_W to $\underline{x}$ by $\phi_W(x_i) = \mathit{true}$ if $x_i \in W$ and $\phi_W(x_i) = \mathit{false}$ if $z_i \in W$, for $1 \leq i \leq m$. Conversely, for any truth assignment ϕ to $\underline{x}$, the set

$$W_\phi = \{x_i : \phi(x_i) = \mathit{true}, 1 \leq i \leq m\} \cup \{z_i : \phi(x_i) = \mathit{false}, 1 \leq i \leq m\}$$

is consistent with p (just assign *false* to $y_1, \ldots, y_n$,r), and $W_\phi \wedge p$ fixes ϕ; it follows immediately that for any ϕ there exists a $W \in W(p, T)$ such that $W \wedge p$ fixes ϕ.

[7] Note that a straightforward simpler construction for $>_G$ obtained by removal of all y_i from T and cancellation of $y_1 \vee \cdots \vee y_n \Rightarrow r$ in p does not allow application of Lemma 6.1.

Now assume that $p >_G q$ is true over T. Then for all $W \in W(p,T)$ we have $W \wedge p \models r$ and thus clearly $r \in W$. Consequently, for every truth assignment ϕ to $\underline{x}$, there is some $W \in W(p,T)$ such that $W \supset W_\phi$ (hence $\phi_W = \phi$) and $r \in W$. Since $W \wedge p$ is satisfiable and $W \wedge p \models r \Rightarrow E(\phi(\underline{x}), \underline{y})$, this implies that some truth assignment ψ to $\underline{y}$ exists such that $E(\phi(\underline{x}), \psi(\underline{y}))$ is true. Thus F is valid.

Conversely, assume F is valid, but $p >_G q$ is false over T. Then $W \in W(p,T)$ exists such that $W \wedge p \not\models r$, which means $r \notin W$. By the conjunct $(y_1 \vee \cdots \vee y_n) \Rightarrow r$ of p, it is immediately clear that $y_i \notin W$, for $1 \leq i \leq n$. Consequently, $W = W_\phi$ for some truth value assignment ϕ to $\underline{x}$. Since F is valid, $E(\phi(\underline{x}), \underline{y})$ is satisfiable, hence clearly $W \wedge p \wedge r$ is satisfiable. Thus $W \notin W(p,T)$, since W is not a maximal subset of T consistent with p. Contradiction. Therefore, if F is valid, $p >_G q$ is true over T, and the claim is proved.

Thus the lemma holds for $>_G$ and by (1) in Lemma 6.1 also for $>_S$. Since $V(p) \subseteq V(T)$ and $q \in T$, it is by (2) and (5) in Lemma 6.1 further true for $> \in \{>_B, >_{Win}, >_{Wid}, >_{Web}\}$. □

We have not considered the cardinality-based change operator by Forbus yet. $\mathbf{\Pi}_2^P$-hardness of CF_F does not follow from construction (1) (this would imply by (3) in Lemma 6.1 $\mathbf{\Pi}_2^P$-hardness of CF_D, which is in Δ_2^P), but is not difficult to show.

Lemma 6.3 *Let T be an atomic knowledge base. To test if $p >_F q$ is true over T is $\mathbf{\Pi}_2^P$-hard.*

Proof: Let $F = \forall \underline{x} \exists \underline{y} E(\underline{x}, \underline{y})$, where $\underline{x} = <x_1, \ldots, x_m>$, $\underline{y} = <y_1, \ldots, y_n>$. Let $z_1, \ldots, z_n$ and r be new propositional variables. Define

$$\begin{array}{rcl} T & = & \{y_1, \ldots, y_n, z_1, \ldots, z_n, r\} \\ p & = & [(y_1 \equiv \neg z_1) \wedge \ldots \wedge (y_n \equiv \neg z_n) \ \wedge \ (r \Rightarrow E(\underline{x}, \underline{y}))] \\ q & = & r \end{array} \tag{2}$$

This transformation is clearly possible in polynomial time. Notice that every model $M \in Mod(T)$ determines a truth assignment ϕ_M to $\underline{x}$, and that every possible truth assignment to $\underline{x}$ is extendible to a model of T. It is not difficult to see that for any model $M \in Mod(T)$, it holds that $n \leq |\triangle|_M^{min}(p) \leq n+1$: The truth value of either y_i or z_i has to be changed to *false*, and if the resulting assignment to the y_i together with ϕ_M does not satisfy $E(\underline{x}, \underline{y})$, just r has to be switched to *false* to get a model of p. Notice that if the truth value of any x_i is changed, then at least $n+1$ changes in total are necessary. Thus we conclude that for ϕ_M there exists a truth assignment ψ to $y_1, \ldots, y_n$ such that $E(\phi_M(\underline{x}), \psi(\underline{y}))$ is true iff $|\triangle|_M^{min}(p) = n$ iff $M' \models r$ for all models $M' \in Mod(p)$ with $|M \triangle M'| = |\triangle|_M^{min}(p)$. □

We thus get the following result.

Theorem 6.4 *Deciding if a counterfactual $p > q$ is true over a knowledge base T is $\mathbf{\Pi}_2^P$-complete for $> \in \{>_G, >_{Win}, >_B, >_S, >_F\}$, and $\mathbf{\Pi}_2^P$-hardness holds even if T is atomic and q is an atom with $q \in T$.*

Proof: By Lemmas 6.2 and 6.3 it remains to show membership in $\mathbf{\Pi}_2^P$. We outline how to disprove nondeterministically $p > q$ in polynomial time with a **NP** oracle, without treatment of special cases (e.g., T, p unsatisfiable).

- $>_G$: Guess $T' \subseteq T$. Call for every $f \in T - T'$ the oracle whether $T' \cup \{f, p\}$ is unsatisfiable. Call the oracle to check that $T' \wedge p \not\models q$.
- $>_{Win}$: Guess $M \in Mod(T)$ and $M' \in Mod(p)$. Test whether $M' \models \neg q$ holds. Call the oracle whether there is no $M'' \in Mod(p)$ such that $M \triangle M'' \subset M \triangle M'$.
- $>_B$: Test by oracle call whether $T \wedge p$ is satisfiable; if this is true, check $T \wedge p \not\models q$ by oracle call, else proceed as for $>_{Win}$.
- $>_S$: Guess $M \in Mod(T)$ and $M' \in Mod(p)$. Test whether $M' \models \neg q$ holds. Call the oracle whether there exist no $M \in Mod(T), M'' \in Mod(p)$ such that $M \triangle M'' \subset M \triangle M'$.
- $>_F$: Guess $M \in Mod(T)$ and $M' \in Mod(p)$. Test whether $M' \models \neg q$ holds. Call the oracle whether there is no $M'' \in Mod(p)$ such that $|M \triangle M''| < |M \triangle M'|$.

□

Note that $>_{Wid}$ and $>_{Web}$ do not appear in Theorem 6.4. In fact, it seems to be difficult to show membership of CF_{Wid}, CF_{Web} in $\mathbf{\Pi}_2^P$. An observation supporting this intuition is the complexity of verifying a guess for $T \circ_{Wid} p$ or, in case of $\circ_{Web}$, $\Omega(T, p)$; an easy corollary to Lemma 6.2 is that this problem is $\mathbf{\Pi}_2^P$-hard, hence a polynomial time algorithm is impossible even if an **NP** oracle is provided, unless $\mathbf{\Pi}_2^P = \Delta_2^P$. A guess for $W \in W(p, T)$ in Ginsberg's approach or for $D \in \triangle^{min}(T, p)$ in Satoh's approach is, in contrast to this, verifiable in polynomial time with a **NP** oracle.

Closer inspection yields the – yet not surprising – result that CF_{Wid} and CF_{Web} are computationally equivalent, at least for satisfiable knowledge changes. (Note that this has been shown for literal bases in Lemma 6.1 already.)

Theorem 6.5 *Let $T = \{f_1, \ldots, f_m\}$ be a satisfiable knowledge base and let p, q be propositional formulae where p is satisfiable. Then, the problem "$p >_{Wid} q$ true over T?" is equivalent to "$p >_{Web} q$ true over T?" under polynomial time transformations.*

Proof: To transform CF_{Wid} polynomially into CF_{Web}, we define a knowledge base T' and a formula p' as follows. Let $s_1, \ldots, s_m$ be new propositional variables. Define

$$T' = \{s_1, \ldots, s_m\}, \quad p' = p \wedge [(s_1 \Rightarrow f_1) \wedge \cdots \wedge (s_m \Rightarrow f_m)] \tag{3}$$

Verify that $\{s_{i_1}, \ldots, s_{i_l}\} \in W(p', T')$ iff $\{f_{i_1}, \ldots, f_{i_l}\} \in W(p, T)$, for any $\{i_1, \ldots, i_l\} \subseteq \{1, \ldots, m\}$. Consequently, by (4) in Lemma 6.1 (T', p' are satisfiable) we get immediately that $p >_{Wid} q$ is true over T iff $p' >_{Wid} q$ is true over T' iff $p' >_{Web} q$ is true over T'.

To transform CF_{Web} polynomially into CF_{Wid}, set $X = \{x_1, \ldots, x_n\} = V(T) \cap V(p)$. Let $y_1, \ldots, y_n$ and $z_1, \ldots, z_n$ be new propositional variables. Let T_1 be the knowledge base obtained by replacing in T each x_i by y_i, for all i. Define

$$T' = \{z_1, \ldots, z_n\}, \quad p' = T_1 \wedge p \wedge [(z_1 \Rightarrow (x_1 \equiv y_1)) \wedge \cdots \wedge (z_n \Rightarrow (x_n \equiv y_n))]. \qquad (4)$$

As T' is atomic and satisfiable, and as p' is satisfiable, by (4) in Lemma 6.1 we immediately obtain that $p' >_{Web} q$ is true over T' iff $p' >_{Wid} q$ is true over T'. Every $W \in W(p', T')$ corresponds one to one to a $D_W \in \triangle^{min}(T, p)$ by $D_W = \{x_i : z_i \notin W, 1 \leq i \leq n\}$. This implies that $x_i \notin \Omega(T, p)$ iff $z_i \in T' \circ_{Wid} p'$, for $1 \leq i \leq n$. One can infer that the models of $T' \circ_{Wid} p'$ are, restricted to $V(T \wedge p \wedge q)$, equivalent to the models of $T \circ_{Web} p$; hence, $p >_{Web} q$ is true over T iff $p' >_{Wid} q$ is true over T'. $\Box$

Deciding whether a formula $f \in T$ is not in $T \circ_{Wid} p$ is in $\mathbf{\Sigma}_2^P$, as well as deciding whether a variable x is in $\Omega(T, p)$ (note that $x \in \Omega(T, p)$ implies $x \in V(T) \cap V(p)$). Therefore, CF_{Wid} and CF_{Web} are solvable in polynomial time with a $\mathbf{\Sigma}_2^P$ oracle as follows:

- $>_{Wid}$: First compute $T \circ_{Wid} p$, eliminating from T each formula $f \in T$ that does not appear in all $W \in W(p, T)$, which is checked with the $\mathbf{\Sigma}_2^P$ oracle, and add p subsequently. Then, check with the oracle whether $T \circ_{Wid} p \models q$ holds.
- $p >_{Web} q$: Check whether T or p is unsatisfiable; if so, check $T \models q$. Otherwise, compute $\Omega(T, p)$, checking for each $x \in V(T) \cap V(p)$ with the $\mathbf{\Sigma}_2^P$ oracle whether $x \in D$ for some $D \in \triangle^{min}(T, p)$. Then, check with the oracle whether there exist no $M \in Mod(T), M' \in Mod(p)$ such that $M - \Omega(T, p) = M' - \Omega(T, p)$ and $M' \models \neg q$.

By Lemma 6.2, we have thus the following result.

Theorem 6.6 CF_{Wid} *and* CF_{Web} *are equivalent to the* $\mathbf{\Sigma}_2^P$*-complete problems under Turing transformations, and thus have roughly the same complexity as those problems.*

In other terms, CF_{Wid} and CF_{Web} are under Turing transformations complete for $\mathbf{P}^{\Sigma_2^P}$, which is the class of decision problems solvable in polynomial time with a $\mathbf{\Sigma}_2^P$ oracle. This characterization does not account for the number of necessary calls to an $\mathbf{\Sigma}_2^P$ oracle, however. The algorithms from above require about m such calls, where $m = |T|$ and $m = |V(T) \cap V(p)|$, respectively. This straightforward upper bound can be improved upon substantially.

Theorem 6.7 *Let m be the number of formulas in a knowledge base T (the number of variables in $V(T) \cap V(p)$). Then* CF_{Wid} *(*CF_{Web}*) is polynomially decidable with $O(\log m)$ calls to an* $\mathbf{\Sigma}_2^P$ *oracle.*

Proof: The basic idea for such an algorithm is to get certain information about $T \circ p$ that is computable with $O(\log m)$ calls to the $\mathbf{\Sigma}_2^P$ oracle and enables verification of a guess for

a model of $T \circ p$ with a Δ_2^P algorithm. If this information is available, one oracle call will suffice to decide $p > q$. In our case, the information is the cardinality of $T \circ_{Wid} p$ and the cardinality of $\Omega(T, p)$, respectively.

We outline an algorithm for CF_{Wid} as claimed, where we assume for sake of simplicity that $p \not\in T$. $|T \circ_{Wid} p|$ is computable by binary search with $O(\log m)$ $\mathbf{\Sigma}_2^P$ oracle calls, since the following problem is in $\mathbf{\Sigma}_2^P$: Given T, p and an integer k, is $|T \circ_{Wid} p| \leq k$? This is true iff at least $l = |T| - (k-1)$ distinct formulae $f_1, \ldots, f_l$ are removed from T. This is decidable as follows: Guess $f_1, \ldots, f_l \in T$ and "witnesses" $W_1, \ldots, W_l \subseteq T$ for the guess of the f_i. W_i is a witness that f_i is removed from T if $W_i \in W(p, T)$ and $f_i \not\in W_i$; this is verifiable in polynomial time with an $\mathbf{NP}$ oracle. Thus the whole procedure sketches an $\mathbf{\Sigma}_2^P$ algorithm for $|T \circ_{Wid} p| \leq k$?. If $|T \circ_{Wid} p|$ is known, call the $\mathbf{\Sigma}_2^P$ oracle whether there exist $f_1, \ldots, f_l \in T$, $W_1, \ldots, W_l \subseteq T$ $(l = |T| - (|T \circ_{Wid} p| - 1))$ and a model M such that the f_i are pairwise distinct, W_i is a witness that f_i is removed from T, and $M \models (T - \{f_1, \ldots, f_l\}) \wedge p \wedge \neg q$. The answer to this call is "no" iff $p >_{Wid} q$ is true over T.

An algorithm for CF_{Web} can be defined in an analogous manner. The case in which T or p is unsatisfiable is treated as above. Concerning the other case, $|\Omega(T, p)|$ is computable by binary search with $O(\log m)$ calls of a $\mathbf{\Sigma}_2^P$ oracle, as deciding $|\Omega(T, p)| \geq k$? is in $\mathbf{\Sigma}_2^P$: a guess for sets $D_1, \ldots, D_k \in \triangle^{min}(T, p)$ such that $|D_1 \cup \cdots \cup D_k| \geq k$ can be verified with a $\mathbf{NP}$ oracle in polynomial time. If $|\Omega(T, p)|$ is known, check with the $\mathbf{\Sigma}_2^P$ oracle whether $D_1, \ldots, D_{|\Omega|} \in \triangle^{min}(T, p)$ and $M \in Mod(T)$, $M' \in Mod(p)$ exist such that $|D| = |\Omega(T, p)|$, $D = D_1 \cup \cdots \cup D_{|\Omega|}$, $M - D = M' - D$, and $M' \models \neg q$ holds. The oracle answers "no" iff $p >_{Web} q$ is true over T. Consequently, $p >_{Web} q$ is polynomially decidable with $O(\log m)$ calls to a $\mathbf{\Sigma}_2^P$ oracle. $\Box$

Recall that $\mathbf{P}^{\Sigma_2^P[O(\log n)]}$ is the class of problems decidable in polynomial time with $O(\log n)$ calls to a $\mathbf{\Sigma}_2^P$ oracle, where n is the input size. Thus we have the following result.

Corollary 6.8 CF_{Wid} *and* CF_{Web} *are in* $\mathbf{P}^{\Sigma_2^P[O(\log n)]}$.

Last but not least we consider Dalal's operator $\circ_D$, for which we provide the following already announced result.

Theorem 6.9 *To decide if* $p >_D q$ *is true over* T *is* $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$*-complete, and* $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$*-hardness holds even if* T *is atomic.*

Proof: We show first membership in $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$ and then completeness for this class.

To prove membership in $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$, we sketch an algorithm that decides $p >_D q$ in polynomial time with $O(\log m)$ calls to an NP oracle, where $m = |V(T) \cap V(p)|$. Check with the oracle whether T or p is unsatisfiable; if so, then $p >_D q$ equals $p \models q$, which is decided by one oracle call. Otherwise, proceed in two steps. First, the distance $d = |\triangle|^{min}(T, p)$ between T and p is determined. Note that $d \leq |V(T) \cap V(p)|$ since for

$M \in Mod(T)$, $M' \in Mod(p)$ such that $|M \triangle M'| = d$ (hence $M' \in Mod(T \circ_D p)$), $M \triangle M' \subseteq V(T) \cap V(p)$ must hold. d can be computed by binary search with $O(\log m)$ oracle calls (deciding $d \leq k$? for given k is obviously in **NP**). The second step is to call the oracle whether there exist $M \in Mod(T)$, $M' \in Mod(p)$ such that $M' \models \neg q$ and $|M \triangle M'| = d$; such M, M' do not exist iff $p >_D q$ is true over T.

For the proof of $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$-hardness, we refer to the problem UOCSAT: Given a set $C = \{C_1, \ldots, C_m\}$ of clauses, decide if all truth assignments that satisfy a maximum number of clauses in C always satisfy the same set of clauses in C. This problem is proved $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$-complete in [27]. It can be transformed into CF_D in polynomial time as follows. Let $c_1, \ldots, c_m$ be new variables that do not appear in C. Define

$$T_1 = \{c_1, \ldots, c_m\}, \quad p_1 = (C_1 \vee \neg c_1) \wedge \cdots \wedge (C_m \vee \neg c_m) \tag{5}$$

It is not difficult to see that a model M is in $Mod(T_1 \circ_D p_1)$ iff the truth assignment by M to the variables in $V(C)$ satisfies as many clauses in C as simultaneously satisfiable (C_i is satisfied by M iff $M \models c_i$). Now let T_1' and p_1' be the knowledge base resp. formula that results if all of its variables are primed, i.e. $T_1' = \{c_1', \ldots, c_m'\}$, and $p_1' = (C_1' \vee \neg c_1') \wedge \cdots \wedge (C_m' \vee \neg c_m')$. ($C_i'$ is the modified clause C_i.) Then, T_1 and T_1' resp. p_1 and p_1' are equivalent upon renaming of variables. Define

$$T = T_1 \cup T_1', \quad p = p_1 \wedge p_1', \quad q = (c_1 \equiv c_1') \wedge \cdots \wedge (c_m \equiv c_m') \tag{6}$$

$T \circ_D p$ "combines" each model of $T_1 \circ_D p_1$ on $V(T_1 \wedge p_1)$ with each model of $T_1' \circ_D p_1'$ on $V(T_1' \wedge p_1')$; thus it is easily verified that $p >_D q$ is true over T iff all truth assignments that satisfy a maximum number of clauses in C always satisfy the same clauses in C. □

7 Horn formulas

So far, we have considered counterfactuals on arbitrary propositional knowledge bases. Since in this case already the implication problem "$T \models p$?" is intractable, the complexity increase for implication testing under counterfactual semantics from co-**NP**-hardness to $\mathbf{\Pi}_2^P$-hardness is not drastic. But what if the implication problem is tractable? It is an interesting question if in this case counterfactual semantics does not raise complexity of implication testing considerably such that tractability is preserved.

The most important restriction of this kind is the Horn case, for which the implication problem $p \models q$ for Horn formulae p, q is decidable in $O(\|p\| \cdot \|q\|)$ time (this follows easily from the linear time algorithms for Horn satisfiability in [9, 42]). It turns out that the complexity of counterfactuals restricted to the Horn case falls back one level in the polynomial hierarchy under all but cardinality based semantics. Horn counterfactuals are (at least) co-**NP**-hard under any of the change operations considered, however, and thus this restriction does not gain tractability as it does for the implication problem.

Recall that a (propositional) Horn clause is a disjunction $l_1 \vee \cdots \vee l_n$ of negated atoms and at most one atom (also written as implication $v_1 \wedge \cdots \wedge v_{n-1} \Rightarrow v_n$, where $l_1 = \neg v_1, \ldots, l_{n-1} =$

$\neg v_{n-1}$ and $l_n = v_n$ for variables $v_1, \ldots, v_n$). A Horn formula is a conjunction of Horn clauses, and a we say that a knowledge base is Horn if it is a set of Horn formulas. Note that atomic knowledge bases and literal bases are Horn. The restriction of CF to the case where T is a set of Horn formulas and p, q are Horn formulae is referred to as HORN-CF.

Now here is the basic result on the lower complexity bound of HORN-CF.

Lemma 7.1 *Let T be a knowledge base, p and q propositional formulae where T, p, and q are Horn. To decide if $p > q$ is true over T is co-***NP***-hard for $> \in \{>_G, >_{Wid}, >_B, >_{Win}, >_S, >_{Web}\}$, even if T is a literal base and $q \in T$.*

Proof: In the spirit of the principle of Minimal Change, we describe one transformation that works for all change semantics for which co-**NP**-hardness is claimed.

We transform the problem UNSAT, which is well-known co-**NP**-complete [16], into HORN-CF for $>_G$. UNSAT is as follows: Given a set $C = \{C_1, \ldots, C_m\}$ of clauses, decide if it is impossible to satisfy all clauses in C simultaneously. We assume without loss of generality that no clause C_i is empty.

The transformation is as follows. Let $V(C) = \{x_1, \ldots, x_n\}$, and let $y_1, \ldots, y_n, z_1, \ldots, z_n$, and s be new variables. Let $F[y]$ denote the formula obtained if in the Boolean formula F all positive occurrences of x_i are replaced by $\neg y_i$, for $1 \leq i \leq n$. Define

$$\begin{aligned} T &= \{x_1, y_1, \ldots, x_n, y_n, \neg z_1, \ldots, \neg z_n, \neg s\} \\ p &= [z_1 \wedge \cdots \wedge z_n \Rightarrow s] \bigwedge_{1 \leq i \leq n} [(\neg x_i \vee \neg y_i) \wedge (x_i \Rightarrow z_i) \wedge (y_i \Rightarrow z_i)] \wedge \bigwedge_{1 \leq j \leq m} C_j[y] \\ q &= \neg s \end{aligned} \tag{7}$$

($C_j[y]$ is certainly a Horn clause). Note that both T and p are satisfiable (to see the satisfiability of p, set all x_i, y_i *false* and s and all z_i *true*).

Let us briefly explain the intuition behind this transformation. The positive occurrences of x_i in C are replaced by $\neg y_i$ to make the clause set Horn; if the truth value of y_i is different from the truth value of x_i for all y_i, then C is satisfied by this assignment iff $C[y]$ is satisfied. z_i indicates if x_i and y_i have actually different truth values; z_i is true in a "possible world" $W \in W(p, T)$ exactly if $x_i \not\equiv y_i$ is valid in W. If C is satisfiable, there will be a "possible world" W in which all z_i are valid, and hence s will be valid in W; otherwise, $\neg s$ will be contained in every "possible world" W.

We claim that $p >_G q$ holds in T if and only if C is unsatisfiable.

Assume C is not unsatisfiable, i.e. there is some truth assignment $\phi(x)$ to $x_1, \ldots, x_n$ such that C is satisfied. Consider $W = \{x_i : \phi(x_i) = \text{true}, 1 \leq i \leq n\} \cup \{y_i : \phi(x_i) = \text{false}, 1 \leq i \leq n\}$. It is easily verified that $W \wedge p$ is satisfiable: $W \wedge p \models [x_i \equiv \neg y_i]$, for $1 \leq i \leq n$, hence each clause $C_j[y]$ is satisfied, $1 \leq j \leq m$; thus if all z_i and s are set *true*, $W \wedge p$ is satisfied. Moreover, W is maximal consistent with p, as $W \wedge p \models z_1 \wedge \cdots \wedge z_n$, and $W \wedge p \models s$. Thus $W \in W(p, T)$. Because $W \wedge p \not\models \neg s$, it follows that $p >_G q$ is false over T.

Conversely, assume that $p >_G q$ does not hold, i.e. for some $W \in W(p,T)$, we have $W \wedge p \not\models \neg s$. Since $\neg s \in T$, we infer that $W \wedge p \wedge \neg s$ is unsatisfiable. The only part of p that forces implication of s is $z_1 \wedge \cdots \wedge z_n \Rightarrow s$. Thus we infer that $W \wedge p \models z_1 \wedge \cdots \wedge z_n$. Maximality of W with respect to set inclusion entails that either x_i or y_i must be in W in this case, which implies that $W \wedge p \models [x_i \equiv \neg y_i]$ for $1 \leq i \leq n$. Consequently, C is satisfied by the truth assignment ϕ defined by $\phi(x_i) = \mathit{true}$ if $x_i \in W$ and $\phi(x_i) = \mathit{false}$ if $x_i \notin W$, $1 \leq i \leq n$, and hence C is not unsatisfiable.

Thus the transformation is proved correct for $>_G$. For the other change semantics, we notice that T is a satisfiable literal base and that $V(p) \subseteq V(T)$ holds; thus by (1),(2) in Lemma 6.1 the transformation is correct for $> \in \{>_S, >_B, >_{Win}\}$, too. Since T is a satisfiable literal base, p is satisfiable, and $q \in T$, by (5) in Lemma 6.1 the transformation works also for $> \in \{>_{Wid}, >_{Web}\}$. □

Counterfactuals in Horn knowledge bases are by this Lemma co-**NP**-hard under any of the considered change semantics (except Forbus's and Dalal's approach, which go uncovered by this Lemma). We show co-**NP**-completeness for the following cases.

Theorem 7.2 *To decide if a counterfactual $p > q$ is true over a Horn knowledge base T, where p and q are Horn formulae is co-**NP**-complete for $> \in \{>_G, >_{Win}, >_B, >_S\}$, even if T is a literal base and $q \in T$.*

Proof: By Lemma 7.1 it remains to show membership in co-**NP**. We sketch how to disprove validity of $p > q$ in T with an **NP** algorithm, without taking care of the rather trivial special cases for T and p. Recall that implication and satisfiability checking is polynomial for Horn knowledge bases resp. Horn formulae.

- $>_G$: Guess $T' \subseteq T$. Check if $T' \cup \{p, f\}$ is unsatisfiable for every $f \in T - T'$. Check if $T' \wedge p \not\models q$.
- $>_{Win}$: Guess $M \in \mathit{Mod}(T)$ and $M' \in \mathit{Mod}(p)$. Test whether $M' \models \neg q$ holds. Let for every variable z, $t(z) = z$ if $M \models z$ and $t(z) = \neg z$ if $M \not\models z$, and let $D = M \triangle M'$. Check if for each $x \in D$, the following Horn formula is unsatisfiable:

$$p \wedge t(x) \wedge \bigwedge_{y \in V(T) - D} t(y).$$

- $>_B$ If $T \wedge p$ is satisfiable, then check $T \wedge p \not\models q$, else proceed as for $>_{Win}$.
- $>_S$: Guess $M \in \mathit{Mod}(T)$ and $M' \in \mathit{Mod}(p)$. Check whether $M' \models \neg q$ holds. Let $D = M \triangle M'$ and $V(T \wedge p) = \{x_1, \ldots, x_n\}$. Substitute y_i for x_i in p for all x_i, where the y_i are new variables that appear nowhere else, and let $p[y]$ denote the resulting formula. Test for each $x_j \in D \cap V(T \wedge p)$ whether the following Horn formula is unsatisfiable:

$$T \wedge p[y] \wedge (x_j \Rightarrow y_j) \wedge (y_j \Rightarrow x_j) \wedge \bigwedge_{x_i \in V(T \wedge p) - D} [(x_i \Rightarrow y_i) \wedge (y_i \Rightarrow x_i)].$$

□

Similar to the issue whether CF_{Wid} and CF_{Web} are in $\mathbf{\Pi}_2^P$ for arbitrary knowledge bases and formulae, it is not clear if HORN-CF_{Wid} and HORN-CF_{Web} are in co-**NP**. The computational equivalence of these problems under polynomial transformations follows easily from constructions (3) and (4) in the proof of Theorem 6.5 (note that $[z_i \Rightarrow (x_i \equiv y_i)]$ is logically equivalent to $[(z_i \wedge x_i \Rightarrow y_i) \wedge (z_i \wedge y_i \Rightarrow x_i)]$, hence p' can be replaced by an equivalent Horn formula, and that satisfiability of T, p is polynomially decidable, thus special cases make no troubles). Moreover, one can show that both problems are solvable with logarithmically many calls to an **NP** oracle: Analogous to the algorithms sketched in the proof of Theorem 6.7, first the number of formulas of $T \circ_{Wid} p$ resp. $\Omega(T,p)$ is computed, which takes $O(\log m)$ oracle calls, where m is the cardinality of T resp. the number of variables in $V(T) \cap V(p)$, and then one final oracle call decides $p > q$. These algorithms settle both problems in $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$.

It remains to consider cardinality based change semantics. Horn counterfactuals have the same complexity as counterfactuals in arbitrary knowledge bases under Forbus's or Dalal's change operation. We show by proper modification of the transformations in the proofs of Lemma 6.3 and Theorem 6.9 that $\mathbf{\Pi}_2^P$-hardness resp. $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$-hardness still stands under this restriction.

For the sake of intelligibility and readability, we introduce a notational shorthand for a formula that fixes equivalence of variables. Let $Y = \{y_1, \ldots, y_n\}$ be a set of variables. We denote by $EQ(Y)$ the formula $(y_1 \Rightarrow y_2) \wedge \cdots \wedge (y_{n-1} \Rightarrow y_n) \wedge (y_n \Rightarrow y_1)$. Note that $EQ(Y)$ is Horn.

Let us turn our attention to Forbus's approach first.

Theorem 7.3 HORN-CF *under Forbus's change semantics is* $\mathbf{\Pi}_2^P$*-complete, and* $\mathbf{\Pi}_2^P$*-hardness holds even if* T *is a literal base.*

Proof: Recall that in the proof of Lemma 6.3 we constructed $T = \{y_1, z_1, \ldots, y_n, z_n, r\}$, $p = [(y_1 \equiv \neg z_1) \wedge \cdots \wedge (y_n \equiv \neg z_n) \wedge (r \Rightarrow E(\underline{x}, \underline{y}))]$ and $q = r$ for formula $F = \forall \underline{x} \exists \underline{y} E(\underline{x}, \underline{y})$ where $\underline{x} = <x_1, \ldots, x_m>$, $\underline{y} = <y_1, \ldots, y_n>$, such that $p >_F q$ is true over T iff F is true. Without loss of generality we may assume that $E(\underline{x}, \underline{y}) = C_1 \wedge \cdots \wedge C_l$ is a conjunction of clauses $C_1, \ldots, C_l$ [54].

Let $s_{y_1}, s_{z_1}, \ldots, s_{y_n}, s_{z_n}$ be additional variables not contained in $V(T \wedge p \wedge q)$. Define

$$\begin{aligned} T_1 &= \{r, x_1, \ldots, x_m, y_1, s_{y_1}, z_1, s_{z_1}, \ldots, y_n, s_{y_n}, z_n, s_{z_n}\} \\ p_1 &= \bigwedge_{1 \le i \le n} [(\neg y_i \vee \neg z_i) \wedge EQ(\{y_i, s_{y_i}\}) \wedge EQ(\{z_i, s_{z_i}\})] \wedge \bigwedge_{1 \le j \le n} (C_j[z] \vee \neg r) \end{aligned}$$

where $C_j[z]$ denotes the clause obtained from C_j replacing every positive y_i by $\neg z_i$, for all i. Note that this transformation is obviously computable in polynomial time.

Consider any model $M \in \mathit{Mod}(T_1)$. Let ϕ be the truth assignment to $\underline{x}$ in M. The distance of M to p_1 satisfies the following relation: $2n \le |\triangle|^{min}(M, p_1) \le 2n + 1$. To verify this,

the lower bound is reached if and only if there exists a truth assignment ψ to $\underline{y}$ such that $E(\phi(\underline{x}), \psi(\underline{y}))$ is true (in this case, merely one of y_i, z_i and the corresponding s_{y_i} resp. s_{z_i} must switch to *false* to get a model of p_1). The upper bound follows from the observation that any model fixing ϕ and satisfying $\neg r$ and $(\neg y_i \vee \neg z_i) \wedge EQ(\{y_i, s_{y_i}\}) \wedge EQ(\{z_i, s_{z_i}\})$ is a model of p_1. This entails that no more than $2n+1$ truth values have to be switched in M to get a model of p_1.

As a consequence, $p_1 >_F q$ is false over T_1 iff there is some model in $Mod(T_1 \circ_F p_1)$ in which r is false iff $|\triangle|_M^{min}(p_1) = 2n+1$ for some $M \in Mod(T_1)$, which means that $E(\phi(\underline{x}), \underline{y})$ is unsatisfiable, i.e. F is false. □

Introducing redundant variables is also appropriate for a proof that the lower complexity bound for CF under Dalal's change semantics remains unchanged in the Horn case.

Theorem 7.4 HORN-CF *under Dalal's change semantics is* $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$*-complete, and* $\mathbf{P}^{\mathbf{NP}[O(\log n)]}$*-hardness holds even if T is a literal base.*

Proof: By Theorem 6.9 we only need to show hardness for this class. Recall that in the proof of Theorem 6.9, the problem UOCSAT for clause set $C = \{C_1, \ldots, C_m\}$ is transformed into evaluation of the counterfactual $p >_D q$ in a knowledge base T defined in (6), (5) by $T = T_1 \cup T_1'$, $p = p_1 \wedge p_1'$ and $q = (c_1 \equiv c_1') \wedge \cdots \wedge (c_m \equiv c_m')$, where $T_1 = \{c_1, \ldots, c_m\}$, $p_1 = (C_1 \vee \neg c_1) \wedge \cdots \wedge (C_m \vee \neg c_m)$, and T_1', p_1' are the "primed versions" of T and p, respectively.

Let $V(C) = X = \{x_1, \ldots, x_n\}$, and let $Y = \{y_1, \ldots, y_n\}$, $Z_{x_1}, Z_{y_1}, \ldots, Z_{x_n}, Z_{y_n}$ be pairwise disjoint sets of additional new propositional variables that do not occur elsewhere such that $|Z_{x,i}| = |Z_{y,i}| = m$. We define

$$\begin{aligned} T_2 &= X \cup Y \cup \{c_1, \ldots, c_m\} \cup Z_{x_1} \cup Z_{y_1} \cup \cdots \cup Z_{x_n} \cup Z_{y_n} \\ p_2 &= \bigwedge_{1 \le i \le n} [(\neg x_i \vee \neg y_i) \wedge EQ(Z_{x_i} \cup \{x_i\}) \wedge EQ(Z_{y_i} \cup \{y_i\})] \wedge \bigwedge_{1 \le j \le m} (C_j[y] \vee \neg c_j), \end{aligned}$$

where $C_j[y]$ denotes the clause C_j with each positive occurrence of x_i replaced by $\neg y_i$, for all i. This transformation is clearly computable in polynomial time.

It is obvious that every model of $T_2 \circ_D p_2$ satisfying $x_i \equiv \neg y_i$ for all i is also a model of $T_1 \circ p_1$. The formulae $EQ(Z_{x_i} \cup \{x_i\})$, $EQ(Z_{y_i} \cup \{y_i\})$ make sure that $x_i \equiv \neg y_i$ must be true in any model of $T_2 \circ p_2$, however: If for some i both x_i and y_i have value *false* in model $M \in Mod(p_2)$, then $|\triangle|_M^{min}(T_2) \ge (n-1)(m+1) + 2(m+1)$. But any model M' satisfying $\neg c_1 \wedge \cdots \wedge \neg c_m$ and $(x_i \equiv \neg y_i) \wedge EQ(Z_{x_i} \cup \{x_i\}) \wedge EQ(Z_{y_i} \cup \{y_i\})$ for all i is a model of p_2 such that $|\triangle|_{M'}^{min}(T_2) = n(m+1)+m < |\triangle|_M^{min}(T_2)$, hence $M \notin Mod(T_2 \circ_D p_2)$. On the other hand, every model of $T_1 \circ_D p_1$ that satisfies $x_i \equiv \neg y_i$ and $EQ(Z_{x_i} \cup \{x_i\})$, $EQ(Z_{y_i} \cup \{y_i\})$ for all i, is a model of $T_2 \circ_D p_2$.

Since none of the variables $c_1', \ldots, c_m'$ appears in $V(T_1 \wedge p_1)$, $V(T_2 \wedge p_2)$, it follows that $Mod(T_1 \circ_D p_1)$ and $Mod(T_2 \circ_D p_2)$ are identical under restriction to the variables in the set $V(T_1 \wedge p_1 \wedge q)$.

Now let T_2' and p_2' be the "primed versions" of T_2 and p_2, and define $T^* = T_2 \cup T_2'$, $p^* = p_2 \wedge p_2'$. It is not difficult to verify that restricted to $V(T \wedge p \wedge q)$, $Mod(T^* \circ_D p^*)$ and $Mod(T \circ_D p)$ are identical; it follows that $p >_D q$ is true over T iff and $p^* >_D q$ is true over T^*. Since $q = (c_1 \equiv c_1') \wedge \cdots \wedge (c_m \equiv c_m')$ can be efficiently rewritten as a Horn formula, this completes the proof. □

8 Bounding the knowledge change

Lemma 6.2 is a strong result on the lower complexity bound of CF, because it states that evaluating $p > q$ in T is $\mathbf{\Pi}_2^P$-hard even in the case in which the knowledge base T and q are very "elementary". The complexity is entirely introduced by the formula p describing the knowledge change. Therefore it is near at hand to investigate into cases in which the knowledge change meets constraints.

A reasonable constraint of this kind is to assume that that p involves only a small number of facts. This can be modeled by a constant upper bound for the length $\|p\|$ of p, cf. [58].

Bounding the knowledge change to a small number of facts yields lower complexity for counterfactual evaluation, at least under model-based change. Nevertheless, this problem remains intractable under all change semantics considered. The following theorem is easily obtained from results in [58].

Theorem 8.1 *Let T be a knowledge base and let p, q be propositional formulae where $\|p\| \leq k$ for a constant k. To decide if $p > q$ is true over T is co-**NP**-complete for $> \in \{>_{Win}, >_F\}$ and is **NP**-hard as well as co-**NP**-hard, but solvable with a constant number of calls to an **NP** oracle for $> \in \{>_B, >_S, >_D, >_{Web}\}$ (two oracle calls for $>_B$).*

Proof: co-**NP**-hardness for all these change semantics is easy: Any set C of clauses is unsatisfiable iff $x > \neg C$ is true over $T = \{\}$, where $x \notin V(C)$.

Winslett presents in [58] algorithms that compute $T \circ p$ in time $O(\log t)$ if $\|p\|$ is bounded by a constant, where t is the number of occurrences of variables in T for $\circ \in \{\circ_{Win}, \circ_F\}$. Thus membership of CF_{Win}, CF_F in co-**NP** for $\|p\| \leq k$ is obvious; hence co-**NP**-completeness is proved.

Obviously, $p >_B q$ is decidable with two calls to an **NP** oracle if $\|p\| \leq k$, as $p >_{Win} q$ is in co-**NP** in this case. Unless **NP** = co-**NP**, to decide $p >_B q$ is neither in **NP** nor in co-**NP**, however, as this is both co-**NP**-hard and **NP**-hard. **NP**-hardness for $>_B$ and $>_S, >_D, >_{Web}$ is easily shown exploiting the fact that for $\circ \in \{>_B, >_S, >_D, >_{Web}\}$, the satisfiability of $T \circ p$ implies that $T \circ p \equiv T \wedge p$ (this follows immediately from the properties of these change operators with respect to the revision postulates in Section 3.2.1). Thus if C is a set of clauses that is without loss of generality falsifiable, $x > C$ is true over $T = \{x \equiv C\}$ iff C is satisfiable, where $x \notin V(C)$ (if C is unsatisfiable, then $T \circ x \equiv x$, otherwise $T \circ x \equiv x \wedge C$ for $\circ \in \{>_B, >_S, >_D, >_{Web}\}$).

The upper complexity bound for the operators $>_S, >_D, >_{Web}$ follows from the observation that $\triangle^{min}(T,p)$ is computable with a *constant* number of **NP** oracle calls if $\|p\| \leq k$, cf. [58]. (The constant depends on $|V(p)|$.) Once $\triangle^{min}(T,p)$ is known, a formula for $T \circ p$ is easily constructed in polynomial time for $\circ \in \{\circ_S, \circ_D, \circ_{Web}\}$, and then $p > q$ is decidable with one **NP** oracle call. □

Under formula-based changes, however, the lower complexity bound of unrestricted CF still stands for small knowledge change.

Theorem 8.2 *It is $\mathbf{\Pi}_2^P$-complete for $>_G$ and $\mathbf{\Pi}_2^P$-hard for $>_{Wid}$ to decide if $p > q$ is true over a complete knowledge base T, and $\mathbf{\Pi}_2^P$-hardness holds in both cases even if $q \in T$ and p, q are atoms.*

Proof: Let s be a new propositional variable. Define T', p', q' from T, p, q in (1) by $T' = T \cup \{\neg s, \neg s \vee p\}$, $p' = s$, $q = r$. It is immediately verified that T is complete for variable set $V = T \cup \{s\}$ and that $p' >_G q'$ is true over T' iff $p >_G q$ is true over T. Thus by (5) in Lemma 6.1 for T', p', q' it follows that $p' >_G q'$ iff $\neg r \in T' \circ_{Wid} p'$ iff $p' >_{Wid} q'$. □

We note that by this theorem the open problem in [58, p.33], if it is $\mathbf{\Sigma}_2^P$-complete to decide if WIDTIO change removes a formula from the knowledge base, where the length of the change formula p is bounded by a constant, is immediately answered positively (a proof of membership in $\mathbf{\Sigma}_2^P$ is already sketched by Winslett). We remark that Nebel's proof of $\mathbf{\Pi}_2^P$-hardness of $p >_G q$ in [45] also would solve the hardness part of this problem, provided that the knowledge base T is allowed to be unsatisfiable.

A positive result

In view of all the negative results from above on tractability of propositional counterfactuals, the question arises which restrictions make the problem tractable. A straightforward trial is to combine the Horn constraint and small knowledge change; this yields tractability for counterfactual evaluation, indeed, but merely for model-based change semantics.

Theorem 8.3 *Let T be a set of Horn formulas and let p, q be Horn formulas, where $\|p\| \leq k$ for k a constant. Deciding the validity of $p > q$ over T is possible in $O(\|T\| \cdot \|q\|)$ time for $> \in \{>_D, >_F, >_B, >_{Win}, >_S, >_{Web}\}$.*

Proof: Denote by $\mathcal{M}(T)$ and $\mathcal{M}(p)$ the models of T and p on the variables $V(p)$. For $\circ_{Win}$ and $\circ_S$, changes of models of T to models of p are restricted to $V(p)$, and the change of $M \in Mod(T)$ is determined by $M \cap V(p)$, which is in $\mathcal{M}(T)$. The key observation is that under the constraint $\|p\| \leq k$, $|\mathcal{M}(p)|$ as well as $|\mathcal{M}(T)|$ is bounded by the constant 2^k, and $\mathcal{M}(p)$, $\mathcal{M}(T)$ can be computed in $O(\|T\|)$ time (using the linear algorithms in [9, 42], for instance). As an immediate consequence, $\triangle_M^{min}(p)$ for every $M \in Mod(T)$ as well as $\triangle^{min}(T,p)$ is also computable in $O(\|T\|)$ time (note that $|\triangle_M^{min}(p)|, |\triangle^{min}(T,p)| < 2^k$).

$T \circ_S p$ is representable as a disjunction $D = D_1 \vee \cdots \vee D_l$ of Horn formulas, which is efficiently constructible in from T, p, $\triangle^{min}(T,p)$ in time linear in the size of T follows. Let for $M \in \mathcal{M}(p)$, $t_M : V(p) \to \{v, \neg v : v \in V(p)\}$ be defined by $t_M(v) = v$ if $M \models v$ and $t_M(v) = \neg v$ if $M \not\models v$. For all $M \in \mathcal{M}(T), M' \in \mathcal{M}(p)$ such that $M \triangle M' \in \triangle^{min}(T,p)$, set up a disjunct $D^{M,M'} = T[M] \wedge \bigwedge_{v \in V(p)} t_{M'}(v)$, where $T[M]$ is a Horn formula obtained from T as follows: First, for each variable v in $V(p)$ all occurrences of v in T are replaced by the truth value of v in M, and then the formula resulting from this step is simplified by removing false literals and clauses with true literals such that a Horn formula is obtained. This is certainly possible in $O(\|T\|)$ time, and hence $D^{M,M'}$ is constructible in $O(\|T\|)$ time. The number of disjuncts $D^{M,M'}$ is trivially bounded by the constant 2^{2k}, thus D can be constructed in $O(\|T\|)$ time. Given D, $p >_S q$ is false over T iff there is some disjunct D_i in D such that $D_i \not\models q$; Since l is bounded by the constant 2^{2k}, and since D_i is Horn and $\|D_i\| \leq \|T\| + k$, this can be checked in $O(\|T\| \cdot \|q\|)$ time. Consequently, $p >_S q$ is evaluable over T in $O(\|T\| \cdot \|q\|)$ time.

For $\circ_{Win}$, note that if $M, M' \in Mod(T)$ such that $M \cap V(p) = M' \cap V(p)$, then $\triangle_M^{min}(p) = \triangle_{M'}^{min}(p)$. Thus as for $T \circ_S p$, a disjunction $D = D_1 \vee \cdots \vee D_l$, $l \leq 2^{2k}$, of Horn formulas D_i that represents $T \circ_{Win} p$ can be constructed in $O(\|T\|)$ time, setting up for each $M \in \mathcal{M}(T)$ and all $M' \in \mathcal{M}(p)$ such that $M \triangle M' \in \triangle_M^{min}(p)$ a disjunct $D^{M,M'}$ in D. Therefore, evaluation of $p >_{Win} q$ over T is possible in $O(\|T\| \cdot \|q\|)$ time; it is immediately verified that the same holds for $p >_B q$, too.

For $\circ_D$, if T is unsatisfiable, then $p >_D q$ is equivalent to $p \models q$; if T is satisfiable, then $\triangle^{min}(T,p)$ just has to be "pruned" to $\{S \in \triangle^{min}(T,p) : |S| = |\triangle|^{min}(T,p)\}$, which can be done in constant time, and then the same procedure as for $\circ_S$ applies. Thus $p >_D q$ can be evaluated over T in $O(\|T\| \cdot \|q\|)$ time (satisfiability of T and $p \models q$ is checkable in $O(\|T\|)$, $O(\|q\|)$ time, respectively).

Similarly, for $\circ_F$ only $\triangle_M^{min}(p)$ has to be "pruned" to $\{S \in \triangle_M^{min}(p) : |S| = |\triangle|_M^{min}(p)\}$, and then one can proceed as in case of $>_{Win}$. Thus, $O(\|T\| \cdot \|q\|)$ time complexity for evaluating $p >_F q$ over T follows.

Last but not least, Weber's set $\Omega(T,p)$ is immediately obtained from $\triangle^{min}(T,p)$, and since $\Omega(T,p) \subseteq V(p)$, $|\Omega(T,p)|$ is bounded by the constant k; Hence, $res_\Omega(T) = D_1 \vee \cdots \vee D_l$ where $l \leq 2^k$ and each D_i can be transformed in $O(\|T\|)$ time into a Horn formula. Consequently, $res_\Omega(T) \wedge p \models q$ is decidable in $O((\|T\| + \|p\|)\|q\|)$ time, hence in $O(\|T\| \cdot \|q\|)$ time. From this, $O(\|T\| \cdot \|q\|)$ time for evaluating $p >_{Web} q$ over T is straightforward. □

Note that by the proof of Theorem 8.3 it is immediately clear that p needs not necessarily be Horn; the theorem still holds for arbitrary formulas p if $\|p\|$ is bounded by a constant. Furthermore, the bound $\|p\| \leq k$ can be relaxed to an $O(\log(\|T\| + \|p\| + \|q\|))$ bound on $|V(p)|$ and we still stay polynomial, since in this case $|\mathcal{M}(T)|$ and $|\mathcal{M}(p)|$ are bounded by a polynomial in the input size and $\mathcal{M}(T)$, $\mathcal{M}(p)$ etc. are computable in polynomial time.

In contrast to Theorem 8.3, bounding the length of p by a constant does not lower worst case complexity for formula-based change.

Theorem 8.4 *Let T be a knowledge base of Horn formulas, and let p, q be Horn formulas where $\|p\| \leq k$ for some constant k. To decide if $p > q$ is true over T is co-***NP***-complete for $>_G$ and co-***NP***-hard for $>_{Wid}$. co-***NP***-hardness holds in both cases even if $q \in T$ and p, q are literals.*

Proof: A slight modification of the construction (7) in Lemma 7.1 will show this. Let t be a new variable, and let p_t be the formula obtained from p adding $\neg t$ to each clause of p. Define T', p', q' from T, p, q by $T' = T \cup \{p_t\}, p' = t, q' = \neg s$. Then $p >_G q$ is true over T iff $p' >_G q'$ is true over T', which holds iff $p' >_{Wid} q'$ is true over T' iff $q' \in T' \circ_{Wid} p'$ by (5) in Lemma 6.1; hence we are done. □

This result can be sharpened to the case in which T is a set of Horn *clauses* rather than Horn *formulas*, at least for Ginsberg's notion of change.

Theorem 8.5 *Let T be a set of Horn clauses, and let p, q be literals. To decide if $p >_G q$ is true over T is co-***NP***-hard.*

Proof: We transform UNSAT into this problem (rf. to the proof of Lemma 7.1 for UNSAT). Let $C = \{C_1, \ldots, C_m\}$ be a set of clauses, where $V(C) = \{x_1, \ldots, x_n\}$. Let $y_1, \ldots, y_n, u, f$ be new variables and define

$$T_f = \{x_1 \wedge y_1 \Rightarrow f, \ldots, x_n \wedge y_n \Rightarrow f, C_1[y] \vee f, \ldots, C_m[y] \vee f\},$$

where $C_j[y]$ is the clause obtained from C_j replacing each positive occurrence of x_i by $\neg y_i$, for all i. Note that T_f is a set of Horn clauses that is satisfied if f is set *true*. Then, define

$$\begin{aligned} T &= T_f \cup \{x_1, y_1, \ldots, x_n, y_n, x_1 \wedge y_1 \Rightarrow u, \ldots, x_n \wedge y_n \Rightarrow u\}, \qquad (8) \\ p &= \neg u, \\ q &= f. \end{aligned}$$

Since f occurs within in T only in formulas of T_f, it is obvious that for every $W \in W(p, T)$ it holds that $T_f \subseteq W$ and that for all $1 \leq i \leq n$, exactly two of x_i, y_i, and $x_i \wedge y_i \Rightarrow u$ are in W. Hence it is easily verified that $p >_G q$ is false over T, i.e. there is some $W \in W(p, T)$ such that $W \wedge p \wedge \neg f$ is satisfiable, if and only if C is satisfied by some truth assignment ϕ to $x_1, \ldots, x_n$. □

Finally, we remark that transformation (8) does not work to show co-**NP**-hardness of CF under WIDTIO change (note that $T \models f$, but $f \notin T$). It is an interesting open question whether HORN-CF$_{Wid}$ remains intractable if the length of p is bounded, especially since this problem goes hand in hand with efficient computation of $T \circ_{Wid} p$.

9 Conclusion

We have considered the computational cost of inferencing a formula q from a propositional knowledge base T that has to be updated or revised with a formula p before, according to

a minimal change approach; this problem amounts to the evaluation of the counterfactual $p > q$ over T under the respective change semantics.

The main finding was that evaluating a counterfactual is for almost all proposed change operators at the second level of the polynomial hierarchy, and in many cases complete for the class $\mathbf{\Pi}_2^P$; roughly speaking, this is due to two different sources of complexity, namely propositional inference and the number of choices to make T consistent with p under minimal change. The complexity results give strong evidence that it is for most approaches unrealistic to expect a polynomial algorithm for evaluating counterfactuals under the respective semantics, even if an algorithm can solve arbitrarily many **NP**-complete problems (e.g., calls to a theorem prover for classical propositional logic) without cost.

We also have considered counterfactuals under the two restrictions of Horn formulas, i.e. all formulas are conjunctions of Horn clauses, and bounded knowledge change, i.e. the case where the change is described by a formula with size bounded by a constant. Note that inference in classical propositional logic is tractable for Horn formulas. These restrictions have different effects on the complexity of counterfactuals according to the underlying change operators. Most noticeably, we found tractability of evaluating counterfactuals only if both restrictions are imposed simultaneously, and only for model-based change operators, but not for formula-based change methods. Counterfactuals are still intractable if only one of the restrictions applies.

Comparing counterfactuals to other common formalisms of nonmonotonic reasoning such as circumscription, default logic, abduction, and truth maintenance systems, we found similar complexity results. Two conclusions from all these complexity results may be drawn. Firstly, nonmonotonic propositional reasoning seems to be inherently harder than classical propositional logic from the computational point of view. Secondly, all of the major formalisms of nonmontonic reasoning are roughly of the same difficulty, and reasoning tasks can be polynomially transformed from one formalism to most of the other formalisms. As a practical consequence, existing algorithms or theorem provers for the one approach can be fruitfully applied to the other.

It remains for the future to find ways for efficient processing of propositional nonmonotonic reasoning. An important issue therefor are suitable restrictions for the various formalisms that eliminate all sources of intractability; a lot work has to be done in that direction.

References

[1] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] K. Apt. Introduction to Logical Programming. Technical report, University of Texas, Department of Computer Science, 1987.

[3] A. Borgida. Language Features for Flexible Handling of Exceptions in Information Systems. *ACM Trans. on Database Syst.*, 10:563–603, 1985.

[4] G. Bossu and P. Siegel. Saturation, Nonmonotonic Reasoning and the Closed World Assumption. *Artificial Intelligence*, 25:13–63, 1985.

[5] M. Cadoli and M. Lenzerini. The Complexity of Closed World Reasoning and Circumscription. In *Proceedings AAAI-90*, pages 550–555, 1990.

[6] M. Dalal. Investigations into a Theory of Knowledge Base Revision: Preliminary Report. In *Proceedings AAAI-88*, pages 475–479, 1988.

[7] M. Dalal. Updates in Propositional Databases. Technical Report DCS-TR-222, Rutgers University, Dept. of Computer Science, Feb. 1988.

[8] J. de Kleer. An Assumption-Based Truth Maintenance System. *Artificial Intelligence*, 28:127–162, 1986.

[9] W. Dowling and J. H. Gallier. Linear-time Algorithms for Testing the Satisfiability of Propositional Horn Theories. *Journal of Logic Programming*, 1(3):267–284, 1984.

[10] T. Eiter and G. Gottlob. Propositional Circumscription and Extended Closed World Reasoning are $\mathbf{\Pi}_2^P$-complete. Technical Report CD-TR 91/20, Christian Doppler Laboratory for Expert Systems, Vienna University of Technology, Austria, May 1991. To appear in *Theoretical Computer Science*.

[11] T. Eiter and G. Gottlob. Higher Level Complexity Results for Abduction. Technical Report CD-TR 92/35, Christian Doppler Laboratory for Expert Systems, TU Vienna, March 1992.

[12] R. Fagin, G. Kuper, J. Ullman, and M. Y. Vardi. Updating Logical Databases. In P. Kanellakis and F. Preparata, editors, *Advances in Computing Research*, volume 3. JAI Press, 1986.

[13] R. Fagin, J. D. Ullman, and M. Y. Vardi. On the Semantics of Updates in Databases. In *Proceedings PODS-83*, pages 352–365, 1983.

[14] K. D. Forbus. Introducing Actions into Qualitative Simulation. In *Proceedings IJCAI-89*, pages 1273–1278, 1989.

[15] P. Gärdenfors. *Knowledge in Flux*. Bradford Books, MIT Press, 1988.

[16] M. Garey and D. S. Johnson. *Computers and Intractability – A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York, 1979.

[17] M. Gelfond, H. Przymusinksa, and T. Przymusinski. On the Relationship Between Circumscription and Negation as Failure. *Artificial Intelligence*, 38:75–94, 1989.

[18] M. Gelfond and H. Przymusinska. Negation as Failure: Careful Closure Procedure. *Artificial Intelligence*, 30:273–287, 1986.

[19] M. L. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30:35–79, 1986.

[20] M. L. Ginsberg. A Circumscriptive Theorem Prover. *Artificial Intelligence*, 39:209–230, 1989.

[21] M. L. Ginsberg and D. E. Smith. Reasoning About Action I: A Possible Worlds Approach. In F. Brown, editor, *Proc. Workshop "The Frame Problem in Artificial Intelligence"*, pages 233–258. Morgan Kaufmann Publ., 1987. also in *Artificial Intelligence*, 35(165–195), 1988.

[22] G. Gottlob. Complexity Results for Nonmonotonic Logics. *Journal of Logic and Computation*, 2:3, June 1992.

[23] G. Grahne. Updates and Counterfactuals. In *Proceedings KR-91*, pages 269–276, 1991.

[24] G. Grahne and A. O. Mendelzon. Updates and Subjunctive Queries. Technical Report KRR-TR-91-4, University of Toronto, CS Dept., July 1991. Also Proceedings Workshop on Nonstandard Queries and Answers, Toulouse, July 1-3, 1991, pp. 33-52.

[25] J. Hartmanis. New Developments in Structural Complexity Theory. *Theoretical Computer Science*, 71:79–93, 1990.

[26] D. S. Johnson. A Catalog of Complexity Classes. volume A of *Handbook of Theoretical Computer Science*, chapter 2. Elsevier Science Publishers B.V. (North-Holland), 1990.

[27] J. Kadin. $P^{NP[O(\log n)]}$ and Sparse Turing-Complete Sets for NP. *Journal of Computer and System Sciences*, 39:282–298, 1989.

[28] H. Katsuno and A. O. Mendelzon. A Unified View of Propositional Knowledge Base Updates. In *Proceedings IJCAI-89*, pages 1413–1419, 1989.

[29] H. Katsuno and A. O. Mendelzon. On the Difference Between Updating a Knowledge Base and Revising it. Technical Report KRR-TR-90-6, University of Toronto, CS Dept., August 1990.

[30] H. Katsuno and A. O. Mendelzon. On the Difference between Updating a Knowledge Base and Revising it. In *Proceedings KR-91*, pages 387 – 395, 1991.

[31] H. Katsuno and A. O. Mendelzon. Propositional Knowledge Base Revision and Minimal Change. *Artificial Intelligence*, 52:253–294, 1991.

[32] A. Keller and M. Winslett-Wilkinson. On the Use of an Extended Relational Model to Handle Changing Incomplete Information. *IEEE Transactions on Software Engineering*, SE:11(7):620–633, 1985.

[33] M. Krentel. The Complexity of Optimization Problems. *Journal of Computer and System Sciences*, 36:490–509, 1988.

[34] D. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, MA, 1973.

[35] V. Lifschitz. Closed-World Databases and Circumscription. *Artificial Intelligence*, 27:229–235, 1985.

[36] W. Marek and M. Truszczyńsky. Modal logic for default reasoning. *Annals of Mathematics and Artificial Intelligence*, 1:275–302, 1990.

[37] J. McCarthy. Circumscription – A Form of Non-Monotonic Reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[38] J. McCarthy. Applications of Circumscription to Formalizing Common-Sense Knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[39] D. McDermott. Non-Monotonic Logic II: Nonmonotonic Modal Theories. *Journal of the ACM*, 29:33–57, 1982.

[40] D. McDermott and J. Doyle. Non-Monotonic Logic I. *Artificial Intelligence*, 13:41–72, 1980.

[41] J. Minker. On Indefinite Data Bases and the Closed World Assumption. In *Proceedings of the 6^{th} Conference on Automated Deduction (CADE)*, pages 292–308, 1982.

[42] M. Minoux. LTUR: a simplified linear time unit resolution for Horn formulae and computer implementation. *Information Processing Letters*, 29:1–12, 1988.

[43] R. Moore. Semantical Considerations on Nonmonotonic Logics. *Artificial Intelligence*, 25:75–94, 1985.

[44] B. Nebel. A Knowledge Level Analysis of Belief Revision. In *Proceedings KR-89*, pages 301–311, 1989.

[45] B. Nebel. Belief Revision and Default Reasoning: Syntax-Based Approaches. In *Proceedings KR-91*, pages 417–428, 1991.

[46] C. H. Papadimitriou and M. Yannakakis. The Complexity of Facets (And Some Facets of Complexity). *Journal of Computer and System Sciences*, 28:244–259, 1984.

[47] T. Przymusinski. An Algorithm to Compute Circumscription. *Artificial Intelligence*, 38:49–73, 1989.

[48] R. Reiter. On Closed-World Databases. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 55–76. Plenum Press, New York, 1978.

[49] R. Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[50] N. Rescher. *Hypothetical Reasoning*. North-Holland, Amsterdam, 1964.

[51] N. Rescher. *The Coherence Theory of Truth*. Oxford University Press, 1973.

[52] V. Rutenburg. Complexity Classification in Truth Maintenance Systems. In *Proceedings STACS-91*, number 480 in Lecture Notes in Computer Science, pages 373–383, 1991.

[53] K. Satoh. Nonmonotonic Reasoning by Minimal Belief Revision. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, pages 455–462, 1988.

[54] L. Stockmeyer and A. Meyer. Word Problems Requiring Exponential Time. In *Proceedings of the Fifth ACM Symposium on the Theory of Computing*, pages 1–9, 1973.

[55] K. Wagner. Bounded query classes. *SIAM J. Comp.*, 19(5):833–846, 1990.

[56] A. Weber. Updating Propositional Formulas. In *Proceedings of the First Conference on Expert Database Systems*, pages 487–500, 1986.

[57] M. Winslett. Reasoning about Action Using a Possible Models Approach. In *Proceedings AAAI-88*, pages 89–93, 1988.

[58] M. Winslett. *Updating Logical Databases.* Cambridge University Press, 1990.

[59] M. Winslett. Circumscriptive Semantics for Updating Databases. *Annals of Mathematics and Artificial Intelligence*, 3:429–450, 1991.

[60] A. Yahya and L. Henschen. Deduction in Non-Horn Databases. *Journal of Automated Reasoning*, 1(2):141–160, 1985.