Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Counting Queries over $\mathcal{ELHI}_\bot$ Ontologies

Meghyn Bienvenu[1], **Quentin Manière**[1], **Michaël Thomazo**[2]
[1]CNRS, University of Bordeaux, Bordeaux INP, LaBRI, Talence, France
[2]Inria, DI ENS, ENS, CNRS, University PSL, Paris, France
{meghyn.bienvenu, quentin.maniere}@u-bordeaux.fr, michael.thomazo@inria.fr

Abstract

While ontology-mediated query answering most often adopts (unions of) conjunctive queries as the query language, some recent works have explored the use of counting queries coupled with DL-Lite ontologies. The aim of the present paper is to extend the study of counting queries to Horn description logics outside the DL-Lite family. Through a combination of novel techniques, adaptations of existing constructions, and new connections to closed predicates, we achieve a complete picture of the data and combined complexity of answering counting conjunctive queries (CCQs) and cardinality queries (a restricted class of CCQs) in $\mathcal{ELHI}_\bot$ and its various sublogics. Notably, we show that CCQ answering is 2EXP-complete in combined complexity for $\mathcal{ELHI}_\bot$ and every sublogic that extends $\mathcal{EL}$ or DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$. Our study not only provides the first results for counting queries beyond DL-Lite, but it also closes some open questions about the combined complexity of CCQ answering in DL-Lite.

1 Introduction

Ontology-mediated query answering (OMQA) facilitates access to data through the use of ontologies, which provide a convenient vocabulary for query formulation and capture domain knowledge that can be exploited to obtain more complete query results. The OMQA approach has been extensively studied over the past fifteen years (Poggi et al. 2008; Bienvenu and Ortiz 2015; Xiao et al. 2018), leading to the identification of ontology languages that are well suited to OMQA due to their attractive computational properties. Particular attention has been paid to Horn description logics of the DL-Lite and $\mathcal{EL}$ families (Calvanese et al. 2007; Baader, Brandt, and Lutz 2005).

While most work on OMQA considers that the user query is a conjunctive query (CQ), there has been significant interest in exploring the possibility of adopting more expressive query languages for OMQA. In particular, several works have investigated ways of equipping CQs with some form of counting (Calvanese et al. 2008; Kostylev and Reutter 2015; Feier, Lutz, and Przybylko 2021). A recent approach, proposed in (Bienvenu, Manière, and Thomazo 2020) as a generalization of (Kostylev and Reutter 2015), considers *counting conjunctive queries (CCQs)* that are syntactically defined like standard CQs except that some variables may be designated as *counting variables*. In each model of the knowledge base, we can count the number of possible assignments to the counting variables that make the query answer hold. As the count value may differ between models, the goal is to identify intervals that provide upper and lower bounds on the count values across all models.

The problem of answering CCQs is intractable, in both data and combined complexity, for common DL-Lite dialects such as DL-Lite$_{\mathsf{core}}$ and DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$(Kostylev and Reutter 2015). Recent works have shown that intractability arises even for simple forms of CCQs (Calvanese et al. 2020a; Bienvenu, Manière, and Thomazo 2021). However, some interesting tractable cases have also been identified, notably, rooted CCQs (Bienvenu, Manière, and Thomazo 2020; Calvanese et al. 2020a; Nikolaou et al. 2019) and cardinality queries (Bienvenu, Manière, and Thomazo 2021) coupled with DL-Lite$_{\mathsf{core}}$ ontologies. Query rewriting techniques have also begun to be explored (Calvanese et al. 2020b). However, despite these advances, we still have only a partial understanding of CCQ answering in common DL-Lite dialects, and the precise combined complexity has remained elusive: the current bounds for DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ are between coNEXP and coN2EXP (Kostylev and Reutter 2015). Moreover, to the best of our knowledge, CCQ answering has not yet been studied for DLs outside the DL-Lite family.

In this paper, we extend the study of CCQ answering to other well-known Horn description logics, such as $\mathcal{EL}$ and the more expressive $\mathcal{ELHI}_\bot$. The techniques used in the DL-Lite context do not readily transfer to $\mathcal{EL}$ due to the presence of conjunction, and in any case, our results show that they do not achieve the optimal combined complexity even for DL-Lite. We therefore develop a new approach based upon the observation that there exists a model minimizing the count value that consists of an arbitrary structure $\mathcal{I}^*$ containing all assignments for the counting variables, augmented with structures that are tree-shaped, provided we ignore edges to and from $\mathcal{I}^*$. Importantly, we can bound the size of the central component $\mathcal{I}^*$, which enables us to explore all possible options for $\mathcal{I}^*$. Checking whether a given $\mathcal{I}^*$ can be extended to a model preserving the minimum count value can be done by specifying a set of *patterns* (intuitively representing a pair of adjacent elements), and testing via local consistency conditions whether they can be coherently assembled. This latter step takes inspiration from a CQ answering technique for existential rules

	Combined complexity						*Data complexity*
	DL-Lite$_{\mathsf{pos}}$	DL-Lite$_{\mathsf{core}}$	DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$	DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$	$\mathcal{EL}(\mathcal{H}_\bot), \mathcal{EL}(\mathcal{HI})$	$\mathcal{EL}(\mathcal{H})\mathcal{I}_\bot$	$\mathcal{EL}(\mathcal{HI}_\bot)$
Concept	NL	coNP	NL	coNP‡	EXP	coNEXP	coNP
Role	NL	coNP	coNP‡	coNP‡	EXP	coNEXP	coNP
CCQ	coNEXP†	coNEXP†	2EXP	2EXP	2EXP	2EXP	coNP

Figure 1: Complexity results for CCQs and cardinality queries, all bounds are tight. †/ ‡: previously known upper / lower bound.

(Thomazo et al. 2012), and is also similar in spirit to type-elimination style procedures, which have been employed for reasoning with expressive DLs, see e.g. (Rudolph, Krötzsch, and Hitzler 2012; Eiter et al. 2009).

Using this new approach, we are able to establish a 2EXP upper bound in combined complexity for $\mathcal{ELHI}_\bot$. A matching lower bound, which applies to both $\mathcal{EL}$ and DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$, is obtained by establishing a novel connection between CCQ answering and OMQA with closed predicates. This yields 2EXP-completeness for a wide range of Horn DLs and closes the combined complexity gap for CCQ answering in DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$. We further prove a coNEXP lower bound for DL-Lite$_{\mathsf{pos}}$, which matches an existing coNEXP upper bound, yielding the precise combined complexity for DL-Lite$_{\mathsf{core}}$ as well. We also explore how to shrink the size of the models implicitly generated by our procedure, producing models with bounded size which we use to show that CCQ answering is coNP-complete in data complexity for all logics between $\mathcal{EL}$ and $\mathcal{ELHI}_\bot$.

In addition to CCQs, we also investigate the special case of cardinality queries, which correspond to Boolean atomic CCQs and allow us to ask for (bounds on) the number of members of a given concept or role. We obtain a complete picture of data and combined complexity of answering cardinality queries in $\mathcal{ELHI}_\bot$ and its various sublogics. While the data complexity is coNP-complete for all considered logics, the combined complexity ranges from NL or coNP in DL-Lite logics to EXP or coNEXP for $\mathcal{EL}$ and its extensions. We achieve these results using a variety of the techniques: refinements of our approach for general CCQs, adaptations of existing constructions, and further reductions involving closed predicates. Figure 1 summarizes the complexity results for both CCQs and cardinality queries.

Paper Organization Section 2 introduces the necessary preliminaries, in particular, the syntax and semantics of the considered DLs and the definition of CCQs. Sections 3 and 4 present our complexity results for CCQs and cardinality queries, respectively, and sketch the underlying techniques (an appendix with full proofs can be found in the long version of this paper, available on arXiv). Section 5 concludes with a discussion of future work.

2 Preliminaries

Knowledge Bases We assume mutually disjoint sets $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ of *concept, role, and individual names*. A *knowledge base (KB)* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ consists of an ABox $\mathcal{A}$ and a TBox $\mathcal{T}$. An *ABox* is a finite set of *concept assertions* $\mathrm{A}(\mathsf{b})$ (with $\mathrm{A} \in \mathsf{N_C}$, $\mathsf{b} \in \mathsf{N_I}$) and *role assertions* $\mathrm{P}(\mathsf{a}, \mathsf{b})$ (with $\mathrm{P} \in \mathsf{N_R}$, $\mathsf{a}, \mathsf{b} \in \mathsf{N_I}$). We denote by $\mathsf{Ind}(\mathcal{A})$ the set of individuals occurring in an ABox $\mathcal{A}$.

A *TBox* is a finite set of axioms. In $\mathcal{ELHI}_\bot$, TBoxes consist of *concept inclusions* $\mathrm{B}_1 \sqsubseteq \mathrm{B}_2$, *positive role inclusions* $\mathrm{R}_1 \sqsubseteq \mathrm{R}_2$, and *negative role inclusions*[1] $\mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$, where the $\mathrm{R_i}$ are *roles* drawn from $\mathsf{N}^{\pm}_{\mathsf{R}} = \{\mathrm{P}, \mathrm{P}^- \mid \mathrm{P} \in \mathsf{N_R}\}$ and the $\mathrm{B_i}$ are *(complex) concepts* constructed as follows:

$$\mathrm{B} := \bot \mid \top \mid \mathrm{A} \mid \mathrm{B}_1 \sqcap \mathrm{B}_2 \mid \exists \mathrm{R}.\mathrm{B} \quad \text{with } \mathrm{A} \in \mathsf{N_C}, \mathrm{R} \in \mathsf{N}^{\pm}_{\mathsf{R}}$$

Various sublogics of $\mathcal{ELHI}_\bot$ can be obtained by disallowing role inclusions, inverse roles, and/or the bottom construct. For example, $\mathcal{EL}$ is obtained by removing all three features, while $\mathcal{ELI}_\bot$ corresponds to disallowing role inclusions (retaining inverse roles and $\bot$). We shall also consider some DL-Lite dialects that are fragments of $\mathcal{ELHI}_\bot$. The most expressive, DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$, allows positive and negative role inclusions, and restricted forms of concept inclusions:

$$\mathrm{D}_1 \sqsubseteq \mathrm{D}_2 \quad \mathrm{D}_1 \sqcap \mathrm{D}_2 \sqsubseteq \bot \quad \mathrm{D_i} := \mathrm{A} \mid \exists \mathrm{R}.\top$$

with $\mathrm{A} \in \mathsf{N_C}, \mathrm{R} \in \mathsf{N}^{\pm}_{\mathsf{R}}$. The logics DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$, DL-Lite$_{\mathsf{core}}$, and DL-Lite$_{\mathsf{pos}}$ are obtained respectively by dropping negative inclusions, role inclusions, or both features.

We shall use $\mathsf{sig}(\mathcal{T})$ (resp. $\mathsf{sig}(\mathcal{K})$) to denote the *signature* of a TBox $\mathcal{T}$ (resp. KB $\mathcal{K}$), i.e. the set of concept and role names appearing in $\mathcal{T}$ (resp. $\mathcal{K}$).

Semantics of KBs An interpretation takes the form $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set (called the domain) and $\cdot^{\mathcal{I}}$ is the interpretation function that maps each $\mathrm{A} \in \mathsf{N_C}$ to $\mathrm{A}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, each $\mathrm{P} \in \mathsf{N_R}$ to $\mathrm{P}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and each $\mathsf{a} \in \mathsf{N_I}$ to $\mathsf{a}^{\mathcal{I}}$. In this paper, we will make the *Standard Names Assumption* by setting $\mathsf{a}^{\mathcal{I}} = \mathsf{a}$. Note however that our results only rely upon the weaker *Unique Names Assumption* (UNA), which stipulates that $\mathsf{a}^{\mathcal{I}} \neq \mathsf{b}^{\mathcal{I}}$ whenever $\mathsf{a} \neq \mathsf{b}$.

The function $\cdot^{\mathcal{I}}$ naturally extends to roles and complex concepts: $(\mathrm{P}^-)^{\mathcal{I}} = \{(y, x) \mid (x, y) \in \mathrm{P}^{\mathcal{I}}\}$, $\bot^{\mathcal{I}} = \emptyset$, $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$, $(\mathrm{B}_1 \sqcap \mathrm{B}_2)^{\mathcal{I}} = \mathrm{B}_1^{\mathcal{I}} \cap \mathrm{B}_2^{\mathcal{I}}$ and $(\exists \mathrm{P}.\mathrm{B})^{\mathcal{I}} = \{d \mid (d, e) \in \mathrm{P}^{\mathcal{I}}, e \in \mathrm{B}^{\mathcal{I}}\}$. An inclusion $\mathrm{G} \sqsubseteq \mathrm{H}$ is satisfied in $\mathcal{I}$ if $\mathrm{G}^{\mathcal{I}} \subseteq \mathrm{H}^{\mathcal{I}}$; an assertion $\mathrm{A}(\mathsf{b})$ (resp. $\mathrm{P}(\mathsf{a}, \mathsf{b})$) is satisfied in $\mathcal{I}$ if $\mathsf{b} \in \mathrm{A}^{\mathcal{I}}$ (resp. $(\mathsf{a}, \mathsf{b}) \in \mathrm{P}^{\mathcal{I}}$). An interpretation is a *model* of a TBox $\mathcal{T}$ (resp. KB $\mathcal{K}$) if it satisfies all axioms in $\mathcal{T}$ (resp. axioms and assertions in $\mathcal{K}$). A KB is *satisfiable* if it has at least one model. An inclusion (resp. assertion) Φ is *entailed* from $\mathcal{T}$ (resp. $\mathcal{K}$), written $\mathcal{T} \models \Phi$ (resp. $\mathcal{K} \models \Phi$), if Φ is satisfied in every model of $\mathcal{T}$ (resp. $\mathcal{K}$).

[1]We follow e.g. (Bienvenu et al. 2014) by including negative role inclusions in $\mathcal{ELHI}_\bot$, so that it has DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ as a sublogic.

Example 1. *Consider the ABox* $\mathcal{A}_e := \{\mathrm{A}_1(\mathsf{a}), \mathrm{B}(\mathsf{b})\}$ *and the* $\mathcal{ELHI}_\perp$ *TBox* $\mathcal{T}_e$*:*

$$\begin{array}{llll} \mathrm{A}_1 \sqsubseteq \exists \mathrm{R}.\mathrm{A}_2 & \mathrm{A}_2 \sqsubseteq \exists \mathrm{R}.\mathrm{A}_1 & \mathrm{B} \sqsubseteq \exists \mathrm{R}.\mathrm{B} & \mathrm{R} \sqcap \mathrm{R}^- \sqsubseteq \bot \\ \mathrm{A}_2 \sqsubseteq \exists \mathrm{R}.\overline{\mathrm{B}} & \overline{\mathrm{B}} \sqsubseteq \exists \mathrm{R}.\mathrm{C} & \mathrm{B} \sqsubseteq \mathrm{C} & \mathrm{B} \sqcap \overline{\mathrm{B}} \sqsubseteq \bot \end{array}$$

Our example KB is $\mathcal{K}_e := (\mathcal{T}_e, \mathcal{A}_e)$*. Figures 2a and 2c depict models of* $\mathcal{K}_e$*.*

We can view an interpretation $\mathcal{I}$ as a (possibly infinite) set of assertions $\mathcal{A}_\mathcal{I} = \{\mathrm{A}(e) \mid e \in \mathrm{A}^\mathcal{I}, \mathrm{A} \in \mathsf{N_C}\} \cup \{\mathrm{P}(e, e') \mid (e, e') \in \mathrm{P}^\mathcal{I}, \mathrm{P} \in \mathsf{N_R}\}$. We say that $\mathcal{I}$ is *$\mathcal{T}$-satisfiable* if $\mathcal{T} \cup \mathcal{A}_\mathcal{I}$ has a model, and it is *$\mathcal{T}$-saturated* if $\mathcal{A}_\mathcal{I}$ contains every assertion entailed by $(\mathcal{T}, \mathcal{A}_\mathcal{I})$.

Counting Queries We consider counting queries as defined in (Bienvenu, Manière, and Thomazo 2020) (which generalizes the queries considered in (Kostylev and Reutter 2015; Calvanese et al. 2020a)). A *counting conjunctive query (CCQ)* takes the form $q(\mathbf{x}) = \exists \mathbf{y} \exists \mathbf{z}\, \psi(\mathbf{x}, \mathbf{y}, \mathbf{z})$, where $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are tuples of *answer, existential, and counting variables*, respectively, and ψ is a conjunction of concept and role atoms with terms from $\mathsf{N_I} \cup \mathbf{x} \cup \mathbf{y} \cup \mathbf{z}$. We use $\mathsf{terms}(q)$ for the set of all terms occurring in q, and we treat queries as sets of atoms when convenient. The usual notion of *conjunctive query (CQ)* is captured by CCQs without counting variables (i.e. $\mathbf{z} = \emptyset$). A CCQ q is *Boolean* if $\mathbf{x} = \emptyset$. *Concept cardinality queries* are Boolean CCQs of the form $\exists z\ \mathrm{A}(z)$ ($\mathrm{A} \in \mathsf{N_C}$), while *role cardinality queries* have the form $\exists z_1, z_2\ \mathrm{R}(z_1, z_2)$ ($\mathrm{R} \in \mathsf{N_R}$).

A *match* for a CCQ q in an interpretation $\mathcal{I}$ is a homomorphism from q into $\mathcal{I}$, i.e. a function π that maps each term in q to an element of $\Delta^\mathcal{I}$ such that $\pi(t) = t$ when $t \in \mathsf{N_I}$, $\pi(t) \in \mathrm{A}^\mathcal{I}$ for every $\mathrm{A}(t) \in q$, and $(\pi(t), \pi(t')) \in \mathrm{P}^\mathcal{I}$ for every $\mathrm{P}(t, t') \in q$. If a match π maps $\mathbf{x}$ to $\mathbf{a}$, then the restriction of π to $\mathbf{z}$ is called a *counting match (c-match)* of $q(\mathbf{a})$ in $\mathcal{I}$. The set of *answers* to q in $\mathcal{I}$, denoted $q^\mathcal{I}$, contains all pairs $(\mathbf{a}, [m, M])$, with $m, M \in \mathbb{N} \cup \{+\infty\}$, such that the number of distinct c-matches of $q(\mathbf{a})$ in $\mathcal{I}$ belongs to the interval $[m, M]$. A *certain answer* to q w.r.t. $\mathcal{K}$ is an answer in every model of $\mathcal{K}$, that is a pair from $\bigcap_{\mathcal{I} \models \mathcal{K}} q^\mathcal{I}$.

As usual, it is sufficient to consider the Boolean case: $(\mathbf{a}, [m, M])$ is a certain answer to a CCQ $q(\mathbf{x})$ iff $(\emptyset, [m, M])$ is a certain answer to the Boolean CCQ $q(\mathbf{a})$ obtained by replacing $\mathbf{x}$ with $\mathbf{a}$. Thus, from now on, *we focus on Boolean CCQs*, and work with candidate answers $[m, M]$ in place of $(\emptyset, [m, M])$.

We further observe that since $\mathcal{ELHI}_\perp$ cannot restrict the size of models, the least upper bound M in a certain answer $[m, M]$ is: 0 if the underlying CQ is unsatisfiable w.r.t. $\mathcal{T}$, 1 if q has a match in every model but $\mathbf{z} = \emptyset$; and $+\infty$ otherwise. As the first two cases can be readily handled using existing techniques, *we focus on identifying certain answers of the form* $[m, +\infty]$.

Example 2. *Let* $q_e := \exists y\ \exists z\ \mathrm{R}(y, z) \wedge \mathrm{C}(z)$ *be a Boolean CCQ. Intervals* $[0, +\infty]$ *and* $[1, +\infty]$ *are certain answers to* q_e *over* $\mathcal{K}_e$*. Interval* $[4, +\infty]$ *is not as the models depicted on Figures 2a and 2c contain only* 3 *matches for* q_e*.*

To clarify how our notion of certain answer relates to standard OMQA semantics, we note that a Boolean CQ q is entailed from $\mathcal{K}$ iff $[1, +\infty]$ is a certain answer to q over $\mathcal{K}$.

Complexity Given a $\mathcal{ELHI}_\perp$ knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a Boolean CCQ q, and an integer $m \geq 0$ (in binary), we are interested in the complexity of deciding whether $[m, +\infty]$ is a certain answer to q w.r.t. $\mathcal{K}$. We will consider the two usual complexity measures: *combined complexity* which is in terms of the size of the whole input, and *data complexity* which is only in terms of the size of $\mathcal{A}$ and m ($\mathcal{T}$ and q are treated as fixed). If O is a TBox, ABox, KB, or CCQ, then the *size* of O, denoted $|O|$, is the number of occurrences of concept and role names in O.

Normal form As is standard (see e.g. (Bienvenu et al. 2014)), we work with $\mathcal{ELHI}_\perp$ TBoxes in a convenient *normal form*, where every concept inclusion has one of the following restricted shapes:

$$\begin{array}{ccc} \mathrm{A} \sqsubseteq \bot & \top \sqsubseteq \mathrm{A} & \mathrm{A}_1 \sqcap \mathrm{A}_2 \sqsubseteq \mathrm{A} \\ \mathrm{A}_1 \sqsubseteq \exists \mathrm{R}.\mathrm{A}_2 & & \exists \mathrm{R}.\mathrm{A}_1 \sqsubseteq \mathrm{A}_2 \end{array}$$

with $\mathrm{A}, \mathrm{A}_1, \mathrm{A}_2 \in \mathsf{N_C}, \mathrm{R} \in \mathsf{N}_\mathsf{R}^\pm$. Through the introduction of fresh concept names, we can transform in polynomial time any TBox $\mathcal{T}$ into a normal-form TBox $\mathcal{T}'$ that is a model-conservative extension of $\mathcal{T}$ (hence, indistinguishable from $\mathcal{T}$ from the point of view of queries). We therefore *assume w.l.o.g. that all considered TBoxes are in normal form.*

Closed Predicates A *KB with closed predicates* consists of a KB $(\mathcal{T}, \mathcal{A})$ and a set $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$ of *closed predicates*. An interpretation $\mathcal{I}$ is a *model of* $(\mathcal{T}, \mathcal{A}, \Sigma)$ if it is a model of $(\mathcal{T}, \mathcal{A})$ which interprets the closed predicates according to $\mathcal{A}$, i.e. $\mathrm{A}^\mathcal{I} = \{\mathsf{a} \mid \mathrm{A}(\mathsf{a}) \in \mathcal{A}\}$ for every $\mathrm{A} \in \Sigma \cap \mathsf{N_C}$ and $\mathrm{P}^\mathcal{I} = \{(\mathsf{a}, \mathsf{b}) \mid \mathrm{P}(a, b) \in \mathcal{A}\}$ for every $\mathrm{P} \in \Sigma \cap \mathsf{N_R}$. Query entailment is then defined as for classical KBs, but using this modified notion of model.

3 General Case of CCQs

This section presents our main contributions: a decision procedure and associated tight complexity bounds for CCQ answering in $\mathcal{ELHI}_\perp$ and its sublogics.

To improve readability, we have split the section into several parts. Section 3.1 presents a double-exponential-time decision procedure, whose correctness proof is detailed in Section 3.2. We explain, in Section 3.3, how to shrink the size of the models implicitly generated by our procedure, which we use to show coNP data complexity. Finally, in Section 3.4, we prove the required lower bounds.

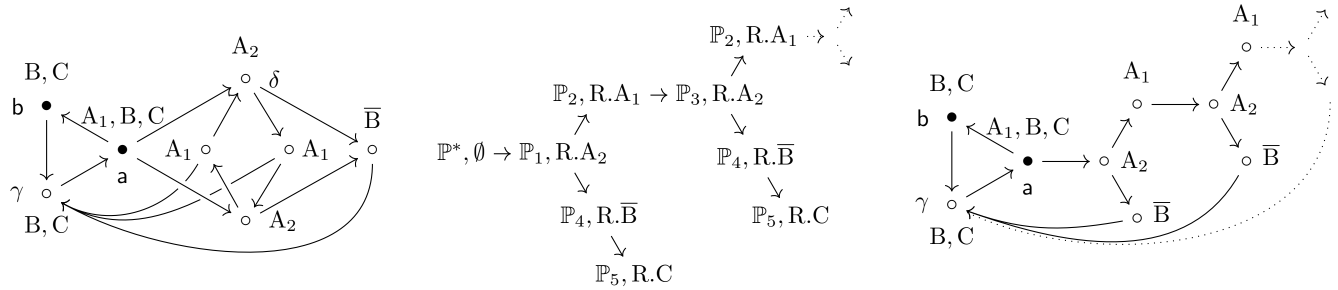
3.1 Decision Procedure

In this subsection, we devise a procedure that computes in double-exponential time the minimum amount of c-matches, which immediately yields the following upper bound:

Theorem 1. *CCQ answering in* $\mathcal{ELHI}_\perp$ *is in* 2EXP *w.r.t. combined complexity.*

Let us fix a satisfiable KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ and a (Boolean) CCQ q. The next lemma provides an upper bound on the minimal number of c-matches.

Lemma 1. *There exists a model of* $\mathcal{K}$ *with less than* $M := (|\mathsf{Ind}(\mathcal{A})| + 3\, |\mathcal{T}|\, 2^{|\mathcal{T}|})^{|q|}$ *c-matches for* q*.*



(a) A model $\mathcal{I}_e$ of $\mathcal{K}_e$. (b) Pattern tree $\mathcal{P}_e$. (c) Model $\mathcal{I}_e^{\diamondsuit}$ built from $\mathbb{P}_e^*$.

Figure 2: Interpretations and pattern tree used along our examples, labels for the only role R have been omitted for readability.

Proof sketch. We can exhibit a model having at most $|\mathsf{Ind}(\mathcal{A})| + 3\,|\mathcal{T}|\,2^{|\mathcal{T}|}$ elements. □

It follows that in any model $\mathcal{I}$ having a minimum number of c-matches, the set $\Delta^* \subseteq \Delta^{\mathcal{I}}$ of elements appearing in the image of a c-match has size at most $M \cdot |q|$. We can thus iterate over all such Δ^*, and even over all induced interpretations $\mathcal{I}^* = \mathcal{I}_{|\Delta^*}$, in double-exponential time w.r.t. combined complexity. The core task will then be to determine, given such a candidate $\mathcal{I}^*$, whether we can extend $\mathcal{I}^*$ into a model of $\mathcal{K}$ *without introducing new c-matches*.

Let us fix our candidate $\mathcal{I}^*$ and see how to check for a suitable extension. The challenging axioms to handle are those of the form $\mathrm{A} \sqsubseteq \exists\mathrm{R.B}$, as they might require us to introduce new elements. We define the set $\Omega := \{\mathrm{R.B} \mid \mathrm{A} \sqsubseteq \exists\mathrm{R.B} \in \mathcal{T}\}$ and call its members *(existential) heads*. Importantly, as our correctness proof will establish, it is sufficient to consider extensions of $\mathcal{I}^*$ which are obtained by adding tree-shaped structures of new elements, plus some edges between the new elements and $\Delta^{\mathcal{I}^*}$ (we may need to use elements from $\Delta^{\mathcal{I}^*}$ as witnesses for existential heads to avoid new query matches). This property enables us to build such an extension by piecing together local interpretations corresponding to the addition of a single edge, using two distinguished symbols $\circledcirc$ and $\circledast$ as placeholders for fresh elements. We shall call these building blocks *patterns*, as they are inspired by a notion of the same name introduced for CQ answering with existential rules (Thomazo et al. 2012).

Patterns not only consist of a local interpretation, but also other information needed to ensure that assembled patterns do not violate any TBox axioms or introduce any new matches. In particular, we shall keep track of (partial) query matches involving the local elements using the notion of a coherent specification. Intuitively, such a specification tells us which matches should be realized in the constructed extension, and naturally contains at least the matches of subqueries of q already realized in the local interpretation.

Definition 1. *Let $\mathcal{I}$ be an interpretation.*

- *The* specification $\mathfrak{M}^{\mathcal{I}}$ induced by $\mathcal{I}$ *is the set of pairs (r,π) such that $r \subseteq q$ and $\pi : r \to \mathcal{I}$ is a (full) match.*
- *A* coherent specification $\mathfrak{M}$ *over $\mathcal{I}$ is a set of pairs (r,π) where $r \subseteq q$ and π is a partial mapping from* $\mathsf{terms}(r)$ *to $\Delta^{\mathcal{I}}$ such that $\mathfrak{M}$ contains $\mathfrak{M}^{\mathcal{I}}$ and if $(r_1,\pi_1),(r_2,\pi_2) \in \mathfrak{M}$ with π_1 and π_2 defined and equal on $\mathsf{var}(r_1)\cap\mathsf{var}(r_2)$, then $(r_1 \cup r_2, \pi_1 \cup \pi_2) \in \mathfrak{M}$.*

To check the compatibility of different specifications, we will need to be able to restrict them to a subdomain:

Definition 2. *The* restriction of a specification $\mathfrak{M}$ *over an interpretation $\mathcal{I}$ to a domain $\Delta \subseteq \Delta^{\mathcal{I}}$, denoted $\mathfrak{M}_{|\Delta}$, is the set of pairs (r,π') such that π' is the restriction of some π to $\pi^{-1}(\Delta)$ for some $(r,\pi) \in \mathfrak{M}$.*

Remark 1. *Induced specifications and restrictions of coherent specifications are both coherent specifications.*

Patterns will contain a further kind of information called a prediction, defined next. The purpose will be explained in more detail once we introduce links between patterns, but roughly it serves to coordinate the successor patterns of a pattern to avoid violating negative role inclusions.

Definition 3. *A* prediction *is a function* $\mathsf{next} : \Omega \to \Delta^{\mathcal{I}^*} \cup \Omega$ *verifying that: for all $\mathrm{R}_1.\mathrm{B}_1, \mathrm{R}_2.\mathrm{B}_2 \in \Omega$, if $\mathcal{T} \models \mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$, then* $\mathsf{next}(\mathrm{R}_1.\mathrm{B}_1) \neq \mathsf{next}(\mathrm{R}_2.\mathrm{B}_2)$.

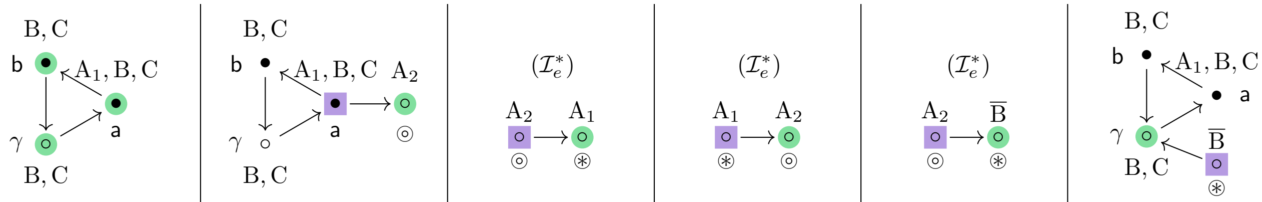
We now formally define the central notion of pattern, relative to $\mathcal{I}^*$ and a candidate specification $\mathfrak{M}^*$ over $\mathcal{I}^*$.

Definition 4. *A* pattern $\mathbb{P}$ *(w.r.t. $\mathcal{I}^*$ and $\mathfrak{M}^*$) is a tuple $(\mathfrak{fr}^{\mathbb{P}}, \mathfrak{gen}^{\mathbb{P}}, \mathfrak{I}^{\mathbb{P}}, \mathfrak{M}^{\mathbb{P}}, \mathsf{next}_{\mathbb{P}})$ where:*

- *The frontier and generated domains $\mathfrak{fr}^{\mathbb{P}}$ and $\mathfrak{gen}^{\mathbb{P}}$ are* disjoints *sets of elements from $\Delta^{\mathcal{I}^*} \cup \{\circledcirc, \circledast\}$;*
- *$\mathfrak{I}^{\mathbb{P}}$ is a $\mathcal{T}$-saturated and $\mathcal{T}$-satisfiable interpretation with $\Delta^{\mathfrak{I}^{\mathbb{P}}} = \Delta^{\mathcal{I}^*} \cup \mathfrak{fr}^{\mathbb{P}} \cup \mathfrak{gen}^{\mathbb{P}}$ and such that $\mathfrak{I}^{\mathbb{P}}_{|\Delta^{\mathcal{I}^*}} = \mathcal{I}^*$;*
- *$\mathfrak{M}^{\mathbb{P}}$ is a coherent specification of q over $\mathfrak{I}^{\mathbb{P}}$ that preserves $\mathfrak{M}^*$, that is $(\mathfrak{M}^{\mathbb{P}})_{|\Delta^{\mathcal{I}^*}} = \mathfrak{M}^*$;*
- $\mathsf{next}_{\mathbb{P}}$ *is a prediction.*

We shall be interested in two types of patterns. The (unique) *initial pattern* $\mathbb{P}^* := (\emptyset, \Delta^{\mathcal{I}^*}, \mathcal{I}^*, \mathfrak{M}^*, \mathsf{Id})$ simply represents $\mathcal{I}^*$ and $\mathfrak{M}^*$. All other patterns of interest represent additions of a pair of adjacent elements, and $\mathfrak{fr}^{\mathbb{P}}$ and $\mathfrak{gen}^{\mathbb{P}}$ will be singletons (representing these two elements).

Example 3. *In our running example, $\Delta^*_e := \{\mathsf{a}, \mathsf{b}, \gamma\}$ (z maps to only these elements). The initial pattern $\mathbb{P}^*_e$ has frontier $\emptyset$, generated terms Δ^*_e, interpretation $\mathcal{I}^*_e := (\mathcal{I}_e)_{|\Delta^*_e}$ depicted in Figure 3a, and specification $\mathfrak{M}^*_e := (\mathfrak{M}^{\mathcal{I}_e})_{|\Delta^*_e}$. Non-initial patterns will be illustrated later.*



(a) Interp. $\mathcal{I}^*_e$. (b) Interp. of $\mathbb{P}^e_1$. (c) Interp. of $\mathbb{P}^e_2$. (d) Interp. of $\mathbb{P}^e_3$. (e) Interp. of $\mathbb{P}^e_4$. (f) Interp. of $\mathbb{P}^e_5$.

Figure 3: Interpretations of patterns from Example 4.

We now define how to combine patterns together, and first, *when* it is necessary to combine them.

Definition 5. *We say that* $\mathrm{R.B} \in \Omega$ *is* applicable *to* e *in a pattern* $\mathbb{P}$ *if* $e \in \mathfrak{gen}^{\mathbb{P}}$ *and there exists* $\mathrm{A} \sqsubseteq \exists \mathrm{R.B} \in \mathcal{T}$ *with* $e \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}}}$ *but* $e \notin (\exists \mathrm{R.B})^{\mathfrak{I}^{\mathbb{P}}}$.

When a head is applicable to a pattern, we need to find another pattern that can realize the head. This is formalized by the following notion of link between patterns, which requires that the two patterns are compatible (Conditions 1, 2, 3), the second pattern realizes the head (Condition 4), and certain consistency conditions hold (Conditions 5, 6).

Definition 6. *Let* $\mathrm{R.B}$ *be an applicable head on* e_1 *in a pattern* $\mathbb{P}_1$. *There is a* $(\mathrm{R.B}, e_1)$*-link from* $\mathbb{P}_1$ *to* $\mathbb{P}_2$ *if:*

1. $\mathfrak{fr}^{\mathbb{P}_2} = \{e_1\}$ *and* $\mathfrak{gen}^{\mathbb{P}_2}$ *is a singleton, say* $\{e_2\}$*;*
2. *For all concept name* A*, we have* $e_1 \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_1}}$ *iff* $e_1 \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_2}}$*;*
3. $\mathfrak{M}^{\mathbb{P}_1}_{|\Delta^{\mathcal{I}^*} \cup \{e_1\}} = \mathfrak{M}^{\mathbb{P}_2}_{|\Delta^{\mathcal{I}^*} \cup \{e_1\}}$*;*
4. $e_2 \in \mathrm{B}^{\mathfrak{I}^{\mathbb{P}_2}}$ *and for all* $\mathrm{P} \in \mathsf{N_R}$*:* $\mathrm{P}^{\mathfrak{I}^{\mathbb{P}_2}} = \mathrm{P}^{\mathcal{I}^*} \cup \{(e_1, e_2) \mid \mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}\} \cup \{(e_2, e_1) \mid \mathcal{T} \models \mathrm{R}^- \sqsubseteq \mathrm{P}\}$
5. *If ever* $e_2 \in \Delta^{\mathcal{I}^*} \cap \mathfrak{fr}^{\mathbb{P}_1}$*, then* $\mathfrak{I}^{\mathbb{P}_1} \cup \mathfrak{I}^{\mathbb{P}_2}$ *is* $\mathcal{T}$*-satisfiable.*
6. *If* $e_2 \in \Delta^{\mathcal{I}^*}$*, then* $e_2 = \mathsf{next}_{\mathbb{P}_1}(\mathrm{R.B})$.

We denote $\mathbb{L}^{\mathrm{R.B}}_{\mathbb{P}_1, e_1}$ *the set of patterns* $\mathbb{P}_2$ *such that there is a* $(\mathrm{R.B}, e_1)$*-link from* $\mathbb{P}_1$ *to* $\mathbb{P}_2$.

Remark 2. *Predictions are used in Condition 6 to avoid problematic situations where two successor patterns merge back to the same element of* $\Delta^{\mathcal{I}^*}$. *Specifically, if we have a* $\mathrm{R}_1.\mathrm{B}_1$*-link from* $\mathbb{P}_0$ *to* $\mathbb{P}_1$ *and a* $\mathrm{R}_2.\mathrm{B}_2$*-link from* $\mathbb{P}_0$ *to* $\mathbb{P}_2$*, with* $\mathcal{T} \models \mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$*, then* $\mathsf{next}_{\mathbb{P}_0}(\mathrm{R}_1.\mathrm{B}_1) \neq \mathsf{next}_{\mathbb{P}_0}(\mathrm{R}_2.\mathrm{B}_2)$*, preventing* $\mathbb{P}_1$ *and* $\mathbb{P}_2$ *from using the same element of* $\Delta^{\mathcal{I}^*}$ *as generated term (which would violate* $\mathcal{T}$*). Condition 5 is similar in spirit, handling the case of the pattern* $\mathbb{P}_1$ *using the frontier element of* $\mathbb{P}_0$ *as a generated term.*

Example 4. *We consider patterns* $\mathbb{P}^e_1, \ldots, \mathbb{P}^e_5$ *whose interpretations are depicted in Figure 3. Frontier terms are indicated by square-purple and generated terms by circle-green. Predictions are* Id *except for* $\mathsf{next}_{\mathbb{P}^e_4}$*, which maps* $\mathrm{R.C}$ *to* γ. *Specifications* $\mathfrak{M}_i$ *are given by:* $\mathfrak{M}_1 = \mathfrak{M}^*_e \cup \{(\alpha_\mathrm{R}, (y, z) \mapsto (\mathsf{a}, \circledcirc)), (\alpha_\mathrm{R}, z \mapsto \circledcirc), (\alpha_\mathrm{R}, y \mapsto \circledcirc)\}$*;* $\mathfrak{M}_5 = \mathfrak{M}^*_e \cup \{(q_e, (y, z) \mapsto (\circledast, \gamma)), (\alpha_\mathrm{R}, (y, z) \mapsto (\circledast, \gamma)), (\alpha_\mathrm{R}, z \mapsto \circledast)\}$*;* $\mathfrak{M}_4 = \mathfrak{M}_1 \cup \mathfrak{M}_5 \cup \{(\alpha_\mathrm{R}, (y, z) \mapsto (\circledcirc, \circledast))\}$*;* $\mathfrak{M}_2 = \mathfrak{M}_4 \setminus \{(\alpha_\mathrm{R}, z \mapsto \circledast)\}$*;* $\mathfrak{M}_3 = \mathfrak{M}_5 \cup \{(\alpha_\mathrm{R}, (y, z) \mapsto (\mathsf{a}, \circledcirc)), (\alpha_\mathrm{R}, (y, z) \mapsto (\circledast, \circledcirc)), (\alpha_\mathrm{R}, y \mapsto \circledcirc)\}$*, where* α_R *denotes* $\mathrm{R}(y, z)$. *Observe that* $\mathfrak{M}_i$ *may include (partial) matches which are not present in* $\mathbb{P}^e_i$*'s interpretation but are useful for linking patterns, e.g.* $(q_e, (y, z) \mapsto (\circledast, \gamma))$ *in* $\mathfrak{M}_4$ *enables a* $(\mathrm{R.C}, \circledast)$*-link from* $\mathbb{P}^e_4$ *to* $\mathbb{P}^e_5$ *(see Example 5).*

We now characterize patterns that cannot be used to satisfy a head without introducing a new c-match.

Definition 7. *A pattern* $\mathbb{P}$ *is* rejecting *if one of the two following conditions holds:*

- *There exists* $(q, \pi) \in \mathfrak{M}^{\mathbb{P}}$ *with* $\pi(\mathbf{z}) \cap \{\circledcirc, \circledast\} \neq \emptyset$*;*
- *There exists an existential head* $\mathrm{R.B}$ *that applies on* e *in* $\mathbb{P}$ *such that all patterns* $\mathbb{P}' \in \mathbb{L}^{\mathrm{R.B}}_{\mathbb{P}, e}$ *are rejecting.*

A pattern is accepting if it is not rejecting.

The acceptance of the initial pattern $\mathbb{P}^*$ is a sufficient condition ensuring $\mathcal{I}^*$ extends to a model having no more c-matches than encoded in $\mathfrak{M}^*$, i.e. the pairs $(q, \pi) \in \mathfrak{M}^*$ such that π is defined for all counting variables.

Lemma 2. *If* $\mathbb{P}^* := (\emptyset, \Delta^*, \mathcal{I}^*, \mathfrak{M}^*, \mathsf{Id})$ *is accepting, then there exists a model* $\mathcal{I}^\diamondsuit$ *such that* $\mathcal{I}^* \subseteq \mathcal{I}^\diamondsuit$ *and if* $\pi : q \to \mathcal{I}^\diamondsuit$ *is a c-match, then* $(q, \pi) \in \mathfrak{M}^*$. *In particular,* $\mathcal{I}^\diamondsuit$ *has at most as many c-matches as those encoded in* $\mathfrak{M}^*$.

Furthermore, the minimum amount of c-matches is reached among initial patterns due to the following result:

Lemma 3. *If* $\mathcal{I}$ *is a model of* $\mathcal{K}$ *with* m *c-matches, then there exists an accepting initial pattern whose specification encodes* exactly m *c-matches.*

Before proving Lemmas 2 and 3, let us recap the overall double-exponential procedure underlying Theorem 1:

Proof of Theorem 1. We consider all possible initial patterns $\mathbb{P}^*$ with an interpretation domain Δ^* such that $\mathsf{Ind}(\mathcal{A}) \subseteq \Delta^*$ and $|\Delta^*| \leq M|q|$ (recall Lemma 1). Every such $\mathbb{P}^*$ is of single-exponential size w.r.t. combined complexity (observe that its specification $\mathfrak{M}^*$ corresponds to a subset of $2^q \times (\Delta^* \cup \{\uparrow\})^q$, where $\uparrow$ is a fresh symbol witnessing the use of *partial* mappings), and thus are double-exponential in number (up to isomorphism) and can be enumerated in double-exponential time. For each such $\mathbb{P}^*$, we construct in double-exponential time the set of all possible descendant patterns of $\mathbb{P}^*$ (which are of single-exponential size, having at most $|\Delta^*| + 2$ elements). We then check whether each possible pattern ($\mathbb{P}^*$ or candidate descendant) is in fact a well-defined pattern, in particular, its interpretation is $\mathcal{T}$-satisfiable and $\mathcal{T}$-saturated. These verifications can be done in double-exponential time, recalling that KB satisfiability and instance checking are in EXP for $\mathcal{ELHI}_\bot$ (even the variant with negative role inclusions, see e.g. (Bienvenu et al. 2014)). Acceptance of $\mathbb{P}^*$ is tested (again in deterministic exponential time) by repeatedly iterating over the set of patterns and removing those that are rejecting either due to their specification, or due to the removal of all patterns that could provide a link for an applicable head. If $\mathbb{P}^*$ is found to be accepting and $\mathfrak{M}^*$ encodes m c-matches, then Lemma 2 ensures the existence of a model with at most m c-matches. Conversely, Lemma 3 ensures that we can find the smallest such m among the accepting initial patterns. □

3.2 Proofs of Lemmas 2 and 3

We now prove the central lemmas of the correctness proof.

From Accepting Patterns to Models To prove Lemma 2, let us suppose we are given an initial pattern $\mathbb{P}^* := (\emptyset, \Delta^*, \mathcal{I}^*, \mathfrak{M}^*, \mathsf{Id})$ that is accepting. Our aim is to construct a model $\mathcal{I}^\diamondsuit$ that extends $\mathcal{I}^*$ and is such that $(q, \pi) \in \mathfrak{M}^*$ for every c-match $\pi : q \to \mathcal{I}^\diamondsuit$.

We proceed as follows. For each accepting descendant pattern $\mathbb{P}$ (w.r.t. $\mathcal{I}^*$ and $\mathfrak{M}^*$) and each head $\mathrm{R.B}$ applicable to e in $\mathbb{P}$, we choose an accepting pattern $\mathsf{ch}^{\mathrm{R.B}}_{\mathbb{P},e}$ from $\mathbb{L}^{\mathrm{R.B}}_{\mathbb{P},e}$. Then, starting from $\mathbb{P}^*$, we build a tree-shaped set of words, whose letters consist of an accepting pattern and existential head, and which witnesses the acceptance of $\mathbb{P}^*$.

Definition 8. *The* pattern tree $\mathcal{P}$ *is defined as the smallest set of words such that:*

- $(\mathbb{P}^*, \emptyset) \in \mathcal{P}$*;*
- *If* $w \cdot (\mathbb{P}, h) \in \mathcal{P}$ *and* $\mathrm{R.B}$ *is applicable to* e *in* $\mathbb{P}$*, then* $w \cdot (\mathbb{P}, h) \cdot (\mathsf{ch}^{\mathrm{R.B}}_{\mathbb{P},e}, \mathrm{R.B}) \in \mathcal{P}$*.*

It remains to 'glue' together the interpretations $\mathfrak{I}^{\mathbb{P}}$ according to the structure of $\mathcal{P}$. Since a pattern $\mathbb{P}$ may occur more than once, we create a copy of $\mathfrak{I}^{\mathbb{P}}$ for each node in $\mathcal{P}$ of the form $w \cdot (\mathbb{P}, h)$. We do not duplicate however elements from $\mathcal{I}^*$ as they precisely are those we want to reuse. Hence only the frontier term and the generated term may be duplicated (provided they do not belong to Δ^*). When a node $w \cdot (\mathbb{P}_1, h_1) \cdot (\mathbb{P}_2, h_2)$ is encountered, we merge the frontier term of $\mathbb{P}_2$ with the already-introduced copy of the generated element from $\mathbb{P}_1$ on which h_2 is applied (which is the only element in $\mathfrak{fr}^{\mathbb{P}_2}$). Therefore, when considering such a node $w \cdot (\mathbb{P}_1, h_1) \cdot (\mathbb{P}_2, h_2)$, the only element we might have to introduce is a copy of the generated term e of $\mathbb{P}_2$ (unless $e \in \Delta^*$), which we shall simply name $w \cdot (\mathbb{P}_1, h_1) \cdot (\mathbb{P}_2, h_2)$. Formally, the copying and merging of elements is achieved by the following family of duplicating functions, defined inductively for each $w \cdot (\mathbb{P}, h) \in \mathcal{P}$.

$$\lambda_{w\cdot(\mathbb{P},h)} : \Delta^{\mathfrak{I}^{\mathbb{P}}} \to \Delta^{\mathcal{I}^*} \cup \{w, w \cdot (\mathbb{P}, h)\}$$
$$e \mapsto \begin{cases} e & \text{if } e \in \Delta^{\mathcal{I}^*} \\ w & \text{if } e \in \mathfrak{fr}^{\mathbb{P}} \setminus \Delta^{\mathcal{I}^*} \\ w \cdot (\mathbb{P}, h) & \text{if } e \in \mathfrak{gen}^{\mathbb{P}} \setminus \Delta^{\mathcal{I}^*} \end{cases}$$

Note that if $e \in \mathfrak{fr}^{\mathbb{P}_2} \setminus \Delta^{\mathcal{I}^*}$, then $e \in \mathfrak{gen}^{\mathbb{P}_1} \setminus \Delta^{\mathcal{I}^*}$, hence

$$\lambda_{w\cdot(\mathbb{P}_1,h_1)\cdot(\mathbb{P}_2,h_2)}(e) = \lambda_{w\cdot(\mathbb{P}_1,h_1)}(e) = w \cdot (\mathbb{P}_1, h_1).$$

The desired model $\mathcal{I}^\diamondsuit$ can then be defined as follows:

$$\mathcal{I}^\diamondsuit := \bigcup_{w\cdot(\mathbb{P},h)\in\mathcal{P}} \lambda_{w\cdot(\mathbb{P},h)}(\mathfrak{I}^{\mathbb{P}}).$$

Example 5. *The patterns introduced in Example 4 are sufficient to witness that* $\mathbb{P}^*_e$ *is accepting. The initial part of* $\mathcal{P}_e$ *is depicted in Figure 2b. The resulting* $\mathcal{I}^\diamondsuit_e$ *is depicted in Figure 2c. Notice how it inherits the tree-shaped structure of* $\mathcal{P}_e$ *up to roles collapsing back in* $\mathcal{I}^*_e$*.*

By definition, each $\lambda_{w\cdot(\mathbb{P},h)}$ is a homomorphism from $\mathfrak{I}^{\mathbb{P}}$ to $\mathcal{I}^\diamondsuit$. Due to Condition 2, the shared element of linked patterns must belong to the same concepts, so concept membership in $\mathcal{I}^\diamondsuit$ transfers back to $\mathfrak{I}^{\mathbb{P}}$:

Lemma 4. *For all* $w \cdot (\mathbb{P}, h) \in \mathcal{P}$*, for all* $e \in \Delta^{\mathfrak{I}^{\mathbb{P}}}$ *and for all* $\mathrm{A} \in \mathsf{N_C}$*, if* $\lambda_{w\cdot(\mathbb{P},h)}(e) \in \mathrm{A}^{\mathcal{I}^\diamondsuit}$*, then* $e \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}}}$*.*

An analogous property fails however for roles, as two patterns $\mathbb{P}_1 = \mathsf{ch}^{h_1}_{\mathbb{P},e}$ and $\mathbb{P}_2 = \mathsf{ch}^{h_2}_{\mathbb{P},e}$ may reuse the same element from Δ^*, that is, $\mathfrak{gen}^{\mathbb{P}_1} = \mathfrak{gen}^{\mathbb{P}_2} \in \Delta^*$. In that case, satisfied roles in $\mathcal{I}^\diamondsuit_{|\Delta}$ where $\Delta := \lambda_{w\cdot(\mathbb{P},h)\cdot(\mathbb{P}_1,h_1)}(\mathfrak{I}^{\mathbb{P}_1})$ may not be satisfied in $\mathfrak{I}^{\mathbb{P}_1}$. Conditions 5 and 6 allow us to show the following weaker property, sufficient for our purposes:

Lemma 5. *For all* $w \cdot (\mathbb{P}, h) \in \mathcal{P}$*,* $d, e \in \Delta^{\mathfrak{I}^{\mathbb{P}}}$*, and* $\mathrm{P} \in \mathsf{N_R}$*: if* $(\lambda_{w\cdot(\mathbb{P},h)}(d), \lambda_{w\cdot(\mathbb{P},h)}(e)) \in \mathrm{P}^{\mathcal{I}^\diamondsuit}$*, then* $\mathfrak{I}^{\mathbb{P}}$ *remains* $\mathcal{T}$*-satisfiable if we add* (d, e) *to* $\mathrm{P}^{\mathfrak{I}^{\mathbb{P}}}$*.*

A similar lemma (given in the appendix) allows us to lift query matches from $\mathcal{I}^\diamondsuit$ to patterns, yielding the following:

Proposition 1. $\mathcal{I}^\diamondsuit$ *is a model of* $\mathcal{K}$ *whose c-matches are included in those encoded in* $\mathfrak{M}^*$*.*

From a Model to an Accepting Initial Pattern We now turn to the proof of Lemma 3. We fix a model $\mathcal{I}$ of $\mathcal{K}$, and our task is to construct an accepting initial pattern having the same number of c-matches as $\mathcal{I}$.

Let Δ^* be the subset of $\Delta^{\mathcal{I}}$ consisting of all individuals in $\mathcal{A}$ and all elements e such that $e = \pi(z)$ for some $\pi : q \to \mathcal{I}$ and counting variable z. Set $\mathcal{I}^* := \mathcal{I}_{|\Delta^*}$ and $\mathfrak{M}^* := (\mathfrak{M}^{\mathcal{I}})_{|\Delta^*}$. Notice in particular that the number of c-matches for q encoded in $\mathfrak{M}^*$ is exactly the number of c-matches for q in $\mathcal{I}$. We claim that $\mathbb{P}^* := (\emptyset, \Delta^*, \mathcal{I}^*, \mathfrak{M}^*, \mathsf{Id})$ is accepting.

To prove this, we shall build a set of patterns, whose every pattern $\mathbb{P}$ is *not trivially rejecting*, i.e. $\mathbb{P}$ does not satisfy the base-case condition of a rejecting pattern, and which is *realized in* $\mathcal{I}$, meaning that $\mathfrak{I}^{\mathbb{P}}$ homomorphically embeds into $\mathcal{I}$. Observe that the initial pattern $\mathbb{P}^*$ satisfies both conditions. To pursue the construction, given any pattern $\mathbb{P}$ satisfying the two conditions and a head h applicable to $\mathbb{P}$, we show how to extract from $\mathcal{I}$ another $\mathbb{Q}$ which satisfies the conditions and which makes h hold for $\mathbb{P}$. Since the number of patterns is finite, every sequence of patterns constructed in such a manner either leads to a trivially accepting pattern (i.e. one with no applicable heads) or loops back to an already explored pattern satisfying the conditions. It follows that all patterns in the set are accepting (in particular, $\mathbb{P}^*$).

To formalize the construction, we shall introduce a function τ associating to each pattern $\mathbb{P}$ a homomorphism $\mathfrak{I}^{\mathbb{P}} \to \mathcal{I}$. Furthermore, we shall assume that, for every $\mathrm{R.A} \in \Omega$, we have chosen a function $\mathsf{succ}^{\mathcal{I}}_{\mathrm{R.A}}$ that maps every element $e \in (\exists \mathrm{R.A})^{\mathcal{I}}$ to an element $e' \in \Delta^{\mathcal{I}}$ such that $(e, e') \in \mathrm{R}^{\mathcal{I}}$ and $e' \in \mathrm{A}^{\mathcal{I}}$. The construction begins with $\mathbb{P}^*$, for which we set $\tau(\mathbb{P}^*) := \mathsf{Id}_{\mathcal{I}^* \to \mathcal{I}}$. Next we take some already constructed pattern $\mathbb{P}_1$ with its associated function $\tau(\mathbb{P}_1)$, and consider a head R.B that is applicable to e_1 in $\mathbb{P}_1$. Since R.B applies to e, there must exist $\mathrm{A} \in \mathsf{N_C}$ such that $e \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_1}}$ and $\mathcal{T} \models \mathrm{A} \sqsubseteq \exists\mathrm{R.B}$. Set $e_1' := \tau(\mathbb{P}_1)(e_1)$. Since $\tau(\mathbb{P}_1)$ is a homomorphism and $\mathcal{I}$ is a model of $\mathcal{T}$, we obtain $e_1' \in (\exists\mathrm{R.B})^{\mathcal{I}}$ and can set $e_2' := \mathsf{succ}^{\mathcal{I}}_{\mathrm{R.B}}(e_1')$. If $e_2' \in \Delta^*$, then we set $e_2 := e_2'$, otherwise we set e_2 to either $\circledcirc$ or $\circledast$ such that $e_1 \neq e_2$.

We can now define the new pattern $\mathbb{P}_2$. Its frontier is e_1 and its generated term is e_2. Its interpretation is given by:

$$\begin{array}{rcl} \mathrm{C}^{\mathfrak{I}^{\mathbb{P}_2}} & := & \mathrm{C}^{\mathcal{I}^*} \cup \{e_k \mid e_k' \in \mathrm{C}^{\mathcal{I}}, k = 1, 2\} \\ \mathrm{P}^{\mathfrak{I}^{\mathbb{P}_2}} & := & \mathrm{P}^{\mathcal{I}^*} \cup \{(e_1, e_2) \mid \mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}\} \\ & & \cup \{(e_2, e_1) \mid \mathcal{T} \models \mathrm{R}^- \sqsubseteq \mathrm{P}\} \end{array}$$

Its specification is $(\mathfrak{M}^{\mathcal{I}})_{|\Delta^* \cup \{e_1', e_2'\}}$ in which e_1' (resp. e_2') has been replaced by e_1 (resp. e_2). Its prediction maps a head h to the value of $\mathsf{succ}^{\mathcal{I}}_h(e_2')$ if it is defined, else to h. Finally, we let $\tau(\mathbb{P}_2)$ be the function that maps elements of Δ^* to themselves, e_1 to e_1' and e_2 to e_2'. Recalling that $\mathcal{I}$ is a model, of $\mathcal{K}$ it is then straightforward to verify that $\mathbb{P}_2$ is a well-defined not-trivially-rejecting pattern, satisfying $\mathbb{P}_2 \in \mathbb{L}^{\mathrm{R.B}}_{\mathbb{P}_1, e_1}$, and such that $\tau(\mathbb{P}_2)$ is indeed a homomorphism.

Example 6. *In the model $\mathcal{I}_e$, depicted in Figure 2a, we can set $\mathsf{succ}^{\mathcal{I}_e}_{\mathrm{R.A}_2}(\mathsf{a}) := \delta$ (other choices of successors are unique), and then apply the preceding construction to obtain the accepting patterns from Example 4.*

3.3 Obtaining Bounded-Size Optimal Models

To obtain optimal models of bounded size, we start from the pattern tree $\mathcal{P}$ and model $\mathcal{I}^{\diamondsuit}$ we constructed from an accepting initial pattern. It remains to merge elements of $\mathcal{I}^{\diamondsuit}$ to obtain a model of the required size. To identify similar elements, we consider their neighbourhoods.

Definition 9. *Consider an interpretation $\mathcal{I}$ and an element $c \in \Delta^{\mathcal{I}}$. Its n-*neighbourhood *$\mathcal{N}^{\mathcal{I},\Delta}_n(c)$ w.r.t. a subdomain $\Delta \subseteq \Delta^{\mathcal{I}}$ is defined inductively as:*

$$\begin{array}{l} \mathcal{N}^{\mathcal{I},\Delta}_0(c) := \{c\} \\ \mathcal{N}^{\mathcal{I},\Delta}_{n+1}(c) := \mathcal{N}^{\mathcal{I},\Delta}_n(c) \cup \left\{ e \;\middle|\; \begin{array}{l} \exists d \in \mathcal{N}^{\mathcal{I},\Delta}_n(c) \setminus \Delta, \\ \exists \mathrm{R} \in \mathsf{N}^{\pm}_{\mathsf{R}},\ (d, e) \in \mathrm{R}^{\mathcal{I}} \end{array} \right\} \end{array}$$

Observe that we stop adding successors when we reach Δ.

To characterize neighbourhoods in $\mathcal{I}^{\diamondsuit}$ (w.r.t. domain Δ^*), we focus on the tree-like structure inherited from $\mathcal{P}$. Recall that we kept a *single* pattern for each head R.B applicable to an element e of a pattern $\mathbb{P}$, namely $\mathsf{ch}^{\mathrm{R.B}}_{\mathbb{P},e}$. We can thus consider the bijection σ mapping $(\mathbb{P}^*, \emptyset) \cdot (\mathbb{P}_1, h_1) \cdots (\mathbb{P}_n, h_n)$ (with $n \geq 1$) to $ah_1 \ldots h_n$, where a is such that $\mathfrak{fr}^{\mathbb{P}_1} = \{a\}$; we extend σ to Δ^* by letting $\sigma(e) = e$ for $e \in \Delta^*$. Inspired by the notion of interleaving used in the DL-Lite setting (Kostylev and Reutter 2015), we define the *interlacing* $\mathcal{I}' := \sigma(\mathcal{I}^{\diamondsuit})$, obtained by renaming elements of $\mathcal{I}^{\diamondsuit}$ using σ. Denote by $\Delta^\circ := \Delta^* \cup \sigma(\mathcal{P} \setminus \mathbb{P}^*)$ the forest-shaped domain that is to $\mathcal{I}'$ what $\mathcal{P}$ is to $\mathcal{I}^{\diamondsuit}$. We define an associated mapping $f' : \Delta^\circ \to \mathcal{I}'$ by setting $f' := \sigma \circ f \circ \sigma^{-1}$ where f maps each element of Δ^* to itself and each $w \cdot (\mathbb{P}, h) \in \mathcal{P}$ to $\lambda_{w \cdot (\mathbb{P},h)}(e)$ where $\mathfrak{gen}^{\mathbb{P}} = \{e\}$.

The definition of $\mathcal{I}'$ ensures that every $c \in \Delta^{\mathcal{I}'} \setminus \Delta^*$ belongs to $\sigma(\mathcal{P} \setminus \mathbb{P}^*)$ and thus $c = aw$ for some $a \in \Delta^*$ and word $w \in \Omega^*$. The tree-shaped structure of Δ° ensures that for all n, there exists a unique prefix $r_{n,c}$ of $\mathsf{a}w$ such that *(i)* $f'(r_{n,c}) \in \mathcal{N}^{\mathcal{I}',\Delta^*}_n(c)$ and *(ii)* for any $d \in \mathcal{N}^{\mathcal{I}',\Delta^*}_n(c)$, there exists a unique word $w^d_{n,c}$ such that $d = f'(r_{n,c} \cdot w^d_{n,c})$.

This leads us to characterize the n-neighbourhood of an element $c \in \mathcal{I}'$ via the following function $\chi_{n,c}$, whose domain Ω_n is the set of words over Ω with length $\leq 2n$. Notice that, departing from (Kostylev and Reutter 2015), we keep track of *sets* of satisfied concepts, in order to handle conjunctions of concepts in the left-hand sides of axioms.

$$\begin{array}{l} \chi_{n,c} : \Omega_n \to \Delta^* \cup 2^{\mathsf{sig}(\mathcal{T})} \cup \{\emptyset\} \\ w \mapsto \begin{cases} \emptyset & \text{if } f'(r_{n,c}w) \text{ undefined} \\ f'(r_{n,c}w) & \text{if } f'(r_{n,c}w) \in \Delta^* \\ \{\mathrm{A} \in \mathsf{sig}(\mathcal{T}) \mid f'(r_{n,c}w) \in \mathrm{A}^{\mathcal{I}'}\} & \text{otherwise} \end{cases} \end{array}$$

We can now introduce the equivalence relation we use to merge elements:

Definition 10. *The equivalence relation $\sim_n$ on $\Delta^{\mathcal{I}'}$ is defined as follows: an element $e \in \Delta^*$ is $\sim_n$-equivalent only to itself; elements c_1, c_2 from $\Delta^{\mathcal{I}'} \setminus \Delta^*$ are $\sim_n$-equivalent iff $w^{c_1}_{n,c_1} = w^{c_2}_{n,c_2}$, $\chi_{n,c_1} = \chi_{n,c_2}$, and $|c_1| = |c_2| \bmod 2|q|+3$.*

We obtain a finite model of the required size by merging elements with respect to $\sim_{|q|+1}$.

Theorem 2. *The interpretation $\mathcal{J} := \mathcal{I}'/{\sim_{|q|+1}}$ is a model of $\mathcal{K}$ that has at most as many c-matches for q as $\mathcal{I}^{\diamondsuit}$. Its size is polynomial w.r.t. data complexity, double-exponential w.r.t. combined complexity, and single-exponential if the size of the CCQ q is fixed.*

Proof sketch. The key to proving that the amount of c-matches does not increase through the quotient operation is to exhibit suitable local homomorphisms. Indeed, a match of q in $\mathcal{J}$ maps each connected component C of q into a $|q|$-neighbourhood $\mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_{|q|}(\overline{c})$, where $\overline{c}$ denotes the equivalence class of c w.r.t. $\sim_{|q|+1}$ and $\overline{\Delta^*}$ stands for the set $\{\overline{e} \mid e \in \Delta^*\}$. By exhibiting a homomorphism ρ_c : $\mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_{|q|}(\overline{c}) \to \mathcal{N}^{\mathcal{I}',\Delta^*}_{|q|}(c)$ such that $\rho^{-1}_{n,c}(\overline{\Delta^*}) \subseteq \overline{\Delta^*}$, we can find a match of C in $\mathcal{I}'$. Such matches for q's connected components together form a match of the full q in $\mathcal{I}'$. It is mostly straightforward to show that $\mathcal{J}$ is a model, except for negative role inclusions, where the homomorphisms ρ_c are needed to move violations of $\mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$ in $\mathcal{J}$ back into $\mathcal{I}'$. The claimed upper bounds are obtained by analyzing the size of $\mathcal{J}$ (i.e. counting the equivalence classes in $\Delta^{\mathcal{J}}$), keeping in mind that due to Lemma 1, we may assume that $|\Delta^*| \leq |\mathsf{Ind}(\mathcal{A})| + |q| \left(|\mathsf{Ind}(\mathcal{A})| + 3\,|\mathcal{T}|\,2^{|\mathcal{T}|}\right)^{|q|}$. ☐

From Theorem 2, it follows that there exists a model minimizing the amount of c-matches with polynomial size w.r.t. data complexity. One can therefore non-deterministically guess this interpretation before verifying it is indeed a model and comparing its amount of c-matches with the input integer. The two latter steps can be done in (deterministic) polynomial time w.r.t. data complexity, yielding an upper bound in data complexity for CCQ answering, matching the corresponding results in the DL-Lite setting (Kostylev and Reutter 2015; Bienvenu, Manière, and Thomazo 2020).

Theorem 3. *CCQ answering in $\mathcal{ELHI}_\bot$ is in* coNP *w.r.t. data complexity.*

3.4 Matching Lower Bounds

We now provide 2EXP lower bounds for $\mathcal{EL}$ and DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$, which together with Theorem 1, establish the 2EXP-completeness of CCQ answering for $\mathcal{ELHI}$ and every sublogic that extends $\mathcal{EL}$ or DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$. The proofs are by reduction from the problem of answering Boolean union of conjunctive queries (BUCQs) over KBs with closed predicates, proven 2EXP-hard in (Ngo, Ortiz, and Šimkus 2016).

Theorem 4. *CCQ answering in $\mathcal{EL}$ is* 2EXP*-hard w.r.t. combined complexity.*

Proof sketch. Consider an $\mathcal{EL}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A}, \Sigma)$ with closed predicates and a BUCQ $q = \bigvee_{k=1}^{l} q_k$. Examining the 2EXP-hardness proof from (Ngo, Ortiz, and Šimkus 2016), we may assume that Σ consists only of concept names and each q_k is connected and has only variables as terms.

Pick a fresh individual aux not used in $\mathcal{A}$, and let $\mathcal{A}'$ be obtained from $\mathcal{A}$ by adding $\mathrm{A}(\mathsf{aux})$ for every concept name A in $\mathcal{K}$ and $\mathrm{P}(\mathsf{aux}, \mathsf{aux})$ for every role name P in $\mathcal{K}$. Consider the KB $\mathcal{K}' = (\mathcal{T}, \mathcal{A}')$ and the CCQ q' built as the conjunction of (i) all of the CQs q_k in q (with all variables treated as counting variables), (ii) the query $q_{\mathrm{A}} = \exists z_{\mathrm{A}}\ \mathrm{A}(z_{\mathrm{A}})$ for each $A \in \Sigma$, and (iii) the queries $q^+_{\mathrm{P}} = \exists z^+_{\mathrm{P}}\ \mathrm{P}(z^+_{\mathrm{P}}, \mathsf{aux})$ and $q^-_{\mathrm{P}} = \exists z^-_{\mathrm{P}}\ \mathrm{P}(\mathsf{aux}, z^-_{\mathrm{P}})$ for each role name P from $\mathcal{K}$. For each $\mathrm{A} \in \Sigma$, let n_{A} be the number of individuals a such that $\mathrm{A}(\mathsf{a}) \in \mathcal{A}$, and set $N := \prod_{\mathrm{A}\in\Sigma}(n_{\mathrm{A}} + 1)$. To complete the proof, one can show that $N + 1$ is a certain answer to q' over $\mathcal{K}'$ iff $\mathcal{K}$ entails q. ☐

Theorem 5. *CCQ answering in DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$ is* 2EXP*-hard w.r.t. combined complexity.*

Proof. As the 2EXP-hardness proof for DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ from (Ngo, Ortiz, and Šimkus 2016) does not involve negative inclusions, we can employ the same approach as for $\mathcal{EL}$ (the added aux assertions cannot lead to inconsistency). ☐

We thus close the open question of the combined complexity of CCQ answering in DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$. Note that our lower bound applies even to the subclass of CCQs whose every variable is a counting variable, as considered in (Kostylev and Reutter 2015; Calvanese et al. 2020a).

The preceding lower bound does not apply to DL-Lite$_{\mathsf{pos}}$, for which coNEXP membership has been shown (Kostylev and Reutter 2015; Bienvenu, Manière, and Thomazo 2020). We pinpoint the exact complexity by giving a matching lower bound, via a reduction from the exponential grid tiling problem. Here again the lower bound holds even when restricted to CCQs with only counting variables.

Theorem 6. *CCQ answering in DL-Lite$_{\mathsf{pos}}$ is* coNEXP*-hard w.r.t. combined complexity.*

4 Cardinality Queries

In this section, we focus on the restricted class of cardinality queries, which allow one to count the number of elements belonging to a given concept or role name.

To reduce the number of cases to be studied, we first notice that role cardinality queries are always harder than concept cardinality queries for the logics we consider.

Theorem 7. *Let $\mathcal{L}$ be a sublogic of $\mathcal{ELHI}_\bot$ that can express* $\mathrm{A} \sqsubseteq \exists \mathrm{P}.\top$ *(*$\mathrm{A} \in \mathsf{N_C}$, $\mathrm{P} \in \mathsf{N_R}$*). Then concept cardinality query answering over $\mathcal{L}$ KBs can be polynomially reduced to role cardinality query answering over $\mathcal{L}$ KBs.*

Proof. Take a concept cardinality query $q_{\mathrm{A}} = \exists z \mathrm{A}(z)$ and a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. We pick a fresh role name $\mathrm{P} \notin \mathsf{sig}(\mathcal{K})$, and consider the role cardinality query $q_{\mathrm{P}} = \exists z_1, z_2\ \mathrm{P}(z_1, z_2)$ and modified TBox $\mathcal{T}' := \mathcal{T} \cup \{\mathrm{A} \sqsubseteq \exists \mathrm{P}.\top\}$.

Any model $\mathcal{I}$ of $\mathcal{K}$ can be extended to a model $\mathcal{I}'$ of $\mathcal{K}' = (\mathcal{T}', \mathcal{A})$ by setting $\mathrm{P}^{\mathcal{I}'} := \{(e, e) \mid e \in \mathrm{A}^{\mathcal{I}}\}$. Indeed, this ensures satisfaction of the additional axiom $\mathrm{A} \sqsubseteq \exists \mathrm{P}.\top$. Moreover, as no new domain elements were introduced, axioms $\top \sqsubseteq \mathrm{B}$ from $\mathcal{T}$ remain satisfied, and all other axioms are not affected since $\mathrm{P} \notin \mathsf{sig}(\mathcal{T})$.

Notice that q_{A} has *exactly* as many matches in $\mathcal{I}$ as q_{P} has in $\mathcal{I}'$, hence an interval $[m, +\infty]$ is a certain answer to q_{A} over $\mathcal{K}$ iff it is a certain answer to q_{P} over $\mathcal{K}'$. ☐

4.1 Results for $\mathcal{EL}$ and its Extensions

The next two results, together with Theorem 7, establish that cardinality query answering is coNEXP-complete w.r.t. combined complexity in $\mathcal{ELHI}_\bot$ and $\mathcal{ELI}_\bot$.

Theorem 8. *Role cardinality query answering in $\mathcal{ELHI}_\bot$ is in* coNEXP *w.r.t. combined complexity.*

Proof. Theorem 2 proves that the minimal number of matches is reached with a model of exponential size. ☐

Theorem 9. *Concept cardinality query answering in $\mathcal{ELI}_\bot$ is* coNEXP*-hard w.r.t. combined complexity.*

Proof sketch. The proof proceeds by reduction from the complement of the Succinct-3COL problem, known to be NEXP-complete (Papadimitriou and Yannakakis 1986). ☐

The coNEXP lower bound relies on KBs that only admit exponentially large models. For logics admitting polynomial-sized models, the complexity slightly decreases.

Theorem 10. *Let $\mathcal{L}$ be a sublogic of $\mathcal{ELHI}_\bot$ for which every satisfiable KB admits a polynomial-sized model. Then role cardinality query answering over $\mathcal{L}$ KBs is in* EXP.

Proof sketch. The key observation is that, for logics with polysize models and single-atom queries, the optimal number of matches is bounded polynomially in the size of the KB. We can thus iterate over all polynomial-sized ABoxes that could represent the restriction of an optimal model to the ABox and elements in matches. We test whether such an ABox extends to a model without new matches by performing a satisfiability check, taking the query role as closed predicate. This gives a deterministic single-exponential time procedure, since satisfiability of $\mathcal{ELHI}_\bot$ KBs with closed predicates is in EXP (Ngo, Ortiz, and Šimkus 2016). □

Corollary 1. *Role cardinality query answering in $\mathcal{ELH}_\bot$ is in* EXP *w.r.t. combined complexity.*

Proof sketch. We observe that a variant of the compact canonical model used in the combined approach (Lutz, Toman, and Wolter 2009), provides a model also for $\mathcal{ELH}_\bot$ KBs with negative role inclusions. □

Corollary 2. *Role cardinality query answering in $\mathcal{ELHI}$ is in* EXP *w.r.t. combined complexity.*

Proof. Existence of polynomial-sized models is trivial due to the absence of negative inclusions. For example, extending $\mathcal{A}$ with every possible fact constructed from $\mathsf{Ind}(\mathcal{A})$ and $\mathsf{sig}(\mathcal{K})$ yields a model of $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. □

We conclude this subsection by providing matching lower bounds for concept cardinality queries in $\mathcal{EL}$.

Theorem 11. *Concept cardinality query answering in $\mathcal{EL}$ is* EXP*-hard w.r.t. combined complexity.*

Proof sketch. The proof is by reduction from the problem of deciding if an $\mathcal{EL}$ KB with closed predicates is satisfiable, proven EXP-hard in (Ngo, Ortiz, and Šimkus 2016). □

Theorem 12. *Concept cardinality query answering in $\mathcal{EL}$ is* coNP*-hard w.r.t. data complexity.*

Proof sketch. We reduce the complement of the graph 3-colorability problem to answering the cardinality query $\exists z\ \mathrm{B}(z)$ w.r.t. the TBox $\mathcal{T}$ containing $\mathrm{A} \sqsubseteq \exists \mathrm{R.B}$ and $\exists \mathrm{R.C_k} \sqcap \exists \mathrm{E}.(\exists \mathrm{R.C_k}) \sqsubseteq \mathrm{B}$ for $k \in \{1, 2, 3\}$. □

4.2 Results for DL-Lite

The data complexity picture being already known from the literature (Bienvenu, Manière, and Thomazo 2021), we focus on combined complexity.

We start by establishing tractabililty for DL-Lite$^{(\mathcal{H})}_{\mathsf{pos}}$ KBs.

Theorem 13. *Concept cardinality query answering in DL-Lite$_{\mathsf{pos}}$ is* NL*-hard w.r.t. combined complexity.*

Proof sketch. We proceed by reduction from the st-connectivity problem, known to be NL-complete (Immerman 1999). □

Theorem 14. *Concept cardinality query answering in DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$ is in* NL *w.r.t. combined complexity.*

Proof. Let $q_\mathrm{C} = \exists z\, \mathrm{C}(z)$ be a concept cardinality query. Starting from the canonical model $\mathcal{C}_\mathcal{K}$ of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, the minimal number of matches can easily be computed.

- If there exists an individual $\mathsf{a} \in \mathsf{Ind}(\mathcal{A})$ such that $\mathcal{K} \models \mathrm{C}(\mathsf{a})$, then we can collapse all anonymous elements onto one such individual (the choice doesn't matter), obtaining a model in which matches are exactly such individuals a, which is clearly minimal (recall we make the UNA). We can check whether $\mathcal{K} \models \mathrm{C}(\mathsf{a})$ in NL (Artale et al. 2009)
- Otherwise, if there exists an anonymous match in $\mathcal{C}_\mathcal{K}$, then we collapse all anonymous elements onto a chosen ABox individual, obtaining a model with a single match for q_C, which is clearly optimal. Existence of an anonymous match can be checked in NL (Artale et al. 2009).
- Otherwise, there are no matches in $\mathcal{C}_\mathcal{K}$, hence 0 is the minimal number of matches.

Notice that we do not need to actually compute the model corresponding to the optimal number of matches, and we only need to compare that number to the input integer. □

Theorem 15. *Role cardinality query answering in DL-Lite$_{\mathsf{pos}}$ is in* NL *w.r.t. combined complexity.*

Proof sketch. The proof relies on the same principle as Theorem 14, with a more sophisticated case analysis. □

Note that the preceding theorem concerns DL-Lite$_{\mathsf{pos}}$ rather than DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$, as role cardinality query answering in DL-Lite$^{\mathcal{H}}_{\mathsf{pos}}$ is coNP-hard even w.r.t. data complexity.

The introduction of disjointness axioms also leads to intractability, even for concept cardinality queries.

Theorem 16. *Concept cardinality query answering in DL-Lite$_{\mathsf{core}}$ is* coNP*-hard w.r.t. combined complexity.*

Proof. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph, and consider

$$\mathcal{T}_\mathcal{G} = \bigcup_{v \in \mathcal{V}} \{\mathrm{A} \sqsubseteq \exists \mathrm{V}, \exists \mathrm{V}^- \sqsubseteq \mathrm{C}\} \cup \bigcup_{\{v_1, v_2\} \in \mathcal{E}} \{\exists \mathrm{V}_1^- \sqsubseteq \neg \exists \mathrm{V}_2^-\}.$$

It is easily verified that $\mathcal{G} \in \mathsf{3COL}$ iff $[4, +\infty] \notin q^{\mathcal{K}_\mathcal{G}}$ for the KB $\mathcal{K}_\mathcal{G} := (\mathcal{T}_\mathcal{G}, \{\mathrm{A}(\mathsf{a})\})$ and query $q = \exists z\, \mathrm{C}(z)$. □

Theorem 17. *Role cardinality query answering in DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ is in* coNP *w.r.t. combined complexity.*

Proof sketch. One guesses a small counterexample to $[m, +\infty]$ being a certain answer, relying on the existence of small models, atomicity of the query, and Theorem 3 of (Ngo, Ortiz, and Šimkus 2016). □

5 Outlook

In this paper, we have extended the study of CCQ answering to Horn DLs outside the DL-Lite family, establishing a complete picture of the combined and data complexity of the problems of answering CCQs and cardinality queries in $\mathcal{ELHI}_\bot$ and its various sublogics. Interestingly, the new techniques we devised also allowed us to close some open questions concerning the combined complexity of CCQ answering in DL-Lite. Going forward, the main challenge is to develop practical algorithms. A first direction is to look for restrictions on the query or ontology that ensure polynomial data complexity for logics of the $\mathcal{EL}$ family. Second, it would be desirable, for $\mathcal{EL}$ but also for DL-Lite, to develop more refined coNP procedures that are amenable to implementation using SAT solvers. We believe that our improved understanding of the structure of optimal models will prove helpful for both of these research directions.

Acknowledgments

Partially supported by ANR project CQFD (ANR-18-CE23-0003).

References

Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The DL-Lite family and relations. *Journal of Artificial Intelligence Research (JAIR)* 36:1–69.

Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ envelope. In *Proceedings of the 19th international joint conference on Artificial intelligence (IJCAI)*, 364–369.

Bienvenu, M., and Ortiz, M. 2015. Ontology-mediated query answering with data-tractable description logics. In *Tutorial Lectures of the 11th Reasoning Web International Summer School (RW)*, 218–307.

Bienvenu, M.; Calvanese, D.; Ortiz, M.; and Simkus, M. 2014. Nested regular path queries in description logics. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 218–227.

Bienvenu, M.; Manière, Q.; and Thomazo, M. 2020. Answering counting queries over DL-Lite ontologies. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 1608–1614.

Bienvenu, M.; Manière, Q.; and Thomazo, M. 2021. Cardinality queries over DL-Lite ontologies. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, 1801–1807.

Calvanese, D.; Giacomo, G. D.; 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)* 39(3):385–429.

Calvanese, D.; Kharlamov, E.; Nutt, W.; and Thorne, C. 2008. Aggregate queries over ontologies. In *Proceedings of the 2nd International Workshop on Ontologies and Information Systems for the Semantic Web (ONISW)*, 97–104.

Calvanese, D.; Corman, J.; Lanti, D.; and Razniewski, S. 2020a. Counting query answers over a DL-Lite knowledge base. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 1658–1666.

Calvanese, D.; Corman, J.; Lanti, D.; and Razniewski, S. 2020b. Rewriting count queries over DL-Lite TBoxes with number restrictions. In *Proceedings of the 33rd International Workshop on Description Logics (DL)*.

Eiter, T.; Lutz, C.; Ortiz, M.; and Simkus, M. 2009. Query answering in description logics: the knots approach. In *Proceedings of the 16th International Workshop on Logic, Language, Information and Computation (WoLLIC)*, 26–36.

Feier, C.; Lutz, C.; and Przybylko, M. 2021. Answer counting under guarded TGDs. In *Proceedings of the 24th International Conference on Database Theory (ICDT)*.

Immerman, N. 1999. Descriptive complexity.

Kostylev, E. V., and Reutter, J. L. 2015. Complexity of answering counting aggregate queries over DL-Lite. *Journal of Web Semantics (JWS)* 94–111.

Lutz, C.; Toman, D.; and Wolter, F. 2009. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, 2070–2075.

Ngo, N.; Ortiz, M.; and Šimkus, M. 2016. Closed predicates in description logics: results on combined complexity. In *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 237–246.

Nikolaou, C.; Kostylev, E. V.; Konstantinidis, G.; Kaminski, M.; Grau, B. C.; and Horrocks, I. 2019. Foundations of ontology-based data access under bag semantics. *Journal of Artificial Intelligence (AIJ)* 91 – 132.

Papadimitriou, C. H., and Yannakakis, M. 1986. A note on succinct representations of graphs. *Information and Control* 71(3):181–185.

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

Rudolph, S.; Krötzsch, M.; and Hitzler, P. 2012. Type-elimination-based reasoning for the description logic $\mathcal{SHIQ}b_s$ using decision diagrams and disjunctive datalog. *Logical Methods in Computer Science (LMCS)* 8(1).

Thomazo, M.; Baget, J.; Mugnier, M.; and Rudolph, S. 2012. A generic querying algorithm for greedy sets of existential rules. In *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

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 the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 5511–5519.