Data Complexity of Querying Description Logic Knowledge Bases under Cost-Based Semantics

Meghyn Bienvenu[1], **Quentin Manière**[2, 3]
[1]Université de Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, Talence, France
[2]Department of Computer Science, Leipzig University, Germany
[3]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany
meghyn.bienvenu@labri.fr, quentin.maniere@uni-leipzig.de

Abstract

In this paper, we study the data complexity of querying inconsistent weighted description logic (DL) knowledge bases under recently-introduced cost-based semantics. In a nutshell, the idea is to assign each interpretation a cost based upon the weights of the violated axioms and assertions, and certain and possible query answers are determined by considering all (resp. some) interpretations having optimal or bounded cost. Whereas the initial study of cost-based semantics focused on DLs between $\mathcal{EL}_\bot$ and $\mathcal{ALCO}$, we consider DLs that may contain inverse roles and role inclusions, thus covering prominent DL-Lite dialects. Our data complexity analysis goes significantly beyond existing results by sharpening several lower bounds and pinpointing the precise complexity of optimal-cost certain answer semantics (no non-trivial upper bound was known). Moreover, while all existing results show the intractability of cost-based semantics, our most challenging and surprising result establishes that if we consider DL-Lite$^{\mathcal{H}}_{\text{bool}}$ ontologies and a fixed cost bound, certain answers for instance queries and possible answers for conjunctive queries can be computed using first-order rewriting and thus enjoy the lowest possible data complexity (AC^0).

Extended version — http://arxiv.org/abs/2511.07095

1 Introduction

Ontology-mediated query answering (OMQA) has been extensively studied within the KR and database communities as a means of improving data access by exploiting semantic information provided by an ontology (Poggi et al. 2008; Bienvenu and Ortiz 2015; Xiao et al. 2018). Ontologies are typically formulated in decidable fragments of first-order logic (FO), with description logics (DLs) being a popular choice (Baader et al. 2017). Given an ontology (or TBox in DL parlance) $\mathcal{T}$, a dataset (or ABox) $\mathcal{A}$, and a query $q(\vec{x})$, the OMQA task boils down to finding the certain answers, i.e. tuples of constants $\vec{a}$ for which the instantiated query $q(\vec{a})$ is entailed from the knowledge base (KB) $(\mathcal{T}, \mathcal{A})$. Observe that if the input KB is inconsistent, every answer tuple is trivially a certain answer, so OMQA trivializes.

A prominent approach to tackling this issue is to adopt alternative inconsistency-tolerant semantics in order to be able to extract meaningful information from inconsistent KBs, cf. (Lembo et al. 2010) and surveys (Bienvenu and Bourgaux 2016; Bienvenu 2020). Many of these semantics are based upon repairs, defined as inclusion-maximal consistent subsets of the ABox. For example, the AR semantics considers the query answers that hold in all repairs, while the brave semantics returns those answers holding in at least one repair. Note that the line of work on repair-based semantics targets scenarios in which the TBox axioms are deemed fully reliable, so inconsistencies derive solely from errors in the ABox. However, in practice, it can be useful to allow for TBox axioms which typically hold but may admit rare exceptions. Such 'soft' ontology axioms can be addressed qualitatively, using generalized notions of repair that have been proposed for existential rule ontologies (Eiter, Lukasiewicz, and Predoiu 2016), or employing non-monotonic extensions of DLs that support defeasible axioms cf. (Bonatti, Lutz, and Wolter 2009; Giordano et al. 2013; Britz et al. 2021). Another option is to adopt a quantitative approach, using the recently proposed cost-based semantics for DL KBs (Bienvenu, Bourgaux, and Jean 2024), henceforth referred to as (BBJ 2024) for succinctness.

In a nutshell, the idea is to annotate axioms and assertions with (possibly infinite) weights, which are used to assign a cost to each interpretation based upon the weights of the violated axioms and assertions (and taking into account also the number of violations of each TBox axiom). To query the KB, we may choose either to consider the set of interpretations achieving the optimal cost, or we may fix a cost bound k and consider all interpretations having cost at most k. We can then define the sets of certain and possible answers as those answers that hold respectively in all or some interpretation of optimal cost or bounded cost. As noted in (BBJ 2024), the optimal-cost certain answer semantics generalizes both the classical certain answer semantics and the AR semantics based upon weighted ABox repairs. Increasing the cost bound k beyond the optimal cost allows one to identify answers that are robust in the sense that they hold not only for the optimal-cost interpretations. Optimal- and bounded-cost possible answers generalize query satisfiability and can serve to compare candidate answers based upon their incompatibility with the KB.

The computational complexity of querying inconsistent weighted KBs under cost-based semantics was investigated in (BBJ 2024). Five central decision problems were considered: bounded-cost satisfiability (does there exist an interpretation with cost at most k?) plus query entailment under the four cost-based semantics. The complexity analysis was fairly comprehensive, considering both combined and data complexity, conjunctive and instance queries, and DLs ranging from the lightweight DL $\mathcal{EL}_\bot$ to the expressive DL $\mathcal{ALCO}$. One important question that was left open, however, was the data complexity of optimal-cost certain semantics (arguably the most useful of the semantics), for which no non-trivial upper bound was provided. Moreover, as the considered DLs allow neither inverse roles nor role inclusions, they do not yield any results for DLs of the DL-Lite family (Calvanese et al. 2007), which are the most commonly utilized in the context of OMQA.

The preceding considerations motivate us to embark on a more detailed data complexity analysis of cost-based semantics, considering various DL-Lite dialects and expressive DLs up to $\mathcal{ALCHIO}$. The results of our study are summarized in Table 1. A first major contribution, detailed in Section 4, is to provide a Δ_2^p upper bound for the optimal-cost certain and possible semantics, matching an existing lower bound for $\mathcal{EL}_\bot$ and a new lower bound we show for DL-Lite$_{\text{core}}$. This result is obtained by using an intricate quotient construction to establish a small interpretation property, which crucially does not depend on the considered cost. In Section 5, we strengthen a number of existing lower bounds for the bounded-cost semantics by showing that they hold even when for cost bound $k = 1$, as well as providing some new lower bounds for DL-Lite$_{\text{core}}$. Finally, our most challenging and surprising technical result (presented in Section 6) is to show that if we consider DL-Lite$^{\mathcal{H}}_{\text{bool}}$ ontologies and a fixed cost bound, then certain answers for instance queries and possible answers for conjunctive queries can be computed using first-order rewriting and thus enjoy the lowest possible data complexity (AC^0). Detailed proofs can be found in the appendix of the extended version.

2 Preliminaries

We recall the syntax & semantics of description logics (DLs) and refer readers to (Baader et al. 2017) for further details.

Description logic knowledge bases We consider countably infinite sets $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ of *concept names*, *role names*, and *individual names*. An *inverse role* has the form r^-, with $\mathsf{r} \in \mathsf{N_R}$. A *role* is either a role name or inverse role. We use $\mathsf{N_R^\pm} = \mathsf{N_R} \cup \{\mathsf{r}^- \mid \mathsf{r} \in \mathsf{N_R}\}$ for the set of roles. If $\mathsf{r} = \mathsf{s}^-$ is an inverse role, then r^- denotes s.

An $\mathcal{ALCIO}$ *concept* C is built according to the grammar $\mathsf{C}, \mathsf{D} ::= \top \mid \bot \mid \mathsf{A} \mid \{\mathsf{a}\} \mid \neg\mathsf{C} \mid \mathsf{C}\sqcap\mathsf{D} \mid \exists\mathsf{r}.\mathsf{D}$ where $\mathsf{A} \in \mathsf{N_C}$, $\mathsf{r} \in \mathsf{N_R^\pm}$, and $\mathsf{a} \in \mathsf{N_I}$. A concept $\{\mathsf{a}\}$ is called a *nominal*. An $\mathcal{ALCI}$ *concept* is a nominal-free $\mathcal{ALCIO}$ concept. An $\mathcal{EL}$ *concept* is an $\mathcal{ALCI}$ concept that uses neither negation, nor $\bot$, nor inverse roles ($\mathcal{EL}_\bot$ concepts may additionally use $\bot$).

An $\mathcal{ALCHIO}$ *TBox* is a finite set of *concept inclusions (CIs)* $\mathsf{C} \sqsubseteq \mathsf{D}$, where C, D are $\mathcal{ALCIO}$ concepts, and *role inclusions (RIs)* $\mathsf{r} \sqsubseteq \mathsf{s}$, where $\mathsf{r}, \mathsf{s} \in \mathsf{N_R^\pm}$. An $\mathcal{EL}$ *TBox* consists only of CIs between $\mathcal{EL}$ concepts. An *ABox* is a finite set of *concept assertions* $\mathsf{A}(\mathsf{a})$ and *role assertions* $\mathsf{r}(\mathsf{a}, \mathsf{b})$ where $\mathsf{A} \in \mathsf{N_C}$, $\mathsf{r} \in \mathsf{N_R}$, and $\mathsf{a}, \mathsf{b} \in \mathsf{N_I}$. We use $\mathsf{Ind}(\mathcal{A})$ for the set of individual names used in $\mathcal{A}$, and $\mathsf{N_C}(\mathcal{T})$ (resp. $\mathsf{N_R}(\mathcal{T})$) for the set of concept (resp. role) names used in $\mathcal{T}$. An $\mathcal{ALCHIO}$ *knowledge base (KB)* takes the form $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ an $\mathcal{ALCHIO}$ TBox and $\mathcal{A}$ an ABox.

We next introduce the syntax of some DLs of the DL-Lite family. A *basic concept* has the form A or $\exists\mathsf{r}$, with $\mathsf{A} \in \mathsf{N_C}$ and $\mathsf{r} \in \mathsf{N_R^\pm}$. A DL-Lite$^{\mathcal{H}}_{\text{core}}$ TBox is a finite set of CIs of the forms $\mathsf{C} \sqsubseteq \mathsf{D}$ and $\mathsf{C} \sqcap \mathsf{D} \sqsubseteq \bot$, with C, D basic concepts, and RIs $\mathsf{r} \sqsubseteq \mathsf{s}$, with $\mathsf{r}, \mathsf{s} \in \mathsf{N_R^\pm}$. We drop superscript $\cdot^{\mathcal{H}}$ if no role inclusions are admitted, and replace $\cdot_{\text{core}}$ by $\cdot_{\text{bool}}$ to indicate that C, D may be built from basic concepts using $\neg, \sqcap, \sqcup$.

The semantics of DL KBs is defined as usual in terms of interpretations $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ with $\Delta^\mathcal{I}$ the non-empty *domain* and $\cdot^\mathcal{I}$ the *interpretation function*. An interpretation satisfies a CI $\mathsf{C} \sqsubseteq \mathsf{D}$ if $\mathsf{C}^\mathcal{I} \subseteq \mathsf{D}^\mathcal{I}$ and likewise for RIs. It satisfies an assertion $\mathsf{A}(\mathsf{a})$ if $\mathsf{a} \in \mathsf{A}^\mathcal{I}$ and $\mathsf{r}(\mathsf{a}, \mathsf{b})$ if $(\mathsf{a}, \mathsf{b}) \in \mathsf{r}^\mathcal{I}$. This notably requires ABox individuals to be interpreted as themselves, thus enforcing the *standard names assumption* (SNA) for the ABox individuals. The interpretation $\mathcal{I}_\mathcal{A}$ associated with an ABox $\mathcal{A}$ has domain $\mathsf{Ind}(\mathcal{A})$ and interprets concept and role names according to the assertions of $\mathcal{A}$:

$$\mathsf{A}^{\mathcal{I}_\mathcal{A}} := \{\mathsf{a} \mid \mathsf{A}(\mathsf{a}) \in \mathcal{A}\} \quad \mathsf{r}^{\mathcal{I}_\mathcal{A}} := \{(\mathsf{a}, \mathsf{b}) \mid \mathsf{r}(\mathsf{a}, \mathsf{b}) \in \mathcal{A}\}.$$

Given any interpretation $\mathcal{I}$, we use $\mathcal{I}|_\Delta$ to denote the restriction of $\mathcal{I}$ to a subdomain $\Delta \subseteq \Delta^\mathcal{I}$.

Queries First-order queries are given by formulas in first-order logic with equality. Since we wish to query DL KBs, we consider queries whose relational atoms can be either concept atoms $\mathsf{A}(t)$ or role atoms $\mathsf{r}(t, t')$, where $\mathsf{A} \in \mathsf{N_C}$, $\mathsf{r} \in \mathsf{N_R}$, and t, t' are *terms* (variables or individuals). We mostly focus on *conjunctive queries (CQs)* which have the form $\exists\vec{y}\psi$, where ψ is a conjunction of concept and role atoms, and $\vec{y}$ a tuple of variables from ψ. *Instances queries (IQs)* are CQs with a single atom. We also consider *acyclic* and *connected* CQs, meaning that the associated undirected graph (containing an edge $\{t, t'\}$ for each role atom $\mathsf{r}(t, t')$) has these properties. A Boolean CQ (BCQ) is a CQ that has no free variables. We write $\mathcal{I} \models q$ to indicate that an interpretation $\mathcal{I}$ satisfies a BCQ q.

Complexity classes For any syntactic object O such as a TBox, ABox, or query, we use $|O|$ to denote the *size* of O, meaning the encoding of O as a word over a suitable alphabet. Our complexity results concern the well-known complexity classes NP and coNP as well as Δ_2^p (deterministic polynomial time with access to an NP oracle) and AC^0. We omit the formal definition of AC^0, which is based upon circuits, as to understand our results, it suffices to know that it is in AC^0 (in data complexity) to test whether a Boolean first-order query is satisfied in a finite interpretation.

3 Cost-Based Semantics for Weighted KBs

In this section, we recall the definition of cost-based semantics from (BBJ 2024) and introduce the associated reasoning tasks. We also recall and prove some basic properties used in later sections.

Throughout the paper, we work with weighted KBs, whose assertions and axioms are annotated with weights:

	BCS^k, IQA_p^k, CQA_p^k	IQA_c^k	CQA_c^k	BCS, IQA_p, CQA_p	IQA_c, CQA_c	$\text{IQA}_{p,c}^{opt}$, $\text{CQA}_{p,c}^{opt}$
$\mathcal{EL}_\bot$ / $\mathcal{ALCHIO}$	NP ‡ Thm. 1, 3	coNP ‡ Thm. 1, 4	coNP ‡ Thm. 1, 4	NP † Thm. 1	coNP † Thm. 1	Δ_2^p † Thm. 2
DL-Lite$_{\text{core}}$ / DL-Lite$_{\text{bool}}^{\mathcal{H}}$	in AC^0 Thm. 8	in AC^0 Thm. 8	coNP Thm. 1, 7	NP Thm. 1, 5	coNP Thm. 1, 5	Δ_2^p Thm. 2, 6

Table 1: All results are completeness results, unless stated otherwise. †: lower bound from (BBJ 2024). ‡: lower bound for $k \geq 3$ from (BBJ 2024), improved to $k \geq 1$ in the present paper. For CQA, lower bounds already hold for connected acyclic BCQs. Results hold both for binary and unary encoding of weights, except for Δ_2^p-hardness (only for binary encoding).

Definition 1. *A* weighted knowledge base *(WKB)* $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ *consists of a knowledge base* $(\mathcal{T}, \mathcal{A})$ *and a weight function* $\omega : \mathcal{T} \cup \mathcal{A} \mapsto \mathbb{N}_{>0} \cup \{\infty\}$. *We can similarly define* weighted TBoxes *($\mathcal{T}_\omega$) and* weighted ABoxes *($\mathcal{A}_\omega$).*

Intuitively, these weights can be viewed as the penalties incurred for violating assertions and axioms: those having cost 1 are the least reliable, while those assigned maximal weight ∞ should definitely be satisfied.

Interpretations will then be assigned costs based upon the sets of violations of the TBox axioms and ABox assertions. Note that differently from prior work, we also define violations of role inclusions, given by pairs of domain elements.

Definition 2. *Given an interpretation $\mathcal{I}$, the* set of violations of a concept inclusion $\mathsf{B} \sqsubseteq \mathsf{C}$ *in $\mathcal{I}$ is* $vio_{\mathsf{B}\sqsubseteq\mathsf{C}}(\mathcal{I}) = \mathsf{B}^\mathcal{I} \setminus \mathsf{C}^\mathcal{I}$, *the* set of violations of a role inclusion $\mathsf{r} \sqsubseteq \mathsf{s}$ *in $\mathcal{I}$ is* $vio_{\mathsf{r}\sqsubseteq\mathsf{s}}(\mathcal{I}) = \mathsf{r}^\mathcal{I} \setminus \mathsf{s}^\mathcal{I}$, *and the* violations of an ABox $\mathcal{A}$ *in $\mathcal{I}$ are* $vio_\mathcal{A}(\mathcal{I}) = \{\alpha \in \mathcal{A} \mid \mathcal{I} \not\models \alpha\}$.

Definition 3. *Let* $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ *be a WKB. The* cost of an interpretation $\mathcal{I}$ *w.r.t.* $\mathcal{K}_\omega$ *is defined by:*

$$\omega(\mathcal{I}) = \sum_{\tau \in \mathcal{T}} \omega(\tau)|vio_\tau(\mathcal{I})| + \sum_{\alpha \in vio_\mathcal{A}(\mathcal{I})} \omega(\alpha)$$

The optimal cost *of* $\mathcal{K}_\omega$ *is* $optc(\mathcal{K}_\omega) = \mathsf{min}_\mathcal{I}(\omega(\mathcal{I}))$. *A WKB* $\mathcal{K}_\omega$ *is* k-satisfiable *if* $\omega(\mathcal{I}) \leq k$ *for some interpretation* $\mathcal{I}$.

We recall next the four cost-based semantics proposed in (BBJ 2024), which depend on whether one considers interpretations whose cost is less than a provided bound, or the interpretations having optimal cost, and whether the query is required to hold in all or at least one such interpretation.

Definition 4. *Let q be a BCQ,* $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ *a WKB, and k an integer. We say that q is entailed by* $\mathcal{K}_\omega$ *under*

- k-cost bounded certain semantics, *written* $\mathcal{K}_\omega \models_c^k q$, *if* $\mathcal{I} \models q$ *for every interpretation* $\mathcal{I}$ *with* $\omega(\mathcal{I}) \leq k$*;*
- k-cost bounded possible semantics, *written* $\mathcal{K}_\omega \models_p^k q$, *if* $\mathcal{I} \models q$ *for some interpretation* $\mathcal{I}$ *with* $\omega(\mathcal{I}) \leq k$*;*
- opt-cost certain semantics, *written* $\mathcal{K}_\omega \models_c^{opt} q$, *if* $\mathcal{I} \models q$ *for every interpretation* $\mathcal{I}$ *with* $\omega(\mathcal{I}) = optc(\mathcal{K}_\omega)$*;*
- opt-cost possible semantics, *written* $\mathcal{K}_\omega \models_p^{opt} q$, *if* $\mathcal{I} \models q$ *for some interpretation* $\mathcal{I}$ *with* $\omega(\mathcal{I}) = optc(\mathcal{K}_\omega)$.

These semantics extend to non-Boolean CQs in the expected way, e.g. the opt-cost certain answers to a CQ $q(\vec{x})$ *w.r.t.* $(\mathcal{T}, \mathcal{A})_\omega$ *are the tuples* $\vec{a}$ *from* $\mathsf{Ind}(\mathcal{A})$ *s.t.* $\mathcal{K}_\omega \models_c^{opt} q(\vec{a})$.

If the underlying KB is satisfiable, then the certain and possible optimal-cost semantics coincide with query entailment and query satisfiability (or with classical notions of certain and possible answers, in the case of non-Boolean queries). These semantics are thus intended to be used when the underlying KB is inconsistent. The opt-cost certain answers identifies those answers that hold in the interpretations deemed most likely and have been shown to generalize previously considered weight-based repair semantics (BBJ 2024). By considering values of k beyond $optc(\mathcal{K}_\omega)$, we can use the k-cost bounded certain semantics to identify 'robust' answers which hold not only in the optimal-cost interpretations but also in those with close-to-optimal cost. By contrast, the opt-cost and k-cost bounded possible answers can serve to rank candidate answers based upon their degree of incompatibility with the WKB.

We now formalize the decision problems for cost-based semantics investigated in this paper, which differ depending on which cost bound is used and whether it is given as input:

- *Bounded cost satisfiability* (BCS) takes as input a WKB $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ and an integer k and decides whether there exists an interpretation $\mathcal{I}$ with $\omega(\mathcal{I}) \leq k$.
- *k-cost satisfiability* (BCS^k) takes as input a WKB $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ and decides whether there exists an interpretation $\mathcal{I}$ with $\omega(\mathcal{I}) \leq k$.
- *Bounded-cost certain (resp. possible) BCQ entailment* (CQA_c / CQA_p) takes as input a WKB $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$, a BCQ q and an integer k and decides whether $\mathcal{K}_\omega \models_c^k q$ (resp. $\mathcal{K}_\omega \models_p^k q$).
- *k-cost certain (resp. possible) BCQ entailment* (CQA_c^k / CQA_p^k) takes as input a WKB $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ and a BCQ q and decides whether $\mathcal{K}_\omega \models_c^k q$ (resp. $\mathcal{K}_\omega \models_p^k q$).
- *Optimal-cost certain (resp. possible) BCQ entailment* (CQA_c^{opt} / CQA_p^{opt}) takes as input a WKB $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ and a BCQ q and decides if $\mathcal{K}_\omega \models_c^{opt} q$ (resp. $\mathcal{K}_\omega \models_p^{opt} q$).

We will also consider the restrictions of the BCQ entailment problems to the case of instance queries, denoted by IQA_c, IQA_p, IQA_c^k, IQA_p^k, IQA_c^{opt} and IQA_p^{opt} respectively.

We shall study the *data complexity* of the preceding reasoning tasks. For the fixed-cost and optimal-cost decision problems, data complexity is measured with respect to the size of the input weighted ABox, while the size of the weighted TBox and query (if present) are treated as constants. For the bounded-cost problems, we measure complexity w.r.t. the weighted ABox and the input integer. Both the ABox weights and the input integer (if present) are assumed to be encoded in binary.

We conclude the section with some easy lemmas, which establish useful reductions between the decision problems.

Lemma 1. *BCS^k for* DL-Lite$_{\text{core}}$ *(resp. $\mathcal{EL}_\bot$) reduces to IQA^k_p for* DL-Lite$_{\text{core}}$ *(resp. $\mathcal{EL}_\bot$). In particular, the same holds for BCS and IQA_p.*

Lemma 2. *IQA^k_p for* DL-Lite$_{\text{core}}$ *(resp. $\mathcal{EL}_\bot$) reduces to the complement of IQA^{k+1}_c for* DL-Lite$_{\text{core}}$ *(resp. $\mathcal{EL}_\bot$). In particular, IQA_p reduces to the complement of IQA_c.*

Lemma 3. *For every $k \geq 0$, IQA^k_p for* DL-Lite$_{\text{core}}$ *(resp. $\mathcal{EL}_\bot$) reduces to IQA^{k+1}_p for* DL-Lite$_{\text{core}}$ *(resp. $\mathcal{EL}_\bot$) WKBs.*

4 An Upper Bound for the General Case

The aim of this section is to establish the next two theorems:

Theorem 1. *CQA_p (resp. CQA_c) for $\mathcal{ALCHIO}$ is in* NP *(resp. in* coNP*).*

Theorem 2. *CQA^{opt}_p and CQA^{opt}_c for $\mathcal{ALCHIO}$ are in Δ^p_2.*

Starting from an interpretation $\mathcal{I}$ that satisfies (or does not satisfy) the query q of interest, the main technical ingredient is the construction of another interpretation $\mathcal{J}$ that behaves as $\mathcal{I}$ w.r.t. q, whose cost is at most the cost of $\mathcal{I}$, and whose domain has a size polynomially bounded by the size of the input ABox $\mathcal{A}$. This is fairly easy if query satisfaction is to be preserved, that is, for CQA$_p$: one can brutally collapse elements together according to their type using the well-known filtration technique (Baader et al. 2017).

Lemma 4. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})_\omega$ be an $\mathcal{ALCHIO}$ WKB, k an integer, and q a BCQ. If there exists an interpretation $\mathcal{I}$ such that $\omega(\mathcal{I}) \leq k$ and $\mathcal{I} \models q$, then there is an interpretation $\mathcal{J}$ such that $\omega(\mathcal{J}) \leq k$ and $\mathcal{J} \models q$, and whose domain $\Delta^{\mathcal{J}}$ has cardinality bounded polynomially in $|\mathcal{A}|$, with $|\mathcal{T}|$ and $|q|$ treated as constants, and independently from k.*

It becomes more challenging if query non-satisfaction is the property to preserve, that is, to address CQA$_c$. In particular, the solution adopted in (BBJ 2024, Proposition 8) for $\mathcal{ALCO}$ WKBs yields a polynomial bound that depends on the bounded cost k. This makes their technique adequate when the cost is fixed, that is, for CQA^k_c, or if the encoding of k is given in unary, but otherwise does not provide a polynomial upper bound w.r.t. data complexity. We deeply rework the approach, not only to support $\mathcal{ALCHIO}$ WKBs, but also to obtain a cost-independent bound as follows.

Lemma 5. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})_\omega$ be an $\mathcal{ALCHIO}$ WKB, k an integer, and q a BCQ. If there exists an interpretation $\mathcal{I}$ such that $\omega(\mathcal{I}) \leq k$ and $\mathcal{I} \not\models q$, then there is an interpretation $\mathcal{J}$ such that $\omega(\mathcal{J}) \leq k$ and $\mathcal{J} \not\models q$, and whose domain $\Delta^{\mathcal{J}}$ has cardinality that is bounded polynomially in $|\mathcal{A}|$, with $|\mathcal{T}|$ and $|q|$ treated as constants, and independently from k.*

With Lemmas 4 and 5, we obtain the NP and coNP upper bounds in Theorem 1 with standard guess-and-check procedures. For Theorem 2, the Δ^p_2 algorithms proceed similarly but first identify the optimal cost via a binary search, using an exponential bound on the optimal cost (whenever finite).

Now, to prove Lemma 5, we rely on the adaptation of a quotient construction from (Manière 2022, Theorem 8) defined to answer *counting conjunctive queries* over $\mathcal{ALCHI}$ KBs. This construction was already reused by (BBJ 2024) for $\mathcal{ALCO}$ WKBs, and it is not too difficult to handle inverse roles and role inclusions by sticking closer to the original version. From the starting interpretation $\mathcal{I}$, our adaptation differs from theirs as it also takes as a parameter a subset $\mathcal{V} \subseteq \mathcal{T}$ of the considered TBox $\mathcal{T}$. In the constructed interpretation, we violate axioms of $\mathcal{V}$ exactly as in the original interpretation $\mathcal{I}$. Intuitively, one can think of these axioms in $\mathcal{V}$ as so expensive to violate that one cannot do better than in $\mathcal{I}$, while potential violations of axioms from $\mathcal{T} \setminus \mathcal{V}$ can be handled in a more systematic and structured manner. Violations of ABox assertions are easier to control and are preserved exactly as in the original interpretation $\mathcal{I}$. For an interpretation $\mathcal{J}$, we denote $vio_{\mathcal{V}}(\mathcal{J}) := \bigcup_{\tau \in \mathcal{V}} vio_\tau(\mathcal{J})$. Adapting the quotient technique yields the following:

Lemma 6. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ALCHIO}$ KB and q a BCQ. Let $\mathcal{V} \subseteq \mathcal{T}$ be a subset of $\mathcal{T}$ and $\mathcal{I}$ an interpretation such that $\mathcal{I} \not\models q$. There exists a polynomial p independent of $\mathcal{A}$, and an interpretation $\mathcal{J}$ satisfying the following:*

1. $\mathcal{J} \not\models q$;
2. $vio_{\mathcal{A}}(\mathcal{J}) = vio_{\mathcal{A}}(\mathcal{I})$;
3. $\forall \tau \in \mathcal{V}, vio_\tau(\mathcal{J}) \subseteq vio_\tau(\mathcal{I})$;
4. $\left|\Delta^{\mathcal{J}}\right| \leq p(|\mathcal{A}| + |vio_{\mathcal{V}}(\mathcal{I})|)$.

We now explain how to obtain Lemma 5, with an approach inspired from (Lutz and Manière 2024), where our Lemmas 5 and 6 respectively play the role of their Proposition 2 and Lemma 1. Let $\mathcal{K}_\omega = (\mathcal{T}, \mathcal{A})_\omega$ be an $\mathcal{ALCHIO}$ WKB, q a BCQ, k an integer, and $\mathcal{I}$ an interpretation such that $\mathcal{I} \not\models q$. For a given $\mathcal{V} \subseteq \mathcal{T}$, we use $\mathcal{J}_{\mathcal{V}}$ to denote the interpretation obtained by applying Lemma 6 with $\mathcal{V}$ the input set of axioms. We prove that there exists a subset $\mathcal{V} \subseteq \mathcal{T}$ such that (i) the size of $vio_{\mathcal{V}}(\mathcal{I})$ is bounded by a polynomial in $|\mathcal{A}|$ independent of k; and (ii) $\omega(\mathcal{J}_{\mathcal{V}}) \leq k$. To do so, we construct a sequence $\mathcal{V}_0 \subsetneq \mathcal{V}_1 \subsetneq \cdots \subsetneq \mathcal{V}_n \subseteq \mathcal{T}$ of $\mathcal{V}$'s that all satisfy item (i) and with $\mathcal{V}_n$ also satisfying item (ii). Note that $\mathcal{J}_{\mathcal{V}_n}$ is then the desired interpretation for Lemma 5: item (i) plus Point 4 from Lemma 6 gives the polynomial bound on the size of $\mathcal{J}_n$, while item (ii) and Point 1 in Lemma 6 ensure the desired properties w.r.t. the cost and query.

Initialization. Set $\mathcal{V}_0 := \emptyset$, which trivially satisfies item (i).

Induction step. Assume that, for some $i \geq 0$, we have successfully constructed $\mathcal{V}_i$ satisfying item (i); thus we have a polynomial p_i independent of k such that $|vio_{\mathcal{V}}(\mathcal{I})| \leq p_i(|\mathcal{A}|)$. If $\mathcal{V}_i$ also satisfies item (ii), then we are done. Otherwise $\omega(\mathcal{J}_{\mathcal{V}_i}) > k$, and since $\omega(\mathcal{I}) \leq k$, there exists an assertion or axiom τ from $\mathcal{K}$ that is violated at least once more in $\mathcal{J}_{\mathcal{V}_i}$ than in $\mathcal{I}$, *i.e.* $|vio_\tau(\mathcal{J}_{\mathcal{V}_i})| > |vio_\tau(\mathcal{I})|$. Note that due to Point 2 in Lemma 6, it is then clear that $\tau \notin \mathcal{A}$. Similarly, due to Point 3 in Lemma 6, we have $\tau \notin \mathcal{V}_i$. Therefore $\tau \in \mathcal{T} \setminus \mathcal{V}_i$. We set $\mathcal{V}_{i+1} := \mathcal{V}_i \cup \{\tau\}$. It remains to verify that $\mathcal{V}_{i+1}$ satisfies item (i). Note that $vio_{\mathcal{V}_{i+1}}(\mathcal{I}) = vio_{\mathcal{V}_i}(\mathcal{I}) \cup vio_\tau(\mathcal{I})$. The size of $vio_{\mathcal{V}_i}(\mathcal{I})$ is bounded adequately by $p_i(|\mathcal{A}|)$. For the size of $vio_\tau(\mathcal{I})$, recall that by choice of τ we have $|vio_\tau(\mathcal{I})| < |vio_\tau(\mathcal{J}_{\mathcal{V}_i})|$. We brutally bound $|vio_\tau(\mathcal{J}_{\mathcal{V}_i})|$ by $\left|\Delta^{\mathcal{J}_{\mathcal{V}_i}}\right|^2$. By Point 4 in Lemma 6, $\Delta^{\mathcal{J}_{\mathcal{V}_i}}$ has size bounded by $p(|\mathcal{A}| + |vio_{\mathcal{V}_i}(\mathcal{I})|)$, thus by $p(|\mathcal{A}| + p_i(|\mathcal{A}|))$. Overall, the size of $vio_{\mathcal{V}_{i+1}}(\mathcal{I})$ is bounded by $p_{i+1}(|\mathcal{A}|)$ where $p_{i+1}(x) := p_i(x) + (p(x + p_i(x)))^2$ is the desired polynomial independent of k.

Note that this procedure is guaranteed to terminate in at most $|\mathcal{T}|$ steps, which concludes the proof.

5 Lower Bounds

We first refine some existing lower bounds for the fixed-cost decision problems in $\mathcal{EL}_\bot$, then prove the lower bounds for DL-Lite listed in Table 1.

5.1 Lower Bounds in Extensions of $\mathcal{EL}_\bot$

We begin with two lower bounds showing that even if the cost k is fixed to 1, all considered reasoning tasks are NP-complete (or coNP-complete, depending on the task) already for $\mathcal{EL}_\bot$ WKBs. These results notably improve those from (BBJ 2024), where the cost was fixed to 3. Lemma 3 lifts our hardness proof to any fixed $k \geq 1$, and since the case of fixed $k = 0$ coincides with the usual semantics of $\mathcal{EL}$ KBs, the complexity w.r.t. fixed cost is now well understood.

Theorem 3. *For every $k \geq 1$, BCS^k, IQA$_p^k$ and CQA$_p^k$ for $\mathcal{EL}_\bot$ are* NP*-hard.*

Proof sketch. The reduction is from 3-SAT. Given a 3-CNF formula $\phi := \bigwedge_{i=1}^{\ell} \bigvee_{j=1}^{3} l_{i,j}$, where each $l_{i,j}$ is a literal over $v_1, \dots, v_n$, we construct a WKB $(\mathcal{T}, \mathcal{A}_\phi)_\omega$. The ABox $\mathcal{A}_\phi$ contains $\mathsf{False}(\mathsf{a})$, $\mathsf{Bool}(\mathsf{a})$, and the additional assertions:

$$\mathsf{Var}(v_k) \text{ for } 1 \leq k \leq n \qquad \mathsf{clause}(\mathsf{a}, c_i) \text{ for } 1 \leq i \leq \ell$$
$$\mathsf{pos}_j(c_i, v_k) \text{ for } l_{i,j} = v_k \qquad \mathsf{neg}_j(c_i, v_k) \text{ for } l_{i,j} = \neg v_k$$

The TBox $\mathcal{T}$ has the following axioms:

$$\mathsf{Bool} \sqsubseteq \mathsf{True} \quad \mathsf{True} \sqcap \mathsf{False} \sqsubseteq \bot \quad \exists \mathsf{clause}.\mathsf{False} \sqsubseteq \mathsf{True}$$
$$\mathsf{Var} \sqsubseteq \exists \mathsf{val}.\mathsf{Bool} \quad \exists \mathsf{val}.\mathsf{True} \sqsubseteq \mathsf{True} \quad \exists \mathsf{val}.\mathsf{False} \sqsubseteq \mathsf{False}$$
$$\exists \mathsf{pos}_1.\mathsf{False} \sqcap \exists \mathsf{pos}_2.\mathsf{False} \sqcap \exists \mathsf{pos}_3.\mathsf{False} \sqsubseteq \mathsf{False}$$

(the six other combinations...)

$$\exists \mathsf{neg}_1.\mathsf{True} \sqcap \exists \mathsf{neg}_2.\mathsf{True} \sqcap \exists \mathsf{neg}_3.\mathsf{True} \sqsubseteq \mathsf{False}$$

The function ω assigns ∞ to all axioms and assertions, except for $\mathsf{Bool} \sqsubseteq \mathsf{True}$, which has weight 1. One can verify that ϕ is satisfiable iff $(\mathcal{T}, \mathcal{A}_\phi)_\omega$ is 1-satisfiable. $\square$

For the case of IQA$_c^k$, we could use Lemma 2 to directly obtain a coNP-hardness proof for $k \geq 2$. We instead re-adapt the above proof to strengthen the result to every $k \geq 1$. Note that concept disjointness axioms are not even needed here.

Theorem 4. *For every $k \geq 1$, IQA$_c^k$ and CQA$_c^k$ for $\mathcal{EL}$ are* coNP*-hard.*

5.2 Lower Bounds in the DL-Lite Family

We now turn to the DL-Lite family, which inherits the upper bounds from Theorems 1 and 2. We begin with a proof that, when k is allowed to vary, all considered reasoning tasks are NP-hard (resp. coNP-hard) already for DL-Lite$_{\text{core}}$ WKBs.

Theorem 5. *BCS, IQA$_p$ and CQA$_p$ for* DL-Lite$_{\text{core}}$ *are* NP*-hard. IQA$_c$ and CQA$_c$ for* DL-Lite$_{\text{core}}$ *WKBs are* coNP*-hard.*

Note that, by virtue of Lemmas 1 and 2, it suffices to prove that BCS for DL-Lite$_{\text{core}}$ is NP-hard.

Proof sketch. We reduce from 3-COL, that is deciding whether a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is 3-colourable. All axioms of $\mathcal{T}$ are given infinite weight by ω and are as follows:

$$\exists s_i \sqcap \exists t_j \sqsubseteq \bot \;\text{ for } s, t \in \{\mathsf{r}, \mathsf{g}, \mathsf{b}\}, s \neq t, \text{ and } i, j \in \{1, 2\}$$
$$\exists s_1^- \sqcap \exists s_2^- \sqsubseteq \bot \;\text{ for } s \in \{\mathsf{r}, \mathsf{g}, \mathsf{b}\}$$

We choose an orientation $\mathcal{E}'$ of $\mathcal{E}$: for each $\{u, v\} \in \mathcal{E}$, we add either (u, v) or (v, u) in $\mathcal{E}'$. For each $e = (u, v) \in \mathcal{E}'$ and each $s \in \{\mathsf{r}, \mathsf{g}, \mathsf{b}\}$, we add $s_1(u, e)$ and $s_2(v, e)$ in the ABox $\mathcal{A}_\mathcal{G}$. Assertions in $\mathcal{A}_\mathcal{G}$ are given weight 1 by ω. It can be verified that $\mathcal{G} \in$ 3-COL iff $(\mathcal{T}, \mathcal{A}_\mathcal{G})_\omega$ is $4|\mathcal{E}|$-satisfiable. $\square$

For optimal cost semantics, we establish a matching Δ_2^p lower bound. The proof adapts an existing construction from (Bourgaux 2016, Proposition 6.2.4) that establishes Δ_2^p-hardness of query entailment for DL-Lite KBs under preferred repair semantics, by reduction from deciding if a given variable is true in the lexicographically maximum truth assignment satisfying a given satisfiable CNF. We point out that, unlike the other lower bounds listed in Table 1, this result crucially relies upon a binary encoding of weights, intuitively because exponentially large weights are needed to perform lexicographic comparison of satisfying valuations.

Theorem 6. *IQA$_p^{opt}$, IQA$_c^{opt}$, CQA$_p^{opt}$, and CQA$_c^{opt}$ for* DL-Lite$_{\text{core}}$ *are Δ_2^p-hard.*

We now move to the case in which k is fixed. Notice indeed that the lower bound from Theorem 5 strongly relies on a varying k. For the certain semantics, we show that coNP-hardness holds even if k is fixed to 1, if we consider CQs. The proof is strongly inspired by the one of Theorem 4, especially with how to simulate the truth value assignment. The main difference is that we use acyclic CQs to circumvent the lack of nested concepts in DL-Lite$_{\text{core}}$.

Theorem 7. *For every $k \geq 1$, CQA$_c^k$ for* DL-Lite$_{\text{core}}$ *is* coNP*-hard. This holds already for connected acyclic BCQs and without concept disjointness axioms.*

6 Positive Results in the DL-Lite Family

In light of the lower bounds established in the previous section, we can only hope to achieve tractability for the DL-Lite family under a fixed cost (see Theorem 5). Furthermore, under the certain semantics, we established that CQA$_c^k$ answering is coNP-hard (Theorem 7) already for DL-Lite$_{\text{core}}$ and $k = 1$. This leaves us with two promising settings to explore: CQA$_p^k$ and IQA$_c^k$ for DL-Lite$_{\text{core}}$. This section establishes that both reasoning tasks enjoy the lowest possible complexity, that is, AC^0. Furthermore, we can even push this positive result to one of the most expressive logics of the DL-Lite family, namely DL-Lite$_{\text{bool}}^{\mathcal{H}}$.

Theorem 8. *For every integer $k \geq 1$, CQA$_p^k$ and IQA$_c^k$ for* DL-Lite$_{\text{bool}}^{\mathcal{H}}$ *are in* AC^0.

Our approach is based on first-order (FO) rewriting: given the weighted TBox $\mathcal{T}_{\omega_\mathcal{T}}$, query q and cost k, we construct an FO-query q' such that for every weighted ABox $\mathcal{A}_{\omega_\mathcal{A}}$, the following, here stated for CQA$_p^k$, holds:

$$(\mathcal{T}, \mathcal{A})_{\omega_\mathcal{T} \cup \omega_\mathcal{A}} \models_p^k q \quad \text{iff} \quad \mathcal{I}_{\mathcal{A}_{\omega_\mathcal{A}}} \models q'.$$

To make this formulation fully precise, we need to define the FO-interpretation $\mathcal{I}_{\mathcal{A}_{\omega_\mathcal{A}}}$ associated with a weighted ABox $\mathcal{A}_{\omega_\mathcal{A}}$, over which the rewritten query q' is evaluated. We argue that $\mathcal{A}_{\omega_\mathcal{A}}$ can be seen as a usual ABox augmented with extra assertions about the weights. We denote by $\mathcal{A}_k^{\omega_\mathcal{A}}$ the extension of $\mathcal{A}$ with special concept and role assertions that encapsulate the relevant information about weights w.r.t. the fixed cost bound k: if a concept assertion $\omega_{\mathcal{A}}(\mathsf{A}(\mathsf{a})) = n$, then we add the assertion $\mathsf{W}^n_{\mathsf{A}}(\mathsf{a})$ if $n \leq k$, or assertion $\mathsf{W}^{\infty}_{\mathsf{A}}(\mathsf{a})$ if $n > k$. We proceed similarly for each role assertion $\mathsf{r}(\mathsf{a},\mathsf{b})$, adding respectively $\mathsf{w}^n_{\mathsf{r}}(\mathsf{a},\mathsf{b})$ or $\mathsf{w}^{\infty}_{\mathsf{r}}(\mathsf{a},\mathsf{b})$. Notice that we only need to introduce $(k+1)(\mathsf{N_C}(\mathcal{A})+\mathsf{N_R}(\mathcal{A}))$ fresh predicates and that computing $\mathcal{A}^{\omega_{\mathcal{A}}}_k$ from any reasonable representation of $\mathcal{A}_{\omega_{\mathcal{A}}}$ can be seen as a pre-processing step achieved by an AC^0 transducer. As $\mathcal{A}^{\omega_{\mathcal{A}}}_k$ is a usual ABox, its corresponding interpretation $\mathcal{I}_{\mathcal{A}^{\omega_{\mathcal{A}}}_k}$ is well defined and can serve as the desired interpretation $\mathcal{I}_{\mathcal{A}_{\omega_{\mathcal{A}}}}$.

Before explaining how to construct q', we sketch the main argument that allows for such a rewriting to exist and that ensures completeness of the claim (*i.e.* the $\Rightarrow$ direction in the formulation above). It relies on a (very!) small interpretation property: if an interpretation witnesses the desired behaviour w.r.t. the query and within the fixed cost, then there exists one that is completely trivial except on a small domain whose size is bounded by a *constant* w.r.t. the input weighted ABox. The different possibilities to interpret such a constant-size domain can thus all be encapsulated in the rewritten query q'. To facilitate the understanding of the rewriting, we first present this small interpretation property.

6.1 A (Very) Small Interpretation Property

Consider a WKB $\mathcal{K} = (\mathcal{T},\mathcal{A})_{\omega}$, a fixed cost k and a BCQ q. We introduce two distinct notions of types. The first is the usual one in DLs: a 1-*type* t is a subset of $\mathsf{N_C}(\mathcal{T}) \cup \{\exists \mathsf{r} \mid \mathsf{r} \in \mathsf{N}^{\pm}_{\mathsf{R}}(\mathcal{T})\}$. The 1-type of an element $e \in \Delta^{\mathcal{I}}$ is $\mathsf{tp}_{\mathcal{I}}(e) := \{\mathsf{A} \mid e \in \mathsf{A}^{\mathcal{I}}\} \cup \{\exists \mathsf{r} \mid e \in (\exists \mathsf{r})^{\mathcal{I}}, \mathsf{r} \in \mathsf{N}^{\pm}_{\mathsf{R}}\}$. This notion of 1-type captures the basic DL-Lite concepts, and thus, if two elements d and e have the same 1-type, then they violate exactly the same DL-Lite$_{\text{bool}}$ CIs.

The second notion of type is intended to capture the concepts that hold due to the ABox assertions. We define the *ABox type* $\mathsf{tp}_{\mathcal{A}}(\mathsf{a})$ of $\mathsf{a} \in \mathsf{Ind}(\mathcal{A})$ as:

$$\mathsf{tp}_{\mathcal{A}}(\mathsf{a}) := \mathsf{tp}_{\mathcal{I}_{\mathcal{A}}}(\mathsf{a}) \cup \{\exists^{>k}\mathsf{r} \mid \mathcal{A} \models \exists^{>k}\mathsf{r}, \mathsf{r} \in \mathsf{N}^{\pm}_{\mathsf{R}}(\mathcal{T})\},$$

where $\mathcal{A} \models \exists^{>k}\mathsf{r}$ means that there exists (at least) $k+1$ distinct individuals $\mathsf{b}_1,\dots,\mathsf{b}_{k+1} \in \mathsf{N_I}$ such that $\mathsf{r}(a,\mathsf{b}_1),\dots,\mathsf{r}(a,\mathsf{b}_{k+1}) \in \mathcal{A}$ (or $\mathsf{r}(\mathsf{b}_1,a),\dots,\mathsf{r}(\mathsf{b}_{k+1},a) \in \mathcal{A}$ if r is an inverse role). Notice that there are at most $2^{|\mathsf{N_C}(\mathcal{T})|+2|\mathsf{N_R}(\mathcal{T})|}$ possible 1-types, and at most $2^{|\mathsf{N_C}(\mathcal{T})|+4|\mathsf{N_R}(\mathcal{T})|}$ possible ABox types.

We write $\mathsf{r} \sqsubseteq_{\mathcal{T}} \mathsf{s}$ if (r,s) is in the transitive closure of $\{(\mathsf{p},\mathsf{p}) \mid \mathsf{p} \in \mathsf{N}^{\pm}_{\mathsf{R}}\} \cup \{(\mathsf{p},\mathsf{q}) \in \mathsf{N}^{\pm}_{\mathsf{R}} \times \mathsf{N}^{\pm}_{\mathsf{R}} \mid \mathsf{p} \sqsubseteq \mathsf{q} \in \mathcal{T}\}$. The following observation motivates the extension of 1-types into ABox types when a fixed cost k is considered:

Lemma 7. *Consider an interpretation $\mathcal{I}$ whose cost is $\leq k$. For every individual a and role $\mathsf{r} \in \mathsf{N}^{\pm}_{\mathsf{R}}$, if $\exists^{>k}\mathsf{r} \in \mathsf{tp}_{\mathcal{A}}(\mathsf{a})$, then $\exists \mathsf{s} \in \mathsf{tp}_{\mathcal{I}}(\mathsf{a})$ for every role $\mathsf{s} \in \mathsf{N}^{\pm}_{\mathsf{R}}$ such that $\mathsf{r} \sqsubseteq_{\mathcal{T}} \mathsf{s}$.*

We now prove that, if there exists an interpretation $\mathcal{I}$ whose cost is $\leq k$, then we find an interpretation $\mathcal{J}$ that behaves as $\mathcal{I}$ w.r.t. the query q, and whose cost is also $\leq k$, with all violations concentrated in a predictable portion of its domain. This portion of the domain of $\mathcal{J}$ is of course small, since the maximum number of violations is k, thus involving at most $2k$ distinct elements. By 'predictable', we mean that these violations take place among a small number of special individuals (constant number with respect to k) that can easily be identified, and on a small set of additional domain elements. To identify these special individuals, we rely on the following intuition: if an ABox type t is realized more than $2k$ times in $\mathcal{A}$, then there is a way to complete t without any 'local' violation. Indeed, if it was impossible to do so, then t being realized more than $2k$ times would always result in more than k violations and thus in a cost exceeding k, contradicting the very existence of $\mathcal{I}$. Therefore, only individuals with a rare ABox type may require a special treatment to keep the cost less than k. Formally, we say that an ABox type t is *rare* in $\mathcal{A}$ if $\#\{\mathsf{a} \in \mathsf{Ind}(\mathcal{A}) \mid \mathsf{tp}_{\mathcal{A}}(\mathsf{a}) = t\} \leq 2k$, and we use $\mathsf{RT}(\mathcal{A})$ for the set of rare ABox types in $\mathcal{A}$.

Now, when we start from an interpretation $\mathcal{I}$ with cost $\leq k$ and attempt to build $\mathcal{J}$, we preserve the interpretation of concepts and roles from $\mathcal{I}$ on those special individuals that have a rare ABox type. For $\mathcal{J}$ to behave like $\mathcal{I}$ with respect to the query q, we also preserve the interpretation on individuals occurring in q. We define the pre-core $\mathsf{pc}(\mathcal{A})$ as the set of individuals from $\mathcal{A}$ that have a rare ABox type, plus those query-related individuals, that is:

$$\mathsf{pc}(\mathcal{A}) := \{\mathsf{a} \mid \mathsf{tp}_{\mathcal{A}}(\mathsf{a}) \in \mathsf{RT}(\mathcal{A})\} \cup \mathsf{Ind}(q).$$

Unfortunately, the pre-core does not contain all the individuals that may be forced to participate in violations. As the following example illustrates, elements 'close' to the pre-core may also be forced to do so.

Example 1. *Consider the fixed cost $k := 3$ and the ABox $\mathcal{A} := \{\mathsf{A}(\mathsf{a}_0)\} \cup \bigcup_{i=0}^{7}\{\mathsf{r}(\mathsf{a}_i,\mathsf{b}_i), \mathsf{t}(\mathsf{b}_i,\mathsf{c}_i)\}$. The ABox type of a_0 is rare, others are not. Consider the TBox $\mathcal{T}$ with axioms:*

$$\mathsf{A} \sqsubseteq \exists \mathsf{u} \sqcap \neg \exists \mathsf{s} \quad \exists \mathsf{u}^- \sqsubseteq \exists \mathsf{r}^- \sqcap \neg \exists \mathsf{s}^- \quad \exists \mathsf{t} \sqsubseteq \exists \mathsf{s}^- \quad \mathsf{r} \sqsubseteq \mathsf{s}$$

and assign cost 1 to all t *assertions, cost 2 to axiom* $\mathsf{r} \sqsubseteq \mathsf{s}$*, and infinite cost to other assertions and axioms. Every interpretation with cost ≤ 3 violates the assertion $\mathsf{t}(\mathsf{b}_0,\mathsf{c}_0)$, which is somewhat surprising as both involved individuals have ABox types that can otherwise be instantiated in a way that does not violate anything. However,* b_0 *and* c_0 *happen to be 'close',* i.e. *at distance less than $k = 3$, to a_0. The rare type of a_0 can then impact b_0 and c_0 as seen above.*

To capture those individuals that may be affected by elements from the pre-core, we essentially explore the neighbourhood of the latter. For $\mathsf{a},\mathsf{b} \in \mathsf{Ind}(\mathcal{A})$, we write $\mathsf{a} \rightsquigarrow_1 \mathsf{b}$ if there exists a role $\mathsf{r} \in \mathsf{N}^{\pm}_{\mathsf{R}}$ and an assertion $\mathsf{r}(\mathsf{a},\mathsf{b})$ in $\mathcal{A}$ and $\exists^{>k}\mathsf{r} \notin \mathsf{tp}_{\mathcal{A}}(\mathsf{a})$. When exploring neighbours, the reason we exclude roles r such that $\exists^{>k}\mathsf{r} \in \mathsf{tp}_{\mathcal{A}}(\mathsf{a})$ comes from Lemma 7: it guarantees that element a satisfies $\exists \mathsf{r}$, so the potential violations on a do not impact the r-edges to b. We then denote $\mathsf{a} \rightsquigarrow_{i+1} \mathsf{c}$ if there exists b such that $\mathsf{a} \rightsquigarrow_i \mathsf{b}$ and $\mathsf{b} \rightsquigarrow_1 \mathsf{c}$. The core of $\mathcal{A}$, denoted $\mathsf{core}(\mathcal{A})$, is now defined as:

$$\mathsf{core}(\mathcal{A}) := \mathsf{pc}(\mathcal{A}) \cup \{\mathsf{b} \mid \mathsf{a} \rightsquigarrow_i \mathsf{b}, \mathsf{a} \in \mathsf{pc}(\mathcal{A}), i \leq k+1\}.$$

Notice that we stop the exploration of the neighbourhood of $\mathsf{pc}(\mathcal{A})$ at depth $k+1$. This is simply because the special behaviour of a pre-core element can only "cascade" to neighbours by enforcing a violation at each layer; thus, impacted elements cannot be further than $(k+1)$-away.

We can now state our key technical lemma. Note that Points 5_p and 5_c are used respectively to handle the possible and certain semantics. Recall that Theorem 7 established coNP-hardness for CQA_c^k, which is why Point 5_c only concerns the case where q is an IQ.

Lemma 8. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})_\omega$ be a WKB, q a BCQ, and k a fixed cost. If there exists an interpretation $\mathcal{I}$ with cost $\leq k$, then there exists an interpretation $\mathcal{J}$ such that:*

1. $\Delta^{\mathcal{J}} = \mathsf{Ind}(\mathcal{A}) \cup W$ *for some* $W \subseteq \{w_t \mid t \text{ is a } 1\text{-type}\}$;
2. $\mathcal{J}|_{\mathsf{pc}(\mathcal{A})} = \mathcal{I}|_{\mathsf{pc}(\mathcal{A})}$;
3. $\omega(\mathcal{J}) = \omega(\mathcal{J}|_{\mathsf{core}(\mathcal{A}) \cup W})$;
4. $\omega(\mathcal{J}) \leq k$;

5_p. *If $\mathcal{I} \models q$, then $\mathcal{J}|_{\mathsf{core}(\mathcal{A}) \cup W} \models q$ (and thus $\mathcal{J} \models q$);*

5_c. *If q is an IQ and $\mathcal{I} \not\models q$, then $\mathcal{J} \not\models q$.*

6.2 Construction of the FO-Rewriting

Consider a DL-Lite$_{\text{bool}}^{\mathcal{H}}$ TBox $\mathcal{T}$, a weight function $\omega_{\mathcal{T}}$ for $\mathcal{T}$, fixed cost k, and BCQ q. We now proceed to the actual construction of the rewritten query q'. Notice that the pre-core always has size at most $P_0 := 2k \times 2^{|\mathsf{N_C}(\mathcal{T})|+4|\mathsf{N_R}(\mathcal{T})|} + |q|$ (at most $2k$ copies of each rare ABox type, plus the individuals in q). For each individual a, there exist at most $2k|\mathsf{N_R}(\mathcal{T})|$ distinct individuals b such that $\mathsf{a} \rightsquigarrow_1 \mathsf{b}$. Therefore, the size of every core, regardless of the specific ABox, is bounded by $P_0 \times (2k|\mathsf{N_R}(\mathcal{T})|)^k$. We let $M_{\mathcal{T},q,k}$ be the above quantity plus the number of possible 1-types, that is:

$$M_{\mathcal{T},q,k} := P_0 \times (2k|\mathsf{N_R}(\mathcal{T})|)^k + 2^{|\mathsf{N_C}(\mathcal{T})|+2|\mathsf{N_R}(\mathcal{T})|}$$

The number $M_{\mathcal{T},q,k}$ gives an upper bound on the maximal size of the domain $\mathsf{core}(\mathcal{A}) \cup W$ involved in the construction of interpretation $\mathcal{J}$ in Lemma 8. Note that Points 3, 5_p and 5_c also guarantee that $\mathsf{core}(\mathcal{A}) \cup W$ contains all the relevant information regarding violations and query satisfaction.

The rewritten query q' considers each relevant interpretation $\mathcal{M}$, up to isomorphism, whose domain $\Delta^{\mathcal{M}}$ has size at most $M_{\mathcal{T},q,k}$ and tries to match the 'individual part' of $\mathcal{M}$ (that is, the $\mathsf{core}(\mathcal{A})$ part, as opposed to the W part) in the input $\mathcal{A}_{\omega_{\mathcal{A}}}$. The rewriting must also check that each remaining individual from $\mathcal{A}_{\omega_{\mathcal{A}}}$ can be interpreted in a manner that is compatible w.r.t. the considered $\mathcal{M}$, in the sense that it shall not introduce any violations (nor satisfy the query, in the case of the certain semantics). To achieve this, we specify along with $\mathcal{M}$ its intended individual part as a subdomain $\Gamma \subseteq \Delta^{\mathcal{M}}$ and the allowed ABox violations as a weighted ABox $\mathcal{V}_\nu$. We call such a triple $(\mathcal{M}, \Gamma, \mathcal{V}_\nu)$ a *strategy* for $(\mathcal{T}_{\omega_{\mathcal{T}}}, q, k)$ if $\mathsf{Ind}(\mathcal{V}) \subseteq \Gamma$ and $\omega_{\mathcal{T}}(\mathcal{M}) + \sum_{\alpha \in \mathcal{V}} \nu(\alpha) \leq k$.

We differentiate between p-strategies, used for CQA_p^k, and c-strategies, used for IQA_c^k. A *p-strategy* is a strategy $(\mathcal{M}, \Gamma, \mathcal{V}_\nu)$ that additionally satisfies $\mathcal{M} \models q$. An ABox type t is p-safe for a p-strategy $\sigma = (\mathcal{M}, \Gamma, \mathcal{V}_\nu)$ if there exists a 1-type t' such that:

1. for every $\mathsf{A} \in \mathsf{N_C}(\mathcal{T})$, if $\mathsf{A} \in t$, then $\mathsf{A} \in t'$;
2. for every $\mathsf{r} \in \mathsf{N_R^{\pm}}(\mathcal{T})$, if $\exists \mathsf{r} \in t$, then $\exists \mathsf{r} \in t'$;
3. t' does not violate any CIs from $\mathcal{T}$;
4. for every $\mathsf{r} \in \mathsf{N_R^{\pm}}(\mathcal{T})$, if $\exists \mathsf{r} \in t'$, then $\exists \mathsf{s} \in t'$ for every $\mathsf{s} \in \mathsf{N_R^{\pm}}(\mathcal{T})$ with $\mathsf{r} \sqsubseteq_{\mathcal{T}} \mathsf{s}$;
5. for every $\mathsf{r} \in \mathsf{N_R^{\pm}}(\mathcal{T})$, if $\exists \mathsf{r} \in t'$, then there exists $d \in \Delta^{\mathcal{M}}$ such that for every $\mathsf{s} \in \mathsf{N_R^{\pm}}(\mathcal{T})$ with $\mathsf{r} \sqsubseteq_{\mathcal{T}} \mathsf{s}$, we have $\exists \mathsf{s}^- \in \mathsf{tp}_{\mathcal{M}}(d)$.

Recall that we need only to define c-strategies for the case where q is an IQ. We may assume w.l.o.g. that q is a concept IQ[1]. A *c-strategy* for $(\mathcal{T}_{\omega_{\mathcal{T}}}, q, k)$ is a strategy $\sigma = (\mathcal{M}, \Gamma, \mathcal{V}_\nu)$ such that $\mathsf{Ind}(q) \subseteq \Gamma$ and $\mathcal{M} \not\models q$, and an ABox type t is c-safe for σ if there exists a 1-type t' that satisfies Conditions 1–5 of p-safe types, plus the following condition:

6. if $q = \exists y\, \mathsf{A}(y)$, then $\mathsf{A} \notin t'$.

The rewritten FO-query q' is now obtained as the disjunction of subqueries q_σ, where each σ is a p-strategy (resp. c-strategy) for $(\mathcal{T}_{\omega_{\mathcal{T}}}, q, k)$. Now, for a given $\sigma := (\mathcal{M}, \Gamma, \mathcal{V}_\nu)$, the subquery q_σ uses one existentially quantified variable v_d for each $d \in \Gamma$, and attempts to identify a subpart of the input weighted ABox $\mathcal{A}_{\omega_{\mathcal{A}}}$ that could be interpreted as $\mathcal{M}|_\Gamma$, up to the violations described in $\mathcal{V}_\nu$. For example, if $d \notin \mathsf{A}^{\mathcal{M}}$ and $\mathsf{A}(d) \in \mathcal{V}$, then q_σ contains the following subquery $\mathsf{A}(v_d) \rightarrow \mathsf{W}_{\mathsf{A}}^{\nu(\mathsf{A}(d))}(v_d)$, enforcing that variable v_d can only be mapped on an individual a such that $\mathsf{A}(\mathsf{a}) \in \mathcal{A}$ if $\omega_{\mathcal{A}}$ and ν agree on the cost of the assertion $\mathsf{A}(\mathsf{a})$. Using a universally quantified variable, q_σ also makes sure that all other individuals in $\mathcal{A}$ have an ABox type that is p-safe (resp. c-safe) w.r.t. σ. It is indeed clear, examining the very local requirements defining safe types that 'having a safe p- (or c-) type' can be verified using an appropriate FO-subquery.

In this manner, we can construct FO-queries q'_p and q'_c, respectively based on p- and c-strategies, such that the following properties hold, concluding the proof of Theorem 8:

Lemma 9. *Let $\mathcal{T}_{\omega_{\mathcal{T}}}$ be a weighted* DL-Lite$_{\text{bool}}^{\mathcal{H}}$ *TBox, k an integer, and q a BCQ. For every weighted ABox $\mathcal{A}_{\omega_{\mathcal{A}}}$:*

$$(\mathcal{T}, \mathcal{A})_{\omega_{\mathcal{T}} \cup \omega_{\mathcal{A}}} \models_p^k q \text{ iff } \mathcal{I}_{\mathcal{A}_{\omega_{\mathcal{A}}}} \models q'_p.$$

Furthermore, if q is an IQ, then:

$$(\mathcal{T}, \mathcal{A})_{\omega_{\mathcal{T}} \cup \omega_{\mathcal{A}}} \not\models_c^k q \text{ iff } \mathcal{I}_{\mathcal{A}_{\omega_{\mathcal{A}}}} \models q'_c.$$

7 Conclusion

Our results significantly improve our understanding of the data complexity of query entailment under recently introduced cost-based semantics. In particular, we have proved a Δ_2^p upper bound for the optimal-cost certain and possible semantics, yielding tight complexity bounds for a wide range of lightweight and expressive DLs, up to $\mathcal{ALCHIO}$ and covering also prominent DL-Lite dialects. Moreover, we obtained surprising tractability results, showing that fixed-cost possible semantics (for CQs) and certain semantics (for IQs) in DL-Lite$_{\text{bool}}^{\mathcal{H}}$ enjoy the same low AC^0 complexity as classical CQ answering in DL-Lite. We expect our upper bounds can be adapted to also handle negative role inclusions (to cover also DL-Lite$_{\mathcal{R}}$). For DLs with functionality or number restrictions, it does not suffice to work with finite interpretations, so wholly different methods are required. Developing a practical implementation of the FO-rewritings for the identified tractable cases is another interesting direction.

[1] IQs of the form $\exists y\, \mathsf{r}(\mathsf{a}, y)$ (resp. $\exists xy\, \mathsf{r}(x, y)$) can be handled by adding infinite-weight CIs $\exists \mathsf{r} \sqsubseteq \mathsf{B}$, $\mathsf{B} \sqsubseteq \exists \mathsf{r}$ for a fresh concept name B, and using the IQ $\mathsf{B}(\mathsf{a})$ (resp. $\exists x.\mathsf{B}(x)$) instead.

Acknowledgments

The authors acknowledge the financial support of the ANR AI Chair INTENDED (ANR-19-CHIA-0014) and the Federal Ministry of Research, Technology and Space of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig", project identification number: ScaDS.AI.

References

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Bienvenu, M. 2020. A Short Survey on Inconsistency Handling in Ontology-Mediated Query Answering. *Künstliche Intell.*, 34(4): 443–451.

Bienvenu, M.; and Bourgaux, C. 2016. Inconsistency-Tolerant Querying of Description Logic Knowledge Bases. In *Reasoning Web Tutorial Lectures*.

Bienvenu, M.; Bourgaux, C.; and Jean, R. 2024. Cost-Based Semantics for Querying Inconsistent Weighted Knowledge Bases. In *Proceedings of KR*.

Bienvenu, M.; and Ortiz, M. 2015. Ontology-Mediated Query Answering with Data-Tractable Description Logics. In *Reasoning Web Tutorial Lectures*.

Bonatti, P. A.; Lutz, C.; and Wolter, F. 2009. The Complexity of Circumscription in DLs. *J. Artif. Intell. Res.*, 35: 717–773.

Bourgaux, C. 2016. *Inconsistency handling in ontology-mediated query answering. (Gestion des incohérences pour l'accès aux données en présence d'ontologies)*. Ph.D. thesis, University of Paris-Saclay, France.

Britz, K.; Casini, G.; Meyer, T.; Moodley, K.; Sattler, U.; and Varzinczak, I. 2021. Principles of KLM-style Defeasible Description Logics. *ACM Trans. Comput. Log.*, 22(1): 1:1–1:46.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable Reasoning and Efficient Query Answering in Description Logics: The DL-Lite Family. *Journal of Automated Reasoning (JAR)*.

Eiter, T.; Lukasiewicz, T.; and Predoiu, L. 2016. Generalized Consistent Query Answering under Existential Rules. In *Proceedings of KR*.

Giordano, L.; Gliozzi, V.; Olivetti, N.; and Pozzato, G. L. 2013. A non-monotonic Description Logic for reasoning about typicality. *Artif. Intell.*, 195: 165–202.

Lembo, D.; Lenzerini, M.; Rosati, R.; Ruzzi, M.; and Savo, D. F. 2010. Inconsistency-Tolerant Semantics for Description Logics. In *Proceedings of RR*.

Lutz, C.; and Manière, Q. 2024. Adding Circumscription to Decidable Fragments of First-Order Logic: A Complexity Rollercoaster. In *Proceedings of KR*, 531–541.

Manière, Q. 2022. *Counting queries in ontology-based data access. (Requêtes de comptage pour l'accès aux données en présence d'ontologies)*. Ph.D. thesis, University of Bordeaux, France.

Poggi, A.; Lembo, D.; Calvanese, D.; De Giacomo, G.; Lenzerini, M.; and Rosati, R. 2008. Linking Data to Ontologies. *Journal of Data Semantics*, 10: 133–173.

Xiao, G.; Calvanese, D.; Kontchakov, R.; Lembo, D.; Poggi, A.; Rosati, R.; and Zakharyaschev, M. 2018. Ontology-Based Data Access: A Survey. In *Proceedings of IJCAI*.