

New inconsistency-tolerant semantics for robust ontology-based data access

Meghyn Bienvenu¹ and Riccardo Rosati²

¹ Laboratoire de Recherche en Informatique
CNRS & Université Paris-Sud, France

² Dipartimento di Ingegneria Informatica, Automatica e Gestionale
Sapienza Università di Roma, Italy

1 Introduction

In ontology-based data access (OBDA) [17], an ontology provides an abstract and formal representation of the domain of interest, which is used as a virtual schema when formulating queries over the data. Current research in OBDA mostly focuses on ontology specification languages for which conjunctive query answering is *first-order (FO) rewritable*. In a nutshell, FO-rewritability means that query answering can be performed by rewriting the input query into a first-order query which encodes the relevant knowledge from the ontology, and then evaluating the resulting query over the data. Among FO-rewritable ontology languages, description logics (DLs) of the *DL-Lite* family [8, 2] have played an especially prominent role and notably served as the inspiration for the OWL 2 QL profile ³ of the latest version of the OWL web ontology language.

In real-world applications involving large amounts of data and/or multiple data sources, chances are that the data will be inconsistent with the ontology. Standard OBDA querying algorithms are next to useless in such circumstances, since first-order logic semantics (upon which DLs and standard ontology languages are based) dictates that everything can be derived from a contradiction. Appropriate mechanisms for handling inconsistent data are thus critical to the success of OBDA in practice. Clearly, the best solution is to restore consistency by removing the pieces of data that are responsible for the inconsistencies. However, this strategy cannot always be applied, since the system may not have enough information to localize the errors, or may lack the authorization to modify the data (as is often the case in information integration applications). Thus, a robust OBDA system must be capable of providing meaningful answers to user queries in the presence of inconsistent data.

Recently, several approaches have pursued the idea of adopting an inconsistency-tolerant semantics for OBDA, taking inspiration from the work on *consistent query answering* in databases [1, 4]. The most well-known and intuitive among such semantics, which we will call the *CQA semantics*, considers as a *repair* of a knowledge base (KB) consisting of an ontology \mathcal{T} and a dataset \mathcal{A} , a maximal subset of \mathcal{A} that is consistent with \mathcal{T} . Query answering under the CQA semantics then amounts to computing those answers that hold in every repair of the KB. Unfortunately, conjunctive query answering (as well as simpler forms of reasoning) under CQA semantics is computationally

³ <http://www.w3.org/TR/owl2-profiles/>

hard, even for extremely simple ontology languages for which reasoning under classical semantics is tractable [13, 5].

To overcome this computational problem, approximations of the CQA semantics have been recently proposed. In particular, [13, 14] introduces a sound approximation (called IAR semantics) that evaluates queries over the intersection of all the repairs of the CQA semantics. It was shown that conjunctive query answering under this semantics is tractable (in particular, it is first-order rewritable) for logics of the *DL-Lite* family. However, the IAR semantics has the drawback that it often constitutes a very rough approximation of the CQA semantics, and desirable query answers may be missed. In an effort to obtain more answers than the IAR semantics, a family of parameterized inconsistency-tolerant semantics, called *k-lazy consistent semantics*, was proposed in [15] and shown to converge in the limit to the CQA semantics. However, since the convergence is not monotone in *k*, these semantics are not sound approximations of the CQA semantics. Moreover, these semantics do not retain the nice computational properties of the IAR semantics: the polynomial data complexity result shown for linear Datalog \pm ontologies only holds for atomic queries, and it follows from results in [5] that conjunctive query answering under *k-lazy consistent semantics* is coNP-hard in data complexity, for every $k \geq 1$.

In this paper, we address the above issues and provide the following contributions:

(i) we propose two new families of inconsistency-tolerant semantics, called *k-defeater* and *k-support* semantics, that approximate the CQA semantics from above (complete approximations) and from below (sound approximations), respectively, and converge to the CQA semantics in the limit;

(ii) we study the data complexity of conjunctive query answering under the new semantics, and show a general tractability result for a broad class of ontology languages that includes all known first-order rewritable languages, in particular almost all DLs of the *DL-Lite* family and several rule-based languages of the Datalog \pm family [6];

(iii) we analyze the combined complexity of conjunctive query answering under the above semantics for ontology languages of the *DL-Lite* family.

The *k-support* and *k-defeater* semantics proposed in this paper provide the basis for a semantically grounded and computationally tractable approximation of the CQA semantics in OBDA systems. In particular, we envision a flexible, iterated execution of query *q* under both *k-support* and *k-defeater* semantics with increasing values of *k*, which stops as soon as the answers to *q* under both semantics coincide, or when the user is not interested in (or does not want to pay further computational cost for) an exact classification of the tuples that are answers to *q* under the CQA semantics.

2 Preliminaries

Ontologies and KBs An *ontology* \mathcal{T} is a finite set of first-order logic sentences, and an *ontology (specification) language* \mathcal{L} is a (typically infinite) set of first-order logic sentences. If $\mathcal{T} \subseteq \mathcal{L}$, then \mathcal{T} is called an \mathcal{L} *ontology*. A *knowledge base (KB)* is a pair consisting of an ontology \mathcal{T} and a finite set \mathcal{A} of ground facts. A KB $\langle \mathcal{T}, \mathcal{A} \rangle$ is said to be *consistent* if the first-order theory $\mathcal{T} \cup \mathcal{A}$ has a model. Otherwise, it is *inconsistent*, which we denote by $\langle \mathcal{T}, \mathcal{A} \rangle \models \perp$.

We are interested in the problem of answering instance queries and conjunctive queries over KBs. Without loss of generality, and for ease of exposition, we only consider Boolean queries (i.e. queries without free variables). A first-order (FO) query, or simply *query*, is a first-order sentence. An *instance query (IQ)* is a FO query consisting of a single ground fact. A *conjunctive query (CQ)* is a FO query of the form $\exists \mathbf{x}(\alpha_1 \wedge \dots \wedge \alpha_n)$ where every α_i is an atom whose arguments are either constants or variables from \mathbf{x} . A query q is *entailed* by a KB \mathcal{K} under classical semantics (denoted by $\mathcal{K} \models q$) if q is satisfied in every model of \mathcal{K} . The *instance checking problem* consists in deciding, for a KB \mathcal{K} and IQ q , whether $\mathcal{K} \models q$. The *conjunctive query entailment problem* is defined analogously, but with q a CQ.

We introduce some terminology for referring to sets of facts which are responsible for inconsistency or query entailment. A set S of ground facts is called \mathcal{T} -consistent if $\langle \mathcal{T}, S \rangle \not\models \perp$. A *minimal \mathcal{T} -inconsistent subset* of \mathcal{A} is any $S \subseteq \mathcal{A}$ such that $\langle \mathcal{T}, S \rangle \models \perp$ and every $S' \subsetneq S$ is \mathcal{T} -consistent. A set of facts $S \subseteq \mathcal{A}$ is said to be a \mathcal{T} -support for query q in \mathcal{A} if S is \mathcal{T} -consistent and $\langle \mathcal{T}, S \rangle \models q$, and it is called a *minimal \mathcal{T} -support* for q in \mathcal{A} if no proper subset of S is a \mathcal{T} -support for q in \mathcal{A} . We sometimes omit “for q ” or “in \mathcal{A} ”, when these are understood.

Given a set of ground facts \mathcal{A} , we define $\mathcal{I}_{\mathcal{A}}$ as the interpretation isomorphic to \mathcal{A} , i.e., the interpretation defined over the domain of constants occurring in \mathcal{A} and such that the interpretation of every relation R in $\mathcal{I}_{\mathcal{A}}$ is equal to the set $\{\mathbf{a} \mid R(\mathbf{a}) \in \mathcal{A}\}$.

DL-Lite ontology languages We focus on DLs of the *DL-Lite* family [8, 2] and recall the syntax and semantics of two specific dialects, called *DL-Lite*⁴ and *DL-Lite_{Horn}*. A *DL-Lite* ontology consists of a finite set of inclusions $B \sqsubseteq C$, where B and C are defined according to the following syntax:

$$B \rightarrow A \mid \exists R \quad C \rightarrow B \mid \neg B \quad R \rightarrow P \mid P^-$$

with A a concept name (unary relation) and P a role name (binary relation). In a *DL-Lite_{Horn}* ontology, inclusions take the form $B_1 \sqcap \dots \sqcap B_n \sqsubseteq C$, with B_1, \dots, B_n and C as above.

The classical semantics of *DL-Lite* and *DL-Lite_{Horn}* ontologies is obtained by translating inclusions into first-order sentences using the following function Φ :

$$\begin{aligned} \Phi(A(x)) &= A(x) & \Phi(\neg B(x)) &= \neg \Phi(B(x)) \\ \Phi(\exists P(x)) &= \exists y(P(x, y)) & \Phi(B_1 \sqcap B_2(x)) &= \Phi(B_1(x)) \wedge \Phi(B_2(x)) \\ \Phi(\exists P^-(x)) &= \exists y(P(y, x)) & \Phi(C \sqsubseteq D) &= \forall x(\Phi(C(x)) \rightarrow \Phi(D(x))) \end{aligned}$$

The classical semantics of a *DL-Lite_{Horn}* KB $\langle \mathcal{T}, \mathcal{A} \rangle$ (and in particular, the notions of model, consistency, and entailment) corresponds to the semantics of the first-order KB $\langle \Phi(\mathcal{T}), \mathcal{A} \rangle$. Note that when considering DL KBs, we assume as is standard that the dataset \mathcal{A} uses only unary and binary relations.

First-order rewritability We say that an ontology \mathcal{T} is *first-order (FO) rewritable* (for CQ answering) under semantics \mathcal{S} if, for every CQ q , there exists an effectively computable FO query q' such that, for every set of ground facts \mathcal{A} , $\langle \mathcal{T}, \mathcal{A} \rangle$ entails q

⁴ This DL is referred to as *DL-Lite_{core}* in [8, 2].

under semantics S iff q' is satisfied in $\mathcal{I}_{\mathcal{A}}$ (in the classical sense). Such a query q' is called a *FO-rewriting* of q relative to \mathcal{T} under semantics S . Moreover, we say that an ontology language \mathcal{L} is *FO-rewritable (for CQ answering)* under semantics S if, for every ontology $\mathcal{T} \subseteq \mathcal{L}$, \mathcal{T} is FO-rewritable for CQ answering under S .

Complexity There are two common ways of measuring the complexity of query entailment. The first, called *combined complexity*, is with respect to the size of the whole input $(\mathcal{T}, \mathcal{A}, q)$, whereas the second, called *data complexity*, is only with respect to the size of \mathcal{A} . Our complexity results utilize standard complexity classes, such as NLSPACE, P, NP, and coNP. We also require the following classes which may be less well-known: AC^0 (problems which can be solved by a family of circuits of constant depth and polynomial size, with unlimited fan-in AND gates and OR gates), Π_2^P (problems whose complement is solvable in non-deterministic polynomial time with access to an NP oracle), and $\Delta_2^P[O(\log n)]$ (problems which are solvable in polynomial time with at most logarithmically many calls to an NP oracle).

3 Inconsistency-tolerant Semantics

In this section, we formally introduce the consistent query answering (CQA) semantics and other relevant inconsistency-tolerant semantics. All of the semantics considered in this paper rely on the notion of a repair, defined as follows:

Definition 1. A repair of a KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ is an inclusion-maximal subset of \mathcal{A} that is \mathcal{T} -consistent. We use $Rep(\mathcal{K})$ to denote the set of repairs of \mathcal{K} .

The repairs of a KB correspond to the different ways of achieving consistency while retaining as much of the original data as possible. Hence, if we consider that the data is mostly reliable, then it is reasonable to assume that one of the repairs accurately reflects the correct portion of the data.

The consistent query answering semantics (also known as the AR semantics [13]) is based upon the idea that, in the absence of further information, a query can be considered to hold if it can be inferred from each of the repairs. Formally:

Definition 2. A query q is entailed by a KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ under the consistent query answering (CQA) semantics, written $\langle \mathcal{T}, \mathcal{A} \rangle \models_{CQA} q$, if $\langle \mathcal{T}, \mathcal{B} \rangle \models q$ for every repair $\mathcal{B} \in Rep(\mathcal{K})$.

Example 1. Consider the DL-Lite ontology \mathcal{T}_{univ} :

$$\begin{array}{l} \text{Prof} \sqsubseteq \text{Faculty} \quad \text{Lect} \sqsubseteq \text{Faculty} \quad \text{Fellow} \sqsubseteq \text{Faculty} \quad \text{Prof} \sqsubseteq \neg \text{Lect} \quad \text{Prof} \sqsubseteq \neg \text{Fellow} \\ \text{Lect} \sqsubseteq \neg \text{Fellow} \quad \text{Prof} \sqsubseteq \exists \text{teaches} \quad \text{Lect} \sqsubseteq \exists \text{teaches} \quad \exists \text{teaches}^- \sqsubseteq \neg \text{Faculty} \end{array}$$

which states that professors, lecturers, and research fellows are disjoint classes of faculty, that professors and lecturers must teach something, and that whatever is taught is not faculty. Now let \mathcal{A}_{sam} be as follows: $\{\text{Prof}(\text{sam}), \text{Lect}(\text{sam}), \text{Fellow}(\text{sam})\}$. It is easy to see that KB $\langle \mathcal{T}_{univ}, \mathcal{A}_{sam} \rangle$ is inconsistent and has 3 repairs: $\mathcal{R}_1 = \{\text{Prof}(\text{sam})\}$, $\mathcal{R}_2 = \{\text{Lect}(\text{sam})\}$ and $\mathcal{R}_3 = \{\text{Fellow}(\text{sam})\}$. Observe that from each of the repairs, we can infer $q_1 = \text{Faculty}(\text{sam})$, so $\langle \mathcal{T}_{univ}, \mathcal{A}_{sam} \rangle \models_{CQA} q_1$. However, $q_2 =$

$\exists x.\text{Faculty}(\text{sam}) \wedge \text{teaches}(\text{sam}, x)$ is not entailed from $\langle \mathcal{T}_{\text{univ}}, \mathcal{R}_3 \rangle$, so we have that $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \not\models_{\text{CQA}} q_2$.

Unfortunately, while the CQA semantics is intuitively appealing, it is well-known that answering queries under this semantics is usually intractable w.r.t. data complexity [13, 5]. This stems from the fact that the number of repairs of $\langle \mathcal{T}, \mathcal{A} \rangle$ may be exponential in the size of \mathcal{A} , even when \mathcal{T} is formulated in extremely simple ontology languages.

To overcome the computational problems of the CQA semantics, a sound approximation of it, called the IAR semantics, was proposed in [13].

Definition 3. A query q is entailed by a KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ under the IAR semantics, written $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{IAR}} q$, if $\langle \mathcal{T}, \mathcal{D} \rangle \models q$ where $\mathcal{D} = \bigcap_{\mathcal{B} \in \text{Rep}(\mathcal{K})} \mathcal{B}$.

The IAR semantics is more conservative than the CQA semantics, as it only uses those facts which are not involved in any contradiction. This has the advantage of yielding query results which are almost surely correct, but also the drawback that some plausible inferences may be missed, as demonstrated by the following example.

Example 2. Reconsider the KB $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle$ and CQ q_1 from Example 1. The intersection of the repairs $\mathcal{R}_1 \cap \mathcal{R}_2 \cap \mathcal{R}_3$ is the empty set, so $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \not\models_{\text{IAR}} q_1$, despite the fact that all the information in \mathcal{A}_{sam} supports q_1 being true.

From the computational perspective, the IAR semantics can be much better-behaved than the CQA semantics. Indeed, it was shown in [14] that $DL\text{-Lite}_A$ is FO-rewritable for CQ answering under the IAR semantics, and this result was recently extended to linear Datalog +/- ontologies [16].

Finally, to obtain a natural overapproximation of the CQA semantics, we introduce its brave version.

Definition 4. A query q is entailed by a KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ under the brave semantics, written $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{brave}} q$, if $\langle \mathcal{T}, \mathcal{B} \rangle \models q$ for some repair $\mathcal{B} \in \text{Rep}(\mathcal{K})$.

Example 3. As q_2 is entailed by $\langle \mathcal{T}_{\text{univ}}, \mathcal{R}_1 \rangle$, we have $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \models_{\text{brave}} q_2$. Also note that every fact in \mathcal{A}_{sam} appear in some repair, hence, all facts in \mathcal{A}_{sam} are entailed under the brave semantics.

As Example 3 demonstrates, the brave semantics has the undesirable feature of allowing contradictory statements to be entailed. Nonetheless, this semantics can still serve a useful role by providing a means of showing that a query is *not entailed* under the CQA semantics.

4 Approximations of the CQA Semantics

In this section, we propose two new families of inconsistency-tolerant semantics, which provide increasingly fine-grained under- and over-approximations of the CQA semantics. As these semantics will be shown in Section 5 to enjoy the same nice computational properties as the IAR semantics, our new approach allows us to marry the advantages of the IAR and CQA semantics.

We begin by presenting our new family of sound approximations of the CQA semantics. The intuition is as follows: if a query q is entailed under the CQA semantics, then this is because there is a set $\{S_1, \dots, S_n\}$ of \mathcal{T} -supports for q such that every repair contains some S_i . The k -support semantics we propose is obtained by allowing a maximum of k different supports to be used.

Definition 5. A query q is entailed by $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ under the k -support semantics, written $\mathcal{K} \models_{k\text{-supp}} q$, if there exist (not necessarily distinct) subsets S_1, \dots, S_k of \mathcal{A} satisfying the following conditions:

- each S_i is a \mathcal{T} -support for q in \mathcal{A}
- for every $R \in \text{Rep}(\mathcal{K})$, there is some S_i with $S_i \subseteq R$

Example 4. The three repairs of $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle$ all use different supports for q_1 . We thus have $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \models_{3\text{-supp}} q_1$, but $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \not\models_{2\text{-supp}} q_1$.

The following theorem resumes the important properties of the family of k -support semantics, showing that they interpolate between the IAR and CQA semantics.

Theorem 1. Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a KB and q a query. Then:

1. $\mathcal{K} \models_{\text{IAR}} q$ if and only if $\mathcal{K} \models_{1\text{-supp}} q$
2. $\mathcal{K} \models_{\text{CQA}} q$ if and only if $\mathcal{K} \models_{k\text{-supp}} q$ for some k
3. for every $k \geq 0$, if $\mathcal{K} \models_{k\text{-supp}} q$, then $\mathcal{K} \models_{k+1\text{-supp}} q$

The k -support semantics allows us to approximate more and more closely the set of queries entailed under the CQA semantics, but provides no way of showing that a particular query is *not entailed* under this semantics. This motivates the study of complete approximations of the CQA semantics.

The observation underlying our new family of complete approximations is the following: if a query q is not entailed under the CQA semantics, this is because there is a \mathcal{T} -consistent set of facts which contradicts all of the \mathcal{T} -supports of q . The k -defeater semantics corresponds to there being no way to construct such a “defeating” set using at most k facts.

Definition 6. A query q is entailed by $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ under the k -defeater semantics, written $\mathcal{K} \models_{k\text{-def}} q$, if there does not exist a \mathcal{T} -consistent subset S of \mathcal{A} with $|S| \leq k$ such that $\langle \mathcal{T}, S \cup C \rangle \models \perp$ for every minimal \mathcal{T} -support $C \subseteq \mathcal{A}$ of q .

Note that if q has no \mathcal{T} -support, then it is not entailed under 0-defeater semantics since one can simply take $S = \emptyset$.

Example 5. We have $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \not\models_{1\text{-def}} q_2$, since by choosing $S = \{\text{Fellow}(\text{sam})\}$, we can invalidate the minimal \mathcal{T} -supports of q_2 , which are $\{\text{Prof}(\text{sam})\}$ and $\{\text{Lect}(\text{sam})\}$.

The next theorem shows that the family of k -defeater semantics provides increasingly closer over-approximations of the CQA semantics, starting from the brave semantics presented in Section 3.

Theorem 2. Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a KB and q a query. Then:

1. $\mathcal{K} \models_{\text{brave}} q$ if and only if $\mathcal{K} \models_{0\text{-def}} q$
2. $\mathcal{K} \models_{\text{CQA}} q$ if and only if $\mathcal{K} \models_{k\text{-def}} q$ for every k
3. for every $k \geq 1$, if $\mathcal{K} \models_{k+1\text{-def}} q$, then $\mathcal{K} \models_{k\text{-def}} q$

5 Data Complexity

In this section, we study the data complexity of conjunctive query answering under the k -support and k -defeater semantics. Our main result is the following theorem which shows that, for a broad class of ontology languages, conjunctive query answering under these semantics can be done using FO-rewriting, and hence is in AC^0 w.r.t. data complexity.

Theorem 3. *Let \mathcal{T} be an ontology that is FO-rewritable for CQ answering under classical semantics and such that for every CQ q , there exist ℓ, m such that for every \mathcal{A} , every minimal \mathcal{T} -support for q relative to \mathcal{A} has cardinality at most ℓ , and every minimal \mathcal{T} -inconsistent subset of \mathcal{A} has cardinality at most m . Then:*

- (i) *for every $k \geq 1$, \mathcal{T} is FO-rewritable for conjunctive query answering under the k -support semantics;*
- (ii) *for every $k \geq 0$, \mathcal{T} is FO-rewritable for conjunctive query answering under the k -defeater semantics.*

Proof (sketch). Let \mathcal{T} be as stated, and let q be a CQ. By assumption, we can find ℓ and m such that for every \mathcal{A} , the minimal \mathcal{T} -supports for q relative to \mathcal{A} have cardinality at most ℓ , and the minimal \mathcal{T} -inconsistent subsets of \mathcal{A} have cardinality bounded by m . For point (i), a FO-rewriting of q relative to \mathcal{T} for the k -support semantics can be obtained by considering the first-order query $\varphi_q = q_1 \vee \dots \vee q_n$, where the disjuncts q_i correspond to the different possible choices of k \mathcal{T} -supports for q of cardinality at most ℓ , and each q_i asserts that the chosen supports are present in \mathcal{A} and that there is no \mathcal{T} -consistent subset of \mathcal{A} of cardinality at most km which conflicts with each of the supports. For point (ii), the desired FO-rewriting of q takes the form $\neg(q_1 \vee \dots \vee q_n)$, where every q_i asserts the existence of a \mathcal{T} -consistent set of facts of cardinality at most k which conflicts with every minimal \mathcal{T} -support for q . Here we again utilize the fact that the size of minimal \mathcal{T} -supports is bounded by ℓ , and hence there are only finitely many types of supports to consider.

Theorem 3 significantly strengthens earlier positive results for the IAR semantics [14, 15] by covering a full range of semantics and an entire class of practically relevant ontology languages. Indeed, it is easy to verify that *all* ontology languages that are currently known to be first-order rewritable under classical semantics satisfy the hypotheses of Theorem 3: that is, all logics of the original *DL-Lite* family [8] and almost all members of the extended *DL-Lite* family [2], as well as all dialects of Datalog+/- that are known to be FO-rewritable under classical semantics [7].

The following examples illustrate the construction of FO-rewritings for the k -support and k -defeater semantics.

Example 6. We consider how to rewrite the CQ q_1 under the k -support semantics. When $k = 1$, we can take as our FO-rewriting the disjunction of the formulas:

$$\begin{aligned}
 & \text{Faculty}(\text{sam}) \wedge \neg \exists x \text{ teaches}(x, \text{sam}) \\
 & \text{Prof}(\text{sam}) \wedge \neg \exists x \text{ teaches}(x, \text{sam}) \wedge \neg \text{Lect}(\text{sam}) \wedge \neg \text{Fellow}(\text{sam}) \\
 & \text{Lect}(\text{sam}) \wedge \neg \exists x \text{ teaches}(x, \text{sam}) \wedge \neg \text{Prof}(\text{sam}) \wedge \neg \text{Fellow}(\text{sam}) \\
 & \text{Fellow}(\text{sam}) \wedge \neg \exists x \text{ teaches}(x, \text{sam}) \wedge \neg \text{Lect}(\text{sam}) \wedge \neg \text{Prof}(\text{sam})
 \end{aligned}$$

Note that each disjunct expresses that one of the four possible \mathcal{T} -supports is present and is not contradicted by other facts. To obtain the rewriting for $k = 2$, we must introduce additional disjuncts which assert that a pair of \mathcal{T} -supports is present and cannot be simultaneously contradicted. We obtain three new disjuncts (the other combinations being subsumed by one of the other disjuncts):

$$\begin{aligned} & \text{Prof}(\text{sam}) \wedge \text{Lect}(\text{sam}) \wedge \neg \exists x \text{teaches}(x, \text{sam}) \wedge \neg \text{Fellow}(\text{sam}) \\ & \text{Lect}(\text{sam}) \wedge \text{Fellow}(\text{sam}) \wedge \neg \exists x \text{teaches}(x, \text{sam}) \wedge \neg \text{Prof}(\text{sam}) \\ & \text{Fellow}(\text{sam}) \wedge \text{Prof}(\text{sam}) \wedge \neg \exists x \text{teaches}(x, \text{sam}) \wedge \neg \text{Lect}(\text{sam}) \end{aligned}$$

Finally, for $k = 3$, we must add further disjuncts to check for the existence of a triple of \mathcal{T} -supports which are present and cannot be defeated. In our case, this leads to one new (non-subsumed) disjunct:

$$\text{Prof}(\text{sam}) \wedge \text{Lect}(\text{sam}) \wedge \text{Fellow}(\text{sam}) \wedge \neg \exists x \text{teaches}(x, \text{sam})$$

Note that this last disjunct is satisfied in $\mathcal{I}_{\mathcal{A}_{\text{sam}}}$, witnessing $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \models_{3\text{-supp}} q_1$. Notice also that in this particular example, the CQA and 3-support semantics coincide, and so the FO-rewriting for $k = 3$ is also a FO-rewriting under the CQA semantics.

Example 7. We now consider how to rewrite the query q_2 under the k -defeater semantics. When $k = 0$, the construction yields the following FO-rewriting:

$$\begin{aligned} & \neg(\neg(\exists x \text{Faculty}(\text{sam}) \wedge \text{teaches}(\text{sam}, x)) \wedge \neg \text{Prof}(\text{sam}) \\ & \quad \wedge \neg \text{Lect}(\text{sam}) \wedge \neg(\exists x \text{Fellow}(\text{sam}) \wedge \text{teaches}(\text{sam}, x))) \end{aligned}$$

Inside the negation, there is a single disjunct which asserts that the empty set conflicts with every \mathcal{T} -support, or equivalently, that there are no \mathcal{T} -supports. When $k = 1$, we must add further disjuncts inside the negation to capture single facts which conflict with all \mathcal{T} -supports. In our case, we must add two new disjuncts:

$$\exists x \text{teaches}(x, \text{sam}) \quad \text{Fellow}(\text{sam}) \wedge \neg \text{teaches}(\text{sam}, x)$$

The first disjunct is needed since any fact of the form $\text{teaches}(x, \text{sam})$ contradicts $\text{Faculty}(\text{sam})$, and hence, every \mathcal{T} -support for q_2 . The second disjunct treats the case where there is no atom $\text{teaches}(x, \text{sam})$, in which case the only possible \mathcal{T} -supports for q_2 are $\text{Prof}(\text{sam})$ and $\text{Lect}(\text{sam})$, both of which are contradicted by $\text{Fellow}(\text{sam})$. Notice that this last disjunct holds in $\mathcal{I}_{\mathcal{A}_{\text{sam}}}$, which proves that $\langle \mathcal{T}_{\text{univ}}, \mathcal{A}_{\text{sam}} \rangle \not\models q_2$.

We briefly remark that polynomial data complexity is not preserved under the new semantics. Indeed, in the lightweight DL \mathcal{EL}_{\perp} , CQ answering and unsatisfiability are P-complete w.r.t. data complexity, but it was shown in [18] that instance checking under the IAR (equiv. 1-support) semantics is coNP-hard w.r.t. data complexity, and it is not hard to show intractability also for the brave (equiv. 0-defeater) semantics.

6 Combined Complexity

To gain further insight into the computational properties of the different inconsistency-tolerant semantics considered in this paper, we study the combined complexity of instance checking and CQ entailment for $DL\text{-Lite}$ and $DL\text{-Lite}_{\text{Horn}}$ KBs under these semantics. The results of our analysis are reported in Figure 1.

		IAR	k-supp ($k > 1$)	CQA	k-def ($k > 0$)	brave
IC	<i>DL-Lite</i>	NLSPACE	NLSPACE	coNP	NLSPACE	NLSPACE
	<i>DL-Lite_{Horn}</i>	coNP	$\geq \text{coNP}$ $\leq \Delta_2^p[O(\log n)]$	coNP	NP	NP
CQ	<i>DL-Lite</i>	NP	NP	Π_2^p	NP	NP
	<i>DL-Lite_{Horn}</i>	$\Delta_2^p[O(\log n)]$	$\Delta_2^p[O(\log n)]$	Π_2^p	NP	NP

Fig. 1. Combined complexity of instance checking (IC) and conjunctive query entailment (CQ) under various inconsistency-tolerant semantics. All results are completeness results, unless otherwise noted.

Before presenting the results in more detail, let us begin with some general observations. First, it is interesting to note that for *DL-Lite* KBs, the complexities obtained for the IAR, k -support, brave, k -defeater, and classical semantics all coincide, and are strictly lower than the complexity w.r.t. the CQA semantics. By contrast, for *DL-Lite_{Horn}* KBs, instance checking under any of the considered inconsistency-tolerant semantics is of higher complexity than under classical semantics. Moreover, we lose the symmetry between the sound and complete approximations. Indeed, for CQ entailment, the complexities of the sound approximations (IAR and k -support) is higher than for the complete approximations (brave and k -defeater semantics).

Finally, we remark that in several cases, and in particular, for the k -support semantics, the complexity for *DL-Lite_{Horn}* is higher than for *DL-Lite*. This can be explained by the fact that for *DL-Lite* KBs, the size of a minimal \mathcal{T} -support of a query is linear in the size of the query and independent of \mathcal{T} , whereas for *DL-Lite_{Horn}* KBs, the bound on minimal \mathcal{T} -supports depends also on the size of \mathcal{T} . Overall, these results suggest that while the k -support and k -defeater semantics are tractable w.r.t. data complexity for both *DL-Lite* and *DL-Lite_{Horn}*, it will likely be much easier to obtain practical algorithms for *DL-Lite* KBs.

We now present our different complexity results and some brief ideas concerning the proofs. We start by showing that for *DL-Lite*, instance checking under the proposed semantics has the same low complexity as under classical semantics.

Theorem 4. *In DL-Lite, instance checking under the k -support semantics is NLSPACE-complete w.r.t. combined complexity, for every $k \geq 1$. The same holds for the k -defeater semantics, for every $k \geq 0$.*

Proof (idea). The proof exploits the fact that when \mathcal{T} is a *DL-Lite* ontology, minimal \mathcal{T} -supports for IQs consist of single facts, and minimal \mathcal{T} -inconsistent subsets contain at most two facts. This means in particular that every k -tuple of minimal \mathcal{T} -supports contains at most k facts, and at most k facts are needed to contradict all k supports. This enables a NLSPACE procedure which guesses k facts and verifies that each fact is a \mathcal{T} -support, and that there is no set with at most k facts which contradicts all of the guessed facts. The upper bound for the k -defeater semantics uses similar ideas.

In *DL-Lite_{Horn}*, instance checking is intractable already for the IAR and brave semantics, and the lower bounds can be used to show intractability also for the k -support

and k -defeater semantics. For the k -defeater semantics, Theorem 6 provides a matching upper bound, while the precise complexity for the k -support semantics remains open.

Theorem 5. *Instance checking in $DL\text{-Lite}_{Horn}$ is coNP-complete w.r.t. combined complexity under the IAR semantics, coNP-hard w.r.t. combined complexity under k -support semantics, and NP-complete w.r.t. combined complexity under both the brave semantics and k -defeater semantics.*

Proof (idea). We sketch the coNP lower bound for the IAR semantics, which is by reduction from UNSAT. Let $\varphi = c_1 \wedge \dots \wedge c_n$ be a propositional CNF over variables x_1, \dots, x_m . Consider the $DL\text{-Lite}_{Horn}$ KB with

$$\begin{aligned} \mathcal{T} = & \{T_i \sqsubseteq C_j \mid x_i \in c_j\} \cup \{F_i \sqsubseteq C_j \mid \neg x_i \in c_j\} \cup \\ & \{T_i \sqcap F_i \sqsubseteq \perp \mid 1 \leq i \leq m\} \cup \{A \sqcap C_1 \sqcap \dots \sqcap C_n \sqsubseteq \perp\} \end{aligned}$$

and $\mathcal{A} = \{A(a)\} \cup \{T_i(a), F_i(a) \mid 1 \leq i \leq m\}$. Then it can be shown that $\langle \mathcal{T}, \mathcal{A} \rangle \models_{IAR} A(a)$ if and only if the formula φ is unsatisfiable.

We next consider the complexity of CQ entailment under our proposed semantics. For $DL\text{-Lite}$, we obtain precisely the same complexity as under the classical semantics.

Theorem 6. *In $DL\text{-Lite}$, CQ entailment under the k -support semantics is NP-complete w.r.t. combined complexity, for every $k \geq 1$. For both $DL\text{-Lite}$ and $DL\text{-Lite}_{Horn}$, CQ entailment under the k -defeater semantics is NP-complete w.r.t. combined complexity, for every $k \geq 1$.*

Proof (idea). We sketch the upper bound for the k -defeater semantics. Fix a $DL\text{-Lite}_{Horn}$ KB $\langle \mathcal{T}, \mathcal{A} \rangle$ and a CQ q . Let S_1, \dots, S_m be the \mathcal{T} -consistent subsets of \mathcal{A} with cardinality at most k . Guess a sequence C_1, \dots, C_m of subsets of \mathcal{A} of cardinality at most $c = 2 \cdot |q| \cdot |\mathcal{T}|$, together with polynomial certificates that $\langle \mathcal{T}, C_i \rangle \models q$, for each C_i . Output yes if for every $1 \leq i \leq m$, the certificate is valid and $S_i \cup C_i$ is \mathcal{T} -consistent. As m is polynomial in $|\mathcal{A}|$ (since k is fixed), and both conditions can be verified in polynomial time for $DL\text{-Lite}_{Horn}$ KBs, we obtain an NP procedure. Correctness relies on the fact that because \mathcal{T} is a $DL\text{-Lite}_{Horn}$ ontology, every minimal \mathcal{T} -support for q has cardinality at most c .

For $DL\text{-Lite}_{Horn}$, CQ entailment under the IAR and k -support semantics rises to $\Delta_2^P[O(\log n)]$ -complete.

Theorem 7. *In $DL\text{-Lite}_{Horn}$, CQ entailment under k -support semantics is $\Delta_2^P[O(\log n)]$ -complete w.r.t. combined complexity, for every $k \geq 1$.*

Proof (idea). The lower bound is by a non-trivial reduction from the Parity(SAT) problem [21]. For the upper bound, consider the following algorithm which takes as input a $DL\text{-Lite}_{Horn}$ KB $\langle \mathcal{T}, \mathcal{A} \rangle$ and CQ q :

1. For every k -tuple $(\alpha_1, \dots, \alpha_k) \subseteq \mathcal{A}^k$ of facts, use an NP oracle to decide whether every repair contains some α_i . Let S contain all k -tuples for which the test succeeds.

2. A final oracle call checks if there is a k -tuple (C_1, \dots, C_k) of subsets of \mathcal{A} of cardinality at most $c = 2 \cdot |\mathcal{T}| \cdot |q|$ such that (i) every C_i is \mathcal{T} -consistent and $\langle \mathcal{T}, C_i \rangle \models q$, and (ii) every k -tuple $(\beta_1, \dots, \beta_k)$ with $\beta_i \in C_i$ belongs to S . Return yes if the call succeeds, else no.

Every minimal \mathcal{T} -support for q contains at most c facts. It follows that the algorithm returns yes if $\langle \mathcal{T}, \mathcal{A} \rangle \models_{k\text{-supp}} q$. Conversely, if the output is yes, with (C_1, \dots, C_n) the k -tuple from Step 2, then by (i), every C_i is a \mathcal{T} -support for q . Moreover, (ii) ensures that every repair contains some C_i , for it not, we could find some k -tuple $(\beta_1, \dots, \beta_k) \in C_1 \times \dots \times C_n$ which does not belong to S , contradicting (ii). Note that the algorithm runs in polynomial time with an NP oracle, since there are only polynomially many k -tuples of facts to consider, for fixed k . As the oracle calls can be organized into a tree, membership in $\Delta_2^p[O(\log n)]$ follows by a result from [11].

Finally, we determine the combined complexity of instance checking and CQ entailment for the CQA semantics (prior results only considered data complexity).

Theorem 8. *For DL-Lite and DL-Lite_{Horn}, instance checking (resp. CQ entailment) under CQA semantics is coNP-complete (resp. Π_2^p -complete) for combined complexity.*

Proof (idea). The upper bounds are easy: guess a repair and show that it does not entail the query. The coNP-lower bound for instance checking follows from the coNP-hardness of this problem w.r.t. data complexity. The Π_2^p -hardness result involves a non-trivial reduction from 2-QBF validity.

We should point out that our proofs are quite generic and can be directly used (or trivially extended) to obtain results for a whole range *DL-Lite* dialects (as well as other ontology languages).

7 Future Work

The present work can be extended in several directions. First, we believe that our approach can have a practical impact on OBDA systems, so we aim to implement and experiment with the approach. It would also be very interesting to investigate the connections between our approach and approximate knowledge compilation [19]: in particular, it would be important (also for practical purposes) to study the possibility of effectively “compiling” our semantics. Moreover, it is also relevant to extend our analysis to more complex OBDA systems, where the ontology elements are related to the data sources through complex mappings [17]. Finally, while the present approach is computationally attractive for all known FO-rewritable ontology languages, tractable approximations of the CQA semantics for other tractable yet non-FO-rewritable ontology languages (like \mathcal{EL}_\perp [3]) are still missing.

Acknowledgments. The first author has been supported by a Université Paris-Sud Attractivité grant and ANR project PAGODA (ANR-12-JS02-007-01). The second author has been partially supported by EU FP7 project Optique – Scalable End-user Access to Big Data (grant n. FP7-318338).

References

1. Arenas, M., Bertossi, L.E., Chomicki, J.: Consistent query answers in inconsistent databases. In: Proc. of PODS. pp. 68–79. ACM Press (1999)
2. Artale, A., Calvanese, D., Kontchakov, R., Zakharyashev, M.: The DL-Lite family and relations. *Journal of Artificial Intelligence Research* 36, 1–69 (2009)
3. Baader, F., Brandt, S., Lutz, C.: Pushing the \mathcal{EL} envelope. In: Proc. of IJCAI. pp. 364–369 (2005)
4. Bertossi, L.E.: *Database Repairing and Consistent Query Answering*. Synthesis Lectures on Data Management, Morgan & Claypool Publishers (2011)
5. Bienvenu, M.: On the complexity of consistent query answering in the presence of simple ontologies. In: Proc. of AAI (2012)
6. Cali, A., Gottlob, G., Pieris, A.: New expressive languages for ontological query answering. In: Proc. of AAI (2011)
7. Cali, A., Gottlob, G., Pieris, A.: Towards more expressive ontology languages: The query answering problem. *Artificial Intelligence* 193, 87–128 (2012)
8. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning* 39(3), 385–429 (2007)
9. Eiter, T., Gottlob, G.: The complexity of logic-based abduction. *J. ACM* 42(1), 3–42 (1995)
10. Eiter, T., Gottlob, G.: The complexity class θ_2^P : Recent results and applications in AI and modal logic. In: Proc. of FCT. pp. 1–18 (1997)
11. Gottlob, G.: Np trees and Carnap’s modal logic. *Journal of the ACM* 42(2), 421–457 (1995)
12. Immerman, N.: Nondeterministic space is closed under complementation. *SIAM Journal of Computing* 17(5), 935–938 (1988)
13. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant semantics for description logics. In: Proc. of RR. pp. 103–117 (2010)
14. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Query rewriting for inconsistent dl-lite ontologies. In: Proc. of RR. pp. 155–169 (2011)
15. Lukasiewicz, T., Martinez, M.V., Simari, G.I.: Inconsistency handling in datalog \pm ontologies. In: Proc. of ECAI. pp. 558–563 (2012)
16. Lukasiewicz, T., Martinez, M.V., Simari, G.I.: Inconsistency-tolerant query rewriting for linear datalog \pm . In: Proc. of Datalog 2.0. pp. 123–134 (2012)
17. Poggi, A., Lembo, D., Calvanese, D., Giacomo, G.D., Lenzerini, M., Rosati, R.: Linking data to ontologies. *Journal of Data Semantics* 10, 133–173 (2008)
18. Rosati, R.: On the complexity of dealing with inconsistency in description logic ontologies. In: Proc. of IJCAI. pp. 1057–1062 (2011)
19. Selman, B., Kautz, H.A.: Knowledge compilation and theory approximation. *Journal of the ACM* 43(2), 193–224 (1996)
20. Szelepcsényi, R.: The method of forcing for nondeterministic automata. *Bulletin of the EATCS* 33, 96–99 (1987)
21. Wagner, K.W.: More complicated questions about maxima and minima, and some closures of NP. *Theoretical Computer Science* 51, 53–80 (1987)

A Proof of Theorem 3

For the purposes of this section, an ontology \mathcal{T} is a sentence in first-order (FO) logic with equality built from a finite set of symbols drawn from a countable set VN of variables, a countable set CN of constant symbols, and a countable set RN of relation symbols. Function symbols, other than constant symbols, are not permitted. A dataset \mathcal{A} is a finite set of ground facts, built using the symbols from CN and RN .

In what follows, we will assume that every ontology \mathcal{T} has an *associated finite signature* $\text{RN}_{\mathcal{T}}$, which can be any finite subset of RN which contains all relations in \mathcal{T} . The signature $\text{RN}_{\mathcal{T}}$ will define the set of relations which may appear in datasets and in queries. Note that the assumption of a finite relational signature is commonplace in databases and realistic from the point of view of applications.

In keeping with common practice in description logics, we make the *unique names assumption (UNA)*, that is, we only consider as models of a KB $\langle \mathcal{T}, \mathcal{A} \rangle$ those interpretations \mathcal{I} in which $c^{\mathcal{I}} \neq d^{\mathcal{I}}$ for every pair c, d of distinct constant symbols appearing in $\mathcal{T} \cup \mathcal{A}$. We point out however that this assumption is not essential, and Theorem 3 applies also to the case where the UNA is not adopted.

We recall that for the sake of readability, we consider only Boolean queries, and so in what follows we will take the term “query” to mean Boolean query. All results and proofs can be straightforwardly extended to handle non-Boolean queries.

An ontology \mathcal{T} is said to be *first-order-rewritable for CQ answering under semantics \mathcal{S}* if, for every CQ q , there exists an effectively computable FO query q' such that, for every $\Sigma_{\mathcal{T}}$ -dataset \mathcal{A} , $\langle \mathcal{T}, \mathcal{A} \rangle$ entails q under semantics \mathcal{S} iff q' is satisfied in the interpretation $\mathcal{I}_{\mathcal{A}}$ (denoted $\mathcal{I}_{\mathcal{A}} \models q'$). Such a query q' is called a *(FO-)rewriting of q relative to \mathcal{T} under semantics \mathcal{S}* . If there exists an FO query q' such that, for every \mathcal{T} -consistent $\Sigma_{\mathcal{T}}$ -dataset \mathcal{A} , $\langle \mathcal{T}, \mathcal{A} \rangle$ entails q under semantics \mathcal{S} iff $\mathcal{I}_{\mathcal{A}} \models q'$, then we say that \mathcal{T} is *first-order-rewritable for CQ answering and consistent datasets under semantics \mathcal{S}* . Note that it is possible for a FO-rewriting to exist relative to consistent datasets, while there is no FO-rewriting w.r.t. arbitrary datasets.

A *union of conjunctive queries with inequality (UCQ_≠)* is a first-order sentence built from atomic formulas, equalities between terms, and inequalities between terms using \wedge , \vee , and \exists . If every CQ possesses a FO-rewriting relative to \mathcal{T} (and consistent datasets) under semantics \mathcal{S} which takes the form of a UCQ_≠, then we say that \mathcal{T} is *UCQ_≠-rewritable for CQ answering (relative to consistent datasets) under semantics \mathcal{S}* . The notions of FO- and UCQ_≠-rewritability for unsatisfiability are defined in the obvious way. For the sake of readability, we will often omit “under semantics \mathcal{S} ”, in which case we mean classical FO semantics.

We consider the following possible properties of an ontology \mathcal{T} :

- B1** there is some ℓ such that for every \mathcal{A} , every minimal \mathcal{T} -inconsistent subset of \mathcal{A} has cardinality at most ℓ
- B2** for every CQ q , there is some m such that for every \mathcal{A} , every minimal \mathcal{T} -support for q has cardinality at most m
- U1** \mathcal{T} is UCQ_≠-rewritable for unsatisfiability
- U2** \mathcal{T} is UCQ_≠-rewritable for CQ answering relative to consistent datasets
- U3** \mathcal{T} is UCQ_≠-rewritable for CQ answering

It is not hard to see that any ontology that satisfies U1 and U2 must also satisfy B1 and B2. Indeed, one can construct the desired bounds on the cardinality of minimal \mathcal{T} -inconsistent sets (resp. minimal \mathcal{T} -supports of q) by considering the maximal number of atoms appearing in the disjuncts of the UCQ $_{\neq}$ -rewriting of unsatisfiability (resp. of q relative to consistent datasets). In preparation for the proof of Theorem 3, we establish that B1 and B2 together imply properties U1, U2, and U3.

Lemma 1. *If \mathcal{T} satisfies condition B1, then it also satisfies condition U1.*

Proof. Suppose that \mathcal{T} satisfies condition B1, and so there is a bound ℓ on the size of minimal \mathcal{T} -inconsistent subsets. Let $\mathcal{U}_1, \mathcal{U}_2, \dots$ be an enumeration of all minimal \mathcal{T} -inconsistent sets, and let $\text{CN}_{\mathcal{T}}$ be the set of constant symbols appearing in \mathcal{T} . Define an equivalence relation \sim on $\text{RN}_{\mathcal{T}}$ -datasets by setting $\mathcal{D}_1 \sim \mathcal{D}_2$ whenever there exists a bijection f from the constant symbols in \mathcal{D}_1 to the constant symbols in \mathcal{D}_2 such that:

- for every relation $R \in \text{RN}_{\mathcal{T}}$: $R(c_1, \dots, c_k) \in \mathcal{D}_1$ iff $R(f(c_1), \dots, f(c_k)) \in \mathcal{D}_2$
- for every $c \in \text{CN}_{\mathcal{T}}$: c occurs in \mathcal{D}_1 iff c occurs in \mathcal{D}_2 , and if c is present in \mathcal{D}_1 , we have $f(c) = c$

We note that since $\text{RN}_{\mathcal{T}}$ and $\text{CN}_{\mathcal{T}}$ are both finite, and the cardinality of the sets \mathcal{U}_i is bounded by ℓ , we can choose a finite set $\{\mathcal{U}'_1, \dots, \mathcal{U}'_k\}$ of sets from the sequence $\mathcal{U}_1, \mathcal{U}_2, \dots$ such that every \mathcal{U}_i is equivalent modulo \sim to one of the \mathcal{U}'_j .

Next we associate a query to each of the datasets \mathcal{D} in $\{\mathcal{U}'_1, \dots, \mathcal{U}'_k\}$. Specifically, if \mathcal{D} has constant symbols d_1, \dots, d_n , then we use the following query:

$$q_{\mathcal{D}} = \exists v_1, \dots, v_n \bigwedge_{1 \leq i < j \leq n} v_i \neq v_j \wedge \bigwedge_{d_i \in \text{CN}_{\mathcal{T}}} v_i = d_i \wedge \bigwedge_{R(d_{i_1}, \dots, d_{i_r}) \in \mathcal{D}} R(v_{i_1}, \dots, v_{i_r})$$

We then define a UCQ $_{\neq}$ -query as follows:

$$q' = \bigvee_{1 \leq i \leq k} q_{\mathcal{U}'_i}$$

To complete the proof, we establish the following claim.

Claim: q' is a rewriting of unsatisfiability relative to \mathcal{T} .

Proof of claim: For the first direction, suppose that \mathcal{A} is a $\text{RN}_{\mathcal{T}}$ -dataset with $\mathcal{I}_{\mathcal{A}} \models q'$. Then there is some $1 \leq i \leq k$ such that $\mathcal{I}_{\mathcal{A}} \models q_{\mathcal{U}'_i}$. It follows that there is some assignment μ to the variables v_1, \dots, v_n of $q_{\mathcal{U}'_i}$ which makes the conjuncts of $q_{\mathcal{U}'_i}$ true in $\mathcal{I}_{\mathcal{A}}$. We aim to show that \mathcal{A} is \mathcal{T} -inconsistent. Suppose for a contradiction that this is not the case. Let \mathcal{I} be a model of \mathcal{T} and \mathcal{A} , and let \mathcal{J} be the interpretation defined as follows:

- the domain of \mathcal{J} is the same as the domain of \mathcal{I}
- every constant symbol c which does not appear in \mathcal{U}'_i is interpreted as in \mathcal{I} : $c^{\mathcal{J}} = c^{\mathcal{I}}$
- every constant symbol d_j in \mathcal{U}'_i is interpreted as follows: $d_j^{\mathcal{J}} = \mu(v_j)^{\mathcal{I}}$
- every relation symbol R is interpreted exactly as in \mathcal{I} : $R^{\mathcal{J}} = R^{\mathcal{I}}$

First we show that \mathcal{J} is a model of \mathcal{U}'_i . If $R(d_{j_1}, \dots, d_{j_r}) \in \mathcal{U}'_i$, then the satisfaction of the third conjunct of $q_{\mathcal{U}'_i}$ yields $R(\mu(v_{j_1}), \dots, \mu(v_{j_r})) \in \mathcal{A}$. We thus have $(\mu(v_{j_1})^{\mathcal{I}}, \dots, \mu(v_{j_r})^{\mathcal{I}}) \in R^{\mathcal{I}}$, hence $(d_{j_1}^{\mathcal{J}}, \dots, d_{j_r}^{\mathcal{J}}) \in R^{\mathcal{J}}$. To show that \mathcal{J} is a model of \mathcal{T} , we first remark that whenever a constant symbol d_j appears both in \mathcal{T} and in \mathcal{U}'_i , the second conjunct of $q_{\mathcal{U}'_i}$ ensures that $\mu(v_j) = d_j$. It follows that all constant symbols in \mathcal{T} are interpreted identically in \mathcal{I} and \mathcal{J} . Moreover, by construction, \mathcal{I} and \mathcal{J} have the same domains and interpret all relation symbols identically. It follows that every FO formula built using only the constant symbols in $\text{CN}_{\mathcal{T}}$ and relation symbols in $\text{RN}_{\mathcal{T}}$ is interpreted identically in \mathcal{I} and \mathcal{J} . In particular, this means that \mathcal{J} is a model of \mathcal{T} . Finally, we note that the UNA is satisfied in \mathcal{J} for the KB $\langle \mathcal{T}, \mathcal{U}'_i \rangle$ since it is satisfied in \mathcal{I} for the KB $\langle \mathcal{T}, \mathcal{A} \rangle$ and for every pair d_j, d_r of distinct constant symbols in \mathcal{U}'_i , the first conjunct of $q_{\mathcal{U}'_i}$ guarantees that $d_j^{\mathcal{J}} = \mu(v_j)^{\mathcal{I}} \neq \mu(v_r)^{\mathcal{I}} = d_r^{\mathcal{J}}$. We have thus constructed a model for $\langle \mathcal{T}, \mathcal{U}'_i \rangle$, contradicting the fact that $\mathcal{T}, \mathcal{U}'_i \models \perp$. We can conclude that $\langle \mathcal{T}, \mathcal{A} \rangle \models \perp$.

For the other direction, suppose that $\langle \mathcal{T}, \mathcal{A} \rangle \models \perp$, and let \mathcal{A}' be a minimal \mathcal{T} -inconsistent subset of \mathcal{A} . It follows that there is some \mathcal{U}'_i such that $\mathcal{A}' \sim \mathcal{U}'_i$, and this is witnessed by some bijection f from the constant symbols in \mathcal{U}'_i to the constant symbols in \mathcal{A}' . Let d_1, \dots, d_n be the constant symbols in \mathcal{U}'_i and v_1, \dots, v_n the variables in $q_{\mathcal{U}'_i}$. Define an assignment μ which sends v_j to $f(d_j)$. The first conjunct of $q_{\mathcal{U}'_i}$ is satisfied in $\mathcal{I}_{\mathcal{A}}$ under assignment μ since f is a bijection, so $f(d_j) \neq f(d_r)$ for $j \neq r$. The second conjunct is also satisfied, since if d is a constant in \mathcal{U}'_i which occurs in \mathcal{T} , then we must have $f(d) = d$. Finally, if the fact $R(d_{j_1}, \dots, d_{j_r})$ belongs to \mathcal{U}'_i , then we must have $R(f(d_{j_1}), \dots, f(d_{j_r}))$, hence the third conjunct is also satisfied in $\mathcal{I}_{\mathcal{A}}$ under μ . We have thus shown that $q_{\mathcal{U}'_i}$ holds in $\mathcal{I}_{\mathcal{A}}$, so $\mathcal{I}_{\mathcal{A}} \models q'$. (*end proof of claim*)

Lemma 2. *If \mathcal{T} satisfies condition B2, then it also satisfies U2.*

Proof. Suppose that \mathcal{T} satisfies condition B2, and let q be a CQ. By condition B2, we have a bound m on the size of minimal \mathcal{T} -supports of q . Let $\mathcal{S}_1, \mathcal{S}_2, \dots$ be an enumeration of all minimal \mathcal{T} -supports of q , and let $\text{CN}_{\mathcal{T}, q}$ be the set of constant symbols appearing in \mathcal{T} or in q . As in the proof of Lemma 1, we define an equivalence relation \sim on $\text{RN}_{\mathcal{T}}$ -datasets by setting $\mathcal{D}_1 \sim \mathcal{D}_2$ whenever there exists a bijection f from the constant symbols in \mathcal{D}_1 to the constant symbols in \mathcal{D}_2 such that:

- for every $R \in \text{RN}_{\mathcal{T}}$: $R(c_1, \dots, c_k) \in \mathcal{D}_1$ iff $R(f(c_1), \dots, f(c_k)) \in \mathcal{D}_2$
- for every $c \in \text{CN}_{\mathcal{T}, q}$: c occurs in \mathcal{D}_1 iff c occurs in \mathcal{D}_2 , and if c is present in \mathcal{D}_1 , we have $f(c) = c$

Since $\text{RN}_{\mathcal{T}}$ and $\text{CN}_{\mathcal{T}, q}$ are both finite, and the cardinality of the sets \mathcal{S}_i is bounded by m , there can be only finitely many distinct equivalence classes which contain some \mathcal{S}_i . For each such equivalence class, we choose a representative dataset \mathcal{S}_i , and we let $\{\mathcal{S}'_1, \dots, \mathcal{S}'_p\}$ be the finite set of representatives. Note that for every $i \geq 1$, we have $\mathcal{S}_i \sim \mathcal{S}'_j$ for some $1 \leq j \leq p$.

Just as in the proof of Lemma 1, to every dataset \mathcal{D} with constant symbols d_1, \dots, d_n , we associate the query

$$q_{\mathcal{D}} = \exists v_1, \dots, v_n \bigwedge_{1 \leq i < j \leq n} v_i \neq v_j \wedge \bigwedge_{d_i \in \text{CN}_{\mathcal{T}, q}} v_i = d_i \wedge \bigwedge_{R(d_{i_1}, \dots, d_{i_r}) \in \mathcal{D}} R(v_{i_1}, \dots, v_{i_r})$$

We then construct the following UCQ_≠-query:

$$q' = \bigvee_{1 \leq i \leq p} q_{S'_i}$$

Claim: q' is a rewriting of q relative to \mathcal{T} and consistent datasets.

Proof of claim: For the first direction, suppose that \mathcal{A} is a \mathcal{T} -consistent RN $_{\mathcal{T}}$ -dataset with $\mathcal{I}_{\mathcal{A}} \models q'$. Then there is some $1 \leq i \leq p$ such that $\mathcal{I}_{\mathcal{A}} \models q_{S'_i}$. We can thus find an assignment μ to the variables v_1, \dots, v_n of $q_{S'_i}$ which makes the conjuncts of $q_{S'_i}$ hold in $\mathcal{I}_{\mathcal{A}}$. We aim to show that $\langle \mathcal{T}, \mathcal{A} \rangle \models q$. Suppose for a contradiction that this is not the case, and let \mathcal{I} be a model of \mathcal{T} and \mathcal{A} in which q does not hold. Let \mathcal{J} be the interpretation which is the same as \mathcal{I} except that:

- every constant symbol d_j in \mathcal{U}'_i is interpreted as follows: $d_j^{\mathcal{J}} = \mu(v_j)^{\mathcal{I}}$

Using the same arguments as in the proof of Lemma 1, one can show that \mathcal{J} is a model of the KB $\langle \mathcal{T}, S'_i \rangle$. By leveraging the fact that all constant symbols and relation symbols appearing in q are interpreted identically in \mathcal{I} and \mathcal{J} , one can further show that q does not hold in \mathcal{J} , contradicting the fact that S'_i is a \mathcal{T} -support for q . We can thus conclude that $\langle \mathcal{T}, \mathcal{A} \rangle \models q$.

For the other direction, suppose that \mathcal{A} is \mathcal{T} -consistent and such that $\langle \mathcal{T}, \mathcal{A} \rangle \models q$. Let $\mathcal{A}' \subseteq \mathcal{A}$ be a minimal \mathcal{T} -support for q . It follows from the definition of the set $\{S'_1, \dots, S'_p\}$ that there is some S'_j such that $\mathcal{A}' \sim S'_j$. By using essentially the same arguments as in the proof of Lemma 1, we can show that $q_{S'_j}$ holds in $\mathcal{I}_{\mathcal{A}}$, which yields $\mathcal{I}_{\mathcal{A}} \models q'$. (*end proof of claim*)

Lemma 3. *If \mathcal{T} satisfies U1 and U3, then it also satisfies U2.*

Proof. Suppose that \mathcal{T} satisfies U1 and U3, and let q be a CQ. Then we can find a UCQ_≠-rewriting Φ_{\perp} of unsatisfiability relative to \mathcal{T} and a UCQ_≠-rewriting Φ_q of q relative to \mathcal{T} and consistent datasets. It is easily verified that $\Phi_{\perp} \vee \Phi_q$ is the desired UCQ_≠-rewriting of q relative to \mathcal{T} .

Combining Lemmas 1, 2, and 3, we have that every ontology \mathcal{T} satisfying the hypotheses of Theorem 3 also satisfies conditions U1, U2, and U3⁵. We now proceed to the main part of the proof of Theorem 3, which consists in exploiting conditions U1 and U2 to construct FO-rewritings under the k -support and k -defeater semantics.

Proof of Theorem 3. Let \mathcal{T} be an ontology satisfying the hypotheses of Theorem 3, and let q be a CQ. We denote by $\text{CN}_{\mathcal{T}}$ (resp. $\text{CN}_{\mathcal{T},q}$) the set of constant symbols appearing in \mathcal{T} (resp. \mathcal{T} or q). By Lemma 1, there is a UCQ_≠-rewriting $\Phi_{\perp} = \kappa_1 \vee \dots \vee \kappa_u$ of unsatisfiability relative to \mathcal{T} . Examining the proof of Lemma 1, we can choose Φ_{\perp} so that: (i) Φ_{\perp} uses only constant symbols from $\text{CN}_{\mathcal{T}}$, and (ii) if a variable v appears in some conjunct of κ_j , then v appears in an atomic formula in κ_j . By Lemma 2, we can also find a UCQ_≠ $\Phi_q = \varphi_1 \vee \dots \vee \varphi_s$ which is a rewriting of q relative to \mathcal{T}

⁵ Actually, we have shown that only two of the three hypotheses of Theorem 3 are required: the hypotheses B1 and B2 guarantee U3, which in turn implies the third hypothesis, namely, FO-rewritability of \mathcal{T} for CQ answering.

and consistent datasets (under classical semantics) We can again assume w.l.o.g. that Φ_q uses only constant symbols from $\text{CN}_{\mathcal{T},q}$, and that every variable v which occurs in φ_j appears in some atom in φ_j . We can further assume that each disjunct φ_i holds in some \mathcal{T} -consistent dataset (indeed, one can simply check whether φ_i is contained in the query Φ_{\perp} and remove it if this is the case). Let m be the maximum number of variables in any disjunct κ_i and n the maximum number of variables in any φ_i . By renaming variables, we ensure that all disjuncts in Φ_q use the same set of variables, say x_1, \dots, x_n . Likewise, we can assume that the query Φ_{\perp} is over the set of variables y_1, \dots, y_m .

The desired FO-rewriting of q relative to \mathcal{T} under k -support semantics will take the following form:

$$\Psi_{k\text{-supp}} = \bigvee_{(i_1, \dots, i_k) \in [1, s]^k} \chi_{(i_1, \dots, i_k)}$$

Intuitively, each tuple $(i_1, \dots, i_k) \in [1, s]^k$ corresponds to a choice of k supports, where the disjuncts φ_i represent the different possible supports. Note that we allow the same index to appear more than once. This allows us to capture the case in which less than k distinct supports are used, or the case where the same type of support occurs multiple times in the data using different sets of constants. The formula $\chi_{(i_1, \dots, i_k)}$ states that these k chosen supports are present in the dataset and that there is no way to simultaneously contradict them. This formula takes the following form:

$$\chi_{(i_1, \dots, i_k)} = \bigexists_{\substack{1 \leq \ell \leq n \\ 1 \leq j \leq k}} x_{\ell}^j \left(\bigwedge_{1 \leq j \leq k} \varphi_{i_j} [x_{\ell} \mapsto x_{\ell}^j] \wedge \neg \delta_{(i_1, \dots, i_k)} \right)$$

where $\varphi_{i_j} [x_{\ell} \mapsto x_{\ell}^j]$ indicates that every variable x_{ℓ} in φ_{i_j} has been replaced with x_{ℓ}^j . The formula $\delta_{(i_1, \dots, i_k)}$ will serve to check for the existence of a consistent set of facts that conflict each of the chosen supports. To define $\delta_{(i_1, \dots, i_k)}$, let \mathcal{S} consist of all pairs (D, E) where D and E are sets of atomic formulas and (in)equalities respectively which satisfy the following conditions:

- the atomic formulas in D use only relation symbols from $\text{RN}_{\mathcal{T}}$ and constant symbols from $\text{CN}_{\mathcal{T}}$
- the set V_D of variables in D is a subset of $\{v_1, \dots, v_o\}$, where $o = km$
- the query

$$\zeta_D = \bigwedge_{\alpha \in D} \alpha \wedge \bigwedge_{\substack{v_{\ell}, v_{\ell'} \in V_D \\ \ell \neq \ell'}} v_{\ell} \neq v_{\ell'} \wedge \bigwedge_{\substack{v_{\ell} \in V_D, \\ a \in \text{CN}_{\mathcal{T}}}} v_{\ell} \neq a$$

is such that⁶

$$\bigexists_{v_{\ell} \in V_D} v_{\ell} \zeta_D \not\subseteq \Phi_{\perp}$$

⁶ As usual, we use \subseteq to denote query containment: $q_1 \subseteq q_2$ signifies that for all $\text{RN}_{\mathcal{T}}$ -datasets D , we have $\mathcal{I}_D \models q_1$ if and only if $\mathcal{I}_D \models q_2$.

- E consists of equalities and inequalities of the forms $x_\ell^j = v_p$, $x_\ell^j \neq v_p$, where $v_p \in V_D$, $1 \leq \ell \leq n$, and $1 \leq j \leq k$
- for every $1 \leq j \leq k$,

$$\bigoplus_{v \in V_j} v (\zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon \wedge \varphi_{i_j}[x_\ell \mapsto x_\ell^j]) \subseteq \Phi_\perp$$

where V_j denotes the set of variables occurring in the left-hand-side query.

The query $\delta_{(i_1, \dots, i_k)}$ is then defined as follows:

$$\delta_{(i_1, \dots, i_k)} = \exists v_1, \dots, v_{km} \bigvee_{(D, E) \in \mathcal{S}} (\zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon)$$

Now that we have completed the definition of the query $\Psi_{k\text{-supp}}$, we turn to the problem of showing that $\Psi_{k\text{-supp}}$ constitutes a rewriting of q relative to \mathcal{T} under the k -support semantics. First suppose that $\langle \mathcal{T}, \mathcal{A} \rangle \models_{k\text{-supp}} q$. Then there exist subsets S_1, \dots, S_k of \mathcal{A} such that each S_i is a \mathcal{T} -support for q in \mathcal{A} , and every $R \in \text{Rep}(\mathcal{K})$ contains some S_i . It follows that for every S_j , we have $\mathcal{I}_{S_j} \models \Phi_q$, so there must exist i_j such that $\mathcal{I}_{S_j} \models \exists x_1, \dots, x_n \varphi_{i_j}$. Thus, for every $1 \leq i \leq k$, we can find a variable assignment μ_j for x_1, \dots, x_n such that \mathcal{I}_{S_j} satisfies φ_{i_j} under μ_j , which we write $\mathcal{I}_{S_j} \models^{\mu_j} \varphi_{i_j}$. We wish to show that $\mathcal{I}_{\mathcal{A}} \models \chi_{(i_1, \dots, i_k)}$. First let μ be the variable assignment for x_1^1, \dots, x_m^k defined as follows: $\mu(x_\ell^j) = \mu_j(x_\ell)$. It is easy to see that $\mathcal{I}_{\mathcal{A}} \models^\mu \bigwedge_{1 \leq j \leq k} \varphi_{i_j}[x_\ell \mapsto x_\ell^j]$. It remains to show that $\mathcal{I}_{\mathcal{A}} \models^\mu \neg \delta_{(i_1, \dots, i_k)}$. Suppose for a contradiction that this is not the case. Then there must exist $(D, E) \in \mathcal{S}$ such that

$$\mathcal{I}_{\mathcal{A}} \models^\mu \exists v_1, \dots, v_{km} (\zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon)$$

and so there is a variable assignment μ' for $x_1^1, \dots, x_m^k, v_1, \dots, v_o$ which extends μ and is such that

$$\mathcal{I}_{\mathcal{A}} \models^{\mu'} \zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon$$

Now let \mathcal{B} be the set of facts obtained by applying μ' to each atom in D . As $\mathcal{I}_{\mathcal{A}} \models^{\mu'} \bigwedge_{\alpha \in D} \alpha$, we know that \mathcal{B} is a subset of \mathcal{A} .

Claim 1. \mathcal{B} is \mathcal{T} -consistent.

Proof. Suppose for a contradiction that $\langle \mathcal{T}, \mathcal{B} \rangle \models \perp$. As Φ_\perp is a rewriting of unsatisfiability w.r.t. \mathcal{T} , we have $\mathcal{I}_{\mathcal{B}} \models \Phi_\perp$. Let \mathcal{D} be any $\text{RN}_{\mathcal{T}}$ -dataset such that $\mathcal{I}_{\mathcal{D}} \models \bigoplus_{v_\ell \in V_D} v_\ell \zeta_D$. Then there must exist an assignment η to the variables in V_D such that $\mathcal{I}_{\mathcal{D}} \models^\eta \zeta_D$. Let \mathcal{D}' be the subset of \mathcal{D} consisting of those facts which can be obtained by applying η to some atom in D . Define a mapping f from the constants in \mathcal{B} to the constants in \mathcal{D} by setting $f(b) = b$ if $b \in \text{CN}_{\mathcal{T}}$, and otherwise, setting $f(\mu'(v_\ell)) = \eta(v_\ell)$. It is easily verified that f defines an isomorphism between $\mathcal{I}_{\mathcal{B}}$ and $\mathcal{I}_{\mathcal{D}}$. Since $\mathcal{I}_{\mathcal{B}} \models \Phi_\perp$, we must also have $\mathcal{I}_{\mathcal{D}'} \models \Phi_\perp$. It follows that \mathcal{D}' is \mathcal{T} -inconsistent, so \mathcal{D} must also be \mathcal{T} -inconsistent, which implies $\mathcal{I}_{\mathcal{D}'} \models \Phi_\perp$. We have thus shown that the query $\bigoplus_{v_\ell \in V(D)} v_\ell \zeta_D$ is contained in Φ_\perp , yielding the desired contradiction.

Claim 2. $\mathcal{B} \cup S_j$ is \mathcal{T} -inconsistent for every $1 \leq j \leq k$.

Proof. Let $1 \leq j \leq k$. It was previously shown that

$$\mathcal{I}_{\mathcal{A}} \models^{\mu'} \zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon \wedge \varphi_{i_j}[x_\ell \mapsto x_\ell^j]$$

Since every variable in the latter query appears in some atom of D or $\varphi_{i_j}[x_\ell \mapsto x_\ell^j]$, and the image of these atoms under μ' is contained in $\mathcal{B} \cup S_j$, we can infer that

$$\mathcal{I}_{\mathcal{B} \cup S_j} \models \bigvee_{v \in V_j} v \left(\zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon \wedge \varphi_{i_j}[x_\ell \mapsto x_\ell^j] \right)$$

We know from the definition of the set \mathcal{S} that the preceding query is contained in Φ_\perp , so we immediately get $\mathcal{I}_{\mathcal{B} \cup S_j} \models \Phi_\perp$. Since Φ_\perp is a rewriting of unsatisfiability w.r.t. \mathcal{T} , the latter implies that $\mathcal{B} \cup S_j$ is \mathcal{T} -inconsistent.

By Claim 1, \mathcal{B} is \mathcal{T} -consistent, so there must exist $\mathcal{R} \in \text{Rep}(\mathcal{T}, \mathcal{A})$ such that $\mathcal{B} \subseteq \mathcal{R}$. It follows from Claim 2 that for every $1 \leq j \leq k$, $\mathcal{R} \cup S_j$ is \mathcal{T} -inconsistent, hence $S_j \not\subseteq \mathcal{R}$. We have thus found a repair \mathcal{R} which contains none of the sets S_j , a contradiction. We may thus conclude that $\mathcal{I}_{\mathcal{A}} \models^{\mu} \neg\delta_{(i_1, \dots, i_k)}$, and hence that $\mathcal{I}_{\mathcal{A}} \models \chi_{(i_1, \dots, i_k)}$.

For the other direction, suppose $\mathcal{I}_{\mathcal{A}} \models \Psi_{k\text{-supp}}$. Then there must exist $(i_1, \dots, i_k) \in [1, s]^k$ such that

$$\mathcal{I}_{\mathcal{A}} \models \bigvee_{\substack{1 \leq \ell \leq n \\ 1 \leq j \leq k}} x_\ell^j \left(\bigwedge_{1 \leq j \leq k} \varphi_{i_j}[x_\ell \mapsto x_\ell^j] \wedge \neg\delta_{(i_1, \dots, i_k)} \right)$$

so we can find an assignment μ to the variables x_ℓ^j such that

$$\mathcal{I}_{\mathcal{A}} \models^{\mu} \bigwedge_{1 \leq j \leq k} \varphi_{i_j}[x_\ell \mapsto x_\ell^j] \wedge \neg\delta_{(i_1, \dots, i_k)}$$

For $1 \leq j \leq k$, let S_j be the set of facts obtained by applying μ to the atoms in the query $\varphi_{i_j}[x_\ell \mapsto x_\ell^j]$.

Claim 3. S_j is a \mathcal{T} -support for q , for every $1 \leq j \leq k$.

Proof. Fix some $1 \leq j \leq k$, and let X_{i_j} be the set of variables in φ_{i_j} . The first step is to show that S_j is \mathcal{T} -consistent. Assume for a contradiction that S_j is \mathcal{T} -inconsistent, in which case $\mathcal{I}_{S_j} \models \Phi_\perp$. Using arguments similar to those in Claim 1, we can show that for any dataset \mathcal{D} , if $\mathcal{I}_{\mathcal{D}} \models \varphi_{i_j}$, then $\mathcal{I}_{\mathcal{D}} \models \Phi_\perp$. This means that the query φ_{i_j} is contained in Φ_\perp , contradicting our earlier assumption. Next we remark that since every variable which participates in φ_{i_j} appears in an atom of φ_{i_j} , and the image of these atoms under μ is contained in $S_j \subseteq \mathcal{A}$, we have

$$\mathcal{I}_{S_j} \models \bigvee_{x_\ell \in X_{i_j}} x_\ell \varphi_{i_j}$$

As φ_{i_j} is a disjunct of Φ_q , it follows that $\mathcal{I}_{S_j} \models \Phi_q$. Since Φ_q is a FO-rewriting of q relative to \mathcal{T} , we get $\langle \mathcal{T}, S_j \rangle \models q$, i.e. S_j is a \mathcal{T} -support for q .

Claim 4. Every $\mathcal{R} \in \text{Rep}(\mathcal{T}, \mathcal{A})$ contains some S_j .

Proof. Suppose for a contradiction that $\mathcal{R} \in \text{Rep}(\mathcal{T}, \mathcal{A})$ is such that $S_j \not\subseteq \mathcal{R}$ for every $1 \leq j \leq k$. Then by the maximality of repairs, it must be the case that $\mathcal{R} \cup S_j$ is \mathcal{T} -inconsistent for every $1 \leq j \leq k$. It follows that for each $1 \leq j \leq k$, $\mathcal{I}_{\mathcal{R} \cup S_j} \models \Phi_{\perp}$, and so we can find some disjunct κ_{h_j} of Φ_{\perp} such that

$$\mathcal{I}_{\mathcal{R} \cup S_j} \models \kappa_{h_j}$$

where Y_j is the set of variables in κ_{h_j} . Thus, for each $1 \leq j \leq k$, we can find an assignment η_j to the variables in Y_j such that $\mathcal{I}_{\mathcal{R} \cup S_j} \models^{\eta_j} \kappa_{h_j}$. Let $d_1, \dots, d_{o'}$ be an enumeration of all constant symbols in $\{\eta_j(y) \mid 1 \leq j \leq k, y \in Y_j\} \setminus \text{CN}_{\mathcal{T}}$. Note that since each Y_j contains at most m variables, $o' \leq o = km$. Define D'_j as the set of atomic formulas in κ_{h_j} whose image under η_j belongs to \mathcal{R} . For every variable y_e in D'_j , do the following:

- if $\eta_j(y_e) \in \text{CN}_{\mathcal{T}}$, then replace y_e by $\eta_j(y_e)$
- if $\eta_j(y_e) = d_g$, then replace y_e by v_g

Call the resulting set D_j , and set D equal to $D_1 \cup \dots \cup D_k$. Next define a set E consisting of the following (in)equalities:

- $v_g = x_{\ell}^j$, for every g, ℓ, j such that $\mu(x_{\ell}^j) = d_g$
- $v_g \neq x_{\ell}^j$, for every g, ℓ, j such that $\mu(x_{\ell}^j) \neq d_g$

Let μ' be the assignment which extends μ to $v_1, \dots, v_{o'}$ as follows: $\mu'(v_g) = d_g$. Now consider the different types of conjuncts in $\zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon$:

- For each atom $\alpha \in D$, there must exist $1 \leq j \leq k$ and an atom $\alpha' \in \kappa_{h_j}$ such that η_j maps α' to \mathcal{R} . Moreover, by construction, the image of α' under η_j is the same as the image of α under μ' . It follows that μ' maps α to a fact in $\mathcal{R} \subseteq \mathcal{A}$, so the atom α is satisfied in $\mathcal{I}_{\mathcal{A}}$ under μ' .
- For each pair of distinct variables $v_{\ell}, v_{\ell'}$ in ζ_D , we have $\mu'(v_{\ell}) = d_{\ell}$ and $\mu'(v_{\ell'}) = d_{\ell'}$, hence $\mu'(v_{\ell}) \neq \mu'(v_{\ell'})$.
- For each variable v_{ℓ} in ζ_D , we have $\mu'(v_{\ell}) = d_{\ell} \notin \text{CN}_{\mathcal{T}}$, so $\mu'(v_{\ell}) \neq c$ for every $c \in \text{CN}_{\mathcal{T}}$.
- For every equality $v_g = x_{\ell}^j$ in E , we have $\mu'(v_g) = d_g$ and $\mu(x_{\ell}^j) = d_g$, so $\mu'(v_g) = \mu'(x_{\ell}^j)$.
- For every inequality $v_g \neq x_{\ell}^j$ in E , we have $\mu'(v_g) = d_g$ and $\mu(x_{\ell}^j) \neq d_g$, so $\mu'(v_g) \neq \mu'(x_{\ell}^j)$.

We have thus shown that every conjunct in $\zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon$ holds in $\mathcal{I}_{\mathcal{A}}$ under the assignment μ' . As μ' extends μ , it follows that

$$\mathcal{I}_{\mathcal{A}} \models^{\mu} \exists v_1, \dots, v_{o'} \zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon$$

It remains to show that $(D, E) \in \mathcal{S}$. By construction, the first, second, and fourth conditions are satisfied by D and E . For the third condition, we note that $\mathcal{I}_{\mathcal{R}} \models \zeta_D$.

Since \mathcal{R} is \mathcal{T} -consistent, we have $\mathcal{I}_{\mathcal{R}} \not\models \Phi_{\perp}$, so ζ_D is not contained in Φ_{\perp} . Finally, for the last condition, one can use $\mathcal{I}_{\mathcal{R} \cup \mathcal{S}_j} \models \kappa_{h_j}$ and the definition of D and E to show that the query $\exists_{v \in V_j} v (\zeta_D \wedge \bigwedge_{\epsilon \in E} \epsilon \wedge \varphi_{i_j}[x_{\ell} \mapsto x'_{\ell}])$ is contained in Φ_{\perp} . We thus have $(D, E) \in \mathcal{S}$, from which we obtain $\mathcal{I}_{\mathcal{A}} \models^{\mu} \delta_{(i_1, \dots, i_k)}$, a contradiction.

From Claims 3 and 4, we immediately obtain $\mathcal{T}, \mathcal{A} \models_{k\text{-supp}} q$. This completes the proof that $\Psi_{k\text{-supp}}$ is an FO-rewriting of q relative to \mathcal{T} under the k -support semantics.

Next we consider how to construct a rewriting of q under the k -defeater semantics. The desired rewriting will take the following form:

$$\Psi_{k\text{-def}} = \bigvee_{D \in \mathcal{S}_k} \chi_D$$

Each $D \in \mathcal{S}_k$ represents a choice of a consistent subset of at most k facts, and the query χ_D checks that these facts are present in the data and conflict with all causes of q . Formally, we let \mathcal{S}_k contain those sets D of at most k atomic formulas which satisfy the following conditions:

- the atomic formulas in D use only relation symbols from $\text{RN}_{\mathcal{T}}$ and constant symbols from $\text{CN}_{\mathcal{T}}$
- the set V_D of variables in D is a subset of $\{v_1, \dots, v_k\}$
- the query

$$\zeta_D = \bigwedge_{\alpha \in D} \alpha \wedge \bigwedge_{\substack{v_{\ell}, v_{\ell'} \in V_D \\ \ell \neq \ell'}} v_{\ell} \neq v_{\ell'} \wedge \bigwedge_{\substack{v_{\ell} \in V_D, \\ a \in \text{CN}_{\mathcal{T}}}} v_{\ell} \neq a$$

is such that

$$\exists_{v_{\ell} \in V_D} \zeta_D \not\subseteq \Phi_{\perp}$$

For each $D \in \mathcal{S}_k$ and $1 \leq i \leq s$, we define another set $\mathcal{C}_{D,i}$ which intuitively corresponds to all ways that a cause of type φ_i can be present in the data without contradicting the chosen set D . Formally, $\mathcal{C}_{D,i}$ contains all sets E of (in)equalities which satisfy the following conditions:

- for every $1 \leq \ell \leq k$ and $1 \leq j \leq n$, either $v_{\ell} = x'_j$ or $v_{\ell} \neq x'_j$ belongs to E
- for every $1 \leq j < \ell \leq k$, either $x'_j = x'_{\ell}$ or $x'_j \neq x'_{\ell}$ is in E
- for every $1 \leq j \leq k$ and $c \in \text{CN}_{\mathcal{T}}$, either $x'_j = c$ or $x'_j \neq c$ is in E
- the query

$$\gamma_{D,i,E} = \varphi_i[x_{\ell} \mapsto x'_{\ell}] \wedge \bigwedge_{\epsilon \in E} \epsilon$$

is such that

$$\exists v_1, \dots, v_k, x'_1 \dots x'_n \zeta_D \wedge \gamma_{D,i,E} \not\subseteq \Phi_{\perp}$$

We can then define the query χ_D as follows:

$$\chi_D = \exists v_1, \dots, v_k \zeta_D \wedge \neg \left(\bigvee_{1 \leq i \leq s} \bigvee_{E \in \mathcal{C}_{D,i}} \exists x'_1 \dots x'_n \gamma_{D,i,E} \right)$$

Here the query ζ_D serves to check that the set of atoms D is instantiated in the data. The query inside the negation is verified if there some φ_i which is made true in the data in such a way that it does not conflict with the instantiation of D . Thus, the negated query will hold just in the case that every cause for q is contradicted by the chosen set of facts represented by D .

This completes the definition of the query $\Psi_{k\text{-def}}$. Using arguments similar to those given for the k -support-semantic, one can show that $\Psi_{k\text{-def}}$ is an FO-rewriting of q relative to \mathcal{T} under the k -defeater-semantic.

B Proofs for Section 6

Theorem 4 *In DL-Lite, instance checking under the k -support semantics is in NLSpace-complete w.r.t. combined complexity, for every $k \geq 1$. The same holds for the k -defeater semantics, for every $k \geq 0$.*

Proof. The lower bounds stem from the NLSpace-completeness of instance checking under the classical semantics [2]. For the upper bounds, let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a *DL-Lite* KB and q be an IQ. Note that because \mathcal{T} is a *DL-Lite* ontology, every minimal \mathcal{T} -support for q in \mathcal{A} contains a single fact, and every minimal \mathcal{T} -inconsistent subset of \mathcal{A} contains at most 2 facts. We first prove the upper bound for the k -support semantics, by giving an NLSpace procedure for deciding whether $\mathcal{T}, \mathcal{A} \models_{k\text{-supp}} q$. The algorithm begins by guessing k facts $\alpha_1, \dots, \alpha_k$ from \mathcal{A} . By using a binary encoding, we need only logarithmic space to store this guess. Next, for each fact α_i , we check whether $\{\alpha_i\}$ is \mathcal{T} -consistent and $\langle \mathcal{T}, \{\alpha_i\} \rangle \models q$, outputting no if either condition fails to hold. Both checks can be done in NLSpace, since satisfiability and instance checking in *DL-Lite* are NLSpace-complete w.r.t. combined complexity. Finally we must verify that every repair $R \in \text{Rep}(\mathcal{K})$ contains some α_i . Because minimal \mathcal{T} -inconsistent subsets contain at most two facts, it follows that there exists a repair which *does not* contain any α_i just in the case that there is a subset U of \mathcal{A} of cardinality at most k which is \mathcal{T} -consistent and such that $\langle \mathcal{T}, U \cup \{\alpha_i\} \rangle \models \perp$ for every α_i . Thus, we can decide in NLSpace whether such a repair exists by guessing such a subset U and checking whether it satisfies the required conditions. Using the well-known result that $\text{NLSpace} = \text{coNLSpace}$ [12, 20], we obtain a NLSpace procedure for testing whether no such subset exists. If the check succeeds, we return yes, else no. It is easy to see that the described procedure returns yes if and only if $\mathcal{T}, \mathcal{A} \models_{k\text{-supp}} q$.

Now we turn to the upper bound for the k -defeater semantics. We give a NLSpace algorithm for checking whether $\langle \mathcal{T}, \mathcal{A} \rangle \not\models_{k\text{-def}} q$, then use $\text{NLSpace} = \text{coNLSpace}$ to obtain an NLSpace algorithm for the complement. The first step is to guess a subset S of \mathcal{A} of cardinality at most k . We then test (in NLSpace) whether S is \mathcal{T} -consistent, and return no if not. Next we must verify that $S \cup \{\alpha\}$ is \mathcal{T} -inconsistent for every minimal \mathcal{T} -support $\{\alpha\}$ of q , returning yes if this is the case, and no otherwise. This can be done by defining a NLSpace procedure for the complementary problem, which works by guessing an assertion $\alpha \in \mathcal{A}$ and verifying that $\{\alpha\}$ is a \mathcal{T} -support of q and that $S \cup \{\alpha\}$ is \mathcal{T} -consistent.

Theorem 5 *Instance checking in $\text{DL-Lite}_{\text{Horn}}$ is coNP-complete w.r.t. combined complexity under the IAR semantics, coNP-hard w.r.t. combined complexity under k -support*

semantics, and NP-complete w.r.t. combined complexity under both the brave semantics and k -defeater semantics.

Proof. We begin with the IAR semantics. For the coNP lower bound, we give a reduction from UNSAT. Let $\varphi = c_1 \wedge \dots \wedge c_n$ be a propositional CNF over variables x_1, \dots, x_m . Consider the $DL\text{-}Lite_{Horn}$ KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ with

$$\begin{aligned} \mathcal{T} = & \{T_i \sqsubseteq C_j \mid x_i \in c_j\} \cup \{F_i \sqsubseteq C_j \mid \neg x_i \in c_j\} \cup \\ & \{T_i \sqcap F_i \sqsubseteq \perp \mid 1 \leq i \leq m\} \cup \{A \sqcap C_1 \sqcap \dots \sqcap C_n \sqsubseteq \perp\} \end{aligned}$$

and $\mathcal{A} = \{A(a)\} \cup \{T_i(a), F_i(a) \mid 1 \leq i \leq m\}$. We aim to show that $\langle \mathcal{T}, \mathcal{A} \rangle \not\models_{IAR} A(a)$ if and only if the formula φ is satisfiable. For the first direction, note that if $\langle \mathcal{T}, \mathcal{A} \rangle \not\models_{IAR} A(a)$, then $A(a) \notin \bigcap_{\mathcal{B} \in \text{Rep}(\mathcal{K})} \mathcal{B}$, and so there is some $\mathcal{R} \in \text{Rep}(\mathcal{K})$ such that $A(a) \notin \mathcal{R}$. Since \mathcal{R} is a repair, it is a maximal \mathcal{T} -consistent subset of \mathcal{A} . It follows from the maximality of \mathcal{R} and the fact that $A(a) \notin \mathcal{R}$ that $\mathcal{R} \cup \{A(a)\}$ is \mathcal{T} -inconsistent. By considering the inclusions in \mathcal{T} , we see that the only possible way for $A(a)$ to provoke a contradiction is for each $C_i(a)$ to be implied by $\langle \mathcal{T}, \mathcal{R} \rangle$. Next we note that the consistency of \mathcal{R} with \mathcal{T} means that for every $1 \leq i \leq m$, at most one of $T_i(a)$ and $F_i(a)$ is present in \mathcal{R} . We can thus find a valuation⁷ V of the variables x_1, \dots, x_m such that $V(x_i) = 1$ whenever $T_i(a) \in \mathcal{R}$ and $V(x_i) = 0$ whenever $F_i(a) \in \mathcal{R}$. Then we remark that $\langle \mathcal{T}, \mathcal{R} \rangle \models C_j(a)$ implies that there is either some $T_i(a) \in \mathcal{R}$ with $T_i \sqsubseteq C_j \in \mathcal{T}$ (hence $x_i \in c_j$) or some $F_i(a) \in \mathcal{R}$ with $F_i \sqsubseteq C_j \in \mathcal{T}$ (hence $\neg x_i \in c_j$). It follows that every clause c_j is satisfied by V , so φ is satisfiable. For the other direction, suppose φ is satisfiable, and let V be a satisfying valuation. Then define a subset \mathcal{R} of \mathcal{A} which contains every $T_i(a)$ such that $V(x_i) = 1$, and every $F_i(a)$ such that $V(x_i) = 0$. Clearly, \mathcal{R} is \mathcal{T} -consistent, and it follows from our choice of V that $\langle \mathcal{T}, \mathcal{R} \rangle \models C_j(a)$ for every $1 \leq j \leq n$. This means that no assertion in $\mathcal{A} \setminus \mathcal{R}$ can be added to \mathcal{R} while preserving consistency. We have thus shown that \mathcal{R} is a repair of $\langle \mathcal{T}, \mathcal{A} \rangle$, and since $\langle \mathcal{T}, \mathcal{R} \rangle \models A(a)$, we obtain $\langle \mathcal{T}, \mathcal{A} \rangle \not\models_{IAR} A(a)$.

For the coNP upper bound, consider a $DL\text{-}Lite_{Horn}$ KB $\langle \mathcal{T}, \mathcal{A} \rangle$ and an IQ α . To show non-entailment, we guess a set $S \subseteq \mathcal{A}$ together with a subset $C_\beta \subseteq \mathcal{A}$ for each assertion $\beta \in S$. Then we return yes if the following conditions hold:

1. For every $\beta \in S$: C_β is a minimal \mathcal{T} -inconsistent subset of \mathcal{A} which contains β
2. $\langle \mathcal{T}, \mathcal{A} \setminus S \rangle \not\models \alpha$

Point (1) ensures that the intersection of repairs contains only assertions from $\mathcal{A} \setminus S$, and (2) shows that we cannot infer α using only these assertions. Thus, if both (1) and (2) hold, then $\langle \mathcal{T}, \mathcal{A} \rangle \not\models_{IAR} \alpha$. Conversely, if α is not IAR-entailed, then we can satisfy both conditions by letting S contain all assertions which do not belong to all repairs, and letting C_β be any minimal \mathcal{T} -inconsistent subset containing β . Finally, we note that conditions (1) and (2) can both be checked in polynomial time, yielding a NP procedure for the complement of instance checking.

⁷ In the following, we call *valuation* of a set of variables X a function $I : X \rightarrow \{0, 1\}$. A valuation I of X satisfies a positive literal z (where $z \in X$) if $I(z) = 1$, and I satisfies a negative literal $\neg z$ if $I(z) = 0$. Finally, I satisfies a disjunction of literals using variables from X if I satisfies at least one such literal.

Now we show the NP-completeness of instance checking under the brave semantics. To show membership in NP, we can utilize a simple algorithm which guesses some subset S of \mathcal{A} , and checks that it is \mathcal{T} -consistent and that it entails the instance query given \mathcal{T} . The NP lower bound is by reduction from the NP-complete solution existence problem for Horn abduction [9]. An instance of this problem takes the form $P = \langle X, H, M, T \rangle$, where X is a set of propositional variables, H and M are subsets of X , and T is a (consistent) propositional Horn theory. The solution existence problem is to determine whether P admits a solution, that is, a subset $S \subseteq H$ such that $T \cup S$ is consistent and $T \cup S \models m$ for every $m \in M$. Given such a problem P , we construct the following $DL\text{-Lite}_{Horn}$ KB

$$\mathcal{T} = T^* \cup \left\{ \prod_{m \in M} m \sqsubseteq Q \right\} \quad \mathcal{A} = \{h(a) \mid h \in H\}$$

where the propositional variables in X are used as concept names, T^* is the obvious translation of the Horn clauses in T into $DL\text{-Lite}_{Horn}$ inclusions, and Q is a fresh concept name. Then it is easily verified that P has a solution if and only if $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{brave}} Q(a)$.

Finally, we explain how the arguments above yield coNP-hardness (resp. NP-hardness) of instance checking in $DL\text{-Lite}_{Horn}$ under the k -support (resp. k -defeater) semantics. First, we note that the coNP lower bound proof for the IAR semantics works also for the k -support semantics (for every $k > 1$). This is because in the KB used for the reduction, there is a unique \mathcal{T} -support for the IQ $A(a)$ (which is $A(a)$ itself). The NP lower bound for the brave semantics needs a slight modification to be used for the k -defeater semantics. Specifically, in the TBox \mathcal{T} , we replace each inclusion $B_1 \sqcap \dots \sqcap B_n \sqsubseteq \perp$ by the inclusion $B_1 \sqcap \dots \sqcap B_n \sqcap N_1 \sqcap \dots \sqcap N_{k+1} \sqsubseteq \perp$, where N_1, \dots, N_{k+1} are $k + 1$ fresh concept names. We also add to \mathcal{A} the facts $N_1(a), \dots, N_{k+1}(a)$. Note that by construction any minimal \mathcal{T} -inconsistent subset must contain all of the assertions $N_1(a), \dots, N_{k+1}(a)$, and hence has cardinality at least $k + 1$. We next remark that if the Horn abduction problem has a solution $S \subseteq H$, then the set $\{s(a) \mid s \in S\}$ is a \mathcal{T} -support which cannot be contradicted by any subset of \mathcal{A} of size at most k , hence, $\langle \mathcal{T}, \mathcal{A} \rangle \models_{k\text{-def}} Q(a)$. Conversely, if no solution exists, then every repair \mathcal{R} is such that $\langle \mathcal{T}, \mathcal{R} \rangle \not\models Q(a)$. It follows that $\langle \mathcal{T}, \mathcal{A} \rangle \not\models_{\text{brave}} Q(a)$, and hence $\langle \mathcal{T}, \mathcal{A} \rangle \not\models_{k\text{-def}} Q(a)$.

For Theorems 6, 7, and 8, we recall that CQ entailment is NP-complete w.r.t. combined complexity for $DL\text{-Lite}$ and $DL\text{-Lite}_{Horn}$, cf. [2]. To simplify the presentation, we will use this result as a black-box, simply assuming that a polynomial-size certificate for CQ entailment exists and that the validity of such a certificate can be checked in polynomial-time.

Theorem 6 *In $DL\text{-Lite}$, CQ entailment under the k -support semantics is NP-complete w.r.t. combined complexity, for every $k \geq 1$. For both $DL\text{-Lite}$ and $DL\text{-Lite}_{Horn}$, CQ entailment under the k -defeater semantics is NP-complete w.r.t. combined complexity, for every $k \geq 0$.*

Proof. The lower bounds follow trivially from the NP-hardness of CQ entailment without an ontology. For the upper bound for the k -support semantics, fix a $DL\text{-Lite}$ KB

$\langle \mathcal{T}, \mathcal{A} \rangle$ and a CQ q . We define a NP procedure for deciding whether $\langle \mathcal{T}, \mathcal{A} \rangle \models_{k\text{-supp}} q$. First, guess subsets S_1, \dots, S_k of \mathcal{A} , as well as certificates that $\langle \mathcal{T}, S_i \rangle \models q$, for each S_i . Verify in polynomial time that each S_i is \mathcal{T} -consistent and that each certificate is valid. Then check that some S_i is present in each repair by considering all subsets $U \subseteq \mathcal{A}$ with $|U| \leq k$ and testing whether U is \mathcal{T} -consistent but $U \cup S_i$ is \mathcal{T} -inconsistent, for every $1 \leq i \leq k$. If no such a set is found, output yes, otherwise output no. Note that since k is fixed, the described procedure runs in non-deterministic polynomial time. Correctness of the procedure utilizes the property that because \mathcal{T} is formulated in *DL-Lite*, every minimal \mathcal{T} -inconsistent subset has at most 2 facts. Hence, if there is some repair \mathcal{R} which does not contain any S_i , then we can find a subset $U \subseteq \mathcal{R}$ of cardinality at most k which contradicts every S_i ($1 \leq i \leq k$).

To show the upper bound for the k -defeater semantics, fix a *DL-Lite_{Horn}* KB $\langle \mathcal{T}, \mathcal{A} \rangle$ and a CQ q . Let S_1, \dots, S_m be the \mathcal{T} -consistent subsets of \mathcal{A} having cardinality at most k . We describe a NP procedure for deciding whether $\langle \mathcal{T}, \mathcal{A} \rangle \models_{k\text{-def}} q$. The first step is to guess a sequence C_1, \dots, C_m of subsets of \mathcal{A} , together with polynomial certificates that $\langle \mathcal{T}, C_i \rangle \models q$, for each C_i . The second step is to verify that for every $1 \leq i \leq m$, the certificate is valid and $S_i \cup C_i$ is \mathcal{T} -consistent. If all of these checks succeed, the procedure outputs yes, and otherwise it outputs no. As m is polynomial in $|\mathcal{A}|$ (since k is fixed), and both conditions can be verified in polynomial time for *DL-Lite_{Horn}* KBs, we obtain an NP procedure. To show correctness, first assume that the procedure outputs yes. Then for every subset S_i of \mathcal{A} containing at most k facts, there exists a subset $C_i \subseteq \mathcal{A}$ such that $S_i \cup C_i$ is \mathcal{T} -consistent and $\langle \mathcal{T}, C_i \rangle \models q$. It follows that there is no S_i satisfying the conditions in Definition 6, so $\langle \mathcal{T}, \mathcal{A} \rangle \models_{k\text{-def}} q$. For the other direction, suppose that $\langle \mathcal{T}, \mathcal{A} \rangle \models_{k\text{-def}} q$. Then for every subset S_i of \mathcal{A} of cardinality at most k , there is some \mathcal{T} -support C_i for q in \mathcal{A} such that $S_i \cup C_i$ is \mathcal{T} -consistent. Thus, in the first step of the procedure, we can guess these sets C_i , together with valid certificates that $\langle \mathcal{T}, C_i \rangle \models q$. Since the guessed sets and certificates satisfy the required conditions, the algorithm will return yes.

Theorem 7 *In *DL-Lite_{Horn}*, CQ entailment under k -support semantics is $\Delta_2^P[O(\log n)]$ -complete w.r.t. combined complexity, for every $k \geq 1$.*

Proof. Since the upper bound was explained in the body of the paper, we focus here on the lower bound, which is by reduction from the $\Delta_2^P[O(\log n)]$ -complete Parity(SAT) problem, cf. [21, 10]. We recall that an instance of this problem consists of a sequence $\varphi_1, \dots, \varphi_n$ of propositional formulas, and the objective is to determine whether the number of satisfiable formulas in $\varphi_1, \dots, \varphi_n$ is odd. It is known that it can be assumed without loss of generality that the φ_i are 3CNF over disjoint sets of variables, such that φ_{i+1} is unsatisfiable whenever φ_i is unsatisfiable. Because of the latter condition, the problem reduces to determining the existence of an odd integer k such that φ_k is satisfiable and φ_{k+1} is unsatisfiable.

Let $\Phi = \varphi_1, \dots, \varphi_n$ be a Parity(SAT) instance satisfying the aforementioned restrictions, where each φ_i is a 3CNF over the variables $x_{i,1}, \dots, x_{i,g_i}$ composed of m_i clauses $c_{i,1}, \dots, c_{i,m_i}$, with each clause $c_{i,j}$ of the form $\ell_{i,j}^1 \vee \ell_{i,j}^2 \vee \ell_{i,j}^3$. We use $v(\ell_{i,j}^u)$ to denote the variable of literal $\ell_{i,j}^u$, i.e., $v(\ell_{i,j}^u) = \ell_{i,j}^u$ if $\ell_{i,j}^u$ is a positive literal, and $v(\ell_{i,j}^u) = x_{i,h}$ if $\ell_{i,j}^u = \neg x_{i,h}$.

We define a $DL\text{-Lite}_{Horn}$ ontology \mathcal{T} which is similar to that used in the proof of Theorem 5, except that now we must include axioms for each of the formulas φ_i :

$$\begin{aligned}\mathcal{T} = & \{T_{i,h} \sqsubseteq C_{i,j} \mid x_{i,h} \in c_{i,j}\} \cup \\ & \{F_{i,h} \sqsubseteq C_{i,j} \mid \neg x_{i,h} \in c_{i,j}\} \cup \\ & \{T_{i,h} \sqcap F_{i,h} \sqsubseteq \perp \mid 1 \leq i \leq n, 1 \leq h \leq g_i\} \cup \\ & \{A \sqcap C_{i,1} \sqcap \dots \sqcap C_{i,m_i} \sqsubseteq \perp \mid 1 \leq i \leq n\}\end{aligned}$$

We define the dataset \mathcal{A} in stages. First, for every pair $(e, i) \in [1, n]^2$, we let $\mathcal{F}_{e,i}$ be the dataset consisting of the following facts:

$$\begin{aligned}& \{R(a_i^e, c_{i,j}^{e,V}), L_j^1(c_{i,j}^{e,V}, V(v(\ell_{i,j}^1))), \\ \bigcup_{1 \leq j \leq m_i} & L_j^2(c_{i,j}^{e,V}, V(v(\ell_{i,j}^2))), L_j^3(c_{i,j}^{e,V}, V(v(\ell_{i,j}^3))) \mid \\ & V \text{ a valuation of } \{v(\ell_{i,j}^1), v(\ell_{i,j}^2), v(\ell_{i,j}^3)\} \text{ satisfying } c_{i,j}\}\end{aligned}$$

and let $\mathcal{G}_{e,i}$ be as follows:

$$\bigcup_{1 \leq j \leq m_i} \{R(a_i^e, c_{i,j}^e)\} \cup \{L_j^h(c_{i,j}^e, 0) \mid 1 \leq h \leq 3\}$$

Next, for every odd $e \in [1, n]$, we define the dataset \mathcal{A}_e as:

$$\begin{aligned}& \{A(a_1^e)\} \cup \{T_{e+1,h}(a_1^e), F_{e+1,h}(a_1^e) \mid 1 \leq h \leq g_{e+1}\} \cup \\ & \{N(a_i^e, a_{i+1}^e) \mid 1 \leq i < n\} \cup \bigcup_{1 \leq i \leq e} \mathcal{F}_{e,i} \cup \bigcup_{e < i \leq n} \mathcal{G}_{e,i}\end{aligned}$$

Then \mathcal{A} is defined as the disjoint union of the \mathcal{A}_e . Finally, we define the CQ q as follows

$$\begin{aligned}A(y_1) \wedge \bigwedge_{i=1}^{n-1} N(y_i, y_{i+1}) \wedge \\ \bigwedge_{i=1}^n \bigwedge_{j=1}^{m_i} \left(R(y_i, w_{i,j}) \wedge \bigwedge_{h=1}^3 L_j^h(w_{i,j}, v(\ell_{i,j}^h)) \right)\end{aligned}$$

where all variables are existentially quantified. Note that \mathcal{T} , \mathcal{A} , and q can be constructed in polynomial time in Φ .

Claim: $\langle \mathcal{T}, \mathcal{A} \rangle \models_{IAR} q$ iff there is some odd f such that φ_f is satisfiable and φ_{f+1} is unsatisfiable.

For the first part of the claim, suppose that $\langle \mathcal{T}, \mathcal{A} \rangle \models_{IAR} q$. Then $\langle \mathcal{T}, \mathcal{B} \rangle \models q$, where \mathcal{B} is the intersection of the repairs of $\langle \mathcal{T}, \mathcal{A} \rangle$. We first remark that none of the relations appearing in the query q occurs on the right-hand-side of an inclusion in \mathcal{T} . Thus, $\langle \mathcal{T}, \mathcal{B} \rangle \models q$ and $\langle \mathcal{T}, \mathcal{B} \rangle \not\models \perp$ implies that $\langle \emptyset, \mathcal{B} \rangle \models q$. Hence, we can find a function π mapping the variables in q to constants in \mathcal{B} such that (i) if $D(u) \in q$, then $D(\pi(u)) \in \mathcal{B}$, and (ii) if $S(u, u') \in q$, then $S(\pi(u), \pi(u')) \in \mathcal{B}$. Since $A(\pi(y_1)) \in \mathcal{B}$,

we must have $\pi(y_1) = a_f^1$ for some odd $f \in [1, n]$. The presence of $A(a_f^1)$ in \mathcal{B} implies that φ_{f+1} is unsatisfiable (cf. proof of Theorem 5). Using the fact that $\pi(y_1) = a_f^1$ and that $N(\pi(y_i), \pi(y_{i+1})) \in \mathcal{B}$ for every $1 \leq i < n$, we can infer that $\pi(y_i) = a_f^i$, for every $1 \leq i \leq n$. Then for every $1 \leq j \leq m_f$, we have $R(\pi(y_f), \pi(w_{f,j})) \in \mathcal{B}$, and so we can find a valuation V_j of $\{v(\ell_{f,j}^1), v(\ell_{f,j}^2), v(\ell_{f,j}^3)\}$ which satisfies $c_{f,j}$ such that $\pi(w_{f,j}) = c_{f,j}^{f, V_j}$. Note that for each variable $x_{f,h}$, either $\pi(x_{f,h}) = 0$ or $\pi(x_{f,h}) = 1$, and so if $c_{f,j}$ and $c_{f,j'}$ both use variable $x_{f,h}$, then V_j and $V_{j'}$ must agree on the value of $x_{f,h}$. It follows that the V_j together define a valuation of $x_{f,1}, \dots, x_{f,g_f}$ which satisfies $\varphi_f = c_{f,1} \wedge \dots \wedge c_{f,g_f}$. We have thus found an odd f such that φ_f is satisfiable and φ_{f+1} unsatisfiable.

To prove the second half of the claim, suppose that f is odd, φ_f is satisfiable, and φ_{f+1} is unsatisfiable. Let \mathcal{B} be the intersection of the repairs of $\langle \mathcal{T}, \mathcal{A} \rangle$. Since φ_f is satisfiable, the conditions we imposed on Φ mean that all φ_i with $1 \leq i < f$ are also satisfiable. For every $1 \leq i \leq f$, let V_i be a valuation of $x_{i,1}, \dots, x_{i,g_i}$ which satisfies φ_i . We now use the valuations V_i to define a function π mapping variables in q to constants in \mathcal{B} :

- for every $1 \leq i \leq n$: $\pi(y_i) = a_i^f$
- for every $1 \leq i \leq f$ and $1 \leq j \leq m_i$: $\pi(w_{i,j}) = c_{i,j}^{f, V_{i,j}}$, where $V_{i,j}$ is the restriction of the valuation V_i to the variables in clause $c_{i,j}$
- for every $1 \leq i \leq f$ and $1 \leq h \leq g_i$: $\pi(x_{i,h}) = V_i(x_{i,h})$
- for every $f < i \leq n$ and $1 \leq j \leq m_i$: $\pi(w_{i,j}) = c_{i,j}^f$
- for every $f < i \leq n$ and $1 \leq h \leq g_i$: $\pi(x_{i,h}) = 0$

Since φ_{f+1} is unsatisfiable, $A(a_f^1)$ does not belong to any minimal \mathcal{T} -inconsistent subset (again, we refer to proof of Theorem 5 for the argument), and so $A(a_f^1)$ belongs to \mathcal{B} . We thus have $A(\pi(y_1)) \in \mathcal{B}$. Using the definition of \mathcal{A} , it is straightforward to show that the image of each of the other atoms in q under mapping π belongs to \mathcal{B} . We thus have $\langle \emptyset, \mathcal{B} \rangle \models q$, and hence $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{IAR}} q$.

Finally, we note that the above reduction also works if we replace the IAR semantics by the k -support semantics, for any $k > 1$.

Theorem 8 *For DL-Lite and DL-Lite_{Hor}, instance checking (resp. CQ entailment) under the CQA semantics is coNP-complete (resp. Π_2^p -complete) w.r.t. combined complexity.*

Proof. The coNP upper bound for instance checking is obtained by guessing subset of the dataset and verifying in polynomial time that it is a repair and that it does not entail the instance query. For the lower bound, we note that instance checking under the CQA semantics is known to be coNP-hard already w.r.t. data complexity, cf. [13]. Membership in Π_2^p for CQ entailment is again shown by guessing a subset of the dataset and verifying that it is a repair and does not entail the query. Since CQ entailment under classical semantics is NP-complete w.r.t. combined complexity, the Π_2^p upper bound follows.

To prove Π_2^p -hardness for CQ entailment, we reduce the problem of validity of 2-QBF formulas to our problem. Let Φ be a 2-QBF formula of the form

$$\Phi = \forall x_1, \dots, x_m \exists y_1, \dots, y_p \varphi$$

where $\varphi = \bigwedge_{i=1}^n c_i$ is a 3-CNF formula over the variables $x_1, \dots, x_m, y_1, \dots, y_p$, where every c_i is a clause of the form $\ell_i^1 \vee \ell_i^2 \vee \ell_i^3$. As in the proof of Theorem 7, we use $v(\ell_i^j)$ to denote the variable of literal ℓ_i^j .

We define a *DL-Lite* KB $\langle \mathcal{T}, \mathcal{A} \rangle$ and CQ q . The ontology \mathcal{T} is as follows:

$$\{\exists GX_i \sqsubseteq \neg \exists GX_i^- \mid 1 \leq i \leq m\}$$

For the dataset \mathcal{A} , we take the following set of facts:

$$\begin{aligned} & \bigcup_{i=1}^n \{L_i^1(c_i^V, I(v(\ell_i^1))), L_i^2(c_i^V, I(v(\ell_i^2))), L_i^3(c_i^V, I(v(\ell_i^3)))\} \mid \\ & I \text{ is a valuation of } \{v(\ell_i^1), v(\ell_i^2), v(\ell_i^3)\} \text{ satisfying } c_i \\ & \cup \{GX_i(0, 1), GX_i(1, 0) \mid 1 \leq i \leq m\} \end{aligned}$$

and for the query q , we use:

$$\bigwedge_{i=1}^n \bigwedge_{h=1}^3 L_i^h(w_i, v(\ell_i^h)) \wedge \bigwedge_{i=1}^m GX_i(x_i, z_i)$$

with all variables existentially quantified. Clearly, \mathcal{T} , \mathcal{A} , and q can all be constructed in time polynomial in Φ .

Claim: Φ is valid if and only if $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{CQA}} q$.

To show the first part of the claim, suppose that Φ is valid. Let \mathcal{R} be a repair of $\langle \mathcal{T}, \mathcal{A} \rangle$. We remark that the minimal \mathcal{T} -inconsistent subsets of $\langle \mathcal{T}, \mathcal{A} \rangle$ are of the form $\{GX_i(0, 1), GX_i(1, 0)\}$ (for $1 \leq i \leq m$). Thus, the repair \mathcal{R} is obtained by removing either $GX_i(0, 1)$ or $GX_i(1, 0)$ (but not both) for every $1 \leq i \leq m$. Define a valuation V_X of the variables $\{x_1, \dots, x_m\}$ by setting $V_X(x_i) = 0$ if $GX_i(0, 1) \in \mathcal{R}$ and $V_X(x_i) = 1$ if $GX_i(1, 0) \in \mathcal{R}$. Because Φ is valid, there exists a valuation V_{XY} of the variables $\{x_1, \dots, x_m, y_1, \dots, y_p\}$ which extends V_X and satisfies every clause in φ . Define a function π mapping the variables in q to the constants in \mathcal{R} , as follows:

- for every $1 \leq i \leq n$: $\pi(w_i) = c_i^{V_i}$ where V_i is the restriction of V_{XY} to variables appearing in clause c_i
- for every $1 \leq i \leq m$: $\pi(x_i) = V_{XY}(x_i)$
- for every $1 \leq i \leq p$: $\pi(y_i) = V_{XY}(y_i)$
- for every $1 \leq i \leq m$: $\pi(z_i) = 0$ if $\pi(x_i) = 1$, and $\pi(z_i) = 1$ if $\pi(x_i) = 0$

It is easily verified that for every atom α in q , the image of α under π belongs to the repair \mathcal{R} . It follows that $\langle \emptyset, \mathcal{R} \rangle \models q$, hence $\langle \mathcal{T}, \mathcal{R} \rangle \models q$. We thus obtain $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{CQA}} q$.

To complete the proof of the claim, suppose that $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{CQA}} q$. Let V_X be a valuation of $\{x_1, \dots, x_m\}$. Consider the set of facts \mathcal{R} defined as

$$\mathcal{A} \setminus (\{GX_i(1, 0) \mid V_X(x_i) = 0\} \cup \{GX_i(0, 1) \mid V_X(x_i) = 1\})$$

It is easy to see that \mathcal{R} is a repair of $\langle \mathcal{T}, \mathcal{A} \rangle$. Since $\langle \mathcal{T}, \mathcal{A} \rangle \models_{\text{CQA}} q$, we must have $\langle \mathcal{T}, \mathcal{R} \rangle \models q$. We remark that the ontology \mathcal{T} does not contain any positive inclusions. Consequently, $\langle \mathcal{T}, \mathcal{R} \rangle \models q$ and $\langle \mathcal{T}, \mathcal{R} \rangle \not\models \perp$ implies that $\langle \emptyset, \mathcal{R} \rangle \models q$. We can thus find

a function π mapping variables in q to constants in \mathcal{R} such that $S(\pi(u), \pi(u')) \in \mathcal{R}$ for every atom $S(u, u') \in q$. Define a valuation V_{XY} of $\{x_1, \dots, x_m, y_1, \dots, y_p\}$ by setting $V_{XY}(x_i) = \pi(x_i)$ and $V_{XY}(y_i) = \pi(y_i)$. Note that for every $1 \leq i \leq m$, $GX_i(\pi(x_i), \pi(z_i)) \in \mathcal{R}$ implies that $\pi(x_i) = V_X(x_i)$, so V_X and V_{XY} coincide on the value of x_i . It follows that V_{XY} extends the valuation V_X . To show that V_{XY} satisfies the formula φ , consider some clause c_i . We must have $\pi(w_i) = c_i^{V_i}$ for some valuation V_i of the variables $\{v(\ell_i^1), v(\ell_i^2), v(\ell_i^3)\}$ which satisfies c_i . For each $1 \leq h \leq 3$, we have that $L_i^h(\pi(w_i), V_i(v(\ell_i^h))) \in \mathcal{R}$. It follows that V_{XY} coincides with V_i on the variables $\{v(\ell_i^1), v(\ell_i^2), v(\ell_i^3)\}$, which means that V_{XY} satisfies clause c_i . We have thus shown that every valuation V_X of the variables $\{x_1, \dots, x_m\}$ can be extended to a valuation of $\{x_1, \dots, x_m, y_1, \dots, y_p\}$ which satisfies φ . We can thus conclude that Φ is valid.