Appendix – Proofs

Theorem 1
Let $\mathscr{T} = \langle \mathscr{P}, \mathscr{A}, \mathscr{C} \rangle$ be an abductive theory and (E,F) a constrained explanation of an observation O. Then every constant symbol occurring in (E,F) occurs in $\mathscr{T}$ or in O.

Proof
Since $F \subseteq \mathscr{P}$ (as required by the definition of an explanation), every constant occurring in F occurs in $\mathscr{P}$. If α is a constant appearing in E but not in $\mathscr{T}$ nor in O, then changing α to a fresh constant ξ results in an explanation[1] and so, contradicts the constrainedness of (E,F). $\square$

Theorem 2
For every abductive theory $\mathscr{T} = \langle \mathscr{P}, \mathscr{A}, \mathscr{C} \rangle$, where $\mathscr{P}$ is stratified and interpreted under the stable-model semantics, and for every observation O,

$$\Psi_A(O,\mathscr{T}) = \Psi_B(O,\mathscr{T}) = \Psi_C(O,\mathscr{T}) = \Psi_D(O,\mathscr{T}).$$

Proof
The assertion follows by the fact that a stratified program admits exactly one stable model. $\square$

Theorem 3
Let $\mathscr{A}$ be a set of abducible predicates, R a (fixed) stratified program with no abducible predicates in the heads of its rules, and $\mathscr{C}$ a (fixed) set of integrity constraints.

1. The following problem is in coNP: given a set B of abducibles, an observation O, and a pair $\Delta = (E,F)$ of sets of abducibles, decide whether Δ is a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \mathscr{C} \rangle$.
2. The following problem is in Σ_2^P: given a set B of abducibles and an observation O, decide whether a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \mathscr{C} \rangle$ exists.

Proof
(1) The complementary problem consists of deciding that (E,F) is not an explanation or that is an arbitrary explanation. The following non-deterministic polynomial-time algorithm decides this problem. Since R is stratified, one can compute the only stable model, say M, of $R \cup ((B \cup E) \setminus F)$ in time linear in the size of B and (E,F). If E and F are not disjoint (which can be checked efficiently), or if $M \not\models O$, or if $M \not\models \mathscr{C}$, the (E,F) is not an explanation. Otherwise, (E,F) is an explanation and we proceed as follows. We non-deterministically guess the set C of occurrences of some constant c occurring in E. We then compute (E',F) by replacing all occurrences of c mentioned in C with a fresh constant ξ and, in the same way as before, determine whether (E',F) is an explanation of O.

(2) If (E,F) is a constrained explanation, then E and F consist of abducibles involving only constants appearing in $\mathscr{T}$ and O (cf. Theorem **??**). It follows, that if (E,F) is a constrained explanation, the size of $E \cup F$ is polynomial in the size of the input. Thus, the problem can be decided by the following non-deterministic polynomial time algorithm with an oracle: guess sets E and F of abducibles and check that (E,F) is a constrained explanation. By (1), that task can be decided by a call to a coNP oracle. $\square$

[1] We tacitly assume here that the semantics of logic programs we consider here are insensitive to the renaming of constants. All standard semantics of programs have this property.

Theorem 4
Let $\mathscr{A}$ be a set of abducible predicates and R a (fixed) non-recursive program with no abducible predicates in the heads of its rules.

1. The following problem is coNP-complete: given a set B of abducibles, an observation O, and a pair (E,F) of sets of abducibles, decide whether (E,F) is a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \emptyset \rangle$.
2. The following problem is Σ_2^P-complete: given a set B of abducibles and an observation O, decide whether a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \emptyset \rangle$ exists.

Proof
(1) The membership part was established in Theorem **??**. Thus, it suffices to show the hardness part.

We note that the following version of the SAT problem is NP-complete (membership is evident, hardness follows by a straightforward reduction from SAT):

Input: A set of atoms Y and a CNF formula F over Y that is not satisfied by the all-false assignment
Question: Is F satisfiable (is the QBF formula $\exists Y\, F$ true)?

We will reduce that problem to the problem whether (under the notation in the statement of the theorem) an explanation (E,F) is arbitrary.

Let then Y be a set of atoms and F a CNF theory that is not satisfied by the all-false assignment on Y. We denote by $Cl(F)$ the set of clauses in F. Let us consider the vocabulary σ consisting of predicate symbols $bad/0$, $in_Y/1$, $clause/1$, $pos/2$, $ngtd/2$, $choose/2$, $gate/1$, $true/1$, $holds/1$, $sometrue/0$, $allfalse/0$, $sat/0$, $clfalse/0$ and $goal/0$, and an abductive theory

$$\mathscr{T}(F) = \langle T(F), \{choose\}, \emptyset \rangle,$$

where $T(F) = R \cup B$, B consists of the atoms

1. $in_Y(a)$, for every $a \in Y$
2. $gate(0)$, where $0 \notin Y$
3. $pos(a,c)$, for every atom $a \in Y$ and clause $c \in Cl(F)$ such that a occurs non-negated in c
4. $ngtd(a,c)$, for every atom $a \in At(F)$ and clause $c \in Cl(F)$ such that a occurs negated in c

and R consists of the rules

1. $clause(C) \leftarrow pos(A,C)$
2. $clause(C) \leftarrow ngtd(A,C)$
3. $true(A) \leftarrow in_Y(A), gate(W), \textbf{not}\ choose(A,W)$
4. $holds(C) \leftarrow pos(A,C), true(A)$
5. $holds(C) \leftarrow ngtd(A,C), \textbf{not}\ true(A)$
6. $clfalse \leftarrow clause(C), \textbf{not}\ holds(C)$
7. $sat \leftarrow \textbf{not}\ clfalse$
8. $sometrue \leftarrow in_Y(A), true(A)$
9. $allfalse \leftarrow \textbf{not}\ sometrue$
10. $bad \leftarrow choose(A,W), \textbf{not}\ in_Y(A)$
11. $goal \leftarrow allfalse, \textbf{not}\ bad$
12. $goal \leftarrow sat, \textbf{not}\ bad.$

Let $\{goal\}$ be the set of observed atoms. It is clear that $U = (\{choose(a,0)\colon a \in Y\}, \emptyset)$ is an explanation (*goal* is derived through the first of its two rules). If F is satisfiable, then let $Y' \subseteq Y$ be (the representation of) an assignment that satisfies F. One can check that $(\{choose(a,0)\colon a \in (Y \setminus Y')\} \cup \{choose(a,\xi)\colon a \in Y'\}, \emptyset)$ is an explanation (now, *goal* can be derived via its second rule). Moreover, $Y' \neq \emptyset$ (by our restriction on the class of formulas). Thus, $(\{choose(a,0)\colon a \in Y\}, \emptyset)$ is arbitrary.

Conversely, let us assume that $U = (\{choose(a,0)\colon a \in Y\}, \emptyset)$ is arbitrary. Then replacing some occurrences of one of the constants must yield an explanation. Replacing a constant $a \in Y$ with fresh constant symbol ξ does not yield an explanation. Indeed, we would have $choose(\xi,0)$ and no $in_Y(\xi)$ in the "add" part of the explanation. Thus, *bad* would hold and would block any possibility of deriving *goal*. It follows that one or more occurrences of 0 can be replaced by ξ so that the result, $(\{choose(a,0)\colon a \in (Y \setminus Y')\} \cup \{choose(a,\xi)\colon a \in Y'\}, \emptyset)$, is an explanation of *goal*. Here $Y' \subseteq Y$ is the *non-empty* set of elements in Y identifying the occurrences of 0 replaced by ξ. Since $Y' \neq \emptyset$, *goal* is derived via the second rule. It follows that *sat* is derivable and so, every clause in F *holds* in the interpretation that assigns true to all elements of Y' and false to all other elements of Y. Thus, F is satisfiable.

It follows that deciding whether an explanation (E,F) is arbitrary is NP-hard. Since every explanation is either arbitrary or constrained, the problem to decide whether (E,F) is constrained is coNP-hard.

(2) As before, the membership part of the assertion follows from Theorem **??**. To prove the hardness part, we note that the following problem is Σ_2^P-complete:

Input: Two disjoint sets X and Y of atoms, and a DNF formula G over $X \cup Y$ such that for every truth assignment v_X to atoms in X, the *all-false* assignment to atoms in Y is a model of formula $G|_{v_X}$

Question: Is the quantified boolean formula $\Phi = \exists X \forall Y\, G$ true.

We will reduce it to our problem.

Let F be the CNF obtained from $\neg G$ by applying the De Morgan's and the double negation laws. Clearly, $F \equiv \neg G$. Let $Cl(F)$ be the set of clauses of F. Let us consider the vocabulary σ consisting of predicate symbols in_X, in_Y, $clause/1$, $pos/2$, $ngtd/2$, $choose/2$, $gate/1$, $true_X/1$, $true_Y/1$, $true/1$, $holds/1$, $sometrue/0$, $allfalse/0$, $sat/0$, $clfalse/0$, $bad/0$, $good/1$, and $goal/0$, and an abductive theory

$$\mathscr{T}(F) = \langle T(F), \{true_X, choose\}, \emptyset \rangle,$$

where $T(F) = R \cup B$, B consists of the atoms:

1. $in_X(a)$, for every $a \in X$
2. $in_Y(a)$, for every $a \in Y$
3. $gate(0)$, where $0 \notin Y$
4. $pos(a,c)$, for every atom $a \in X \cup Y$ and clause $c \in Cl(F)$ such that a occurs non-negated in c
5. $ngtd(a,c)$, for every atom $a \in X \cup Y$ and clause $c \in Cl(F)$ such that a occurs negated in c

and R consists of the rules

1. $clause(C) \leftarrow pos(A,C)$
2. $clause(C) \leftarrow ngtd(A,C)$
3. $true_Y(A) \leftarrow in_Y(A), gate(W),$ **not** $choose(A,W)$

4. $true(A) \leftarrow true_X(A)$
5. $true(A) \leftarrow true_Y(A)$
6. $holds(C) \leftarrow pos(A,C), true(A)$
7. $holds(C) \leftarrow ngtd(A,C)$, **not** $true(A)$
8. $clfalse \leftarrow clause(C)$, **not** $holds(C)$
9. $sat \leftarrow$ **not** $clfalse$
10. $sometrue \leftarrow in_Y(A), true_Y(A)$
11. $allfalse \leftarrow$ **not** $sometrue$
12. $bad \leftarrow choose(A,W)$, **not** $in_Y(A)$
13. $bad \leftarrow true_X(A)$, **not** $in_X(A)$
14. $good(A) \leftarrow in_Y(A), choose(A,W)$
15. $bad \leftarrow in_Y(A)$, **not** $good(A)$
16. $goal \leftarrow allfalse$, **not** bad
17. $goal \leftarrow sat$, **not** bad.

Let $O = \{goal\}$ be an observation. We will prove that Φ is true if and only if *goal* has a constrained explanation from $\mathscr{T}(F)$.
($\Rightarrow$) Let v_X be an assignment of truth values to variables in X such that the formula $\forall Y\, G|_{v_X}$ is true. Here by $G|_{v_X}$ we denote the formula obtained from G by substituting the truth values given by v_X for the corresponding variables from X, and then by simplifying these values away. We understand the formula $F|_{v_X}$ in the same way. Clearly, $F|_{v_X} \equiv \neg G|_{v_X}$. Thus, $\exists Y F|_{v_X}$ is false.
Let us define

$$E = \{true_X(a) : a \in X \text{ and } v_X(a) = true\} \cup$$
$$\{choose(a,0) \colon a \in Y\}.$$

We will show that $(E, \emptyset)$ is a constrained explanation of *goal*. First, it is evident that $(E, \emptyset)$ is an explanation as *goal* can be derived through the first of its two rules. Next, we note that we cannot replace any constant a appearing in atoms $true_X(a)$ with a new constant ξ. Indeed, if $true_X(\xi)$ were to be a part of a the “add” part of an explanation, *bad* would hold (via the rule (13)) and *goal* would not! Similarly, we cannot replace $a \in Y$ in any atom $choose(a,0)$, as only elements of Y must show on these positions, the property forced by rule (12). Finally, we cannot replace any non-empty set of 0’s with ξ. If any such replacement resulted in an explanation, *goal* could only be derived through its second clause (*allfalse* cannot be derived now). However, that would imply that $\exists Y F|_{v_X}$ is true, with the “witness” assignment assigning *true* to every $y \in Y$ such that $choose(y, \xi)$ is a part of the modified explanation, and *false* to all other elements of Y.
($\Leftarrow$) Let us assume that *goal* has a constrained explanation. It must have a form $(E, \emptyset)$, where

$$E = \{true_X(a) : a \in U\} \cup \{choose(a,b) \colon a \in Y, b \in U_a\},$$

where U is some subset of X and where for every $a \in Y$, U_a, is some *nonempty* set. Indeed, if for some $a \in Y$, there is no b such that $choose(a,b) \in E$, $good(a)$ cannot be derived from the revised program and, consequently, *bad* would follow (by the rule (15)). That would make it impossible to derive *goal*.

Now, if for some $a \in Y$ there is $b \in U_a$ such that $b \neq 0$, then $(E, \emptyset)$ is not constrained (indeed, that constant b could be replaced by a new constant ξ without any effect on the derivability of *goal*). Thus, for every $a \in Y$, $U_a = \{0\}$ and so,

$$E = \{true_X(a) : a \in U\} \cup \{choose(a,0) \colon a \in Y\}.$$

Since this explanation is constrained, there is no subset of positions where 0 occurs that can be substituted with ξ. Therefore, $\exists Y F|_{v_X}$, where v_X is the truth assignment determined by U, is false. One can show that by following the argument used in part (1) of the theorem (due to our assumption on G, the all-false assignment to atoms in Y is not a model of $F|_{v_X}$). Thus, Φ is true. $\square$

Theorem 5
Let $\mathscr{A}$ be a set of abducible predicates and R a (fixed) Horn program with no abducible predicates in the heads of its rules.

1. The following problem is coNP-complete: given a set B of abducibles, an observation O, and a pair (E,F) of sets of abducibles, decide whether (E,F) is a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \emptyset \rangle$.
2. The following problem is Σ_2^P-complete: given a set B of abducibles and an observation O, decide whether a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \emptyset \rangle$ exists.

Proof
(1) The membership part follows by Theorem **??**. To prove hardness, we show that the problem to decide whether (E,F) is arbitrary is NP-hard. That is sufficient, as every explanation is either arbitrary or constrained. To show NP-hardness of the problem to decide whether an explanation is arbitrary, we reduce the SAT problem to it.

Thus, let Y be a (finite) set of atoms, say $Y = \{y_1, \ldots, y_n\}$, and F a CNF consisting of clauses $c_1, \ldots, c_m$. We denote by $Cl(F)$ the set of clauses in F, that is, $Cl(F) = \{c_1, \ldots, c_m\}$. Let us also consider three additional distinct symbols t, f and 0. We define the vocabulary σ to consist of predicate symbols $in_Y/1$, $clause/1$, $pos/2$, $ngtd/2$, $p/2$ $true/1$, $false/0$, $ok/1$, $next/2$, $next_C/2$, $clsat/0$, $sat/1$, and $goal/0$, and an abductive theory

$$\mathscr{T}(F) = \langle T(F), \{p\}, \emptyset \rangle,$$

where $T(F)$ consists of the following atoms:

1. $in_Y(a)$, for every $a \in Y \cup \{t, f\}$
2. $pos(a,c)$, for every atom $a \in Y$ and clause $c \in Cl(F)$ such that a occurs non-negated in c
3. $ngtd(a,c)$, for every atom $a \in Y$ and clause $c \in Cl(F)$ such that a occurs negated in c
4. $next(y_i, y_{i+1})$, for $i = 1, \ldots, n-1$, $next(t, y_1)$, and $next(y_n, f)$
5. $next_C(c_i, c_{i+1})$, for $i = 1, \ldots, m-1$
6. $p(t, 0)$

and of the following rules

1. $clause(C) \leftarrow pos(A,C)$
2. $clause(C) \leftarrow ngtd(A,C)$
3. $true(A) \leftarrow in_Y(A), p(A,Z), p(t,Z)$
4. $false(A) \leftarrow in_Y(A), p(A,Z), p(f,Z)$
5. $clsat(C) \leftarrow pos(A,C), true(A)$
6. $clsat(C) \leftarrow ngtd(A,C), false(A)$
7. $ok(t)$
8. $ok(A) \leftarrow ok(A'), next(A',A), true(A)$
9. $ok(A) \leftarrow ok(A'), next(A',A), false(A)$
10. $ok(f) \leftarrow ok(A'), next(A',f)$

11. $sat(c_1) \leftarrow clsat(c_1)$
12. $sat(C) \leftarrow sat(C'), next_C(C', C), clsat(C)$
13. $goal \leftarrow ok(f), sat(c_m), p(f, Z)$.

Clearly, the pair $(E, \emptyset)$, where $E = \{p(x,0) \colon x \in Y \cup \{f\}\}$, is an explanation of $goal$. Indeed, for every $x \in Y$, both $true(x)$ and $false(x)$ can be derived from $T(F) \cup E$ (because $p(t,0)$ and $p(f,0)$ both hold in $T(F) \cup E$). Thus, for every clause c, $clsat(C)$ can be derived, too. These two observations imply that $ok(f)$ and $sat(c_m)$ can both be derived from $T(F) \cup E$. Consequently, $goal$ is explained by $(E, \emptyset)$.

Let us assume that E is arbitrary. We will prove that F is satisfiable. By the definition, one of the constants appearing in E can be replaced by a fresh constant ξ so that the resulting pair $(E', \emptyset)$ is an explanation of $goal$ wrt $\mathscr{T}(F)$. It follows that $ok(f)$ can be derived from $T(F) \cup E'$, that is, that for every $x \in Y$, at least one of $true(x)$ and $false(x)$ can be derived. This, implies that for every $x \in Y$, $p(x,0) \in T(F) \cup E'$, that is, ξ is substituted for f or 0 in E.

Since, by (13), every explanation of $goal$ contains at least one atom of the form $p(f,z)$, ξ is not substituted for f in E to produce E'. Thus, E' is obtained from E by substituting ξ for some occurrences of 0. Let $U = \{u \in Y \cup \{f\} \colon p(u, \xi) \in E'\}$. If $f \notin U$, then let y denote any element in $U \cap Y$ (such an element exists as $U \neq \emptyset$). Since $p(t, \xi)$ and $p(f, \xi)$ are not in $T(F) \cup E'$, neither $true(y)$ nor $false(y)$ can be derived from $T(F) \cup E'$. Thus, neither $ok(f)$ nor $goal$ can be derived from $T(F) \cup E'$. It follows that $f \in U$. Consequently, for every $x \in U$, $false(x)$ can be derived from $T(F) \cup E'$, and $true(x)$ cannot be. Similarly, for every $x \in Y \setminus U$, $true(x)$ can be derived from $T(F) \cup E'$, and $false(x)$ cannot be. Thus, the atoms $true(x)$ and $false(x)$ in $T(F) \cup E'$ determine a truth assignment on atoms of Y. Since $sat(c_m)$ can be derived from $T(F) \cup E'$, $clsat(c)$ can be derived form $T(F) \cup E'$, for every clause c in F. It follows that the truth assignment determined by the atoms $true(x)$ and $false(x)$ in $T(F) \cup E'$ satisfies F.

Conversely, let us assume that F is satisfiable. Let us consider any satisfying assignment for F and let U comprises f and those atoms in Y that are false under this assignment. Let E' be obtained from E by substituting ξ for the occurrences of 0 in atoms $p(y,0)$, $y \in U$. It is easy to verify that for every $y \in U$, $false(y)$ can be derived from $T(F) \cup E'$, and $true(y)$ cannot be. Similarly, for every $y \in Y \setminus U$, $true(y)$ can be derived from $T(F) \cup E'$, and $false(y)$ cannot be. Moreover, $clsat(c)$ can be derived from $T(F) \cup E'$, for every clause c of F. Consequently, $ok(f)$ and $sat(c_m)$ can be derived from $T(F) \cup E'$. Since $p(f, \xi) \in E'$, $goal$ can be derived from $T(F) \cup E'$, that is, $(E', \emptyset)$ is an explanation of $goal$ wrt $\mathscr{T}$. Thus, $(E, \emptyset)$ is arbitrary.

(2) The argument for the membership part follows by Theorem **??**.

We prove hardness. The problem to decide whether a QBF $\Phi = \exists X \forall Y G$, where G is a DNF formula over variables in $X \cup Y$, is true, is Σ_2^P-complete. We will reduce it to the problem in question.

Below, we understand v_X, F, $Cl(F)$ and $G|_{v_X}$ as in the proof of Theorem **??**. We assume that $X = \{x_1, \ldots, x_k\}$, $Y = \{y_1, \ldots, y_n\}$ and $C = \{c_1, \ldots, c_m\}$.

We define σ to consist of predicate symbols $in_X/1$, $in_Y/1$, $clause/1$, $pos/2$, $ngtd/2$, $next_X/2$, $next_Y/2$, $next_C/2$, $true_X/1$, $false_X/1$, $true/1$, $false/1$, $ok_X/1$, $ok_Y/1$, $sat/1$, $f_X/1$, $l_X/1$, $f_Y/1$, $l_Y/1$, $f_C/1$, $l_C/1$, $tr/1$, $fa/1$, $assign/2$, $good_X/0$, $good_Y/0$, $good_C/0$, $good_f/0$, $goal/0$. We assume three new distinct constants 0, t and f and consider an abductive theory

$$\mathscr{T}(F) = \langle T(F), \{true_X, false_X, assign, fa\}, \emptyset \rangle,$$

where $T(F)$ consists of the following atoms (part B):

1. $in_X(a)$, for every $a \in X$
2. $in_Y(a)$, for every $a \in Y \cup \{t, f\}$
3. $pos(a,c)$, for every atom $a \in X \cup Y$ and clause $c \in Cl(F)$ such that a occurs non-negated in c
4. $ngtd(a,c)$, for every atom $a \in X \cup Y$ and clause $c \in Cl(F)$ such that a occurs negated in c
5. $f_X(x_1)$, $l_X(x_k)$
6. $f_Y(y_1)$, $l_Y(y_n)$
7. $f_C(c_1)$, $l_C(c_m)$
8. $next_X(x_i, x_{i+1})$, for $i = 1, \ldots, k-1$
9. $next_Y(y_i, y_{i+1})$, for $i = 1, \ldots, n-1$
10. $next_C(c_i, c_{i+1})$, for $i = 1, \ldots, m-1$
11. $tr(0)$

and of the following rules (part R)

1. $clause(C) \leftarrow pos(A,C)$
2. $clause(C) \leftarrow ngtd(A,C)$
3. $true(A) \leftarrow true_X(A)$
4. $false(A) \leftarrow false_X(A)$
5. $true(B) \leftarrow in_X(B), true_X(A), false_X(A)$
6. $false(B) \leftarrow in_X(B), true_X(A), false_X(A)$
7. $true(B) \leftarrow in_Y(B), true_X(A), false_X(A)$
8. $false(B) \leftarrow in_Y(B), true_X(A), false_X(A)$
9. $true(A) \leftarrow in_Y(A), assign(A,Z), tr(Z)$
10. $false(A) \leftarrow in_Y(A), assign(A,Z), fa(Z)$
11. $clsat(C) \leftarrow pos(A,C), true(A)$
12. $clsat(C) \leftarrow ngtd(A,C), false(A)$
13. $ok_X(A) \leftarrow f_X(A), true(A)$
14. $ok_X(A) \leftarrow f_X(A), false(A)$
15. $ok_X(A) \leftarrow ok_X(A'), next_X(A',A), true(A)$
16. $ok_X(A) \leftarrow ok_X(A'), next_X(A',A), false(A)$
17. $good_X \leftarrow ok_X(A), l_X(A)$
18. $ok_Y(A) \leftarrow f_Y(A), true(A)$
19. $ok_Y(A) \leftarrow f_Y(A), false(A)$
20. $ok_Y(A) \leftarrow ok_Y(A'), next_Y(A',A), true(A)$
21. $ok_Y(A) \leftarrow ok_Y(A'), next_Y(A',A), false(A)$
22. $good_Y \leftarrow ok_Y(A), l_Y(A)$
23. $sat(C) \leftarrow clsat(C), f_C(C)$
24. $sat(C) \leftarrow sat(C'), next_C(C',C), clsat(C)$
25. $good_C \leftarrow sat(C), l_C(C)$
26. $goal \leftarrow good_X, good_Y, good_C, fa(Z)$
27. $goal \leftarrow good_X, good_Y, in_Y(A), false(A), true(A), fa(Z)$.

Let v_X be an assignment of truth values to variables in X such that the formula $\forall Y\ G|_{v_X}$ is true, and let

$$\begin{aligned} E = & \{true_X(a) \colon a \in X \text{ and } v_X(a) = true\} \cup \\ & \{false_X(a) \colon a \in X \text{ and } v_X(a) = false\} \cup \\ & \{assign(y,0) \colon y \in Y\} \cup \{fa(0)\}. \end{aligned}$$

It is clear that $(E, \emptyset)$ is an explanation for *goal* wrt $\mathscr{T}(F)$. Indeed, since for every $y \in Y$ we have *true*(y) and *false*(y), *goal* can be derived by means of the rule (27). Let us assume that E is arbitrary. Then, there is a constant, say a, appearing in E such that replacing some occurrences of a with a fresh constant ξ results in another explanation of *goal*. However, if $a \in X$, then neither *true*(a) nor *false*(a) can be derived after the replacement. Consequently, we cannot derive $good_X$ and so, we cannot derive *goal* either. If $a \in Y$, then again neither *true*(a) nor *false*(a) can be derived. Now, $good_Y$ cannot be derived and so, neither can *goal*. Thus, $a = 0$. If we do not replace the occurrence of 0 in *fa*(0) with ξ, then there is $y \in Y$ such that we replace the occurrence of 0 in *assign*$(y, 0)$ with ξ. For that y, after the replacement we cannot derive *true*(y) nor *false*(y) and so, $good_Y$ and *goal* cannot be derived. It follows that there is a set $Y' \subseteq Y$ such that

$$\begin{aligned} E' =& \{true_X(a)\colon a \in X \text{ and } v_X(a) = true\} \cup \\ & \{false_X(a)\colon a \in X \text{ and } v_X(a) = false\} \cup \\ & \{assign(y,0)\colon y \in Y \setminus Y'\} \cup \{assign(y,\xi)\colon y \in Y'\} \cup \{fa(\xi)\} \end{aligned}$$

gives rise to an explanation $(E', \emptyset)$ of *goal*. Clearly, after applying $(E', \emptyset)$, for no $y \in Y$, both *true*(y) and *false*(y) can be derived. Thus, *goal* must be derivable by means of the rule (26). Moreover, for every $y \in Y$, we have exactly one of *true*(y) and *false*(y) hold: *true*(y) holds in $y \in Y \setminus Y'$, and *false*(y) holds if $y \in Y'$. Since *goal* can be derived, it follows that $good_C$ can be derived. Consequently, the truth assignment on Y defined by the atoms *true*(y) and *false*(y), where $y \in Y$, satisfies the set of clauses of $F|_{v_X}$, that is $\exists Y F|_{v_X}$ is true. This is a contradiction since $\exists Y F|_{v_X} \equiv \neg \forall Y\, G|_{v_X}$. Hence, $(E, \emptyset)$ is constrained.

Conversely, let $(E, \emptyset)$ be a constrained explanation of the goal. Clearly, E consists of facts of the form $true_X(a)$, $false_X(b)$, *assign*(y,z) and *fa*(w). For every element $x \in X$, at least one of $true_X(x)$ and $false_X(x)$ must be present in E (otherwise, we cannot derive $good_X$). Moreover, if for at least one element $a \in X$ we have $true_X(a)$ and $false_X(a)$ in E, then changing these two occurrences of a to ξ does not affect derivability of *goal* (indeed, by the rules (5)-(8) both before and after the change we have *true*(x) and *false*(x) hold for all $x \in X \cup Y$). Thus, $(E, \emptyset)$ would not be constrained. Finally if $true_X(a)$ or $false_X(a)$ is in E, $a \in X$. Otherwise, that a could be replaced by ξ without affecting the derivability of *goal*, contradicting again the assumption that $(E, \emptyset)$ is constraied. It follows that if $true_X(x)$ or $false_X(x)$ is in E, $x \in X$ and that the atoms $true_X(x)$ and $false_X(x)$ that belong to E determine a truth assignment on X, say v_X.

Next, let us assume that for some $\alpha \neq 0$ we have $assign(y, \alpha) \in E$. Then replacing all occurrences of α by ξ (including possibly an occurrence if α in *fa*(α)) has no effect on the derivability of *goal*. As before, we get a contradiction. Thus, if E contains facts *assign*(y,z), they are of the form *assign*$(y, 0)$. If any of these y's is not in Y, it can be changed to ξ without affecting the derivability of *goal*.

Next, we note that if E contains *fa*(α), where $\alpha \neq 0$, that α can be changed to ξ without affecting the derivability of *goal*.

If for some $y \in Y$, *assign*$(y, 0)$ is not in E, then for that y we can derive neither *true*(y) nor *false*(y). Thus, we cannot derive $good_Y$ and, consequently, we cannot derive *goal* either. It follows that E contains all facts *assign*$(y, 0)$, $y \in Y$, and no other facts based on the relation symbol *assign*.

If *fa*(0) is not in E, *goal* cannot be derived. Thus, E is of the form we considered above. Let $Y' \subseteq Y$ and let E' be as above. Since $(E, \emptyset)$ is constrained, $(E', \emptyset)$ is not an explanation of *goal*. That is a truth assignment on Y such that elements in $Y \setminus Y'$ are assigned *true* and those in Y'

are assigned *false* is not a satisfying assingment for $F|_{v_X}$. Consequently, it follows that $\exists Y F|_{v_X}$ is false and so, $\forall Y G|_{v_X}$ is true. This last property implies that $\exists X \forall Y G$ is true. □

Theorem 6
Let $\mathscr{A}$ be a set of abducible predicates, R a (fixed) non-recursive Horn program with no abducible predicates in the heads of its rules, and $\mathscr{C}$ a (fixed) set of integrity constraints.

1. The following problem is coNP-complete: given a set B of abducibles, an observation O, and a pair (E,F) of sets of abducibles, decide whether (E,F) is a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \mathscr{C} \rangle$.
2. The following problem is Σ_2^P-complete: given a set B of abducibles and an observation O, decide whether a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \mathscr{C} \rangle$ exists.

Proof
(1) The membership part follows by Theorem **??**. To prove hardness, we show that the problem to decide whether (E,F) is arbitrary is NP-hard. That is sufficient, as every explanation is either arbitrary or constrained. To show NP-hardness of the problem to decide whether an explanation is arbitrary, we reduce the SAT problem to it. Thus, let Y be a (finite) set of atoms and F a CNF theory over Y. As before, we denote by $Cl(F)$ the set of clauses in F. Let us also consider three additional distinct symbols t, f and 0. We define the vocabulary σ to consist of predicate symbols $in_Y/1$, $clause/1$, $pos/2$, $ngtd/2$, $p/2$ $true/1$, $false/0$, $ok/1$, $next/2$, $next_C/2$, $clsat/0$, $sat/1$, and $goal/0$, and an abductive theory

$$\mathscr{T}(F) = \langle T(F), \{p\}, \mathscr{C} \rangle,$$

where $T(F)$ consists of the following atoms:

1. $in_Y(a)$, for every $a \in Y \cup \{t, f\}$
2. $pos(a,c)$, for every atom $a \in Y$ and clause $c \in Cl(F)$ such that a occurs non-negated in c
3. $ngtd(a,c)$, for every atom $a \in Y$ and clause $c \in Cl(F)$ such that a occurs negated in c
4. $p(t,0)$

and of the rules

1. $clause(C) \leftarrow pos(A,C)$
2. $clause(C) \leftarrow ngtd(A,C)$
3. $true(A) \leftarrow in_Y(A), p(A,Z), p(t,Z)$
4. $false(A) \leftarrow in_Y(A), p(A,Z), p(f,Z)$
5. $clsat(C) \leftarrow pos(A,C), true(A)$
6. $clsat(C) \leftarrow ngtd(A,C), false(A)$
7. $goal \leftarrow p(f,X)$

and where $\mathscr{C}$ consists of

1. $\forall C\ clause(C) \supset clsat(C)$
2. $\forall A\ in_Y(A) \supset false(A) \vee true(A)$.

Clearly, the pair $(E,\emptyset)$, where $E = \{p(x,0) \colon x \in Y \cup \{f\}\}$, is an explanation of *goal*. Indeed, for every $x \in Y$, both $true(x)$ and $false(x)$ can be derived from $T(F) \cup E$ (because $p(t,0)$ and $p(f,0)$ both hold in $T(F) \cup E$). Thus, for every clause c, $clsat(C)$ can be derived, too. Consequently, the two integrity constraints in the theory hold for the least model of the program

$T(F) \cup E$. Moreover, *goal* belongs to this unique model and so, it is entailed by the revised theory.

Let us assume that E is arbitrary. We will prove that F is satisfiable. By the definition of arbitrariness, one of the constants appearing in E can be replaced by a fresh constant ξ so that the resulting pair $(E', \emptyset)$ is an explanation of *goal* wrt $\mathscr{T}(F)$, that is, the least model of $T(F) \cup E'$ satisfies both integrity constraints of the abductive theory and contains an atom of the form $p(f,z)$ (in order for *goal* to hold.

If f is replaced with ξ in E, then the least model of $T(F) \cup E'$ does not contain any fact of the form $p(f,X)$, and $(E', \emptyset)$ is not an explanation. If some atom $x \in Y$ is replaced by ξ, then for that atom neither $true(x)$ nor $false(x)$ belongs to the least model of $T(F) \cup E'$, which means that this model violates the second integrity constraint, contrary to the fact that $(E', \emptyset)$ is an explanation.

Thus, there is a non-empty set $U \subseteq Y \cup \{f\}$, such that when each occurrence of 0 in $p(x,0)$, where $u \in U$, is replaced by ξ, the resulting set $E' = \{p(x,\xi) \colon x \in U\} \cup \{p(x,0) \colon x \in (Y \setminus U)\}$ gives rise to an explanation $(E', \emptyset)$. Let us assume that $f \notin U$. Since $U \neq \emptyset$, $U \cap Y \neq \emptyset$. Let $x \in U \cap Y$. For this x, the least model of $T(F) \cup E'$ contains neither $false(x)$ nor $true(x)$, violating the second integrity constraint. Thus, $f \in U$ and, consequently, the least model of $T(F) \cup E'$, contains atoms $false(x)$, where $x \in U \setminus \{f\}$, and $true(x)$, where $x \in Y \setminus U$. It follows that this set of atoms defines a valuation on Y. Moreover, since the first integrity constraint holds, this valuation satisfies all clauses of F.

Conversely, let us assume that F is satisfiable. Let us consider any satisfying assignment for F and let U comprises f and those atoms in Y that are false under this assignment. Let E' be obtained from E by substituting ξ for the occurrences of 0 in atoms $p(y,0)$, $y \in U$. It is easy to verify that for every $y \in U$, $false(y)$ can be derived from $T(F) \cup E'$, and $true(y)$ cannot be. Similarly, for every $y \in Y \setminus U$, $true(y)$ can be derived from $T(F) \cup E'$, and $false(y)$ cannot be. Moreover, $clsat(c)$ can be derived from $T(F) \cup E'$, for every clause c of F. Thus, both integrity constraints are satisfied by the least model of $T(F) \cup E'$, and that model also contains $p(f,\xi)$ and so, also *goal*. Thus, $(E, \emptyset)$ is an arbitrary explanation of goal.

(2) The membership part follows by Theorem **??**. To prove hardness we proceed similarly as in the proofs of Theorems 4 and 5. $\square$

Theorem 7
Let $\mathscr{A}$ be a set of abducible predicates and R a (fixed) non-recursive Horn program with no abducible predicates in the heads of its rules. The following problems are in P.

1. Given a set B of abducibles, an observation O, and a pair (E,F) of sets of abducibles, decide whether (E,F) is a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \emptyset \rangle$.
2. Given a set B of abducibles and an observation O, decide whether a constrained explanation for O wrt $\langle R \cup B, \mathscr{A}, \emptyset \rangle$ exists.

Proof
(1) Let us consider an explanation (E,F). Since R is non-recursive, there is a constant, say k, such that any proof of o based on the rules in R and facts in B revised by (E,F) has length bounded from above by k. Thus, the total number of facts used in any such proof is bounded by k, too.

Since at most k atoms in E are relevant to any proof, if E contains more than k abducibles with predicate symbols of positive arity, it is not constrained. Indeed, at least one of these abducibles does not play any role in the proof. For for this abducible, say $a = p(c_1, \ldots, c_m)$, we have that replacing c_1 with ξ in a results in an explanation.

If on the other hand, the number of abducibles with predicate symbols of positive arity in E is less than or equal to k, then the total number of constants occurring in all abducibles in E is bounded by a constant k' dependent on R only (independent of the size of input, that is, of the size of B). Thus, there is only a fixed number of possible selections of occurrences of a constant for replacement by a new symbol ξ. For each of them, we can test in polynomial time whether it leads to an explanation. Thus, we can decide whether (E,F) is constrained in polynomial time.

(2) If (E,F) is a constrained explanation, then $(E,\emptyset)$ is a constrained explanation. Moreover, we can assume that E contains all zero arity abducibles in B. Thus, each such constrained explanation is determined by its non-zero arity abducibles. Non-zero arity abducibles in a constrained explanation use only constants appearing in O and P (Theorem **??**). It follows, the set of all possible non-zero arity abducibles that might be chosen to form E has size that is polynomial in the size of B (the input size). Since by an argument from the previous proof, we can assume that E contains no more than k non-zero arity abducibles (where k is a constant depending only on R), the set of all candidate explanations that need to be tested to decide the problem is polynomial in the size of input. Since each such candidate explanation can be tested for constrainedness in polynomial time (by the previous result), the assertion follows. □