# Outlier Detection using Default Logic*

Fabrizio Angiulli
ICAR-CNR c/o DEIS
Università della Calabria
87030 Rende (CS), Italy
angiulli@icar.cnr.it

Rachel Ben-Eliyahu - Zohary
Comm. Systems Engineering Dept.
Ben-Gurion University of the Negev
Beer-Sheva 84105, Israel
rachel@bgumail.bgu.ac.il

Luigi Palopoli
DIMET
Università di Reggio Calabria
Loc. Feo di Vito
89100 Reggio Calabria, Italy
palopoli@ing.unirc.it

December 9, 2003

**Abstract**

In AI theory and applications, default logic is used to describe regular behavior and normal properties. In this paper, we suggest to exploit default logics in a somewhat different way, that is, using this formalism for detecting *outliers*, that denote individuals who behave in an unexpected way or feature abnormal properties. The ability to locate outliers can help in keeping knowledge base integrity and singling out *irregularities* in stored knowledge about individuals. In this paper we first formally define the notion of an *outlier* and an outlier witness. Then, we illustrate potential interesting applications for the presented notions. We then analyze the computational complexity associated with finding outliers. We show that several versions of the outlier detection problem all lie over the second level of the polynomial hierarchy. For an example, the question of establishing if at least one outlier can be detected in a given propositional default theory is $\Sigma_3^P$-complete under polynomial time transformation. The fact that outlier detection involves heavy computation is a challenge, but many times the queries involved can be executed off-line, thus relieving the problem in some sense. In addition we show that outlier detection can be done in polynomial time for the class of acyclic normal

*A preliminary an partial version of this work appears in the Proceedings of the International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 2003.

unary defaults and the class of acyclic dual normal unary defaults. Finally, we also discuss the relationship of outlier detection and abduction in default theories.

**Keywords:** computational complexity, data mining, knowledge representation, nonmonotonic reasoning.

# 1 Introduction

Default logics were developed as a tool for representing and reasoning with incomplete knowledge. Using the default rules, we are able to describe how things work in general. Then, using the default rules, we can make some assumptions about individuals and draw conclusions about their properties and behaviour.

In this paper, we would like to suggest a somewhat different usage for default rules. The basic idea is as follows. Since default rules are used for describing regular behaviour, we can exploit them for detecting individuals or elements who *do not* behave normally according to the default theory at hand. In a sense, an *outlier* is a property of an element to which no logical justification can be associated within the theory.

This can be of help in several applications context, e.g., to single out exceptional behaving individuals or system components. Note that, here, exceptions are not *explicitly* listed in the theory as "abnormals", as often done in logical-based abduction. Rather, the "abnormality" is singled out exactly because some of the properties characterizing them does not have a justification within the theory at hand.

For example, suppose that it usually takes about two seconds to download a one-megabyte file from some server. Then, one day, the system is slower - instead four seconds are needed to perform the same task. While four seconds may be a good performance it is helpful to find the source of the delay. Another example might be that someone's car breaks are making a strange noise. Although they seem to be functioning properly, this is not normal behavior and the car should be serviced. So, if the truth of the fact denoting the occurrence of the noise is not supported by the rest of the theory, then we would be allowed to conclude that such a noise wouldn't be there, singling out an exceptional situation, that is, an outlier.

Another usage of outlier detection would be for examining database integrity. If an abnormal property is discovered in a database, the source who reported this observation would have to be double-checked.

Detecting abnormal properties, that is, detecting outliers, can also lead to an update of the default rules. Suppose we have the rule that birds fly, and we observe a bird, say Tweety, that does not fly. So, we might report to the knowledge engineer the occurrence of such outlier in the theory, which should lead the engineer to update the knowledge base, for instance, with the properties that Tweety is a penguin, and penguins do not fly.

In this paper, we shall formally state the ideas briefly sketched above within the context of Reiter's default logic. In this paper, we concentrate on the propositional fragment of default logic although the generalization of such ideas to the realm of first-order defaults is also worth exploring. So, whenever we use a default theory with variables (e.g., in some of the following examples), we refer to it as an abbreviation of its grounded version.

The rest of this paper is organized as follows. In Section 2 we give preliminary definitions as well as we formally define the concepts of outlier and related notions. In Section 3 we discuss the complexity of finding outliers in general propositional as well as in disjunction free default logics. Then, in Section 4 we describe some tractable cases. Section 5 reports about the relationship of outlier detection and abduction in default theories and of outlier detection in logic and in data. Finally, in Section 6 we are drawing our conclusions.

# 2 Definitions

In this section we provide preliminary as well as new definitions for concepts we will be using throughout the paper.

## 2.1 Preliminaries

### 2.1.1 Default logics

Let $T$ be a propositional theory. Then $T^*$ denotes its logical closure. Let $S$ be a set of propositional formulas, then $\neg S$ denotes the set of all formulas that are the negation of some formula in $S$.

Default logic was introduced by Reiter [25]. A *propositional default theory* $\Delta$ is a pair $(D, W)$ consisting of a set $W$ of propositional formulas and a set $D$ of default rules. A *default rule* $\delta$ has the form

$$\frac{\alpha : \beta_1, \ldots, \beta_m}{\gamma}$$

where $\alpha$, each $\beta_i$, $1 \leq i \leq m$, and $\gamma$ are propositional formulas. In particular, $\alpha$ is called the *prerequisite*, $\beta_1, \ldots, \beta_m$ the *justification*, and $\gamma$ the *consequent* (or *conclusion*) of $\delta$. The prerequisite could be missing, while justification and consequent are required. If the conclusion of a default rule occurs in the justification, the rule is said to be *semi-normal*, while if the conclusion is identical to the justification the rule is said to be *normal*. A default theory containing only (semi-)normal defaults is said to be (*semi-*)*normal*.

Given a default rule $\delta$, we denote by $pre(\delta)$, $just(\delta)$, and $concl(\delta)$, respectively, the prerequisite, justification, and consequent of $\delta$. Given a set $R = \delta_1, \ldots, \delta_n$ of default rules, we denote by $pre(R)$, $just(R)$, and $concl(R)$, respectively, the sets $\{pre(\delta_1), \ldots, pre(\delta_n)\}$, $\{just(\delta_1), \ldots, just(\delta_n)\}$, and $\{concl(\delta_1), \ldots, concl(\delta_n)\}$.

The informal meaning of a default rule $\delta$ is the following: if $pre(\delta)$ is known, and if it is consistent to assume $just(\delta)$, then conclude $concl(\delta)$. The semantics of a default theory is defined in terms of *extensions*, that are maximal sets of conclusions that can be drawn from a theory. Formally, $\mathcal{E}$ is an extension for the theory $(D, W)$ if and only if it satisfies the following equations:

- $E_0 = W$,
- for $i \geq 0$, $E_{i+1} = E_i^* \cup \left\{\gamma \mid \frac{\alpha:\beta_1,\ldots,\beta_m}{\gamma} \in D, \alpha \in E_i, \neg\beta_1 \notin \mathcal{E}, \ldots, \neg\beta_m \notin \mathcal{E}\right\}$,
- $\mathcal{E} = \bigcup_{i=0}^{\infty} E_i$.

Thus, by definition, an extension is a deductively closed set of formulas, hence infinite. Nonetheless, by results of [31], an extension $\mathcal{E}$ of a propositional default theory $\Delta = (D, W)$ can be finitely characterized through the set $D_{\mathcal{E}}$ of *generating defaults* for $\mathcal{E}$ w.r.t. $\Delta$. Indeed in [31] the authors show that a propositional default theory $\Delta = (D, W)$ has an extension $\mathcal{E}$ iff there exists a set $D_{\mathcal{E}} \subseteq D$, the generating defaults for $\mathcal{E}$ w.r.t. $\Delta$, that can be partitioned into a finite number of strata $D_{\mathcal{E}}^{(0)}, D_{\mathcal{E}}^{(1)}, \ldots, D_{\mathcal{E}}^{(n)}$, such that:

- $D_{\mathcal{E}}^{(0)} = \{\delta \mid \delta \in D_{\mathcal{E}}, pre(\delta) \in W^*\}$,
- for each $i$, $1 \leq i \leq n$, $D_{\mathcal{E}}^{(i)} = \{\delta \mid \delta \in D_{\mathcal{E}} - \bigcup_{j=0}^{i-1} D_{\mathcal{E}}^{(j)}, pre(\delta) \in (W \cup concl(\bigcup_{j=0}^{i-1} D_{\mathcal{E}}^{(j)}))^*\}$,
- $(\forall\delta \in D_{\mathcal{E}})(\forall\beta \in just(\delta))(\neg\beta \notin (W \cup concl(D_{\mathcal{E}}))^*)$, and
- $(\forall\delta \in D)(pre(\delta) \in (W \cup concl(D_{\mathcal{E}}))^* \wedge (\forall\beta \in just(\delta))(\neg\beta \notin (W \cup concl(D_{\mathcal{E}}))^* \Rightarrow \delta \in D_{\mathcal{E}})$.

If such a set $D_{\mathcal{E}}$ exists, then $\mathcal{E} = (W \cup concl(D_{\mathcal{E}}))^*$ is an extension of $\Delta$.

A finite propositional default theory $\Delta = (D, W)$ is *disjunction free* (DF for short), if $W$ is a set of literals, and the precondition, justification and consequence of each default in $D$ is a conjunction of literals. It is useful to rewrite the definition of extension provided in [31] for the special case of disjunction-free theories [16]. Let $\Delta = (D, W)$ be a disjunction free default theory, then $\mathcal{E}$ is an extension of $\Delta$ iff there exists a sequence of rules $\delta_1, \ldots, \delta_n$ from $D$, and a sequence of sets $E_0, E_1, \ldots, E_n$, such that for all $i > 0$:

- $E_0 = W$,
- $E_i = E_{i-1} \cup concl(\delta_i)$,
- $pre(\delta_i) \subseteq E_{i-1}$,
- $(\nexists c \in just(\delta_i))(\neg c \in E_n)$,
- $(\nexists\delta \in D)(pre(\delta) \subseteq E_n \wedge concl(\delta) \nsubseteq E_n \wedge (\nexists c \in just(\delta))(\neg c \in E_n))$,

and $\mathcal{E}$ is the logical closure of $E_n$. We call the set of literals $E_n$, the *signature set* of $\mathcal{E}$, and denote it by $liter(\mathcal{E})$. For each extension $\mathcal{E}$ of a DF theory, the sequence of rules $\delta_1, \ldots, \delta_n$ described above is the set $D_{\mathcal{E}}$ of generating defaults of $\mathcal{E}$.

A DF default theory is *normal mixed unary* (NMU in short) iff its set of defaults contains only rules of the form $\frac{\alpha:\beta}{\beta}$, where $\alpha$ is either empty or a literal and $\beta$ is a literal. An NMU default theory is *normal unary* (NU for short) iff the prerequisite of each default is either empty or positive. An NMU default theory is *dual normal* (DNU for short) unary iff the prerequisite of each default is either empty or negative.

Although default theories are *nonmonotonic*, normal default theories satisfy the property on *semi-monotonicity* (see Theorem 3.2 of [25]). Semi-monotonicity in default logic means the following: let $\Delta = (D, W)$ and $\Delta' = (D', W)$ be two default theories such that $D \subseteq D'$; then for every extension $E$ of $\Delta$ there is an extension $E'$ of $\Delta'$ such that $E \subseteq E'$.

A default theory may not have any extensions. For example the default theory $(\{\frac{:\beta}{\neg\beta}\}, \emptyset)$ has no extensions. A default theory is said to be *coherent* if it has at least one extension, and incoherent otherwise. In particular, normal default theories are always coherent. A coherent propositional default theory $\Delta = (D, W)$ may have one (and only one) extension which is inconsistent. In this case the theory is said to be *inconsistent*. In particular, it can be shown (see Theorem 2.2 of [25] for details) that $\Delta$ is inconsistent iff $W$ is inconsistent. In general, a coherent propositional default theory $\Delta$ has more than one extension. Thus, given a propositional formula $\phi$, two basic questions involving default theories are the following:

**Membership:** does there exist an extension of $\Delta$ that contains $\phi$ ?

**Entailment:** does every extension of $\Delta$ contain $\phi$ ?

In particular, entailment is closely related to that reasoning called *skeptical* (or *cautious*) *reasoning*, where a literal is believed iff it is included in all extensions of the theory.

Let $\Delta$ be a default theory and $\phi$ be a formula. Then $\Delta \models \phi$ means that $\phi$ is entailed by $\Delta$. Similarly, for a set of formulas $S$, $\Delta \models S$ means that every formula $\phi \in S$ belongs to every extension of $\Delta$.

### 2.1.2 Complexity Theory

We recall some basic definitions about complexity theory, particularly, the polynomial time hierarchy. The reader is referred to [15, 21] for more on this.

The class P is the set of decision problems that can be answered by a Turing machine in polynomial time. The class of decision problems that can be solved by a nondeterministic Turing machine in polynomial time is denote by NP, while the class of decision problems whose complementary problem is in NP, is denote by co-NP. The classes $\Sigma_k^P$ and $\Pi_k^P$, constituting the *polynomial hierarchy*, are defined as follows: $\Sigma_0^P = \Pi_0^P = \mathrm{P}$ and for all $k \geq 1$, $\Sigma_k^P = \mathrm{NP}^{\Sigma_{k-1}^P}$, and $\Pi_k^P = \text{co-}\Sigma_k^P$. $\Sigma_k^P$ models computability by a nondeterministic polynomial time Turing machine which may use an oracle, that is, loosely speaking, a subprogram, that can be run with no computational cost, for solving a problem in $\Sigma_{k-1}^P$. The class $D_k^P$, $k \geq 1$, is defined as the class of problems that consist of the conjunction of two independent problems from $\Sigma_k^P$ and $\Pi_k^P$, respectively. Note that, for all $k \geq 1$, $\Sigma_k^P \subseteq D_k^P \subseteq \Sigma_{k+1}^P$.

Let $\Gamma$ denote the set of all the strings over a given finite set of symbols. A function $f : \Gamma \mapsto \Gamma$ is said to be *polynomial time computable* if there exists a polynomial time Turing machine that computes it. Let $L_1, L_2$ be two subsets of $\Gamma$. A polynomial time computable function $\tau : \Gamma \mapsto \Gamma$ is called a *polynomial time transformation* from $L_1$ to $L_2$ if for each $x \in \Gamma$ the following holds: $x \in L_1$ iff $\tau(x) \in L_2$. The *language* $L(A)$ associated to a decision problem $A$, accepting inputs from $\Gamma$, is the set constituted by the strings $x \in \Gamma$ such that $A$ returns "yes" on the input $x$. A problem $A$ is *polynomially reducible* to a problem $B$ if there exists a polynomial time transformation from $L(A)$ to $L(B)$. A problem $A$ is *complete* for the class $\mathcal{C}$ of the polynomial hierarchy iff $A$ belongs to $\mathcal{C}$ and every problem in $\mathcal{C}$ is polynomially reducible to $A$.

A well known $\Sigma_k^P$-complete problem is to decide the validity of a formula $QBE_{k,\exists}$, that is, a formula of the form $\exists X_1 \forall X_2 \ldots QX_k f(X_1, \ldots, X_k)$, where $Q$ is $\exists$ if $k$ is odd and is $\forall$ if $k$ is even, $X_1, \ldots, X_k$ are disjoint set of variables, and $f(X_1, \ldots, X_k)$ is a propositional formula in $X_1, \ldots, X_k$. Analogously, the validity of a formula $QBE_{k,\forall}$, that is a formula of the form $\forall X_1 \exists X_2 \ldots QX_k f(X_1, \ldots, X_k)$, where $Q$ is $\forall$ if $k$ is odd and is $\exists$ if $k$ is even, is complete for $\Pi_k^P$. Deciding the conjunction $\Phi \wedge \Psi$, where $\Phi$ is a $QBE_{k,\exists}$ formula and $\Psi$ is a $QBE_{k,\forall}$ formula, is complete for $D_k^P$.

## 2.2 Defining outliers

Next we will formalize the notion of an outlier in default logic. In order to motivate the definition and make it easy to understand, we will first look at an example.

**Example 2.1** Consider the following default theory, representing the knowledge that birds fly and penguins are birds that do not fly, and the observations that Tweety is bird, Pini is a pinguin, and Tweety is beautiful and does not fly.

$$D = \left\{ \frac{Bird(x) : Fly(x)}{Fly(x)}, \frac{Penguin(x) : Bird(x)}{Bird(x)}, \right.$$
$$\left. \frac{Penguin(x) : \neg Fly(x)}{\neg Fly(x)} \right\}$$
$$W = \{Bird(Tweety), Beautiful(Tweety),$$
$$Penguin(Pini), \neg Fly(Tweety)\}$$

This theory has two extensions. One extension is the logical closure of $W \cup \{Bird(Pini), \neg Fly(Pini)\}$ and the other is the logical closure of $W \cup \{Bird(Pini), Fly(Pini)\}$.
If we look carefully at the extensions, we note that Tweety not flying is quite strange, since we know that birds fly and Tweety is a bird. Therefore, there is no apparent justification to the fact that Tweety does not fly (other than the fact $\neg Fly(Tweety)$ belonging to $W$). Had we been told that Tweety is a penguin, we could have explained the fact that Tweety does not fly. But, as the theory stands now, we are not able to explain why Tweety does not fly, and, thus, Tweety has, as to say, a exceptional property. If we are trying to nail down what induces such an exception, we notice that if we would have dropped the observation $\neg Fly(Tweety)$ from $W$, we would have concluded the exact opposite, that is, that Tweety does fly. Thus, $\neg Fly(Tweety)$ induces such an exceptionality (we will call *witness* such a literal like $\neg Fly(Tweety)$). Furthermore, if we drop from $W$ both

$\neg Fly(Tweety)$ and $Bird(Tweety)$, we are no longer able to conclude that Tweety flies. This implies that, in this context, $Fly(Tweety)$ derives from the fact that Tweety is a bird. Thus $Bird(Tweety)$ denotes the exceptional property characterizing Tweety as an outlier.

We note that, following the above example, one could be induced to define an outlier as an individual, i.e. a constant, in our case *Tweety*, that possesses an exceptional property, denoted by a literal having the individual as one of its arguments, in our case $Bird(Tweety)$. However, it is certainly more general and flexible define outliers as to single out a property of an individual which is exceptional, rather than simply the individual itself. That assumed, we also note that, within the propositional context we deal with here, we do not explicitly have individuals distinct from their properties and, therefore, the choice is anyway immaterial.

Therefore we define outliers and witnesses as follows.

**Definition 2.2** Let $\Delta = (D, W)$ be a propositional default theory such that $W$ is consistent and let $l \in W$ be a literal. If there exists a non empty set of literals $S \subseteq W$ such that:

1. $(D, W_S) \models \neg S$, and
2. $(D, W_{S,l}) \not\models \neg S$.

where $W_S = W \setminus S$ and $W_{S,l} = W_S \setminus \{l\}$, then we say that $l$ is an *outlier* in $\Delta$ and $S$ is an *outlier witness set* for $l$ in $\Delta$.

Thus, according to this definition, in the example theory reported above, we should conclude that $Bird(Tweety)$ denotes an outlier and $\{\neg Fly(Tweety)\}$ is its witness.

Note that we have defined an outlier witness to be a set, not necessarily a single literal. The reason is that for in some theories taking a single literal does not suffice to "form" a witness for a given outlier, having all witnesses of such an outlier a cardinality strictly larger than one.

**Example 2.3** Consider the following default theory $\Delta = (D, W)$, where the set of default rules $D$ convey the following information about weather and traffic in a small town in southern California:

1. $\frac{July \wedge Weekend : \neg Traffic\_Jam \wedge \neg Rain}{\neg Traffic\_Jam \wedge \neg Rain}$ - that is, normally in a July weekend there is no traffic jam and no rain.
2. $\frac{January : Rain}{Rain}, \frac{January : \neg Rain}{\neg Rain}$ - in January it sometimes rains and sometimes it doesn't rain.
3. $\frac{Weekend \wedge Traffic\_Jam : Accident \vee Rain}{Accident \vee Rain}$ - If there is a traffic jam in the weekend then normally it must be raining or there have been an accident.

Suppose also, that $W = \{July,\ Weekend,\ Traffic\_Jam,\ Rain\}$. Then, the set $S = \{Traffic\_Jam,\ Rain\}$ is an outlier witness for both $Weekend$ and $July$. Moreover, $S$ is a *minimal* outlier witness set for either of $Weekend$ or $July$, since deleting one of the members from $S$ will render $S$ not being a witness set.

Some more examples are reported next.

**Example 2.4** Consider the following default theory $\Delta$:

$$
\begin{aligned}
D = \Bigg\{ & \frac{Income(x) \wedge Adult(x) : Works(x)}{Works(x)}, \\
& \frac{FlyingS(x) : InterestTakeOff(x)}{InterestTakeOff(x)}, \\
& \frac{FlyingS(x) : InterestNavigate(x)}{InterestNavigate(x)} \Bigg\} \\
W = \{ & Income(Johnny), Adult(Jhony), \\
& \neg Works(Johnny), FlyingS(Johnny), \\
& \neg InterestTakeOff(Johnny)\}
\end{aligned}
$$

This theory claims that normally, adults who have monthly income work, and students who take flying lessons are interested in take off and in navigating. The observations are that Johnny is an adult who has monthly income, but he does not work. He is also a student in a flying school but he is not interested in take off. After we have learned some lessons from the September 11 events, we'd like our system to conclude that Johnny is an individual involved in, more than one, outlier. Indeed, the reader can verify that the following facts are true:

1. $(D, W_{\neg Works(Johnny)}) \models Works(Johnny)$,
2. $(D, W_{\neg InterestTakeOff(Johnny)}) \models InterestTakeOff(Johnny)$,
3. $(D, W_{\neg Works(Johnny), Adult(Johnny)}) \not\models Works(Johnny)$, and
4. $(D, W_{\neg InterestTakeOff(Johnny), FlyingS(Johnny)}) \not\models InterestTakeOff(Johnny)$

Hence, both $\neg Works(Johnny)$ and $\neg InterestTakeOff(Johnny)$ are outlier witnesses, while $Adult(Johnny)$ and $FlyingS(Johnny)$ are the correspondent outliers. Note that $Income(Johnny)$ is also an outlier, with the witness $\neg Works(Johnny)$.

## 2.3 Defining outlier detection problems

Having defined outliers, an important question is how much complex is to single them out. This is dealt with in the following Sections 3 and 4. There, we shall refer to following problems (that we shall also called *queries*) defined for an input default theory $\Delta = (D, W)$:

$Q0$: Given $\Delta$, does there exist at least one outlier in $\Delta$ ?

$Q1$: Given $\Delta$ and a literal $l \in W$, is $l$ an outlier in $\Delta$ ?

$Q2$: Given $\Delta$ and a set of literals $S \subseteq W$, is $S$ a witness for any outlier $l$ in $\Delta$ ?

$Q3$: Given $\Delta$, a set of literals $S \subseteq W$, and a literal $l \in W$, is $S$ a witness for $l$ in $\Delta$ ?

# 3 Complexity Results

In this and in the following section we will analyze the complexity associated with detecting outliers in general, DF, NMU, NU and DNU propositional default theories (this section) and illustrate some tractable classes of theories (Section 4). Detailed results proof are reported in the Appendix. Here, we limit ourselves to provide, for each of the outlier problems defined above (i.e. $Q0$, $Q1$, $Q2$, and $Q3$), a general discussion of the proof techniques we employ. The complexity results are summarized in Table 1, where $\mathcal{C}$-c stands for $\mathcal{C}$-complete.

| **Theory \ Query** | $Q0$ | $Q1$ | $Q2$ | $Q3$ |
|---|---|---|---|---|
| Propositional | $\Sigma_3^P$-c | $\Sigma_3^P$-c | $D_2^P$-c | $D_2^P$-c |
| DF, NMU | $\Sigma_2^P$-c | $\Sigma_2^P$-c | $D^P$-c | $D^P$-c |
| NU, DNU | NP-c | NP-c | P | P |
| Acy. NU, Acy. DNU | P | P | P | P |

Table 1: Complexity results for outlier detection

## 3.1 Queries Q0 and Q1

We start commenting about query $Q0$, the most general form of query that we have defined above. Given a default theory, this query asks for the existence of an outlier in the theory. When general propositional default theories are considered, this query is rather complex as it lies at the third level of the polynomial hierarchy.

**Theorem 3.1** *$Q0$ on general propositional default theories is $\Sigma_3^P$-complete under polynomial time transformations.*

We note that a problem lying at the $k$-th level of the polynomial hierarchy is characterized by exactly $k$ independent "sources of complexity". Each source of complexity consists of a search space composed by an exponential number of candidate solutions. In the case of general propositional default theories, two of the three sources underly to the associated entailment problem, that are $(i)$ the exponential number of generating defaults $D_\mathcal{E} \subseteq D$ and, thus, of possible extensions $\mathcal{E}$ of the default theory $\Delta = (D, W_S)$ ($\Delta = (D, W_{S,l})$

resp.), and $(ii)$ the propositional deductive inference needed to check that $D_{\mathcal{E}}$ generates an extension $\mathcal{E}$ of $\Delta$ and that $\neg S \in \mathcal{E}$ ($\neg S \notin \mathcal{E}$ resp.). The third one is determined by the exponential number of subsets of literals $S \cup \{l\}$ of $W$ candidate to play the role of an outlier witness set (the set $S$) and an outlier (the literal $l$) in $\Delta$.

If we restrict query $Q0$ to DF theories then the following holds.

**Theorem 3.2** $Q0$ *restricted to DF propositional default theories is* $\Sigma_2^P$*-complete under polynomial time transformations.*

The complexity of $Q0$ for DF theories goes down one level in the polynomial hierarchy w.r.t. general theories as, in this case, the deductive inference check reduces to simple set operations, and, therefore, we are left with only two sources of complexity.

The complexity associated with $Q0$ does not decrease even if we consider such a simplified form of DF theories as NMU theories, as stated below.

**Theorem 3.3** $Q0$ *restricted to NMU propositional default theories is* $\Sigma_2^P$*-complete under polynomial time transformations.*

This result is explained since the complexity of the entailment problem for NMU theories is the same for DF theories (see Lemma 1 in Appendix for the proof of this statement).

To obtain a further reduction in complexity, we have to consider simpler theories than the NMU ones.

**Theorem 3.4** $Q0$ *restricted to propositional NU default theories is NP-complete under polynomial time transformations.*

**Theorem 3.5** $Q0$ *restricted to propositional DNU default theories is NP-complete under polynomial time transformations.*

Query $Q0$ on these theories lies at the first level of the polynomial hierarchy since the entailment problem for NU and DNU default theories is polynomial time decidable. Note that, however, this NP-completeness result tells that searching for outliers even in these simple form of theories is a very complex task.

Next, we comment the proof techniques we have used (detailed proofs are reported in the Appendix) to prove the above statements.

The $\mathcal{C}$-membership of query $Q0$ on a propositional default theory $\Delta = (D, W)$, can be proved by building a nondeterministic Turing machine $T$ that guesses simultaneously a literal $l$ in $W$ and a subset $S = \{s_1, \ldots, s_n\}$ of $W$, and then verifies that

$$(D, W_S) \models \neg s_1 \wedge \ldots \wedge \neg s_n \text{ (query } q'\text{), and}$$
$$(D, W_{S,l}) \not\models \neg s_1 \wedge \ldots \wedge \neg s_n \text{ (query } q''\text{).}$$

Let co-$\mathcal{C}_e$ the class of complexity of the entailment problem for $\Delta$, then the query $q'$ is in the class co-$\mathcal{C}_e$, while the query $q''$ is in the class $\mathcal{C}_e$. Thus, $T$ can employ a $\mathcal{C}_e$ oracle to solve both query $q'$ and query $q''$. Hence, $Q0$ is in the class $\mathcal{C} = NP^{\mathcal{C}_e}$. We recall that the entailment problem is in $\Pi_2^P = \text{co-}\Sigma_2^P$ for general propositional default theories [28, 13], is in co-NP for DF [16] and NMU (see Lemma 1 in Appendix) propositional default theories, and is in P for NU [16] and DNU [32] propositional default theories. As a consequence, query $Q0$ is respectively in the classes $\text{NP}^{\Sigma_2^P} = \Sigma_3^P$, $\text{NP}^{\text{NP}} = \Sigma_2^P$, and $\text{NP}^P = \text{NP} = \Sigma_1^P$ for such classes of theories.

To prove the completeness of the query $Q0$ in the above reported classes, we reduce $Q0$ to the $\Sigma_k^P$-complete ($k \in \{1, 2, 3\}$) problem of deciding the validity of a formula $QBE_{k,\exists}$. The reductions described in the proofs of Theorems 3.1, 3.2, 3.3, 3.4 and 3.5, associate with the formula $\Phi$ the default theory $\Delta(\Phi) = (D(\Phi), W(\Phi))$ such that:

- there exists one and only one literal $l$ in $W(\Phi)$ that can be an outlier, while the literals in $W(\Phi) \setminus \{l\}$ can only belong to an outlier witness set $S$;
- there exists a bijection between all the possible outlier witness sets $S$ coming from $W(\Phi)$ and all the possible truth values of the variables in the set $X_1$;
- $\Delta(\Phi)$ encodes $\Phi$ itself, and is such that $(D(\Phi), W(\Phi)_S) \models \neg S$ iff $\forall X_2 \ldots QX_k f(X_1, \ldots, X_k)$ is valid, subject to the truth value assignment of $X_1$ induced by $S$;
- $l$ acts as a switch, i.e. if it is removed from $W(\Phi)_S$ then $(D(\Phi), W(\Phi)_{S,l}) \not\models \neg S$, for each admissible outlier witness set $S$.

Summarizing, the query $Q0$ is complete for the class $\Sigma_3^P$ for general propositional default theories, is complete for the class $\Sigma_2^P$ for DF and NMU propositional default theories, and is complete for the class NP for NU and DNU default theories, hence its complexity lies, in the polynomial hierarchy, exactly one level above the level associated to the corresponding entailment problems.

As for the query $Q1$, considerations analogous to that drawn for query $Q0$ hold and, thus, the complexity results for these two queries coincide, as summarized in the following results.

**Theorem 3.6** *$Q1$ on general propositional default theories is $\Sigma_3^P$-complete under polynomial time transformations.*

**Theorem 3.7** *$Q1$ restricted to DF propositional default theories is $\Sigma_2^P$-complete under polynomial time transformations.*

**Theorem 3.8** *$Q1$ restricted to NMU propositional default theories is $\Sigma_2^P$-complete under polynomial time transformations.*

**Theorem 3.9** *$Q1$ restricted to NU and DNU propositional default theories is NP-complete under polynomial time transformations.*

Intuitively, this can be justified noting that none of the sources of complexity involved for the various cases, with the query $Q0$ is cancelled by knowing the outlier literal $l$ in advance. In particular, the number of possible outlier witness set $S \subseteq W \setminus \{l\}$ for $l$ is still exponential.

## 3.2 Queries Q2 and Q3

Given a default theory and a set of literals $S$, query $Q2$ asks whether $S$ is a witness for any outlier in the theory. The complexity of $Q2$ lies in the polynomial hierarchy one level below the complexity of $Q0$. Indeed, one of the sources of complexity involved with query $Q0$, that is the exponential number of outlier witnesses, falls off when query $Q2$ is considered.

In particular, $Q2$ on general theories is the conjunction of two independent problems, one from $\Pi_2^P$ and one from $\Sigma_2^P$.

**Theorem 3.10** *$Q2$ on general propositional default theories is $D_2^P$-complete under polynomial time transformations.*

For DF and NMU default theories, $Q2$ is the conjunction of two independent problems from co-NP and NP respectively.

**Theorem 3.11** *$Q2$ restricted to DF propositional default theories is $D^P$-complete under polynomial time transformations.*

**Theorem 3.12** *$Q2$ restricted to NMU propositional default theories is $D^P$-complete under polynomial time transformations.*

Finally, $Q2$ on NU and DNU theories is the conjunction of two polynomial time solvable problems, and, hence, in these cases, this query is in P.

**Theorem 3.13** *$Q2$ restricted to NU and DNU propositional default theories is in $P$.*

Next, we comment on the techniques adopted to prove the above statements. Detailed proofs can be found in the Appendix.

We start with membership. Also in this case we have to answer the two queries $q'$ and $q''$ reported in Section 3.1 above, but this time the outlier witness set $S$ is given in input. We recall that for general propositional default theories $q'$ is in $\Pi_2^P$, while for DF and NMU propositional default theories it is in co-NP. As for query $q''$, it is respectively in $\Sigma_2^P$ and NP, provided that $l$ is known. Nevertheless, it is possible to show membership of $q''$ in these classes also when $l$ is unknown. Indeed, query $q''$ can be answered showing that there exists a literal $l$ in $W$ and an extension $E$ of the theory $(D, W_{S,l})$ such that $\neg S \notin E$. Thus, we can build a nondeterministic polynomial time Turing machine that guesses simultaneously the literal $l \in W_S$ and the subset $D_E \subseteq D$ of generating defaults of an extension $E$ of $(D, W_{S,l})$ together with an ordering of the rules in $D_E$, and then:

- for general propositional default theories: uses an NP oracle to (a) check the conditions that $D_E$ must satisfy to be a set of generating defaults for $E$ (see [31] or Section 2.1 for details), and to (b) verify that $\neg s_1 \wedge \ldots \wedge \neg s_n \notin E$. These steps can be performed executing a polynomially bounded number of calls to the NP oracle;

- for DF and NMU default theories: (a) checks the conditions that $D_E$ must satisfy to be a set of generating defaults for an extension $E$ of a disjunction-free theory (see [16] or Section 2.1 for details), and (b) verifies that $\neg s_1 \wedge \ldots \wedge \neg s_n \notin E$, by checking that there exists $i$, $1 \leq i \leq n$, such that $\neg s_i$ is not the conclusion of any default in $D_E$. These steps can be performed in polynomial time.

Thus, $Q2$ is the conjunction of two independent problems from $\Pi_2^P$ and $\Sigma_2^P$ for general theories, and from co-NP and NP for DF and NMU theories, and hence query $Q2$ is, respectively, in $D_2^P$ and in $D^P$.

As for hardness part, we reduce to $Q2$ the problem of deciding the problem of

$$\Delta_1 \models s_1 \wedge \Delta_2 \not\models s_2 \text{ (query } q\text{)}$$

where $\Delta_1$ and $\Delta_2$ are two independent general (resp. NMU) propositional default theories, and $s_1$ and $s_2$ are two letters. We note that $\Delta_1 \models s_1$ is a $\Pi_2^P$-complete (resp. co-NP-complete) problem, while $\Delta_2 \not\models s_2$ is a $\Sigma_2^P$-complete (resp. NP-complete) problem. In particular, we associate to $q$ the theory $\Delta(q) = (D(q), W(q))$ such that $\neg s_1, s_2 \in W(q)$, and $q$ is true iff $\{\neg s_1\}$ is an outlier witness set for $s_2$ in $\Delta(q)$ (see Theorems 3.10, 3.11, and 3.12 for details).

Finally, we consider query $Q3$. This query is important as it may constitute the basic operator to be implemented in a system of outlier detection on propositional default theories. Indeed, given a default theory, set of literals $S$ and a literal $l$, this query simply asks whether $S$ is an outlier witness set for $l$ in the input theory.

Fixing the literal $l$ in advance does not decrease the implied computational effort and, thus, the complexity figures associated to query $Q3$ are the same as for query $Q2$.

**Theorem 3.14** *$Q3$ on general propositional default theories is $D_2^P$-complete under polynomial time transformations.*

**Theorem 3.15** *$Q3$ restricted to DF propositional default theories is $D^P$-complete under polynomial time transformations.*

**Theorem 3.16** *$Q3$ restricted to NMU propositional propositional default theories is $D^P$-complete under polynomial time transformations.*

**Theorem 3.17** *$Q3$ restricted to NU and DNU propositional default theories is in P.*

# 4 Tractable Cases

In this section we show that the class of acyclic normal unary default theories is tractable with respect to the computational tasks involved in outlier detection using default logic.

Next, we recall the complexity of the entailment problem for NU and DNU default theories, and then we define acyclic (dual) normal unary theories.

**Theorem 4.1** [16, 32] *Let $\Delta$ be a NU or a DNU propositional default theory and let $l$ be a literal. Then, the problem of deciding if $\Delta \models l$ is $\mathcal{O}(n^2)$, where $n$ is the length of the theory.*

**Definition 4.2** Let $\Delta = (D, W)$ be a NMU default theory. The *atomic dependency graph* $(V, E)$ of $\Delta$ is the directed graph such that

$V = \{l \mid l$ is a letter occurring in $\Delta\}$, and

$E = \{(x, y) \mid \frac{x:y}{y} \in D \vee \frac{x:\neg y}{\neg y} \in D \vee \frac{\neg x:y}{y} \in D \vee \frac{\neg x:\neg y}{\neg y} \in D\}$.

**Definition 4.3** A (dual) normal unary default theory is *acyclic* iff its atomic dependency graph is acyclic.

The following result, together with the polynomial time solvability of the entailment problem for (dual) normal unary theories above recalled, permit us to state the tractability of queries $Q0-Q3$ when restricted to acyclic (dual) normal unary theories. Formal proof is given in Appendix.

**Theorem 4.4** *Let $(D, W)$ be a consistent acyclic NMU default and let $l$ be a literal in $W$. Then any minimal outlier witness set for $l$ in $(D, W)$ is of size at most 1.*

**Theorem 4.5** *For the class of acyclic normal unary default theories and the class of acyclic dual normal unary default theories, queries $Q0-Q3$ can be answered in polynomial time in the size of the theory.*

**Proof:** Follows from Theorem 4.1 and Theorem 4.4. □

# 5 Related work

Research work related to what we have presented in this paper can be divided into two groups, that are, $(i)$ work done on abduction, which is quite relevant to our own, and $(ii)$ work done on outlier detection from data, which is, counterwisely, less related to concepts discusses in this paper. In the following of this section we shall first survey on papers belonging to group $(i)$ and then we shall be dealing with papers of group $(ii)$.

## 5.1 Abduction

The research on logical-based abduction [23, 8, 10] is closely related to outlier detection. In the framework of logic-based abduction, the domain knowledge is described using a logical theory $T$. A subset $X$ of hypotheses is an abduction explanation to a set of manifestations $M$ if $T \bigcup X$ is a consistent theory that entails $M$. Abduction resembles outlier detection in that it "deals" with exceptional situations.

The work most relevant to ours is perhaps the paper by Eiter, Gottlob, and Leone on abduction from default theories [11]. In that paper, the authors have presented a basic model of abduction from default logic and analyzed the complexity of the main abductive reasoning tasks. They presented two modes of abductions: one based on brave reasoning and the other on cautious reasoning. According to these authors, a default abduction problem (DAP) is a tuple $\langle H, M, W, D\rangle$ where $H$ is a set of ground literals called *hypotheses*, $M$ is a set of ground literals called *observations*, and $(D, W)$ is a default theory. The goal, in general, is to explain some observations from $M$ using some of the hypotheses, in the context of the default theory $(D, W)$. Eiter, Gottlob, and Leone suggest the following definition for a skeptical explanation:

**Definition 5.1 ([11])** *Let $P = \langle H, M, D, W\rangle$ be a DAP and let $E \subseteq H$. Then, $E$ is a* skeptical explanation *for $P$ iff*

1. $(D, W \bigcup E) \models M$*, and*
2. $(D, W \bigcup E)$ *has a consistent extension.*

There is a relationship between outliers and skeptical explanations in the context of normal default theories, as the following theorem states. The theorem also holds for ordered semi-normal default theories [12].

**Theorem 5.2** *Let $\Delta = (D, W)$ be a normal default theory, where $W$ is consistent. Let $l \in W$ and $S \subseteq W$. $S$ is an outlier witness set for $l$ iff $\{l\}$ is a minimal skeptical explanation for $\neg S$ in the DAP $P = \langle\{l\}, \neg S, D, W_{S,l}\rangle$*

**Proof:** Let $\Delta = (D, W)$ be a normal default theory. Let $l \in W$ and let $S \subseteq W$ be an outlier witness set for $l$. By the definition of an outlier, it must be the case that $(D, W_S) \models \neg S$, or in other words, $(D, W_{S,l} \bigcup \{l\}) \models \neg S$. Moreover, since $(D, W)$ is a normal default theory, so is $(D, W_{S,l} \bigcup \{l\})$. In addition, since $W$ is consistent, so is $W_S$. Hence, $(D, W_S)$ has a consistent extension. So by definition of explanation, $\{l\}$ is a skeptical explanation for $\neg S$ in the DAP $P$. Note that by the definition of an outlier, we also know that $(D, W_{S,l}) \not\models \neg S$; hence $\{l\}$ is a *minimal* explanation.

On the opposite direction, suppose $\{l\}$ is a minimal skeptical explanation for $\neg S$ in the DAP $P = \langle\{l\}, \neg S, D, W_{S,l}\rangle$. By definition, we know that:

1. $(D, W_S) \models \neg S$, and
2. $(D, W_S)$ has a consistent extension.

Moreover, since $\{l\}$ is a *minimal* explanation, at least one of the following must be true:

1. $(D, W_{S,l}) \not\models \neg S$, or
2. $(D, W_{S,l})$ does not have a consistent extension.

Since $\Delta = (D, W)$ is a normal default theory and $W$ is a consistent theory, it must be the case that $\Delta = (D, W_{S,l})$ is also a normal default theory and $W_{S,l}$ is consistent. Hence, the default theory $(D, W_{S,l})$ has a consistent extension. So it must be the case that $(D, W_{S,l}) \not\models \neg S$. Therefore we can conclude that $S$ is an outlier witness set for $l$ in $(D, W)$.

$\square$

Hence, we can say that $S$ is an outlier witness for $l$ if $l \in W$, $l$ is a skeptical explanation for $S$, but still $\neg S$ holds in every extension of the theory.

It is clear that, by Theorem 5.2, there exists a sort of duality relationship between outlier detection and abduction with propositional default theories. This is due to the fact that in outlier detection problems we have to guess the outlier witness set $S$, which then plays the role of observations in Theorem 5.2, while observations in abduction constitutes a part of the input. Furthermore, in abduction problems it is needed to guess an explanation, i.e. a subset of the hypotheses, whose role in Theorem 5.2 is, on the contrary, played by the outlier $l$, and we have seen, in Section 3, that in outlier detection knowing the outlier in advance does not relieve any source of complexity.

Despite this close relationship between the two problems, we notice that the construction given in the proof of Theorem 5.2 does not depict a technique to solve outlier detection problems using abduction, since for outlier detection we have to single out both the outlier $l$ and its outlier witness set $S$, while in abduction both hypotheses and observations are fixed sets. Indeed, outlier detection is a knowledge discovery technique: the task in outlier detection is to lean who the exceptionals (the outliers), or the suspects, if you wish, are, and to justify the suspicion (that is, list the outlier witnesses). Rather, we believe that this result emphasizes the common property of these two techniques, that is the fact that both deal with exceptional situations.

## 5.2 Outlier detection from data

The literature concerning outlier detection is mainly related to the statistical, machine learning and data mining fields, hence, in almost all cases the approaches presented deal with data that can be organized in a single relational table, often with all the attributes being numerical, while a metrics relating each pair of rows in the table is required. The approaches to outlier detection can be classified in *supervised*-learning based methods, where each example must be labelled as exceptional or not [19, 26], and the *unsupervised*-learning based ones, where the label is not required. The latter approach is more general because in real situations we do not have such information. As the technique proposed in this work is unsupervised, in the following we deal only with unsupervised methods. Unsupervised-learning based methods for outlier detection can be categorized in several approaches.

The first is *statistical-based* and assumes that the given data set has a distribution model. Outliers are those objects that satisfies a discordancy test, that is that are significantly larger (or smaller) in relation to the hypothesized distribution [4].

*Deviation-based* techniques identify outliers by inspecting the characteristics of objects and consider an object that deviates from these features an outlier [3, 27].

A completely different approach that finds outliers by observing *low dimensional projections* of the search space is presented in [1]. Thus a point is considered an outlier, if it is located in some low density subspace.

Yu et al. [9] introduced a method based on *wavelet transform*, that identifies outliers by removing clusters from the original data set. Wavelet transform has also been used in [30] to detect outliers in stochastic processes.

Another category is the *density-based*, presented in [7] where a notion of *local outlier* is introduced that measures the degree of an object to be an outlier with respect to the density of the local neighborhood. To reduce the computational load, Jin et al. in [14] proposed a method to determine only the top-$n$ local outliers.

*Distance-based* outlier detection has been introduced by Knorr and Ng [17, 18] to overcome the limitations of statistical methods. A *distance-based* outlier is defined as follows: *A point $p$ in a data set is an outlier with respect to parameters $k$ and $\delta$ if at least $k$ points in the data set lies greater than distance $\delta$ from $p$.* This definition generalizes the definition of outlier in statistics and it is suitable when the data set does not fit any standard distribution. Ramaswamy et al. [24] modified this definition of outlier, as it does not provide a ranking of the outliers. The new definition of outlier is based on the distance of the $k$-th nearest neighbor of a point $p$, denoted with $D^k(p)$, and it is the following: *Given a $k$ and $n$, a point $p$ is an outlier if no more than n-1 other points in the data set have a higher value for $D^k$ than $p$.* This means that the top $n$ points having the maximum $D^k$ values are considered outliers. In [2] a new definition of outlier that considers for each point the sum of the distances from its $k$ nearest neighbors is proposed. The authors presented an algorithm using the Hilbert space-filling curve that exhibits scaling results close to linear. In [5] a near linear time algorithm for the detection of distance-based outliers exploiting randomization is presented.

The general differences and analogies between the approaches described above and our own should be quite understood. In fact, those approached deal with knowledge, as encoded within one single relational table that is, in a sense, flat, i.e. such that does there not exist any explicit relationship linking the objects (tuples) of the data set under examination. Vice versa, the technique proposed in this work deal with complex knowledge bases, which may well comprise relational-like information, but generally also include semantically richer forms of knowledge, such as axioms, default rules and so forth: in this latter case several complex relations relating objects (atoms) of the underlying theory are explicitly available. As a consequence, even if the intuitive and general sense of computing outliers in the two contexts is analogous, the specific definitions that are used are quite different as well as different are the formal properties of computed outliers. Even disregarding such important distinctions in the semantics associated with the concept of outlier detection within the two frameworks, there are further differences thereof. For instance, being our reference framework far richer than that of relational table, it turns out that outlier detection in our context is, from the computational point of view, much more difficult: in fact, the most complex outlier detection tasks that we address are $\Sigma_3^P$-

complete, while almost all outlier detection problems within the relational data context are polynomial time solvable and only few of them are NP-complete.

# 6 Conclusion

Suppose you are walking in the street and you see a blind person walking in the opposite direction. You believe he is blind because he is feeling his way with a walking stick. Suddenly, something falls out of his bag, and to your surprise, he finds it immediately without probing around with his fingers, as is customary for a blind person. This kind of behavior will render the "blind" person walking towards you suspicious.

The purpose of this paper has been to formally mimic this type of reasoning using default logic. We have formally defined the notion of outlier and outlier witness, and analyzed the complexities involved, pointing out some non-trivial tractable cases. As explained in the introduction, outlier detection can also be used for maintaining knowledge base integrity and completeness.

This work can be extended in several ways. First, we can develop the concept of outliers in other frameworks of default databases, like System Z [22] and Circumscription [20]. Second, we can look for intelligent heuristics that will enable us to perform the heavy computational task involved more efficiently. Third, we can study the problem from the perspective of looking at default theories as "semantic check tool-kit" for relational databases.


### Acknowledgments

The authors gratefully thank Michael Gelfond for some fruitful discussions on the subject of the paper and Francesco Scarcello for providing insights about some of the computational complexity tools we have used.


# References


[1] C. C. Aggarwal and P.S. Yu. Outlier detection for high dimensional data. In *Proc. ACM Int. Conference on Managment of Data (SIGMOD'01)*, 2001.

[2] F. Angiulli and C. Pizzuti. Fast outlier detection in high dimensional spaces. In *Proc. Int. Conf. on Principles of Data Mining and Knowledge Discovery (PKDD 2002)*, pages 15–26, 2002.

[3] A. Arning, C. Aggarwal, and P. Raghavan. A linear method for deviation detection in large databases. In *Proc. Int. Conf. on Knowledge Discovery and Data Mining (KDD'96)*, pages 164–169, 1996.

[4] V. Barnett and T. Lewis. *Outliers in Statistical Data.* John Wiley & Sons, 1994.

[5] S.D. Bay and M. Schwabacher. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In *Proc. Int. Conf. on Knowledge Discovery in Databases (KDD 2003)*, 2003.

[6] Rachel Ben-Eliyahu and Rina Dechter. Propositional semantics for disjunctive logic programs. *Annals of Mathematics and Artificial Intelligence*, 12:53–87, 1994. A short version appears in Proceedings of the Joint International Conference and Symposium on Logic Programming, 1992.

[7] M. M. Breunig, H. Kriegel, R.T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In *Proc. ACM Int. Conf. on Managment of Data (SIGMOD'00)*, 2000.

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

[9] Yu D., Sheikholeslami S., and A. Zhang. Findout: Finding outliers in very large datasets. In *Tech. Report, 99-03, Univ. of New York, Buffalo*, pages 1–19, 1999.

[10] Thomas Eiter and Georg Gottlob. The complexity of logic-based abduction. *JACM*, 42(1):3–42, 1995.

[11] Thomas Eiter, Georg Gottlob, and Nicola Leone. Semantics and complexity of abduction from default theories. *Artificial Intelligence*, 90(1-2):177–223, 1997.

[12] David W. Etherington. Formalizing nonmonotonic reasoning systems. *Artificial Intelligence*, 31:41–85, 1987.

[13] George Gottlob. Complexity results for nonmonotonic logics. *The Journal of Logic and Computation*, 2(3):397–425, 1992.

[14] H.V. Jagadish. Linear clustering of objects with multiple atributes. In *Proc. ACM Int. Conf. on Managment of Data (SIGMOD'90)*, pages 332–342, 1990.

[15] D. S. Johnson. A catalog of complexity classes. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science,* Volume A: Algorithms and Complexity, chapter 9, pages 67–161. Elsevier and The MIT Press (co-publishers), 1990.

[16] Henry A. Kautz and Bart Selman. Hard problems for simple default logics. *Artificial Intelligence*, 49:243–279, 1991.

[17] E. Knorr and R. Ng. Algorithms for mining distance-based outliers in large datasets. In *Proc. Int. Conf. on Very Large Databases (VLDB98)*, pages 392–403, 1998.

[18] E. Knorr, R. Ng, and V. Tucakov. Distance-based outlier: algorithms and applications. *VLDB Journal*, 8(3-4):237–253, 2000.

[19] W. Lee, S.J. Stolfo, and K.W. Mok. Mining audit data to build intrusion detection models. In *Proc. Int. Conf on Knowledge Discovery and Data Mining (KDD-98)*, pages 66–72, 1998.

[20] John McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[21] Christos H. Papadimitriou. *Computatational Complexity.* Addison-Wesley, Reading, Mass., 1994.

[22] Judea Pearl. System z: a natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 121–135, Monterey, CA, 1990.

[23] David Poole. Normality and faults in logic-based diagnosis. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1304–1310, 1989.

[24] S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. In *Proc. ACM Int. Conf. on Managment of Data (SIGMOD'00)*, pages 427–438, 2000.

[25] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[26] S. Rosset, U. Murad, E. Neumann, Y. Idan, and G. Pinkas. Discovery of fraud rules for telecommunications-challenges and solutions. In *Proc. Int. Conf on Knowledge Discovery and Data Mining (KDD-99)*, pages 409–413, 1999.

[27] S. Sarawagi, R. Agrawal, and N. Megiddo. Discovery-driven exploration of olap data cubes. In *Proc. Sixth Int. Conf on Extending Database Thecnology (EDBT)*, Valencia, Spain, March 1998.

[28] Jonathan Stillman. The complexity of propositional default logics. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 794–799, 1992.

[29] Jonathan Stillman. The complexity of propositional default logics. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 794–800, San Jose, CA, July 1992.

[30] Z.R. Struzik and A. Siebes. Outliers detection and localisation with wavelet based multifractal formalism. In *Tech. Report, CWI,Amsterdam, INS-R0008*, 2000.

[31] A. Zhang and W. Marek. On the classification and existence of structures in default logic. *Fundamenta Informaticae*, 13(4):485–499, 1990.

[32] Rachel Ben-Eliyahu Zohary. Yet some more complexity results for default logic. *Artificial Intelligence*, 139(1):1–20, 2002.

# Appendix: Proofs

Next, we report the proofs of the theorems discussed in Sections 3 and 4. Before starting, we introduce some notations that will be used in the following.

Let $L$ be a consistent set of literals. Then we denote with $\mathcal{T}_L$ the truth assignment on the set of letters occurring in $L$ such that, for each positive literal $p \in L$, $\mathcal{T}_L(p) = \mathbf{true}$, and for each negative literal $\neg p \in L$, $\mathcal{T}_L(p) = \mathbf{false}$.

Let $T$ be a truth assignment of the set $x_1, \ldots, x_n$ of variables. Then we denote with $Lit(T)$ the set of literals $\{\ell_1, \ldots, \ell_n\}$, such that $\ell_i$ is $x_i$ if $T(x_i) = \mathbf{true}$ and is $\neg x_i$ if $T(x_i) = \mathbf{false}$, for $i = 1, \ldots, n$.

## Proofs of Section 3

### Query Q0

**Theorem 3.1:** $Q0$ on general propositional default theories is $\Sigma_3^P$-complete under polynomial time transformations.

**Proof of Theorem 3.1:** (Membership) Given a a theory $\Delta = (D, W)$, we must show that there exists a literal $l$ in $W$ and a subset $S = \{s_1, \ldots, s_n\}$ of $W$ such that $(D, W_S) \models \neg s_1 \wedge \ldots \wedge \neg s_n$ (query $q'$) and $(D, W_{S,l}) \not\models \neg s_1 \wedge \ldots \wedge s_n$ (query $q''$). Query $q'$ is $\Pi_2^P$-complete, while query $q''$ is $\Sigma_2^P$-complete [13, 29]. Thus, we can build a polynomial-time nondeterministic Turing machine with a $\Sigma_2^P$ oracle, solving query $Q0$ as follows: the machine guesses both the literal $l$ and the set $S$ and then solves queries $q'$ and $q''$ by two calls to the oracle.

(Hardness) Let $\Phi = \exists X \forall Y \exists Z f(X, Y, Z)$ be a quantified boolean formula, where $X = x_1, \ldots, x_n$, $Y = y_1, \ldots, y_m$, and $Z$ are disjoint set of variables. We associate with $\Phi$ the default theory $\Delta(\Phi) = (D(\Phi), W(\Phi))$, where $W(\Phi)$ is the set of letters $\{l, s_1, \overline{s}_1, \ldots, s_n, \overline{s}_n\}$ consisting of new letters distinct from those occurring in $\Phi$, and $D(\Phi) = D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5$:

$$D_1 = \left\{ \delta_{1,i} = \frac{: \neg s_i \wedge x_i \wedge e_i}{x_i \wedge e_i}, \overline{\delta}_{1,i} = \frac{: \neg \overline{s}_i \wedge \neg x_i \wedge e_i}{\neg x_i \wedge e_i} \mid i = 1, \ldots, n \right\}$$

$$D_2 = \left\{ \delta_{2,i} = \frac{: \neg s_i \wedge \neg \overline{s}_i \wedge \neg \alpha \wedge \beta}{\beta} \mid i = 1, \ldots, n \right\} \cup \left\{ \delta_2 = \frac{\beta : \alpha}{\alpha} \right\}$$

$$D_3 = \left\{ \delta_{3,j} = \frac{: y_j}{y_j}, \overline{\delta}_{3,j} = \frac{: \neg y_j}{\neg y_j} \mid j = 1, \ldots, m \right\}$$

$$D_4 = \left\{ \delta_4 = \frac{l \wedge e_1 \wedge \ldots \wedge e_n : f(X, Y, Z) \wedge g}{g} \right\}$$

$$D_5 = \left\{ \delta_{5,i} = \frac{g : \neg s_i}{\neg s_i}, \overline{\delta}_{5,i} = \frac{g : \neg \overline{s}_i}{\neg \overline{s}_i} \mid i = 1, \ldots, n \right\}$$

where also $\alpha$, $\beta$, $g$, $e_1, \ldots, e_n$ are new letters distinct from those occurring in $\Phi$. Clearly, $W(\Phi)$ is consistent and $\Delta(\Phi)$ can be built in polynomial time. Now we show that $\Phi$ is valid iff there exists an outlier in $\Delta(\Phi)$.

In the rest of the proof we denote by $\sigma(s_i)$ ($\widehat{\sigma}(x_i)$ resp.) the literal $x_i$ ($s_i$ resp.) and by $\sigma(\overline{s}_i)$ ($\widehat{\sigma}(\neg x_i)$ resp.) the literal $\neg x_i$ ($\overline{s}_i$ resp.), for $i = 1, \ldots, n$. Let $S$ be a subset of $\{s_1, \overline{s}_1, \ldots, s_n, \overline{s}_n\}$ ($\{x_1, \neg x_1, \ldots, x_n, \neg x_n\}$ resp.), we denote by $\sigma(S)$ ($\widehat{\sigma}(S)$ resp.) the set $\{\sigma(s) \mid s \in S\}$ ($\{\widehat{\sigma}(s) \mid s \in S\}$ resp.).

**Claim 1** *Let $S = \{s'_1, \ldots, s'_n\}$, with $s'_i$ either $s_i$ or $\overline{s}_i$, for $i = 1, \ldots, n$, and let $E$ be an extension of $(D(\Phi), W(\Phi)_S)$ and $D_E$ its associated set of generating defaults. Then:*

1. *$D_E \cap D_1 = \{\delta'_{1,i} \mid i = 1, \ldots, n\}$, where $\delta'_{1,i}$ is either $\delta_{1,i}$ or $\overline{\delta}_{1,i}$ depending on $s'_i$ being $s_i$ or $\overline{s}_i$, for $i = 1, \ldots, n$;*
2. *$D_E \cap D_2 = \emptyset$;*
3. *$D_E \cap D_4$ is $\{\delta_4\}$ if $\neg f(X, Y, Z) \notin E$, and $\emptyset$ otherwise;*
4. *$D_E \cap D_5$ is $\{\delta'_{5,i} \mid i = 1, \ldots, n\}$, where $\delta'_{5,i}$ is either $\delta_{5,i}$ or $\overline{\delta}_{5,i}$ depending on $s'_i$ being $s_i$ or $\overline{s}_i$, for $i = 1, \ldots, n$, if $\neg f(X, Y, Z) \notin E$, and $\emptyset$ otherwise.*

**Proof of Claim 1:** (1) and (2) are immediate. As for (3) and (4) simply note that the precondition of rule $\delta_4$ always belong to $E$, because $e_1, \ldots, e_n \in E$ by rules in the set $D_E \cap D_1$. Thus $\delta_4 \in D_E$ and $g \in E$ iff $\neg f(X, Y, Z) \notin E$. □

The previous claim states that the set $S$, together with the formula $f(X, Y, Z)$, uniquely identifies the generating defaults coming from the set $D(\Phi) \setminus D_3$ of an extension of $(D(\Phi), W(\Phi)_S)$. We denote the set $D_E \cap (D(\Phi) \setminus D_3)$ with $D_S(\Phi)$.

**Claim 2** *Let $S = \{s'_1, \ldots, s'_n\}$, with $s'_i$ either $s_i$ or $\overline{s}_i$, for $i = 1, \ldots, n$, and let $\ell_j$ be either $y_j$ or $\neg y_j$, for $j = 1, \ldots, m$. Then there exists a bijection between the sets $L = \{\ell_1, \ldots, \ell_m\}$ and the extensions $E_L$ of $(D(\Phi), W(\Phi)_S)$.*

**Proof of Claim 2:** ($\Rightarrow$) Consider a generic set $L$. Let $D_L$ be the set of defaults containing rule $\delta_{3,j}$, if $\ell_j = y_j$, and rule $\overline{\delta}_{3,j}$, otherwise, for $j = 1, \ldots, m$, and such that $D_L \supset D_S(\Phi)$. As for the rules of $D_S(\Phi)$ coming from the sets $D_4$ and $D_5$, take $\delta_4$ and $\delta'_{5,1}, \ldots, \delta'_{5,n}$, as defined in Claim 1, if $\mathcal{T}_{\sigma(S) \cup L}$ satisfies $f(X, Y, Z)$, and $\emptyset$ otherwise. It is easy to verify that $D_L$ is the set of generating defaults of an extension $E_L$ of $(D(\Phi), W(\Phi)_S)$ such that $E_L \supseteq L$ and that no other set of generating defaults can be associated to an extension of $(D(\Phi), W(\Phi)_S)$ containing $L$.

($\Leftarrow$) Let $E_L$ be an extension of $(D(\Phi), W(\Phi)_S)$ and let $D_L$ its associated set of generating defaults. From Claim 1, $D_L$ must contain the set $D_S(\Phi)$ and not other rule from the sets $D_1, D_2, D_4, D_5$. As for $D_3$, suppose that there exists $k \in \{1, \ldots, m\}$ such that both $\delta_{3,k}$ and $\overline{\delta}_{3,k}$ do not belong to the set $D_L$. Clearly it follows that both $\neg y_k \notin E_L$ and $y_k \notin E_L$, thus $E_L$ is not closed under the application of defaults in $D(\Phi)$, i.e. it is not an extension, a contradiction. Thus $E_L$ must contain a set $L$. Furthermore $L$ is unique, as both $y_j$ and $\neg y_j$ cannot belong to $E_L$, for $j = 1, \ldots, m$. $\square$

Thus, the extension $E_L$ associated to $L$ is the unique extension of $(D(\Phi), W(\Phi)_S)$ such that $E_L \supseteq L$. Now we can proceed with the main proof.

($\Rightarrow$) Suppose that $\Phi$ is valid. Now we show that $l$ is an outlier in $\Delta(\Phi)$. As $\Phi$ is valid, then there exists a truth assignment $T_X$ on the set $X$ of variables such that $T_X$ satisfies $\forall Y \exists Z f(X, Y, Z)$. Let $S = \widehat{\sigma}(Lit(T_X))$. It follows from Claim 2 that we can associate to each truth assignment $T_Y$ on the set $Y$ of variables, one and only one extension $E_Y$ of $(D(\Phi), W(\Phi)_S)$. In particular, $E_Y \supseteq Lit(T_X) \cup Lit(T_Y)$. As $\Phi$ is valid, then $\neg f(X, Y, Z) \notin E_Y$ and $E_Y \models \neg S$. Furthermore, from Claim 2, these are all the extensions of $(D(\Phi), W(\Phi)_S)$ and thus $(D(\Phi), W(\Phi)_S) \models \neg S$.

Consider now the theory $(D(\Phi), W(\Phi)_{S,l})$. We note that the literal $l$ appears in the precondition of rule $\delta_4$, whose conclusion $g$ represents, in turn, the precondition of the rules in the set $D_5$, rules that allow to conclude $\neg S$, and that $l$ does not appear in the conclusion of any rule of $D(\Phi)$. Thus $(D(\Phi), W(\Phi)_{S,l}) \not\models \neg S$.

Hence $l$ is an outlier in $\Delta(\Phi)$.

**Claim 3** *Let $S \subseteq W(\Phi)$ be an outlier witness for a literal $o \in W(\Phi)$ in $\Delta(\Phi)$. Then $S = \{s'_1, \ldots, s'_n\}$, where $s'_i$ is either $s_i$ or $\overline{s}_i$, for $i = 1, \ldots, n$.*

**Proof of Claim 3:** First, we note that $l$ cannot belong to $S$ as $\neg l$ does not appear in the consequence of any default of $D(\Phi)$.

Suppose that there exists $k \in \{1, \ldots, n\}$ such that both $s_k$ and $\overline{s}_k$ occur in $S$. Then the default $\delta_{2,k}$ adds the special letter $\beta$ to the candidate extension. But rule $\delta_2$, having $\beta$ as precondition, has a conclusion inconsistent with the justification of $\delta_{2,k}$, the rule which added $\beta$. Hence $(D(\Phi), W(\Phi)_{S,o})$ is incoherent and $(D(\Phi), W(\Phi)_{S,o}) \models \neg S$, a contradiction.

Now, suppose that there exists $a \in \{1, \ldots, n\}$, such that both $s_a$ and $\overline{s}_a$ do not occur in $S$. In this case $S$ cannot be an outlier witness in $\Delta(\Phi)$. Indeed, the previous condition implies that $e_a$ cannot belong to every extension of $(D(\Phi), W(\Phi)_S)$. Furthermore, as $S$ is non empty by definition, then there exists $b \in \{1, \ldots, n\}$ such that either $s_b$ or $\overline{s}_b$ belong to $S$, call this letter $s'_b$. In order that $(D(\Phi), W(\Phi)_S) \models \neg s'_b$, it is needed that either rule $\delta_{5,b}$ or $\overline{\delta}_{5,b}$ belongs to the set of generating defaults of every extension of this theory. But the precondition of this rule is the letter $g$, which, in turn, is the consequence of rule $\delta_4$ having $e_a$ among its preconditions. We note that $g \notin W(\Phi)_S$ and that $g$ appears only in rule $\delta_4$ as a consequence. We have previously stated that $e_a$ cannot belong to every extension of $(D(\Phi), W(\Phi)_S)$, thus we can conclude that $S$ is not an outlier witness in $\Delta(\Phi)$, a contradiction. $\square$

**Claim 4** *Let $s \in W(\Phi) \setminus \{l\}$. Then $s$ is not an outlier in $\Delta(\Phi)$.*

**Proof of Claim 4:** By contradiction, suppose that there exists $s \in W(\Phi) \setminus \{l\}$ such that $s$ is an outlier in $\Delta(\Phi)$. Then, there exists $S \subseteq W(\Phi) \setminus \{l, s\}$ such that $S$ is an outlier witness for $s$ in $\Delta(\Phi)$. From Claim 3, $\overline{s} \in S$, where $\overline{s}$ is either $\overline{s}_k$, if $s = s_k$, or $s_k$, if $s = \overline{s}_k$ ($k \in \{1, \dots, n\}$). But this implies that $(D(\Phi), W(\Phi)_{S,s})$ is incoherent, thus $(D(\Phi), W(\Phi)_{S,s}) \models \neg S$, a contradiction. $\square$

($\Leftarrow$) Suppose that there exists an outlier in $\Delta(\Phi)$. Then, from Claim 4, it must be equal to $l$. Hence, there exists a non empty set of literals $S \subseteq W(\Phi) \setminus \{l\}$ such that $S$ is an outlier witness for $l$ in $\Delta(\Phi)$. From Claim 3, $S = \{s'_1, \dots, s'_n\}$, where $s'_i$ is either $s_i$ or $\overline{s}_i$, for $i = 1, \dots, n$. Now we show that $\mathcal{T}(\sigma(S))$ satisfies $\forall Y \exists Z f(X, Y, Z)$, i.e. that $\Phi$ is valid. From Claim 2, for each set $L = \{\ell_1, \dots, \ell_m\}$ there exists one extension $E_L$ of $(D(\Phi), W(\Phi)_S)$ such that $E_L \supseteq L$. We note also that $E_L \supseteq \sigma(S)$. Thus, in order to be $l$ an outlier in $\Delta(\Phi)$, it must be the case that, for each set $L$, $\neg f(X, Y, Z) \notin E_Y$, i.e. that $\mathcal{T}(\sigma(S)) \circ \mathcal{T}(L)$ satisfies $f(X, Y, Z)$. Hence, we can conclude that $\Phi$ is valid. $\square$

**Theorem 3.2:** $Q0$ restricted to DF propositional default theories is $\Sigma_2^P$-complete under polynomial time transformations.
**Proof of Theorem 3.2:** This result follows immediately from Theorem 3.3. $\square$

**Lemma 1** *Let $\Delta$ be a NMU propositional default theory and let $l_1, \dots, l_m$ be a set of literals. Then the entailment problem $\Delta \models l_1 \wedge \dots \wedge l_m$ is co-NP-complete under polynomial time transformations.*

**Proof of Lemma 1:** (Membership) Membership in co-NP follows from the membership in co-NP of the entailment problem for disjunction free propositional default theories [16].

(Hardness) Let $\Phi$ be a boolean formula on the set of variables $X = x_1, \dots, x_n$, such that $\Phi = C_1 \wedge \dots \wedge C_r$, with $C_k = t_{k,1} \vee \dots \vee t_{k,u_k}$. and each $t_{k,1}, \dots, t_{k,u_k}$ is a literal on the set $X$, for $k = 1, \dots, r$. We associate to $\Phi$ the default theory $\Delta(\Phi) = (D(\Phi), \emptyset)$, where $\Delta(\Phi)$ is $D_1 \cup D_2 \cup D_3$ and

$$D_1 = \left\{ \frac{: x_i}{x_i}, \frac{: \neg x_i}{\neg x_i} \mid i = 1, \dots, n \right\}$$

$$D_2 = \left\{ \frac{t_{k,j} : c_k}{c_k} \mid k = 1, \dots, r; j = 1, \dots, u_k \right\}$$

$$D_3 = \left\{ \frac{: \neg c_k}{\neg c_k}, \frac{\neg c_k : l_1}{l_1}, \dots, \frac{\neg c_k : l_m}{l_m} \mid k = 1, \dots, r \right\}$$

where $l_1, \dots, l_m$ are new letters distinct from those occurring in $\Phi$. Now we show that $\Phi$ is unsatisfiable iff $\Delta(\Phi) \models l_1 \wedge \dots \wedge l_m$.

Consider a generic extension $E$ of $\Delta(\Phi)$. From the rules in the set $D_1$, $E$ is such that, for each $i = 1, \dots, n$, either $x_i \in E$ or $\neg x_i \in E$.

($\Rightarrow$) Suppose that $\Phi$ is unsatisfiable. Then for each truth assignment $T$ on the set of variables $X$ there exists at least a clause $C_{f(T)}$, $1 \leq f(T) \leq r$, that is not satisfied by $T$. From rules in the set $D_2$ and from what above stated, $c_{f(\mathcal{T}_{E \cap (X \cup \neg X)})} \notin E$, and from rules in the set $D_3$, $\neg c_{f(\mathcal{T}_{E \cap (X \cup \neg X)})} \in E$ and $l_1, \dots, l_m \in E$.

($\Leftarrow$) Suppose that $\Delta(\Phi) \models l_1 \wedge \ldots \wedge l_m$. Then it is the case that, for each extension $E$ of $\Delta(\Phi)$ there exists $g(E)$, $1 \leq g(E) \leq r$, such that $\neg c_{g(E)} \in E$. For each truth assignment $T$ on the set of variables $X$, let $\mathcal{E}(T)$ denote the set containing all the extensions $E$ of $\Delta(\Phi)$ such that $E \supseteq Lit(T)$. Then, for each $E \in \mathcal{E}(T)$, $\neg c_{g(E)} \in E$, implies that the none of the rules in the set $D_2$ having $c_{g(E)}$ as their conclusion belong to the set of generating defaults of $E$. Thus, the clause $C_{g(E)}$ is not satisfied by $T$. As this holds for each truth assignment $T$, then it is the case that $\Phi$ is unsatisfiable. □

**Theorem 3.3:** $Q0$ restricted to NMU propositional default theories is $\Sigma_2^P$-complete under polynomial time transformations.

**Proof of Theorem 3.3:** (Membership) Given a NMU theory $\Delta = (D, W)$, we must show that there exists a literal $l$ in $W$ and a subset $S = \{s_1, \ldots, s_n\}$ of $W$ such that $(D, W_S) \models \neg s_1 \wedge \ldots \wedge \neg s_n$ (query $q'$) and $(D, W_{S,l}) \not\models \neg s_1 \wedge \ldots \wedge s_n$ (query $q''$). Query $q'$ is NP-complete, while query $q''$ is co-NP-complete (see Lemma 1 above). Thus, we can build a polynomial-time nondeterministic Turing machine with a NP oracle, solving query $Q0$ as follows: the machine guesses both the literal $l$ and the set $S$ and then solves queries $q'$ and $q''$ by two calls to the oracle.

(Hardness) Let $\Phi = \exists X \forall Y f(X, Y)$ be a quantified boolean formula in conjunctive normal form, where $X = x_1, \ldots, x_n$ and $Y = y_1, \ldots, y_m$ are disjoint set of variables, and $f(X, Y) = C_1 \wedge \ldots \wedge C_r$, with $C_k = t_{k,1} \vee \ldots \vee t_{k,u_k}$, and each $t_{k,1}, \ldots, t_{k,u_k}$ is a literal on the set $X \cup Y$, for $k = 1, \ldots, r$. We associate to $\Phi$ the default theory $\Delta(\Phi) = (D(\Phi), W(\Phi))$, where $W(\Phi)$ is the set $\{l, x_1, \ldots, x_n, c_1, \ldots, c_r\}$ of letters, with $l$ and $c_1, \ldots, c_r$ new letters distinct from those occurring in $\Phi$, and $D(\Phi) = D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5$:

$$D_1 = \left\{ \delta_{1,i} = \frac{: \neg x_i}{\neg x_i}, \delta'_{1,i} = \frac{\chi : x_i}{x_i} \mid i = 1, \ldots, n \right\}$$

$$D_2 = \left\{ \delta_{2,i} = \frac{: y_j}{y_j}, \overline{\delta}_{2,i} = \frac{: \neg y_j}{\neg y_j} \mid j = 1, \ldots, m \right\}$$

$$D_3 = \left\{ \delta_{3,k}^{(h)} = \frac{t_{k,h} : \neg c_k}{\neg c_k} \mid k = 1, \ldots, r;\ h = 1, \ldots, u_k \right\}$$

$$D_4 = \left\{ \delta_4 = \frac{: \neg l}{\neg l}, \delta'_4 = \frac{\neg l : \chi}{\chi} \right\}$$

$$D_5 = \left\{ \delta_{5,k} = \frac{c_k : \chi}{\chi}, \delta'_{5,k} = \frac{\chi : c_k}{c_k} \mid k = 1, \ldots, r \right\}$$

where also $\chi$ is a new letter distinct from those occurring in $\Phi$. Clearly, $W(\Phi)$ is consistent and $\Delta(\Phi)$ can be built in polynomial time. Now we show that $\Phi$ is valid iff there exists an outlier in $\Delta(\Phi)$.

Given a set of literals $S$, in the rest of the proof we denote by $\sigma(S)$ the set of literals $(X \setminus S) \cup \neg(X \cap S)$.

**Claim 5** *Let $S$ be a set of letters such that $\{c_1, \ldots, c_r\} \subseteq S \subseteq W(\Phi) \setminus \{l\}$, and let $E$ be an extension of $(D(\Phi), W(\Phi)_S)$ and $D_E$ its associated set of generating defaults. Then:*

1. $D_E \cap D_1 = \{\delta_{1,v} \mid x_v \in S\}$
2. $D_E \cap D_3 \supseteq \{\delta_{3,v}^{(w)} \mid t_{v,w} \in \sigma(S)\}$
3. $D_E \cap D_4 = \emptyset$
4. $D_E \cap D_5 = \emptyset$

**Proof of Claim 5:** We start from (3): $l \in W(\Phi)_S$ implies that both $\delta_4, \delta'_4 \notin E$. (4) We note that $c_k \notin W(\Phi)_S$, and that $c_k$ is the consequence of $\delta'_{5,k}$ having $\chi$ as precondition, for each $k = 1, \ldots, r$. But $\chi$ is the consequence of $\delta'_4$, that do not belong to $E$ as shown above, and of $\delta_{5,k}$, for $k = 1, \ldots, r$. Thus $c_k \notin W(\Phi)_S$ implies that $\delta'_{5,k} \notin D_E$, and in turn that $\delta_{5,k} \notin D_E$, for each $k = 1, \ldots, r$. (1) As $\chi \notin E$, then $\delta'_{1,i} \notin D_E$, for each $i = 1, \ldots, n$, and $x_v \in S$ implies that $x_v \notin E$. Thus, for each $x_v \in S$, $\delta_{1,v} \in D_E$. (2) From what above stated, any extension $E$ of $(D(\Phi), W(\Phi)_S)$ contains the set of literals $\sigma(S)$, while $c_k \notin E$, for $k = 1, \ldots, r$. $\square$

The previous claim states that the set $S$ uniquely identifies the generating defaults of the theory $(D(\Phi), W(\Phi)_S)$ coming from the sets $D_1$, $D_4$, and $D_5$, together with those coming from the set $D_3$ which have a literal on the set $X$ as their precondition. We denote this set of generating defaults with $D_S(\Phi)$.

**Claim 6** *Let $S$ be a set of literals such that $\{c_1, \ldots, c_r\} \subseteq S \subseteq W(\Phi) \setminus \{l\}$, and let $\ell_j$ be either $y_j$ or $\neg y_j$, for $j = 1, \ldots, m$. Then there exists a bijection between the sets $L = \{\ell_1, \ldots, \ell_m\}$ and the extensions $E_L$ of $(D(\Phi), W(\Phi)_S)$.*

**Proof of Claim 6:** ($\Rightarrow$) Consider a generic set $L$. Let $D_L$ be the set of defaults (a) containing rule $\delta_{2,j}$, if $\ell_j = y_j$, and rule $\overline{\delta}_{2,j}$, otherwise, for $j = 1, \ldots, m$, (b) such that $D_L \supseteq \{\delta_{3,v}^{(w)} \mid t_{v,w} \in L\}$, and (c) such that $D_L \supseteq D_S(\Phi)$. It is easy to verify that $D_L$ is the set of generating defaults of an extension $E_L$ of $(D(\Phi), W(\Phi)_S)$ such that $E_L \supseteq L$ and that no other set of generating defaults can be associated to an extension of the same theory containing $L$.

($\Leftarrow$) Let $E_L$ be an extension of $(D(\Phi), W(\Phi)_S)$ and let $D_L$ be its associated set of generating defaults. From Claim 5, $D_L$ must contain the set $D_S(\Phi)$. It follows immediately from the rules in the set $D_2$, that $E_L$ must contain an unique set $L$. Furthermore, the rules of $D_L$ coming from $D_3$ but not in $D_S(\Phi)$ are uniquely identified by $L$, as $c_k \notin E_L$, for $k = 1, \ldots, r$. $\square$

**Claim 7** *Let $S$ be a subset of $W(\Phi)$, and $\Delta' = (D(\Phi), W(\Phi)_S)$. Then*

1. $S \supseteq \{l\}$, *or*
2. $S \not\supseteq \{c_1, \ldots, c_r\}$

*implies that* $\Delta' \not\models \neg s$ *for each* $s \in W(\Phi) \setminus \{l\}$.

**Proof of Claim 7:** First, we note that, if the letter $\chi$ belong to $E$, then there exists an extension $E$ of $\Delta'$ such that $E \supseteq \{x_1, \ldots, x_n, c_1, \ldots, c_r\}$ obtained applying both defaults $\delta'_{1,i}$, for $i = 1, \ldots, n$, and $\delta'_{5,k}$, for $k = 1, \ldots, r$. To conclude the proof (1) if $l \in S$ then $\chi \in E$ by rules $\delta_4$ and $\delta'_4$, and (2) if exists $\overline{k} \in \{1, \ldots, r\}$ such that $c_{\overline{k}} \notin S$ then $\chi \in E$ by rule $\delta_{5,\overline{k}}$. $\square$

($\Rightarrow$) Suppose that $\Phi$ is valid. Then there exists a truth assignment $T_X$ on the set $X$ of variables such that $T_X$ satisfies $\forall Y f(X,Y)$. Let $S = \{s \in X \mid T_X(s) = \textbf{false}\} \cup \{c_1, \ldots, c_r\}$. Now we show that $S$ is an outlier witness for $l$ in $\Delta(\Phi)$.

From Claim 6 it follows that we can associate to each truth assignment $T_Y$ on the set $Y$ of variables, one and only one extension $E_Y$ of $(D(\Phi), W(\Phi)_S)$. In particular, $E_Y \supseteq Lit(T_X) \cup Lit(T_Y)$. As $\Phi$ is valid, then at least a literal in each clause of $\Phi$ must be true, thus $\neg c_k \in E_Y$, for $k = 1, \ldots, r$. We note that the rules $\delta_{1,i}$, $i \in \{1, \ldots, n\}$, add to $E_Y$ the negation of the variables in $X \cap S$. We have above stated that these are all the extensions of $(D(\Phi), W(\Phi)_S)$ and thus $(D(\Phi), W(\Phi)_S) \models \neg S$.

To conclude the proof, from Claim 7 it follows that $(D(\Phi), W(\Phi)_{S,l}) \not\models \neg S$. Hence $l$ is an outlier in $\Delta(\Phi)$.

**Claim 8** *Let* $S \subseteq W(\Phi)$ *be an outlier witness for a literal* $o \in W(\Phi)$ *in* $\Delta(\Phi)$. *Then* $\{c_1, \ldots, c_r\} \subseteq S \subseteq W(\Phi) \setminus \{l\}$.

**Proof of Claim 8:** Let $\Delta'$ be the theory $(D(\Phi), W(\Phi)_S)$ and $\Delta''$ be $(D(\Phi), W(\Phi)_{S,o})$.

Suppose that $l \in S$. From Claim 7, $l \in S$ implies that $\Delta' \not\models \neg s$, for each $s \in W(\Phi) \setminus \{l\}$. Hence, we can conclude that $l \in S$ implies that $S = \{l\}$. But, cause rule $\delta_4$ and as there not exists a rule in $D(\Phi)$ having $l$ as its consequence, both $\Delta' \models \neg l$ and $\Delta'' \models \neg l$, no matter what is the value of $o$. Thus, $\{l\}$ cannot be an outlier witness for any literal in $W(\Phi)$, and $S \subseteq W(\Phi) \setminus \{l\}$.

By absurd, suppose that $S \not\supseteq \{c_1, \ldots, c_r\}$. Then, by Claim 7 then $S$ must be empty, a contradiction. $\square$

($\Leftarrow$) Suppose that there exists an outlier $o$ in $\Delta(\Phi)$. Then there exists a nonempty set of literals $S$, $\{c_1, \ldots, c_r\} \subseteq S \subseteq W(\Phi) \setminus \{l\}$ from Claim 8, such that $S$ is an outlier witness for $o$ in $\Delta(\Phi)$.

Now we show that $\mathcal{T}_{\sigma(S)}$ satisfies $\forall Y f(X,Y)$. i.e. that $\Phi$ is valid. From Claim 6 for each set $L = \{\ell_1, \ldots, \ell_m\}$, where each $\ell_j$ is either $y_j$ or $\neg y_j$, for $j = 1, \ldots, m$, there exists one extension $E_L$ of $(D(\Phi), W(\Phi)_S)$ such that $E_L \supseteq L$. We note also that $E_L \supseteq \sigma(S)$. Thus, in order to be $o$ an outlier in $\Delta(\Phi)$, it must be the case that, for each set $L$, $\neg c_1 \wedge \ldots \wedge \neg c_r \in E_L$, i.e. that $\mathcal{T}_{\sigma(S) \cup L}$ satisfies $f(X,Y)$. Hence, we can conclude that $\Phi$ is valid.

As for the value of $o$, we note that $S$ is always an outlier witness for $o = l$ in $\Delta(\Phi)$. Indeed, consider the theory $\Delta'' = (D(\Phi), W(\Phi)_{S,l})$. It follows from Claim 7 that $\Delta'' \not\models \neg S$.

Hence, we can conclude that $\Phi$ is valid. $\square$

**Theorem 3.4:** $Q0$ restricted to propositional NU default theories is NP-complete under polynomial time transformations.

**Proof of Theorem 3.4:** (Membership) Consider a normal unary default theory $\Delta = (D, W)$ and a literal $q$. The entailment problem $\Delta \models q$ is polynomial time decidable [16]. Thus, query $Q0$ can be solved in nondeterministic polynomial time guessing both an outlier $l \in W$ and an outlier witness $S \subseteq W$ and then asking for $(D, W_S) \models \neg S \wedge (D, W_{S,l}) \not\models \neg S$.

(Hardness) Let $\Phi = f(X)$ be a quantified boolean formula in conjunctive normal form, where $X = x_1, \ldots, x_n$ is a set of variables, and $f(X) = C_1 \wedge \ldots \wedge C_m$, with $C_j = t_{j,1} \vee \ldots \vee t_{j,u_j}$, and each $t_{j,1}, \ldots, t_{j,u_j}$ is a literal on the set $X$, for $j = 1, \ldots, m$. We associate to $\Phi$ the default theory $\Delta(\Phi) = (D(\Phi), W(\Phi))$, where $W(\Phi)$ is the set $\{l, x_1, \ldots, x_n, c_1, \ldots, c_{m+1}\}$ of letters, with $l, c_1, \ldots, c_{m+1}$ new letters distinct from those occurring in $\Phi$, and $D(\Phi) = D_1 \cup D_2 \cup D_3 \cup D_4$:

$$D_1 = \left\{\delta_{1,1,i} = \frac{x_i : \neg\overline{x}_i}{\neg\overline{x}_i}, \delta_{1,2,i} = \frac{\chi : x_i}{x_i}, \delta_{1,3,i} = \frac{: \neg x_i}{\neg x_i}, \delta_{1,4,i} = \frac{: \overline{x}_i}{\overline{x}_i} \mid i = 1, \ldots, n\right\}$$

$$D_2 = \left\{\frac{\ell(t_{j,k}) : \neg c_j}{\neg c_j} \mid j = 1, \ldots, m;\ k = 1, \ldots, u_j\right\}$$

$$D_3 = \left\{\delta_3 = \frac{l : \neg c_{m+1}}{\neg c_{m+1}}\right\}$$

$$D_4 = \left\{\delta_{4,j} = \frac{c_j : \chi}{\chi}, \delta'_{4,j} = \frac{\chi : c_j}{c_j} \mid j = 1, \ldots, m+1\right\}$$

where also $\overline{x}_1, \ldots, \overline{x}_n$ and $\chi$ are new letters distinct from those occurring in $\Phi$, and $\ell(x_i) = x_i$ and $\ell(\neg x_i) = \overline{x}_i$, for $i = 1, \ldots, n$. Clearly, $W(\Phi)$ is consistent and $\Delta(\Phi)$ can be built in polynomial time. Now we show that $\Phi$ is satisfiable iff there exists an outlier in $\Delta(\Phi)$.

Let $S$ be a subset of $\{x_1, \ldots, x_n\}$, in the rest of the proof we denote by $\ell(S)$ the set $\{\ell(s) \mid s \in S\}$.

($\Rightarrow$) Suppose that $\Phi$ is satisfiable. Then there exists a truth assignment $T_X$ on the set $X$ of variables such that $T_X$ satisfies $f(X)$. Let $S = \{s \in X \mid T_X(s) = \textbf{false}\} \cup \{c_1, \ldots, c_{m+1}\}$. Now we show that $S$ is an outlier witness for $l$ in $\Delta(\Phi)$. Consider a generic extension $E$ of $\Delta' = (D(\Phi), W(\Phi)_S)$. Clearly $E \supseteq \ell(Lit(T_X))$, as rules $\delta_{1,4,i}$, $1 \leq i \leq n$, add to $E$ the letters $\ell(\neg(X \cap S))$, while $\ell(X \setminus S) \subseteq W(\Phi)_S$. Furthermore, as $T_X$ satisfies $\Phi$, then $\neg c_j \in E$, for $j = 1, \ldots, m$. We note that the rules $\delta_{1,3,i}$, $1 \leq i \leq n$, add to $E$ the negation of the variables in $X \cap S$, while rule $\delta_3$ adds the literal $\neg c_{m+1}$ to $E$. Thus $\Delta' \models \neg S$. To conclude the proof, it is easy to verify that $(D(\Phi), W(\Phi)_{S,l}) \not\models \neg S$. Hence $l$ is an outlier in $\Delta(\Phi)$.

($\Leftarrow$) Let $S \subseteq W(\Phi)$ be an outlier witness for a literal $o \in W(\Phi)$ in $\Delta(\Phi)$. Now we show that $\{c_1, \ldots, c_{m+1}\} \subseteq S \subseteq W(\Phi) \setminus \{l\}$. First, $l \notin S$ as $\neg l$ does not appear in the conclusion of any rule of $D(\Phi)$. We note that, if the letter $\chi$ belongs to $E$, then there exists an extension $E$ of $(D(\Phi), W(\Phi)_S)$ such that $E \supseteq \{x_1, \ldots, x_n, c_1, \ldots, c_{m+1}\}$, thus $S$ must be empty, a contradiction. But, $\chi \in E$ iff $S \not\supseteq \{c_1, \ldots, c_{m+1}\}$.

Let $\sigma(S)$ denote the set of literals $(X \setminus S) \cup \neg(X \cap S)$. Now we show that $\mathcal{T}_{\sigma(S)}$ satisfies $\Phi$. As $\Delta' = (D(\Phi), W(\Phi)_S) \models \neg S$, then it is the case that, for each extension $E$ of $\Delta'$, $\neg c_1 \wedge \ldots \wedge \neg c_m \in E$. Among these extensions, there is at least one extension $E'$, with associated set of generating defaults $D_{E'}$, such that

- $(\forall s \in \ell(\sigma(S)))(s \in E')$ as $E' \supseteq W(\Phi)_S \supseteq (X \setminus S)$ and $D_{E'} \supseteq \{\delta_{1,4,k} \mid x_k \in (X \cap S)\}$
- $(\forall s \in \ell(\neg\sigma(S)))(s \notin E')$ as $D_{E'} \supseteq \{\delta_{1,1,k} \mid x_k \in (X \setminus S)\}$ and $x_k \in (X \cap S)$ implies that $x_k \notin E'$ (remember that $\chi \notin E'$)

Thus, in order to be $\neg c_1 \wedge \ldots \wedge \neg c_m \in E'$ it is the case that $\mathcal{T}_{\sigma(S)}$ satisfies $\Phi$.

As for the value of $o$, we note that $S$ is always an outlier witness for $o = l$ in $\Delta(\Phi)$. Indeed, consider the theory $\Delta'' = (D(\Phi), W(\Phi)_{S,l})$. It is easy to verify that $\Delta'' \not\models \neg c_{m+1}$, thus $\Delta'' \not\models \neg S$. □

**Theorem 3.5:** $Q0$ restricted to propositional DNU default theories is NP-complete under polynomial time transformations.
**Proof of Theorem 3.5:** (Membership) Polynomial time decidability of the entailment problem for DNU propositional default theories is stated in [32]. The rest of the membership part is analogous to that of Theorem 3.4. (Hardness) This part is analogous to that of Theorem 3.4. □

**Query Q1**

**Theorem 3.6:** $Q1$ on general propositional default theories is $\Sigma_3^P$-complete under polynomial time transformations.
**Proof of Theorem 3.6:** (Membership) The proof is analogous to that used in Theorem 3.1. (Hardness) The reduction is the same as that of Theorem 3.1. Clearly, $\Phi$ is valid iff $l$ is an outlier for $\Delta(\Phi)$. □

**Theorem 3.7:** $Q1$ restricted to DF propositional default theories is $\Sigma_2^P$-complete under polynomial time transformations.
**Proof of Theorem 3.7:** This result follows immediately from Theorem 3.8. □

**Theorem 3.8:** $Q1$ restricted to NMU propositional default theories is $\Sigma_2^P$-complete under polynomial time transformations.
**Proof of Theorem 3.8:** The proof is analogous to that used in Theorem 3.3. □

**Theorem 3.9:** $Q1$ restricted to NU and DNU propositional default theories is NP-complete under polynomial time transformations.
**Proof of Theorem 3.9:** The proof is analogous to that of Theorems 3.4 and 3.5 respectively. □

**Query Q2**

**Theorem 3.10:** $Q2$ on general propositional default theories is $D_2^P$-complete under polynomial time transformations.

**Proof of Theorem 3.10:** (Membership) Given a theory $\Delta = (D, W)$ and a subset $S = \{s_1, \ldots, s_n\} \subseteq W$, we must show that $(D, W_S) \models \neg s_1 \wedge \ldots \wedge \neg s_n$ (query $q'$) and there exists a literal $l \in W$ such that $(D, W_{S,l}) \not\models \neg s_1 \wedge \ldots \wedge \neg s_n$ (query $q''$). Solving $q'$ is in $\Pi_2^P$. As for query $q''$, it can be decided by a polynomial time nondeterministic Turing machine, with an oracle in NP, that (a) guesses both the literal $l \in W$ and the set $D_E \subseteq D$ of generating defaults of an extension $E$ of $(D, W_{S,l})$ together with an order of these defaults, (b) checks the necessary and sufficient conditions that $D_E$ must satisfy to be a set of generating defaults for $E$ (see [31] or Section 2.1 for a detailed description of these conditions), by multiple calls to the oracle, and (c) verifies that $\neg s_1 \wedge \ldots \wedge \neg s_n \notin E$, by other calls to the oracle. The total number of calls to the oracle is polynomially bounded. Thus, $Q2$ is the conjunction of two independent problems from $\Pi_2^P$ ($q'$) and $\Sigma_2^P$ ($q''$), i.e. it is in $D_2^P$.

(Hardness) Let $\Delta_1 = (D_1, W_1)$ and $\Delta_2 = (D_2, W_2)$ two consistent propositional default theories, let $s_1, s_2$ be two letters, and let $q$ be the query $\Delta_1 \models s_1 \wedge \Delta_2 \not\models s_2$. W.l.o.g, we can assume that $\Delta_1$ and $\Delta_2$ contain different letters, that the letter $s_1$ occurs in $D_1$ but not in $W_1$ (and, from the previous condition, not in $\Delta_2$), and that the letter $s_2$ occurs in $D_2$ but not in $W_2$ (and hence not in $\Delta_1$). We associate with $q$ the default theory $\Delta(q) = (D(q), W(q))$ defined as follows. Let $D_1 = \{\frac{\alpha_i:\beta_i}{\gamma_i} \mid i = 1, \ldots, n\}$ and let $L_1 = \{\ell_1, \ldots, \ell_m\} \subseteq W_1$ be all the literals belonging to $W_1$, then $D(q) = \{\frac{s_2 \wedge \alpha_i:\beta_i}{\gamma_i} \mid i = 1, \ldots, n\} \cup \{\delta_j = \frac{:\neg\ell_i \wedge \neg\mu \wedge \nu}{\nu} \mid j = 1, \ldots, m\} \cup \{\delta_0 = \frac{\nu:\mu}{\mu}\} \cup D_2$, and $W(q) = W_1 \cup W_2 \cup \{\neg s_1, s_2\}$, where $\nu$ and $\mu$ are new letters distinct from those occurring in $\Delta_1$ and $\Delta_2$, and from $s_1$ and $s_2$. Now we show that $q$ is true iff $\{\neg s_1\}$ is a witness for any outlier in $\Delta(q)$. We note that $q$ is the conjunction of a $\Pi_2^P$-hard and a $\Sigma_2^P$-hard problem, thus this will prove $D_2^P$-hardness.

($\Rightarrow$) Suppose that $q$ is true. Now we show that $\{\neg s_1\}$ is an outlier witness for $s_2$ in $\Delta(q)$. Consider the theory $\Delta' = (D(q), W(q)_{\{\neg s_1\}})$. From $\Delta_1 \models s_1$ and $s_2 \in W(q)_{\{\neg s_1\}}$, we can conclude that $\Delta' \models s_1$. Consider now the theory $\Delta'' = (D(q), W(q)_{\{\neg s_1\}, s_2})$. As $\Delta_2 \not\models s_2$, then $s_2$ cannot belong to any extension $E$ of $\Delta''$, and its associated set $D_E$ of generating defaults does not contain any rule coming from $D(q) \setminus D_2$. We also note that $\Delta''$ is consistent, as both $\Delta_1$ and $\Delta_2$ are consistent. Thus we can conclude that $\Delta'' \not\models s_1$. Hence, $\{\neg s_1\}$ is an outlier witness for $s_2$ in $\Delta(q)$.

($\Leftarrow$) Suppose that $\{\neg s_1\}$ is a witness for any outlier $o$ in $\Delta(q)$. We denote by $\Delta'$ and $\Delta''$ the theories $(D(q), W(q)_{\{\neg s_1\}})$ and $(D(q), W(q)_{\{\neg s_1\}, o})$ respectively. First, we note that $\Delta' \models s_1$. As $s_1$ occurs only in the rules of $D(q)$ coming from $D_1$, and the rules in $D_2$ have no letter in common with the rules in $D(q) \setminus D_2$, except for $s_2$, and $s_2 \in W(q)_{\{\neg s_1\}}$, then it is the case that $\Delta_1 \models s_1$. Now we show that $o$ is equal to $s_2$. In order to $\Delta'' \not\models s_1$, from what above stated, then $o$ must be either $s_2$ or a literal in $L_1$. Suppose that $o = \ell_k$ ($k \in \{1, \ldots, m\}$), then the rules $\delta_k$ and $\delta_0$ together make the theory $\Delta''$ incoherent, and $\Delta'' \models s_1$. Thus, $o$ is equal to $s_2$. Clearly, it must be also the case that $\Delta'' \not\models s_2$, i.e. that

$\Delta_2 \not\models s_2$. This proves that the query $q$ is true. $\square$

**Theorem 3.11:** $Q2$ restricted to DF propositional default theories is $D^P$-complete under polynomial time transformations.
**Proof of Theorem 3.11:** This result follows immediately from Theorem 3.12. $\square$

**Theorem 3.12:** $Q2$ restricted to NMU propositional default theories is $D^P$-complete under polynomial time transformations.
**Proof of Theorem 3.12:** (Membership) Given an NMU default theory $\Delta = (D, W)$ and a subset $S = \{s_1, \ldots, s_n\}$ of $W$, we must show that $(D, W_S) \models \neg s_1 \wedge \ldots \wedge \neg s_n$ (query $q'$) and that there exists a literal $l \in W$ such that $(D, W_{S,l}) \not\models \neg s_1 \wedge \ldots \wedge \neg s_n$ (query $q''$). From Lemma 1 it follows immediately that query $q'$ is in co-NP. As for query $q''$, it can be decided by a polynomial time nondeterministic Turing machine that (a) guesses both the literal $l$ and the set $D_E \subseteq D$ of generating defaults of an extension $E$ of $(D, W_{S,l})$ together with an order of these defaults, (b) checks the necessary and sufficient conditions that $D_E$ must satisfy to be a set of generating defaults for a disjunction free theory $E$ (see [16] or Section 2.1 for a detailed description of these conditions), and (c) verifies that $\neg s_1 \wedge \ldots \wedge \neg s_n \notin E$, by checking that there exists $i$, $1 \leq i \leq n$, such that $\neg s_i$ is not the conclusion of any default in $D_E$. Both points (b) and (c) can be performed in deterministic polynomial time, thus $Q2$ restricted to normal mixed unary theories is the conjunction of two independent problems from co-NP ($q'$) and NP ($q''$), i.e. it is in $D^P$.

(Hardness) Let $\Delta_1 = (D_1, W_1)$ and $\Delta_2 = (D_2, W_2)$ be two normal mixed unary default theories such that both $W_1$ and $W_2$ are consistent, let $s_1, s_2$ be two letters, and let $q$ be the query $\Delta_1 \models s_1 \wedge \Delta_2 \not\models s_2$. W.l.o.g. we can assume that $\Delta_1$ and $\Delta_2$ contain different letters, that the letter $s_1$ occurs in $D_1$ but not in $W_1$ (and, from the previous condition, not in $\Delta_2$), and the letter $s_2$ occurs in $D_2$ but not in $W_2$ (and hence not in $\Delta_1$). We associate with $q$ the default theory $\Delta(q) = (D(q), W(q))$ defined as follows.

Let $W_1 = \{\ell_1, \ldots, \ell_m\}$, let $D'$ be $\{\frac{\alpha:\neg\ell}{\neg\ell} \in D_1 \mid \ell \in W_1\}$ where $\alpha$ is empty or denotes an arbitrary literal, and let $D''$ be $\{\delta_\ell = \frac{:\ell}{\ell} \in D_1 \mid \neg\ell \notin W_1\}$, then $D(q) = \{\delta_1 = \frac{s_2:\ell_1}{\ell_1}, \ldots, \delta_m = \frac{s_2:\ell_m}{\ell_m}\} \cup \{\delta'_\ell = \frac{s_2:\ell}{\ell} \mid \delta_\ell \in D''\} \cup (D_1 \setminus (D' \cup D'')) \cup D_2$, and $W(q) = W_2 \cup \{\neg s_1, s_2\}$. Now we show that $q$ is true iff $\{\neg s_1\}$ is a witness for any outlier in $\Delta(q)$. We note that $q$ is the conjunction of a NP-hard and a co-NP-hard problem, thus this will prove $D^P$-hardness.

We note that $\Delta(q)$ is consistent, as $W_2$ is consistent. Furthermore, we note that the rules in the set $D_1 \setminus (D' \cup D'')$ all have non empty prerequisite. Indeed, let $\frac{:\ell}{\ell}$ a prerequisite free rule of $D_1$, then either $\neg\ell \in W_1$, and in the case this rule belongs to $D'$, or $\neg\ell \notin W_1$, and in this case the rule belongs to $D''$. In the following, we will denote by $\Delta'$ the theory $(D(q), W(q)_{\{\neg s_1\}})$.

**Claim 9** *For each literal $l$ occurring in $\Delta_1$, $\Delta_1 \models l$ iff $\Delta' \models l$.*

**Proof of Claim 9:** We start considering the case $l \in W_1$. Let $i$, $1 \leq i \leq m$, be such that $l = \ell_i$. First, we note that $\neg l$ cannot belong to any extension $E'$ of $\Delta'$, as every rule of the form $\frac{\alpha:\neg l}{\neg l}$ coming from $D_1$ (where $\alpha$ is possibly empty), is not present in $D(q)$. Thus, as the rule $\delta_i$ of $D(q)$ has $l$ as its justification and conclusion and $s_2$ as its prerequisite, and $s_2 \in W(q)_{\{\neg s_1\}}$, then $l$ belongs to every extension of $\Delta'$.

Now we consider a literal $l$ occurring in $D_1$ but not in $W_1$. The claim statement follows immediately by noting that $\Delta' \models \ell_i$, $1 \leq i \leq m$, as above stated, and that $D(q)$ contains all the defaults in $D_1$ except those in the sets $D'$ and $D''$ (the latter rules are replaced by slightly modified rules). Indeed, the rules in $D'$ do not belong to the set of generating defaults of any extension of $\Delta_1$, while each rule $\delta_\ell$ in $D''$ is replaced by a new rule $\frac{s_2:\ell}{\ell}$ in $\Delta'$, but $s_2 \in W(q)_{\{\neg s_1\}}$ implies that these rules can be considered, loosely speaking, equivalent to those in $D''$. □

Now we can resume with the main proof.

($\Rightarrow$) Suppose that $q$ is true. Now we show that $\{\neg s_1\}$ is an outlier witness for $s_2$ in $\Delta(q)$. From $\Delta_1 \models s_1$ and Claim 9, we can conclude that $\Delta' \models s_1$. Consider now the theory $\Delta'' = (D(q), W(q)_{\{\neg s_1\}, s_2})$. As $\Delta_2 \not\models s_2$ then $\Delta'' \not\models s_1$. Indeed, $D_1 \setminus (D' \cup D'')$ contains only default rules with non empty prerequisite, while the rules $\delta_i$ ($1 \leq i \leq m$) and $\delta'_\ell$ ($\delta_\ell \in D''$) all have $s_2$ as their prerequisite. Furthermore, we note that $\Delta''$ is consistent, as $W_2$ is consistent. Thus we can conclude that $\Delta'' \not\models s_1$. Hence, $\{\neg s_1\}$ is a witness for $s_2$ in $\Delta(q)$.

($\Leftarrow$) Suppose that $\{\neg s_1\}$ is a witness for any outlier $o$ in $\Delta(q)$. From $\Delta' \models s_1$ and Claim 9, we can conclude that $\Delta_1 \models s_1$. Let $\Delta''$ be the theory $(D(q), W(q)_{\{\neg s_1\}, o})$. In order to be $\Delta'' \not\models s_1$, then $o$ must be $s_2$. Indeed, until $s_2 \in W(q)_{\{\neg s_1\}, o}$, then $\Delta'' \models s_1$ by following the same line of reasoning of Claim 9. □

**Theorem 3.13:** $Q2$ restricted to NU and DNU propositional default theories is in $P$.
**Proof of Theorem 3.13:** The entailment problem for both NU and DNU propositional default theories can be decided in polynomial time [16, 32]. Thus $Q2$, on the input $\Delta = (D, W)$ and $S$, can be solved by a deterministic polynomial time Turing machine verifying that there exists a literal $l \in W_S$ such that $(D, W_S) \models \neg S \wedge (D, W_{S,l}) \not\models \neg S$. To conclude the proof we note that the number of literals in $W_S$ is linearly related to the size of the theory. □

**Query Q3**

**Theorem 3.14:** $Q3$ on general propositional default theories is $D_2^P$-complete under polynomial time transformations.
**Proof of Theorem 3.14:** Both hardness and membership proofs are analogous to that of Theorem 3.10. □

**Theorem 3.15:** $Q3$ restricted to DF propositional default theories is $D^P$-complete under polynomial time transformations.
**Proof of Theorem 3.15:** This result follows immediately from Theorem 3.16. □

**Theorem 3.16:** $Q3$ restricted to NMU propositional propositional default theories is $D^P$-complete under polynomial time transformations.
**Proof of Theorem 3.16:** Both membership and hardness proofs are analogous to that of Theorem 3.12. □

**Theorem 3.17:** $Q3$ restricted to NU and DNU propositional default theories is in P.
**Proof of Theorem 3.17:** The proof is analogous to that of Theorem 3.13. □

# 7 Proofs of Section 4

**Definition 7.1** Let $\Delta = (D, W)$ be an NMU default theory, let $l$ be a literal, and $E$ be a set clauses. A *proof* of $l$ w.r.t. $\Delta$ and $E$ is a either $l$ by itself, if $l \in W$, or a sequence of defaults $\delta_1, ..., \delta_n$, such that the following holds:

1. $l$ is the consequence of $\delta_n$,
2. $\neg l \notin E$, and
3. for each $\delta_i$, $1 \leq i \leq n$, either $\delta_i$ is prerequisite-free, or $\delta_1, ..., \delta_{i-1}$ is a proof of the prerequisite of $\delta_i$ w.r.t. $\Delta$ and $E$.

**Lemma 7.2** [6] *Let $\Delta = (D, W)$ be an NMU default theory, let $l$ be a literal and let $E$ be an extension of $\Delta$. Then $l$ is in $E$ iff there is a proof of $l$ in $E$.*

**Definition 7.3** We say that a set of literals $E$ *satisfies* an NMU default $\delta = \frac{y:x}{x}$ iff at least one of the following three conditions hold:

1. $y \notin E$,
2. $\neg x \in E$,
3. $x \in E$.

**Theorem 7.4** [6] *A logically closed set of clauses $E$ is an extension of a DF consistent default theory $(D, W)$ iff the following holds:*

1. *$W \subseteq E$,*
2. *$E$ satisfies every rule in $D$,*
3. *every literal in $E$ has a proof w.r.t $(D, W)$ and $E$.*

Let $(D, W)$ be a default theory. The process of *crossing out defaults* from a sequence of defaults from $D$ according to a literal $l$ removes from the sequence defaults that might become inapplicable in case $l$ is added to $W$. The formal definition follows.

**Definition 7.5** Let $(D, W)$ be an NMU default theory. Given a literal $l$ and a sequence of defaults $\delta_1, ..., \delta_n \subseteq D$, the process of *crossing out defaults* from this sequence according to a literal $l$ goes as follows:

1. for each literal $x$ such that there is a path from $l$ to $x$ in the dependency graph of $(D, W)$ (this includes $l$ itself), if $\neg x$ is a consequence of some default $\delta_i$, then cross out $\delta_i$;

2. repeat the following until no default is crossed out:

   for each $j$, $1 \leq j \leq n$, if $\delta_j$ was not crosses out, then if the prerequisite $y$ of $\delta_j$ does not belong to $W$ and $y$ is not a consequence of any default $\delta_h$, with $h < j$, then cross out $\delta_j$.

**Lemma 7.6** *Let $\Delta = (D, W)$ be an NMU default theory and let $l$ be a literal such that $l \notin W$ and $W' = W \bigcup \{l\}$ is consistent. Assume further that $E_1, ..., E_n$ is a list of all the extensions of $\Delta$. Then, all the extensions of the default theory $\Delta' = (D, W')$ can be computed using the following* incremental procedure*:*

1. *$\mathcal{E} = \emptyset$; ($\mathcal{E}$ will accumulate all extensions)*

2. *For each $E_i$ in $E_1, ..., E_n$ do*

   *(a) Let $\sigma = \delta_{i_1}, ..., \delta_{i_k}$ be a sequence of generating defaults of $E_i$ as described in Section 2.1.*

   *(b) Cross out defaults from $\sigma$ according to the literal $l$, as described in Definition 7.5. Let $\sigma_l$ be the set of all defaults that were* **not** *crossed out from $\sigma$.*

   *(c) Let $E'_i$ be the extension of the default theory $(\sigma_l, W)$ (this theory has exactly one extension, and $\sigma_l$ is the set of its generating default).*

   *(d) $\mathcal{E} = \mathcal{E} \bigcup \mathcal{E}_i$, where $\mathcal{E}_i$ is the set of all the extensions of the default theory $(D - \sigma_l, liter(E'_i) \bigcup l)$.*

**Proof:** First, we have to show that every set in $\mathcal{E}$ is indeed an extension of $\Delta'$. We will use Theorem 7.4. Let $E \in \mathcal{E}$. Clearly, $W' \subseteq E$. Next we show that $E$ satisfies every default in $D$. Let $\delta = \frac{y:x}{x} \in D$ ($y$ possibly empty). If $\delta \in D - \sigma_l$ then it is clearly satisfied by $E$. If $\delta \in \sigma_l$ then $x \in E'_i$, and so $x \in E$ so $E$ satisfies $\delta$. It is left to show that each literal $x \in E$ has a proof with respect to $E$ and $\Delta$‘. Let $x \in E$, let $\Delta'' = (D - \sigma_l, liter(E'_i) \bigcup l)$. Since $E$ is an extension of $\Delta''$, $x$ has a minimal proof w.r.t $E$ and $\Delta''$ (Lemma 7.2). the proof goes by induction on $i$, the length of that proof:

**case** $i = 1$ then there are two possibilities

1. $x \in liter(E'_i) \bigcup l$. Then either $x = l$, and clearly has a proof, or $x \in liter(E'_i)$ and then the proof of $x$ w.r.t $E'_i$ and $(\sigma'_l, W)$ is a proof of $x$.
2. There is a default $\frac{:x}{x} \in D - \sigma_l$. Since $E$ is consistent, $x \in E$ and $\frac{:x}{x} \in D$, $\frac{:x}{x}$ is a proof of $x$ w.r.t $E$ and $\Delta'$.

**case** $i > 1$ Let $\delta = \frac{y:x}{x}$ be the last default in a minimal proof of $x$ w.r.t $E$ and $\Delta''$. By the induction hypothesis, $y$ has a proof w.r.t. $E$ and $\Delta'$. Since $E$ is consistent, $x \in E$ and $\frac{y:x}{x} \in D$, the concatenation of the proof of $y$ with the default $\frac{y:x}{x}$ yields a proof of $x$ w.r.t $E$ and $\Delta'$.

Second, we have to show that every extension of $\Delta'$ is indeed generated by the incremental procedure. Let $E$ be an extension of $\Delta'$. Then $E$ has a sequence $\sigma = \delta_1, ..., \delta_n$ of generating defaults. We will modify $\sigma$ as follows.

1. Let $i$ be the minimum index such that $l$ is a prerequisite of $\delta_i$. $H = \{\delta_i\}$, delete $\delta_i$ from $\sigma$.
2. For $h = i + 1$ to $n$ do

   If the prerequisite of $\delta_h$ is not in $W$ (note that $W = W' - \{l\}$) and not a consequence of any default which is currently before $\delta_h$ then

   (a) $H = H \bigcup \{\delta_h\}$,

   (b) delete $\delta_h$ from $\sigma$.
3. Let $\sigma'_l$ be the sequence of defaults left in $\sigma$.
4. While there is a default $\delta \in H$ and a default $\delta' \in D$ such that cons($\delta'$) = cons($\delta$) and pre($\delta'$) in $W$ or pre( $\delta'$) = cons($\delta''$) for some $\delta'' \in \sigma'_l$ do:
   - $H = H - \delta$;
   - add $\delta'$ to the end of $\sigma'_l$;

Let $E$ be the extension of $(\sigma'_l, W)$. Clearly, $E' \subseteq E$.

**Claim 10** *Every consequence of a default in $\sigma'_l$ belongs to $E'$.*

**Proof:** It is easy to see that every consequence of a default in $\sigma'_l$ has a proof w.r.t $(\sigma'_l, W)$ and $E'$. Hence the claim follows by Theorem 7.4. □

**Claim 11** *$\sigma'_l$ is the set of generating defaults of $E'$.*

By Theorem 3.2 of [25], there is an extension $E''$ of $(D, W)$ such that $E' \subseteq E''$. We will show that $E$ is added to $\mathcal{E}$ at Step 2($d$) when $E_i$ of Step 2 is equel to $E''$. Let $\pi$ be the sequence of generating defaults of $E''$ picked at Step 2($a$). Let $\pi_l$ be the sequence left after crossing out defaults from $\pi$ according to $l$. Let $Ex$ be the extension of $(\pi_l, W)$.

**Claim 12** *Every consequence of a default in $\pi_l$ belongs to $Ex$.*

**Proof:** It is easy to see that every consequence of a default in $\pi_l$ has a proof w.r.t $(\pi_l, W)$ and $Ex$. Hence the claim follows by Theorem 7.4. □

**Claim 13** *$Ex$ is a subset of $E'$.*

**Proof:** If $x \in Ex \bigcap W$ then clearly $x \in E'$. Assume $x \notin W$, $x \in Ex$. By Claim 10 and Claim 12 it is enough to show that for every default $\delta$ in $\pi_l$ there is default in $\sigma_l'$ with the same consequence. Note that all the defaults of $\sigma$ appear in $\sigma_l'$, except the ones that require $l$ in order to be applicable. The proof goes by induction on $i$, where $i$ is the index of $\delta$ in the sequence $\pi_l$.

$i = 1$ Then $\delta = \frac{y:x}{x}$ with an empty $y$ or $y \in W$. Assume conversely that there is no default in $\sigma_l'$ having $x$ as a consequence. We consider two case:

**There is a default $\delta' \in \sigma$ with $x$ as a consequence:** Since $y \in W$, by the way $\sigma_l'$ was constructed from $\sigma$ $\delta = \frac{y:x}{x} \in \sigma_l'$, a contradiction.

**There is no default in $\sigma$ with $x$ as a consequence:** Since $\delta \in D$, $y \in W$ and $\sigma$ is a set of generating defaults, it must be the case that there is a default $\frac{z:\sim x}{\sim x} \in \sigma$ for some $z$ that might be empty. Since $x \in Ex$ and $Ex \subseteq E''$ and $E' \subseteq E''$ (all of them consistent extensions), and by Claim 10, $\frac{z:\sim x}{\sim x} \notin \sigma_l'$. Since $\frac{z:\sim x}{\sim x} \in \sigma$ but $\frac{z:\sim x}{\sim x} \notin \sigma_l'$, there must be a path in the dependency graph of $(D, W')$ from $l$ to $\sim x$. So there is also a path from $l$ to $\sim x$ in the dependency graph of $(D, W)$. By the way $\pi_l$ is constructed, it cannot be the case that $\delta = \frac{y:x}{x} \in \pi_l$.

**induction step** Assume $j > 1$, $\delta_j = \frac{y:x}{x}$ is in the sequence $\pi_l$. By the induction hypothesis, $y \in E'$. Assume conversely that there is no default in $\sigma_l'$ having $x$ as a consequence. We consider two case:

**There is a default $\delta' \in \sigma$ with $x$ as a consequence:** By the way $\sigma_l'$ was constructed from $\sigma$ and by Claim 11 $\delta = \frac{y:x}{x} \in \sigma_l'$, a contradiction.

**There is no default in $\sigma$ with $x$ as a consequence:** Since $\delta \in D$, $y \in E'$, $E' \subseteq E$ and $\sigma$ is a set of generating defaults of $E$, it must be the case that there is a default $\frac{z:\sim x}{\sim x} \in \sigma$ for some $z$ that might be empty. We proceed as in the case $i = 1$ to get a contradiction.

□

In order to show that $E$ is generated, it is now enough to show that $E$ is an extension of $(D - \pi_l, liter(Ex) \bigcup \{l\})$. We will use Theorem 7.4.

First, we need to show that $liter(Ex) \bigcup \{l\}$ is a subset of $E$. Clearly, $l \in E$. The rest follows from Claim 13, since $E' \subseteq E$.

Second, we need to show that every default in $D - \pi_l$ is satisfied by $E$. This is obvious because $D - \pi_l \subseteq D$ and $E$ is an extension of $(D, W')$ and hence satisfies every default from $D$.

Third, we have to show that if $x \in E$ then $x$ has a proof with respect to $(D - \pi_l, liter(Ex) \bigcup \{l\})$ and $E$. $E$ is an extension of $(D, W')$. Therefore, if $x \in E$ then $x$ has a proof with respect to $(D, W')$ (Lemma 7.2). By induction on the length of a minimal proof of $x$ with respect to $(D, W')$ we will show that it has a proof w.r.t $(D - \pi_l, liter(Ex) \bigcup \{l\})$ and $E$.

Assume $x \in W'$. Since $W' = W \bigcup \{l\}$ and $Ex$ is an extension of $(\pi_l, W)$ it must be the case that $x \in Ex \bigcup \{l\}$, so $x$ has a proof w.r.t $(D - \pi_l, liter(Ex) \bigcup \{l\})$ and $E$. Suppose, using the induction hypothesis that if $x$ has a minimal proof of length $n$ with respect to $(D, W')$ and $E$ then it has a proof w.r.t $(D - \pi_l, liter(Ex) \bigcup \{l\})$ and $E$. Assume $x$ has a proof of length $n + 1$ with respect to $(D, W')$ and $E$. Let $\delta = \frac{y:x}{x}$ be the last default in the proof. $y$ has a proof of size $\leq n$ with respect to $(D, W')$ and $E$, and so, by the induction hypothesis $y$ has a proof w.r.t. $(D - \pi_l, liter(Ex) \bigcup \{l\})$ and $E$. If $\delta \in D - \pi_l$, then clearly $x$ has a proof w.r.t. $(D - \pi_l, liter(Ex) \bigcup \{l\})$ and $E$ (the proof is the proof of $y$ concatenated with $\delta$). If $\delta \in \pi_l$, then by Claim 12 $x \in Ex$, and so $x$ has a proof w.r.t $(D - \pi_l, liter(Ex) \bigcup \{l\})$. $\square$

**Theorem 4.4:** Let $(D, W)$ be a consistent acyclic NMU default and let $l$ be a literal in $W$. Then any minimal outlier witness set for $l$ in $(D, W)$ is of size at most 1.

**Proof of Theorem 4.4:** Let $S$ be an outlier witness set for $l$ in $(D, W)$ such that $|S| > 1$. By definition, the following must be true:

1. $(D, W_S) \models \neg S$, and
2. $(D, W_{S,l}) \not\models \neg S$.

Let $v$ be a smallest literal in $S$, by the partial ordering induced by the atomic dependency graph of $(D, W)$, which is acyclic. We claim that $\{v\}$ is an outlier witness set for $l$. We will show that the following holds:

1. $(D, W_v) \models \neg v$, and
2. $(D, W_{v,l}) \not\models \neg v$.

Item 2 clearly holds. Next we show Item 1. Let $S = x_1, ..., x_n, v$. We know that $(D, W_S) \models \neg S$. We will use the incremental procedure of Lemma 7 $n$ times, each time taking $l$ to be one of $S - \{v\}$. We will show by induction on $i$ $(1 \leq i \leq n)$that after each step $(D, W_{x_1,...,x_i}) \models \neg v$. By Lemma 7, we can use the incremental procedure in order to compute all the extensions of $(D, W_S \bigcup \{x_1\})$ out of all the extensions of $(D, W_S)$. Since $(D, W_S) \models \neg v$, for any extension $E$ of $(D, W_S)$, a sequence $\sigma$ of generating defaults of $E$ must contain a proof of $\neg v$. Since $(D, W)$ is acyclic, and since $v$ is a smallest literal in $S$ by the partial ordering induced by the atomic dependency graph of $(D, W)$, there is a proof of $v$ left also after crossing out defaults from $\sigma$ according to the literal $x_1$. Hence every extension of $(D, W_S \bigcup \{x_1\})$ must have $\neg v$ in it.

Assume by induction that after applying the incremental procedure $n - 1$ times $(D, W_{\{x_1,...,x_{n-1}\}}) \models \neg v$. By similar arguments we can show that $(D, W_{\{x_1,...,x_n\}}) \models \neg v$. $\square$