I N F S Y S
RESEARCH
R E P O R T

INSTITUT FÜR INFORMATIONSSYSTEME

ABTEILUNG WISSENSBASIERTE SYSTEME

GENERATING ALL ABDUCTIVE EXPLANATIONS FOR QUERIES ON PROPOSITIONAL HORN THEORIES

Thomas Eiter and Kazuhisa Makino

INFSYS RESEARCH REPORT 1843-03-09

JUNE 2003

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

TECHNISCHE UNIVERSITÄT WIEN

INFSYS RESEARCH REPORT
INFSYS RESEARCH REPORT 1843-03-09, JUNE 2003

GENERATING ALL ABDUCTIVE EXPLANATIONS FOR QUERIES ON PROPOSITIONAL HORN THEORIES
(PRELIMINARY REPORT)

Thomas Eiter[1] and Kazuhisa Makino[2]

Abstract. Abduction is a fundamental mode of reasoning, which has taken on increasing importance in Artificial Intelligence (AI) and related disciplines. Computing abductive explanations is an important problem, and there is a growing literature on this subject. We contribute to this endeavor by presenting new results on computing multiple resp. all of the possibly exponentially many explanations of an abductive query from a propositional Horn theory represented by a Horn CNF. Here the issues are whether a few explanations can be generated efficiently and, in case of all explanations, whether the computation is possible in *polynomial total time* (or *output-polynomial time*), i.e., in time polynomial in the combined size of the input and the output. We explore these issues for queries in CNF and important restrictions thereof. Among the results, we show that computing all explanations for a negative query literal from a Horn CNF is not feasible in polynomial total time unless $\mathrm{P} = \mathrm{NP}$, which settles an open issue. However, we show how to compute under restriction to acyclic Horn theories polynomially many explanations in input polynomial time and all explanations in polynomial total time, respectively. Complementing and extending previous results, this draws a detailed picture of the computational complexity of computing multiple explanations for queries on Horn theories.
Keywords: Computational logic, abduction, propositional logic, Horn theories, polynomial total time computation, NP-hardness.

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

[2]Division of Systems Science, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan. E-mail: makino@sys.es.osaka-u.ac.jp

Acknowledgements: This work was partly supported by the Austrian Science Fund (FWF) project Z29-N04, by TU Wien through a scientific collaboration grant, and by the Scientific Grant in Aid of the Ministry of Education, Science, Sports and Culture of Japan.

A short version of this paper appears in: Proceedings 12th Annual Conference of the EACSL (CSL 2003), J. Makowsky and M. Baaz, editors, Springer LNCS, 2003.

Contents

1 Introduction

Abduction is a fundamental mode of reasoning, which was extensively studied by C.S. Peirce [18]. It has taken on increasing importance in Artificial Intelligence (AI) and related disciplines, where it has been recognized as an important principle of common-sense reasoning (see e.g. [3]). Abduction has applications in many areas of AI and Computer Science including diagnosis, database updates, planning, natural language understanding, learning etc. (see e.g. references in [10]), where it is primarily used for generating explanations.

In a logic-based setting, abduction can be viewed as the task to find, given a set of formulas Σ (the *background theory*) and a formula χ (the *query*), a minimal set of formulas E (an *explanation*) from a set of hypotheses H such that Σ plus E is satisfiable and logically entails χ. Often considered is a scenario where Σ is a propositional Horn theory, χ is a single literal or a conjunction of literals, and H contains literals (see [23, 10] and references therein). For use in practice, the computation of abductive explanations in this setting is an important problem, for which well-known early systems such as Theorist [19] or ATMS solvers [6, 21] have been devised. Since then, there has been a growing literature on this subject, indicating the need for efficient abductive procedures. We refer to [17], which gives an excellent survey on intimately closely related problems in computational logic. Note that much effort has been spent on studying various input restrictions, cf. [13, 4, 12, 24, 8, 7, 10, 22, 23].

While computing *some* explanation of a query χ has been studied extensively in the literature, the issue of computing multiple or even *all* explanations for χ has received less attention. This problem is important since often one would like to select one out of a set of alternative explanations according to a preference or plausibility relation; this relation may be based on subjective intuition and thus difficult to formalize. As easily seen, exponentially many explanations may exist for a query, and thus computing all explanations inevitably requires exponential time in general, even in propositional logic. However, it is of interest whether the computation is possible in *polynomial total time* (or *output-polynomial time*), i.e., in time polynomial in the combined size of the input and the output. Furthermore, if exponential space is prohibitive, it is of interest to know whether a few explanations (e.g., polynomially many) can be generated in polynomial time, as studied by Selman and Levesque [23].

Computing some explanation for a query χ which is a literal from a Horn theory is a well-known polynomial problem. Selman and Levesque conjectured [23] that generating $O(n)$ many explanations for a positive literal is NP-hard, where n is the number of propositional atoms in the language, even if it is guaranteed that there are only few explanations overall. As shown in [11], this conjecture is not true unless $\mathrm{P}{=}\mathrm{NP}$. This follows from the result of [11] that all explanations for an atom can be generated in polynomial total time.

The status of generating all explanations for a negative literal $\chi = \overline{q}$ from a Horn CNF φ, however, remained open in [11]. Moreover, it was unclear whether a resolution-style procedure similar to the one for query atoms in [11] could solve the problem in polynomial total time. In this paper, we provide a negative answer to this question, by showing that given a collection of explanations for a query $\chi = \overline{q}$ from a Horn CNF φ, deciding whether there is an additional explanation is NP-complete. Consequently, the existence of a polynomial total time algorithm for computing all explanations implies P=NP. However, for the well-known class of acyclic Horn theories (see e.g. [5, 23, 20, 1]) we present an algorithm which enumerates all explanations for $\overline{q}$ with incremental polynomial delay (i.e., in time polynomial in the size of the input and output so far), and thus solves the problem in polynomial total time. Compared to explanations for an atomic query q, intuitively cyclic dependencies between atoms make the problem difficult. For completeness, a resolution-style procedure as in [11] needs to consider besides the input and output clauses also auxiliary clauses (see Example 7), whose derivation may cause a lot of overhead, since it is not a priori clear which such clauses are needed.

We furthermore address computing all explanations for queries χ beyond literals, where we consider CNF and important special cases such as a clause and a term (i.e., a conjunction of literals). Note that the explanations for single clause queries correspond to the minimal support clauses for a clause in Clause Management Systems [21]. In the light of the negative results from above, we aim at elucidating the tractability frontier and present positive as well as negative results for such queries.

Our results shed new light on the computational nature of abduction and Horn theories in particular. They imply that, e.g., generating all minimal support clauses for a given clause (cf. [21]) from an acyclic Horn CNF is feasible in polynomial total time. The intractability result for negative literal queries $\overline{q}$ is somewhat unexpected, and the tractability result for acyclic Horn theories is more difficult to obtain than in case of atomic queries. As a byproduct, we also obtain results for computing all prime implicates of Horn theories containing a certain literal, which complement and refine previous results for computing all prime implicates of a Horn theory [2].

The rest of this paper is organized as follows. The next section contains preliminaries and fixes notation. In Section 3, we show that generating all explanations for a negative query literal is coNP-hard, while in Section 4, we present our positive result for acyclic Horn theories. The final Section 6 concludes the paper. Technical details of some proofs have been moved to the Appendix.

2 Preliminaries and Notation

We assume a standard propositional language with atoms $x_1, x_2, \ldots, x_n$ from a set At, where each x_i takes either value 1 (true) or 0 (false). Negated atoms are denoted by $\overline{x}_i$, and the opposite of a literal ℓ by $\overline{\ell}$. Furthermore, we use $\overline{A} = \{\overline{\ell} \mid \ell \in A\}$ for any set of literals A and set $Lit = At \cup \overline{At}$.

A clause is a disjunction $c = \bigvee_{p\in P(c)} p \vee \bigvee_{p\in N(c)} \overline{p}$ of literals, where $P(c)$ and $N(c)$ are the sets of atoms occurring positively and negated in c and $P(c) \cap N(c) = \emptyset$. Dually, a term is conjunction $t = \bigwedge_{p\in P(t)} p \wedge \bigwedge_{p\in N(t)} \overline{p}$ of literals, where $P(t)$ and $N(t)$ are similarly defined. We also view clauses and terms as sets of literals $P(c) \cup N(c)$ and $P(t) \cup N(t)$, respectively. A clause c is *Horn*, if $|P(c)| \leq 1$; *definite*, if $|P(c)| = 1$; and *negative* (resp., *positive*), if $|P(c)| = 0$ (resp., $|N(c)| = 0$). A conjunctive normal form (CNF) is a conjunction of clauses. It is *Horn* (resp., *definite, negative, positive*), if it contains only Horn clauses (resp., definite, negative, positive clauses). A *theory* Σ is any finite set of formulas; it is *Horn*, if it is a set of Horn clauses. As usual, we identify Σ with $\varphi = \bigwedge_{c\in\Sigma} c$, and write $c \in \varphi$ etc.

A *model* is a vector $v \in \{0, 1\}^n$, whose i-th component is denoted by v_i. For $B \subseteq \{1, \ldots, n\}$, we let x^B be the model v such that $v_i = 1$, if $i \in B$ and $v_i = 0$, if $i \notin B$, for $i \in \{1, \dots, n\}$. Satisfaction $v \models \varphi$ and logical consequence $\varphi \models c$, $\varphi \models \psi$ etc. are defined as usual (i.e., $\varphi(v) = 1$ etc.).

Example 1 *The CNF* $\varphi = (\overline{x}_1 \vee \overline{x}_4) \wedge (\overline{x}_4 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee x_2) \wedge (\overline{x}_4 \vee \overline{x}_5 \vee x_1) \wedge (\overline{x}_2 \vee \overline{x}_5 \vee x_3)$ *over* $At = \{x_1, x_2, \dots, x_5\}$ *is Horn. The vector* $u = (0, 1, 0, 1, 0)$ *is a model of* φ. □

The following proposition is well-known.

Proposition 2.1 *Given a Horn CNF* φ *and a clause* c*, deciding whether* $\varphi \models c$ *is possible in polynomial time* (*in fact, in linear time, cf.* [9]).

Recall that two clauses c and c' resolve on a pair of literals $x, \overline{x}$ if $x, \overline{x} \in c \cup c'$ and $c \cup c' \setminus \{x, \overline{x}\}$ is a legal clause (thus, x must occur in exactly one of c and c', and same for $\overline{x}$); c and c' resolve if their is a pair of literals $x, \overline{x}$ on which they resolve. Note that this pair, if it exists, is unique. In that case, we denote by $c \oplus c'$ the clause $c \cup c' \setminus \{x, \overline{x}\}$, which is their *resolvent* (otherwise, $c \oplus c'$ is undefined). A *resolution proof* of a clause c from a CNF φ is a sequence $c_1, c_2, \dots, c_l$ of clauses such that $c_l = c$ and, for all $i = 1, \dots, l$, either $c_i \in \varphi$ or $c_i = c_j \oplus c_k$ for clauses c_j and c_k such that $j, k < i$. It is well-known that resolution proofs are sound and complete with respect to clause inference in the following sense (cf. [17]): For any CNF φ and clause c, $\varphi \models c$ holds iff there is a clause $c' \subseteq c$ which has a resolution proof from φ. For further background on resolution, we refer to [16, 15].

2.1 Abductive explanations

The notion of an abductive explanation can be formalized as follows (cf. [23, 10]).

Definition 1 Given a (Horn) theory Σ, called the background theory, a CNF χ (called *query*), an *explanation of* χ is a minimal set of literals $E \subseteq Lit$ such that

(i) $\Sigma \cup E \models \chi$, and

(ii) $\Sigma \cup E$ is satisfiable.

Example 2 Reconsider the Horn CNF $\varphi = (\overline{x}_1 \vee \overline{x}_4) \wedge (\overline{x}_4 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee x_2) \wedge (\overline{x}_4 \vee \overline{x}_5 \vee x_1) \wedge (\overline{x}_2 \vee \overline{x}_5 \vee x_3)$ from above. Suppose we want to explain $\chi = x_2$ from $A = \{x_1, x_4\}$. Then, we find that $E = \{x_1\}$ is an explanation. Indeed, $\Sigma \cup \{x_1\} \models x_2$, and $\Sigma \cup \{x_1\}$ is satisfiable; moreover, E is minimal. On the other hand, $E' = \{x_1, \overline{x}_4\}$ satisfies (i) and (ii) for $\chi = x_2$, but is not minimal. □

More restricted forms of explanations require that E must be formed over a given set of abducible letters (cf. [23]); however, in such a setting, generating all explanations is easily seen to be coNP-hard for the cases that we consider from results in the literature.

The following characterization of explanations is immediate by the monotonicity of classical logic.

Proposition 2.2 *For any theory Σ, any query χ, and any $E \subseteq Lit$, E is an explanation for χ from Σ iff the following conditions hold: (i) $\Sigma \cup E$ is satisfiable, (ii) $\Sigma \cup E \models \chi$, and (iii) $\Sigma \cup (E \setminus \{\ell\}) \not\models \chi$, for every $\ell \in E$.*

From Proposition 2.2, we thus obtain the following easy lemma.

Lemma 2.1 *Given a Horn CNF φ, a set $E \subseteq Lit$, and a CNF query χ, deciding whether E is an explanation for χ w.r.t. A is possible in polynomial time.*

3 Intractability of Negative Literal Queries

In this section, we show that computing all explanations of a negative query $\chi = \overline{q}$ is not possible in polynomial total time unless $\mathrm{P} = \mathrm{NP}$. This result follows by standard arguments from the following theorem.

Theorem 3.1 *Given a Horn CNF φ, a query $\chi = \overline{q}$, and explanations $E_1, E_2, \ldots, E_k$ for χ, deciding whether χ has some additional explanation E_{k+1} different from each E_i, $1 \leq i \leq k$, is* NP*-complete. Hardness holds even if φ is definite Horn.*

In the proof of Theorem 3.1, we use the following well-known lemma, which links prime implicates of a theory to explanations. Recall that a *prime implicate* of a theory Σ is a minimal (w.r.t. inclusion) clause c such that $\Sigma \models c$. Let us call an explanation E for a literal query $\chi = \ell$ *trivial*, if $E = \{\ell\}$.

Lemma 3.1 (cf. [21, 14]) *Given a theory Σ, a set $E \subseteq Lit$ is a nontrivial explanation of a query literal χ iff the clause $c = \bigvee_{\ell \in E} \overline{\ell} \vee \chi$ is a prime implicate of Σ.*

Note $\chi = \ell$ has a trivial explanation iff $\Sigma \not\models \ell$ and $\Sigma \not\models \overline{\ell}$, which can be checked in polynomial time. Hence, as for NP-hardness we can without loss of generality focus on generating the nontrivial explanations of $\overline{q}$, i.e., all prime implicates containing $\overline{q}$.

Proof of Theorem 3.1. As for membership in NP, an additional explanation E_{k+1} can be guessed and, by Lemma 2.1, be verified in polynomial time.

We show the NP-hardness by a reduction from 3SAT. Let $\gamma = c_1 \wedge \cdots \wedge c_m$, $m \geq 2$, be a 3CNF over atoms $x_1, \ldots, x_n$, where $c_i = \ell_{i,1} \vee \ell_{i,2} \vee \ell_{i,3}$. We introduce for each clause c_i a new atom y_i, for each x_j a new atom x'_j (which intuitively corresponds to $\overline{x}_i$), and special atoms q and z. The Horn CNF φ contains the following clauses:

1. $c_{i,j} = \overline{q} \vee \overline{\ell^\star_{i,j}} \vee y_i$, for all $i = 1, \ldots, m$ and $j = 1, 2, 3$;
2. $d_{i,j} = \overline{\ell^\star_{i,j}} \vee y_i \vee \overline{y}_{i+1}$, for all $i = 1, \ldots, m$ and $j = 1, 2, 3$;
3. $\overline{x}_i \vee \overline{x'_i} \vee z$, for all $i = 1, \ldots, n$;
4. $e = \overline{y}_1 \vee \overline{y}_2 \vee \cdots \vee \overline{y}_m \vee z$,

where $\ell^\star_{i,j} = x_k$ if $\ell_{i,j} = x_k$ and $\ell^\star_{i,j} = x'_k$ if $\ell_{i,j} = \overline{x}_i$, and $y_{m+1} = y_1$.

Note that φ is definite Horn, and thus all prime implicates of φ are definite Horn. Informally, the clauses $c_{i,j}$ and $d_{i,j}$ stand for selection of literal $\ell_{i,j}$ in clause c_i. The clause in 4., which is needed to produce any negative prime implicate c containing $\overline{q}$, and the minimality of a prime implicate will effect that a literal is chosen from each clause c_i, and the clauses in 3. will ensure that the choice is consistent, such that γ is satisfied. Since the positive prime implicates containing $\overline{q}$ happen to be just the clauses $c_{i,j}$, the existence of a further prime implicate of φ containing $\overline{q}$ is equivalent to satisfiability of γ.

We establish the following properties of φ, which are proved in the appendix.

Lemma 3.2 *Any prime implicate c of φ such that $q \in N(c)$ and $P(c) \neq \{z\}$ is of the form $c_{i,j}$, where $i \in \{1, \ldots, m\}$ and $j \in \{1, 2, 3\}$.*

Lemma 3.3 *Any prime implicate c of φ such that $P(c) = \{z\}$ and $q \in N(c)$ satisfies (i) $\{x_i, x'_i\} \not\subseteq N(c)$, for all $i = 1, \ldots, n$, and (ii) $y_i \notin N(c)$, for all $i = 1, \ldots, m$.*

From Lemma 3.2, it is now easy to see that all prime implicates of φ given by the clauses in 1. correspond to nontrivial explanations of $\overline{q}$, and from Lemma 3.3 that an additional nontrivial explanation for $\overline{q}$ exists if and only if some prime implicate of φ of form $c = \overline{q} \vee \bigvee_{x_i \in X} \overline{x}_i \vee \bigvee_{x'_i \in X'} \overline{x'}_i \vee z$ exists iff the CNF γ is satisfiable. As for the last equivalence, note that for each smallest (w.r.t. $\subseteq$) choice $\ell_i \in c_i$ of a consistent collection of literals $\ell_1, \ldots, \ell_m$, we have an additional prime implicate of φ of form $\overline{q} \vee \bigvee_{\ell_i} \overline{\ell^\star_i} \vee z$. Conversely, each additional prime implicate c containing $\overline{q}$ gives rise to a consistent set of literals $\{x_j \mid x_{i,j} \in N(c)\} \cup \{\overline{x}_j \mid x'_{i,j} \in N(c)\}$ which satisfies γ.

Clearly, φ is constructible in polynomial time from γ. Since φ is definite, this proves the NP-hardness under the asserted restriction. □

We notice that in the Horn CNF φ in the hardness proof of Theorem 3.1, the atom z occurs only positive. Thus, φ remains Horn if the polarity of z is changed. We obtain therefore easily the following result.

Theorem 3.2 *Given a Horn CNF φ and a query $\chi = \overline{q}$, deciding whether there exists some positive explanation for χ, i.e., explanation E which consists only of positive literals, is* NP*-complete. Hardness holds even if all non-positive explanations for χ from φ are part of the input.*

In contrast, deciding the existence of some positive explanation for an atomic query $\chi = q$ is feasible in polynomial time (make φ prime and check whether some clause afterwards contains q). Theorem 3.1 strengthens the result in [11] that in general, deciding the existence of some positive explanation for a negative literal query is NP-complete, to the case where all other explanations are known. As shown in the next section, this can not be further strengthened to the case of acyclic Horn theories similar as in [11].

4 Negative Literal Queries on Acyclic Horn Theories

Since as shown in the previous section, a polynomial total time procedure for generating all explanations of a negative literal query is infeasible in general unless P=NP, it becomes an issue to find restricted input classes for which this is feasible. In this section, we show a positive result for the important class of acyclic Horn theories, which has been studied extensively in the literature (see, e.g., [5, 23, 20, 1]).

We first recall the concept of acyclic Horn theories (see e.g. [5, 23]).

Definition 2 *For any Horn CNF φ over atom set At, its dependency graph is the directed graph $G(\varphi) = (V, E)$, where $V = At$ and $E = \{x_i \to x_j \mid c \in \varphi, x_i \in N(c),\, x_j \in P(c)\}$, i.e., E contains an arc from each atom in a negative literal to the positive literal in a clause (if such a literal exists). A Horn CNF φ is* acyclic *if $G(\varphi)$ has no directed cycle.*

Example 3 *As easily seen, the edges of $G(\varphi)$ for the CNF φ in Examples 1 and 2 are $x_1 \to x_2$, $x_4 \to x_1$, $x_5 \to x_1$, $x_2 \to x_3$, and $x_5 \to x_3$. Hence, φ is acyclic.* □

Since the trivial explanation $E = \{\overline{q}\}$ can be easily generated (if it applies), we focus on generating all nontrivial explanations. For a negative query on an acyclic Horn theory, this is accomplished by Algorithm N-EXPLANATIONS in Figure 1. It first converts the input into an equivalent prime Horn CNF $\varphi^\star$, and then applies a restricted resolution procedure, in which pairs (c, c') of clauses are considered of which at least one is a prime implicate containing $\overline{q}$ and the other is either a clause of this form or a clause from the converted input $\varphi^\star$. In case their resolvent $d := c \oplus c'$ exists and, as implied by condition (ii) in Definition 1, includes only prime implicates containing $\overline{q}$, any such prime implicate d' is computed. If d' is recognized as a new prime implicate which has not been generated so far, a corresponding explanation is output and the set of candidate pairs is enlarged.

Example 4 Reconsider the Horn CNF $\varphi = (\overline{x}_1 \vee \overline{x}_4) \wedge (\overline{x}_4 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee x_2) \wedge (\overline{x}_4 \vee \overline{x}_5 \vee x_1) \wedge (\overline{x}_2 \vee \overline{x}_5 \vee x_3)$, and apply N-EXPLANATIONS for $\chi = \overline{x}_1$. All clauses of φ are prime except $\overline{x}_4 \vee \overline{x}_5 \vee x_1$, which contains the prime implicate $\overline{x}_4 \vee \overline{x}_5$. Thus, $\varphi^\star = (\overline{x}_1 \vee \overline{x}_4) \wedge (\overline{x}_4 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee x_2) \wedge (\overline{x}_4 \vee \overline{x}_5) \wedge (\overline{x}_2 \vee \overline{x}_5 \vee x_3)$, and S contains the clauses $\overline{x}_1 \vee \overline{x}_4$ and $\overline{x}_1 \vee x_2$; the corresponding explanations $E_1 = \{x_4\}$ and $E_2 = \{\overline{x}_2\}$ are output.

In Step 2, the pair $(\overline{x}_2 \vee \overline{x}_5 \vee x_3, \overline{x}_1 \vee x_2)$ is found in O which satisfies condition (i) in Definition 1. Moreover, (ii) is satisfied, since $\varphi^\star \not\models \overline{x}_5 \vee x_3$. Thus, a prime implicate contained in $d = \overline{x}_1 \vee \overline{x}_5 \vee x_3$ is computed; in fact, d is prime. Therefore, $E_3 = \{x_5, \overline{x}_3\}$ is output and d is added to S, and subsequently O is enlarged. Eventually, the pair $(\overline{x}_3 \vee \overline{x}_4,\ \overline{x}_1 \vee \overline{x}_5 \vee x_3)$ from O, which satisfies condition (i), will be considered. However, $\varphi^\star \models \overline{x}_4 \vee \overline{x}_5$, and thus S remains unchanged. Hence, the output of N-EXPLANATIONS is E_1, E_2, and E_3. As can be seen, these are all nontrivial explanations for $\overline{x}_1$ from φ. □

We remark that our algorithm is similar in spirit to an algorithm for computing all prime implicates of a Horn CNF in polynomial total time [2]. Our algorithm solves a more constrained problem, though.

Algorithm N-EXPLANATIONS

Input: An acyclic Horn CNF φ and an atom q.

Output: All nontrivial explanations of the query $\chi = \overline{q}$ from φ.

Step 1. $\varphi^\star := \emptyset$, $S := \emptyset$, and $O := \emptyset$;

Step 2. **for each** $c \in \varphi$ **do**
add any prime implicate $c' \subseteq c$ of φ to $\varphi^\star$;
for each $c' \in \varphi^\star$ with $q \in N(c')$ and $c' \notin S$ **do**
begin output $(\{\overline{\ell} \mid \ell \in c' \setminus \{\overline{q}\}\}$;
$S := S \cup \{c'\}$; $O := O \cup \{(c, c') \mid c \in \varphi^\star, q \notin P(c)\}$
end;

Step 3. **while** some $(c, c') \in O$ exists **do**
begin $O := O \setminus \{(c, c')\}$;
if (1) c and c' resolve and (2) $\varphi^\star \not\models (c \oplus c' \setminus \{\overline{q}\})$
then begin $d := c \oplus c'$;
compute any prime implicate $d' \subseteq d$ of φ;
if $d' \notin S$ **then**
begin output $(\{\overline{\ell} \mid \ell \in d' \setminus \{\overline{q}\}\}$; $S := S \cup \{d'\}$;
$O := O \cup \{(d'', d') \mid d'' \in \varphi^\star, q \notin P(d'')\} \cup \{(d'', d') \mid d'' \in S\}$
end
end
end. □

Figure 1: Algorithm computing all nontrivial explanations of a query $\chi{=}\overline{q}$ on an acyclic Horn theory

In the rest of this section, we show that Algorithm N-EXPLANATIONS generates all explanations in polynomial total time. For that, we first show its correctness, which splits into a soundness and completeness part, and then analyze its time complexity.

As for soundness, it is easily seen that Algorithm N-EXPLANATIONS produces output only if d' is some prime implicate of $\varphi^\star$ (and thus of φ) such that $q \in N(d')$. Thus, from Lemma 3.1, we immediately obtain

Lemma 4.1 (Soundness of N-EXPLANATIONS**)** *Algorithm* N-EXPLANATIONS *outputs only nontrivial explanations for $\overline{q}$ from φ.*

It is much more difficult to show the completeness, i.e., that Algorithm N-EXPLANATIONS actually generates all nontrivial explanations. Intuitively, the difficulty stems from the fact that the restricted resolution procedure retains only prime clauses containing $\overline{q}$, and, moreover, may skip relevant prime implicates $d' \subseteq c \oplus c'$ in Step 3 if condition (ii) fails, i.e., $c \oplus c'$ is an implicate of φ (which is tantamount to the condition that $c \oplus c'$ contains some prime implicate of φ that does not contain $\overline{q}$). To see that no explanation is missed requires a careful analysis of how the desired explanations are generated, and leads to a nontrivial argument which takes the complex interaction between clauses into account.

We need a number of preliminary technical lemmas on which our proof builds, which are interesting in their own right. In what follows, we call for any literal ℓ a clause c an *ℓ-clause*, if c contains ℓ.

The following propositions are well-known.

Proposition 4.1 *Let c_1, c_2 be Horn implicates of a Horn CNF φ that resolve. Then, $c = c_1 \oplus c_2$ is Horn, and if c_1 contains a negative implicate of φ, then also $c_1 \oplus c_2$ contains a negative implicate of φ.*

Proposition 4.2 (cf. [2]) *Every prime implicate c of a Horn CNF φ has an input resolution proof from it, i.e., a resolution proof $c_1, c_2, \ldots, c_l\,(= c)$ such that either $c_i \in \varphi$ or $c_i = c_j \oplus c_k$ where $j, k < l$ and either $c_j \in \varphi$ or $c_k \in \varphi$, for all $i \in \{1, \ldots, l\}$.*

We start with the following lemma.

Lemma 4.2 *Let φ be a prime Horn CNF, and let c be any prime implicate of φ such that $c \notin \varphi$. Then, $c = c_1 \oplus c_2$, where c_1 is a prime implicate contained in φ, and either (i) c_2 is a prime implicate of φ, or (ii) $c_2 = c \cup \{\ell\}$ where $c_1 \setminus \{\overline{\ell}\} \subset c$ and c is the unique prime implicate of φ contained in c_2.*

Proof. input resolution proof. Thus, $c = c_1 \oplus c_2$ where, without loss of generality, $c_1 \in \varphi$ and thus c_1 is a prime implicate. Suppose that c_2 is not prime. Then, it contains a prime implicate d of φ. If c_1 and d resolve, then (i) holds for c_1 and d since the primality of c implies $c = c_1 \oplus d$. Otherwise, $d \subseteq c_2 \setminus \{\ell\}$ where c_1 and c_2 resolve on the pair of ℓ and $\overline{\ell}$. Thus, $d \subseteq c$ holds. The primality of c then implies $d = c$. Hence, c_2 must be of form $c \cup \{\ell\}$, and $c_1 \setminus \{\overline{\ell}\} \subseteq c$ must hold; the primality of c_1 implies that this inclusion is proper. If there exist no c_1 and c_2 as in (i), the primality of d and c_1 implies that d (=c) is the unique prime implicate contained in $c_2 = d \cup \{\ell\}$ (note that if ℓ were a prime implicate, then $c_1 \setminus \{\overline{\ell}\}$ would be a prime implicate, which is a contradiction). $\Box$ Note that item (ii) is needed in this lemma, as shown by the following example.

Example 5 Consider the Horn CNF $\varphi = (\overline{x}_0 \vee \overline{x}_1 \vee x_2)(\overline{x}_2 \vee \overline{x}_3)(\overline{x}_3 \vee x_0)$. As easily checked, φ is prime and has a further prime implicate $\overline{x}_1 \vee \overline{x}_3$, which can not be derived as the resolvent of any two prime implicates of φ. Note that φ is acyclic. $\Box$

Next we state some important properties of acyclic Horn CNFs under resolution.

Proposition 4.3 *Let φ be an acyclic Horn CNF, and let $c = c_1 \oplus c_2$ where $c_1, c_2 \in \varphi$. Then, $\varphi' = \varphi \wedge c$ is acyclic Horn, and the dependency graphs $G(\varphi)$ and $G(\varphi')$ have the same transitive closure. Furthermore, any subformula $\varphi'' \subseteq \varphi$ is acyclic Horn.*

Thus, adding repeatedly clauses derived by resolution preserves the acyclicity of a CNF, and, moreover, the possible topological sortings of the dependency graph.

The following proposition captures that for an acyclic Horn CNF, resolution cannot be blocked because of multiple resolving pairs of x_i and $\overline{x}_i$ of literals.

Proposition 4.4 *Let φ be an acyclic Horn CNF, and let c_1 and c_2 be any implicates of φ derived from φ by resolution. Then, c_1 and c_2 do not resolve iff $P(c_1) \cap N(c_2) = \emptyset$ and $P(c_2) \cap N(c_1) = \emptyset$.*

We define an ordering on Horn clauses as follows. Suppose that $\leq$ imposes a total ordering on the atoms ($x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{i_n}$). Then, for any Horn clauses c_1 and c_2, define $c_1 \leq c_2$ iff $c_1 = c_2$ or one of the following conditions holds:

(i) $P(c_1) \neq \emptyset$ and $P(c_2) = \emptyset$;

(ii) $P(c_1) = \{x_i\}$ and $P(c_2) = \{x_j\}$ and $x_i < x_j$;

(iii) $P(c_1) = P(c_2)$ and $\max N(c_1) \triangle N(c_2) \in c_1$, where "$\triangle$" denotes standard symmetric difference (i.e., $S_1 \triangle S_2 = (S_1 \cup S_2) \setminus (S_1 \cap S_2)$).

As usual, we write $c_1 < c_2$ if $c_1 \leq c_2$ and $c_1 \neq c_2$, $c_1 > c_2$ for $c_2 < c_1$ etc. Note that $\leq$ orders first all definite Horn clauses along their positive literals, followed by the negative clauses. Notice that $c_1 \subset c_2$ implies $c_2 < c_1$, for any Horn clauses c_1 and c_2.

The following proposition is not difficult to establish:

Proposition 4.5 *Every total ordering $\leq$ of the atoms At induces a total ordering $\leq$ of all Horn clauses over At as described.*

With respect to acyclic Horn CNFs φ, in the rest of this paper we assume an arbitrary but fixed total ordering $\leq$ of the atoms which is compatible with some topological sorting of the dependency graph $G(\varphi)$.

Proposition 4.6 *Let c_1 and c_2 be Horn clauses such that $c = c_1 \oplus c_2$ exists. Then, $c_1 < c$ and $c_2 < c$ hold.*

Proof. Without loss of generality, $c_1 = \overline{x_{1_1}} \vee \cdots \vee \overline{x_{1_{n_1}}} \vee x_i$ and $c_2 = \overline{x_{2_1}} \vee \cdots \vee \overline{x_{2_{n_2}}} \vee \overline{x}_i \vee\ (\vee x_k)$, where the x_k part in c_2 is optional. If c_2 is negative, then clearly $c_1 < c_2$. Otherwise, the underlying ordering on atoms implies $x_i < x_k$, and hence we also have $c_1 < c_2$. Now let us compare c_2 and $c = c_1 \oplus c_2$. Observe that $c = \overline{x_{1_1}} \vee \cdots \vee \overline{x_{1_{n_1}}} \vee \overline{x_{2_1}} \vee \cdots \vee \overline{x_{2_{n_2}}}\ (\vee x_k)$. Thus, $\max N(c) \triangle N(c_2) = x_i$. Since $x_i \in N(c_2)$, it follows that $c_2 < c$. Since $\leq$ is a total ordering, we thus have $c_1 < c_2 < c$, which proves the claim. $\square$

Corollary 4.1 *Let φ be an acyclic Horn CNF, and let c, c_1 and c_2 be any implicates of φ derived from φ such that $c \subseteq c_1 \oplus c_2$. Then, $c > c_1$ and $c > c_2$ holds.*

Consequently, in any input resolution proof of a clause from an acyclic Horn CNF the derived clauses increase monotonically. As for the derivation of prime implicates, we find for such CNFs a more general form than in Lemma 4.2:

Lemma 4.3 *Let φ be an acyclic prime Horn CNF, and let c be any prime implicate of φ such that $c \notin \varphi$. Then, there are prime implicates c_1 and c_2 of φ and, for some $k \geq 0$, prime implicates $d_1, d_2, \ldots, d_k$ and literals $\ell_1, \ell_2, \ldots, \ell_k$, respectively, such that: (i) $c_1, d_1, \ldots, d_k \in \varphi$, and (ii) $c = c_1 \oplus e_1$, where $e_i = c \cup \{\ell_i\} = d_i \oplus e_{i+1}$, for $i \in \{1, \ldots, k\}$, and $e_{k+1} = c_2$, such that e_i contains the single prime implicate c.*

Proof. By Proposition 4.2, clause c has an input resolution proof from φ. For such a proof, let $c = c_1 \oplus e_1$ for a prime implicate $c_1 \in \varphi$. If e_1 is a prime implicate of φ, then the statement holds for $k = 0$ (we have $c_2 = e_1$). Otherwise, e_1 contains some prime implicate e'_1 of φ. If $c_1 \oplus e'_1$ exists, then the statement also holds for $k = 0$. Otherwise (i.e., c_1 and e'_1 do not resolve), then by applying a similar argument in the proof of Lemma 4.2, we have $e_1 = c \cup \{\ell_1\}$, where $(\ell_1, \overline{\ell_1})$ is the resolving pair of literals, and c is the unique prime implicate of φ contained in e_1.

Consider now in turn the ancestors d_1, e_2 of e_1 in the input resolution proof of c. If both are prime, then the claimed statement holds for $k = 1$. Otherwise, w.l.o.g. $d_1 \in \varphi$ and also e_2 contains some prime implicate e'_2 of φ. Again, there are two cases:
1) d_1 and e'_2 resolve. If $d_1 \oplus e'_2 \subseteq c$ (note that $d_1 \oplus e'_2 \subseteq e_1$), then we have $d_1 \oplus e'_2 = c$ and the claimed statement holds for $k = 0$. Otherwise, $\ell_1 \in d_1 \oplus e'_2$. If $d_1 \oplus e'_2 \subset e_1$, then there exists a prime implicate $c' \subset e_1$ other than c, a contradiction. Thus we have $d_1 \oplus e'_2 = e_1$, which implies that the statement holds for $k = 1$.
2) d_1 and e'_2 do not resolve. Thus $e'_2 \subseteq e_1$. Since e_1 contains the unique prime implicate c of φ, it follows that $e'_2 = c$. Thus, e_2 is of form $e_2 = c \cup \{\ell_2\}$ as claimed. By continuing this argument, we can construct sequences of clauses $d_1, \ldots, d_k$ and literals $\ell_1, \ell_2, \ldots, \ell_k$ as claimed. Notice that for an acyclic φ, the sequence eventually must terminate since $e_i > e_{i+1}$ holds for every $i \geq 1$ by Proposition 4.6. $\square$ An immediate consequence of this result is that prime implicates of an acyclic Horn CNF can be generated from two prime implicates as follows.

Corollary 4.2 *Let φ be an acyclic prime Horn CNF, and let c be any prime implicate of φ such that $c \notin \varphi$. Then, there exist prime implicates c_1 and c_2 of φ which resolve such that either (i) $c = c_1 \oplus c_2$ or (ii) $c_1 \oplus c_2 = c \cup \{\ell\}$, where $\ell \notin c$ and c is the unique prime implicate of φ contained in $c_1 \oplus c_2$.*

In Example 5, the further prime implicate $\overline{x}_1 \vee \overline{x}_3$ can be derived as in case (ii) of Corollary 4.2: For $c_1 = \overline{x}_0 \vee \overline{x}_1 \vee x_2$ and $c_2 = \overline{x}_2 \vee \overline{x}_3$, we have $c_1 \oplus c_2 = \overline{x}_1 \vee \overline{x}_3 \vee \overline{x}_0$, and $c = \overline{x}_1 \vee \overline{x}_3$ is the unique prime implicate of φ contained in $c_1 \oplus c_2$.

After the preparatory results, we now show that Algorithm N-EXPLANATIONS is complete. Using an inductive argument on clause orderings, we show that all explanations are generated by taking into account possible derivations of prime implicates as established in Lemma 4.3 and Corollary 4.2. However, an inductive proof along $\leq$ encounters two major difficulties: First, the resolvent $c = c_1 \oplus c_2$ of two clauses is *larger* than c_1 and c_2, thus we cannot simply rearrange resolution steps and appeal to smaller clauses. Second, Algorithm N-EXPLANATIONS does not generate prime implicates d' by a resolution step alone, but using *minimization* in Step 3; that is, a prime implicate *included* in the resolvent $d = c \oplus c'$. A respective statement is much more difficult to prove than the one if d were prime.

In order to overcome these difficulties, we use a more sophisticated ordering of clause pairs (c, c') and establish as a stepping stone the following key lemma. For ease of reference, let us say that resolvable implicates c_1 and c_2 of a Horn CNF φ satisfy the (technical) property $(*)$, if the following conditions hold:

1. At least one of c_1 and c_2 is prime.
2. If c_i is not prime, then it is of form $c_i = c' \cup \{\ell\}$, where c' is the unique prime implicate of φ contained in c_i ($i \in \{1, 2\}$), and c_i occurs in some derivation of c' as in Lemma 4.3.
3. There is no implicate $c'_1 \subset c_1$ (resp., $c'_2 \subset c_2$) of φ such that $c = c'_1 \oplus c_2$ (resp., $c = c_1 \oplus c'_2$).

Lemma 4.4 (Key Lemma) *Let φ be a prime acyclic Horn CNF, and let c_1 and c_2 be resolvable clauses satisfying $(*)$ such that $\overline{q} \in c := c_1 \oplus c_2$. Suppose that $c_i \in S$ if c_i is prime and $q \in N(c_i)$ (resp., $c'_i \in S$ if $c_i = c'_i \cup \{\ell_i\}$ where c'_i is prime and $q \in N(c'_i)$) for $i \in \{1, 2\}$. Then at least one of the following conditions hold:* (i) *$c \setminus \{\overline{q}\}$ is an implicate of φ, or* (ii) *c contains a $\overline{q}$-clause from S.*

Proof. (Outline) We prove the statement using an inductive argument which involves clause orderings and takes into account how the clauses c_1 and c_2 are recursively generated. Depending on the shape of c_1 and c_2, we consider different cases.

Consider first the case in which both c_1 and c_2 contain $\overline{q}$. Then, w.l.o.g. $c = \overline{q} \vee \overline{a} \vee \overline{b}\,(\vee x_i)$, $c_1 = \overline{q} \vee \overline{a} \vee \overline{x}\,(\vee x_i)$, and $c_2 = \overline{b} \vee x \vee \overline{q}$. Here, $\overline{a}$ and $\overline{b}$ are disjunctions of negative literals, while x is a single atom; "$(\vee x_i)$" means the optional presence of x_i.

Both c_1 and c_2 contain a unique prime implicate c'_1 resp. c'_2 of φ (where possibly $c'_i = c_i$). If $\overline{q} \in c'_i$, then by assertion we have $c'_i \in S$. Thus, if both c'_1 and c'_2 contain $\overline{q}$, Algorithm N-EXPLANATIONS considers $c'_1 \oplus c'_2$, which implies the statement. No other cases are possible, since either c'_1 or c'_2 must contain $\overline{q}$ (since c_1 or c_2 is prime) and condition 3 of $(*)$ excludes that exactly one of c'_1 and c'_2 contains $\overline{q}$. This proves the statement if both c_1 and c_2 contain $\overline{q}$.

For the other cases, assume that $\overline{q} \in c_1$ and $\overline{q} \notin c_2$ and prove the statement by induction along the lexicographic ordering of the pairs (c_1, c_2), where the clauses c_1 are in *reverse* ordering $\geq$ and the clauses c_2 in regular ordering $\leq$. We distinguish the following cases:

Definite/Negative Case 1 (DN1): $c = \overline{q} \vee \overline{a} \vee \overline{b}\,(\vee x_i)$, $c_1 = \overline{q} \vee \overline{a} \vee \overline{x}\,(\vee x_i)$, and $c_2 = \overline{b} \vee x$. That is, the $\overline{q}$-clause c is generated by resolving a $\overline{q}$-clause c_1 with a non-$\overline{q}$-clause c_2, where the positive resolution literal x is in c_2.

Definite/Negative Case 2 (DN2): $c = \overline{q} \vee \overline{a} \vee \overline{b}\,(\vee x_i)$, $c_1 = \overline{q} \vee \overline{a} \vee x$, and $c_2 = \overline{b} \vee \overline{x}\,(\vee x_i)$. That is, the $\overline{q}$-clause c is generated by resolving a $\overline{q}$-clause c_1 with a non-$\overline{q}$-clause c_2, where the positive resolution literal x is in c_1.

The statement is shown by a careful analysis of parent clauses of c_1 and c_2, and by reordering and adapting resolution steps. The case DN1 recursively only involves cases of the same kind (in fact, for negative c_1 we need to appeal only to smaller instances (c'_1, c'_2) where c'_1 is negative), while case DN2 recursively involves itself as well as DN1. More details are given in the appendix. □

By combining Lemma 4.4 with Proposition 4.6 and Corollary 4.2, we obtain by an inductive argument on the clause ordering $\leq$ the desired completeness result.

Lemma 4.5 (Completeness of N-EXPLANATIONS) *Algorithm* N-EXPLANATIONS *outputs all the nontrivial explanations for a query $\chi = \overline{q}$ from an acyclic Horn CNF φ.*

Proof. We prove by induction on $\leq$ that S contains each $\overline{q}$-prime implicate c of φ.
(Basis) Let c be the least prime implicate of φ which contains $\overline{q}$. From Proposition 4.6 and Corollary 4.2, we conclude that $c \in \varphi$ must hold. Hence, $c \in S$.
(Induction) Suppose the claim holds for all $\overline{q}$-prime implicates c' of φ such that $c' < c$, and consider c. By Corollary 4.2, there exist prime implicates c_1 and c_2 such that either (i) $c = c_1 \oplus c_2$ or (ii) c is the unique prime implicate contained in $c_1 \oplus c_2 = c \cup \{\ell\}$ where $\ell \notin c$. By Proposition 4.6 and the induction hypothesis, we have $c_i \in S$ if $q \in N(c_i)$ holds for $i \in \{1, 2\}$. Consequently, c_1 and c_2 satisfy the conditions of Lemma 4.4. Hence, either (a) $c_1 \oplus c_2 \setminus \{\overline{q}\}$ is an implicate of φ, or (b) $c_1 \oplus c_2$ contains a $\overline{q}$-clause c' from S. Since $q \in N(c)$ and c is the unique prime implicate contained in $c_1 \oplus c_2$, we have (b). It follows from the uniqueness of c that $c' = c$, which proves the statement. $\Box$ We are now in a position to establish the main result of this section. Let $||\varphi||$ denote the size (number of symbols) of any CNF φ.

Theorem 4.1 *Algorithm* N-EXPLANATIONS *incrementally outputs, without duplicates, all non-trivial explanations of $\chi = \overline{q}$ from φ. Moreover, the next output (respectively termination) occurs within $O(s \cdot (s+m) \cdot n \cdot ||\varphi||)$ time, where m is the number of clauses in φ, n the number of atoms, and s the number of explanations output so far.*

Proof. By Lemmas 4.1 and 4.5, it remains to verify the time bound of Algorithm N-EXPLANATIONS. Computing a prime implicate $c' \subseteq c$ and $d' \subseteq d$ of φ in Steps 2 and 3, respectively, is feasible in time $O(n \cdot ||\varphi||)$ (cf. Proposition 2.1), and thus the outputs in Step 2 occur with $O(m \cdot n \cdot ||\varphi||)$ delay. As for Step 3, note that O contains only pairs (c, c') where $c \in \varphi^\star \cup S$ and $c' \in S$ such that the explanation corresponding to c' was generated, and each such pair is added to O only once. Thus, the next output or termination follows within $s \cdot (s+m)$ runs of the while-loop, where s is the number of solutions output so far. The body of the loop can be done, using proper data structures, in $O(n \cdot ||\varphi||)$ time (for checking $d' \notin S$ efficiently, we may store S in a prefix tree). Thus, the time until the next output resp. termination is bounded by $O(s \cdot (s+m) \cdot n \cdot ||\varphi||)$. $\Box$

We note some easy corollaries to the previous theorem.

Corollary 4.3 *Computing polynomially many explanations for a negative query $\chi = \overline{q}$ from an acyclic Horn CNF φ is feasible in polynomial time (in the size of the input).*

Corollary 4.4 *Given an acyclic Horn CNF φ and all non-positive explanations $E_1, \dots, E_k$ for a negative query $\chi = \overline{q}$ from φ, deciding the existence of any positive explanation for χ from φ (and computing one) is feasible in polynomial time.*

The last result contrasts with the similar result in [11] that the problem is NP-complete, if the non-positive explanations $E_1, \dots, E_k$ are not part of the input. Intuitively, this may be explained by the fact that acyclicity corresponds to a kind of factored representation which may generate a

lot of (exponentially many) non-positive explanations, which compensate the exponential search space for deciding the existence of a positive explanation in "difficult" cases.
We conclude this section with some remarks on Algorithm N-EXPLANATIONS.

(1) As for implementation, standard data structures and marking methods can be used to realize efficient update of the sets O and S, to determine resolvable clauses, and to eliminate symmetric pairs (c, c') an (c', c) in O.

(2) Algorithm N-EXPLANATIONS is incomplete for cyclic Horn theories, as shown by the following example.

Example 6 Consider the Horn CNF $\varphi = (\overline{x}_0 \vee \overline{x}_1 \vee x_2)(\overline{x}_0 \vee \overline{x}_1 \vee x_3)(\overline{x}_1 \vee \overline{x}_2 \vee x_3)(\overline{x}_1 \vee x_2 \vee \overline{x}_3)(\overline{x}_2 \vee \overline{x}_3 \vee x_4)$ over $x_0, \dots, x_4$. Note that all clauses in φ are prime, and that x_2 and x_3 are symmetric. There are three further prime implicates, viz. $c_1 = \overline{x}_1 \vee \overline{x}_2 \vee x_4$, $c_2 = \overline{x}_1 \vee \overline{x}_3 \vee x_4$, and $c_3 = \overline{x}_0 \vee \overline{x}_1 \vee x_4$. Thus, $\overline{q} = \overline{x}_0$ has the nontrivial explanations $E_1 = \{x_1, \overline{x}_2\}$, $E_2 = \{x_1, \overline{x}_3\}$, and $E_3 = \{x_1, \overline{x}_4\}$. Apply then algorithm N-EXPLANATIONS on input φ and $q = x_0$. While it outputs E_1 and E_2, it misses explanation E_3. □

Algorithm N-EXPLANATIONS may be extended to handle this example and others correctly by adding in Step 2 some prime implicates to $\varphi^\star$ which are generated in polynomial time (e.g., by minimizing clauses derived by resolution proofs from $\varphi^\star$ whose number of steps is bounded by a constant).

(3) Algorithm N-EXPLANATIONS is no longer complete if we constrain the resolution process to input resolution, i.e., consider only pairs (c, c') in Step 3 where at least one of c and c' is from φ (which means that in the update of O in Step 3, the part "$\{(d'', d') \mid d'' \in S\}$" is omitted). This is shown by the following example.

Example 7 Consider the Horn CNF $\varphi = (\overline{x}_0 \vee x_1)(\overline{x}_1 \vee \overline{x}_2 \vee x_3)(\overline{x}_1 \vee \overline{x}_3 \vee x_4)$ over $x_0, \dots, x_4$. As easily seen, φ is acyclic. Moreover, φ is prime. There are three further prime implicates containing $\overline{x}_0$, viz. $c_1 = \overline{x}_0 \vee \overline{x}_2 \vee x_3$, $c_2 = \overline{x}_0 \vee \overline{x}_3 \vee x_4$, and $c_3 = \overline{x}_0 \vee \overline{x}_2 \vee x_4$. Hence, $\overline{q} = \overline{x}_0$ has the nontrivial explanations $E_1 = \{\overline{x}_1\}$, $E_2 = \{x_2, \overline{x}_3\}$, $E_3 = \{x_3, \overline{x}_4\}$, and $E_4 = \{x_2, \overline{x}_4\}$. If at least one of the clauses (c, c') in Step 3 must be from φ, then E_2 and E_3 are generated from $(\overline{x}_1 \vee \overline{x}_2 \vee x_3, \overline{x}_0 \vee x_1)$ and $(\overline{x}_1 \vee \overline{x}_3 \vee x_4, \overline{x}_0 \vee x_1)$, respectively, while E_4 is missed: The pairs $(\overline{x}_1 \vee \overline{x}_3 \vee x_4, \overline{x}_0 \vee \overline{x}_2 \vee x_3)$ and $(\overline{x}_1 \vee \overline{x}_2 \vee x_3, \overline{x}_0 \vee \overline{x}_3 \vee x_4)$ yield the same resolvent $\overline{x}_0 \vee \overline{x}_1 \vee \overline{x}_2 \vee x_4$, for which the test $\varphi^\star \not\models (c \oplus c' \setminus \{\overline{q}\})$ fails since $\overline{x}_1 \vee \overline{x}_2 \vee x_4$, which is the resolvent of the second and third clause in φ, is an implicate. Note that E_4 is generated from each of the excluded symmetric pairs $(\overline{x}_0 \vee \overline{x}_2 \vee x_3, \overline{x}_0 \vee \overline{x}_3 \vee x_4)$ and $(\overline{x}_0 \vee \overline{x}_3 \vee x_4, \overline{x}_0 \vee \overline{x}_2 \vee x_3)$. □

In terms of generating prime implicates, this contrasts with the cases of computing all prime implicates of a Horn CNF and all prime implicates that contain a positive literal q , for which input-resolution style procedures are complete, cf. [2, 11].

5 Compound Queries

In this section, we consider generating all explanations for queries beyond literals. Theorem 3.1 implies that this problem is intractable for any common class of CNF queries which admits a negative literal. However, also for positive CNFs, the problem is intractable.

Theorem 5.1 *Deciding whether a given CNF χ has an explanation from a Horn CNF φ is* NP*-complete. Hardness holds even if χ is positive and φ is negative (thus acyclic).*

Proof. Membership in NP easily follows from Lemma 2.1. Hardness is shown via a reduction from the classical EXACT HITTING SET problem. Let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a collection of subsets $S_i \subseteq U$ of a finite set U. Construct $\chi = \bigwedge_i \left(\bigvee_{u \in S_i}\right)$ and $\varphi = \bigwedge_i \bigwedge_{x \neq y \in S_i} (\overline{x} \vee \overline{y})$. Then χ has an explanation from φ iff there exists an exact hitting set for $\mathcal{S}$, i.e., a set $H \subseteq U$ such that $|H \cap S_i| = 1$ for all $i \in \{1, \ldots, m\}$. $\Box$ For important special cases of positive CNFs, we obtain positive results. In particular, this holds if the query χ is restricted to be a clause or a term.

Theorem 5.2 *Computing polynomially many (resp., all) explanations for a query χ which is either a positive clause or a positive term from a Horn CNF φ is feasible in polynomial time (resp., polynomial total time).*

Proof. Let us first consider the case in which χ is a positive clause $c = \bigvee_{x \in P(c)} x$. Then let $\varphi^* = \varphi \wedge \bigwedge_{x \in P(c)} (\overline{x} \vee x^*)$, where x^* is a new letter. It is not difficult to see that φ^* is a Horn CNF, and there is a one-to-one correspondence between explanations for a query χ from φ and the ones for x^* form φ^* (except for a trivial explanation x^*). This, together with the result in [11] that all explanations for a query $\chi = q$ where q is an atom from a Horn CNF can be generated with incremental polynomial delay, proves the theorem.

Similarly, if χ is a positive term $t = \bigwedge_{x \in P(t)} x$, one can consider explanations for x^* from the Horn CNF $\varphi^* = \varphi \wedge (\bigvee_{x \in P(t)} \overline{x} \vee x^*)$, where x^* is a new letter. $\Box$ In case of acyclic Horn theories, the positive result holds also in the case where negative literals are present in a clause query.

Theorem 5.3 *Computing polynomially many (resp., all) explanations for a query $\chi = c$ where c is a clause from an acyclic Horn CNF φ is feasible in polynomial time (resp., polynomial total time).*

Proof. Let $\chi = \bigvee_{x \in P(c)} x \vee \bigvee_{x \in N(c)} \overline{x}$. Then let $\varphi^* = \varphi \wedge \bigwedge_{x \in P(c)} (\overline{x} \vee \overline{x^*}) \wedge \bigwedge_{x \in N(c)} (x \vee \overline{x^*})$, where x^* is a new letter. It is not difficult to see that φ^* is an acyclic Horn CNF, and there is a ono-to-one correspondence between explanations for a query χ from φ and the ones for $\overline{x^*}$ from φ^* (except for a trivial explanation $\overline{x^*}$). This together with Theorem 4.1 proves the theorem. $\Box$

Note that explanations for a single clause query $\chi = c$ correspond to the minimal support clauses for c as used in Clause Management Systems [21]. Thus, from Theorems 3.1 and 5.3 we obtain that while in general, generating all minimal support clauses for a given clause c is not possible in polynomial total time unless $\mathrm{P} = \mathrm{NP}$, it is feasible with incremental polynomial delay for acyclic Horn theories.

The presence of negative literals in a query $\chi = t$ for a term t from an acyclic Horn theory is more involved; a similar reduction technique as for a clause to a single literal seems not to work.

Query χ / Horn theory Σ, by	CNF		single literal		single clause		single term	
	general	positive	atom q	$\overline{q}$	positive	general	positive	general
Horn CNF φ	NPTT	NPTT	PTT[a]	NPTT	PTT	NPTT	PTT	NPTT
Acyclic Horn CNF φ	NPTT	NPTT	PTT[a]	PTT	PTT	PTT	PTT	–

[a] By the results of [11].

Table 1: Complexity of computing all abductive explanations for a query χ from a Horn theory (PTT = polynomial total time, NPTT = not polynomial total time unless P=NP)

We can show that generating all nontrivial explanations E (i.e., satisfying $E \cap \chi = \emptyset$) for a term is intractable; the case of all explanations is currently open.

6 Conclusion

We considered computing all abductive explanations for a query χ from a propositional Horn CNF φ, which is an important problem that has many applications in AI and Computer Science. In this paper, we presented a number of new complexity results, which complement and extend previous results in the literature; they are compactly summarized in Table 1.

We showed the intractability of computing all abductive explanations for a negative literal query χ from a general Horn CNF φ (thus closing an open issue), while we presented a polynomial total time algorithm for acyclic Horn CNFs. Since this amounts to computing all prime implicates of φ which contain $\overline{q}$, we have obtained as a byproduct also new results on computing all such prime implicates from a Horn CNF. Note that our intractability result contrasts with the result in [2] that all prime implicates of a Horn CNF are computable in polynomial total time. Furthermore, our results on clause queries imply analogous results for generating all minimal support clauses for a clause in a Clause Management System [21].

It remains for further work to complete the picture and to find further meaningful input classes of cyclic Horn theories which permit generating a few resp. all explanations in polynomial total time. For example, this holds for clause queries from quadratic Horn CNFs (i.e., each clause is Horn and has at most 2 literals) and for literal queries from Horn CNFs in which each clause contains the query literal. Another issue is a similar study for the case of predicate logic.

A Appendix: Proofs

Proof of Lemma 3.2. Let c be a prime implicate of φ such that $q \in N(c)$ and $P(c) \neq \emptyset$. Then c can be generated by a resolution proof involving clauses $c^{(1)}, c^{(2)}, \ldots, c^{(k)} = c$ where each $c^{(i)}$ either is from φ or is the resolvent of clauses $c^{(i_1)}$ and $c^{(i_2)}$ such that $i_1, i_2 < i$. Since $P(c) \neq \emptyset$, we have that $P(c^{(i)}) \neq \emptyset$ for each $i \leq k$ in a shortest such resolution proof.

We notice that only atoms y_i occur positively in φ. Therefore, $P(c) = \{y_i\}$ must hold for some $i \in \{1, \ldots, m\}$. Furthermore, since y_i occurs positively only in the clauses $c_{i,j}$ and $d_{i,j}$, we conclude that c must also contain at least one of the literals $\overline{\ell^\star_{i,1}}$, $\overline{\ell^\star_{i,2}}$, and $\overline{\ell^\star_{i,3}}$. It follows that $c_{i,j} \subseteq c$ holds for some $i \in \{1,\dots,m\}$ and $j \in \{1,2,3\}$. In order to complete the proof of the lemma, it thus remains to show that no clause $c_{i,j} \setminus \{\ell\}$, $\ell \in c_{i,j}$ is an implicate of φ. Indeed, $c_{i,j} \setminus \{\overline{q}\}$ is not an implicate of φ, since $\varphi(u) = 1$ for the model u obtained by assigning 0 to $y_1, y_2, \dots, y_m$ and 1 to all other atoms. Next, $c_{i,j} \setminus \{\overline{\ell^\star_{i,j}}\}$ is not an implicate of φ, since $\varphi(v) = 1$ for the model v obtained by assigning 0 to all atoms except q and z. Finally, $c_{i,j} \setminus \{y_i\}$ is not an implicate of φ, since $\varphi(w) = 1$ for the model w obtained by assigning 1 to all atoms except z. Thus, $c_{i,j}$ is a prime implicate of φ. This proves the lemma. □

Proof of Lemma 3.3. Let c be any prime implicate of φ such that $P(c) = \{z\}$ and $q \in N(c)$. Then c can be generated by a resolution proof involving clauses $c^{(1)}$, $c^{(2)}$, …, $c^{(k)} = c$. Since only the atoms y_i and z occur positively in φ and $P(c) = \{z\}$, there exists some $i^* \in \{1,\dots,k\}$ such that $c^{(i^*)} = \overline{y}_1 \vee \overline{y}_2 \vee \cdots \vee \overline{y}_m \vee z$, i.e., $c^{(i^*)}$ is the clause e from φ, and that $z \in P(c)$ must hold. Now if c contains both $\overline{x}_i$ and $\overline{x'_i}$, then we have $\overline{x}_i \vee \overline{x'_i} \vee z \models c$, which implies that c is not a prime implicate (since $\overline{q} \in c$), which is a contradiction. This proves item (i) of the lemma.

To show (ii), observe that each clause d in φ such that $y_i \in P(d)$ satisfies $\ell^\star_{i,j} \in N(d)$, for some $j \in \{1,2,3\}$. Hence, $y_i \notin N(c)$ implies $\ell^\star_{i,j} \in N(c)$ for some $j \in \{1,2,3\}$, for every $i \in \{1,\dots,m\}$. Let us write c as $c = \overline{q} \vee \bigvee_{y_i \in Y} \overline{y}_i \vee \bigvee_{x_i \in X} \overline{x}_i \vee \bigvee_{x'_i \in X'} \overline{x'}_i \vee z$. It follows from the discussion above that $(*)$ $y_i \notin Y$ implies $\{\ell^\star_{i,1}, \ell^\star_{i,2}, \ell^\star_{i,3}\} \cap (X \cup X') \neq \emptyset$ for every $i \in \{1,\dots,m\}$.

We show now that $Y = \emptyset$ must hold. Otherwise, using the clause e $(= \overline{y}_1 \vee \overline{y}_2 \vee \cdots \vee \overline{y}_m \vee z)$ from 4. and suitable clauses $d_{i,j}$ from 2., we can generate by resolution an implicate c^* of φ which contains literal $\overline{y}_i$ only if $y_i \in Y$ and such that $c^* \subseteq c \setminus \{\overline{q}\}$, which contradicts the minimality of c. Indeed, suppose that some $y_i \in Y$ exists. Let $i' < i$ be the least index in the cyclic ordering of $y_1, \dots, y_m$ $(y_{m+1} = y_1)$ such that $y_{i'}, y_{i'+1}, \dots, y_{i-1} \notin Y$. Then, using clauses $d_{k,j_k} = \overline{\ell^\star_{k,j_k}} \vee y_k \vee \overline{y}_{k+1}$ such that $\overline{\ell^\star_{k,j_k}} \in c$ as implied by $(*)$, we can remove the literals $\overline{y}_{i'}, \overline{y}_{i'+1}, \dots, \overline{y}_{i-1}$ from e while adding $\overline{\ell^\star_{i',j_{i'}}}$, $\overline{\ell^\star_{i'+1,j_{i'+1}}}$, …, $\overline{\ell^\star_{i-1,j_{i-1}}}$. After performing this for every $y_i \in Y$ in arbitrary order, we obtain a clause c^* with $P(c^*) = \{z\}$ such that $c^* \subseteq c \setminus \{\overline{q}\}$. However, this contradicts the minimality of c. Therefore, $Y = \emptyset$ must hold. It follows that item (ii) holds, which proves the lemma. □

Proof of Lemma 4.4 (continued). Recall that we use a set-oriented view of clauses, where disjunction symbols correspond to set unions and parentheses for singletons are occasionally omitted; accordingly, $c \setminus \ell$ denotes the clause c with literal ℓ removed.

(Basis) Let (c_1, c_2) the least pair such that $\overline{q} \in c_1$, $\overline{q} \notin c_2$ and the conditions of the lemma are satisfied, and suppose that c_1 and c_2 resolve on literal pair of ℓ and $\overline{\ell}$, where $\ell \in c_1$. Note that $\ell \notin \{q, \overline{q}\}$.

If $c_1 \setminus \overline{q}$ is an implicate of φ, then $(c_1 \setminus \overline{q}) \oplus c_2 \subseteq c \setminus \overline{q}$ is an implicate of φ, and condition (i) holds. Otherwise, each prime implicate $c'_1 \subseteq c_1$ contains $\overline{q}$, and thus $c'_1 \in S$. Suppose that c_1 is not prime. Then, if $\ell \in c'_1$, the pair (c'_1, c_2) satisfies the conditions of the lemma; since $c'_1 > c_1$, this contradicts the minimality of (c_1, c_2). On the other hand, if $\ell \notin c'_1$, then $c'_1 \subseteq c$ and condition (i) holds. For the rest, assume that c_1 is prime.

If $c_2 \in \varphi$, then Algorithm N-EXPLANATIONS considers $c_1 \oplus c_2$, and the statement holds. If

$c_2 \notin \varphi$, consider any proof of c_2 as in Lemma 4.3, such that $c_2 = c_{2,1} \oplus c_{2,2}$. Without loss of generality, $\overline{\ell} \in c_{2,1}$. Hence, $c_1 \oplus c_{2,1}$ exists. Since $c_{2,1} < c_2$, the minimality of (c_1, c_2) implies that the conditions of the lemma do not hold for $(c_1, c_{2,1})$; hence condition 3 of $(*)$ must be violated. This implies that $c_{2,1}$ is not prime and of form $c_{2,1} = c'_{2,1} \cup \{\ell'\}$ where $c'_{2,1}$ is the unique prime implicate contained in $c'_{2,1}$, and that $c_1 \oplus c'_{2,1} = c_1 \oplus c_{2,1}$. Note that $\ell' \in c_1$. Furthermore, by the structure of the proof of c_2, the clauses $c_{2,1}$ and $c_{2,2}$ must resolve on $\ell', \overline{\ell'}$, which implies $\overline{\ell'} \in c_{2,2}$. Therefore, also c_1 and $c_{2,2}$ resolve on $\ell', \overline{\ell'}$. Note that $c_{2,2}$ must be prime. The pair $(c_1, c_{2,2})$ thus satisfies $(*)$. Note that the acyclicity of φ implies $\ell' \neq q$. If $\ell' = \overline{q}$, then $(c_1 \oplus c_{2,2}) \oplus c_{2,1} =: d$ exists and $d \subseteq c \setminus \overline{q}$, and hence condition (i) holds. Otherwise, c_1 and $c_{2,2}$ satisfy the conditions of the lemma. However, since $c_{2,2} < c_2$, this raises a contradiction to the minimality of (c_1, c_2). This proves the base case.

(Induction) Suppose the statement holds for all pairs (c'_1, c'_2) ordered before (c_1, c_2) and consider the pair (c_1, c_2). Depending on the structure of c_1 and c_2, we consider the following cases.
Case DN1. $c = \overline{q} \vee \overline{a} \vee \overline{b}\,(\vee x_i)$, $c_1 = \overline{q} \vee \overline{a} \vee \overline{x}_k\,(\vee x_i)$, and $c_2 = \overline{b} \vee x_k$.

We consider the different cases depending on the primality of c_1 and c_2.
DN1.1: c_1 is prime while c_2 is not prime, and thus $c_2 = c'_2 \cup \{\ell\}$, where c'_2 is prime. If $\ell = x_k$, then $c'_2 = \overline{b} \subseteq c \setminus \overline{q}$ and thus condition (i) holds. For the rest, assume that $\ell \neq x_k$. Note that ℓ is a negative literal, c_1 and c'_2 resolve, and $c_1 \oplus c'_2 \subseteq c$ holds. Since condition 3 of $(*)$ excludes $c_1 \oplus c'_2 = c$, we must have $c_1 \oplus c'_2 = c \setminus \ell$. This means $\ell \notin c_1$.

Reconsidering the clauses c, c_1, and c_2, they have form $c = \overline{q} \vee \overline{\alpha} \vee \overline{\beta} \vee \ell\,(\vee x_i)$, $c_1 = \overline{q} \vee \overline{\alpha} \vee \overline{\ell}_0\,(\vee x_i)$, and $c_2 = \underbrace{\overline{\beta} \vee \ell_0}_{=c'_2} \vee \ell$, where c'_2 is a prime implicate of φ and ℓ_0 is a positive literal. Clause c_2 in turn, is generated by resolving clauses $c_{2,1}$ and $c_{2,2}$ in a proof of c'_2 as in Lemma 4.3.

If both $c_{2,1}$ and $c_{2,2}$ are prime, then the argument is the same as in case DN1.3 below where c_2 is prime and generated by resolving prime clauses $c_{2,1}$ and $c_{2,2}$.

Otherwise, w.l.o.g we have that $c_{2,1} = \overline{\beta}_1 \vee \overline{\ell'} \vee \ell$ and $c_{2,2} = \overline{\beta} \vee \ell_0 \vee \ell'$, where $c_{2,1}$ is prime. Note that $c_{2,1}, c_{2,2} < c_2$ and that ℓ' must be a negative literal. Furthermore, we can see that $\ell, \ell', \overline{\ell'} \notin c'_2$; $\overline{\beta}_1 \subseteq \overline{\beta}$, $\overline{\ell'} \notin c_1$; $\overline{q} \notin \overline{\beta}$; $q \notin \{\ell, \overline{\ell}, \ell_0, \ell'\}$, since otherwise c_1 and $c_{2,2}$ have two different resolving pairs of literals, which contradicts the acyclicity.

Suppose that $q = \overline{\ell'}$. Then, $(c_1 \oplus c_{2,1}) \oplus c'_2 = (\overline{\beta}_1 \vee \overline{\alpha} \vee \overline{\ell}_0 \vee \ell\,(\vee x_i)) \oplus c'_2 = \overline{\beta}_1 \vee \overline{\alpha} \vee \overline{\beta} \vee \ell\,(\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ (note that $\ell \in c$), and thus condition (i) holds. Thus, for the rest we assume that $q \neq \overline{\ell'}$.

According to the membership of ℓ' in c_1, we consider the possible cases:

Case 1) $\ell' \in c_1$: Then, c_1 and $c_{2,1}$ resolve on $\ell', \overline{\ell'}$ and $c_1 \oplus c_{2,1} = \overline{q} \vee (\overline{\alpha} \setminus \ell') \vee \overline{\beta}_1 \vee \ell \vee \overline{\ell}_0\,(\vee x_i)$. Both c_1 and $c_{2,1}$ are prime and satisfy $(*)$. Moreover, $q \notin N(c_{2,1})$. Hence, by the induction hypothesis, we have

(1.1) $d = \underbrace{(\overline{\alpha} \setminus \ell')}_{=\overline{\alpha}_0} \vee \overline{\beta}_1 \vee \ell \vee \overline{\ell}_0\,(\vee x_i)$ is an implicate of φ, or

(1.2) $\overline{q} \vee \overline{\alpha}_0 \vee \overline{\beta}_1 \vee \ell \vee \overline{\ell}_0\,(\vee x_i)$ contains a $\overline{q}$-clause $e \in S$.

In case (1.1), we have $d \oplus c_2 = \overline{\alpha}_0 \vee \overline{\beta}_1 \vee \ell \vee \overline{\beta}\,(\vee x_i) \subseteq \overline{\alpha} \vee \overline{\beta} \vee \ell\,(\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ, and thus condition (i) holds. In case (1.2), condition (ii) holds if $\overline{\ell}_0 \notin e$, since then $e \subseteq c$. Otherwise, e resolves with c_2' on $\ell_0, \overline{\ell}_0$ and $(*)$ is satisfied for e and c_2'. Note that $c_1 < e$, $\overline{q} \in e$. and $\overline{q} \notin c_2'$. Thus, by the induction hypothesis, we can show that $e \oplus c_2' \subseteq \overline{q} \vee \overline{\alpha} \vee \ell \vee \overline{\beta}\,(\vee x_i)$ satisfies the claim.

Case 2) $\ell' \notin c_1$: This implies that $c_1 \oplus c_2' \subset c_1 \oplus c_{2,2}$. Since c_2' is the unique prime implicate contained in $c_{2,2}$, the clauses c_1 and $c_{2,2}$ satisfy $(*)$. Moreover, $\overline{q} \notin c_{2,2}$ (note that, by assumption, $\ell' \neq \overline{q}$). Hence, e and $c_{2,2}$ satisfy the conditions of the lemma. Since $c_{2,2} < c_2$, by the induction hypothesis $c_1 \oplus c_{2,2}$ satisfies

(2.1) $d = \overline{\alpha} \vee \overline{\beta} \vee \ell'\,(\vee x_i)$ is an implicate of φ, or

(2.2) $\overline{q} \vee \overline{\alpha} \vee \overline{\beta} \vee \ell'\,(\vee x_i)$ contains a $\overline{q}$-clause $e \in S$.

In case (2.1), d resolves with $c_{2,1}$, and $d \oplus c_{2,1} = \overline{\alpha} \vee \overline{\beta} \vee \overline{\beta}_1 \vee \ell\,(\vee x_i)$ is an implicate contained in $c \setminus \overline{q}$, thus condition (i) holds.

In case (2.2), if $\ell' \notin e$, then $e \subseteq c$, and condition (ii) holds. Otherwise, if $\ell' \in e$, then e resolves with $c_{2,1}$ on $\ell', \overline{\ell'}$, and $e \oplus c_{2,1} \subseteq \overline{q} \vee \overline{\alpha} \vee \overline{\beta} \vee \overline{\beta}_1 \vee \ell\,(\vee x_i) \subseteq c$. Both e and $c_{2,1}$ are prime, and they satisfy $(*)$. Moreover, $\overline{q} \notin c_{2,1}$. Since $c_1 < e$, thus the claim holds by the induction hypothesis.

DN1.2: c_1 $(= \overline{q} \vee \overline{a} \vee \overline{x}_k\,(\vee x_i))$ is not prime while c_2 $(= \overline{b} \vee x_k)$ is prime.

Then, c_1 is form $c_1 = c_1' \cup \{\ell\}$, where c_1' is the unique prime implicate contained in c_1. If $\ell = \overline{q}$, then $c_1' \oplus c_2 = \overline{a} \vee \overline{b}\,(\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ, and condition (i) holds. Next, if $\ell = \overline{x}_k$, then $c_1' = \overline{q} \vee \overline{a}\,(\vee x_i) \in S$, and thus condition (ii) holds. Finally, if $\ell \in \overline{a} \cup \{x_i\}$, then $c_1' \in S$, and c_1' and c_2 satisfy $(*)$ and $c_1' \oplus c_2 \subseteq c$. Since $c_1 < c_1'$, by the induction hypothesis, the claim follows.

DN1.3: both c_1 $(= \overline{q} \vee \overline{a} \vee \overline{x}_k\,(\vee x_i))$ and c_2 $(= \overline{b} \vee x_k)$ are prime.

Then, $c_1 \in S$ holds by assertion. If $c_2 \in \varphi$, then $c_1 \oplus c_2$ is considered by Algorithm N-EXPLANATIONS and the claimed statement holds. Otherwise, c_2 is the resolvent of clauses $c_{2,1} = \overline{b}_1 \vee x_k \vee \overline{y}$ and $c_{2,2} = \overline{b}_2 \vee y$ in some proof of c_2 as in Lemma 4.3. At least one of $c_{2,1}$ and $c_{2,2}$ is prime. Since $c_{2,2}$ differs from c_2 in at least two literals (as $y \in c_{2,2} \setminus c_2$ and $x_k \in c_2 \setminus c_{2,2}$), it must be prime. Hence, if $c_{2,1}$ is not prime, it has form $c_{2,1} = \overline{b} \vee x_k \vee \overline{y}$ (i.e., $\overline{b}_1 = \overline{b}$).

We remark that in what follows, the primality of c_2 is not used for the argument where both $c_{2,1}$ and $c_{2,2}$ are prime. Thus, the proof in this case also applies to DN1.1 where c_2 is not prime but the resolvent of prime implicates $c_{2,1}$ and $c_{2,2}$.

We note that $y \neq x_i$ if $x_i \in c_1$, since otherwise c_1 and $c_{2,1}$ violate the acyclicity of φ.

Consider the case in which $y = q$. Then, $c_{2,2} = \overline{b}_2 \vee q$, and c_1 and $c_{2,2}$ resolve yielding $c_1 \oplus c_{2,2}$ $= \overline{a} \vee \overline{x}_k \vee \overline{b}_2\,(\vee x_i)$. Resolving this clause with c_2, we obtain $\overline{a} \vee \overline{b}_2 \vee \overline{b}\,(\vee x_i) = \overline{a} \vee \overline{b}\,(\vee x_i) \subseteq c \setminus \overline{q}$. Thus condition (i) holds. For the rest of case D1.3, we thus assume $y \neq q$.

The clauses c_1 and $c_{2,1}$ resolve and yield $c_1 \oplus c_{2,1} = \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee \overline{y}\,(\vee x_i) =: d$. If c_1 and $c_{2,1}$ violate $(*)$, then condition 3 of it must be violated. This implies that $c_{2,1}$ is not prime and that $c_1 \oplus (c_{2,1} \setminus \overline{y}) = d$ since $c_{2,1} = c_2 \cup \overline{y}$ in this case. Hence, $\overline{y} \in c_1$, which means that c_1 and $c_{2,2}$ resolve yielding $c_1 \oplus c_{2,2} = \overline{q} \vee (\overline{a} \setminus \overline{y}) \vee \overline{x}_k \vee \overline{b}_2\,(\vee x_i)$. Moreover, c_1 and $c_{2,2}$ satisfy $(*)$, and $q \notin N(c_{2,2})$. Since $c_{2,2} < c_2$, by the induction hypothesis we obtain

(1) $f = (\overline{a} \setminus \overline{y}) \vee \overline{x}_k \vee \overline{b}_2\,(\vee x_i)$ is an implicate of φ, or

(2) $\overline{q} \vee (\overline{a} \setminus \overline{y}) \vee \overline{x}_k \vee \overline{b}_2\,(\vee x_i)$ contains a $\overline{q}$-clause $e \in S$.

In case (1), we obtain that $f \oplus c_2 = \overline{a} \vee \overline{b}\,(\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ. Thus, condition (i) holds. In case (2), we obtain $e \subseteq c$. If $\overline{x}_k \notin e$, condition (ii) holds. Otherwise, the pair of e and c_2 satisfies $(*)$ and, moreover, the conditions of the lemma. Since $e \oplus c_2 \subseteq c.$ and $c_1 < e$, the claim thus follows from the induction hypothesis.

Finally, let us assume that c_1 and $c_{2,1}$ satisfy $(*)$. Then, by the induction hypothesis, we have

(1) $e := \overline{a} \vee \overline{b}_1 \vee \overline{y}\,(\vee x_i)$ is an implicate of φ, or

(2) $c_1 \oplus c_{2,1} = \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee \overline{y}\,(\vee x_i)$ contains a $\overline{q}$-clause d from S.

In case (1), $e \oplus c_{2,2} = \overline{a} \vee \overline{b}_1 \vee \overline{b}_2\,(\vee x_i) = \overline{a} \vee \overline{b}\,(\vee x_i)$ is an implicate of φ, and thus condition (i) is true.

In case (2), if $\overline{y} \notin d$, then $d \subseteq c$ and condition (ii) holds. Otherwise, if $\overline{y} \in d$, then d resolves with $c_{2,2}$ on $y, \overline{y}$ and $d \oplus c_{2,2} \subseteq \overline{q} \vee \overline{a} \vee \overline{b}\,(\vee x_i) \subseteq c$. The clauses d and $c_{2,2}$ satisfy $(*)$, $\overline{q} \in d \oplus c_{2,2}$ and $\overline{q} \notin c_{2,2}$. Hence they satisfy the conditions of the lemma, and since $c_1 < d$, we obtain by the induction hypothesis that the claim holds.

This concludes the proof of Case DN1.

Case DN2: $c_1 = \overline{q} \vee \overline{a} \vee x$, $c_2 = \overline{b} \vee \overline{x}\,(\vee x_i)$, and $c = \overline{q} \vee \overline{a} \vee \overline{b}\,(\vee x_i)$.

Suppose that c_1 is not prime. Then, it is of form $c_1 = c_1' \cup \{\ell\}$, where c_1' is the unique prime implicate contained in c_1. If $\ell = \overline{q}$, then $c_1' \oplus c_2 = \overline{a} \vee \overline{b}\,(\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ, and thus condition (i) holds. If $\ell = x$, then c contains $\overline{q} \vee \overline{a} \in S$ and condition (ii) holds. Otherwise (i.e., $\ell \neq \overline{q}, x$), consider c_1' and c_2. Then c_1' and c_2 satisfy $(*)$ and since $\overline{q} \in c_1' \oplus c_2$ also the conditions of the lemma. Since $c_1' > c_1$, by the induction hypothesis, the claim holds.

Thus, for the rest assume that c_1 is prime.

If $c_2 \in \varphi$, then $c_1 \oplus c_2$ is considered by Algorithm N-EXPLANATIONS and thus the claim holds. Otherwise, c_2 is the resolvent of clauses $c_{2,1} = \overline{b}_1(\vee\overline{x}) \vee y$ and $c_{2,2} = \overline{b}_2(\vee\overline{x}) \vee \overline{y}\,(\vee x_i)$ as in Lemma 4.3, where $\overline{x}$ occurs in at least one of c_1 and c_2. Note that $x \neq q$ holds (otherwise, $q, \overline{q} \in c_1$ would hold).

We first show that the claim holds if $y = q$. We must have $\overline{x} \notin c_{2,1}$; otherwise, we have $x < q$ via $c_{2,1}$ and $q < x$ via c_1, which contradicts the acyclicity of φ. Thus, $(c_1 \oplus c_{2,1}) \oplus c_2 = (\overline{a} \vee \overline{b}_1 \vee x) \oplus (\overline{b} \vee \overline{x}\,(\vee x_i)) = \overline{a} \vee \overline{b}\,(\vee x_i) \subseteq c \setminus \overline{q}$, and thus condition (i) holds.

Thus, for the rest we assume $y \neq q$, and consider possible occurrences of $\overline{x}$ in $c_{2,1}$ resp. $c_{2,2}$.

Case DN2.1: $\overline{x} \in c_{2,1}$. Then, $c_1 \oplus c_{2,1}$ exists and yields $c_1 \oplus c_{2,1} = \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee y$. We show that c_1 and $c_{2,1}$ satisfy $(*)$. Towards a contradiction, suppose it fails. Then, condition 3 of $(*)$ must be violated, which implies that $c_{2,1}$ is not prime. Consequently, it must be of form $c_{2,1} = c_{2,1}' \cup \{\ell\}$ where $c_{2,1}'$ is the unique prime implicate contained in $c_{2,1}$. Note that $y \neq x_i$ when $x_i \in c_2$. Therefore, we have $y \notin c_2$. It follows $\ell = y$ and $c_{2,1}' = c_{2,1} \setminus y$. However, $c_1 \oplus c_{2,1} \neq c_1 \oplus c_{2,1}'$, since $\overline{x} \in c_{2,1}$ implies $x \neq y$ and thus $y \notin c_1$. Thus, condition 3 of $(*)$ cannot be violated, which is a contradiction. It follows that c_1 and $c_{2,1}$ satisfy $(*)$. Moreover, since $\overline{q} \in c_1 \oplus c_{2,1}$ $(=: d)$ and $\overline{q} \notin c_{2,1}$, they satisfy the conditions of the lemma. Since $c_{2,1} < c_2$, from the induction hypothesis (case DN2), we have

(A) $d \setminus \overline{q}$ is an implicate of φ, or

(B) d contains a $\overline{q}$-clause $e \in S$.

In case (A), $d' = d \setminus \overline{q} = \overline{a} \vee \overline{b}_1 \vee y$ is an implicate of φ, and so is $e' = d' \oplus c_{2,2} = \overline{a} \vee \overline{b}_1 \vee \overline{b}_2 (\vee \overline{x})\, (\vee x_i)$. If $\overline{x} \notin c_{2,2}$, then $e' \subseteq c \setminus \overline{q}$ and condition (i) holds. Consider thus $\overline{x} \in c_{2,2}$. If $d' \setminus y$ is an implicate of φ, then condition (i) holds. Thus, in what follows we assume that each prime implicate $f \subseteq d'$ contains y.

Since $\overline{x} \in c_{2,2}$, the clauses c_1 and $c_{2,2}$ resolve on $x, \overline{x}$ and yield $c_1 \oplus c_{2,2} = \overline{q} \vee \overline{a} \vee \overline{b}_2 \vee \overline{y}\, (\vee x_i)$. Moreover, c_1 and $c_{2,2}$ satisfy $(*)$. To see this, note that otherwise condition 3 of $(*)$ must fail. If $c_{2,2}$ is not prime, then it is of form $c_{2,2} = c'_{2,2} \cup \{\ell\}$ where $c'_{2,2}$ is the unique prime implicate contained in $c_{2,2}$ and c_2 has either form $c_2 = c'_{2,2} \cup \{\ell'\}$ or $c_2 = c'_{2,2}$. Since $\overline{y} \notin c_2$, it follows that $\ell = \overline{y}$. Therefore, violation of condition 3) of $(*)$ implies that $c_1 \oplus c'_{2,2} = c_1 \oplus c_{2,2}$ and $\overline{y} \in c_1$. But then $y < x$ via c_1 and $x < y$ via $c_{2,1}$, which contradicts the acyclicity of φ. Hence, $(*)$ must hold for c_1 and $c_{2,2}$. Moreover, $c_{2,2} < c_2$. Thus, by the induction hypothesis, we have

(A.1) $\overline{a} \vee \overline{b}_2 \vee \overline{y}\, (\vee x_i)$ is an implicate of φ, or

(A.2) $\overline{q} \vee \overline{a} \vee \overline{b}_2 \vee \overline{y}\, (\vee x_i)$ contains a $\overline{q}$-clause $d'' \in S$.

In case (A.1), by resolution with $d' = \overline{a} \vee \overline{b}_1 \vee y$ we can see that $\overline{a} \vee \overline{b}_1 \vee \overline{b}_2\, (\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ, thus condition (i) holds. In case (A.2), if $\overline{y} \notin d''$, then $d'' \subseteq c$ and condition (ii) holds. Otherwise, some prime implicate $e' \subseteq d'$ exists which by assertion contains y. The pair of d'' and e' resolves on $y, \overline{y}$ and satisfies $(*)$, and, moreover $d'' \oplus e' \subseteq \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee \overline{b}_2\, (\vee x_i)$, such that $\overline{q} \in d'' \oplus e'$ (as $y \neq q$) and $\overline{q} \notin e'$. Hence d'' and e' satisfy the conditions of the lemma. Since $d'' > c_1$, by the induction hypothesis (case N1 resp. P2), it follows that the claim holds.

In case (B), if $y \notin e$ $(\subseteq \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee y)$, then $e \subseteq c$ and condition (ii) holds. Otherwise (i.e., $y \in e$), consider the membership of $\overline{x}$ in $c_{2,2}$.

If $\overline{x} \notin c_{2,2}$, consider e and $c_{2,2}$. The resolvent $e \oplus c_{2,2} \subseteq \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee \overline{b}_2\, (\vee x_i)$ exists. If $c_{2,2}$ contains some implicate f which does not contain $\overline{y}$, then $f \subseteq c \setminus \overline{q}$ and condition (i) holds. Otherwise, let $f \subseteq c_{2,2}$ be any prime implicate. The clauses e and f satisfy $(*)$ and $\overline{q} \in e \oplus f$, $\overline{q} \notin f$. Since $e \oplus f \subseteq c$ and $c_1 < e$, the claim follows from the induction hypothesis (appealing to case DN2).

If $\overline{x} \in c_{2,2}$, consider c_1 and $c_{2,2}$. Then $c_1 \oplus c_{2,2} = \overline{q} \vee \overline{a} \vee \overline{b}_2 \vee \overline{y}\, (\vee x_i) =: d$. Moreover, c_1 and $c_{2,2}$ satisfy $(*)$. Indeed, similar as above, we can conclude that violation of condition 3) of $(*)$ is only possible if $c_{2,2} = c'_{2,2} \cup \{\ell\}$ where $c'_{2,2}$ is prime and $\ell = \overline{y}$ such that $\overline{y} \in c_1$, which would contradict the acyclicity of φ. Furthermore, $\overline{q} \in d$ and $\overline{q} \notin c_{2,2}$. Since $c_{2,2} < c_2$, by the induction hypothesis, we have

(B.1) $d \setminus \overline{q} = \overline{a} \vee \overline{b}_2 \vee \overline{y}\, (\vee x_i)$ is an implicate of φ, or

(B.2) d contains a $\overline{q}$-clause $d' \in S$.

In case (B.1), $d \setminus \overline{q}$ contains some prime implicate d''. If $\overline{y} \notin d''$, then $d'' \subseteq c \setminus \overline{q}$ and condition (i) holds. Otherwise, e $(\subseteq \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee y)$ from (B) and d'' satisfy $(*)$ and $\overline{q} \in e \oplus d''$ and $\overline{q} \notin d''$. Since $e \oplus d'' \subseteq c$ and $c_1 < e$, the claim follows from the induction hypothesis.

In case (B.2), e and d' are prime implicates in S. If $\overline{y} \notin d'$, then $d' \subseteq c$ and condition (ii) holds. Otherwise, e and d' resolve and $e \oplus d' \subseteq c$; since $e \oplus d'$ is considered by Algorithm N-EXPLANATIONS, the claim holds. This concludes the proof of Case DN2.1.

Case DN2.2: $\overline{x} \notin c_{2,1}$. Then, $\overline{x} \in c_{2,2}$ must hold. We conclude that $c_{2,1} = \overline{b}_1 \vee y$ must be prime, since it differs from c_2 in at least two literals.

The clauses c_1 and $c_{2,2}$ resolve and $c_1 \oplus c_{2,2} = \overline{q} \vee \overline{a} \vee \overline{b}_2 \vee \overline{y}\,(\vee x_i)$. Consider property $(*)$ for them.

Assume first that c_1 and $c_{2,2}$ violate $(*)$. Then condition 3 of $(*)$ must be violated and hence $c_{2,2}$ must not be prime. Therefore, $c_{2,2} = c'_{2,2} \cup \{\ell\}$ where $c'_{2,2}$ is prime. Since $\overline{y} \in c_{2,2}$ but $\overline{y} \notin c_2$, it follows $\ell = \overline{y}$. Violation of condition 3 thus implies that $c_1 \oplus (c_{2,2} \setminus \overline{y}) = c_1 \oplus c_{2,2}$. Hence, $\overline{y} \in c_1$. This means c_1 and $c_{2,1}$ resolve and yield $c_1 \oplus c_{2,1} = \overline{q} \vee (\overline{a} \setminus \overline{y}) \vee \overline{b}_1 \vee x$. The clauses c_1 and $c_{2,1}$ satisfy $(*)$ (both are prime) and $\overline{q} \notin c_{2,1}$. Since $c_{2,1} < c_2$, by the induction hypothesis (case D1), we have

(C.1) $d = (\overline{a} \setminus \overline{y}) \vee \overline{b}_1 \vee x$ is an implicate of φ, or

(C.2) $\overline{q} \vee (\overline{a} \setminus \overline{y}) \vee \overline{b}_1 \vee x$ contains a $\overline{q}$-clause $d' \in S$.

In case (C.1), $d \oplus c_2 \subseteq \overline{a} \vee \overline{b}_1 \vee \overline{b}_2\,(\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ, and condition (i) holds.

In case (C.2), if $x \notin d'$, then $d' \subseteq c$ holds, and thus condition (ii) holds. If $c_2 \setminus \overline{x}$ contains an implicate f of φ, then $f \subseteq c \setminus \overline{q}$ and condition (i) holds. Otherwise, let $c'_2 \subseteq c_2$ be any prime implicate. Then d' and c'_2 resolve on $x, \overline{x}$. We note that d' and c'_2 satisfy $(*)$, and that $\overline{q} \in d' \oplus c'_2$ and $\overline{q} \notin c'_2$. Hence, the conditions of the lemma are satisfied. Since $c_1 < d'$ and $d' \oplus c'_2 \subseteq c$, from the induction hypothesis the claim follows. This shows that the claim holds if c_1 and $c_{2,2}$ violate $(*)$.

Assume finally that c_1 and $c_{2,2}$ satisfy $(*)$. Then, since $\overline{q} \notin c_{2,2}$ and $c_{2,2} < c_2$, by the induction hypothesis, we have

(D.1) $d = \overline{a} \vee \overline{b}_2 \vee \overline{y}\,(\vee x_i)$ is an implicate of φ, or

(D.2) $\overline{q} \vee \overline{a} \vee \overline{b}_2 \vee \overline{y}\,(\vee x_i)$ contains a $\overline{q}$-clause $d' \in S$.

In case (D.1), $d \oplus c_{2,1} = \overline{a} \vee \overline{b}_1 \vee \overline{b}_2\,(\vee x_i) \subseteq c \setminus \overline{q}$ is an implicate of φ, and thus condition (i) holds. In case (D.2), if $\overline{y} \notin d'$, then $d' \subseteq c$, and thus condition (ii) holds. Otherwise, d' resolves with $c_{2,1}$ and $d' \oplus c_{2,1} \subseteq \overline{q} \vee \overline{a} \vee \overline{b}_1 \vee \overline{b}_2\,(\vee x_i) = c$. The clauses d' and $c_{2,1}$ satisfy $(*)$ (both are prime) and $\overline{q} \in d' \oplus c_{2,1}$, $\overline{q} \notin c_{2,1}$. Since $d' > c_1$, by the induction hypothesis (case N1 resp. D1), the claim holds.

This concludes the proof of Case DN2.2, and thus also of DN2 and Lemma 4.4. □

References

[1] K. Apt and M. Bezem. Acyclic programs. *New Generation Computing*, 9(3-4):335–364, 1991.

[2] E. Boros, Y. Crama, and P. L. Hammer. Polynomial-time inference of all valid implications for Horn and related formulae. *Annals of Mathematics and Artificial Intelligence*, 1:21–32, 1990.

[3] G. Brewka, J. Dix, and K. Konolige. *Nonmonotonic Reasoning – An Overview*. Number 73 in CSLI Lecture Notes. CSLI Publications, Stanford University, 1997.

[4] T. Bylander. The monotonic abduction problem: A functional characterization on the edge of tractability. In *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning (KR-91)*, pages 70–77, 1991.

[5] L. Console, D. Theseider Dupré, and P. Torasso. On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1(5):661–690, 1991.

[6] J. de Kleer. An assumption-based truth maintenance system. *Artificial Intelligence*, 28:127–162, 1986.

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

[8] A. del Val. The complexity of restricted consequence finding and abduction. In *Proceedings AAAI-2000*, pages 337–342, 2000.

[9] W. Dowling and J. H. Gallier. Linear-time algorithms for testing the satisfiability of propositional Horn theories. *Journal of Logic Programming*, 3:267–284, 1984.

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

[11] T. Eiter and K. Makino. On computing all abductive explanations. In *Proceedings Eighteenth National Conference on Artificial Intelligence (AAAI '02), July 28 – August 1, 2002, Edmonoton, Alberta (Canada)*, pages 62–67. AAAI Press, July 2002.

[12] K. Eshghi. A tractable class of abduction problems. In *Proceedings IJCAI-93*, pages 3–8, Chambery, France, August 1993.

[13] G. Friedrich, G. Gottlob, and W. Nejdl. Hypothesis classification, abductive diagnosis, and therapy. In *Proceedings International Workshop on Expert Systems in Engineering*, LNCS/LNAI 462, pages 69–78, September 1990 1990.

[14] K. Inoue. Linear resolution for consequence finding. *Artificial Intelligence*, 56(2-3):301–354, 1992.

[15] H. Kleine Büning and T. Lettmann. *Aussagenlogik - Deduktion und Algorithmen*. B.G. Teubner, Stuttgart, 1994.

[16] A. Leitsch. *The Resolution Calculus*. Springer, 1997.

[17] P. Marquis. Consequence finding algorithms. In D. Gabbay and Ph.Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume V: Algorithms for Uncertainty and Defeasible Reasoning, pages 41–145. Kluwer Academic, 2000.

[18] C. S. Peirce. Abduction and induction. In J. Buchler, editor, *Philosophical Writings of Peirce*, chapter 11. Dover, New York, 1955.

[19] D. Poole. Explanation and prediction: An architecture for default and abductive reasoning. *Computational Intelligence*, 5(1):97–110, 1989.

[20] D. Poole. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64(1):81–130, 1993.

[21] R. Reiter and J. de Kleer. Foundations of assumption-based truth maintenance systems: Preliminary report. In *Proceedings AAAI-87*, pages 183–188, 1982.

[22] B. Selman and H. J. Levesque. Abductive and default reasoning: A computational core. In *Proceedings AAAI-90*, pages 343–348, July 1990.

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

[24] B. Zanuttini. New polynomial classes for logic-based abduction. *Journal of Artificial Intelligence Research*, 2003. To appear. Presented at *Huitièmes Journées Nationales sur la résolution de Problèmes NP-Complets"(JNPC'02)*, Nice, May 27-29, 2002.