Proceedings of the 21st International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Cost-Based Semantics for Querying Inconsistent Weighted Knowledge Bases

Meghyn Bienvenu[1,2] , **Camille Bourgaux**[3] , **Robin Jean**[1]
[1]Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, Talence, France
[2]Japanese-French Laboratory for Informatics, CNRS, NII, IRL 2537, Tokyo, Japan
[3]DI ENS, ENS, CNRS, PSL University & Inria, Paris, France
{meghyn.bienvenu,robin.jean}@u-bordeaux.fr, camille.bourgaux@ens.fr

Abstract

In this paper, we explore a quantitative approach to querying inconsistent description logic knowledge bases. We consider weighted knowledge bases in which both axioms and assertions have (possibly infinite) weights, which are used to assign a cost to each interpretation based upon the axioms and assertions it violates. Two notions of certain and possible answer are defined by either considering interpretations whose cost does not exceed a given bound or restricting attention to optimal-cost interpretations. Our main contribution is a comprehensive analysis of the combined and data complexity of bounded cost satisfiability and certain and possible answer recognition, for description logics between $\mathcal{EL}_\bot$ and $\mathcal{ALCO}$.

1 Introduction

Ontology-mediated query answering (OMQA) is a framework for improving data access through the use of an ontology, which has been extensively studied by the KR and database communities (Poggi et al. 2008; Bienvenu and Ortiz 2015; Xiao et al. 2018). Much of the work on OMQA considers ontologies formulated in description logics (DLs) (Baader et al. 2017). In the DL setting, OMQA consists in finding the answers that are logically entailed from the knowledge base (KB), consisting of the ABox (data) and TBox (ontology). Due to the use of classical first-order semantics, whereby everything is entailed from a contradiction, classical OMQA semantics fails to provide informative answers when the KB is inconsistent.

The issue of handling inconsistencies, or more generally unwanted consequences, in DL KBs has been explored from many angles. One solution is to modify the KB in order to render it consistent, and there has been significant research on how to aid users in the debugging process, e.g. by generating justifications that pinpoint the sources of the inconsistency (Parsia, Sirin, and Kalyanpur 2005; Peñaloza 2020). This line of work mostly focuses on helping knowledge engineers to debug the TBox before deployment in applications, but some recent work specifically target ABoxes (Baader and Kriegel 2022). However, in an OMQA setting, where the ABox can be very large and subject to frequent updates, it is unrealistic to assume that we can always restore consistency. This has motivated a substantial line of research on inconsistency-tolerant semantics to obtain meaningful answers from inconsistent KBs, surveyed in (Bienvenu and Bourgaux 2016; Bienvenu 2020). Many of these semantics are based upon repairs, defined as inclusion-maximal subsets of the ABox that are consistent w.r.t. the TBox. Two of the most commonly considered repair semantics are the AR semantics (Lembo et al. 2010), which asks for those answers that hold in every repair, and the brave semantics (Bienvenu and Rosati 2013), which considers those answers that hold in at least one repair. Note that the work on repair-based semantics typically assumes that the TBox is reliable, which is why repairs are subsets of the ABox, with the TBox left untouched. A notable exception is the work of Eiter, Lukasiewicz, and Predoiu (2016), which considers generalized notions of repair for existential rule ontologies composed of hard and soft rules, in which contradictions may be resolved by removing or minimally violating soft rules.

In this paper, we explore a novel quantitative approach to querying inconsistent description logic KBs, which combines the idea of soft ontology axioms from (Eiter, Lukasiewicz, and Predoiu 2016) with a recent cost-based approach to repairing databases w.r.t. soft constraints (Carmeli et al. 2021). The idea is to associate with every TBox axiom and ABox assertion a (possibly infinite) weight. 'Hard' axioms and assertions, which must be satisfied, are assigned a weight of ∞, and the remaining 'soft' axioms and assertions are assigned weights based upon their reliability, with higher weights indicating greater trust. The cost of an interpretation is defined by taking into account the number of violations of an axiom (assertions can be violated at most once) and the weights of the violated axioms and assertions. When determining the query answers, we shall use the cost to select a set of interpretations, either by considering all interpretations whose cost is below a given threshold, or considering only those interpretations having an optimal (i.e. minimum) cost. We shall then consider both the certain answers, which hold in all of the selected interpretations, and the possible answers, which hold w.r.t. at least one selected interpretation. When restricted to consistent KBs, the optimal-cost certain and possible semantics coincide with the classical certain and possible answer semantics, cf. (Andolfi et al. 2024). By varying the cost bounds, we can identify answers that are robust (i.e. hold not only in all optimal-cost interpretations but also 'close-to-optimal' ones) or rank candidate answers based upon their incompatibility with the KB.

We perform a comprehensive analysis of the complexity of the main decision problems in our setting, namely, bounded-cost satisfiability of weighted KBs and recognition of certain and possible answers w.r.t. the set of k-cost-bounded or optimal-cost interpretations. Our study covers lightweight and expressive description logics, ranging from $\mathcal{EL}_\bot$ to $\mathcal{ALCO}$, and queries given either as instance (IQs) or conjunctive queries (CQs). We consider both the combined and data complexity measures, as well as the impact of unary and binary encodings of the cost bound and weights. Our results are summarized in Table 1. For combined complexity, most problems are EXPTIME-complete, except for those involving certain answers to CQs, which are 2EXPTIME-complete. For data complexity, we identify problems which are (co)NP-complete, Θ_2^p-complete, and Δ_2^p-complete[1], depending on the encoding and maximal value of the weights.

The paper is organized as follows. Following the preliminaries in Section 2, we introduce in Section 3 our formal framework and the associated decision problems. Sections 4 and 5 present respectively our combined and data complexity results. We discuss related work in Section 6 and conclude in Section 7 with some directions for future work.

Omitted proofs are provided in the appendix of the long version (Bienvenu, Bourgaux, and Jean 2024).

2 Preliminaries

We briefly recall the syntax and semantics of DL.

Syntax A DL *knowledge base* (KB) $\mathcal{K} = \langle\mathcal{T}, \mathcal{A}\rangle$ consists of an ABox $\mathcal{A}$ and a TBox $\mathcal{T}$, both of which are constructed from three mutually disjoint countable sets $\mathsf{N_C}$ of *concept names* (unary predicates), $\mathsf{N_R}$ of *role names* (binary predicates), and $\mathsf{N_I}$ of *individual names* (constants). The *ABox* is a finite set of *concept assertions* of the form $A(a)$ with $A \in \mathsf{N_C}, a \in \mathsf{N_I}$ and *role assertions* of the form $R(a, b)$ with $R \in \mathsf{N_R}, a, b \in \mathsf{N_I}$. The *TBox* is a finite set of axioms whose form depends on the DL in question. In $\mathcal{ALCOQ}u$ TBox axioms are *concept inclusions* $C \sqsubseteq D$ where C and D are complex concepts formed using the following syntax:

$$C := A \mid \{a\} \mid \top \mid \bot \mid C \sqcap C \mid C \sqcup C \mid \neg C \\ \mid \exists R.C \mid \forall R.C \mid \leq nR.C \mid \geq nR.C$$

where $A \in \mathsf{N_C}$, $a \in \mathsf{N_I}$, $R \in \mathsf{N_R} \cup \{U\}$, with U the special *universal role*.[2]

The DL $\mathcal{ALCO}$ is the restriction of $\mathcal{ALCOQ}u$ disallowing the use of qualified number restrictions ($\leq nR.C$ or $\geq nR.C$) and of the universal role U. The DL $\mathcal{EL}_\bot$ further disallows the use of universal restrictions ($\forall R.C$), negations ($\neg C$), unions ($C \sqcup C$) and nominals ($\{a\}$).

We denote by $\mathsf{Ind}(\mathcal{A})$ (resp. $\mathsf{Ind}(\mathcal{K})$) the set of individuals that occur in $\mathcal{A}$ (resp. in $\mathcal{K}$), and by $\mathsf{sig}(\mathcal{T})$ (resp. $\mathsf{sig}(\mathcal{K})$) the set of concept and role names that occur in $\mathcal{T}$ (resp. in $\mathcal{K}$).

[1] Δ_2^p is the class of decision problems solvable in polynomial time with access to an NP oracle, and Θ_2^p (aka $\Delta_2^p[log\ n]$) the subclass allowing only logarithmically many NP oracle calls.

[2] Usually the universal role cannot occur in qualified number restrictions ($\leq nR.C$ or $\geq nR.C$) but nominals allow us to simulate such number restrictions as explained in (Ortiz and Simkus 2012).

Semantics An *interpretation* has the form $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where the *domain* $\Delta^\mathcal{I}$ is a non-empty set and $\cdot^\mathcal{I}$ maps each $a \in \mathsf{N_I}$ to $a^\mathcal{I} \in \Delta^\mathcal{I}$, each $A \in \mathsf{N_C}$ to $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, each $R \in \mathsf{N_R}$ to $R^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$ and interprets the universal role U by $U^\mathcal{I} = \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The function $\cdot^\mathcal{I}$ is extended to general concepts, e.g., $(\exists R.D)^\mathcal{I} = \{c \mid \exists d \in D^\mathcal{I} : (c, d) \in R^\mathcal{I}\}$; $\{a\}^\mathcal{I} = \{a^\mathcal{I}\}$; $\top^\mathcal{I} = \Delta^\mathcal{I}$; $\bot^\mathcal{I} = \emptyset$; $(\leq nR.C)^\mathcal{I} = \{c \mid \#\{d \in C^\mathcal{I} \mid (c, d) \in R^\mathcal{I}\} \leq n\}$ and $(\geq nR.C)^\mathcal{I} = \{c \mid \#\{d \in C^\mathcal{I} \mid (c, d) \in R^\mathcal{I}\} \geq n\}$. An interpretation $\mathcal{I}$ satisfies an assertion $A(a)$ (resp. $R(a, b)$) if $a \in A^\mathcal{I}$ (resp. $(a, b) \in R^\mathcal{I}$); we thus make a weak version of the *standard names assumption* (SNA).[3] $\mathcal{I}$ satisfies an inclusion $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$ and $\{a\}^\mathcal{I} = \{a\}$ for every nominal occurring in C or D. We write $\mathcal{I} \models \tau$ (resp. $\mathcal{I} \models \alpha$) to indicate that $\mathcal{I}$ satisfies an axiom τ (resp. assertion α). An interpretation $\mathcal{I}$ is a *model* of $\mathcal{K} = \langle\mathcal{T}, \mathcal{A}\rangle$, denoted $\mathcal{I} \models \mathcal{K}$, if $\mathcal{I}$ satisfies all inclusions in $\mathcal{T}$ ($\mathcal{I} \models \mathcal{T}$) and all assertions in $\mathcal{A}$ ($\mathcal{I} \models \mathcal{A}$). A KB $\mathcal{K}$ is *consistent* if it has a model.

Queries We consider *conjunctive queries* (CQs) which take the form $\exists\vec{y}\psi$, where ψ is a conjunction of atoms of the forms $A(t)$ or $R(t, t')$, where t, t' are variables or individuals, and $\vec{y}$ is a tuple of variables from ψ. A CQ is called *Boolean* (BCQ) if all of its variables are existentially quantified; a CQ consisting of a single atom is an *instance query* (IQ). When we use the generic term query, we mean a CQ. For a BCQ q and an interpretation $\mathcal{I}$, we denote by $\mathcal{I} \models q$ the fact that $\mathcal{I}$ satisfies q. A BCQ q is *entailed* from $\mathcal{K}$, written $\mathcal{K} \models q$, if $\mathcal{I} \models q$ for every model $\mathcal{I}$ of $\mathcal{K}$. A BCQ q is *satisfiable w.r.t.* $\mathcal{K}$ if there exists a model $\mathcal{I}$ of $\mathcal{K}$ such that $\mathcal{I} \models q$. For a non-Boolean CQ $q[\vec{x}]$ with free variables $\vec{x} = (x_1, \ldots, x_k)$, a tuple of individuals $\vec{a} = (a_1, \ldots, a_k)$ is a *certain answer* for $q[\vec{x}]$ w.r.t. $\mathcal{K}$ just in the case that $\mathcal{K} \models q[\vec{a}]$, where $q[\vec{a}]$ is the BCQ obtained by replacing each x_i by a_i. Tuple $\vec{a}$ is said to be a *possible answer* for $q[\vec{x}]$ w.r.t. $\mathcal{K}$ if the BCQ $q[\vec{a}]$ is satisfiable w.r.t. $\mathcal{K}$. Observe that certain and possible answer recognition corresponds to BCQ entailment and satisfiability respectively.

To simplify the presentation, we shall focus on BCQs. However, all definitions and results are straightforwardly extended to non-Boolean queries, and we shall thus sometimes speak of 'query answers' when providing intuitions.

3 Weighted Knowledge Bases

We consider a quantitative way of integrating the notion of soft constraints by giving weights to axioms and assertions. Intuitively, these weights represent penalties associated to each violation of the axioms or assertions. They will allow us to assign a cost to interpretations based upon the axioms and assertions they violate, and use this cost to select which interpretations to consider when answering queries.

Definition 1. *A* weighted knowledge base *(WKB)* $\mathcal{K}_\omega = (\langle\mathcal{T}, \mathcal{A}\rangle, \omega)$ *consists of a knowledge base* $\langle\mathcal{T}, \mathcal{A}\rangle$ *and a cost function* $\omega : \mathcal{T} \cup \mathcal{A} \mapsto \mathbb{N}_{>0} \cup \{+\infty\}$. *We denote by* $\mathcal{T}_\infty$

[3] The usual SNA requires that $a^\mathcal{I} = a$ for every $a \in \mathsf{N_I}$, hence that $\mathsf{N_I} \subseteq \Delta^\mathcal{I}$, so all interpretations have an infinite domain. To be able to bound the size of interpretations, we adopt this 'weak' version of the SNA, used e.g., by Lutz, Manière, and Nolte (2023).

(resp. $\mathcal{A}_\infty$) the set of TBox axioms (resp. ABox assertions) that have an infinite cost and let $\mathcal{K}_\infty = \langle \mathcal{T}_\infty, \mathcal{A}_\infty \rangle$. We sometimes use ω_χ as a shorthand for $\omega(\chi)$.

Example 1. *Consider the following WKB about visa requirements to enter some country c: $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A} \rangle, \omega)$ where $\mathcal{T} = \{\tau_1, \tau_2, \tau_3\}$, $\mathcal{A} = \{\alpha_1, \alpha_2\}$ and*

$$\begin{array}{ll} \tau_1 = \mathsf{Visa} \sqcap \mathsf{NoVisa} \sqsubseteq \bot & \omega(\tau_1) = \infty \\ \tau_2 = \exists \mathsf{hasNat}.\{c\} \sqcap \exists \mathsf{hasNat}.\{b\} \sqsubseteq \bot & \omega(\tau_2) = \infty \\ \tau_3 = \forall \mathsf{hasNat}.\neg\{c\} \sqsubseteq \mathsf{Visa} & \omega(\tau_3) = 1 \\ \alpha_1 = \mathsf{hasNat}(p, b) & \omega(\alpha_1) = 1 \\ \alpha_2 = \mathsf{NoVisa}(p) & \omega(\alpha_2) = 2 \end{array}$$

Two 'absolute' constraints τ_1 and τ_2 express that one cannot both need a visa (Visa*) and not need one (*NoVisa*) and that it is not possible to have both nationalities (*hasNat*) c and b. A 'soft' constraint τ_3 expresses that someone that does not have nationality c normally needs a visa. The ABox states that a person p has nationality b and does not need a visa, and the second assertion is more reliable than the first one.*

To measure how far an interpretation is from being a model of the KB, we rely on the following sets of *violations*.

Definition 2. *The* set of violations of a concept inclusion *$B \sqsubseteq C$ in an interpretation $\mathcal{I}$ is the set*

$$vio_{B \sqsubseteq C}(\mathcal{I}) = (B \sqcap \neg C)^{\mathcal{I}}.$$

The violations of an ABox $\mathcal{A}$ *in an interpretation $\mathcal{I}$ are*

$$vio_{\mathcal{A}}(\mathcal{I}) = \{\alpha \in \mathcal{A} \mid \mathcal{I} \not\models \alpha\}.$$

These sets of violations can be used to associate a *cost* to interpretations, by taking into account the weights assigned by the WKB to the violated inclusions and assertions.

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

$$cost_{\mathcal{K}_\omega}(\mathcal{I}) = \sum_{\tau \in \mathcal{T}} \omega_\tau |vio_\tau(\mathcal{I})| + \sum_{\alpha \in vio_{\mathcal{A}}(\mathcal{I})} \omega_\alpha.$$

We say that $\mathcal{K}_\omega$ is k-satisfiable *if there exists an interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$ and define the* optimal cost *of $\mathcal{K}_\omega$ as $optc(\mathcal{K}_\omega) = \min_{\mathcal{I}}(cost_{\mathcal{K}_\omega}(\mathcal{I}))$.*

Remark 1. *Note that $cost_{\mathcal{K}_\omega}(\mathcal{I})$ will be ∞ if any infinite-weight assertion or inclusion is violated in $\mathcal{I}$ and/or if any inclusion has an infinite set of violations in $\mathcal{I}$.*

Example 2 (Ex.1 cont'd)**.** *Consider the following interpretations over domain $\Delta^{\mathcal{I}} = \{p, b, c\}$ that correspond to different possibilities for p's nationalities and need for a visa.*

- *Case p has nationality b and needs a visa: $\mathsf{hasNat}^{\mathcal{I}_b^v} = \{(p, b)\}$, $\mathsf{Visa}^{\mathcal{I}_b^v} = \{p\}$ and $\mathsf{NoVisa}^{\mathcal{I}_b^v} = \emptyset$. $\mathcal{I}_b^v$ violates only α_2 so $cost_{\mathcal{K}_\omega}(\mathcal{I}_b^v) = 2$.*
- *Case p has nationality b and does not need a visa: $\mathsf{hasNat}^{\mathcal{I}_b^n} = \{(p, b)\}$, $\mathsf{Visa}^{\mathcal{I}_b^n} = \emptyset$ and $\mathsf{NoVisa}^{\mathcal{I}_b^n} = \{p\}$. $\mathcal{I}_b^n$ violates only τ_3 so $cost_{\mathcal{K}_\omega}(\mathcal{I}_b^n) = 1$.*
- *Case p has nationality c and needs a visa: $\mathsf{hasNat}^{\mathcal{I}_c^v} = \{(p, c)\}$, $\mathsf{Visa}^{\mathcal{I}_c^v} = \{p\}$ and $\mathsf{NoVisa}^{\mathcal{I}_c^v} = \emptyset$. $\mathcal{I}_c^v$ violates only α_1 and α_2 so $cost_{\mathcal{K}_\omega}(\mathcal{I}_c^v) = 3$.*
- *Case p has nationality c and does not need a visa: $\mathsf{hasNat}^{\mathcal{I}_c^n} = \{(p, c)\}$, $\mathsf{Visa}^{\mathcal{I}_c^n} = \emptyset$ and $\mathsf{NoVisa}^{\mathcal{I}_c^n} = \{p\}$. $\mathcal{I}_c^n$ violates only α_1 so $cost_{\mathcal{K}_\omega}(\mathcal{I}_c^n) = 1$.*
- *Case p has nationality b and c and does not need a visa: $\mathsf{hasNat}^{\mathcal{I}_{bc}^n} = \{(p, b), (p, c)\}$, $\mathsf{Visa}^{\mathcal{I}_{bc}^n} = \emptyset$ and $\mathsf{NoVisa}^{\mathcal{I}_{bc}^n} = \{p\}$. $\mathcal{I}_{bc}^n$ violates τ_2 so $cost_{\mathcal{K}_\omega}(\mathcal{I}_{bc}^n) = \infty$.*

Since $\mathcal{K}_\omega$ is inconsistent and the smallest weight is 1, it follows that $\mathcal{I}_b^n$ and $\mathcal{I}_c^n$ are of optimal cost and $optc(\mathcal{K}_\omega) = 1$.

It is now possible to define variants of the classical certain and possible answers, by considering either only interpretations whose cost does not exceed a given bound, or only optimal-cost interpretations. For simplicity, we state the definitions in terms of BCQ entailment.

Definition 4. *Let q be a BCQ and $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A} \rangle, \omega)$ be a WKB. 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 $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$;*
- k-cost-bounded possible semantics*, written $\mathcal{K}_\omega \models_p^k q$, if $\mathcal{I} \models q$ for some interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$;*
- opt-cost certain semantics*, written $\mathcal{K}_\omega \models_c^{opt} q$, if $\mathcal{I} \models q$ for every interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) = optc(\mathcal{K}_\omega)$;*
- opt-cost possible semantics*, written $\mathcal{K}_\omega \models_p^{opt} q$, if $\mathcal{I} \models q$ for some interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) = optc(\mathcal{K}_\omega)$.*

Example 3 (Ex.1 cont'd)**.** *Since $optc(\mathcal{K}_\omega) = 1$, weights of axioms different from τ_3 and α_1 are greater than 1 and $\mathcal{I}_b^n$ and $\mathcal{I}_c^n$ are interpretations of cost 1 that violate τ_3 and α_1 respectively, it follows that interpretations of optimal cost violate exactly one axiom in $\{\tau_3, \alpha_1\}$. In particular, they all satisfy α_2, i.e., $\mathcal{K}_\omega \models_c^{opt} \mathsf{NoVisa}(p)$. Since all interpretations of minimal cost satisfy τ_1, it follows that $\mathcal{K}_\omega \not\models_p^{opt} \mathsf{Visa}(p)$. On the other hand, we obtain that $\mathcal{K}_\omega \models_p^{opt} \mathsf{hasNat}(p, b)$ (because of $\mathcal{I}_b^n$) and $\mathcal{K}_\omega \models_p^{opt} \mathsf{hasNat}(p, c)$ (because of $\mathcal{I}_c^n$).*

If we now consider interpretations of cost bounded by 2, we obtain that $\mathcal{K}_\omega \not\models_c^2 \mathsf{NoVisa}(p)$ and $\mathcal{K}_\omega \models_p^2 \mathsf{Visa}(p)$ (because of $\mathcal{I}_b^v$), hence we cannot conclude anymore whether p needs a visa or not using the certain semantics. However, we can still exclude some statements even under possible semantics. For example, $\mathcal{K}_\omega \not\models_p^2 \mathsf{hasNat}(p, c) \wedge \mathsf{Visa}(p)$, since this holds only in interpretations of cost at least 3.

When the underlying KB is consistent, the certain and possible optimal-cost semantics coincide with classical query entailment and query satisfiability (or classical certain and possible answers in the case of non-Boolean queries):

Proposition 1. *Let $\mathcal{K}_\omega$ be such that $optc(\mathcal{K}_\omega) = 0$. Then:*

- *$\mathcal{K}_\omega \models_c^{opt} q$ iff $\mathcal{K} \models q$*
- *$\mathcal{K}_\omega \models_p^{opt} q$ iff q is satisfiable w.r.t. $\mathcal{K}$*

It is also interesting to consider how the k-cost-bounded semantics vary with different values of k:

Proposition 2. *Consider a WKB $\mathcal{K}_\omega$, BCQ q, and $k \geq 0$.*

- *If $\mathcal{K}_\omega \models_c^k q$, then $\mathcal{K}_\omega \models_c^{k'} q$ for every $0 \leq k' \leq k$*
- *If $\mathcal{K}_\omega \not\models_p^k q$, then $\mathcal{K}_\omega \not\models_p^{k'} q$ for every $0 \leq k' \leq k$*

Moreover, $\mathcal{K}_\omega \not\models_p^k q$ and $\mathcal{K}_\omega \models_c^k q$ if $k < optc(\mathcal{K}_\omega)$.

	BCS	IQA_p^b, CQA_p^b	IQA_c^b	CQA_c^b	IQA_p^{opt}, CQA_p^{opt}	IQA_c^{opt}	CQA_c^{opt}
Combined	ExpTime	ExpTime	ExpTime	2ExpTime	ExpTime	ExpTime	2ExpTime
Data	NP	NP	coNP	coNP	Δ_2^{p*} / $\Theta_2^{p\dagger}$	Δ_2^p-hard* / $\Theta_2^{p\dagger}$	Δ_2^p-hard* / $\Theta_2^{p\dagger}$

Table 1: Overview of complexity results for description logics between $\mathcal{EL}_\perp$ and $\mathcal{ALCO}$. All bounds are tight except the two '-hard' cases. Lower bounds hold even if the weights (and the input integer in the case of combined complexity) are encoded in unary, except those marked with *. Upper bounds hold even if the weights (and the input integer in the case of combined complexity) are encoded in binary. †: Θ_2^p-complete if the finite weights on the assertions are either bounded (independently from $|\mathcal{A}|$), or encoded in unary.

The preceding result shows, unsurprisingly, that k-cost-bounded semantics are only informative for $k \geq optc(\mathcal{K}_\omega)$. Increasing k beyond $optc(\mathcal{K}_\omega)$ leads to fewer and fewer queries being entailed under the k-cost-bounded certain semantics, which may be useful in identifying query answers that are robust in the sense that they continue to hold even if we consider a larger set of 'close-to-optimal' interpretations. By contrast, as k grows, so does the set of entailed queries under k-cost-bounded possible semantics. Being quite permissive, the opt-cost and k-cost-bounded possible semantics will entail many queries, and thus are not suitable replacements for standard (certain answer) querying semantics. Instead, non-entailment under these semantics can serve to eliminate or rank candidate tuples of individuals (or the corresponding instantiated Boolean queries) based upon how incompatible they are w.r.t. the expressed information.

Relationship With Preferred Repair Semantics We show that opt-cost certain semantics generalizes the $\leq_\omega$-AR semantics defined by Bienvenu, Bourgaux, and Goasdoué (2014) for KBs with weighted ABoxes, where $\omega : \mathcal{A} \to \mathbb{N}_{>0}$ models the reliability of the assertions while the TBox axioms are considered absolute. In this context, $\leq_\omega$-repairs are subsets of the ABox consistent with the TBox and maximal for the preorder defined over ABox subsets by $\mathcal{A}_1 \leq_\omega \mathcal{A}_2$ if $\sum_{\alpha\in\mathcal{A}_1}\omega_\alpha \leq \sum_{\alpha\in\mathcal{A}_2}\omega_\alpha$. A BCQ q is entailed under $\leq_\omega$-AR (resp. $\leq_\omega$-brave) semantics if $\langle\mathcal{T},\mathcal{A}'\rangle \models q$ for every (resp. some) $\leq_\omega$-repair $\mathcal{A}'$ of $\mathcal{A}$.

Proposition 3. *Let $\mathcal{K}_\omega = (\langle\mathcal{T},\mathcal{A}\rangle,\omega)$ be a WKB such that $\mathcal{T}$ is satisfiable, $\omega(\tau) = \infty$ for every $\tau \in \mathcal{T}$ and $\omega(\alpha) \neq \infty$ for every $\alpha \in \mathcal{A}$, and let q be a BCQ.*

$$\mathcal{K}_\omega \models_c^{opt} q \iff \langle\mathcal{T},\mathcal{A}\rangle \models_{\leq_\omega\text{-}AR} q$$

Proof sketch. Since $\mathcal{T}$ is satisfiable, there is a model $\mathcal{I}$ of $\mathcal{T}$, and since $\omega(\alpha) \neq \infty$ for every $\alpha \in \mathcal{A}$, $cost_{\mathcal{K}_\omega}(\mathcal{I}) \neq \infty$. It follows that $optc(\mathcal{K}_\omega) \neq \infty$ and that every $\mathcal{I}$ of optimal cost is such that $\mathcal{I} \models \mathcal{T}$ and $cost_{\mathcal{K}_\omega}(\mathcal{I}) = \sum_{\alpha\in vio_\mathcal{A}(\mathcal{I})}\omega_\alpha$.

Hence, for every $\mathcal{I}$ of optimal cost, $\mathcal{A}' = \mathcal{A} \setminus vio_\mathcal{A}(\mathcal{I})$ is a $\leq_\omega$-repair. Indeed, since $\mathcal{I} \models \langle\mathcal{T},\mathcal{A}'\rangle$, $\langle\mathcal{T},\mathcal{A}'\rangle$ is consistent. Moreover, $\sum_{\alpha\in\mathcal{A}'}\omega_\alpha = \sum_{\alpha\in\mathcal{A}}\omega_\alpha - cost_{\mathcal{K}_\omega}(\mathcal{I})$ and $cost_{\mathcal{K}_\omega}(\mathcal{I})$ is minimal, so $\sum_{\alpha\in\mathcal{A}'}\omega_\alpha$ is maximal. It also follows that every $\leq_\omega$-repair $\mathcal{A}'$ is such that $\sum_{\alpha\in\mathcal{A}'}\omega_\alpha = \sum_{\alpha\in\mathcal{A}}\omega_\alpha - optc(\mathcal{K}_\omega)$. ☐

Note however that opt-cost possible semantics, does *not* generalize $\leq_\omega$-brave, but only over-approximates it:

$$\langle\mathcal{T},\mathcal{A}\rangle \models_{\leq_\omega\text{-brave}} q \implies \mathcal{K}_\omega \models_p^{opt} q.$$

Indeed, we have shown that to each $\leq_\omega$-repair corresponds at least one interpretation with optimal cost but given an interpretation $\mathcal{I}$ with optimal cost w.r.t. $\mathcal{K}_\omega = (\langle\mathcal{T},\mathcal{A}\rangle,\omega)$, if $B \in \mathsf{N_C} \setminus \mathsf{sig}(\mathcal{K})$ and $b \in \mathsf{N_I} \setminus \mathsf{Ind}(\mathcal{K})$, then one can add $b^\mathcal{I}$ to $B^\mathcal{I}$ without changing the cost of $\mathcal{I}$ w.r.t. $\mathcal{K}_\omega$, so $\mathcal{K}_\omega \models_p^{opt} B(b)$ while $\langle\mathcal{T},\mathcal{A}\rangle \not\models_{\leq_\omega\text{-brave}} B(b)$.

Decision Problems In our complexity analysis, we will consider the following decision problems:

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

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

Complexity Measures It is customary to consider *combined complexity* and *data complexity* when studying decision problems related to query answering over DL KBs. Data complexity considers only the size of the ABox while combined complexity takes into account the size of the whole input. In the case of WKBs, we consider the assertion weights as part of the ABox and inclusion weights as part of the TBox. We will use the following notation: given a WKB $\mathcal{K}_\omega = (\langle\mathcal{T},\mathcal{A}\rangle,\omega)$, $|\mathcal{A}|$ (resp. $|\mathcal{T}|$, $|\mathcal{K}|$) is the length of the string representing $\mathcal{A}$ (resp. $\mathcal{T}$, $\mathcal{K}$), where elements of $\mathsf{N_C}$, $\mathsf{N_R}$ and $\mathsf{N_I}$ are considered of length one, and $|\mathcal{A}_\omega|$ (resp. $|\mathcal{T}_\omega|$, $|\mathcal{K}_\omega|$) is the length of the string representing the set $\{(\alpha,\omega(\alpha)) \mid \alpha \in \mathcal{A}\}$ (resp. $\{(\tau,\omega(\tau)) \mid \tau \in \mathcal{T}\}$, $\{(\chi,\omega(\chi)) \mid \chi \in \mathcal{T}\cup\mathcal{A}\}$), where elements of $\mathsf{N_C}$, $\mathsf{N_R}$ and $\mathsf{N_I}$ are considered of length one and weights are encoded either in unary or in binary. Note that if the TBox contains qualified number restrictions, the numbers can also be encoded in unary or binary. We will also make this encoding distinction for the integer k taken as input by some of the decision problems we consider. If $|k|$ denotes the size of the encoding of k, $k = |k|$ when encoded in unary and $k \leq 2^{|k|}$ when encoded in binary. Finally, for a BCQ q, $|q|$ is the length of the string representing q where elements of $\mathsf{N_C}$, $\mathsf{N_R}$, $\mathsf{N_I}$ and variables are considered of size one. Note that when we use $|\cdot|$ over a set which is not a (weighted) ABox or TBox, we simply means the set cardinality.

4 Combined Complexity

In this section we study the combined complexity of bounded cost satisfiability and certain and possible answer recognition, for DLs between $\mathcal{EL}_\bot$ and $\mathcal{ALCO}$. The first line of Table 1 gives an overview of the results.

4.1 Upper Bounds

To characterize the cost of interpretations, we define the notion of k-*configuration*. Intuitively, a k-configuration specifies how to allocate a cost of k between possible violations.

Definition 5 (k-configuration). *Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be a WKB and k be an integer. A* k-configuration *for $\mathcal{K}_\omega$ is a function $\gamma : \mathcal{T} \cup \mathcal{A} \mapsto \mathbb{N}$ such that:*

- $\gamma(\tau) \in \mathbb{N}$ *for every* $\tau \in \mathcal{T}$,
- $\gamma(\alpha) \in \{0, 1\}$ *for every* $\alpha \in \mathcal{A}$,
- $\sum_{\chi \in \mathcal{T}\cup\mathcal{A}} \gamma(\chi)\omega_\chi \leq k$.

An interpretation $\mathcal{I}$ satisfies the k-configuration γ if $|vio_\tau(\mathcal{I})| \leq \gamma(\tau)$ for every $\tau \in \mathcal{T}$ and $\mathcal{I} \models \alpha$ for every $\alpha \in \mathcal{A}$ such that $\gamma(\alpha) = 0$.

The definition of k-configurations implies in particular that $\gamma(\chi) = 0$ for every $\chi \in \mathcal{T}_\infty \cup \mathcal{A}_\infty$.

Lemma 1. *Let $\mathcal{K}_\omega$ be a WKB and $\mathcal{I}$ be an interpretation.*

$$cost_{\mathcal{K}_\omega}(\mathcal{I}) = \min\{k \mid \exists\gamma \text{ } k\text{-configuration s.t. } \mathcal{I} \text{ satisfies } \gamma\}$$

Proof. If $\mathcal{I}$ satisfies a k-configuration γ, $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$. Indeed, for every $\tau \in \mathcal{T}$, $|vio_\tau(\mathcal{I})| \leq \gamma(\tau)$, and for every $\alpha \in vio_\mathcal{A}(\mathcal{I})$, $\gamma(\alpha) = 1$ because $\gamma(\alpha) = 0$ implies $\mathcal{I} \models \alpha$. Thus $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq \sum_{\tau\in\mathcal{T}} \gamma(\tau)\omega_\tau + \sum_{\alpha\in\mathcal{A}} \gamma(\alpha)\omega_\alpha \leq k$. Moreover, if $cost_{\mathcal{K}_\omega}(\mathcal{I}) = k$, we can define a k-configuration γ such that $\mathcal{I}$ satisfies γ by setting $\gamma(\tau) = |vio_\tau(\mathcal{I})|$ for every $\tau \in \mathcal{T}$, and $\gamma(\alpha) = 0$ if $\mathcal{I} \models \alpha$, $\gamma(\alpha) = 1$ otherwise for every $\alpha \in \mathcal{A}$. ☐

We now define a new KB in a more expressive DL in such a way that the models of the new KB will be interpretations that satisfy a given k-configuration.

Given a concept inclusion $\tau = B \sqsubseteq C$ we define the *violation concept* $V_\tau := B \sqcap \neg C$ such that for every interpretation $\mathcal{I}$, it holds that $vio_\tau(\mathcal{I}) = V_\tau^\mathcal{I}$.

Definition 6. *Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB, k an integer and γ a k-configuration for $\mathcal{K}_\omega$. We define the $\mathcal{ALCOQ}u$ KB $\mathcal{K}_\gamma = \langle \mathcal{T}_\gamma, \mathcal{A}_\gamma\rangle$ associated to $\mathcal{K}_\omega$ and γ as:*

$$\begin{aligned}\mathcal{T}_\gamma =& \mathcal{T}_\infty \cup \{\top \sqsubseteq \leq \gamma(\tau)\, U.V_\tau \mid \tau \in \mathcal{T} \setminus \mathcal{T}_\infty\}\\ \mathcal{A}_\gamma =& \{\alpha \in \mathcal{A} \mid \ \gamma(\alpha) = 0\}.\end{aligned}$$

Proposition 4. *Let $\mathcal{K}_\omega$ be a WKB and γ be a k-configuration for $\mathcal{K}_\omega$. For every interpretation $\mathcal{I}$, $\mathcal{I} \models \mathcal{K}_\gamma$ iff $\mathcal{I}$ satisfies γ.*

Proof. Suppose $\mathcal{I} \models \mathcal{K}_\gamma$. For every $\tau \in \mathcal{T} \setminus \mathcal{T}_\infty$, since $\mathcal{I} \models \top \sqsubseteq \leq \gamma(\tau)U.V_\tau$, then $|vio_\tau(\mathcal{I})| = |V_\tau^\mathcal{I}| \leq \gamma(\tau)$. For every $\tau \in \mathcal{T}_\infty$, since $\mathcal{I} \models \mathcal{T}_\infty$, $|vio_\tau(\mathcal{I})| = 0 = \gamma(\tau)$. Finally, as $\mathcal{I} \models \mathcal{A}_\gamma$, $\mathcal{I}$ satisfies all $\alpha \in \mathcal{A}$ such that $\gamma(\alpha) = 0$.

Conversely, suppose that $\mathcal{I}$ satisfies γ. For every $\tau \in \mathcal{T} \setminus \mathcal{T}_\infty$, $|V_\tau^\mathcal{I}| = |vio_\tau(\mathcal{I})| \leq \gamma(\tau)$ thus $\mathcal{I} \models \top \sqsubseteq \leq \gamma(\tau)U.V_\tau$. For every $\tau \in \mathcal{T}_\infty$, $|vio_\tau(\mathcal{I})| \leq \gamma(\tau) = 0$ thus $\mathcal{I} \models \tau$. Therefore $\mathcal{I} \models \mathcal{T}_\gamma$. As $\mathcal{I}$ satisfies all $\alpha \in \mathcal{A}$ such that $\gamma(\alpha) = 0$, we also have $\mathcal{I} \models \mathcal{A}_\gamma$, so $\mathcal{I} \models \mathcal{K}_\gamma$. ☐

This construction allows us to decide bounded cost satisfiability via $\mathcal{ALCOQ}u$ satisfiability.

Theorem 1. *BCS for $\mathcal{ALCO}$ is in* ExpTime *in combined complexity (even if the bound k and the weights are encoded in binary).*

Proof. Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be a WKB and k be an integer. By Lemma 1 and Proposition 4, $\mathcal{K}_\omega$ is k-satisfiable iff there exists a k-configuration γ such that $\mathcal{K}_\gamma$ is satisfiable. The number of k-configurations γ is bounded by $(k+1)^{|\mathcal{T}|}2^{|\mathcal{A}|}$ (hence by $2^{|k+1||\mathcal{T}|+|\mathcal{A}|}$ if k is encoded in binary and $|k+1|$ is the length of the encoding of $k+1$) as there are at most $k+1$ possibilities for the value of $\gamma(\tau)$ for $\tau \in \mathcal{T}$ and 2 possibilities for the value of $\gamma(\alpha)$ for $\alpha \in \mathcal{A}$. Moreover, for a given γ, $\mathcal{K}_\gamma$ is of polynomial size and can be constructed in polynomial time w.r.t. $|\mathcal{K}_\omega|$ and $|k|$ by encoding the number restrictions in binary (since the numbers occurring in such restrictions are bounded by k). Therefore, as satisfiability in $\mathcal{ZOQ}$ (which extends $\mathcal{ALCOQ}u$) is in ExpTime even with binary encoding in number restrictions (Calvanese, Eiter, and Ortiz 2009), checking for every k-configuration γ whether $\mathcal{K}_\gamma$ is satisfiable is a decision procedure for BCS that runs in exponential time w.r.t. combined complexity. Note that the complexity results for $\mathcal{ALCOQ}u$ apply even if they are shown without the standard name assumption because $\mathcal{K}_\gamma$ is satisfiable under our weak SNA iff $\mathcal{K}_\gamma \cup \{\{a\} \sqcap \{b\} \sqsubseteq \bot \mid a, b \in \mathsf{Ind}(\mathcal{K}), a \neq b\}$ is satisfiable without any assumption on the interpretation of individuals. ☐

To prove the upper bounds on query entailment, we need to first show some results on the computation of the optimal cost of an $\mathcal{ALCO}$ WKB. Since the number of violations of a concept inclusion in an interpretation $\mathcal{I}$ is bounded by the cardinality of its domain $\Delta^\mathcal{I}$, the following proposition is useful to bound the optimal cost of a WKB.

Proposition 5. *Let $\mathcal{K}$ be an $\mathcal{ALCO}$ KB. If $\mathcal{K}$ is satisfiable, then it has a model $\mathcal{I}$ such that $|\Delta^\mathcal{I}| \leq |\mathsf{Ind}(\mathcal{K})| + 2^{|\mathcal{T}|}$.*

Proof sketch. We adapt the proof of $\mathcal{ALC}$ bounded model property by Baader et al. (2017). It is based on the notion of filtration that 'merges' elements that belong to the same concepts and is easily extended to handle nominals. ☐

The following lemma is a consequence of Proposition 5 and the definition of the cost of an interpretation.

Lemma 2. *The optimal cost for an $\mathcal{ALCO}$ WKB $\mathcal{K}_\omega$ (such that $\mathcal{K}_\infty$ is satisfiable) is exponentially bounded in $|\mathcal{K}_\omega|$ (even if the weights are encoded in binary).*

Proof. Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB such that $\mathcal{K}_\infty$ is satisfiable. By Proposition 5, there exists a model $\mathcal{I}$

of $\mathcal{K}_\infty$ such that $|\Delta^\mathcal{I}| \leq |\mathsf{Ind}(\mathcal{K})| + 2^{|\mathcal{T}|} := l$.

$$\begin{aligned} cost_{\mathcal{K}_\omega}(\mathcal{I}) &= \sum_{\tau\in\mathcal{T}} \omega_\tau |vio_\tau(\mathcal{I})| + \sum_{\alpha\in vio_\mathcal{A}(\mathcal{I})} \omega_\alpha \\ &\leq l(\sum_{\tau\in\mathcal{T}\setminus\mathcal{T}_\infty} \omega_\tau) + \sum_{\mathcal{A}\setminus\mathcal{A}_\infty} \omega_\alpha \\ &\leq l|\mathcal{T}| \max_{\tau\in\mathcal{T}\setminus\mathcal{T}_\infty}(\omega_\tau) + |\mathcal{A}| \max_{\alpha\in\mathcal{A}\setminus\mathcal{A}_\infty}(\omega_\alpha) \end{aligned}$$

It follows that $optc(\mathcal{K}_\omega) \leq cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq L$ where $L := (|\mathcal{T}||\mathsf{Ind}(\mathcal{K})| + |\mathcal{T}|2^{|\mathcal{T}|})\max_{\tau\in\mathcal{T}\setminus\mathcal{T}_\infty}(\omega_\tau) + |\mathcal{A}|\max_{\alpha\in\mathcal{A}\setminus\mathcal{A}_\infty}(\omega_\alpha)$. Moreover, since the maximal (finite) weights are at most exponential in $|\mathcal{K}_\omega|$ (even if weights are encoded in binary), L is exponential in $|\mathcal{K}_\omega|$. ☐

Since the optimal cost is exponentially bounded and BCS is in ExpTime, we obtain the following result.

Lemma 3. *Computing the optimal cost of an $\mathcal{ALCO}$ WKB $\mathcal{K}_\omega$ can be done in exponential time in the size of the WKB $|\mathcal{K}_\omega|$ (even if the weights are encoded in binary).*

Proof. Let $\mathcal{K}_\omega = (\langle\mathcal{T},\mathcal{A}\rangle,\omega)$ be an $\mathcal{ALCO}$ WKB. If $\mathcal{K}_\infty$ is not satisfiable, $optc(\mathcal{K}_\omega) = \infty$. Otherwise, by Lemma 2, $optc(\mathcal{K}_\omega) \leq L$ for some $L := 2^{p(|\mathcal{K}_\omega|)}$ where p is a polynomial function. To compute $optc(\mathcal{K}_\omega)$, one can check for every $0 \leq i \leq L$ whether there exists $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq i$. By Theorem 1, each call to BCS takes exponential time w.r.t. $|\mathcal{K}_\omega|$ and the size of the binary encoding of i, which is bounded by $p(|\mathcal{K}_\omega|)$. The whole computation thus takes exponential time w.r.t. $|\mathcal{K}_\omega|$. ☐

We show that BCQ entailment (hence also BIQ entailment) under our variants of the possible semantics can be decided through an exponential number of calls to BCS.

Theorem 2. *CQA^b_p and CQA$^{opt}_p$ for $\mathcal{ALCO}$ are in* ExpTime *in combined complexity (even if the bound k and the weights are encoded in binary).*

Proof. Let $\mathcal{K}_\omega = (\langle\mathcal{T},\mathcal{A}\rangle,\omega)$ be a WKB, k an integer and $q = \exists\vec{y}\psi$ a BCQ with $\psi = \bigwedge_{i=1}^n \varphi_i$ where each φ_i is an atom of the form $A(t)$ or $R(t,t')$ with $t,t' \in \mathsf{N_I} \cup \vec{y}$.

Let $\mathsf{N}_{\vec{y}} \subseteq \mathsf{N_I} \setminus \mathsf{Ind}(\mathcal{K})$ such that $|\mathsf{N}_{\vec{y}}| = |\vec{y}|$, and for every valuation $v : \vec{y} \mapsto \mathsf{Ind}(\mathcal{K}) \cup \mathsf{N}_{\vec{y}}$ let $v(\varphi_i)$ denote the fact obtained by replacing each variable x by $v(x)$ in φ_i and define a WKB: $\mathcal{K}^v_{\omega_v} = (\langle\mathcal{T},\mathcal{A}_v\rangle,\omega_v)$ with $\mathcal{A}_v = \mathcal{A} \cup \{v(\varphi_i) \mid 1 \leq i \leq n\}$ and ω_v extends ω with $\omega_v(v(\varphi_i)) = \infty$ for $1 \leq i \leq n$. We show that $\mathcal{K}_\omega \models^k_p q$ iff there exists v such that $\mathcal{K}^v_{\omega_v}$ is k-satisfiable.
($\Leftarrow$) If there exists v such that $\mathcal{K}^v_{\omega_v}$ is k-satisfiable, let $\mathcal{I}$ be such that $cost_{\mathcal{K}^v_{\omega_v}}(\mathcal{I}) \leq k$. By construction of $\mathcal{K}^v_{\omega_v}$, $\mathcal{I} \models v(\varphi_i)$ for $1 \leq i \leq n$ so v is a match for q in $\mathcal{I}$, i.e., $\mathcal{I} \models q$. Moreover, $cost_{\mathcal{K}_\omega}(\mathcal{I}) = cost_{\mathcal{K}^v_{\omega_v}}(\mathcal{I}) \leq k$. Hence $\mathcal{K}_\omega \models^k_p q$.
($\Rightarrow$) If $\mathcal{K}_\omega \models^k_p q$, there exists $\mathcal{I} \models q$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$. Let π be a match for q in $\mathcal{I}$ (note that $\pi(c) = c$ for every $c \in \mathsf{N_I}$). Consider $\mathsf{D}^\pi_{\vec{y}} := \{\pi(x) \mid x \in \vec{y}\} \setminus \mathsf{Ind}(\mathcal{K})$. Since $|\mathsf{D}^\pi_{\vec{y}}| \leq |\mathsf{N}_{\vec{y}}|$, we can define an injective function f from $\mathsf{D}^\pi_{\vec{y}}$ to $\mathsf{N}_{\vec{y}}$. Let $v : \vec{y} \mapsto \mathsf{Ind}(\mathcal{K}) \cup \mathsf{N}_{\vec{y}}$ such that $v(x) = \pi(x)$ if $\pi(x) \in \mathsf{Ind}(\mathcal{K})$ and $v(x) = f(\pi(x))$ otherwise, and define $\mathcal{I}_v$ by $\Delta^{\mathcal{I}_v} = \Delta^\mathcal{I} \setminus \mathsf{D}^\pi_{\vec{y}} \cup \mathsf{N}_{\vec{y}}$, $c^{\mathcal{I}_v} = c$ for every $c \in \mathsf{N}_{\vec{y}}$, and for every $A \in \mathsf{N_C}$ and $R \in \mathsf{N_R}$, substitute $\pi(x) \in \mathsf{D}^\pi_{\vec{y}}$ with $v(x)$ in $A^\mathcal{I}$ (resp. $R^\mathcal{I}$) to obtain $A^{\mathcal{I}_v}$ (resp. $R^{\mathcal{I}_v}$). By construction of $\mathcal{I}_v$, $\mathcal{I}_v \models v(\varphi_i)$ for $1 \leq i \leq n$. Moreover, for every $\alpha \in \mathcal{A}$, $\mathcal{I}_v \models \alpha$ iff $\mathcal{I} \models \alpha$ and for every $\tau \in \mathcal{T}$, $vio_\tau(\mathcal{I}_v) = vio_\tau(\mathcal{I}) \setminus \mathsf{D}^\pi_{\vec{y}} \cup \{f(e) \mid e \in vio_\tau(\mathcal{I}) \cap \mathsf{D}^\pi_{\vec{y}}\}$. Hence $cost_{\mathcal{K}^v_{\omega_v}}(\mathcal{I}_v) = cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$ and $\mathcal{K}^v_{\omega_v}$ is k-satisfiable.

Therefore, checking for every valuation v (there are at most $(|\mathsf{Ind}(\mathcal{K})| + |q|)^{|q|}$ such valuations) whether $\mathcal{K}^v_{\omega_v}$ is k-satisfiable (in exponential time w.r.t. $|\mathcal{K}^v_{\omega_v}|$ and $|k|$ by Theorem 1, even with binary encoding of k and the weights) yields an ExpTime procedure to decide CQA^b_p.

Regarding CQA$^{opt}_p$, we obtain an ExpTime decision procedure by first computing $optc(\mathcal{K}_\omega)$ in exponential time w.r.t. $|\mathcal{K}_\omega|$ using Lemma 3, then applying the ExpTime procedure for CQA^b_p using $optc(\mathcal{K}_\omega)$ as the bound (since by Lemma 2 $optc(\mathcal{K}_\omega)$ is exponentially bounded in $|\mathcal{K}_\omega|$, its binary encoding is polynomial in $|\mathcal{K}_\omega|$). ☐

Regarding our variants of the certain semantics, we need to distinguish between IQs and CQs.

Theorem 3. *CQA^b_c and CQA$^{opt}_c$ for $\mathcal{ALCO}$ are in* 2ExpTime *in combined complexity and IQA^b_c and IQA$^{opt}_c$ for $\mathcal{ALCO}$ are in* ExpTime *in combined complexity (even if the bound k and the weights are encoded in binary).*

Proof. Let $\mathcal{K}_\omega = (\langle\mathcal{T},\mathcal{A}\rangle,\omega)$ be a WKB, k an integer and q a BCQ. We have that $\mathcal{K}_\omega \models^k_c q$ iff $\mathcal{I} \models q$ for every interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$. By Lemma 1, this is the case iff for every k-configuration γ of $\mathcal{K}_\omega$, for every $\mathcal{I}$ satisfying γ, $\mathcal{I} \models q$. By Proposition 4, this holds iff for every k-configuration γ of $\mathcal{K}_\omega$, for every $\mathcal{I} \models \mathcal{K}_\gamma$, $\mathcal{I} \models q$. Hence we obtain that $\mathcal{K}_\omega \models^k_c q$ iff $\mathcal{K}_\gamma \models q$ for every k-configuration γ of $\mathcal{K}_\omega$. Therefore, checking for every k-configuration γ for $\mathcal{K}_\omega$ whether $\mathcal{K}_\gamma \models q$ yields a decision procedure for CQA^b_c.

To obtain that CQA^b_c is in 2ExpTime in combined complexity and IQA^b_c is in ExpTime in combined complexity, even if the bound k and the weights are encoded in binary, we use the following facts: (i) the number of k-configurations is exponentially bounded and each $\mathcal{K}_\gamma$ is of polynomial size and can be constructed in polynomial time (cf. proof of Theorem 1), (ii) satisfiability of $\mathcal{ALCOQ}u$ is in ExpTime even with binary encoding in number restrictions (cf. proof of Theorem 1), and (iii) BCQ entailment in tame $\mathcal{ZOIQ}$ (which extends $\mathcal{ZOQ}$, hence $\mathcal{ALCOQ}u$) is in 2ExpTime, and in ExpTime in the case of BIQ, even with binary encoding in number restrictions (Bednarczyk and Rudolph 2019, Theorem 8). Note that the complexity results for tame $\mathcal{ZOIQ}$ apply even if they are shown in a context where the SNA is not made because $\mathcal{K}_\gamma \models q$ under our version of the SNA iff $\mathcal{K}_\gamma \cup \{\{a\} \sqcap \{b\} \sqsubseteq \bot \mid a,b \in \mathsf{Ind}(\mathcal{K}), a \neq b\} \models q$ without any assumption on the way the individual names are interpreted.

Regarding CQA$^{opt}_c$ (resp. IQA$^{opt}_c$), we obtain a 2ExpTime (resp. ExpTime) decision procedure as we did in the proof of Theorem 2 by first computing $optc(\mathcal{K}_\omega)$ in exponential time then applying the procedure for CQA^b_c

(resp. IQA$_c^b$) using $optc(\mathcal{K}_\omega)$ as the bound (using the fact that its binary encoding is polynomial in $|\mathcal{K}_\omega|$). $\square$

4.2 Lower Bounds

We first prove the hardness of bounded cost satisfiability for $\mathcal{EL}_\perp$ using a reduction from concept cardinality query answering for $\mathcal{EL}$ KBs (Bienvenu, Manière, and Thomazo 2022). Given a *concept cardinality query* q_A, where $A \in \mathsf{N_C}$, the answer to q_A in an interpretation $\mathcal{I}$, denoted $q_A^\mathcal{I}$, is equal to the cardinality of $A^\mathcal{I}$. A certain answer to q_A w.r.t. a KB $\mathcal{K}$ is an interval $[m, M] \in (\mathbb{N} \cup \{\infty\})^2$ such that $q_A^\mathcal{I} \in [m, M]$ for every model $\mathcal{I}$ of $\mathcal{K}$.

Theorem 4. *BCS for $\mathcal{EL}_\perp$ is* ExpTime*-hard in combined complexity (even if the bound k and weights are in unary).*

Proof. Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be an $\mathcal{EL}$ KB and q_A be a concept cardinality query. Define the following $\mathcal{EL}_\perp$ WKB:

$$\mathcal{K}'_\omega = (\langle \mathcal{T} \cup \{A \sqsubseteq \perp\}, \mathcal{A} \rangle, \omega)$$
$$\omega(\chi) = \infty \text{ for } \chi \in \mathcal{T} \cup \mathcal{A} \qquad \omega(A \sqsubseteq \perp) = 1$$

For every model $\mathcal{I}$ of $\mathcal{K}$, $q_A^\mathcal{I} = |A^\mathcal{I}| = |vio_{A \sqsubseteq \perp}(\mathcal{I})|$, so $cost_{\mathcal{K}'_\omega}(\mathcal{I}) = |vio_{A \sqsubseteq \perp}(\mathcal{I})| = q_A^\mathcal{I}$. It follows that $[m, \infty]$ is a certain answer to q_A iff $cost_{\mathcal{K}'_\omega}(\mathcal{I}) \in [m, \infty]$ for every model $\mathcal{I}$ of $\mathcal{K}$, i.e., iff there is no $\mathcal{I}$ with $cost_{\mathcal{K}'_\omega}(\mathcal{I}) < m$.

As deciding if $[m, \infty]$ is a certain answer to a cardinality query in $\mathcal{EL}$ is ExpTime-hard (Manière 2022, Theorem 42), BCS for $\mathcal{EL}_\perp$ is ExpTime-hard in combined complexity. Moreover, our reduction only uses weights independent from $|\mathcal{K}|$ and the proof of (Manière 2022, Theorem 42) uses an m linear in $|\mathcal{A}|$, so BCS ExpTime-hardness holds even if the bound k and the weights are encoded in unary. $\square$

We next reduce BCS in $\mathcal{EL}_\perp$ to IQA$_p^b$ and IQA$_c^b$ for $\mathcal{EL}_\perp$ to leverage this hardness result to IQ (and thus CQ) answering under the k-cost-bounded semantics.

Theorem 5. *IQA$_p^b$ and IQA$_c^b$ for $\mathcal{EL}_\perp$ are* ExpTime*-hard in combined complexity (even if the bound k and the weights are encoded in unary).*

Proof. Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A} \rangle, \omega)$ be an $\mathcal{EL}_\perp$ WKB and k an integer. Let $B \in \mathsf{N_C} \setminus \mathsf{sig}(\mathcal{K})$ and $b \in \mathsf{N_I} \setminus \mathsf{Ind}(\mathcal{K})$.

First note that $k < optc(\mathcal{K}_\omega)$ iff $\mathcal{K}_\omega \models_c^k B(b)$, due to Proposition 2 and the fact that B and b do not occur in $\mathcal{K}$. As it is ExpTime-hard to decide whether $k < optc(\mathcal{K}_\omega)$ (Theorem 4), this yields the lower bound for IQA$_c^b$.

Now for IQA$_p^b$, let $\mathcal{K}'_{\omega'} = (\langle \mathcal{T}, \mathcal{A} \cup \{B(b)\} \rangle, \omega')$ where ω' extends ω with $\omega'(B(b)) = \infty$. Then $\mathcal{K}'_{\omega'} \models_p^k B(b)$ iff there exists an interpretation $\mathcal{I}$ such that $cost_{\mathcal{K}'_{\omega'}}(\mathcal{I}) \leq k$ and $\mathcal{I} \models B(b)$. For every $\mathcal{I}$ such that $\mathcal{I} \models B(b)$, $cost_{\mathcal{K}_\omega}(\mathcal{I}) = cost_{\mathcal{K}'_{\omega'}}(\mathcal{I})$. Moreover, we can add $b^\mathcal{I}$ to $B^\mathcal{I}$ in any interpretation without changing $cost_{\mathcal{K}_\omega}(\mathcal{I})$ since b and B do not occur in $\mathcal{K}$. Thus $\mathcal{K}'_{\omega'} \models_p^k B(b)$ iff there exists $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$. $\square$

To show the lower bounds for IQ (and thus CQ) answering under the opt-cost semantics, we use a reduction from the problem of deciding if an $\mathcal{EL}$ KB with closed concept names is satisfiable: given a KB and a set of concept names decide if there exists a model $\mathcal{I}$ of the KB such that for every closed concept name A, if $d \in A^\mathcal{I}$ then $A(d)$ is in the ABox.

Theorem 6. *IQA$_p^{opt}$ and IQA$_c^{opt}$ for $\mathcal{EL}_\perp$ are* ExpTime*-hard in combined complexity (even if the weights are encoded in unary).*

Proof sketch. We reduce the ExpTime-hard problem of deciding if an $\mathcal{EL}$ KB with closed concept names is satisfiable (Ngo, Ortiz, and Simkus 2016) to IQA$_p^{opt}$ and IQA$_c^{opt}$. Our reductions are adapted from the proof of the ExpTime-hardness of concept cardinality query answering from (Manière 2022, Theorem 42). $\square$

Finally, we strengthen the lower bounds for CQs and the variants of certain semantics, matching the upper bounds, by adapting a proof for the 2ExpTime-hardness of CQ evaluation on circumscribed $\mathcal{EL}$ KBs. A circumscribed KB specifies that some predicates should be minimized, that is, the extensions of these predicates in models of the circumscribed KB must be minimal regarding set inclusion.

Theorem 7. *CQA$_c^b$ and CQA$_c^{opt}$ for $\mathcal{EL}_\perp$ are* 2ExpTime*-hard in combined complexity (even if the bound k and the weights are encoded in unary).*

Proof sketch. We adapt the proof of the 2ExpTime-hardness of CQ evaluation on circumscribed $\mathcal{EL}$ KBs from (Lutz, Manière, and Nolte 2023, Theorem 2), which proceeds by reduction from the 2ExpTime-hard problem of BCQ entailment for $\mathcal{EL}$ KBs with closed predicates (Ngo, Ortiz, and Simkus 2016). $\square$

5 Data Complexity

We now turn our attention to the data complexity of the decision problems we consider. The second line of Table 1 gives an overview of the results. Recall that data complexity takes only into account the size of the weighted ABox $|\mathcal{A}_\omega|$. In particular, for problems that have an integer k as part of their input, we consider that k is fixed. We will discuss the complexity w.r.t. $|\mathcal{A}_\omega|$ and k at the end of the section.

5.1 Upper Bounds

To obtain the data complexity upper bounds, our general approach is to guess a 'small' interpretation of bounded or optimal cost, that satisfies or does not satisfy the query. It follows from Proposition 5 that if $\mathcal{K}_\infty$ is satisfiable, then $\mathcal{K}_\omega$ admits an interpretation of finite cost whose domain is linearly bounded in the size of the ABox. The following proposition shows that there is always such a bounded-cardinality interpretation which is also of optimal cost.

Proposition 6. *Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A} \rangle, \omega)$ be an $\mathcal{ALCO}$ WKB and k an integer. If there exists an interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$, then there is $\mathcal{I}'$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}') \leq k$ and $|\Delta^{\mathcal{I}'}| \leq |\mathsf{Ind}(\mathcal{K})| + 2^{|\mathcal{T}|}$.*

The upper data complexity bound for bounded cost satisfiability follows directly.

Theorem 8. *BCS for $\mathcal{ALCO}$ is in* NP *in data complexity (even with a binary encoding of weights).*

Proof. Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB and k an integer. By Proposition 6, guessing an interpretation $\mathcal{I}$ of cardinality at most $|\mathsf{Ind}(\mathcal{K})| + 2^{|\mathcal{T}|}$ and checking if $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$ (which can be done in a polynomial time w.r.t $|\mathcal{A}_\omega|$) is an NP procedure to decide BCS. □

To obtain the results for query entailment, we need to refine Proposition 6 to preserve also query (non) entailment.

Proposition 7. *Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB, k an integer and q a BCQ. If there exists an interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$ and $\mathcal{I} \models q$, then there is one whose domain has cardinality at most $|\mathsf{Ind}(\mathcal{K})| + 2^{|\mathcal{T}|}$.*

Preserving query non-entailment is more complex. Contrary to the proofs of Propositions 6 and 7 that build on the notion of filtration used to show the bounded model property of $\mathcal{ALC}(\mathcal{O})$, the following proposition relies on a non-trivial adaptation of constructions introduced in the context of counting queries (Bienvenu, Manière, and Thomazo 2022; Manière 2022). The latter work shows how to convert a (potentially infinite) interpretation into a finite one while avoiding the introduction of new query matches. In our case, we must further prevent new violations of soft TBox axioms.

Proposition 8. *Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB, k an integer, and q a BCQ. If there exists an interpretation $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$ and $\mathcal{I} \not\models q$, then there is one whose domain has cardinality that is bounded polynomially in $|\mathcal{A}|$ and k (with $|\mathcal{T}|$ and $|q|$ treated as constants).*

We are now ready to prove the data complexity upper bounds for the k-cost-bounded semantics.

Theorem 9. *CQA^b_p for $\mathcal{ALCO}$ is in* NP *in data complexity and CQA^b_c for $\mathcal{ALCO}$ is in co*NP *in data complexity (even if the weights are encoded in binary).*

Proof. Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB, k an integer, and q a BCQ. For CQA^b_p, we know from Proposition 7 that it suffices to guess an interpretation $\mathcal{I}$ whose domain has cardinality at most $|\mathsf{Ind}(\mathcal{K})| + 2^{|\mathcal{T}|}$, and check that $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq k$ and $\mathcal{I} \models q$ (both checks being possible in polynomial time w.r.t $|\mathcal{A}_\omega|$), yielding an NP procedure. The argument is similar for CQA^b_c, but uses Proposition 8, which gives the desired polynomial bound in $|\mathcal{A}|$ on the interpretation domain, since k is treated as fixed. □

For the opt-cost semantics, we use the bound on the optimal cost to compute it by binary search before guessing a 'small' interpretation of optimal cost that entails (or does not entail) the query. We recall that Θ^p_2 is the class of problems which are solvable in polynomial time with at most logarithmically many calls to an NP oracle.

Theorem 10. *If there is an ABox-independent bound on the finite weights or weights are encoded in unary, then CQA^{opt}_p and CQA^{opt}_c for $\mathcal{ALCO}$ are in Θ^p_2 in data complexity.*

Proof. Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB, and q a BCQ. Suppose there is a bound on the maximum finite weights that is independent from $|\mathcal{A}|$, or alternatively, that the weights are encoded in unary. Observe that under this assumption, the bound L on the optimal cost given in Lemma 2 is polynomial in $|\mathcal{A}_\omega|$. By doing a binary search using calls to an NP oracle that decides BCS, we can compute the value of $optc(\mathcal{K}_\omega)$ with a logarithmic numbers of such calls. It then suffices to make a final call to an NP (resp. coNP) oracle that decides CQA^b_p (resp. CQA^b_c) (Theorem 9) with $k = optc(\mathcal{K}_\omega)$. Indeed, it is easily seen from the proof of Theorem 9 that the NP/ coNP upper bounds hold not only for fixed k, but also when k is polynomial in $|\mathcal{A}_\omega|$. We obtain Θ^p_2 procedures for deciding CQA^{opt}_p and CQA^{opt}_c. □

In the case where the weights are encoded in binary and not bounded independently from the ABox size, we can further show that opt-cost possible semantics is in Δ^p_2 w.r.t. data complexity (solvable in polynomial time with an NP oracle).

Theorem 11. *CQA^{opt}_p for $\mathcal{ALCO}$ is in Δ^p_2 in data complexity.*

Proof. Let $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ be an $\mathcal{ALCO}$ WKB, q a BCQ, and let L be the bound on the optimal cost given in Lemma 2. The weights of assertions are encoded in $\mathcal{A}_\omega$ so $\max_{\alpha \in \mathcal{A}\setminus\mathcal{A}_\infty}(\omega_\alpha)$ and thus L is at most exponential in $|\mathcal{A}_\omega|$ (in the case of binary encoding). By doing a binary search using calls to an NP oracle that decides BCS we can compute $optc(\mathcal{K}_\omega)$ with a numbers of such calls logarithmic in L (so a polynomial w.r.t. $|\mathcal{A}_\omega|$). Therefore using a final call to an NP oracle that decides CQA^b_p (Theorem 9) with $k = optc(\mathcal{K}_\omega)$ we obtain a Δ^p_2 procedure to decide CQA^{opt}_p. Indeed, even if k might be exponential w.r.t. $|\mathcal{A}_\omega|$, the proof of Theorem 9 for the k-cost-bounded *possible* semantics relies on guessing an interpretation whose domain cardinality is bounded independently from k, so the NP upper bound for CQA^b_p holds even if k depends arbitrarily on $|\mathcal{A}_\omega|$. □

We leave open the question of whether the same upper bound can be obtained for opt-cost certain semantics. The reason is that the proof of Theorem 9 for the k-cost-bounded *certain* semantics relies on guessing an interpretation whose domain cardinality is bounded polynomially in k, hence exponentially in $|\mathcal{A}_\omega|$ when $k = optc(\mathcal{K}_\omega)$ and the weights are encoded in binary and not bounded independently from $|\mathcal{A}|$.

5.2 Lower Bounds

We start by showing that bounded cost satisfiability is NP-hard, using an adaptation of the proof of the NP-hardness of cardinality query answering (Manière 2022, Theorem 48).

Theorem 12. *BCS for $\mathcal{EL}_\bot$ is* NP*-hard in data complexity (even with a unary encoding of weights).*

Proof. We reduce the 3-colorability problem to BCS. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph, and define $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ with:

$$\begin{aligned}
\mathcal{T} =& \{\exists R.C_i \sqcap \exists E.(\exists R.C_i) \sqsubseteq B \mid 1 \leq i \leq 3\} \cup \\
& \{A \sqsubseteq \exists R.B, B \sqsubseteq \bot\} \\
\mathcal{A} =& \{A(v) \mid v \in \mathcal{V}\} \cup \{E(v_1, v_2) \mid \{v_1, v_2\} \in \mathcal{E}\} \cup \\
& \{C_1(c_1), C_2(c_2), C_3(c_3), B(c_1), B(c_2), B(c_3)\}
\end{aligned}$$

$\omega(B \sqsubseteq \bot) = 1$ and $\omega(\chi) = \infty$ for all other $\chi \in \mathcal{T} \cup \mathcal{A}$.

We show that $\mathcal{K}_\omega$ is 3-satisfiable iff $\mathcal{G}$ is 3-colorable.

($\Leftarrow$) If $\mathcal{G}$ is 3-colorable, let $\rho : \mathcal{V} \mapsto \{c_1, c_2, c_3\}$ be a 3-coloring of $\mathcal{G}$ and let $\mathcal{I}_\rho$ be the interpretation that satisfies exactly the assertions of $\mathcal{A} \cup \{R(v, \rho(v)) \mid v \in \mathcal{V}\}$. We show that $\mathcal{I}_\rho \models \mathcal{K}_\infty$, so that $cost_{\mathcal{K}_\omega}(\mathcal{I}_\rho) = |B^{\mathcal{I}_\rho}| = 3$. By construction, $\mathcal{I}_\rho \models \mathcal{A}$ and since $A^{\mathcal{I}_\rho} = \mathcal{V}$, the $R(v, \rho(v))$ assertions ensure that $\mathcal{I}_\rho \models A \sqsubseteq \exists R.B$. Moreover, by definition of ρ, there is no monochromatic edge, which ensures that $(\exists R.C_i \sqcap \exists E.(\exists R.C_i))^{\mathcal{I}_\rho} = \emptyset$ for $i \in \{1, 2, 3\}$.
($\Rightarrow$) If $\mathcal{K}_\omega$ is 3-satisfiable, since $\{c_1, c_2, c_3\} \subseteq B^{\mathcal{I}}$ for any $\mathcal{I}$ of finite cost, then there exists $\mathcal{I}$ such that $cost_{\mathcal{K}_\omega}(\mathcal{I}) = 3$ and $B^{\mathcal{I}} = \{c_1, c_2, c_3\}$. For every $v \in \mathcal{V}$, since $\mathcal{I} \models A(v)$ and $\mathcal{I} \models A \sqsubseteq \exists R.B$, then there exists $c \in \{c_1, c_2, c_3\}$ such that $(v, c) \in R^{\mathcal{I}}$. Define a coloring $\rho : \mathcal{V} \mapsto \{c_1, c_2, c_3\}$ by arbitrarily selecting one such c per v: $(v, \rho(v)) \in R^{\mathcal{I}}$ for each $v \in \mathcal{V}$. We show that ρ is a 3-coloring of $\mathcal{G}$. Otherwise, if there was an edge (v_1, v_2) such that $\rho(v_1) = \rho(v_2) = c_i$, then $\mathcal{I} \models E(v_1, v_2) \wedge R(v_1, c_i) \wedge R(v_2, c_i) \wedge C_i(c_i)$, i.e., $v_1 \in (\exists R.C_i \sqcap \exists E.(\exists R.C_i))^{\mathcal{I}}$. It would follow that $v_1 \in B^{\mathcal{I}}$ and $cost_{\mathcal{K}_\omega}(\mathcal{I}) \geq 4$, contradicting $cost_{\mathcal{K}_\omega}(\mathcal{I}) = 3$. ☐

A direct adaptation of the last proof gives NP and coNP lower bounds for IQ (and thus CQ) answering under the k-cost-bounded semantics.

Theorem 13. *IQA^b_p (resp. IQA^b_c) for $\mathcal{EL}_\perp$ is* NP-*hard (resp. co*NP-*hard) in data complexity (even with a unary encoding of weights).*

Proof. Given a graph $\mathcal{G}$, consider the WKB $\mathcal{K}_\omega$ defined in the proof of Theorem 12 and let $D \in \mathsf{N_C} \setminus \mathsf{sig}(\mathcal{K})$ and $d \in \mathsf{N_I} \setminus \mathsf{Ind}(\mathcal{K})$. Then $\mathcal{G}$ is *not* 3-colorable iff $\mathcal{K}_\omega$ is *not* 3-satisfiable, iff $\mathcal{K}_\omega \models^3_c D(d)$, yielding the lower bound for IQA^b_c. Now for IQA^b_p, let $\mathcal{K}'_{\omega'} = (\langle \mathcal{T}, \mathcal{A} \cup \{D(d)\}\rangle, \omega')$ where ω' extends ω with $\omega'(D(d)) = \infty$. Then $\mathcal{K}'_{\omega'} \models^3_p D(d)$ iff there exists $\mathcal{I}$ with $cost_{\mathcal{K}_\omega}(\mathcal{I}) \leq 3$, iff $\mathcal{G}$ is 3-colorable. ☐

The reduction used for the next theorem is inspired by the proof of Θ^p_2-hardness of the $\leq$-AR semantics (based upon cardinality-maximal repairs, or equivalently, weight-based ABox repairs with all assertions assigned equal weight) for some existential rule languages (Lukasiewicz, Malizia, and Vaicenavičius 2019, Theorem 8).

Theorem 14. *IQA$^{opt}_p$ and IQA$^{opt}_c$ for $\mathcal{EL}_\perp$ are Θ^p_2-hard in data complexity (even if the finite weights are bounded independently from $|\mathcal{A}|$ and encoded in unary).*

Proof sketch. We use a reduction from the problem of deciding whether a given vertex belongs to all the independent sets of maximum size of a graph. ☐

When the assertions weights can be exponential w.r.t. $|\mathcal{A}|$, one can encode priority between the assertions, which we use to prove the following result.

Theorem 15. *IQA$^{opt}_p$ and IQA$^{opt}_c$ for $\mathcal{EL}_\perp$ are Δ^p_2-hard in data complexity (if the weights are encoded in binary).*

Proof sketch. The result is proved by reduction from the following Δ^p_2-hard problem: given a satisfiable Boolean formula $\varphi = c_1 \wedge \ldots \wedge c_m$ over variables $x_1, \ldots, x_n$ such that each clause c_i has exactly two positive and two negative literals (with `true` and `false` admitted as literals) and given $k \in \{1, \ldots, n\}$, decide whether the lexicographically maximum truth assignment ν_{max} satisfying φ w.r.t. $(x_1, \ldots, x_n)$ fulfills $\nu_{max}(x_k) = 1$. It follows from (Krentel 1988) and from the reductions from SAT to 3SAT and from 3SAT to 2+2SAT (Schaerf 1993) that this problem is Δ^p_2-hard.

Let φ be an instance of the problem as previously defined. We define an $\mathcal{EL}_\perp$WKB $\mathcal{K}_\omega = (\langle \mathcal{T}, \mathcal{A}\rangle, \omega)$ as follows.

$$\begin{aligned}
\mathcal{A} =& \{S(\varphi, c_j) \mid 1 \leq j \leq m\} \cup \\
& \{P_\ell(c_k, x_j) \mid \ell \in \{1, 2\}, x_j \text{ is the } \ell^{th} \text{ pos. lit. of } c_k\} \cup \\
& \{N_\ell(c_k, x_j) \mid \ell \in \{1, 2\}, \neg x_j \text{ is the } \ell^{th} \text{ neg. lit. of } c_k\} \cup \\
& \{F(x_i), T(x_i), T'(x_i) \mid 1 \leq i \leq n\} \cup \{F(f), T(t)\} \\
\mathcal{T} =& \{F \sqcap T \sqsubseteq \bot, F \sqcap T' \sqsubseteq \bot\} \cup \\
& \{\exists S.(\exists P_1.F \sqcap \exists P_2.F \sqcap \exists N_1.T \sqcap \exists N_2.T) \sqsubseteq \bot\}
\end{aligned}$$

We set $\omega(\tau) = \infty$ for every $\tau \in \mathcal{T}$, and $\omega(\alpha) = \infty$ for every assertion α built on P_ℓ and N_ℓ as well as for $T(t)$ and $F(f)$. For the remaining assertions, we define the weights through the following prioritization (Bourgaux 2016, Lemma 6.2.5):

- $L_1 = \{T(x_i), F(x_i) \mid 1 \leq i \leq n\}$
- $L_2 = \{S(\varphi, c_j) \mid 1 \leq j \leq m\}$
- $L_3 = \{T'(x_1)\}$, $L_4 = \{T'(x_2)\}, \ldots, L_{n+2} = \{T'(x_n)\}$

Let $u = max(2n, m) + 1$ and define $\omega(\alpha) = u^{n+2-i}$ for every $\alpha \in L_i$. We show that $\mathcal{K}_\omega \models^{opt}_c T'(x_k)$ iff $\mathcal{K}_\omega \models^{opt}_p T'(x_k)$ iff $\nu_{max}(x_k) = 1$. ☐

Interestingly, while the preceding proof crucially relies upon a binary encoding of exponential weights, all of our other lower bounds only employ 1 and ∞ as weights.

5.3 Taking k Into Account

We defined the data complexity as the complexity w.r.t. $|\mathcal{A}_\omega|$ but it is interesting to investigate the impact of considering k as part of the input. Naturally, all data complexity lower bounds hold when k is also part of the input. Moreover, it can be checked that the data complexity upper bounds also hold for the complexity w.r.t. $|\mathcal{A}_\omega|$ and $|k|$, *except for the k-cost-bounded certain semantics* (Theorem 9). The difficulty comes from the proof of Proposition 8: the bound obtained for the cardinality of the domain of the new interpretation is polynomial in k thus becomes exponential in $|k|$ if k is encoded in binary. Therefore, if $|k|$ is treated as part of the input, the proof of Theorem 9 only yields a coNP upper bound for CQA^b_c if we assume a *unary encoding of k*.

6 Related Work

Our cost-based framework for reasoning on inconsistent KBs shares features with a number of existing formalisms. Our axioms with finite weights can be seen as quantitative versions of the soft rules considered in (Eiter, Lukasiewicz, and Predoiu 2016). More generally, the idea of allowing TBox axioms to be (exceptionally) violated shares high-level similarities with defeasible axioms, minimized concepts, and typicality operators considered in non-monotonic extensions of DLs, cf. (Bonatti, Lutz, and Wolter 2009; Giordano et al. 2013; Britz et al. 2021). Although we differ in adopting a quantitative semantics for axiom violations, we were nevertheless able to import some techniques from circumscription (Lutz, Manière, and Nolte 2023) and reasoning with closed predicates (Ngo, Ortiz, and Simkus 2016).

Our approach is related to existing quantitative notions of repair. Indeed, our assignment of costs to interpretations can be seen as adapting the database repairs for soft constraints from (Carmeli et al. 2021) to the DL setting. As detailed in Section 3, our approach is also related to weight-based ABox repairs: when TBox axioms have weight ∞, our opt-cost certain semantics coincides with the $\leq_\omega$-AR semantics from (Bienvenu, Bourgaux, and Goasdoué 2014).

Our proofs rely upon techniques devised for other forms of quantitative reasoning in DLs. We exploit results on counting and cardinality queries (Bienvenu, Manière, and Thomazo 2022; Manière 2022) and reduce some of our problems to reasoning in DLs with number restrictions. We could also have obtained some of our combined complexity upper bounds by using results on *cardinality constraints* $\leq nC$ (which enforce that $|C^{\mathcal{I}}| \leq n$), previously studied in (Tobies 2000) and expressible in the much more expressive quantitative logics in (Baader, Bednarczyk, and Rudolph 2020). We expect that the techniques from these works may prove useful in future studies of WKBs.

7 Conclusion

We have introduced a new cost-based framework to querying inconsistent DL KBs, in which both TBox axioms and ABox assertions may be violated and notions of possible and certain query answers are defined w.r.t. the bounded or minimal-cost interpretations. By exploiting connections to other DL reasoning tasks, we were able to establish an almost complete picture of the complexity of query answering in our framework, for DLs between $\mathcal{EL}_\bot$ and $\mathcal{ALCO}$.

The main aim of future work will be to extend our results to cover DLs involving other common constructs, like inverse roles ($\mathcal{I}$), role inclusions ($\mathcal{H}$), functional roles ($\mathcal{F}$), and number restrictions ($\mathcal{Q}$). We will focus first on the DL-Lite family due to its widespread use in OMQA. While it is not difficult to adapt the definitions of violations and WKBs (although there may be several ways to count violations of a functionality axiom, as one could count as a single violation a domain element that has multiple successors or instead count every pair of two distinct successors), the techniques underlying our upper bounds do not readily transfer. For example, violations of role inclusions cannot be encoded as concepts (at least in standard DLs), and adding functionality or number restrictions may lead to every optimal-cost interpretation having an infinite domain. Inverse roles (present in even the simplest DL-Lite logics) are also non-trivial to handle, though we expect that the techniques employed for our data complexity upper bounds can be suitably adapted. Finally, another important but challenging direction is to devise practical algorithms that are amenable to implementation.

Acknowledgements

We thank the reviewers for their careful reading and useful feedback. This work was partially supported by the ANR AI Chair INTENDED (ANR-19-CHIA-0014) and the ANR project CQFD (ANR-18-CE23-0003).

References

Andolfi, L.; Cima, G.; Console, M.; and Lenzerini, M. 2024. What does a query answer tell you? Informativeness of query answers for knowledge bases. In *Proceedings of AAAI*.

Baader, F., and Kriegel, F. 2022. Pushing optimal ABox repair from $\mathcal{EL}$ towards more expressive Horn-DLs. In *Proceedings of KR*.

Baader, F.; Bednarczyk, B.; and Rudolph, S. 2020. Satisfiability and query answering in description logics with global and local cardinality constraints. In *Proceedings of ECAI*.

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

Bednarczyk, B., and Rudolph, S. 2019. Worst-case optimal querying of very expressive description logics with path expressions and succinct counting. In *Proceedings of IJCAI*.

Bienvenu, M., and Bourgaux, C. 2016. Inconsistency-tolerant querying of description logic knowledge bases. In *Reasoning Web, Tutorial Lectures*.

Bienvenu, M., and Ortiz, M. 2015. Ontology-mediated query answering with data-tractable description logics. In *Reasoning Web*.

Bienvenu, M., and Rosati, R. 2013. Tractable approximations of consistent query answering for robust ontology-based data access. In *Proceedings of IJCAI*.

Bienvenu, M.; Bourgaux, C.; and Goasdoué, F. 2014. Querying inconsistent description logic knowledge bases under preferred repair semantics. In *Proceedings of AAAI*.

Bienvenu, M.; Bourgaux, C.; and Jean, R. 2024. Cost-based semantics for querying inconsistent weighted knowledge bases (Long version with appendix). http://arxiv.org/abs/2407.20754.

Bienvenu, M.; Manière, Q.; and Thomazo, M. 2022. Counting queries over $\mathcal{ELHI}_\bot$ ontologies. In *Proceedings of KR*.

Bienvenu, M. 2020. A short survey on inconsistency handling in ontology-mediated query answering. *Künstliche Intell.* 34(4):443–451.

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. Dissertation, 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.; Eiter, T.; and Ortiz, M. 2009. Regular path queries in expressive description logics with nominals. In *Proceedings of IJCAI*.

Carmeli, N.; Grohe, M.; Kimelfeld, B.; Livshits, E.; and Tibi, M. 2021. Database repairing with soft functional dependencies. In *Proceedings of ICDT*.

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.

Krentel, M. W. 1988. The complexity of optimization problems. *J. Comput. Syst. Sci.* 36(3):490–509.

Lembo, D.; Lenzerini, M.; Rosati, R.; Ruzzi, M.; and Savo, D. F. 2010. Inconsistency-tolerant semantics for description logics. In *Proceedings of RR*.

Lukasiewicz, T.; Malizia, E.; and Vaicenavičius, A. 2019. Complexity of inconsistency-tolerant query answering in Datalog+/– under cardinality-based repairs. *Proceedings of AAAI*.

Lutz, C.; Manière, Q.; and Nolte, R. 2023. Querying circumscribed description logic knowledge bases. In *Proceedings of KR*.

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

Ngo, N.; Ortiz, M.; and Simkus, M. 2016. Closed predicates in description logics: Results on combined complexity. In *Proceedings of KR*.

Ortiz, M., and Simkus, M. 2012. Reasoning and query answering in description logics. In *Reasoning Web*.

Parsia, B.; Sirin, E.; and Kalyanpur, A. 2005. Debugging OWL ontologies. In *Proceedings of WWW*.

Peñaloza, R. 2020. Axiom pinpointing. In *Applications and Practices in Ontology Design, Extraction, and Reasoning*, volume 49 of *Studies on the Semantic Web*. IOS Press. 162–177.

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.

Schaerf, A. 1993. On the complexity of the instance checking problem in concept languages with existential quantification. *J. Intell. Inf. Syst.* 2(3):265–278.

Tobies, S. 2000. The complexity of reasoning with cardinality restrictions and nominals in expressive description logics. *J. Artif. Intell. Res.* 12:199–217.

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*.