R A P P O R T   D E   R E C H E R C H E

L R I

EXPLAINING QUERY ANSWERS UNDER INCONSISTENCY-TOLERANT SEMANTICS OVER DESCRIPTION LOGIC KNOWLEDGE BASES

BIENVENU M / BOURGAUX C / GOASDOUE F

Unité Mixte de Recherche 8623
CNRS-Université Paris Sud-LRI

05/2015

Rapport de Recherche N° 1580

CNRS – Université de Paris Sud
Centre d'Orsay
LABORATOIRE DE RECHERCHE EN INFORMATIQUE
Bâtiment 650
91405 ORSAY Cedex (France)

Explaining Query Answers under Inconsistency-Tolerant Semantics over Description Logic Knowledge Bases

Meghyn Bienvenu[1], Camille Bourgaux[1], and François Goasdoué[2]

[1]LRI, CNRS & Université Paris-Sud, Orsay, France
[2]IRISA, Université de Rennes 1, Lannion, France

Abstract. Several inconsistency-tolerant semantics have been introduced for querying inconsistent description logic knowledge bases. This paper addresses the problem of explaining why a tuple is a (non-)answer to a query under such semantics. We define explanations for positive and negative answers under the brave, AR and IAR semantics. We then study the computational properties of explanations in the lightweight description logic DL-Lite$_\mathcal{R}$. For each type of explanation, we analyze the data complexity of recognizing (preferred) explanations and deciding if a given assertion is relevant or necessary. We establish tight connections between intractable explanation problems and variants of propositional satisfiability (SAT), enabling us to compute explanations by exploiting solvers for Boolean satisfaction and optimization problems. We implemented the proposed algorithms and empirically showed that the method works.

1 Introduction

Description logic (DL) knowledge bases (KBs) consist of a TBox (ontology) that provides conceptual knowledge about the application domain and an ABox (dataset) that contains facts about particular entities [3]. The problem of querying such KBs using database-style queries (in particular, conjunctive queries) has been a major focus of recent DL research. Since scalability is a key concern, much of the work has focused on lightweight DLs for which query answering can be performed in polynomial time w.r.t. the size of the ABox. The DL-Lite family of lightweight DLs [13] is especially popular due to the fact that query answering can be reduced, via query rewriting, to the problem of standard database query evaluation.

Since the TBox is usually developed by experts and subject to extensive debugging, it is often reasonable to assume that its contents are correct. By contrast, the ABox is typically substantially larger and subject to frequent modifications, making errors almost inevitable. As such errors may render the KB inconsistent, several inconsistency-tolerant semantics have been introduced in order to provide meaningful answers to queries posed over inconsistent KBs.

The most well-known is the *AR semantics* [22], inspired by work on consistent query answering in databases (cf. [7] for a survey). Query answering under AR semantics amounts to considering those answers (w.r.t. standard semantics) that can be obtained from every *repair*, i.e. inclusion-maximal subset of the ABox that is consistent with the TBox. A more cautious semantics, called *IAR semantics* [22], queries the intersection of the repairs and provides a lower bound on AR semantics. The *brave* semantics [10], which considers the answers holding in some repair, provides a natural upper bound.

The complexity of inconsistency-tolerant query answering in the presence of ontologies is now well understood (e.g. [30, 8, 24]), so attention has turned to the problem of implementing these alternative semantics. There are currently two systems for querying inconsistent DL-Lite KBs: the QuID system of [31] implements the IAR semantics, using either query rewriting or ABox cleaning, and the CQAPri system of [9] implements the AR, IAR and brave semantics, using tractable methods to obtain the answers under IAR and brave semantics and calls to a SAT solver to identify the answers holding under AR semantics.

The need to equip reasoning systems with explanation services is widely acknowledged by the DL community (see Section 6 for discussion and references), and such facilities are all the more essential when using inconsistency-tolerant semantics, as recently argued in [1, 2]. Indeed, the brave, AR, and IAR semantics allow one to classify query answers into three categories of increasing reliability, and a user may naturally wonder why a given tuple was assigned to, or excluded from, one of these categories. In this paper, we address this issue by proposing and exploring a framework for explaining query answers under these three semantics. Our contributions are as follows:

- We define explanations of positive and negative query answers under brave, AR and IAR semantics. Intuitively, such explanations pinpoint the portions of the ABox that, in combination with the TBox, suffice to obtain the considered query answer. We focus on ABox assertions since inconsistencies are assumed to stem from errors in the ABox, and because this yields a simple but non-trivial framework to explore.
- We study the main decision problems related to explanations: checking if an assertion is relevant / necessary (i.e. appears in some / all explanations), and recognizing explanations, resp. most preferred explanations according to some natural ranking criteria. We determine the data complexity of these problems for DL-Lite$_\mathcal{R}$. For both the intractable decision problems and the task of generating explanations, we establish tight connections to known SAT-based reasoning tasks, thereby enabling the use of SAT solvers.
- Finally, we present our implementation of the explanation services in a prototype system, and experiments that show that explanations for a query answer can be computed very quickly, typically a few milliseconds, and almost always in less than half a second.

Proof details are provided in the appendix.

2 Preliminaries

We briefly recall the syntax and semantics of description logics (DLs), and the inconsistency-tolerant semantics we use.

Syntax A DL *knowledge base (KB)* consists of an ABox and a TBox, both constructed from a set $\mathsf{N_C}$ of *concept names* (unary predicates), a set of $\mathsf{N_R}$ of *role names* (binary predicates), and a set $\mathsf{N_I}$ of *individuals* (constants). The *ABox* (dataset) consists of a finite number of *concept assertions* of the form $A(a)$ and *role assertions* of the form $R(a,b)$, where $A \in \mathsf{N_C}$, $R \in \mathsf{N_R}$, $a, b \in \mathsf{N_I}$. The *TBox* (ontology) consists of a set of axioms whose form depends on the DL in question.

In the logic DL-Lite$_\mathcal{R}$, TBox axioms are either *concept inclusions* $B \sqsubseteq C$ or *role inclusions* $Q \sqsubseteq S$ formed according to the following syntax (where $A \in \mathsf{N_C}$ and $R \in \mathsf{N_R}$):

$$B := A \mid \exists Q \qquad C := B \mid \neg B \qquad Q := R \mid R^- \qquad S := Q \mid \neg Q$$

Semantics An *interpretation* has the form $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set and $\cdot^\mathcal{I}$ maps each $a \in \mathsf{N_I}$ to $a^\mathcal{I} \in \Delta^\mathcal{I}$, each $A \in \mathsf{N_C}$ to $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, and each $R \in \mathsf{N_R}$ to $R^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The function $\cdot^\mathcal{I}$ is straightforwardly extended to general concepts and roles, e.g. $(R^-)^\mathcal{I} = \{(c,d) \mid (d,c) \in R^\mathcal{I}\}$ and $(\exists Q)^\mathcal{I} = \{c \mid \exists d : (c,d) \in Q^\mathcal{I}\}$. An interpretation $\mathcal{I}$ satisfies an inclusion $G \sqsubseteq H$ if $G^\mathcal{I} \subseteq H^\mathcal{I}$; it satisfies $A(a)$ (resp. $R(a,b)$) if $a^\mathcal{I} \in A^\mathcal{I}$ (resp. $(a^\mathcal{I}, b^\mathcal{I}) \in R^\mathcal{I}$). An interpretation $\mathcal{I}$ is a *model* of $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if $\mathcal{I}$ satisfies all axioms in $\mathcal{T}$ and assertions in $\mathcal{A}$. A KB $\mathcal{K}$ is *consistent* if it has a model; otherwise it is inconsistent, denoted $\mathcal{K} \models \bot$. An ABox $\mathcal{A}$ is $\mathcal{T}$*-consistent* if the KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is consistent.

Example 1. As a running example, we consider a simple KB $\mathcal{K}_\mathsf{ex} = (\mathcal{T}_\mathsf{ex}, \mathcal{A}_\mathsf{ex})$ about the university domain that contains concepts for postdoctoral researchers ($\mathsf{Postdoc}$), professors (Pr) of two levels of seniority ($\mathsf{APr}, \mathsf{FPr}$), and PhD holders ($\mathsf{PhD}$), and roles to link advisors to their students (Adv) and instructors to their courses (Teach). The ABox $\mathcal{A}_\mathsf{ex}$ provides information about an individual a:

$$\begin{aligned}\mathcal{T}_\mathsf{ex} = &\{\mathsf{Postdoc} \sqsubseteq \mathsf{PhD}, \mathsf{Pr} \sqsubseteq \mathsf{PhD}, \mathsf{Postdoc} \sqsubseteq \neg\mathsf{Pr}, \\ &\mathsf{FPr} \sqsubseteq \mathsf{Pr}, \mathsf{APr} \sqsubseteq \mathsf{Pr}, \mathsf{APr} \sqsubseteq \neg\mathsf{FPr}, \exists\mathsf{Adv} \sqsubseteq \mathsf{Pr}\} \\ \mathcal{A}_\mathsf{ex} = &\{\mathsf{Postdoc}(a), \mathsf{FPr}(a), \mathsf{APr}(a), \mathsf{Adv}(a,b), \\ &\mathsf{Teach}(a,c_1), \mathsf{Teach}(a,c_2), \mathsf{Teach}(a,c_3)\}\end{aligned}$$

Observe that $\mathcal{A}_\mathsf{ex}$ is $\mathcal{T}_\mathsf{ex}$-inconsistent.

Queries We focus on *conjunctive queries* (CQs) which take the form $\exists \boldsymbol{y}\, \psi$, where ψ is a conjunction of atoms of the forms $A(t)$ or $R(t,t')$, t, t' are variables or individuals, and $\boldsymbol{y}$ is a tuple of variables from ψ. When we use the generic term *query*, we mean a CQ. Given a CQ q with free variables $x_1, \dots, x_k$ and a tuple of individuals $\mathbf{a} = (a_1, \dots, a_k)$, we use $q(\mathbf{a})$ to denote the first-order sentence resulting from replacing each x_i by a_i. A tuple $\mathbf{a}$ is a *certain answer* to q over $\mathcal{K}$, written $\mathcal{K} \models q(\mathbf{a})$, iff $q(\mathbf{a})$ holds in every model of $\mathcal{K}$.

Causes and conflicts A *cause* for $q(\mathbf{a})$ w.r.t. KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is a minimal $\mathcal{T}$-consistent subset $\mathcal{C} \subseteq \mathcal{A}$ such that $\mathcal{T}, \mathcal{C} \models q(\mathbf{a})$. We use $\mathsf{causes}(q(\mathbf{a}), \mathcal{K})$ to refer to the set of causes for $q(\mathbf{a})$. A *conflict* for $\mathcal{K}$ is a minimal $\mathcal{T}$-inconsistent subset of $\mathcal{A}$, and $\mathsf{confl}(\mathcal{K})$ denotes the set of conflicts for $\mathcal{K}$.

When $\mathcal{K}$ is a DL-Lite$_\mathcal{R}$ KB, every conflict for $\mathcal{K}$ has at most two assertions. We can thus define the set of *conflicts of a set of assertions* $\mathcal{C} \subseteq \mathcal{A}$ as follows:

$$\mathsf{confl}(\mathcal{C}, \mathcal{K}) = \{\beta \mid \exists \alpha \in \mathcal{C}, \{\alpha, \beta\} \in \mathsf{confl}(\mathcal{K})\}.$$

Inconsistency-tolerant semantics A *repair* of $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is an inclusion-maximal subset of $\mathcal{A}$ that is $\mathcal{T}$-consistent. We consider three previously studied inconsistency-tolerant semantics based upon repairs. Under *AR semantics*, a tuple $\mathbf{a}$ is an answer to q over $\mathcal{K}$, written $\mathcal{K} \models_{\text{AR}} q(\mathbf{a})$, just in the case that $\mathcal{T}, \mathcal{R} \models q(\mathbf{a})$ for *every repair* $\mathcal{R}$ of $\mathcal{K}$ (equivalently: every repair contains some cause of $q(\mathbf{a})$). If there exists *some repair* $\mathcal{R}$ such that $\mathcal{T}, \mathcal{R} \models q(\mathbf{a})$ (equivalently: $\mathsf{causes}(q(\mathbf{a}), \mathcal{K}) \neq \emptyset$), then $\mathbf{a}$ is an answer to q under *brave semantics*, written $\mathcal{K} \models_{\text{brave}} q(\mathbf{a})$. For *IAR semantics*, we have $\mathcal{K} \models_{\text{IAR}} q(\mathbf{a})$ iff $\mathcal{T}, \mathcal{R}_\cap \models q(\mathbf{a})$ (equivalently, $\mathcal{R}_\cap$ contains some cause for $q(\mathbf{a})$), where $\mathcal{R}_\cap$ is the *intersection of all repairs* of $\mathcal{K}$. The three semantics are related as follows:

$$\mathcal{K} \models_{\text{IAR}} q(\mathbf{a}) \quad \Rightarrow \quad \mathcal{K} \models_{\text{AR}} q(\mathbf{a}) \quad \Rightarrow \quad \mathcal{K} \models_{\text{brave}} q(\mathbf{a})$$

For $S \in \{\text{AR}, \text{brave}, \text{IAR}\}$, we call $\mathbf{a}$ a *(positive) S-answer* (resp. *negative S-answer*) if $\mathcal{K} \models_{\text{S}} q(\mathbf{a})$ (resp. $\mathcal{K} \not\models_{\text{S}} q(\mathbf{a})$).

Example 2. The example KB $\mathcal{K}_{\mathsf{ex}}$ has three repairs:

$$\mathcal{R}_1 = \mathcal{A}_{\mathsf{ex}} \setminus \{\mathsf{FPr}(a), \mathsf{APr}(a), \mathsf{Adv}(a, b)\}$$
$$\mathcal{R}_2 = \mathcal{A}_{\mathsf{ex}} \setminus \{\mathsf{Postdoc}(a), \mathsf{FPr}(a)\}$$
$$\mathcal{R}_3 = \mathcal{A}_{\mathsf{ex}} \setminus \{\mathsf{Postdoc}(a), \mathsf{APr}(a)\}$$

We consider the following example queries: $q_1 = \mathsf{Prof}(x)$, $q_2 = \exists y\, \mathsf{PhD}(x) \wedge \mathsf{Teach}(x, y)$, and $q_3 = \exists y \mathsf{Teach}(x, y)$. Evaluating these queries on $\mathcal{K}_{\mathsf{ex}}$ yields the following results:

- $\mathcal{K}_{\mathsf{ex}} \models_{\text{brave}} q_1(a)$ but $\mathcal{K}_{\mathsf{ex}} \not\models_{\text{AR}} q_1(a)$
- $\mathcal{K}_{\mathsf{ex}} \models_{\text{AR}} q_2(a)$ but $\mathcal{K}_{\mathsf{ex}} \not\models_{\text{IAR}} q_2(a)$
- $\mathcal{K}_{\mathsf{ex}} \models_{\text{IAR}} q_3(a)$

3 Explaining Query Results

The inconsistency-tolerant semantics from the preceding section allows us to identify three types of positive query answer:

$$\text{IAR-answers} \quad \subseteq \quad \text{AR-answers} \quad \subseteq \quad \text{brave-answers}$$

The goal of the present work is to help the user understand the classification of a particular tuple, e.g. why is $\mathbf{a}$ an AR-answer, and why is it not an IAR-answer? To this end, we introduce the notion of *explanation* for positive and negative query answers under brave, AR, and IAR semantics. Note that for consistent KBs, these three semantics collapse into classical semantics, so existing techniques for explaining query answers can be used instead [11, 14, 16].

Formally, the explanations we consider will take either the form of a set of ABox assertions (viewed as a conjunction) or a set of sets of assertions (interpreted as a disjunction of conjunctions). We chose to focus on ABox assertions, rather than TBox axioms, since we target scenarios in which inconsistencies are due to errors in the ABox, so understanding the link between (possibly faulty) ABox assertions and query results is especially important. Moreover, as we shall see in Sections 4 and 5, our 'ABox-centric' explanation framework already poses non-trivial computational challenges.

The simplest answers to explain are positive brave- and IAR-answers. For the former, we can use the query's causes as explanations, and for the latter, we consider the causes that do not participate in any contradictions. Note that in what follows we suppose that $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is a KB and q is a query.

Definition 1. *An explanation for* $\mathcal{K} \models_{brave} q(\mathbf{a})$ *is a cause for* $q(\mathbf{a})$ *w.r.t.* $\mathcal{K}$. *An explanation for* $\mathcal{K} \models_{IAR} q(\mathbf{a})$ *is a cause* $\mathcal{C}$ *for* $q(\mathbf{a})$ *w.r.t.* $\mathcal{K}$ *such that* $\mathcal{C} \subseteq \mathcal{R}$ *for every repair* $\mathcal{R}$ *of* $\mathcal{K}$.

Example 3. There are three explanations for $\mathcal{K}_{\mathsf{ex}} \models_{\text{brave}} q_1(a)$: $\mathsf{FPr}(a)$, $\mathsf{APr}(a)$, and $\mathsf{Adv}(a, b)$. There are twelve explanations for $\mathcal{K}_{\mathsf{ex}} \models_{\text{brave}} q_2(a)$: $\mathsf{Postdoc}(a) \wedge \mathsf{Teach}(a, c_j)$, $\mathsf{FPr}(a) \wedge \mathsf{Teach}(a, c_j)$, $\mathsf{APr}(a) \wedge \mathsf{Teach}(a, c_j)$, and $\mathsf{Adv}(a, b) \wedge \mathsf{Teach}(a, c_j)$, for each $j \in \{1, 2, 3\}$. There are three explanations for $\mathcal{K}_{\mathsf{ex}} \models_{\text{IAR}} q_3(a)$: $\mathsf{Teach}(a, c_1)$, $\mathsf{Teach}(a, c_2)$, and $\mathsf{Teach}(a, c_3)$.

To explain why a tuple is an AR-answer, it is no longer sufficient to give a single cause, since different repairs may use different causes. We will therefore define explanations as (minimal) disjunctions of causes that 'cover' all repairs.

Definition 2. *An explanation for* $\mathcal{K} \models_{AR} q(\mathbf{a})$ *is a set* $\mathcal{E} = \{\mathcal{C}_1, \ldots, \mathcal{C}_m\} \subseteq \mathsf{causes}(q(\mathbf{a}), \mathcal{K})$ *such that (i) every repair* $\mathcal{R}$ *of* $\mathcal{K}$ *contains some* $\mathcal{C}_i$*, and (ii) no proper subset of* $\mathcal{E}$ *satisfies this property.*

Example 4. There are 36 explanations for $\mathcal{K}_{\mathsf{ex}} \models_{\text{AR}} q_2(a)$, each taking one of the following two forms:

$$\begin{aligned}
\mathcal{E}_{ij} =& (\mathsf{Postdoc}(a) \wedge \mathsf{Teach}(a, c_i)) \vee (\mathsf{Adv}(a, b) \wedge \mathsf{Teach}(a, c_j)) \\
\mathcal{E}'_{ijk} =& (\mathsf{Postdoc}(a) \wedge \mathsf{Teach}(a, c_i)) \vee (\mathsf{FPr}(a) \wedge \mathsf{Teach}(a, c_j)) \\
& \vee (\mathsf{APr}(a) \wedge \mathsf{Teach}(a, c_k))
\end{aligned}$$

for some $i, j, k \in \{1, 2, 3\}$.

We next consider how to explain negative AR- and IAR-answers, that are brave-answers not entailed under AR or IAR semantics. For AR semantics, the idea is to give a (minimal) subset of the ABox that is consistent with the TBox and contradicts every cause of the query, since any such subset can be extended to a repair that omits all causes. For IAR semantics, the formulation is slightly different as we need only ensure that every cause is contradicted by some consistent subset.

Definition 3. *An explanation for* $\mathcal{K} \not\models_{AR} q(\mathbf{a})$ *is a* $\mathcal{T}$*-consistent subset* $\mathcal{E} \subseteq \mathcal{A}$ *such that (i)* $\mathcal{T}, \mathcal{E} \cup \mathcal{C} \models \bot$ *for every* $\mathcal{C} \in \mathsf{causes}(q(\mathbf{a}), \mathcal{K})$*, and (ii) no proper subset of* $\mathcal{E}$ *has this property. An explanation for* $\mathcal{K} \not\models_{IAR} q(\mathbf{a})$ *is a (possibly* $\mathcal{T}$*-inconsistent) subset* $\mathcal{E} \subseteq \mathcal{A}$ *such that (i) for every* $\mathcal{C} \in \mathsf{causes}(q(\mathbf{a}), \mathcal{K})$*, there exists a* $\mathcal{T}$*-consistent subset* $\mathcal{E}' \subseteq \mathcal{E}$ *with* $\mathcal{T}, \mathcal{E}' \cup \mathcal{C} \models \bot$*, and (ii) no proper subset of* $\mathcal{E}$ *has this property.*

Example 5. The unique explanation for $\mathcal{K}_{\mathsf{ex}} \not\models_{\mathrm{AR}} q_1(a)$ is $\mathsf{Postdoc}(a)$, which contradicts the three causes of $q_1(a)$. The explanations for $\mathcal{K}_{\mathsf{ex}} \not\models_{\mathrm{IAR}} q_2(a)$ are: $\mathsf{FPr}(a) \wedge \mathsf{Postdoc}(a)$, $\mathsf{APr}(a) \wedge \mathsf{Postdoc}(a)$, and $\mathsf{Adv}(a, b) \wedge \mathsf{Postdoc}(a)$.

When there are a large number of explanations for a given result, it may be impractical to present them all to the user. In such cases, one may choose instead to rank the explanations according to some preference criteria, and to present one or a small number of most *preferred explanations*. In the present work, we will use *cardinality* to rank explanations for brave- and IAR-answers and negative AR- and IAR-answers. For positive AR-answers, we consider two ways of ranking explanations: the *number of disjuncts*, and the total *number of assertions*. Another interesting criterion is the difficulty of the associated TBox reasoning. For example, we may compute for each cause the minimum number of TBox axioms needed to show that the cause yields the query, and then use this number to rank explanations for brave- and IAR-answers.

Example 6. Reconsider explanations $\mathcal{E}_{1\,1}$ and $\mathcal{E}'_{1\,2\,3}$ for $\mathcal{K}_{\mathsf{ex}} \models_{\mathrm{AR}} q_2(a)$. There are at least two reasons why $\mathcal{E}_{1\,1}$ may be considered easier to understand than $\mathcal{E}'_{1\,2\,3}$. First, $\mathcal{E}_{1\,1}$ contains fewer disjuncts, hence requires less disjunctive reasoning. Second, both disjuncts of $\mathcal{E}_{1\,1}$ use the same Teach assertion, whereas $\mathcal{E}'_{1\,2\,3}$ uses three different Teach assertions, which may lead the user to (wrongly) believe all are needed to obtain the query result. Preferring explanations having the fewest number of disjuncts, and among them, those involving a minimal set of assertions, leads to focusing on the explanations of the form $\mathcal{E}_{i\,i}$, where $i \in \{1, 2, 3\}$.

A second complementary approach is to concisely summarize the set of explanations in terms of the *necessary assertions* (i.e. those which appear in every explanation) and the *relevant assertions* (i.e. appear in at least one explanation).

Example 7. If we tweak the example KB to include n courses taught by a, then there would be $n^2 + n^3$ explanations for $\mathcal{K}_{\mathsf{ex}} \models_{\mathrm{AR}} q_2(a)$, built using only $n + 4$ assertions. Presenting the necessary assertions (in this case, $\mathsf{Postdoc}(a)$) and relevant ones ($\mathsf{FPr}(a)$, $\mathsf{APr}(a)$, $\mathsf{Adv}(a, b)$, $\mathsf{Teach}(a, c_i)$) gives a succinct overview of the set of explanations.

	brave, IAR	AR	neg. IAR	neg. AR
REL	in P	Σ_2^p-co	in P	NP-co
NEC	in P	NP-co	in P	coNP-co
REC	in P	BH_2-co	in P	in P
BEST REC†	in P	Π_2^p-co‡	coNP-co*	coNP-co*

† upper bounds hold for ranking criteria that can be decided in P
‡ Π_2^p-hard for smallest disjunction or fewest assertions
* coNP-hard for cardinality-minimal explanations

Fig. 1: Data complexity results for conjunctive queries.

4 Algorithms and Complexity Results

We next study the computational properties of the different notions of explanation defined in Section 3. In addition to the problem of computing explanations, we consider four natural decision problems: decide whether a given assertion appears in some explanation (REL) or in every explanation (NEC), decide whether a candidate is an explanation (REC), resp. a best explanation according a given criteria (BEST REC). As we target applications in which the ABox is significantly larger than the TBox and query, we use *data complexity*, which is only with respect to the size of the ABox, to measure the difficulty of these reasoning tasks.

Here and in the following section, we focus on KBs expressed in the lightweight logic DL-Lite$_\mathcal{R}$ since it is a popular choice for ontology-based data access and the only DL for which the three considered semantics have been implemented. We recall that in DL-Lite$_\mathcal{R}$, KB satisfiability and query answering are both in P w.r.t. data complexity [13], and conflicts are of size at most two.

Theorem 1. *The results in Figure 1 hold*[1]*.*

In what follows, we sketch the proof of the complexity results from Theorem 1, and we also explain how to compute explanations and relevant and necessary assertions. Often this will take the form of a reduction to a SAT-related reasoning task, which can be performed by a SAT solver.

Positive brave- and IAR-answers It is possible to compute the causes and their conflicts in polytime w.r.t. data complexity. It follows that recognizing a best explanation and computing the union and intersection of explanations to identify relevant and necessary assertions can also be done in P.

Positive AR-answers We relate explanations of AR-answers to minimal unsatisfiable subsets of a set of propositional clauses. Let us recall that, given sets F and H of soft and hard clauses respectively, a subset $M \subseteq F$ is a *minimal*

[1] The P upper bounds for REL and NEC can be improved to AC^0, but the proofs involve cumbersome query rewriting constructions that are less suited for use in practice than the methods used to show P membership.

unsatisfiable subset (MUS) of F w.r.t. H if (i) $M \cup H$ is unsatisfiable, and (ii) $M' \cup H$ is satisfiable for every $M' \subsetneq M$.

To explain $\mathcal{K} \models_{\text{AR}} q(\mathbf{a})$, we consider the soft clauses

$$\varphi_{\neg q} = \{ \bigvee_{\beta \in \mathsf{confl}(\mathcal{C},\mathcal{K})} x_\beta \mid \mathcal{C} \in \mathsf{causes}(q(\mathbf{a}),\mathcal{K})\}$$

and the hard clauses

$$\varphi_{cons} = \{\neg x_\alpha \vee \neg x_\beta \mid x_\alpha, x_\beta \in \mathsf{vars}(\varphi_{\neg q}), \{\alpha,\beta\} \in \mathsf{confl}(\mathcal{K})\}$$

It was proven in [9] that $\mathcal{K} \models_{\text{AR}} q(\mathbf{a})$ iff $\varphi_{\neg q} \cup \varphi_{cons}$ is unsatisfiable, and we can further show:

Proposition 1. *The explanations for $\mathcal{K} \models_{AR} q(\mathbf{a})$ are the sets of causes which correspond to the MUSes of $\varphi_{\neg q}$ w.r.t. φ_{cons}.*

By the preceding proposition, we can compute all explanations for $\mathcal{K} \models_{\text{AR}} q(\mathbf{a})$ by computing all MUSes of $\varphi_{\neg q}$ w.r.t. φ_{cons}.

The upper bounds for REL, NEC, and REC follow immediately from Proposition 1 and known complexity results for MUSes [23]. For BEST REC, we show that an explanation is not a best one by guessing a better candidate and checking in BH_2 that it is an explanation.

For the lower bounds, we use the following reduction from the corresponding MUSes problems. Let $\varphi_0 = \{C_1, ..., C_n\}$ be an unsatisfiable set of clauses over $\{X_1, ..., X_p\}$, and consider the following KB:

$$\mathcal{T}_0 = \{\exists P^- \sqsubseteq \neg\exists N^-, \exists U^- \sqsubseteq \neg\exists P, \exists U^- \sqsubseteq \neg\exists N, \exists U \sqsubseteq A\}$$
$$\mathcal{A}_0 = \{P(c_i, x_j) \mid X_j \in C_i\} \cup \{N(c_i, x_j) \mid \neg X_j \in C_i\} \cup \{U(a, c_i) \mid 1 \leq i \leq n\}$$

It was proven in [8] that φ is unsatisfiable iff $\mathcal{T}_0, \mathcal{A}_0 \models_{\text{AR}} A(a)$. We further observe that the explanations of $\mathcal{T}_0, \mathcal{A}_0 \models_{\text{AR}} A(a)$ correspond to the MUSes of φ_0 w.r.t. $\emptyset$. The lower bounds for REL, NEC, and REC follow immediately, and the proof of [23] of Σ_2^p-hardness of deciding if there exists a MUS of size at most k also shows that deciding if a set of clauses is a smallest MUS is Π_2^p-hard, so deciding if an explanation contains a smallest number of causes is Π_2^p-complete. Since here a cause contains only one assertion, deciding if an explanation contains a smallest number of assertions is also Π_2^p-complete.

Negative IAR-answers To compute the explanations and cardinality-minimal explanations for negative IAR-answers, we rely on the following proposition.

Proposition 2. *The explanations (resp. cardinality-minimal explanations) for $\mathcal{K} \not\models_{IAR} q(\mathbf{a})$ are the sets of assertions corresponding to the inclusion-minimal (resp. cardinality-minimal) models of $\varphi_{\neg q}$.*

The relevant and necessary assertions can however be computed in P, from the causes and their conflicts, computable in P w.r.t. data complexity. An assertion

α is necessary iff it is the only conflict of some cause, and it is relevant iff it is in conflict with a cause $\mathcal{C}$ such that for every other cause $\mathcal{C}'$, if $\mathsf{confl}(\mathcal{C}',\mathcal{K}) \subseteq \mathsf{confl}(\mathcal{C},\mathcal{K})$, then $\alpha \in \mathsf{confl}(\mathcal{C}',\mathcal{K})$. We can therefore compute the necessary and relevant assertions by examining the conflicts $\mathsf{confl}(\mathcal{C},\mathcal{K})$ of each cause $\mathcal{C}$:

- If $\mathsf{confl}(\mathcal{C},\mathcal{K})$ contains only one assertion, that assertion is necessary
- We compute the set of assertions relevant for $\mathcal{C}$ as follows:
  - let $\mathsf{Relevant}_{\mathcal{C}} = \mathsf{confl}(\mathcal{C},\mathcal{K})$
  - for all $\mathsf{confl}(\mathcal{C}',\mathcal{K})$ such that $\mathsf{confl}(\mathcal{C}',\mathcal{K}) \subseteq \mathsf{confl}(\mathcal{C},\mathcal{K})$, $\mathsf{Relevant}_{\mathcal{C}} \leftarrow \mathsf{Relevant}_{\mathcal{C}} \cap \mathsf{confl}(\mathcal{C}',\mathcal{K})$

  All assertions in $\mathsf{Relevant}_{\mathcal{C}}$ are relevant for explaining the negative answer.

Recognizing an explanation for a negative answer under IAR semantics can be done in P by checking that it contains at least one conflict of each cause, and that it is minimal. For BEST REC, an explanation is not a best one if we can guess a better one, so for ranking criteria that can be decided in P, the problem BEST REC can be decided in coNP.

The lower bound for BEST REC for cardinality-minimal explanations is by reduction from the problem of deciding if a valuation ν that satisfies a monotone 3-SAT formula $\varphi = C_1 \wedge ... \wedge C_n$ over $\{X_1, ..., X_p\}$ assigns a minimal number of variables to true. Take the KB and query

$$\begin{aligned}
\mathcal{T} &= \{\exists P_k^- \sqsubseteq \neg T \mid 1 \leq k \leq 3\}\\
\mathcal{A} &= \{T(x_i) \mid 1 \leq i \leq p\} \cup \{P_k(c_j, x_i) \mid X_i\ k^{th} \text{ term of } C_j\}\\
q &= \exists y z_1 z_2 z_3\, P_1(y, z_1) \wedge P_2(y, z_2) \wedge P_3(y, z_3)
\end{aligned}$$

An explanation for $\mathcal{K} \not\models_{\text{IAR}} q$ is a set $\mathcal{E}$ of T such that for every c_j, there is at least one $X_i \in C_j$ such that $T(x_i) \in \mathcal{E}$. Thus, ν assigns a minimal number of variables to true iff $\mathcal{E} = \{T(x_i) \mid \nu(X_i) = \mathsf{true}\}$ is a smallest explanation.

Negative AR-answers Similarly to negative IAR-answers, there is a correspondence between the explanations and the models of propositional formulas.

Proposition 3. *The explanations (resp. cardinality-minimal explanations) for $\mathcal{K} \not\models_{AR} q(\mathbf{a})$ are the sets of assertions corresponding to the inclusion-minimal (resp. cardinality-minimal) models of $\varphi_{\neg q} \wedge \varphi_{cons}$.*

Recognizing an explanation $\mathcal{E}$ for a negative AR-answer can be done by checking consistency of $\mathcal{E}$, inconsistency of $\mathcal{E} \cup \mathcal{C}$ for every cause $\mathcal{C}$, and minimality of $\mathcal{E}$. Since each can be done in P, REL is in NP and BEST REC in coNP for ranking criteria that can be decided in P (guess an explanation that contains the assertion, resp. is a better explanation). Deciding if an assertion is necessary is in coNP: α is necessary iff $\varphi_{\neg q} \wedge \varphi_{cons} \wedge \neg x_\alpha$ is unsatisfiable. To compute all necessary assertions, we test every assertion involved in $\varphi_{\neg q} \wedge \varphi_{cons}$.

The lower bounds are proved by reduction from (UN)SAT:

- For NEC, we consider $\mathcal{T}_1 = \mathcal{T}_0 \cup \{\exists U \sqsubseteq \neg S\}$ and $\mathcal{A}_1 = \mathcal{A}_0 \cup \{S(a)\}$. It can be shown that φ_0 is satisfiable iff $S(a)$ is not necessary for explaining $\mathcal{T}_1, \mathcal{A}_1 \not\models_{\text{AR}} A(a)$.

– For REL, we use $\mathcal{T}_1$ and $\mathcal{A}_2 = \mathcal{A}_1 \cup \{U(a, c_{n+1}), P(c_{n+1}, x_{p+1})\}$, and we show that φ_0 is satisfiable iff $P(c_{n+1}, x_{p+1})$ is relevant for $\mathcal{T}_1, \mathcal{A}_2 \not\models_{\mathrm{AR}} A(a)$.
– For BEST REC, we consider the KB:

$$\begin{aligned}\mathcal{T}_3 =& \mathcal{T}_0 \cup \{U_1 \sqsubseteq U, U_2 \sqsubseteq U, \exists U_1^- \sqsubseteq \neg T, \exists U_2 \sqsubseteq \neg S\}\\ \mathcal{A}_3 =& \{P(c_i, x_j) \mid X_j \in C_i\} \cup \{N(c_i, x_j) \mid \neg X_j \in C_i\} \cup \\ & \{U_1(a, c_i), U_2(a, c_i), T(c_i) \mid 1 \leq i \leq n\} \cup \{S(a)\}\end{aligned}$$

One can show that $\mathcal{E} = \{S(a), T(c_1), ..., T(c_m)\}$ is a smallest explanation for $\mathcal{T}_3, \mathcal{A}_3 \not\models_{\mathrm{AR}} A(a)$ iff φ_0 is unsatisfiable.

5 Prototype and experiments

We implemented our explanation framework in Java using the CQAPri system (`www.lri.fr/~bourgaux/CQAPri`) and the SAT4J v2.3.4 SAT solver (`www.sat4j.org`). CQAPri supports querying of DL-Lite$_\mathcal{R}$ KBs using a variety of inconsistency-tolerant semantics [9], including those we focus on: brave, AR and IAR. In addition to satisfiability testing, SAT4J can be used to compute MUSes and cardinality-minimal models [6].

Our prototype runs in two modes: either it explains *some* selected query answers among those computed by CQAPri, or *all* the answers while they are being computed by CQAPri. These answers are divided into three classes:

– Possible: $\mathcal{K} \models_{\text{brave}} q(\mathbf{a})$ and $\mathcal{K} \not\models_{\mathrm{AR}} q(\mathbf{a})$
– Likely: $\mathcal{K} \models_{\mathrm{AR}} q(\mathbf{a})$ and $\mathcal{K} \not\models_{\mathrm{IAR}} q(\mathbf{a})$
– (Almost) sure: $\mathcal{K} \models_{\mathrm{IAR}} q(\mathbf{a})$

Concretely, explaining an answer **a** consists in providing, for the relevant semantics S, S' according to the class of **a**: (*i*) *all* explanations of $\mathcal{K} \models_S q(\mathbf{a})$, as well as necessary and relevant assertions, and (*ii*) *one* smallest explanation of $\mathcal{K} \not\models_{S'} q(\mathbf{a})$, with necessary and relevant assertions when $S' = IAR$, and necessary assertions when $S' = AR$ together with necessary and relevant assertions for $\mathcal{K} \not\models_{IAR} q(\mathbf{a})$. Positive explanations are ranked as explained in Section 3. For explanations of $\mathcal{K} \models_{AR} q(\mathbf{a})$, the user chooses the priority of the two criteria (i.e., number of disjuncts and total number of assertions).

To explain the answers to a query, we rely on the algorithms of the preceding section.The system therefore needs the *causes* of the query answers as well as their *conflicts*. The conflicts are directly available from CQAPri but not the causes: CQAPri uses query rewriting to identify consistent (but not necessarily minimal) subsets of the ABox entailing the answers, from which we must prune the non-minimal ones.

Experimental setting We used the CQAPri benchmark [9] available at `www.lri.fr/~bourgaux/CQAPri`, which builds on the DL-Lite$_\mathcal{R}$ version [25] of the Lehigh University Benchmark (`swat.cse.lehigh.edu/projects/lubm`). It extends the DL-Lite$_\mathcal{R}$ TBox with negative inclusions and describes how to obtain an ABox with

Query id	shape	#atoms	#variables	#rewritings
g2	atomic	1	1	44
g3	atomic	1	1	44
q1	dag	5	2	6401
q2	tree	3	2	450
q3	tree	2	3	155
q4	dag	6	4	202579

Table 1: Characteristics of the test queries: query shape, numbers of atoms, number of variables, and number of CQs in the UCQ-rewriting.

a natural repartition of conflicts by adding assertions to an initial ABox consistent with the enriched TBox. The LUBM benchmark has been used to evaluate close works such [16, 17] on explanations and [31] which evaluates a system that computes IAR-answers. Since the explanations to an answer generally involve only a limited part of the ABox linked to the individuals of the answer, we choose to use datasets of limited size to get a reasonable number of answers and be able to systematically compute explanations for all these answers.

We picked the so-called u5p0 consistent ABox from the CQAPri benchmark, because of its reasonable size (∼500K facts), from which we generated ten inconsistent ABoxes with different ratios of assertions in conflicts by adding from 384 to 22873 assertions. These ABoxes are denoted cX, with X the ratio of conflicts varying from a realistic value of 2% to a value of 48% challenging our approach. Also, the way we generate conflicts ensures $\mathtt{cX} \subseteq \mathtt{cY}$ if $\mathtt{X} \leq \mathtt{Y}$.

The following six queries were used in our experiments:

$$
\begin{aligned}
g2 &= \mathsf{Organization}(x)\\
g3 &= \mathsf{Employee}(x)\\
q1 &= \exists y\, \mathsf{Person}(x) \wedge \mathsf{takesCourse}(x,y) \wedge \mathsf{GraduateCourse}(y) \wedge\\
&\quad \mathsf{takesCourse}(GraduateStudent131, y) \wedge \mathsf{Person}(GraduateStudent131)\\
q2 &= \exists x\, \mathsf{Employee}(x) \wedge \mathsf{memberOf}(x, Department4.University0) \wedge\\
&\quad \mathsf{degreeFrom}(x,y)\\
q3 &= \exists y\, \mathsf{teacherOf}(x,y) \wedge \mathsf{degreeFrom}(x,z)\\
q4 &= \exists yz\, \mathsf{Employee}(x) \wedge \mathsf{degreeFrom}(x,y) \wedge \mathsf{memberOf}(x,z) \wedge \mathsf{Employee}(u) \wedge\\
&\quad \mathsf{degreeFrom}(u,y) \wedge \mathsf{memberOf}(u,z)
\end{aligned}
$$

Table 1 displays the characteristics of these queries, which have (*i*) a variety of structural aspects and number of rewritings, and (*ii*) answers in the three considered classes (see Table 2). We borrowed g2 and g3 from [9] and designed the other queries ourselves.

Our experiments ran on an Intel i5-3470 CPU server at 3.20 GHz, with 8 Go of RAM, and running Windows 7. Reported times are averaged over 10 runs.

Experimental results We summarize below the general tendencies we observed. Table 2 shows for each query the number of answers of each classe it has, as well as the distribution of the explanation times of these answers. Figures 2 and 3 show the proportion of time spent in the different phases and the total

	number of answers	< 1 ms	$[1, 10[$ ms	$[10, 100[$ ms	$[100, 500[$ ms	> 500 ms
g2 Sure	1764	99.8	0.2	0	0	0
Likely	681	52	45.8	2.1	0	0.1
Poss.	818	44.9	22.3	27.1	5.7	0
g3 Sure	6637	99.9	0.1	0	0	0
Likely	433	81.1	18.7	0.2	0	0
Poss.	895	90.3	8.8	0.9	0	0
q1 Likely	10	20	30	50	0	0
q2 Sure	30	100	0	0	0	0
Poss.	103	95.1	4.9	0	0	0
q3 Sure	2571	99.5	0.5	0	0	0
Likely	4	50	50	0	0	0
Poss.	7736	91.8	8.1	0.1	0	0
q4 Sure	6712	99.8	0.2	0	0	0
Likely	186	72.1	27.4	0.5	0	0
Poss.	2960	90.5	9.4	0.1	0	0

Table 2: Number of answers of each class and distribution of explanation times (in milliseconds) per query on c28.

time to explain all query answers over ABoxes with different proportion of assertions involved in some conflict. The total cost of explanations is given by the two upper bars which represent the additional cost during the execution phase (add. exec. cost) and the time spent in explaining (explain); the three lower bars relate to the query answering phase, which consists in rewriting the query (rewrite), executing the rewritten query (execute), and identifying IAR, AR and brave answers (CQA). Regarding explanation costs, additional cost during the execution phase consists mainly in pruning non-minimal consistent subsets of the ABox entailing the answers to get the causes, and explaining is computing the explanations from the causes and conflicts.

The main conclusion is that explaining some selected or all the answers to a query, as described above, is always feasible and quite fast when there is a *few* percent of conflicts in the ABox (Figure 2, c2 case), as is likely to be the case in most real applications. Moreover, in *all* the experiments we made, explaining an answer takes often less than 1ms and almost never more than 0.5s.

	g2			g3			q1			q2			q3			q4		
	c2	c28	c48	c2	c28	c48	c2	c28	c48	c2	c28	c48	c2	c28	c48	c2	c28	c48
Sure	95.9	54	28	98.9	83.3	56.1	100	0	0	89.8	22.6	0	92	24.93	1.3	95.7	68.1	45.8
Likely	2.4	20.9	19.1	0.3	5.5	13.2	0	100	0	0	0	0	0	0.04	0.1	0.3	1.9	1.1
Possible	1.7	25.1	52.9	0.8	11.2	30.7	0	0	100	10.2	77.4	100	8	75.03	98.6	4	30	53.1

Table 3: Distribution of answers in the different classes

In more detail, adding conflicts to the ABox complicates the explanations of answers, due to their shift from the Sure to the Likely and Possible classes, as Table 3 shows. Explaining such answers indeed come at higher computational cost. Figures 2, 3 illustrate this phenomenon. The general trend is illustrated with the case of q3 in Figure 3: adding more conflicts makes the difficulty of explaining

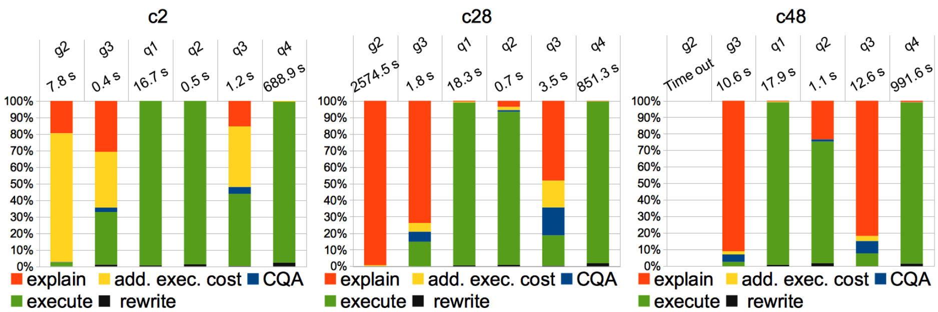

Fig. 2: Proportion of time spent in the different phases and total time (in second) to explain all query answers on three ABoxes.

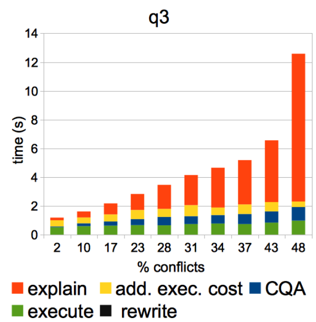

Fig. 3: Time spent for explaining all answers of q3 w.r.t. proportion of assertions involved in some conflict.

grow more rapidly. Observe however that the shift of answers from the Likely to the Possible classes does not necessarily complicates their explanations, as shows the shift of the 10 answers of q1 from the Likely class to the Possible one (Figure 2, c28 and c48, and Table 3). A striking case is that of g2, which leads to a long explanation time on c28 and even a timeout on c48 (>60min). On c28, most of the time (∼40min) is spent in explaining the only Likely answer in the >500 column of Table 2: the difficulty is to compute a smallest explanation for $\mathcal{K} \not\models_{\text{IAR}} q(\mathbf{a})$ due to the unusual size of the explanation (24 assertions, whereas the sizes encountered in our experiments were up to 6).

Finally, we observed that the average number of explanations per answer is often reasonably low, although some answers have a large number of explanations (e.g., on c28, less than 10 on average, but 654 for an IAR-answer to g2, 243 for an AR-answer to q1, and 693 for a brave-answer to g2).

6 Related Work on Explanations

As mentioned in Section 1, there has been significant interest in equipping DL reasoning systems with explanation facilities. The earliest work proposed formal proof systems as a basis for explaining concept subsumptions [26, 12]. The post-2000 literature mainly focuses on *axiom pinpointing* [32, 21], in which the problem is to generate minimal subsets of the KB that yield a given (surprising or undesirable) consequence; such subsets are often called *justifications*. An extensive experimental evaluation of algorithms for computing all justifications for expressive DLs is presented in [20]. For the lightweight DL $\mathcal{EL}+$, justifications have been shown to correspond to minimal models of propositional Horn formulas and can computed using SAT solvers [33], and a polynomial algorithm has been proposed to compute one justification in [4]. In DL-Lite, the problem is simpler: all justifications of a TBox axiom can be enumerated in polynomial delay [29].

It should be noted that work on axiom pinpointing has thus far focused on explaining entailed TBox axioms (or possibly ABox assertions), but not answers to conjunctive queries. The latter problem is considered by Borgida et al. [11], who introduce a proof-theoretic approach to explaining positive answers to CQs over DL-Lite$_\mathcal{A}$ KBs. The approach outputs a single proof, involving both TBox axioms and ABox assertions, using minimality criteria to select a 'simplest' proof.

More recently, the problem of explaining negative query answers over DL-Lite$_\mathcal{A}$ KBs has been studied by Calvanese et al. [14]. Formally, the explanations for $\mathcal{T}, \mathcal{A} \not\models q(\mathbf{a})$ correspond to sets $\mathcal{A}'$ of ABox assertions such that $\mathcal{T}, \mathcal{A} \cup \mathcal{A}' \models q(\mathbf{a})$. Practical algorithms and an implementation for computing such explanations were described in [16] (where the problem is called ABox abduction). The latter work was recently extended in [17] to the case of inconsistent KBs. Essentially the idea is to add a set of ABox assertions that will lead to the answer holding under IAR semantics (in particular, the new assertions must not introduce any inconsistencies). By contrast, in our setting, negative query answers result not from the absence of supporting facts, but rather the presence of conflicting assertions, and thus, our explanations are composed of assertions from the original ABox.

Probably the closest related work is [1, 2] which addresses the problem of explaining positive and negative answers under the inconsistency-tolerant semantics ICR [8] in the argumentation framework. The authors define explanations of positive answers as sets of facts and rules from which the query can be derived and propose a dialectical explanation for query failure, which consists of a dialogue between the user and the reasoner that gives counter-argument to arguments supporting the query and may explain why they are conflicting. We do not consider the same semantics but their ideas can be used to present the TBox axioms that link our causes to the query.

Finally, we note that the problem of explaining query results has been studied in the database community, cf. [15] for a survey pertaining to positive query answers and [18] for negative answers.

7 Conclusion and Future Work

In this paper, we proposed a framework for explaining query answers and non-answers over DL KBs under three commonly considered inconsistency-tolerant semantics (brave, AR, IAR). We then investigated the computational properties and practical feasibility of the framework, focusing on the lightweight description logic DL-Lite$_\mathcal{R}$ that underpins the OWL 2 QL profile [27]. While some of the explanation tasks were shown to be intractable, we exhibited tight connections with variants of propositional satisfiability, which allowed us to exploit the facilities of modern SAT solvers. The experimental evaluation of our prototype system shows that explanations of query (non-)answers can be generated very quickly (typically less than 1ms).

There are several natural directions for future work. First, we plan to accompany our explanations with details on the TBox reasoning involved, using the work of [11] on proofs of positive query answers as a starting point. As mentioned in Section 3, the difficulty of such proofs could provide an additional criteria for ranking explanations. The work on the cognitive complexity of justifications [19] may provide a starting point in defining a suitable measure for quantifying the difficulty of proofs. Second, our experiments showed that a query answer can possess a very large number of explanations, many of which are quite similar in structure. We therefore plan to investigate ways of improving the presentation of explanations, e.g. by identifying and grouping similar explanations (as has been done for justifications in [5]), by defining a notion of representative explanation as in [16], or by adopting a factorized representation (like in [28]). Finally, it would be interesting to explore how explanations can be used to partially repair the data based upon the user's feedback. For instance, in our running example, if $q_2(a)$ is sure according to the user, and should therefore hold under IAR semantics, the system may propose to remove $\mathsf{Postdoc}(a)$, which is necessary for explaining $\mathcal{K}_{\mathsf{ex}} \not\models_{\text{IAR}} q_2(a)$.

References

1. Arioua, A., Tamani, N., Croitoru, M.: On conceptual graphs and explanation of query answering under inconsistency. In: Proc. of ICCS (2014)
2. Arioua, A., Tamani, N., Croitoru, M., Buche, P.: Query failure explanation in inconsistent knowledge bases using argumentation. In: Proc. of COMMA (2014)
3. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press (2003)
4. Baader, F., Peñaloza, R., Suntisrivaraporn, B.: Pinpointing in the description logic EL^+. In: Proc. of KI (2007)
5. Bail, S., Parsia, B., Sattler, U.: The logical diversity of explanations in OWL ontologies. In: Proc. of CIKM (2013)
6. Berre, D.L., Parrain, A.: The sat4j library, release 2.2. JSAT 7(2-3), 59–64 (2010)
7. Bertossi, L.E.: Database Repairing and Consistent Query Answering. Synthesis Lectures on Data Management, Morgan & Claypool Publishers (2011)
8. Bienvenu, M.: On the complexity of consistent query answering in the presence of simple ontologies. In: Proc. of AAAI (2012)
9. Bienvenu, M., Bourgaux, C., Goasdoué, F.: Querying inconsistent description logic knowledge bases under preferred repair semantics. In: Proc. of AAAI (2014)

10. Bienvenu, M., Rosati, R.: Tractable approximations of consistent query answering for robust ontology-based data access. In: Proc. of IJCAI (2013)
11. Borgida, A., Calvanese, D., Rodriguez-Muro, M.: Explanation in the DL-Lite family of description logics. In: Proc. of OTM (2008)
12. Borgida, A., Franconi, E., Horrocks, I.: Explaining ALC subsumption. In: Proc. of ECAI (2000)
13. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The DL-Lite family. J. Autom. Reasoning (JAR) 39(3), 385–429 (2007)
14. Calvanese, D., Ortiz, M., Simkus, M., Stefanoni, G.: Reasoning about explanations for negative query answers in DL-Lite. J. Artif. Intell. Res. (JAIR) 48, 635–669 (2013)
15. Cheney, J., Chiticariu, L., Tan, W.C.: Provenance in databases: Why, how, and where. Foundations and Trends in Databases 1(4), 379–474 (2009)
16. Du, J., Wang, K., Shen, Y.: A tractable approach to abox abduction over description logic ontologies. In: Proc. of AAAI (2014)
17. Du, J., Wang, K., Shen, Y.: Towards tractable and practical abox abduction over inconsistent description logic ontologies. In: Proc. of AAAI (2015)
18. Herschel, M., Hernández, M.A.: Explaining missing answers to SPJUA queries. PVLDB 3(1), 185–196 (2010)
19. Horridge, M., Bail, S., Parsia, B., Sattler, U.: The cognitive complexity of OWL justifications. In: Proc. of ISWC (2011)
20. Horridge, M., Parsia, B., Sattler, U.: Extracting justifications from bioportal ontologies. In: Proc. of ISWC (2012)
21. Kalyanpur, A., Parsia, B., Sirin, E., Hendler, J.A.: Debugging unsatisfiable classes in OWL ontologies. J. Web Sem. 3(4), 268–293 (2005)
22. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant semantics for description logics. In: Proc. of RR (2010)
23. Liberatore, P.: Redundancy in logic I: Cnf propositional formulae. Artif. Intell. (AIJ) (2005)
24. Lukasiewicz, T., Martinez, M.V., Simari, G.I.: Complexity of inconsistency-tolerant query answering in datalog+/-. In: Proc. of OTM. pp. 488–500 (2013)
25. Lutz, C., Seylan, I., Toman, D., Wolter, F.: The combined approach to OBDA: Taming role hierarchies using filters. In: Proc. of ISWC (2013)
26. McGuinness, D.L., Borgida, A.: Explaining subsumption in description logics. In: Proc. of IJCAI (1995)
27. Motik, B., Cuenca Grau, B., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C.: OWL 2 Web Ontology Language profiles. W3C Recommendation (11 December 2012), available at `http://www.w3.org/TR/owl2-profiles/`
28. Olteanu, D., Zavodny, J.: Factorised representations of query results: size bounds and readability. In: Proc. of ICDT (2012)
29. Peñaloza, R., Sertkaya, B.: Complexity of axiom pinpointing in the dl-lite family of description logics. In: Proc. of ECAI (2010)
30. Rosati, R.: On the complexity of dealing with inconsistency in description logic ontologies. In: Proc. of IJCAI (2011)
31. Rosati, R., Ruzzi, M., Graziosi, M., Masotti, G.: Evaluation of techniques for inconsistency handling in OWL 2 QL ontologies. In: Proc. of ISWC (2012)
32. Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: Proc. of IJCAI (2003)
33. Sebastiani, R., Vescovi, M.: Axiom pinpointing in lightweight description logics via horn-sat encoding and conflict analysis. In: Proc. of CADE (2009)

A Proofs for Section 4

In what follows, we will not distinguish between sets of clauses and conjunctions of clauses.

A.1 Proofs of Propositions 1 and 2, 3

We recall that the encoding is composed of the set of soft clauses:

$$\varphi_{\neg q} = \bigwedge_{\mathcal{C} \in \mathsf{causes}(q(\mathbf{a}),\mathcal{K})} \Big(\bigvee_{\beta \in \mathsf{confl}(\mathcal{C},\mathcal{K})} x_\beta \Big)$$

and the set of hard clauses:

$$\varphi_{cons} = \bigwedge_{x_\alpha, x_\beta \in \mathsf{vars}(\varphi_{\neg q}), \{\alpha,\beta\} \in \mathsf{confl}(\mathcal{K})} (\neg x_\alpha \vee \neg x_\beta)$$

Proposition 1. The explanations of $\mathcal{K} \models_{\text{AR}} q(\mathbf{a})$ are the sets of causes which correspond to the MUSes of $\varphi_{\neg q} \wedge \varphi_{cons}$.

Propositions 2, 3. The explanations (resp. of smallest cardinality) of $\mathcal{K} \not\models_{\text{AR}} q(\mathbf{a})$ are the sets of assertions which correspond to the inclusion-minimal (resp. cardinality-minimal) models of $\varphi_{\neg q} \wedge \varphi_{cons}$, and the explanations (resp. of smallest cardinality) of $\mathcal{K} \not\models_{\text{IAR}} q(\mathbf{a})$ are the sets of assertions which correspond to the inclusion-minimal (resp. cardinality-minimal) models of $\varphi_{\neg q}$.

Proof. It is shown in [9] that $\mathcal{K} \models_{\text{AR}} q(\mathbf{a})$ iff $\varphi_{\neg q} \wedge \varphi_{cons}$ is unsatisfiable. This is because the assertions whose corresponding variables are assigned to true in a valuation that satisfies $\varphi_{\neg q} \wedge \varphi_{cons}$ form a subset of the ABox which: (i) contradicts every cause, since $\varphi_{\neg q}$ states that for every cause, one conflicting assertion is selected, and (ii) is consistent, since φ_{cons} states that two assertions in a conflict cannot be selected together. Thus, the inclusion-minimal models of $\varphi_{\neg q} \wedge \varphi_{cons}$ are precisely the explanations for negative AR-answers. For negative IAR-answers, we drop the consistency condition, so explanations correspond instead to inclusion-minimal models of the (trivially satisfiable) $\varphi_{\neg q}$. We have thus established Propositions 2, 3.

We next consider the case of positive AR-answers (Proposition 1). Since φ_{cons} is composed of hard clauses (and trivially satisfied by assigning all variables to false), when $\varphi_{\neg q} \wedge \varphi_{cons}$ is unsatisfiable, the MUSes are sets of clauses of $\varphi_{\neg q}$ which cannot be satisfied together with φ_{cons}. A set of clauses of $\varphi_{\neg q}$ is unsatisfiable together with φ_{cons} just in the case where it is not possible to select one conflicting assertion for each corresponding cause in a consistent way. It follows that the MUSes are the explanations of $\mathcal{K} \models_{\text{AR}} q(\mathbf{a})$.

A.2 Complexity upper bounds

The following two lemmas will be used for the upper complexity bounds.

Lemma 1. *An assertion α is relevant for $\mathcal{K} \not\models_{IAR} q(\mathbf{a})$ iff it is in conflict with a cause $\mathcal{C}$ such that for every other cause $\mathcal{C}'$, if $\mathsf{confl}(\mathcal{C}',\mathcal{K}) \subseteq \mathsf{confl}(\mathcal{C},\mathcal{K})$, then $\alpha \in \mathsf{confl}(\mathcal{C}',\mathcal{K})$.*

Proof. For the first direction, suppose α is relevant for $\mathcal{K} \not\models_{\mathrm{IAR}} q(\mathbf{a})$. Then there is a subset $\mathcal{E} \subseteq \mathcal{A}$ with $\alpha \in \mathcal{E}$ such that every cause of $q(\mathbf{a})$ is in conflict with some assertion in $\mathcal{E}$, and no proper subset of $\mathcal{E}'$ possesses this property. Since $\mathcal{E}$ is a minimal set of assertions having this property, we know that there is some cause $\mathcal{C}$ that does not conflict with any assertion in $\mathcal{E} \setminus \{\alpha\}$, and so there cannot exist another cause $\mathcal{C}'$ such that $\mathsf{confl}(\mathcal{C}',\mathcal{K}) \subseteq \mathsf{confl}(\mathcal{C},\mathcal{K})$ and $\alpha \notin \mathsf{confl}(\mathcal{C}',\mathcal{K})$.

For the second direction, suppose that α is in conflict with a cause $\mathcal{C}$ of $q(\mathbf{a})$ and for every other cause $\mathcal{C}'$, $\mathsf{confl}(\mathcal{C}',\mathcal{K}) \subseteq \mathsf{confl}(\mathcal{C},\mathcal{K})$ implies $\alpha \in \mathsf{confl}(\mathcal{C}',\mathcal{K})$. It follows that for every cause $\mathcal{C}'$ of $q(\mathbf{a})$, either $\alpha \in \mathsf{confl}(\mathcal{C}',\mathcal{K})$, or there exists an assertion $\beta_{\mathcal{C}'} \in \mathsf{confl}(\mathcal{C}',\mathcal{K})$ such that $\beta_{\mathcal{C}'} \notin \mathsf{confl}(\mathcal{C},\mathcal{K})$. We can therefore construct an explanation for $\mathcal{K} \not\models_{\mathrm{IAR}} q(\mathbf{a})$ by taking α together with some of the assertions $\beta_{\mathcal{C}'}$.

Lemma 2. *An assertion α is necessary for $\mathcal{K} \not\models_{AR} q(\mathbf{a})$ iff $\varphi_{\neg q} \wedge \varphi_{cons} \wedge \neg x_\alpha$ is unsatisfiable.*

Proof. The proof of Proposition 3 shows that the explanations of $\mathcal{K} \not\models_{\mathrm{AR}} q(\mathbf{a})$ correspond to the minimal models of $\varphi_{\neg q} \wedge \varphi_{cons}$. It follows that $\varphi_{\neg q} \wedge \varphi_{cons} \wedge \neg x_\alpha$ is unsatisfiable iff every model of $\varphi_{\neg q} \wedge \varphi_{cons}$ contains x_α, i.e., iff every explanation for $\mathcal{K} \not\models_{\mathrm{AR}} q(\mathbf{a})$ contains α.

A.3 Complexity lower bounds

We next detail the proofs of the complexity lower bounds, which are broken into several propositions.

Proposition 4. *Regarding explanations for AR-answers,* REC *is* BH_2*-hard,* NEC *is* NP*-hard,* REL *is* Σ^p_2*-hard, and* BEST REC *is* Π^p_2*-hard w.r.t. data complexity.*

Proof. We link the explanations of AR-answers with the minimum unsatisfiable subsets (MUSes) of a CNF formula.

Let $\varphi_0 = C_1 \wedge ... \wedge C_n$ be an unsatisfiable set of clauses over $\{X_1, ..., X_p\}$. Consider the following KB and query (borrowed from [8]):

$$\begin{aligned}
\mathcal{T}_0 &= \{\exists P^- \sqsubseteq \neg\exists N^-, \exists U^- \sqsubseteq \neg\exists P, \exists U^- \sqsubseteq \neg\exists N, \exists U \sqsubseteq A\}\\
\mathcal{A}_0 &= \{P(c_i, x_j) \mid X_j \in C_i\} \cup \{N(c_i, x_j) \mid \neg X_j \in C_i\} \cup \{U(a, c_i) \mid 1 \leq i \leq n\}\\
q_0 &= A(x)
\end{aligned}$$

The causes for $q_0(a)$ are given by the assertions $U(a, c_i)$, which are in conflict with assertions of the form $P(c_i, x_j)$ or $N(c_i, x_j)$. It was shown in [8] that $\mathcal{T}_0, \mathcal{A}_0 \models_{\mathrm{AR}} A(a)$ iff φ_0 is unsatisfiable. To prove the proposition, we will require the following stronger claim:

Claim. The following are equivalent:

1. the set of clauses $\{C_{i_1}, ..., C_{i_k}\}$ is unsatisfiable
2. every repair of $(\mathcal{T}_0, \mathcal{A}_0)$ contains some assertion from $\{U(a, c_{i_1}), ..., U(a, c_{i_k})\}$

Proof of claim. It will be more convenient to show that the negations of the two statements are equivalent. First suppose that $\{C_{i_1}, ..., C_{i_k}\}$ is satisfiable, as witnessed by the satisfying assignment ν. Define a repair $\mathcal{R}_\nu$ of $(\mathcal{T}_0, \mathcal{A}_0)$ by including the assertion $P(c_i, v_j)$ if $\nu(v_j) = \mathsf{true}$, including $N(c_i, v_j)$ if $\nu(v_j) = \mathsf{false}$, and then adding as many other assertions as needed to obtain a maximal $\mathcal{T}_0$-consistent subset. Since ν satisfies every clause in $\{C_{i_1}, ..., C_{i_k}\}$, it follows that for every index $\ell \in \{i_1, \ldots, i_k\}$, the clause C_ℓ contains a positive literal v_ℓ such that $\nu(v_\ell) = \mathsf{true}$, or a negative literal $\neg v_\ell$ such that $\nu(v_\ell) = \mathsf{false}$. In the former case, $\mathcal{R}_\nu$ contains the assertion $P(c_\ell, v_\ell)$, and in the latter case, $\mathcal{R}_\nu$ contains $N(c_\ell, v_\ell)$. In both cases, there is an assertion in $\mathcal{R}_\nu$ that conflicts with $U(a, c_\ell)$, so the latter assertion cannot appear in $\mathcal{R}_\nu$. We have thus shown that $\mathcal{R}_\nu$ does not contain any of the assertions in $\{U(a, c_{i_1}), ..., U(a, c_{i_k})\}$.

Next suppose that there is a repair $\mathcal{R}$ that have an empty intersection with $\{U(a, c_{i_1}), ..., U(a, c_{i_k})\}$. By the maximality of $\mathcal{R}$, it follows that for every $\ell \in \{i_1, \ldots, i_k\}$, there must exist an assertion in $\mathcal{R}$ of the form $P(c_\ell, v_j)$ or $N(c_\ell, v_j)$. Define a (possibly partial) assignment $\nu_\mathcal{R}$ by setting by X_j to true if $\mathcal{R}$ contains some $P(c_i, x_j)$ and to false if $\mathcal{R}$ contains some $N(c_i, x_j)$ (recall that $\mathcal{R}$ is consistent with $\mathcal{T}_0$, and so it cannot contain both $P(c_i, x_j)$ and $N(c_k, x_j)$). By construction, $\nu_\mathcal{R}$ satisfies all of the clauses in $\{C_{i_1}, ..., C_{i_k}\}$, i.e. $\{C_{i_1}, ..., C_{i_k}\}$ is satisfiable. (*end proof of claim*)

It follows from the preceding claim that the explanations for $\mathcal{T}_0, \mathcal{A}_0 \models_{\text{AR}} q_0(a)$, i.e., the minimal sets of causes for $q_0(a)$ that cover all repairs, correspond precisely to the MUSes of φ_0. We can therefore exploit known complexity results for MUSes ([23]):

– Deciding if a clause belongs to a MUS is Σ_2^p-complete, so deciding if $U(a, c_i)$ belongs to an explanation for $\mathcal{T}_0, \mathcal{A}_0 \models_{\text{AR}} q_0(a)$ is Σ_2^p-hard w.r.t. data complexity.
– Deciding if a clause belongs to every MUS is NP-complete, so deciding if $U(a, c_i)$ belongs to every explanation for $\mathcal{T}_0, \mathcal{A}_0 \models_{\text{AR}} q_0(a)$ is NP-hard w.r.t. data complexity.
– Deciding if a set of clauses is a MUS is BH_2-complete, so deciding if $\{\{U(a, c_{i_1})\}, ..., \{U(a, c_{i_k})\}\}$ is an explanation is BH_2-hard w.r.t. data complexity.

The proof of [23] for Σ_2^p hardness of deciding if there exists a MUS of size at most k also shows that deciding if a set of clause is a smallest MUS is Π_2^p-hard. It follows that deciding if an explanation for an AR-answer contains a smallest number of causes is Π_2^p-complete. Moreover, since every cause in the considered KB consists of a single assertion, deciding if an explanation for an AR-answer contains a smallest number of assertions is also Π_2^p-complete.

Proposition 5. *Regarding explanations for negative AR-answers,* NEC *is co*NP*-hard,* REL *is* NP*-hard, and* BEST REC *is co*NP*-hard (for explanations of smallest cardinality) w.r.t. data complexity.*

Proof. All reductions are from SAT. Let $\varphi = C_1 \wedge ... \wedge C_n$ be a set of clauses over propositional variables $\{X_1, ..., X_p\}$.

• NEC: Let $\mathcal{T}_0$, $\mathcal{A}_0$, and q_0 be as in Proposition 4. Define a new TBox $\mathcal{T}_1 = \mathcal{T}_0 \cup \{\exists U \sqsubseteq \neg S\}$ and ABox $\mathcal{A}_1 = \mathcal{A}_0 \cup \{S(a)\}$. By construction, the assertion $S(a)$ contradicts every cause for $q_0(a)$, so $\mathcal{T}_1, \mathcal{A}_1 \not\models_{\text{AR}} q_0(a)$. We show that deciding whether φ is satisfiable is equivalent to deciding if $S(a)$ is *not* necessary for explaining $\mathcal{T}_1, \mathcal{A}_1 \not\models_{\text{AR}} q_0(a)$. This establishes the coNP-hardness of checking necessity.

$\Rightarrow$ Let ν be a satisfying valuation for φ. It can be easily verified that the set $\{P(c_i, v_j) \in \mathcal{A}_0 \mid \nu(v_j) = \mathsf{true}\} \cup \{N(c_i, v_j) \in \mathcal{A}_0 \mid \nu(v_j) = \mathsf{false}\}$ conflicts with every cause of $q_0(a)$, and so by choosing a subset of these assertions, we can construct an explanation for $\mathcal{T}_1, \mathcal{A}_1 \not\models_{\text{AR}} q_0(a)$ that does not contain $S(a)$.

$\Leftarrow$ An explanation $\mathcal{E}$ that does not contain $S(a)$ forms a $\mathcal{T}_1$-consistent set of P- and N-assertions such that every c_i has an outgoing P- or N-edge. We obtain a (partial) assignment $\nu_{\mathcal{E}}$ that satisfies φ by setting $\nu_{\mathcal{E}}(v_j) = \mathsf{true}$ if $\mathcal{E}$ contains an assertion $P(c_i, v_j)$ and $\nu_{\mathcal{E}}(v_j) = \mathsf{false}$ if $\mathcal{E}$ contains an assertion $N(c_i, v_j)$.

• REL: We use the TBox $\mathcal{T}_1$ and the ABox $\mathcal{A}_2 = \mathcal{A}_1 \cup \{U(a, c_{n+1}), P(c_{n+1}, x_{p+1})\}$. Again, we note that $S(a)$ contradicts every cause for $q_0(a)$, so $\mathcal{T}_1, \mathcal{A}_2 \not\models_{\text{AR}} q_0(a)$. We show that φ is satisfiable iff $P(c_{n+1}, x_{p+1})$ is relevant for explaining $\mathcal{T}_1, \mathcal{A}_2 \not\models_{\text{AR}} q_0(a)$; it follows that relevance is NP-hard.

$\Rightarrow$ If φ is satisfiable, then we can obtain an explanation for $\mathcal{T}_1, \mathcal{A}_2 \not\models_{\text{AR}} q_0(a)$ by adding $P(c_{n+1}, x_{p+1})$ to a minimal subset of the P- and N-assertions corresponding to a satisfying truth assignment for φ.

$\Leftarrow$ If φ is unsatisfiable, then every explanation must contain $S(a)$. It follows that $\{S(a)\}$ is the only explanation, so $P(c_{n+1}, x_{p+1})$ is not relevant.

• BEST REC: We consider the following KB:

$$\begin{aligned}
\mathcal{T}_3 =& \mathcal{T}_0 \cup \{U_1 \sqsubseteq U, U_2 \sqsubseteq U, \exists U_1^- \sqsubseteq \neg T, \exists U_2 \sqsubseteq \neg S\} \\
\mathcal{A}_3 =& \{P(c_i, x_j) \mid X_j \in C_i\} \cup \{N(c_i, x_j) \mid \neg X_j \in C_i\} \cup \\
& \{U_1(a, c_i), U_2(a, c_i), T(c_i) \mid 1 \leq i \leq n\} \cup \{S(a)\}
\end{aligned}$$

We claim that $\mathcal{E} = \{S(a), T(c_1), ..., T(c_m)\}$ is a smallest explanation for $\mathcal{T}_3, \mathcal{A}_3 \not\models_{\text{AR}} q_0(a)$ iff φ is unsatisfiable.

$\Rightarrow$ If φ is satisfiable, then we can use a satisfying truth assignment to define a consistent set of m P- and N-edges such that every c_i has an outgoing edge. This set is an explanation for $\mathcal{T}_3, \mathcal{A}_3 \not\models_{\text{AR}} q_0(a)$, and it has fewer assertions than $\mathcal{E}$.

$\Leftarrow$ If there exists an explanation of size at most m, it contains necessarily only P- and N-edges, since m assertions (P, N or T) are needed to conflict all U_1, and $S(a)$ is needed as soon as one of the U_1-assertions is conflicted only by a T-assertion. It follows that there exists a consistent set of P- and N-assertions such that every c_i has an outgoing edge, from which we can construct a satisfying assignment for φ.

Lemma 3. *Deciding if a truth assignment that satisfies a monotone 3-SAT formula assigns a smallest number of variables to true is co*NP*-hard.*

Proof. The proof is by reduction from the following NP-hard problem: given a set $\varphi = \{C_1, \ldots, C_n\}$ of monotone 3-clauses over variables $\{X_1, ..., X_p\}$, is there a satisfying truth assignment for φ that has at most k variables assigned to true?

Consider the set of clauses $\varphi' = \{C_i \vee Y_j \mid 1 \leq i \leq p, 1 \leq j \leq k+1\}$ where $Y_1, ..., Y_{k+1}$ are fresh variables. We claim that there is a satisfying truth assignment for φ that assigns at most k variables to true iff the following truth assignment ν' is not minimal for φ': $\nu'(X_i) = \mathsf{false}$ $(1 \leq i \leq p)$ and $\nu'(Y_j) = \mathsf{true}$ $(1 \leq j \leq k{+}1)$. Indeed, if there exists a truth assignment ν of the X_i that satisfies φ and contains at most k variables assigned to true, the truth assignment that extends ν by assigning every Y_j to false satisfies φ', since it makes every C_l true. It follows that ν' is not minimal. For the other direction, if ν' is not minimal, there exists a truth assignment of the X_i and Y_j that satisfies φ' and assigns at most k variables to true. The valuation ν' makes $C_1, ..., C_n$ true, otherwise it should satisfy $Y_1 \wedge ... \wedge Y_{k+1}$, which is not possible with only k variables assigned to true. It follows that there exists a truth assignment of $\{X_1, ..., X_p\}$ that satisfies φ and contains at most k variables assigned to true.

Proposition 6. *Regarding explanations for negative IAR-answers,* BEST REC *is co*NP*-hard (for explanations of lowest cardinality) w.r.t. data complexity.*

Proof. We use a reduction from the coNP-hard problem of deciding if a truth assignment that satisfies a monotone 3-SAT formula assigns a smallest number of variables to true (Lemma 3).

Let $\varphi = C_1 \wedge ... \wedge C_n$ be a monotone 3-CNF over the variables $\{X_1, ..., X_p\}$, and let ν be a truth assignment that satisfies φ. Consider the following KB:

$$\mathcal{T} = \{\exists P_k^- \sqsubseteq \neg T \mid 1 \leq k \leq 3\}$$
$$\mathcal{A} = \{T(x_i) \mid 1 \leq i \leq p\} \cup \{P_k(c_j, x_i) \mid X_i \ k^{th} \text{ term of } C_j\}$$
$$q = \exists y z_1 z_2 z_3\, P_1(y, z_1) \wedge P_2(y, z_2) \wedge P_3(y, z_3)$$

The causes for $q(a)$ take the form $\{P_1(c_j, x_{i_1}), P_2(c_j, x_{i_2}), P_3(c_j, x_{i_3})\}$. It follows that an explanation for $\mathcal{T}, \mathcal{A} \not\models_{\mathrm{IAR}} q$ is a set $\mathcal{E}$ of T-assertions such that for every c_j, there is at least one $X_i \in C_j$ such that $T(x_i) \in \mathcal{E}$.

Deciding if ν assigns a minimal number of variables to true is equivalent to deciding if $\mathcal{E} = \{T(x_i) \mid \nu(X_i) = \mathsf{true}\}$ is a smallest explanation.