Counting Queries over $\mathcal{ELHI}_\perp$ 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 $\mathcal{ELHI}_\perp$ and its various sublogics. Notably, we show that CCQ answering is 2EXP-complete in combined complexity for $\mathcal{ELHI}_\perp$ 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 $\mathcal{ELHI}_\perp$. 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 R_i are *roles* drawn from $\mathsf{N}^{\pm}_{\mathsf{R}} = \{\mathrm{P}, \mathrm{P}^- \mid \mathrm{P} \in \mathsf{N_R}\}$ and the 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:

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

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 $\mathcal{ELHI}_\bot$ 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 $\mathcal{ELHI}_\bot$ 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 $\mathcal{ELHI}_\bot$ 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 $\mathcal{ELHI}_\bot$ TBoxes in a convenient *normal form*, where every concept inclusion has one of the following restricted shapes:

$$\mathrm{A} \sqsubseteq \bot \qquad \top \sqsubseteq \mathrm{A} \qquad \mathrm{A}_1 \sqcap \mathrm{A}_2 \sqsubseteq \mathrm{A}$$
$$\mathrm{A}_1 \sqsubseteq \exists\mathrm{R}.\mathrm{A}_2 \qquad \exists\mathrm{R}.\mathrm{A}_1 \sqsubseteq \mathrm{A}_2$$

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 $\mathcal{ELHI}_\bot$ 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.

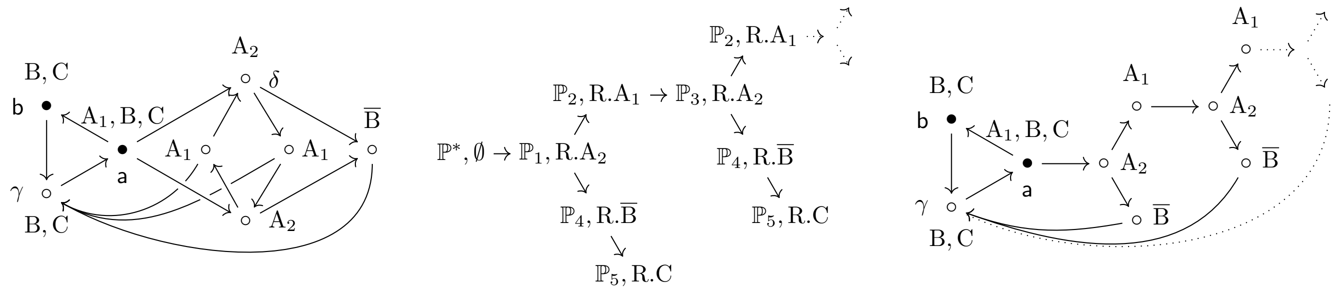
3.1 Decision Procedure

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 $\mathcal{ELHI}_\bot$ 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 ⊚ and ⊛ 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* $\mathsf{next}(\mathrm{R}_1.\mathrm{B}_1) \neq \mathsf{next}(\mathrm{R}_2.\mathrm{B}_2)$.

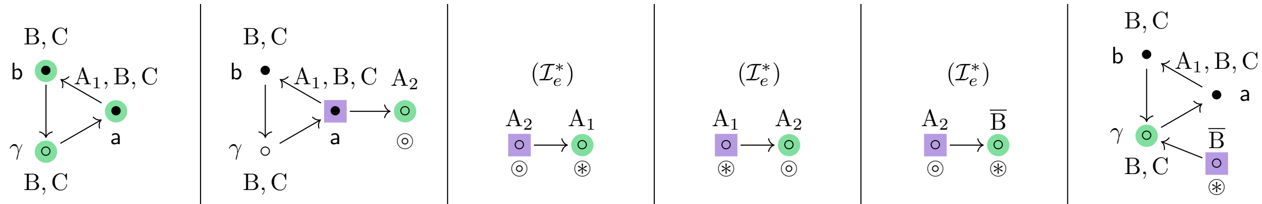
We now formally define the central notion of pattern, relative to $\mathcal{I}^*$ and a candidate specification $\mathfrak{M}^*$ over $\mathcal{I}^*$.

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 $\Delta^{\mathcal{I}^*} \cup \{⊚, ⊛\}$;*
- *$\mathfrak{I}^{\mathbb{P}}$ 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 $\mathbb{P}_1^e$. (c) Interp. of $\mathbb{P}_2^e$. (d) Interp. of $\mathbb{P}_3^e$. (e) Interp. of $\mathbb{P}_4^e$. (f) Interp. of $\mathbb{P}_5^e$.

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* $\mathbb{P}_1^e, \dots, \mathbb{P}_5^e$ *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* $\mathsf{next}_{\mathbb{P}_4^e}$*, 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* $\mathbb{P}_i^e$*'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* $\mathbb{P}_4^e$ *to* $\mathbb{P}_5^e$ *(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:

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

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:*

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

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 $ah_1 \dots h_n$, 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.

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

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 $|\Delta^*| \leq |\mathsf{Ind}(\mathcal{A})| + |q| \left(|\mathsf{Ind}(\mathcal{A})| + 3\,|\mathcal{T}|\,2^{|\mathcal{T}|}\right)^{|q|}$. $\square$

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.

A Proofs for Section 3 (General Case of CCQs)

A.1 Proofs for Section 3.1 (Decision Procedure)

The notion of *canonical model* of a KB will be used several times in the proofs, so we start by recalling its definition.

Canonical model. It is well known that every satisfiable $\mathcal{ELHI}_\bot$ KB admits a canonical (or universal) model that embeds homomorphically into each of its models. We recall how such a model $\mathcal{C}_\mathcal{K}$ can be constructed (see e.g. (Bienvenu and Ortiz 2015)). The domain $\Delta^{\mathcal{C}_\mathcal{K}}$ consists of all sequences $\mathsf{a}\mathrm{R}_1.\mathrm{M}_1 \ldots \mathrm{R}_\mathsf{n}.\mathrm{M}_\mathsf{n}$ ($\mathsf{n} \geq 0$) such that $\mathsf{a} \in \mathsf{Ind}(\mathcal{A})$, each R_i belongs to $\mathsf{N}_\mathsf{R}^\pm$, each M_i is a conjunction of concepts from $\mathsf{N}_\mathsf{C} \cup \{\top\}$ (treated as a set when convenient), and the following conditions hold:

- If $\mathsf{n} \geq 1$, then $\mathcal{T} \models \mathrm{M}_0 \sqsubseteq \exists \mathrm{R}_1.\mathrm{M}_1$ where $\mathrm{M}_0 = \{\mathrm{A} \in \mathsf{N}_\mathsf{C} \cup \{\top\} \mid \mathcal{K} \models \mathrm{A}(\mathsf{a})\}$ and M_1 is maximal, as a set of concept names, for this property.
- For every $1 \leq \mathsf{i} < \mathsf{n}$, $\mathcal{T} \models \mathrm{M}_\mathrm{i} \sqsubseteq \exists \mathrm{R}_{\mathrm{i}+1}.\mathrm{M}_{\mathrm{i}+1}$ and $\mathrm{M}_{\mathrm{i}+1}$ is maximal, as a set of concept names, for this property.

Individual names are interpreted as themselves ($\mathsf{a}^{\mathcal{C}_\mathcal{K}} = \mathsf{a}$), and concept and role names are interpreted as follows:

$$\begin{aligned} \mathrm{A}^{\mathcal{C}_\mathcal{K}} &= \{\mathsf{a} \mid \mathcal{K} \models \mathrm{A}(\mathsf{a})\} \cup \{e \cdot \mathrm{R.M} \mid \mathrm{A} \in \mathrm{M}\} \\ \mathrm{P}^{\mathcal{C}_\mathcal{K}} &= \{(\mathsf{a}, \mathsf{b}) \mid \mathcal{K} \models \mathrm{P}(\mathsf{a}, \mathsf{b})\} \cup \{(e, e \cdot \mathrm{P}_0.\mathrm{M}) \mid \mathcal{T} \models \mathrm{P}_0 \sqsubseteq \mathrm{P}\} \cup \{(e \cdot \mathrm{P}_0.\mathrm{M}, e) \mid \mathcal{T} \models \mathrm{P}_0 \sqsubseteq \mathrm{P}^-\} \end{aligned}$$

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.*

Proof. Consider the canonical model $\mathcal{C}_\mathcal{K}$ of $\mathcal{K}$. For each element of $\Delta^{\mathcal{C}_\mathcal{K}}$, we define its size: the size $|\mathsf{a}|$ of an individual a is 1, the size $|w \cdot \mathrm{R.M}|$ of a non-individual element $w \cdot \mathrm{R.M}$ is $|w| + 1$. We now equip $\Delta^{\mathcal{C}_\mathcal{K}}$ with the following equivalence relation $\sim$: each individual is only equivalent to itself, while two non-individual elements $w_1 \cdot \mathrm{R}_1.\mathrm{M}_1$ and $w_2 \cdot \mathrm{R}_2.\mathrm{M}_2$ are equivalent iff $\mathrm{R}_1.\mathrm{M}_1 = \mathrm{R}_2.\mathrm{M}_2$ and $|w_1| = |w_2| \mod 3$. Let $\widetilde{u}$ denote the equivalence class of the element u w.r.t $\sim$ and $\nu : d \mapsto \widetilde{d}$ the canonical projection.

We claim that the interpretation $\mathcal{M} := \mathcal{C}_\mathcal{K} \,/\, \sim$ with domain $\Delta^{\mathcal{C}_\mathcal{K}} \,/\, \sim$ and interpretation of atomic concepts and roles given by $\cdot^\mathcal{M} := \nu \circ \cdot^{\mathcal{C}_\mathcal{K}}$ is a model. Notice it has the desired amount of elements as each equivalence class is: either a single individual, or fully characterized by an integer modulo 3, a role from $\mathsf{sig}(\mathcal{T})$ and a set of concepts from $\mathsf{sig}(\mathcal{T})$.

We consider each of the possible kinds of assertions and axioms occurring in $\mathcal{K}$:

$\mathrm{A}(\mathsf{a})$. Since $\mathcal{C}_\mathcal{K}$ is a model, we have $\mathsf{a} \in \mathrm{A}^{\mathcal{C}_\mathcal{K}}$. Therefore, the definition of $\mathrm{A}^\mathcal{M}$ gives $\widetilde{\mathsf{a}} = \mathsf{a} \in \mathrm{A}^\mathcal{M}$.

$\mathrm{P}(\mathsf{a}, \mathsf{b})$. Since $\mathcal{C}_\mathcal{K}$ is a model, we have $(\mathsf{a}, \mathsf{b}) \in \mathrm{P}^{\mathcal{C}_\mathcal{K}}$. Therefore, the definition of $\mathrm{P}^\mathcal{M}$ gives $(\widetilde{\mathsf{a}}, \widetilde{\mathsf{b}}) = (\mathsf{a}, \mathsf{b}) \in \mathrm{P}^\mathcal{M}$.

$\mathrm{A} \sqsubseteq \bot$. Since $\mathcal{C}_\mathcal{K}$ is a model, we have $\mathrm{A}^{\mathcal{C}_\mathcal{K}} = \emptyset$. Therefore, the definition of $\mathrm{A}^\mathcal{M}$ gives $\mathrm{A}^\mathcal{M} = \emptyset$.

$\top \sqsubseteq \mathrm{A}$. Let $u \in \top^\mathcal{M} = \Delta^\mathcal{M}$. By definition of $\Delta^\mathcal{M}$, there exists $u_0 \in \Delta^{\mathcal{C}_\mathcal{K}}$ such that $\widetilde{u_0} = u$. Since $u_0 \in \top^{\mathcal{C}_\mathcal{K}}$ and $\mathcal{C}_\mathcal{K}$ is a model, we have $u_0 \in \mathrm{A}^{\mathcal{C}_\mathcal{K}}$. Therefore the definition of $\mathrm{A}^\mathcal{M}$ gives $u = \widetilde{u_0} \in \mathrm{A}^\mathcal{M}$.

$\mathrm{A}_1 \sqcap \mathrm{A}_2 \sqsubseteq \mathrm{A}$. Let $u \in (\mathrm{A}_1 \sqcap \mathrm{A}_2)^\mathcal{M}$. By definition of $\mathrm{A}_1^\mathcal{M}$ and $\mathrm{A}_2^\mathcal{M}$, there exist $u_1 \in \mathrm{A}_1^{\mathcal{C}_\mathcal{K}}$ and $u_2 \in \mathrm{A}_2^{\mathcal{C}_\mathcal{K}}$ with $\widetilde{u_1} = \widetilde{u_2} = u$. Since $\widetilde{u_1} = \widetilde{u_2}$, elements u_1 and u_2 satisfy the same concepts: either because they both are the same individual, or because they end with the same R.M, which fully determines the concepts they satisfy. In particular $u_1 \in (\mathrm{A}_1 \sqcap \mathrm{A}_2)^{\mathcal{C}_\mathcal{K}}$. Since $\mathcal{C}_\mathcal{K}$ is a model, we have $u_1 \in \mathrm{A}^{\mathcal{C}_\mathcal{K}}$, yielding by definition of $\mathrm{A}^\mathcal{M}$ that $u = \widetilde{u_1} \in \mathrm{A}^\mathcal{M}$.

$\mathrm{A}_1 \sqsubseteq \exists \mathrm{R}.\mathrm{A}_2$. Let $u \in \mathrm{A}_1^\mathcal{M}$. By definition of ${\mathrm{A}_1}^\mathcal{M}$ there exists $u_0 \in {\mathrm{A}_1}^{\mathcal{C}_\mathcal{K}}$ with $\widetilde{u_0} = u$. Since $\mathcal{C}_\mathcal{K}$ is a model, it ensures there exists $v_0 \in {\mathrm{A}_2}^{\mathcal{C}_\mathcal{K}}$ with $(u_0, v_0) \in \mathrm{R}^{\mathcal{C}_\mathcal{K}}$. By definition of ${\mathrm{A}_2}^\mathcal{M}$ and $\mathrm{R}^\mathcal{M}$, the element $v := \widetilde{v_0}$ satisfies both $v \in {\mathrm{A}_2}^\mathcal{M}$ and $(u, v) \in \mathrm{R}^\mathcal{M}$, that is $u \in (\exists \mathrm{R}.\mathrm{A}_2)^\mathcal{M}$.

$\exists \mathrm{R}.\mathrm{A}_1 \sqsubseteq \mathrm{A}_2$. Let $u \in (\exists \mathrm{R}.\mathrm{A}_1)^\mathcal{M}$, that is there exists $v \in \mathrm{A}_1^\mathcal{M}$ with $(u, v) \in \mathrm{R}^\mathcal{M}$. By definition of $\mathrm{A}_1^\mathcal{M}$ and $\mathrm{R}^\mathcal{M}$, there exist $(u_0, v_0) \in \mathrm{R}^{\mathcal{C}_\mathcal{K}}$ and $v_1 \in \mathrm{A}_1^{\mathcal{C}_\mathcal{K}}$ such that $\widetilde{u_0} = u$ and $\widetilde{v_0} = \widetilde{v_1} = v$. Again, since $\widetilde{v_0} = \widetilde{v_1}$ both v_0 and v_1 satisfy the same concepts, that is in particular $u_0 \in (\exists \mathrm{R}.\mathrm{A}_1)^{\mathcal{C}_\mathcal{K}}$. Since $\mathcal{C}_\mathcal{K}$ is a model, it ensures $u_0 \in \mathrm{A}_2^{\mathcal{C}_\mathcal{K}}$, yielding by definition of $\mathrm{A}_2^\mathcal{M}$ that $u = \widetilde{u_0} \in \mathrm{A}_2^\mathcal{M}$.

$\mathrm{P} \sqsubseteq \mathrm{R}$. Let $(u, v) \in \mathrm{P}^\mathcal{M}$. By definition of $\mathrm{P}^\mathcal{M}$, there exists $(u_0, v_0) \in \mathrm{P}^{\mathcal{C}_\mathcal{K}}$ such that $\widetilde{u_0} = u$ and $\widetilde{v_0} = v$. Since $\mathcal{C}_\mathcal{K}$ is a model, it ensures $(u_0, v_0) \in \mathrm{R}^{\mathcal{C}_\mathcal{K}}$, hence $(\widetilde{u_0}, \widetilde{v_0}) = (u, v) \in \mathrm{R}^\mathcal{M}$ by definition of $\mathrm{R}^\mathcal{M}$.

$\mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$. By contradiction, assume one can find $(u, v) \in (\mathrm{R}_1 \sqcap \mathrm{R}_2)^\mathcal{M}$. By definition of $\mathrm{R}_1^\mathcal{M}$ and $\mathrm{R}_2^\mathcal{M}$, there exists $(u_1, v_1) \in \mathrm{R}_1^{\mathcal{C}_\mathcal{K}}$ and $(u_2, v_2) \in \mathrm{R}_2^{\mathcal{C}_\mathcal{K}}$ such that $\widetilde{u_1} = \widetilde{u_2} = u$ and $\widetilde{v_1} = \widetilde{v_2} = v$.
If either $\mathcal{K} \models \mathrm{R}_1(u_1, v_1)$ or $\mathcal{K} \models \mathrm{R}_2(u_2, v_2)$, then, each individual being alone in its equivalence class, we have $u_1 = u_2$ and $v_1 = v_2$. In particular it gives $(u_1, v_1) \in (\mathrm{R}_1 \sqcap \mathrm{R}_2)^{\mathcal{C}_\mathcal{K}}$, contradicting $\mathcal{C}_\mathcal{K}$ being a model.
Otherwise we distinguish the four possible cases:

- $v_1 = u_1 \cdot \mathrm{P}_1.\mathrm{M}_1$ and $\mathcal{T} \models \mathrm{P}_1 \sqsubseteq \mathrm{R}_1$.
  - $v_2 = u_2 \cdot \mathrm{P}_2.\mathrm{M}_2$ and $\mathcal{T} \models \mathrm{P}_2 \sqsubseteq \mathrm{R}_2$. Since $\widetilde{v_1} = \widetilde{v_2}$ we have $\mathrm{P}_1.\mathrm{M}_1 = \mathrm{P}_2.\mathrm{M}_2$. In particular $(u_1, v_1) \in {\mathrm{R}_2}^{\mathcal{C}_\mathcal{K}}$, which contradicts $\mathcal{C}_\mathcal{K}$ being a model.

* $u_2 = v_2 \cdot \mathrm{P}_2.\mathrm{M}_2$ and $\mathcal{T} \models \mathrm{P}_2 \sqsubseteq \mathrm{R}_2^-$. In particular $|v_1| = |u_1| + 1 \mod 3$ and $|u_2| = |v_2| + 1 \mod 3$. Recall that $\widetilde{u_1} = \widetilde{u_2}$ and $\widetilde{v_1} = \widetilde{v_2}$, hence $|u_1| = |u_2| \mod 3$ and $|v_1| = |v_2| \mod 3$. It yields $0 = 2 \mod 3$, contradiction.

- $u_1 = v_1 \cdot \mathrm{P}_1.\mathrm{M}_1$ and $\mathcal{T} \models \mathrm{P}_1 \sqsubseteq \mathrm{R}_1^-$.
  * $v_2 = u_2 \cdot \mathrm{P}_2.\mathrm{M}_2$ and $\mathcal{T} \models \mathrm{P}_2 \sqsubseteq \mathrm{R}_2$. Symmetric to the previous case, leading to a contradiction.
  * $u_2 = v_2 \cdot \mathrm{P}_2.\mathrm{M}_2$ and $\mathcal{T} \models \mathrm{P}_2 \sqsubseteq \mathrm{R}_2^-$. Since $\widetilde{u_1} = \widetilde{u_2}$ we have $\mathrm{P}_1.\mathrm{M}_1 = \mathrm{P}_2.\mathrm{M}_2$. In particular $(u_1, v_1) \in {\mathrm{R}_2}^{\mathcal{C}_\mathcal{K}}$, which contradicts $\mathcal{C}_\mathcal{K}$ being a model.

□

A.2 Proofs for Section 3.2 (Proofs of Lemmas 2 and 3)

From Accepting Patterns to Models

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 $\lambda_{w\cdot(\mathbb{P},h)}(e) \in \mathrm{A}^{\mathcal{I}^\diamond}$, then $e \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}}}$.*

Proof. Let $w_1 \cdot (\mathbb{P}_1, h_1) \in \mathcal{P}$ be a relevant pattern, e_1 an element from $\Delta^{\mathfrak{I}^{\mathbb{P}_1}}$ and A a concept name. Assume $\lambda_{w_1\cdot\mathbb{P}_1}(e_1) \in \mathrm{A}^{\mathcal{I}^\diamond}$. By definition of $\mathrm{A}^{\mathcal{I}^\diamond}$ there exists a relevant pattern $w_2 \cdot (\mathbb{P}_2, h_2)$, and an element $e_2 \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_2}}$ such that $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1) = \lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(e_2)$. We further refer to this latter equality as $(*)$. We distinguish 5 cases.

1. $e_1 \in \Delta^{\mathcal{I}^*}$ or $e_2 \in \Delta^{\mathcal{I}^*}$.
   $(*)$ yields $e_1 = e_2$. Interpretation $\mathfrak{I}^{\mathbb{P}_2}$ preserves $\mathcal{I}^*$, hence $e_2 \in \mathrm{A}^{\mathcal{I}^*}$. Interpretation $\mathfrak{I}^{\mathbb{P}_1}$ preserves $\mathcal{I}^*$, hence $e_1 \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_1}}$.
   In the remaining cases, we assume $e_1, e_2 \notin \Delta^{\mathcal{I}^*}$, which ensures $\mathbb{P}_1 \neq \mathbb{P}^*$ and $\mathbb{P}_2 \neq \mathbb{P}^*$. In particular, $\mathfrak{fr}^{\mathbb{P}_1}$, $\mathfrak{gen}^{\mathbb{P}_1}$, $\mathfrak{fr}^{\mathbb{P}_2}$ and $\mathfrak{gen}^{\mathbb{P}_2}$ are singletons.
2. $e_1 \in \mathfrak{gen}^{\mathbb{P}_1} \setminus \Delta^{\mathcal{I}^*}$ and $e_2 \in \mathfrak{gen}^{\mathbb{P}_2} \setminus \Delta^{\mathcal{I}^*}$.
   $(*)$ yields $\mathbb{P}_1 = \mathbb{P}_2$. Recall $\mathfrak{gen}^{\mathbb{P}_1}$ is a singleton hence $e_1 = e_2$, which concludes.
3. $e_1 \in \mathfrak{fr}^{\mathbb{P}_1} \setminus \Delta^{\mathcal{I}^*}$ and $e_2 \in \mathfrak{gen}^{\mathbb{P}_2} \setminus \Delta^{\mathcal{I}^*}$.
   $(*)$ yields $w_1 = w_2 \cdot (\mathbb{P}_2, h_2)$. In particular $w_2 \cdot (\mathbb{P}_2, h_2) \cdot (\mathbb{P}_1, h_1) \in \mathcal{P}$, hence $\mathbb{P}_1 = \mathsf{ch}^{h_1}_{\mathbb{P}_2,e_2}$. By definition of a link we obtain $e_1 = e_2$ (Condition 1) and e_1 satisfies the same concepts in both interpretations (Condition 2) hence $e_1 \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_1}}$.
4. $e_1 \in \mathfrak{gen}^{\mathbb{P}_1} \setminus \Delta^{\mathcal{I}^*}$ and $e_2 \in \mathfrak{fr}^{\mathbb{P}_2} \setminus \Delta^{\mathcal{I}^*}$.
   Same arguments as for Case 3 but this time with $\mathbb{P}_2 = \mathsf{ch}^{h_2}_{\mathbb{P}_1,e_1}$.
5. $e_1 \in \mathfrak{fr}^{\mathbb{P}_1} \setminus \Delta^{\mathcal{I}^*}$ and $e_2 \in \mathfrak{fr}^{\mathbb{P}_2} \setminus \Delta^{\mathcal{I}^*}$.
   $(*)$ yields the existence of $w \cdot (\mathbb{Q}, h)$ such that $w_1 = w_2 = w \cdot (\mathbb{Q}, h)$. In particular $w \cdot (\mathbb{Q}, h) \cdot (\mathbb{P}_2, h_2) \in \mathcal{P}$, hence $\mathbb{P}_2 = \mathsf{ch}^{h_2}_{\mathbb{Q},e_2}$. By definition of a link e_2 satisfies the same concepts in both interpretations (Condition 2) hence $e_2 \in \mathrm{A}^{\mathfrak{I}^{\mathbb{Q}}}$. Similarly, $w \cdot (\mathbb{Q}, h) \cdot (\mathbb{P}_1, h_1) \in \mathcal{P}$, hence $\mathbb{P}_1 = \mathsf{ch}^{h_1}_{\mathbb{Q},e_2}$. By definition of a link we obtain $e_1 = e_2$ (Condition 1) and e_1 satisfies the same concepts in both interpretations (Condition 2) hence $e_1 \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_1}}$.

□

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 $(\lambda_{w\cdot(\mathbb{P},h)}(d), \lambda_{w\cdot(\mathbb{P},h)}(e)) \in \mathrm{P}^{\mathcal{I}^\diamond}$, then $\mathfrak{I}^{\mathbb{P}}$ remains $\mathcal{T}$-satisfiable if we add (d, e) to $\mathrm{P}^{\mathfrak{I}^{\mathbb{P}}}$.*

Proof. Let $w_1 \cdot (\mathbb{P}_1, h_1) \in \mathcal{P}$ and $d_1, e_1 \in \Delta^{\mathfrak{I}^{\mathbb{P}_1}}$ two elements. Let $\mathrm{P} \in \mathsf{N_R}$ be a role name. Assume $(\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(d_1), \lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1)) \in \mathrm{P}^{\mathcal{I}^\diamond}$. By definition of $\mathrm{P}^{\mathcal{I}^\diamond}$, there exist $w_2 \cdot (\mathbb{P}_2, h_2) \in \mathcal{P}$ and $(d_2, e_2) \in \mathrm{P}^{\mathfrak{I}^{\mathbb{P}_2}}$ with $\lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(d_2) = \lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(d_1)$ and $\lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(e_2) = \lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1)$. We further refer to these two equalities as $(*_d)$ and $(*_e)$. We distinguish 5 main cases.

1. ($d_1 \in \Delta^{\mathcal{I}^*}$ or $d_2 \in \Delta^{\mathcal{I}^*}$) and ($e_1 \in \Delta^{\mathcal{I}^*}$ or $e_2 \in \Delta^{\mathcal{I}^*}$).
   $(*_d)$ yields $d_1 = d_2$ and $(*_e)$ yields $e_1 = e_2$. Interpretation $\mathfrak{I}^{\mathbb{P}_2}$ preserves $\mathcal{I}^*$, hence $(d_2, e_2) \in \mathrm{P}^{\mathcal{I}^*}$. Interpretation $\mathfrak{I}^{\mathbb{P}_1}$ preserves $\mathcal{I}^*$, hence $(d_1, e_1) \in \mathrm{P}^{\mathfrak{I}^{\mathbb{P}_1}}$. It then suffices to recall that $\mathfrak{I}^{\mathbb{P}_1}$ is $\mathcal{T}$-satisfiable.
   In the remaining cases, we assume that $e_1, e_2 \notin \Delta^{\mathcal{I}^*}$ or $d_1, d_2 \notin \Delta^{\mathcal{I}^*}$, which ensures $\mathbb{P}_1 \neq \mathbb{P}^*$ and $\mathbb{P}_2 \neq \mathbb{P}^*$. In particular, $\mathfrak{fr}^{\mathbb{P}_1}$, $\mathfrak{gen}^{\mathbb{P}_1}$, $\mathfrak{fr}^{\mathbb{P}_2}$ and $\mathfrak{gen}^{\mathbb{P}_2}$ are singletons. Furthermore, the conditions on roles for a non-initial pattern (Condition 4) ensures $d_2 \neq e_2$ (recall we assume $(d_2, e_2) \in \mathrm{P}^{\mathfrak{I}^{\mathbb{P}_2}}$).
2. ($d_1 \in \Delta^{\mathcal{I}^*}$ or $d_2 \in \Delta^{\mathcal{I}^*}$) and ($e_1, e_2 \notin \Delta^{\mathcal{I}^*}$).
   $(*_d)$ yields $d_1 = d_2$, we distinguish 4 remaining subcases.

(a) $e_1 \in \mathfrak{gen}^{\mathbb{P}_1}$ and $e_2 \in \mathfrak{gen}^{\mathbb{P}_2}$.
We have $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1) = w_1 \cdot (\mathbb{P}_1, h_1)$ and $\lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(e_2) = w_2 \cdot (\mathbb{P}_2, h_2)$. Hence $(*_e)$ yields in particular $\mathbb{P}_1 = \mathbb{P}_2$. Recall that $\mathfrak{gen}^{\mathbb{P}_1}$ is a singleton, so $e_1 = e_2$. Therefore $\mathfrak{I}^{\mathbb{P}_1}$ already contains the fact $\mathrm{P}(d_1, e_1)$. Recalling that $\mathfrak{I}^{\mathbb{P}_1}$ is satisfiable concludes this case.

(b) $e_1 \in \mathfrak{fr}^{\mathbb{P}_1}$ and $e_2 \in \mathfrak{gen}^{\mathbb{P}_2}$.
We have $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1) = w_1$ and $\lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(e_2) = w_2 \cdot (\mathbb{P}_2, h_2)$. Hence $(*_e)$ yields $w_1 = w_2 \cdot (\mathbb{P}_2, h_2)$. In particular $w_2 \cdot (\mathbb{P}_2, h_2) \cdot (\mathbb{P}_1, h_1) \in \mathcal{P}$, therefore $\mathbb{P}_1 = \mathsf{ch}^{h_1}_{\mathbb{P}_2,e_2}$ and $e_1 = e_2$. Notice e_1, that is also e_2, satisfies the same concepts in $\mathfrak{I}^{\mathbb{P}_1}$ and in $\mathfrak{I}^{\mathbb{P}_2}$ (Lemma 4 applies to e_1 seen in $\mathfrak{I}^{\mathbb{P}_1}$ and e_1 seen in $\mathfrak{I}^{\mathbb{P}_2}$), and same for d_1, that is also d_2. Therefore, the $\mathcal{T}$-satisfiability of $\mathfrak{I}^{\mathbb{P}_2}$ ensures that adding fact $\mathrm{P}(d_1, e_1)$ to $\mathfrak{I}^{\mathbb{P}_1}$ does not violate any negative *concept* inclusion from $\mathcal{T}$. We make a case analysis to show the same is true for negative role inclusions:

- First suppose $\mathfrak{gen}^{\mathbb{P}_1} = \{d_1\}$. Since $(d_1, e_1) \in \mathrm{P}^{\mathfrak{I}^{\mathbb{P}_2}}$ and $e_1 \in \mathfrak{gen}^{\mathbb{P}_2}$, then we must have $\mathfrak{fr}^{\mathbb{P}_2} = \{d_1\}$ (Condition 4). We can hence apply Condition 5 from the definition of the link given by $\mathbb{P}_1 = \mathsf{ch}^{h_1}_{\mathbb{P}_2,e_2}$, ensuring that $\mathfrak{I}^{\mathbb{P}_1} \cup \mathfrak{I}^{\mathbb{P}_2}$, which contains $\mathfrak{I}^{\mathbb{P}_1}$ and fact $\mathrm{P}(d_1, e_1)$, is $\mathcal{T}$-satisfiable.
- If $\mathfrak{gen}^{\mathbb{P}_1} \neq \{d_1\}$, then there are no roles between d_1 and e_1 in $\mathfrak{I}^{\mathbb{P}_1}$ (Condition 4), hence no negative role inclusion is violated by adding fact $\mathrm{P}(d_1, e_1)$ in $\mathfrak{I}^{\mathbb{P}_1}$.

(c) $e_1 \in \mathfrak{gen}^{\mathbb{P}_1}$ and $e_2 \in \mathfrak{fr}^{\mathbb{P}_2}$.
Same arguments as for Case 2.b but with $\mathbb{P}_2 = \mathsf{ch}^{h_2}_{\mathbb{P}_1,e_1}$.

(d) $e_1 \in \mathfrak{fr}^{\mathbb{P}_1}$ and $e_2 \in \mathfrak{fr}^{\mathbb{P}_2}$.
We have $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1) = w_1$ and $\lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(e_2) = w_2$. Hence $(*_e)$ yields the existence of $w \cdot (\mathbb{Q}, h)$ such that $w_1 = w_2 = w \cdot (\mathbb{Q}, h)$. In particular $w \cdot (\mathbb{Q}, h) \cdot (\mathbb{P}_1, h_1) \in \mathcal{P}$, hence $\mathbb{P}_1 = \mathsf{ch}^{h_1}_{\mathbb{Q},e_1}$. Similarly we obtain $\mathbb{P}_2 = \mathsf{ch}^{h_2}_{\mathbb{Q},e_2}$. As $e_1, e_2 \notin \Delta^*$, the pattern $\mathbb{Q}$ must be different from $\mathbb{P}^*$, hence its generated term is unique, which gives $e_1 = e_2$. Notice e_1, that is also e_2, satisfies the same concepts in $\mathfrak{I}^{\mathbb{P}_1}$ and in $\mathfrak{I}^{\mathbb{P}_2}$ (Lemma 4 applies to e_1 seen in $\mathfrak{I}^{\mathbb{P}_1}$ and e_1 seen in $\mathfrak{I}^{\mathbb{P}_2}$), and same for d_1, that is also d_2. Therefore, the $\mathcal{T}$-satisfiability of $\mathfrak{I}^{\mathbb{P}_2}$ ensures adding fact $\mathrm{P}(d_1, e_1)$ in $\mathfrak{I}^{\mathbb{P}_1}$ does not violate any negative *concept* inclusion from $\mathcal{T}$. It remains to treat the case of negative *role* inclusions. Notice that due to Condition 4 of links, and the facts that $(d_2, e_2) \in \mathrm{P}^{\mathfrak{I}^{\mathbb{P}_2}}$, $e_2 \notin \Delta^{\mathcal{I}^*}$, and $d_2 \in \Delta^{\mathcal{I}^*}$, we must have $\mathfrak{gen}^{\mathbb{P}_2} = \{d_2\} \subseteq \Delta^{\mathcal{I}^*}$. It follows then from Condition 6 that $\mathsf{next}_{\mathbb{Q}}(h_2) = d_2$. We consider two cases:

- If $\mathfrak{gen}^{\mathbb{P}_1} = \{d_1\}$, we obtain similarly $\mathsf{next}_{\mathbb{Q}}(h_1) = d_1$. Denoting $h_1 := \mathrm{R}_1.\mathrm{B}_1$ and $h_2 := \mathrm{R}_2.\mathrm{B}_2$, we obtain, by definition of a prediction, that R_1 and R_2 are non-contradictory. Due to Condition 4 (on the link between $\mathbb{Q}$ and $\mathbb{P}_2$), we have $\mathcal{T} \models \mathrm{R}_2 \sqsubseteq \mathrm{P}$. Therefore $\mathrm{P}(d_1, e_1)$ is non-contradictory with $\mathrm{R}_1(d_1, e_1)$ and hence with $\mathfrak{I}^{\mathbb{P}_1}$ as all roles between d_1 and e_1 in $\mathfrak{I}^{\mathbb{P}_1}$ are consequences of $\mathrm{R}_1(d_1, e_1)$ (Condition 4 on the link given by $\mathbb{P}_1 = \mathsf{ch}^{h_1}_{\mathbb{Q},e_1}$).
- If $\mathfrak{gen}^{\mathbb{P}_1} \neq \{d_1\}$, then there are no roles between d_1 and e_1 (Condition 4), hence no negative role inclusion is violated by adding fact $\mathrm{P}(d_1, e_1)$ in $\mathfrak{I}^{\mathbb{P}_1}$.

3. $(d_1, d_2 \notin \Delta^{\mathcal{I}^*})$ and $(e_1 \in \Delta^{\mathcal{I}^*}$ or $e_2 \in \Delta^{\mathcal{I}^*})$.
This case is symmetric to Case 2.

4. $d_1, d_2, e_1, e_2 \notin \Delta^{\mathcal{I}^*}$.
If ($d_1 \in \mathfrak{gen}^{\mathbb{P}_1}$ and $d_2 \in \mathfrak{gen}^{\mathbb{P}_2}$) or ($e_1 \in \mathfrak{gen}^{\mathbb{P}_1}$ and $e_2 \in \mathfrak{gen}^{\mathbb{P}_2}$), then $(*_d)$ (resp $(*_e)$) yields $\mathbb{P}_1 = \mathbb{P}_2$ and we are easily done. Recalling from the note at the end of Case 1 that we may assume that $d_2 \neq e_2$, we are left with 4 subcases, each immediately leading to a contradiction.

(a) $d_2 \in \mathfrak{gen}^{\mathbb{P}_2}$ (thus $d_1 \in \mathfrak{fr}^{\mathbb{P}_1}$ and $e_2 \in \mathfrak{fr}^{\mathbb{P}_2}$) and $e_1 \in \mathfrak{gen}^{\mathbb{P}_1}$. $(*_d)$ yields $w_1 = w_2 \cdot (\mathbb{P}_2, h_2)$ and $(*_e)$ yields $w_2 = w_1 \cdot (\mathbb{P}_1, h_1)$, contradiction.

(b) $d_2 \in \mathfrak{gen}^{\mathbb{P}_2}$ (thus $d_1 \in \mathfrak{fr}^{\mathbb{P}_1}$ and $e_2 \in \mathfrak{fr}^{\mathbb{P}_2}$) and $e_1 \in \mathfrak{fr}^{\mathbb{P}_1}$. $(*_d)$ yields $w_1 = w_2 \cdot (\mathbb{P}_2, h_2)$ and $(*_e)$ yields $w_2 = w_1$, contradiction.

(c) $d_2 \in \mathfrak{fr}^{\mathbb{P}_2}$ (thus $e_2 \in \mathfrak{gen}^{\mathbb{P}_2}$, thus $e_1 \in \mathfrak{fr}^{\mathbb{P}_1}$) and $d_1 \in \mathfrak{gen}^{\mathbb{P}_1}$. $(*_d)$ yields $w_2 = w_1 \cdot (\mathbb{P}_1, h_1)$ and $(*_e)$ yields $w_1 = w_2 \cdot (\mathbb{P}_2, h_2)$, contradiction.

(d) $d_2 \in \mathfrak{fr}^{\mathbb{P}_2}$ (thus $e_2 \in \mathfrak{gen}^{\mathbb{P}_2}$, thus $e_1 \in \mathfrak{fr}^{\mathbb{P}_1}$) and $d_1 \in \mathfrak{fr}^{\mathbb{P}_1}$. $(*_d)$ yields $w_2 = w_1$ and $(*_e)$ yields $w_1 = w_2 \cdot (\mathbb{P}_2, h_2)$, contradiction.

□

Lemmas 4 and 5 in hand, it is then a technicality to verify that $\mathcal{I}^{\diamond}$ is a model of $\mathcal{K}$.

Lemma 6. *$\mathcal{I}^{\diamond}$ is a model of $\mathcal{K}$.*

Proof. We consider each possible shape of assertion and axiom in $\mathcal{K}$:

$\mathrm{A}(\mathsf{a})$. Since $\mathcal{I}^*$ is a model of $\mathcal{A}$, we have $\mathsf{a} \in \mathrm{A}^{\mathcal{I}^*}$. Recall $\mathcal{I}^*$ is the interpretation of the initial pattern. Therefore the definition of $\mathrm{A}^{\mathcal{I}^\diamondsuit}$ gives $\mathsf{a} = \lambda_{\mathbb{P}^*,\emptyset}(\mathsf{a}) \in \mathrm{A}^{\mathcal{I}^\diamondsuit}$.

$\mathrm{P}(\mathsf{a},\mathsf{b})$. Since $\mathcal{I}^*$ is a model of $\mathcal{A}$, we have $(\mathsf{a},\mathsf{b}) \in \mathrm{P}^{\mathcal{I}^*}$. Recall $\mathcal{I}^*$ is the interpretation of the initial pattern. Therefore the definition of $\mathrm{P}^{\mathcal{I}^\diamondsuit}$ gives $(\mathsf{a},\mathsf{b}) = (\lambda_{\mathbb{P}^*,\emptyset}(\mathsf{a}), \lambda_{\mathbb{P}^*,\emptyset}(\mathsf{b})) \in \mathrm{P}^{\mathcal{I}^\diamondsuit}$.

$\mathrm{A} \sqsubseteq \bot$. Let $u \in \mathrm{A}^{\mathcal{I}^\diamondsuit}$. By definition of $\mathrm{A}^{\mathcal{I}^\diamondsuit}$, there exist $w \cdot (\mathbb{P}, h) \in \mathcal{P}$ and an element $e \in \Delta^{\mathfrak{I}^\mathbb{P}}$ such that $e \in \mathrm{A}^{\mathfrak{I}^\mathbb{P}}$ and $\lambda_{w\cdot(\mathbb{P},h)}(e) = u$. Since $\mathfrak{I}^\mathbb{P}$ is $\mathcal{T}$-satisfiable, it yields a contradiction.

$\top \sqsubseteq \mathrm{A}$. Let $u \in \top^{\mathcal{I}^\diamondsuit} = \Delta^{\mathcal{I}^\diamondsuit}$. By definition of $\Delta^{\mathcal{I}^\diamondsuit}$, we have $w \cdot (\mathbb{P}, h) \in \mathcal{P}$ and an element $e \in \Delta^{\mathfrak{I}^\mathbb{P}}$ such that $\lambda_{w\cdot\mathbb{P},h}(e) = u$. Since $e \in \top^{\mathfrak{I}^\mathbb{P}}$ and $\mathfrak{I}^\mathbb{P}$ is $\mathcal{T}$-saturated, it ensures $e \in \mathrm{A}^{\mathfrak{I}^\mathbb{P}}$. Therefore the definition of $\mathrm{A}^{\mathcal{I}^\diamondsuit}$ gives $u = \lambda_{w\cdot(\mathbb{P},h)}(e) \in \mathrm{A}^{\mathcal{I}^\diamondsuit}$.

$\mathrm{A}_1 \sqcap \mathrm{A}_2 \sqsubseteq \mathrm{A}$. Let $u \in \mathrm{A}_1 \sqcap \mathrm{A}_2{}^{\mathcal{I}^\diamondsuit}$. By definition of $\Delta^{\mathcal{I}^\diamondsuit}$, there exist $w\cdot(\mathbb{P}, h) \in \mathcal{P}$ and an element $e \in \Delta^{\mathfrak{I}^\mathbb{P}}$ such that $\lambda_{w\cdot(\mathbb{P},h)}(e) = u$. Lemma 4 applied on both concepts A_1 and A_2 ensures $e \in \mathrm{A}_1 \sqcap \mathrm{A}_2{}^{\mathfrak{I}^\mathbb{P}}$. Since $\mathfrak{I}^\mathbb{P}$ is $\mathcal{T}$-saturated, it ensures $e \in \mathrm{A}^{\mathfrak{I}^\mathbb{P}}$. Therefore the definition of $\mathrm{A}^{\mathcal{I}^\diamondsuit}$ gives $u = \lambda_{w\cdot(\mathbb{P},h)}(e) \in \mathrm{A}^{\mathcal{I}^\diamondsuit}$.

$\mathrm{A}_1 \sqsubseteq \exists\mathrm{R}.\mathrm{A}_2$. Let $u \in \mathrm{A}_1^{\mathcal{I}^\diamondsuit}$. By definition of $\mathrm{A}_1{}^{\mathcal{I}^\diamondsuit}$, there exist $w \cdot (\mathbb{P}, h) \in \mathcal{P}$ and an element $e \in \Delta^{\mathfrak{I}^\mathbb{P}}$ such that $e \in \mathrm{A}_1{}^{\mathfrak{I}^\mathbb{P}}$ and $\lambda_{w\cdot(\mathbb{P},h)}(e) = u$.
If $e \in \mathfrak{fr}^\mathbb{P}$, then w cannot be empty (recall the initial pattern has an empty frontier!). Hence we have $w = w' \cdot (\mathbb{P}_0, h_0)$ with $e \in \mathfrak{gen}^{\mathbb{P}_0}$ and $\lambda_{w'\cdot(\mathbb{P}_0,h_0)}(e) = u$. Lemma 4 gives $e \in \mathrm{A}_1{}^{\mathfrak{I}^{\mathbb{P}_0}}$. Therefore we can assume w.l.o.g. that $e \in \mathfrak{gen}^\mathbb{P}$, by considering w instead of $w \cdot (\mathbb{P}, h)$.
If $\mathrm{R}.\mathrm{A}_2$ is not applicable to e in $\mathbb{P}$, then this is because there exists $e' \in \mathrm{A}_2{}^{\mathfrak{I}^\mathbb{P}}$ with $(e, e') \in \mathrm{R}^{\mathfrak{I}^\mathbb{P}}$. Set $v := \lambda_{w\cdot(\mathbb{P},h)}(e')$. By definition of $\mathrm{R}^{\mathcal{I}^\diamondsuit}$ and $\mathrm{A}_2{}^{\mathcal{I}^\diamondsuit}$, we obtain $v \in \mathrm{A}_2^{\mathcal{I}^\diamondsuit}$ and $(u, v) \in \mathrm{R}^{\mathcal{I}^\diamondsuit}$.
If $\mathrm{R}.\mathrm{A}_2$ is applicable to e in $\mathbb{P}$, then since $\mathbb{P}$ is accepting there must exist an accepting pattern $\mathbb{P}_1 \in \mathsf{ch}_{\mathbb{P},e}^{\mathrm{R}.\mathrm{A}_2}$. In particular $w \cdot (\mathbb{P}, h) \cdot (\mathbb{P}_1, \mathrm{R}.\mathrm{A}_2) \in \mathcal{P}$. Let e' be the generated term of $\mathbb{P}_1$. From the definition of a link between patterns, we have $(e, e') \in \mathrm{R}^{\mathfrak{I}^{\mathbb{P}_1}}$ and $e' \in \mathrm{A}_2^{\mathfrak{I}^{\mathbb{P}_1}}$. Set $v := \lambda_{w\cdot(\mathbb{P},h)\cdot(\mathbb{P}_1,\mathrm{R}.\mathrm{A}_2)}e'$. Noticing $\lambda_{w\cdot(\mathbb{P},h)}(e) = \lambda_{w\cdot(\mathbb{P},h)\cdot(\mathbb{P}_1,\mathrm{R}.\mathrm{A}_2)}(e)$ and by definition of $\mathrm{R}^{\mathcal{I}^\diamondsuit}$ and $\mathrm{A}_2{}^{\mathcal{I}^\diamondsuit}$, we obtain $v \in \mathrm{A}_2^{\mathcal{I}^\diamondsuit}$ and $(u, v) \in \mathrm{R}^{\mathcal{I}^\diamondsuit}$.

$\exists\mathrm{R}.\mathrm{A}_1 \sqsubseteq \mathrm{A}_2$. Let $u \in (\exists\mathrm{R}.\mathrm{A}_1)^{\mathcal{I}^\diamondsuit}$, that is, there exists $v \in \mathrm{A}_1^{\mathcal{I}^\diamondsuit}$ with $(u, v) \in \mathrm{R}^{\mathcal{I}^\diamondsuit}$. By definition of $\mathrm{R}^{\mathcal{I}^\diamondsuit}$, there exist $w \cdot (\mathbb{P}, h) \in \mathcal{P}$ and elements $e, e' \in \Delta^{\mathfrak{I}^\mathbb{P}}$ such that $(e, e') \in \mathrm{R}^{\mathfrak{I}^\mathbb{P}}$, $\lambda_{w\cdot(\mathbb{P},h)}(e) = u$ and $\lambda_{w\cdot(\mathbb{P},h)}(e') = v$. By Lemma 4 we obtain $e' \in \mathrm{A}_1{}^{\mathfrak{I}^\mathbb{P}}$. Since $\mathfrak{I}^\mathbb{P}$ is $\mathcal{T}$-saturated, we have $e \in \mathrm{A}_2{}^{\mathfrak{I}^\mathbb{P}}$. Therefore by definition of $\mathrm{A}_2{}^{\mathcal{I}^\diamondsuit}$ we obtain $u \in \mathrm{A}_2{}^{\mathcal{I}^\diamondsuit}$.

$\mathrm{P} \sqsubseteq \mathrm{R}$. Let $(u, v) \in \mathrm{P}^{\mathcal{I}^\diamondsuit}$. By definition of $\mathrm{P}^{\mathcal{I}^\diamondsuit}$, there exist $w(\mathbb{P}, h) \in \mathcal{P}$ and elements $e, e' \in \Delta^{\mathfrak{I}^\mathbb{P}}$ such that $(e, e') \in \mathrm{P}^{\mathfrak{I}^\mathbb{P}}$, $\lambda_{w\cdot(\mathbb{P},h)}(e) = u$ and $\lambda_{w\cdot(\mathbb{P},h)}(e') = v$. Since $\mathfrak{I}^\mathbb{P}$ is $\mathcal{T}$-saturated, we have $(e, e') \in \mathrm{R}^{\mathfrak{I}^\mathbb{P}}$. Therefore by definition of $\mathrm{R}^{\mathcal{I}^\diamondsuit}$ we obtain $(u, v) \in \mathrm{R}^{\mathcal{I}^\diamondsuit}$.

$\mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$. Let $(u, v) \in \mathrm{R}_1 \sqcap \mathrm{R}_2{}^{\mathcal{I}^\diamondsuit}$. By definition of $\mathrm{R}_1^{\mathcal{I}^\diamondsuit}$, there exist $w_1 \cdot (\mathbb{P}_1, h_1) \in \mathcal{P}$ and elements $d_1, e_1 \in \Delta^{\mathfrak{I}^{\mathbb{P}_1}}$ such that $(d_1, e_1) \in \mathrm{R}_1^{\mathfrak{I}^{\mathbb{P}_1}}$, $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(d_1) = u$ and $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1) = v$. Similarly, by definition of $\mathrm{R}_2^{\mathcal{I}^\diamondsuit}$, there exist a pattern $w_2 \cdot (\mathbb{P}_2, h_2)$ and elements $d_2, e_2 \in \Delta^{\mathfrak{I}^{\mathbb{P}_2}}$ such that $(d_2, e_2) \in \mathrm{R}_2^{\mathfrak{I}^{\mathbb{P}_2}}$, $\lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(d_2) = u$ and $\lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(e_2) = v$. In particular we have $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(d_1) = \lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(d_2)$ and $\lambda_{w_1\cdot(\mathbb{P}_1,h_1)}(e_1) = \lambda_{w_2\cdot(\mathbb{P}_2,h_2)}(e_2)$. By Lemma 5, we can add (d_1, e_1) to $\mathrm{R}_2^{\mathfrak{I}^{\mathbb{P}_1}}$ while retaining $\mathcal{T}$-satisfiability, contradicting the fact that $\mathcal{T}$ contains $\mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$.

□

It remains to verify that there are no additional c-matches for q in $\mathcal{I}^\diamondsuit$, that is, no more than encoded in $\mathfrak{M}^*$. The inherited tree-like structure of $\mathcal{I}^\diamondsuit$, along with the specifications having to be preserved between linked patterns, ensures that if a match $\pi : q \to \mathcal{I}^\diamondsuit$ exists, then it is actually already taken into account in the specification of the patterns from $\mathcal{P}$. Therefore, if a match maps a counting variable z onto an element of shape $w \cdot (\mathbb{P}, h)$ in $\mathcal{I}^\diamondsuit$, we shall ensure that $(q, z \mapsto s)$, with s either $\circledcirc$ or $\circledast$, belongs to $\mathfrak{M}^\mathbb{P}$. This would contradict $\mathbb{P}$ being accepting. The exact (stronger) statement is as follows.

Lemma 7. *For all $w\cdot(\mathbb{P}, h) \in \mathcal{P}$, if $\pi : r \to \mathcal{I}^\diamondsuit$ is a match of $r \subseteq q$, then we have $(r, \pi') \in \mathfrak{M}^\mathbb{P}$ where $\pi' := (\lambda_{w\cdot(\mathbb{P},h)})^{-1} \circ \pi_{|\Delta}$ with $\Delta := \pi^{-1}(\lambda_{w\cdot(\mathbb{P},h)}(\Delta^{\mathfrak{I}^\mathbb{P}}))$.*

Proof. Considering a breadth-first total order $\leqslant$ on $\mathcal{P}$, and given $W \in \mathcal{P}$, define $\mathcal{I}_W^\diamondsuit$ as follows:

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

We prove by induction on $W \in \mathcal{P}$ that for all $r \subseteq q$, all matches $\pi : r \to \mathcal{I}^{\diamondsuit}_W$ and for all $w \cdot (\mathbb{P}, h) \leqslant W$, we have $(r, \pi') \in \mathfrak{M}^{\mathbb{P}}$ where $\pi' := (\lambda_{w\cdot(\mathbb{P},h)})^{-1} \circ \pi_{|\Delta}$ with $\Delta := \pi^{-1}(\lambda_{w\cdot(\mathbb{P},h)}(\Delta^{\mathfrak{I}^{\mathbb{P}}}))$.

- Assume $W = (\mathbb{P}^*, \emptyset)$, we have $\mathcal{I}^{\diamondsuit}_W = \mathcal{I}^*$. Consider $r \subseteq q$ and a match $\pi : r \to \mathcal{I}^{\diamondsuit}_W$. The only $w \leqslant W$ is $w = W = (\mathbb{P}^*, \emptyset)$. Recalling $\lambda_{(\mathbb{P}^*,\emptyset)} = \mathsf{Id}$, we have $\pi' = \pi$. Therefore (r, π) belongs to the induced specification of $\mathcal{I}^*$. Since $\mathfrak{M}^*$ is coherent, it contains in particular (r, π'), which concludes the base case.
- Assume $W \in \mathcal{P}$ with $(\mathbb{P}^*, \emptyset) < W$ and the statement holds for all $w_0 < W$ (Induction hypothesis 1). Consider $r \subseteq q$ and a match $\pi : r \to \mathcal{I}^{\diamondsuit}_W$. Consider $w \cdot (\mathbb{P}, h) \leqslant W$. Denote d the distance from W to $w \cdot (\mathbb{P}, h)$ in the tree $\mathcal{P}$, that is the number of links required to move from W to $w \cdot (\mathbb{P}, h)$. We prove by induction on d that $(r, \pi') \in \mathfrak{M}^{\mathbb{P}}$ where $\pi' := (\lambda_{w\cdot(\mathbb{P},h)})^{-1} \circ \pi_{|\Delta}$ and with $\Delta := \pi^{-1}(\lambda_{w\cdot(\mathbb{P},h)}(\Delta^{\mathfrak{I}^{\mathbb{P}}}))$.
  - When $d = 0$, we have $W = w \cdot (\mathbb{P}, h)$. Let W' the predecessor of W w.r.t. $\leqslant$. We partition r into r_1 the atoms α from r such that π is a match for α in $\lambda_W(\mathfrak{I}^{\mathbb{P}})$ and r_2 the other atoms, which are hence necessarily mapped by π into $\mathcal{I}^{\diamondsuit}_{W'}$. We denote by $\pi_1 := \pi_{|\mathsf{var}(r_1)}$ and $\pi_2 := \pi_{|\mathsf{var}(r_2)}$ the corresponding restrictions of π.

    First note that since $\mathfrak{M}^{\mathbb{P}}$ is coherent, it contains the pair (r_1, π'_1) where $\pi'_1 := (\lambda_{w\cdot(\mathbb{P},h)})^{-1} \circ (\pi_1)_{|\Delta^1}$ with $\Delta^1 := (\pi_1)^{-1}(\lambda_{w\cdot(\mathbb{P},h)}(\Delta^{\mathfrak{I}^{\mathbb{P}}}))$.

    Letting $w = w' \cdot (\mathbb{Q}, h')$, we next note that applying the Induction Hypothesis 1 on W' with w (which is indeed $\leqslant W'$) and r_2 and π_2, gives us $(r_2, \pi'_2) \in \mathfrak{M}^{\mathbb{Q}}$ where $\pi'_2 := (\lambda_{w'\cdot(\mathbb{Q},h')})^{-1} \circ (\pi_2)_{|\Delta^2}$ with $\Delta^2 := (\pi_2)^{-1}(\lambda_{w'\cdot(\mathbb{Q},h')}(\Delta^{\mathfrak{I}^{\mathbb{Q}}}))$.

    Since $w' \cdot (\mathbb{Q}, h') \cdot (\mathbb{P}, h) \in \mathcal{P}$, we can consider $\mathbb{P} = \mathsf{ch}^h_{\mathbb{Q},e}$, where e denotes the frontier of $\mathbb{P}$. Condition 3 in the definition of a link therefore ensures $(r_2, (\pi'_2)_{|\Delta^{\mathcal{I}^*} \cup \{e\}}) \in \mathfrak{M}^{\mathbb{P}}$. We'd like to form the union of this latter pair with (r_1, π'_1).

    Consider $v \in \mathsf{var}(r_1) \cap \mathsf{var}(r_2)$. Since r_1 contains only atoms that are mapped on $\lambda_W(\mathfrak{I}^{\mathbb{P}})$ by π, the variable v is thus mapped either to an element of Δ^*, to w or to $w \cdot (\mathbb{P}, h)$. The latter is excluded as r_2 only contains atoms that are mapped in $\mathcal{I}^{\diamondsuit}_{W'}$ but $w \cdot (\mathbb{P}, h) \notin \Delta^{\mathcal{I}^{\diamondsuit}_{W'}}$ since $\leqslant$ is breath-first and $W' < W = w \cdot (\mathbb{P}, h)$. If $\pi(v) \in \Delta^*$, then it is clear that π'_1 and $(\pi'_2)_{|\Delta^{\mathcal{I}^*} \cup \{e\}}$ are defined and equal on v. Otherwise $\pi(v) = w$, which yields that $\lambda_W(e) = w$ and $\lambda_w(e) = w$. The first ensures π'_1 is defined on v and equal to w, while the second ensures the same for $(\pi'_2)_{|\Delta^{\mathcal{I}^*} \cup \{e\}}$. As this holds for each variable in $v \in \mathsf{var}(r_1) \cap \mathsf{var}(r_2)$, and that $\mathfrak{M}^{\mathbb{P}}$ is coherent we have $(r_1 \cup r_2, \pi'_1 \cup (\pi'_2)_{|\Delta^{\mathcal{I}^*} \cup \{e\}}) \in \mathfrak{M}^{\mathbb{P}}$, which is the desired pair.
  - Assume now the property holds for all w at distance $d \geq 0$ from W (Induction Hypothesis 2). Let $w_{d+1} \leqslant W$ be exactly at distance $d + 1$ from W. In particular, notice that $w_{d+1} < W$. There exists a link between w_{d+1} and some $w_d \leqslant W$ at distance exactly d from W. We distinguish two cases:
    * $w_{d+1} = w_d \cdot (\mathbb{P}, h)$. We exhibit another suitable partition of r. Denote w^+_{d+1} the elements $w' \cdot (\mathbb{Q}, h') \in \mathcal{P}$ such that w_{d+1} is a prefix of $w' \cdot (\mathbb{Q}, h')$ and $w' \cdot (\mathbb{Q}, h') \leqslant W$. Define r_{d+1} as the atoms α from r such that π is a match for α in some $\lambda_{w'\cdot(\mathbb{Q},h')}(\mathfrak{I}^{\mathbb{Q}})$ with $w' \cdot (\mathbb{Q}, h') \in w^+_{d+1}$. Let r_d consists of the remaining atoms, which are hence mapped on elements that *cannot* admit w_{d+1} as a prefix. Denote by π_{d+1} and π_d the corresponding restrictions of π.
      We first note that $W \notin w^+_{d+1}$, as it would contradict w_d being closer to W than w_{d+1}. Therefore π_{d+1} maps r_{d+1} in $\mathcal{I}^{\diamondsuit}_{W'}$ and we can apply Induction Hypothesis 1 with w_{d+1}, r_{d+1} and π_{d+1}, which provides $(r_{d+1}, \pi'_{d+1}) \in \mathfrak{M}^{\mathbb{P}}$ where $\pi'_{d+1} := (\lambda_{w_d\cdot(\mathbb{P},h)})^{-1} \circ (\pi_{d+1})_{|\Delta^{d+1}}$ with $\Delta^{d+1} := (\pi_{d+1})^{-1}(\lambda_{w_d\cdot(\mathbb{P},h)}(\Delta^{\mathfrak{I}^{\mathbb{P}}}))$.
      Letting $w_d = w_0 \cdot (\mathbb{P}_d, h_d)$, we next note that Induction Hypothesis 2 applied on w_d, r_d and π_d provides $(r_d, \pi'_d) \in \mathbb{P}_d$ where $\pi'_d := (\lambda_{w_0\cdot(\mathbb{P}_d,h_d)})^{-1} \circ (\pi_d)_{|\Delta^d}$ with $\Delta^d := (\pi_d)^{-1}(\lambda_{w_0\cdot(\mathbb{P}_d,h_d)}(\Delta^{\mathfrak{I}^{\mathbb{P}_d}}))$. The link between w_{d+1} and w_d then ensures that $(r_d, (\pi'_d)_{|\Delta^{\mathcal{I}^*} \cup \{e\}}) \in \mathbb{P}$ where e denotes the frontier term of $\mathbb{P}$.
      Consider $v \in \mathsf{var}(r_{d+1}) \cap \mathsf{var}(r_d)$. Since r_{d+1} contains only atoms that are mapped on $\lambda_{w'\cdot(\mathbb{Q},h')}(\mathbb{Q})$ by π for some $w' \cdot (\mathbb{Q}, h') \in w^+_{d+1}$, the variable v is thus mapped either to an element of Δ^*, to w_d or to elements $w' \cdot (\mathbb{Q}, h')$ admitting w_{d+1} as a prefix. But since r_d contains only terms that can not map on elements admitting w_{d+1} as a prefix, only Δ^* or w_d remain possible. Noticing $\lambda_{w_{d+1}}(e) = \lambda_{w_d}(e) = w_d$ if ever $\pi(v) = w_d$ allows to conclude as in the Case $d = 0$.
    * $w_d = w_{d+1} \cdot (\mathbb{P}_d, h_d)$. We exhibit another suitable partition of r. Denote w^+_d the elements $w' \cdot (\mathbb{Q}, h') \in \mathcal{P}$ such that w_d is a prefix of $w' \cdot (\mathbb{Q}, h')$ and $w' \cdot (\mathbb{Q}, h') \leqslant W$. Define r_d as the atoms α from r such that π is a match for α in some $\lambda_{w'\cdot(\mathbb{Q},h')}(\mathfrak{I}^{\mathbb{Q}})$ with $w' \cdot (\mathbb{Q}, h') \in w^+_d$. Let r_{d+1} consists of the remaining atoms, which are hence mapped on elements that *cannot* admit w_d as a prefix. Denote by π_d and π_{d+1} the corresponding restrictions of π.
      We first note that $W \in w^+_d$, as w_d is closer to W than w_{d+1}. Therefore π_{d+1} maps r_{d+1} in $\mathcal{I}^{\diamondsuit}_{W'}$ and we can apply Induction Hypothesis 1 with w_{d+1}, r_{d+1} and π_{d+1}, which provides $(r_{d+1}, \pi'_{d+1}) \in \mathfrak{M}^{\mathbb{P}}$ where $\pi'_{d+1} := (\lambda_{w_d\cdot(\mathbb{P},h)})^{-1} \circ (\pi_{d+1})_{|\Delta^{d+1}}$ with $\Delta^{d+1} := (\pi_{d+1})^{-1}(\lambda_{w_d\cdot(\mathbb{P},h)}(\Delta^{\mathfrak{I}^{\mathbb{P}}}))$.
      We next note that Induction Hypothesis 2 applied on w_d, r_d and π_d provides $(r_d, \pi'_d) \in \mathbb{P}_d$ where $\pi'_d := (\lambda_{w_0\cdot(\mathbb{P}_d,h_d)})^{-1} \circ$

$(\pi_d)_{|\Delta^d}$ with $\Delta^d := (\pi_d)^{-1}(\lambda_{w_0\cdot(\mathbb{P}_d,h_d)}(\Delta^{\mathfrak{I}^{\mathbb{P}_d}}))$. The link between w_{d+1} and w_d then ensures that $(r_d, (\pi'_d)_{|\Delta^{\mathcal{I}^*}\cup\{e\}}) \in \mathbb{P}$ where e denotes the frontier term of $\mathbb{P}_d$.
Consider $v \in \mathsf{var}(r_{d+1}) \cap \mathsf{var}(r_d)$. Since r_d contains only atoms that are mapped on $\lambda_{w'\cdot(\mathbb{Q},h')}(\mathbb{Q})$ by π for some $w' \cdot (\mathbb{Q}, h') \in w_d^+$, the variable v is thus mapped either to an element of Δ^*, to w_{d+1} or to elements $w' \cdot (\mathbb{Q}, h')$ admitting w_d as a prefix. But since r_d contains only terms that can not map on elements admitting w_d as a prefix, only Δ^* or w_{d+1} remain possible. Noticing $\lambda_{w_{d+1}}(e) = \lambda_{w_d}(e) = w_{d+1}$ if ever $\pi(v) = w_{d+1}$ allows to conclude as in the previous cases.

□

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

Proof. Modelhood follows directly from Lemma 6, which is based on Lemmas 4 and 5 presented in the paper. The amount of c-matches is handle by Lemma 7 just above. □

From a Model to an Accepting Initial Pattern We here prove the various claims building an accepting initial pattern from $\mathcal{I}$. The base case consisting of verifying that $\mathbb{P}^*$ is non-trivially rejecting is trivial, and $\mathsf{Id}_{\mathcal{I}^*\to\mathcal{I}}$ is indeed an homomorphism.

We now move to the induction case: assume $\mathbb{P}_1$ has been obtained by the described procedure and has all the desired properties (especially those two mentioned just above). Let $\mathrm{R}.\mathrm{B}$ an existential head that applies on e_1 in $\mathbb{P}_1$.

We first verify that $\tau(\mathbb{P}_2)$ is an homomorphism:

- Let $u \in \mathrm{A}^{\mathfrak{I}^{\mathbb{P}_2}}$. If $u \in \mathrm{A}^{\mathcal{I}^*}$, then in particular $u \in \Delta^*$ hence $\tau(\mathbb{P}_2)(u) = u \in \mathrm{A}^{\mathcal{I}^*} \subseteq \mathrm{A}^{\mathcal{I}}$. Otherwise, $u = e_k$ for $k = 1$ or $k = 2$ with $e'_k \in \mathrm{A}^{\mathcal{I}}$. In that case, notice $\tau(\mathbb{P}_2)(u) = e'_k$ which concludes.
- Let $(u, v) \in \mathrm{P}^{\mathfrak{I}^{\mathbb{P}_2}}$. If $(u, v) \in \mathrm{P}^{\mathcal{I}^*}$, then in particular $u, v \in \Delta^*$, hence $(\tau(\mathbb{P}_2)(u), \tau(\mathbb{P}_2)(v)) = (u, v) \in \mathrm{P}^{\mathcal{I}^*} \subseteq \mathrm{P}^{\mathcal{I}}$. Otherwise, if $(u, v) = (e_1, e_2)$ with $\mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}$, then notice that $(\tau(\mathbb{P}_2)(u), \tau(\mathbb{P}_2)(v)) = (e'_1, e'_2)$. Since e'_2 is the successor of e'_1 for $\mathrm{R}.\mathrm{B}$ in $\mathcal{I}$, and that $\mathcal{I}$ models $\mathcal{T}$, it yields $(e'_1, e'_2) \in \mathrm{P}^{\mathcal{I}}$ which concludes. Otherwise we have $(u, v) = (e_2, e_1)$ with $\mathcal{T} \models \mathrm{R}^- \sqsubseteq \mathrm{P}$, then notice that $(\tau(\mathbb{P}_2)(u), \tau(\mathbb{P}_2)(v)) = (e'_1, e'_1)$. Since e'_2 is the successor of e'_1 for $\mathrm{R}.\mathrm{B}$ in $\mathcal{I}$, and that $\mathcal{I}$ models $\mathcal{T}$, it yields $(e'_2, e'_1) \in \mathrm{P}^{\mathcal{I}}$ which concludes.

We now verify that $\mathbb{P}_2$ as defined in the core paper is indeed a well-defined pattern.

- The frontier e_1 and the generated term e_2 of $\mathbb{P}_2$ are indeed elements from $\Delta^* \cup \{\circledcirc, \circledast\}$.
- The interpretation $\mathfrak{I}^{\mathbb{P}_2}$ is clearly $\mathcal{T}$-satisfiable as it embeds homomorphically by $\tau(\mathbb{P}_2)$ in the *model* $\mathcal{I}$ being a model of $\mathcal{T}$. It is $\mathcal{T}$-saturated since: concepts and roles on $\mathcal{I}^*$ are fully-preserved and they come from the *model* $\mathcal{I}$. The additional concepts on e_1 and e_2 are also all preserved from those on e'_1 and e'_2. The additional roles between e_1 and e_2 are all defined as induced by $\mathrm{R}(e_1, e_2)$ which ensures this edge is also saturated. Finally, it indeed preserves $\mathcal{I}^*$ (we can't have e_1 and e_2 in Δ^* at the same time as it would contradict $\mathrm{R}.\mathrm{B}$ being applicable on $e_1 = e'_1$ since $e_2 = e'_2$ is the $\mathrm{R}.\mathrm{B}$ successor in $\mathcal{I}$!)
- Restrictions of induced specification are coherent hence $\mathfrak{M}^{\mathbb{P}_2}$ is indeed coherent. We verify it preserves $\mathfrak{M}^*$: $(\mathfrak{M}^{\mathbb{P}_2})_{\Delta^*} := ((\mathfrak{M}^{\mathcal{I}})_{|\Delta^*\cup(\tau(\mathbb{P}_2))^{-1}(\{e'_1,e'_2\})})_{|\Delta^*} = (\mathfrak{M}^{\mathcal{I}})_{|\Delta^*} = \mathfrak{M}^*$.
- Let $\mathrm{R}_1.\mathrm{B}_1$ and $\mathrm{R}_2.\mathrm{B}_2$ be two heads such that $\mathcal{T} \models \mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$. By definition of next_2, if it maps $\mathrm{R}_1.\mathrm{B}_1$ and $\mathrm{R}_2.\mathrm{B}_2$ to the same element, it means that the successors of e'_1 for these two heads in $\mathcal{I}$ are equal, which would contradict $\mathcal{I}$ being a model.

The fact that $\mathbb{P}_2$ is not trivially rejecting is trivial as its specification is a restriction of the induced specification of $\mathcal{I}$, which doesn't contain pairs (q, π) that map π outside Δ^* (that is literally the definition of Δ^*!).

We finally verify $\mathbb{P}_2 \in \mathbb{L}^{\mathrm{R}.\mathrm{B}}_{\mathbb{P}_1,e_1}$.

1. We indeed have $\mathfrak{fr}^{\mathbb{P}_2} = \{e_1\}$ and $\mathfrak{gen}^{\mathbb{P}_2} = \{e_2\}$ singletons.
2. $\mathbb{P}_1$ can either be the initial pattern of a built one. In both cases the concepts satisfied on e_1 in $\mathbb{P}_1$ are inherited from those on e'_1. Since it is also the case for $\mathbb{P}_2$, this condition holds.
3. $\mathbb{P}_1$ can either be the initial pattern of a built one. In both cases the specification is the induced specification of $\mathcal{I}$ restricted to the domain of $\mathfrak{I}^{\mathbb{P}_1}$. We directly have the desired equality as both e_1 (seen in $\mathbb{P}_1$ and $\mathbb{P}_2$) comes from the *same* e'_1.
4. This condition matches the definition of $\mathfrak{I}^{\mathbb{P}_2}$.
5. A violation of this condition would imply that e'_1 is the successor of e'_2 for an head incompatible with h would contradict $\mathcal{I}$ being a model.
6. Recall we fixed the choice of successor once and for all!

A.3 Proofs for Section 3.3 (Obtaining Bounded-Size Optimal Models)

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.*

We here focus on proving $\mathcal{J}$ is indeed a model and contains at most as many c-matches as $\mathcal{I}'$. Notice the latter point is equivalent to having at most as many c-matches as $\mathcal{I}^{\diamondsuit}$, as $\mathcal{I}'$ is simply obtained by a renaming of elements from $\mathcal{I}^{\diamondsuit}$. Let us first formulate two remarks concerning the constructed interpretation $\mathcal{J}$.

Remark 3. *The set of concepts from $\mathsf{sig}(\mathcal{T})$ satisfied by $c \in \Delta^{\mathcal{I}'}$ is exactly $\chi_{n,c}(w^c_{n,c})$. Therefore, if $c \sim_n c'$, then c and c' satisfy the same concept names.*

Remark 4. *If $c \sim_n c'$, then $c \sim_m c'$ for any $m \leq n$.*

We now define homomorphisms ρ_c, mentioned in the proof sketch, inductively on $\mathcal{N}_k^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$ with k increasing from 0 to $|q|$. Starting from the element $\overline{c} \in \mathcal{N}_0^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$, we can naturally carry it back as $\rho_c(\overline{c}) = c \in \mathcal{N}_0^{\mathcal{I}',\overline{\Delta^*}}(c)$. Assume now that we have defined $\rho_c(\overline{d})$ for some $\overline{d} \in \mathcal{N}_n^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$ and that we are moving further to an element $\overline{e} \in \mathcal{N}_{n+1}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$ along an edge $(\overline{d},\overline{e})$ in $\mathcal{J}$. In the case of $\overline{e} \notin \overline{\Delta^*}$, the following lemma produces a candidate $\rho_c(\overline{e})$, namely e', which is to $\rho_c(\overline{d})$, namely d', what $\overline{e}$ is to $\overline{d}$.

Lemma 8. *Given two elements $\overline{d},\overline{e} \in \Delta^{\mathcal{J}} \setminus \overline{\Delta^*}$, if there exists a role P from $\mathsf{N}_\mathsf{R}^{\pm}$ such that $(\overline{d},\overline{e}) \in \mathrm{P}^{\mathcal{J}}$, then there exists a unique element $\mathrm{R.B} \in \Omega$ such that one of the two following conditions is satisfied:*

edge$^+$. $|e| = |d| + 1 \mod 2|q|+3$, $w^e_{|q|+1,e} = w^d_{|q|+1-1,d} \cdot \mathrm{R.B}$ *and* $\mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}$.
Furthermore, for all $d' \sim_k d$, the element $e' := d' \cdot \mathrm{R.B}$ belongs to $\Delta^{\mathcal{I}'}$ and satisfies $e' \sim_{k-1} e$.

edge$^-$. $|d| = |e| + 1 \mod 2|q|+3$, $w^d_{|q|+1,d} = w^e_{|q|+1-1,e} \cdot \mathrm{R.B}$ *and* $\mathcal{T} \models \mathrm{R}^- \sqsubseteq \mathrm{P}$.
Furthermore, for all $d' \sim_k d$, we have e' such that $d' = e' \cdot \mathrm{R.B}$ and the prefix e' satisfies $e' \sim_{k-1} e$.

Proof. **Unicity.** Notice the two conditions are mutually exclusive: $|e| = |d|+1 \mod 2|q|+3$ and $|d| = |e|+1 \mod 2|q|+3$ would imply $0 = 2 \mod 2|q|+3$, which is impossible as $2|q|+3 > 2$. Furthermore, in each case R.B is defined as the last letter of the word $w^e_{|q|+1,e}$ (resp $w^d_{|q|+1,d}$), which is unique and does not depends on the choice of e (resp d) nor on P.

Existence and additional property. From the definition of $\mathrm{P}^{\mathcal{J}}$, there exist $(d_0,e_0) \in \mathrm{P}^{\mathcal{I}'}$ such that $\overline{d_0} = \overline{d}$ and $\overline{e_0} = \overline{e}$. Recall $\overline{d},\overline{e} \notin \overline{\Delta^*}$, hence $d_0, e_0 \notin \Delta^*$. In that case the definition of f' ensures the only antecedent of d_0 (resp e_0) by f' is itself. Therefore the definition of $\mathrm{P}^{\mathcal{I}'}$, that is $\sigma(\mathrm{P}^{\mathcal{I}^{\diamondsuit}})$, yields two cases:

- We have $e_0 = d_0 \cdot \mathrm{R.B}$ with $\mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}$. It follows that $|e_0| = |d_0|+1 \mod 2|q|+3$ and $w^{e_0}_{|q|+1,e_0} = w^{d_0}_{|q|+1-1,d_0} \cdot \mathrm{R.B}$, immediately yielding the same properties for d and e as $(\overline{d_0},\overline{e_0}) = (\overline{d},\overline{e})$.
  Let now $1 \leq k \leq |q|+1$ be an integer and $d' \sim_k d$. Transitivity gives $d' \sim_k d_0$, and we have in particular $\chi_{k,d'} = \chi_{k,d_0}$ and $w^{d'}_{k,d'} = w^{d_0}_{k,d_0}$. Recall that $e_0 = d_0 \cdot \mathrm{R.B}$, hence we have $\chi_{k,d_0}(w^{d_0}_{k,d_0} \cdot \mathrm{R.B}) \neq \emptyset$, hence $\chi_{k,d'}(w^{d'}_{k,d'} \cdot \mathrm{R.B}) \neq \emptyset$, that is $d' \cdot \mathrm{R.B}$ is well-defined.
  Notice it is now sufficient to prove $d' \cdot \mathrm{R.B} \sim_{k-1} e_0$: that is because $\overline{e} = \overline{e_0}$, hence transitivity will conclude the proof. It should be clear that $w^{d'\cdot\mathrm{R.B}}_{k-1,d'\cdot\mathrm{R.B}} = w^{e_0}_{k-1,e_0}$ and $|d' \cdot \mathrm{R.B}| = |e_0| \mod 2|q|+3$. Hence we are only left proving that $\chi_{k,e_0} = \chi_{k,d'\cdot\mathrm{R.B}}$.
  First, $e_0 = d_0 \cdot \mathrm{R.B}$ ensures that χ_{k,d_0} fully-determines χ_{k-1,e_0}. Moreover, $\chi_{k,d'}$ fully-determines $\chi_{k-1,d'\cdot\mathrm{R.B}}$. But since $\chi_{k,d_0} = \chi_{k,d'}$ and $w^{e_0}_{k,d_0} = w^{d'\cdot\mathrm{R.B}}_{k,d'}$, we obtain: $\chi_{k,e_0} = \chi_{k,d'\cdot\mathrm{R.B}}$, concluding the proof.
- We have $d_0 = e_0 \cdot \mathrm{R.B}$ with $\mathcal{T} \models \mathrm{R}^- \sqsubseteq \mathrm{P}$. It follows that $|d_0| = |e_0|+1 \mod 2|q|+3$ and $w^{d_0}_{|q|+1,d_0} = w^{e_0}_{|q|+1-1,e_0} \cdot \mathrm{R.B}$, immediately yielding the same properties for d and e as $(\overline{d_0},\overline{e_0}) = (\overline{d},\overline{e})$.
  Let now $1 \leq k \leq |q|+1+1$ be an integer and $d' \sim_k d$. Transitivity gives $d' \sim_k d_0$, and we have in particular $w^{d'}_{1,d'} = w^{d_0}_{1,d_0} = \mathrm{R.B}$ (very important to have $k \geq 1$ here!). That is d' ends by R.B, and therefore we can indeed have prefix e' such that $d' = e' \cdot \mathrm{R.B}$.
  The rest of the proof follows the previous Case 1, this time focusing on the proof of $e' \sim_{k-1} e_0$, based on $d' \sim_k d_0$.

□

Notice the "strength" of the equivalence relation $\sim_k$ between $\overline{e}$ and $\rho_c(\overline{e})$ decreases as we move further in the neighbourhood of $\overline{c}$. However, since we start from $\rho_c(\overline{c}) := c \sim_{|q|+1} c$ and explore a $|q|$-neighbourhood, the index remains at least 1. This is essential as $\sim_1$ encodes relations to elements of $\overline{\Delta^*}$ as the next lemma shows. It allows in particular to treat the case of $\overline{e} \in \overline{\Delta^*}$.

Lemma 9. *If $(\overline{d},\overline{e}) \in \mathrm{R}^{\mathcal{J}}$ for some $e \in \Delta^*$, and if $d' \sim_1 d$, then $(d',e) \in \mathrm{R}^{\mathcal{I}'}$.*

Proof. Recall that since $e \in \Delta^*$ we have $\overline{e} = \{e\}$. The definition of $\mathrm{R}^{\mathcal{J}}$ and further of $\mathrm{R}^{\mathcal{I}'}$ provide $d_0, e_0 \in \Delta^\circ$ such that : $\overline{f'(d_0)} = \overline{d}$, $f'(e_0) = e$ and satisfying $(f'(d_0), f'(e_0)) \in \mathrm{R}^{\mathcal{I}'}$ from one of the following three cases:

- $f'(d_0), f'(e_0) \in \mathrm{R}^{\mathcal{I}^*}$. In particular $f'(d_0) \in \Delta^*$, hence $f'(d_0) = d = d'$. Therefore $(d', e) = (f'(d_0), f'(e_0)) \in \mathrm{R}^{\mathcal{I}'}$.
- $e_0 = d_0 \cdot \mathrm{P.B}$ with $\mathcal{T} \models \mathrm{P} \sqsubseteq \mathrm{R}$. If $f'(d_0) \in \Delta^*$, then we again have $f'(d_0) = d = d'$ immediately yielding $(d', e) \in \mathrm{R}^{\mathcal{I}'}$. Otherwise we have $\chi_{1,f'(d_0)}(w^{f'(d_0)}_{1,f'(d_0)} \cdot \mathrm{P.B}) = f'(e_0) = e$. But since $f'(d_0) \sim_1 d \sim_1 d'$, we have $\chi_{1,d'} = \chi_{1,f'(d_0)}$ and $w^{d'}_{1,d'} = w^{f'(d_0)}_{1,f'(d_0)}$. Therefore $e = \chi_{1,f'(d_0)}(w^{f'(d_0)}_{1,f'(d_0)} \cdot \mathrm{P.B}) = \chi_{1,d'}(w^{d'}_{1,d'}\mathrm{P.B}) = f'(r_{1,d'} w^{d'}_{1,d'}\mathrm{P.B})$. Recalling that $d' = f'(r_{1,d'} w^{d'}_{1,d'})$, we hence obtain $(d', e) = (f'(r_{1,d'} w^{d'}_{1,d'}), f'(r_{1,d'} w^{d'}_{1,d'}\mathrm{P.B})) \in \mathrm{P}^{\mathcal{I}'} \subseteq \mathrm{R}^{\mathcal{I}'}$.
- $d_0 = e_0 \cdot \mathrm{P.B}$ with $\mathcal{T} \models \mathrm{P} \sqsubseteq \mathrm{R}^-$. If $f'(d_0) \in \Delta^*$, then we again have $f'(d_0) = d = d'$ immediately yielding $(d', e) \in \mathrm{R}^{\mathcal{I}'}$. Otherwise the 1-root of $f'(d_0) = d_0$ is e_0 and $w^d_{1,d} = \mathrm{P.B}$. We thus have: $\chi_{1,f'(d_0)}(\varepsilon) = f'(e_0) = e$ (where ε denotes the empty word). But since $f'(d_0) \sim_1 d \sim_1 d'$, we have $\chi_{1,d'} = \chi_{1,f'(d_0)}$ and $w^{d'}_{1,d'} = w^d_{1,d}$. Combining the preceding facts, we obtain $(d', e) = (f'(r_{1,d'} w^{d'}_{1,d'}), \chi_{1,d'}(\varepsilon)) = (f'(r_{1,d'} \cdot \mathrm{P.B}), f'(r_{1,d'})) \in (\mathrm{P}^-)^{\mathcal{I}'} \subseteq \mathrm{R}^{\mathcal{I}'}$.

□

It remains to free ourselves from the particular choice of $\overline{d}$, which is likely not to be the only element of $\mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_n(\overline{c})$ connected to $\overline{e}$. Taking a closer look at Lemma 8, we observe that $\rho_c(\overline{e})$, that is e', is obtained either by adding a letter to $\rho_c(\overline{d})$, that is d', or by removing the last letter of $\rho_c(\overline{d})$, and that these letters coincide with those in the suffixes of elements d and e. Therefore, when moving from $\overline{c}$ to $\overline{e}$ and ignoring self-cancelling steps, each added letter must appear in the suffix of e and, similarly, each removed letter must appear in the suffix of c.

The challenge is therefore to quantify the number of additions and removals to build $\rho_c(\overline{e})$ directly from c and $\overline{e}$. The next definition captures the relative difference of letters between $\overline{c}$ and $\overline{e}$, encoded in $|c|$ and $|e| \mod 2|q|+3$.

Definition 11. *Let $\overline{c} \in \Delta^{\mathcal{J}}$ and $n \leq |q|$. The* relative depth *of $\overline{e} \in \mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_n(\overline{c})$ from $\overline{c}$ is the integer $\delta_{\overline{c}}(\overline{e}) \in [-n, n]$ such that $|e| = |c| + \delta_{\overline{c}}(\overline{e}) \mod 2|q|+3$.*

Remark 5. *By induction on $n \leq |q|$, it is straightforward to see that $\delta_{\overline{c}}(\overline{e})$ is well defined. Unicity is ensured by $\delta_{\overline{c}}(\overline{e}) \leq n \leq |q|$. A consequence of Lemma 8 is that for the smallest $n \leq |q|$ such that $\overline{e} \in \mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_n(\overline{c})$ we have $\delta_{\overline{c}}(\overline{e}) = n \mod 2$.*

We can now identify how many additions and removals cancelled each other. Indeed, if it takes n steps to reach $\overline{e}$ from $\overline{c}$, with relative difference of $\delta := \delta_{\overline{c}}(\overline{e})$, then $n - |\delta|$ is the length of the self-cancelling path, hence: $\frac{n-|\delta|}{2}$ cancelled additions and $\frac{n-|\delta|}{2}$ cancelled removals. Therefore, the actual amount of additions is $\frac{n-|\delta|}{2} + \delta$ if $\delta \geq 0$, or $\frac{n-|\delta|}{2}$ if $\delta \leq 0$, that is in both cases $\frac{n+\delta}{2}$. Similarly we obtain $\frac{n-\delta}{2}$ for the actual amount of removals. The next theorem formalizes all these intuitions: $\rho_{n,c}(\overline{e})$ (in non-trivial cases) is obtained by removing the $\frac{n-\delta}{2}$ last letters of c and keeping the $\frac{n+\delta}{2}$ last letters from the suffix of e. It is then a technicality to verify these syntactical operations on words make sense in the domain of $\mathcal{I}'$.

Theorem 18. *For all $c \in \Delta^{\mathcal{I}'}$ and all $n \leq |q|$, the following mapping:*

$$\rho_{n,c}(\overline{e}) : \mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_n(\overline{c}) \to \mathcal{N}^{\mathcal{I}',\Delta^*}_n(c) \qquad \overline{e} \mapsto \begin{cases} \rho_{n-1,c}(\overline{e}) & \text{if } \overline{e} \in \mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_{n-1}(\overline{c}) \\ e & \text{if } \overline{e} \in \overline{\Delta^*} \\ r_{\frac{n-\delta_{\overline{c}}(\overline{e})}{2},c} \cdot w^e_{\frac{n+\delta_{\overline{c}}(\overline{e})}{2},e} & \text{otherwise} \end{cases}$$

is a homomorphism satisfying $\rho_{n,c}(\overline{e}) \sim_{|q|+1-n} e$ and $\rho^{-1}_{n,c}(\overline{\Delta^*}) \subseteq \overline{\Delta^*}$.

Proof. Let $c \in \Delta^{\mathcal{I}'}$. We proceed by induction on $n \leq |q|$ and prove along a technical statement. Property $\rho_{n,c}(\overline{e}) \sim_{|q|+1-n} e$ will already ensure $w^{\rho_{n,c}(\overline{e})}_{|q|+1-n,\rho_{n,c}(\overline{e})} = w^e_{|q|+1-n,e}$; we reinforce this latter fact as follows. If $e \in \mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_n(\overline{c}) \setminus \mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_{n-1}(\overline{c})$, then:

$$w^{\rho_{n,c}(\overline{e})}_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2},\rho_{n,c}(\overline{e})} = w^e_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2},e} \tag{$*$}$$

It is indeed a stronger statement since $-n \leq \delta_{\overline{c}}(\overline{e}) \leq n$ leads to $0 \leq \frac{n-\delta_{\overline{c}}(\overline{e})}{2} \leq n$, hence $|q|+1-n \leq |q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2}$. Property $*$ therefore provides a more precise information about the suffix of $\rho_{n,c}e$.

Base case: $n = 0$. Let $\overline{e} \in \mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_0(\overline{c})$, hence $\overline{e} = \overline{c}$. If $\overline{c} \in \overline{\Delta^*}$, then $\rho_{0,c}e = e = c$. Otherwise we have $\delta_{\overline{c}}(\overline{e}) = 0$, hence $\rho_{0,c}e = r_{0,c} \cdot w^c_{0,c} = c$. In both cases $\rho_{0,c}e = c$, and it is straightforward that all the desired properties hold. In particular, agreeing that $\mathcal{N}^{\mathcal{J},\overline{\Delta^*}}_{-1}(\overline{c})$ can reasonably be set to $\emptyset$, our technical statement holds.

Induction case. Assume the statement holds for $0 \leq n-1 < |q|$. Let $\overline{e} \in \mathcal{N}_n^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$. If $\overline{e} \in \mathcal{N}_{n-1}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$, then the induction hypothesis applies directly on $\overline{e}$ and provides (stronger versions of) the desired properties. Otherwise, we have by definition of neighbourhoods an element $\overline{d} \in \mathcal{N}_{n-1}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$, not belonging to $\overline{\Delta^*}$ nor to $\mathcal{N}_{n-2}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$, and a role $\mathrm{P} \in \mathsf{N}_{\mathsf{R}}^{\pm}$ such that $(\overline{d}, \overline{e}) \in \mathrm{P}^{\mathcal{J}}$. We apply the induction hypothesis on $\overline{d}$, which gives $\rho_{n-1,c}(\overline{d}) = r_{\frac{n-1-\delta_{\overline{c}}(\overline{d})}{2},d} \cdot w^d_{\frac{n-1+\delta_{\overline{c}}(\overline{d})}{2},d}$ since $\overline{d} \notin \overline{\Delta^*}$. We further distinguish between $\overline{e} \in \Delta^*$ and $\overline{e} \notin \Delta^*$, the latter subcase yielding two subcases by applying Lemma 8 and distinguishing between Cases edge$^+$and edge$^-$. We have therefore three cases to treat.

$\overline{e} \in \overline{\Delta^*}$. We have $\rho_{n,c}(\overline{e}) = e$ and the only non-trivial property to prove is that $e \in \mathcal{N}_n^{\mathcal{I}',\Delta^*}(c)$. Recall the induction hypothesis ensures in particular $\rho_{n-1,c}(\overline{d}) \sim_1 d$. Lemma 9 applies and ensures $(\rho_{n-1,c}(\overline{d}), e) \in \mathrm{P}^{\mathcal{I}'}$, which provides the desired property.

edge$^+$. Case edge$^+$ensures $|e| = |d| + 1 \mod 2|q| + 3$, hence $\delta_{\overline{c}}(\overline{e}) = \delta_{\overline{c}}(\overline{d}) + 1$, and $w^e_{|q|+1,e} = w^d_{|q|+1-1,d} \cdot \mathrm{R.B}$. Therefore, our element $\rho_{n,c}(\overline{e})$ of interest simplifies as:

$$\begin{aligned}
\rho_{n,c}(\overline{e}) &= r_{\frac{n-\delta_{\overline{c}}(\overline{e})}{2},c} \cdot w^e_{\frac{n+\delta_{\overline{c}}(\overline{e})}{2},e} \\
&= r_{\frac{n-(\delta_{\overline{c}}(\overline{d})+1)}{2},c} \cdot w^e_{\frac{n+(\delta_{\overline{c}}(\overline{d})+1)}{2},e} \\
&= r_{\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2},c} \cdot w^e_{\frac{(n-1)+\delta_{\overline{c}}(\overline{d})}{2}+1,e} \\
&= r_{\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2},c} \cdot w^d_{\frac{(n-1)+\delta_{\overline{c}}(\overline{d})}{2},d} \cdot \mathrm{R.B} \\
&= \rho_{n-1,c}(\overline{d}) \cdot \mathrm{R.B},
\end{aligned}$$

which is well-defined and satisfies $\rho_{n,c}(\overline{e}) \sim_{|q|+1-n} e$ from Lemma 8. Recalling that the induction hypothesis gives $\rho_{n-1,c}(\overline{d}) \in \mathcal{N}_{n-1}^{\mathcal{I}',\Delta^*}(c)$, it follows that $\rho_{n,c}(\overline{e}) \in \mathcal{N}_n^{\mathcal{I}',\Delta^*}(c)$. Furthermore, notice that $\overline{e}$ and $\overline{d}$ satisfy all conditions of our additional statement. Since in Case edge$^+$we have $\mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}$, reusing $\rho_{n,c}(\overline{e}) = \rho_{n-1,c}(\overline{d}) \cdot \mathrm{R.B}$ immediately yields $(\rho_{n-1,c}(\overline{d}), \rho_{n,c}(\overline{e})) \in \mathrm{P}^{\mathcal{I}'}$.

Checking that Property $*$ holds is now a technicality, and recall that since $d \in \mathcal{N}_{n-1}^{\mathcal{J},\overline{\Delta^*}}(\overline{c}) \setminus \mathcal{N}_{n-2}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$, we can apply it to d by induction hypothesis. We hence have:

$$\begin{aligned}
w^{\rho_{n,c}(\overline{e})}_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2},\rho_{n,c}(\overline{e})} &= w^{\rho_{n-1,c}(\overline{d})}_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2}-1,\rho_{n-1,c}(\overline{d})} \cdot \mathrm{R.B} \\
&= w^{\rho_{n-1,c}(\overline{d})}_{|q|+1-\frac{(n-1)+1-(\delta_{\overline{c}}(\overline{d})+1)}{2}-1,\rho_{n-1,c}(\overline{d})} \cdot \mathrm{R.B} \\
&= w^{\rho_{n-1,c}(\overline{d})}_{|q|+1-\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2}-1,\rho_{n-1,c}(\overline{d})} \cdot \mathrm{R.B} \\
&= w^d_{|q|+1-\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2}-1,d} \cdot \mathrm{R.B} \\
&= w^e_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2},e}.
\end{aligned}$$

edge$^-$. Case edge$^-$ensures $|e| = |d| - 1 \mod 2|q| + 3$, hence $\delta_{\overline{c}}(\overline{e}) = \delta_{\overline{c}}(\overline{d}) - 1$, and $w^d_{|q|+1,d} = w^e_{|q|+1-1,e} \cdot \mathrm{R.B}$. By induction hypothesis, element $\rho_{n-1,c}(\overline{d}) = r_{\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2},d} \cdot w^d_{\frac{(n-1)+\delta_{\overline{c}}(\overline{d})}{2},d}$ is well-defined. Notice Property $*$ on d (which, again can be applied as $d \in \mathcal{N}_{n-1}^{\mathcal{J},\overline{\Delta^*}}(\overline{c}) \setminus \mathcal{N}_{n-2}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$) gives more precise information on the suffix of $\rho_{n-1,c}(\overline{d})$ than the definition of $\rho_{n-1,c}(\overline{d})$, because $n \leq |q| + 1$ leads to $\frac{(n-1)+\delta_{\overline{c}}(\overline{d})}{2} + 1 \leq |q| + 1 - \frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2}$. Therefore, $w^d_{\frac{(n-1)+\delta_{\overline{c}}(\overline{d})}{2}+1,d}$ is itself a suffix of $w^d_{|q|+1-\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2},d}$, which equals $w^{\rho_{n-1,c}(\overline{d})}_{|q|+1-\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2},\rho_{n-1,c}(\overline{d})}$. Hence we obtain:

$$\begin{aligned}
\rho_{n-1,c}(\overline{d}) &= r_{\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2}+1,d} \cdot w^d_{\frac{(n-1)+\delta_{\overline{c}}(\overline{d})}{2}+1,d} \\
&= r_{\frac{n-\delta_{\overline{c}}(\overline{e})}{2},d} \cdot w^d_{\frac{n+\delta_{\overline{c}}(\overline{e})}{2}+1,d} \\
&= r_{\frac{n-\delta_{\overline{c}}(\overline{e})}{2},d} \cdot w^e_{\frac{n+\delta_{\overline{c}}(\overline{e})}{2},e} \cdot \mathrm{R.B} \\
&= \rho_{n,c}(\overline{e}) \cdot \mathrm{R.B}
\end{aligned}$$

Lemma 8 now ensures $\rho_{n,c}(\overline{e}) \sim_{|q|+1-n} e$ (and could already ensure we can find this suffix of $\rho_{n,c}(\overline{d})$! However, we had to check that the formula still works here, in particular that the suffix of $\rho_{n-1,c}(\overline{d})$ matches long enough the suffix of d) .

Recalling that the induction hypothesis gives $\rho_{n-1,c}(\overline{d}) \in \mathcal{N}_{n-1}^{\mathcal{I}',\Delta^*}(c)$, it follows that $\rho_{n,c}(\overline{e}) \in \mathcal{N}_n^{\mathcal{I}',\Delta^*}(c)$. Furthermore, notice that $\overline{e}$ and $\overline{d}$ satisfy all conditions of our additional statement. Since in Case edge$^-$we have $\mathcal{T} \models \mathrm{R}^- \sqsubseteq \mathrm{P}$, reusing $\rho_{n-1,c}(\overline{d}) = \rho_{n,c}(\overline{e}) \cdot \mathrm{R.B}$ immediately yields $(\rho_{n-1,c}(\overline{d}), \rho_{n,c}(\overline{e})) \in \mathrm{P}^{\mathcal{I}'}$.

Again, we check Property $*$ holds:

$$
\begin{array}{rcl}
w^{\rho_{n,c}(\overline{e})}_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2},\rho_{n,c}(\overline{e})} \cdot \mathrm{R.B} & = & w^{\rho_{n-1,c}(\overline{d})}_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2}+1,\rho_{n-1,c}(\overline{d})} \\
& = & w^{\rho_{n-1,c}(\overline{d})}_{|q|+1-\frac{(n-1)+1-(\delta_{\overline{c}}(\overline{d})-1)}{2}+1,\rho_{n-1,c}(\overline{d})} \\
& = & w^{\rho_{n-1,c}(\overline{d})}_{|q|+1-\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2},\rho_{n-1,c}(\overline{d})} \\
& = & w^{d}_{|q|+1-\frac{(n-1)-\delta_{\overline{c}}(\overline{d})}{2},d} \\
& = & w^{e}_{|q|+1-\frac{n-\delta_{\overline{c}}(\overline{e})}{2},e} \cdot \mathrm{R.B}
\end{array}
$$

We now verify that $\rho_{n,c}$ is a homomorphism.

- Let $\overline{u} \in \mathrm{A}^{\mathcal{J}} \cap \mathcal{N}_n^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$. By definition of $\mathrm{A}^{\mathcal{J}}$, we have $e \in \mathrm{A}^{\mathcal{I}'}$. Since $n \leq |q|$ we have $\rho_{n,c}(\overline{u}) \sim_1 e$, hence applying Remark 3 we obtain $\rho_{n,c}(\overline{u}) \in \mathrm{A}^{\mathcal{I}'}$.
- Let $(\overline{u}, \overline{v}) \in \mathrm{R}^{\mathcal{J}} \cap (\mathcal{N}_n^{\mathcal{J},\overline{\Delta^*}}(\overline{c}) \times \mathcal{N}_n^{\mathcal{J},\overline{\Delta^*}}(\overline{c}))$. If $\overline{u} \in \overline{\Delta^*}$ or $\overline{v} \in \overline{\Delta^*}$, then Lemma 9 applies on $\rho_{n,c}(\overline{u})$ or on $\rho_{n,c}(\overline{v})$ (recall $\rho_{n,c}(\overline{u}) \sim_1 u$ and $\rho_{n,c}(\overline{v}) \sim_1 v$) and gives $(\rho_{n,c}(\overline{u}), \rho_{n,c}(\overline{v})) \in \mathrm{R}^{\mathcal{J}}$. Otherwise $\overline{u} \notin \overline{\Delta^*}$ and $\overline{v} \notin \overline{\Delta^*}$. Let n_1, n_2 be the minimum integers such that $\overline{u} \in \mathcal{N}_{n_1}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$ and $\overline{v} \in \mathcal{N}_{n_2}^{\mathcal{J},\overline{\Delta^*}}(\overline{c})$. Since $(\overline{u}, \overline{v}) \in \mathrm{R}^{\mathcal{J}}$, we have $n_1 - n_2 \in \{-1, 0, 1\}$. Definitions of $\delta_{\overline{c}}(\overline{u})$ and $\delta_{\overline{c}}(\overline{v})$ lead to $|u| - |v| = \delta_{\overline{c}}(\overline{u}) - \delta_{\overline{c}}(\overline{v}) \mod 2|q|+3$. Lemma 8 gives $|u| = |v| \pm 1 \mod 2|q|+3$. Recall $\delta_{\overline{c}}(\overline{u}), \delta_{\overline{d}}(\overline{v}) \in [-|q|, |q|]$, hence $-2|q|-1 \leq \delta_{\overline{c}}(\overline{u}) - \delta_{\overline{c}}(\overline{u}) \mp 1 \leq 2|q|+1$. Since $\delta_{\overline{c}}(\overline{u}) - \delta_{\overline{d}}(\overline{v}) \mp 1 = 0 \mod 2|q|+3$ and $2|q|+1 < 2|q|+3$, we must have $\delta_{\overline{c}}(\overline{u}) - \delta_{\overline{c}}(\overline{v}) = \pm 1$. Joint to Remark 5, it excludes the case $n_1 - n_2 = 0$. We are hence left with $n_1 = n_2 \pm 1$. Applying our additional property with $k := \max(n_1, n_2)$ gives $(\rho_{n,c}(\overline{u}), \rho_{n,c}(\overline{v})) \in \mathrm{R}^{\mathcal{I}'}$.

Finally, $\rho_{n,c}^{-1}(\Delta^*) \subseteq \overline{\Delta^*}$ is a straightforward consequence of $\rho_{n,c}(\overline{u}) \sim_1 u$ (and again, recall elements from Δ^* are alone in their equivalent class!).

□

Let us clarify how Theorem 18 concludes our proof.

Proof of Theorem 2.

Modelhood. We first prove that $\mathcal{J}$ is indeed a model by considering each possible shape of assertions and axioms:

$\mathrm{A}(\mathsf{a})$. Since $\mathcal{I}'$ is a model, we have $\mathsf{a} \in \mathrm{A}^{\mathcal{I}'}$. Therefore, the definition of $\mathrm{A}^{\mathcal{J}}$ gives $\overline{\mathsf{a}} = \mathsf{a} \in \mathrm{A}^{\mathcal{J}}$.

$\mathrm{P}(\mathsf{a}, \mathsf{b})$. Since $\mathcal{I}'$ is a model, we have $(\mathsf{a}, \mathsf{b}) \in \mathrm{P}^{\mathcal{I}'}$. Therefore, the definition of $\mathrm{P}^{\mathcal{J}}$ gives $(\overline{\mathsf{a}}, \overline{\mathsf{b}}) = (\mathsf{a}, \mathsf{b}) \in \mathrm{P}^{\mathcal{J}}$.

$\mathrm{A} \sqsubseteq \bot$. Since $\mathcal{I}'$ is a model, we have $\mathrm{A}^{\mathcal{I}'} = \emptyset$. Therefore, the definition of $\mathrm{A}^{\mathcal{J}}$ gives $\mathrm{A}^{\mathcal{J}} = \overline{\emptyset} = \emptyset$.

$\top \sqsubseteq \mathrm{A}$. Let $u \in \top^{\mathcal{J}} = \Delta^{\mathcal{J}}$. By definition of $\Delta^{\mathcal{J}}$, there exists $u_0 \in \Delta^{\mathcal{I}'}$ such that $\overline{u_0} = u$. Since $u_0 \in \top^{\mathcal{I}'}$ and $\mathcal{I}'$ is a model, it ensures $u_0 \in \mathrm{A}^{\mathcal{I}'}$. Therefore the definition of $\mathrm{A}^{\mathcal{J}}$ gives $u = \overline{u_0} \in \mathrm{A}^{\mathcal{J}}$.

$\mathrm{A}_1 \sqcap \mathrm{A}_2 \sqsubseteq \mathrm{A}$. Let $u \in (\mathrm{A}_1 \sqcap \mathrm{A}_2)^{\mathcal{J}}$. By definition of $\mathrm{A}_1^{\mathcal{J}}$ and $\mathrm{A}_2^{\mathcal{J}}$, there exists $u_1 \in \mathrm{A}_1^{\mathcal{I}'}$ and $u_2 \in \mathrm{A}_2^{\mathcal{I}'}$ with $\overline{u_1} = \overline{u_2} = u$. Remark 3 ensures u_1 and u_2 satisfy the same concepts, that is in particular $u_1 \in (\mathrm{A}_1 \sqcap \mathrm{A}_2)^{\mathcal{I}'}$. Since $\mathcal{I}'$ is a model, it ensures $u_1 \in \mathrm{A}^{\mathcal{I}'}$, yielding by definition of $\mathrm{A}^{\mathcal{J}}$ that $u = \overline{u_1} \in \mathrm{A}^{\mathcal{J}}$.

$\mathrm{A}_1 \sqsubseteq \exists \mathrm{R}.\mathrm{A}_2$. Let $u \in \mathrm{A}_1^{\mathcal{J}}$. By definition of ${\mathrm{A}_1}^{\mathcal{J}}$ there exists $u_0 \in {\mathrm{A}_1}^{\mathcal{I}'}$ with $\overline{u_0} = u$. Since $\mathcal{I}'$ is a model, it ensures there exists $v_0 \in {\mathrm{A}_2}^{\mathcal{I}'}$ with $(u_0, v_0) \in \mathrm{R}^{\mathcal{I}'}$. By definition of ${\mathrm{A}_2}^{\mathcal{J}}$ and $\mathrm{R}^{\mathcal{J}}$, the element $v := \overline{v_0}$ satisfies both $v \in {\mathrm{A}_2}^{\mathcal{J}}$ and $(u, v) \in \mathrm{R}^{\mathcal{J}}$, that is $u \in (\exists \mathrm{R}.\mathrm{A}_2)^{\mathcal{J}}$.

$\exists \mathrm{R}.\mathrm{A}_1 \sqsubseteq \mathrm{A}_2$. Let $u \in (\exists \mathrm{R}.\mathrm{A}_1)^{\mathcal{J}}$, that is there exists $v \in \mathrm{A}_1^{\mathcal{J}}$ with $(u, v) \in \mathrm{R}^{\mathcal{J}}$. By definition of $\mathrm{A}_1^{\mathcal{J}}$ and $\mathrm{R}^{\mathcal{J}}$, there exist $(u_0, v_0) \in \mathrm{R}^{\mathcal{I}'}$ and $v_1 \in \mathrm{A}_1^{\mathcal{I}'}$ such that $\overline{u_0} = u$ and $\overline{v_0} = \overline{v_1} = v$. Remark 3 ensures v_0 and v_1 satisfy the same concepts, that is in particular $u_0 \in (\exists \mathrm{R}.\mathrm{A}_1)^{\mathcal{I}'}$. Since $\mathcal{I}'$ is a model, it ensures $u_0 \in \mathrm{A}_2^{\mathcal{I}'}$, yielding by definition of $\mathrm{A}_2^{\mathcal{J}}$ that $u = \overline{u_0} \in \mathrm{A}_2^{\mathcal{J}}$.

$\mathrm{P} \sqsubseteq \mathrm{R}$. Let $(u, v) \in \mathrm{P}^{\mathcal{J}}$. By definition of $\mathrm{P}^{\mathcal{J}}$, there exists $(u_0, v_0) \in \mathrm{P}^{\mathcal{I}'}$ such that $\overline{u_0} = u$ and $\overline{v_0} = v$. Since $\mathcal{I}'$ is a model, it ensures $(u_0, v_0) \in \mathrm{R}^{\mathcal{I}'}$, hence $(\overline{u_0}, \overline{v_0}) = (u, v) \in \mathrm{R}^{\mathcal{J}}$ by definition of $\mathrm{R}^{\mathcal{J}}$.

$\mathrm{R}_1 \sqcap \mathrm{R}_2 \sqsubseteq \bot$. By contradiction, assume one can find $(u, v) \in (\mathrm{R}_1 \sqcap \mathrm{R}_2)^{\mathcal{J}}$. By definition of $\mathrm{R}_1^{\mathcal{J}}$ and $\mathrm{R}_2^{\mathcal{J}}$, there exists $(u_1, v_1) \in \mathrm{R}_1^{\mathcal{I}'}$ and $(u_2, v_2) \in \mathrm{R}_2^{\mathcal{I}'}$ such that $\overline{u_1} = \overline{u_2} = u$ and $\overline{v_1} = \overline{v_2} = v$.
If $u_1, v_1 \in \Delta^*$, then, each element from Δ^* being alone in its equivalence class, we have $u_1 = u_2$ and $v_1 = v_2$. In particular it gives $(u_1, v_1) \in (\mathrm{R}_1 \sqcap \mathrm{R}_2)^{\mathcal{I}'}$, contradicting $\mathcal{I}'$ being a model.

Otherwise say $u_1 \notin \Delta^*$ (the case of $v_1 \notin \Delta^*$ is symmetrical), hence $\overline{v_1} \in \mathcal{N}_1^{\mathcal{J},\overline{\Delta^*}}(\overline{u_1})$. Theorem 18 gives a homomorphism from $\mathcal{N}_1^{\mathcal{J},\overline{\Delta^*}}(\overline{u_1})$ to $\mathcal{N}_1^{\mathcal{I}',\Delta^*}(u_1)$. But since $(\overline{u_1},\overline{v_1}) \in (\mathrm{R}_1 \sqcap \mathrm{R}_2)^{\mathcal{J}}$, we obtain a contradiction with $\mathcal{I}'$ being a model.

Amount of c-matches. We now prove $\mathcal{J}$ contains at most as many matches as $\mathcal{I}'$ by building an injection from matches in $\mathcal{J}$ to matches in $\mathcal{I}'$. Assume we have a match $\pi : q \to \mathcal{J}$. Consider the set of variables $\mathbf{v}_\pi := \{v \mid v \in \mathbf{y} \cup \mathbf{z}, \pi(v) \notin \overline{\Delta^*}\}$. Let $\mathcal{C}$ denote the set of connected components of $\mathbf{v}_\pi$ in $q_{|\mathbf{v}_\pi}$ (that is the query obtained by keeping only those atoms containing variables from $\mathbf{v}_\pi$). For each connected component $C \in \mathcal{C}$, chose a reference variable $v_C \in C$. Since π is a homomorphism and that $|C| \leq |q|$, every variable $v \in C$ satisfies $\pi(v) \in \mathcal{N}_{|q|}^{\mathcal{J},\overline{\Delta^*}}(\pi(v_C))$. Let $d_C \in \Delta^{\mathcal{I}'}$ denote your favourite representative for the class of $\pi(v_C)$ (that is $\overline{d_C} = \pi(v_C)$). From Theorem 18, we have a homomorphism $\rho_C : \mathcal{N}_{|q|}^{\mathcal{J},\overline{\Delta^*}}(\pi(v_C)) \to \mathcal{N}_{|q|}^{\mathcal{I}',\Delta^*}(d_C)$. Using these ρ_C, one per $C \in \mathcal{C}$, we define:

$$\begin{array}{rcl} \pi' : \mathbf{x} \cup \mathbf{y} \cup \mathbf{z} & \to & \Delta^{\mathcal{I}'} \\ v & \mapsto & \begin{cases} \rho_C(\pi(v)) & \text{if } v \in C, C \in \mathcal{C} \\ e & \text{if } \pi(v) = \overline{e} \in \overline{\Delta^*} \end{cases} \end{array}$$

Since each ρ_C is a homomorphism (again Theorem 18), we can check the overall π' is also a homomorphism:

- Consider $\mathrm{A}(v) \in q$. If $v \in C$ for some $C \in \mathcal{C}$, then ρ_C being a homomorphism gives $\pi'(v) \in \mathrm{A}^{\mathcal{I}'}$. Otherwise $\pi(v) = \overline{e} \in \overline{\Delta^*}$, but since π is a homomorphism we have $\pi(v) \in \mathrm{A}^{\mathcal{J}}$. Since $\overline{e} = \{e\}$ and by definition of $\mathrm{A}^{\mathcal{J}}$, it ensures $e \in \mathrm{A}^{\mathcal{I}'}$, that is $\pi'(v) \in \mathrm{A}^{\mathcal{I}'}$.
- Consider $\mathrm{R}(u, v) \in q$.
  - If both $\pi(u), \pi(v) \notin \overline{\Delta^*}$, then we can find $C \in \mathcal{C}$ such that $u, v \in C$, and then we use ρ_C being a homomorphism.
  - If both $\pi(u), \pi(v) \in \overline{\Delta^*}$, then the definition of $\mathrm{R}^{\mathcal{J}}$ provides $(u_0, v_0) \in \mathrm{R}^{\mathcal{I}'}$ with $\overline{u_0} = \pi(u) \in \overline{\Delta^*}$ and $\overline{v_0} = \pi(v) \in \overline{\Delta^*}$. Hence $\overline{u_0} = \{u_0\}$ and $\overline{v_0} = \{v_0\}$, which gives $(\pi'(u), \pi'(v)) \in \mathrm{R}^{\mathcal{I}'}$.
  - If $\pi(u) \notin \overline{\Delta^*}$ and $\pi(v) \in \overline{\Delta^*}$, then we have $\pi'(u) = \rho_C(\pi(u))$ for some $C \in \mathcal{C}$. Theorem 18 ensures $\pi'(u) \sim_1 \pi(u)$, and since π is a homomorphism, we also have $(\pi(u), \pi(v)) \in \mathrm{R}^{\mathcal{J}}$. Therefore we can apply Lemma 9 and we obtain $(\pi'(u), \pi'(v)) \in \mathrm{R}^{\mathcal{I}'}$.

In particular, π' is a match, hence $\pi'(\mathbf{z}) \subseteq \Delta^*$. Using property $\rho_C^{-1}(\Delta^*) \subseteq \overline{\Delta^*}$ for each $C \in \mathcal{C}$, provided by Theorem 18 along with definition of π', we obtain that $\pi(\mathbf{z}) \subseteq \overline{\Delta^*}$. Since $\rho_{|\overline{\Delta^*}} = \mathsf{Id}$, we have that the application $\pi_{|\mathbf{z}} \mapsto \pi'_{|\mathbf{z}}$ is injective. Therefore $\mathcal{J}$ contains at most as much matches as $\mathcal{I}'$ does.

Size of the model. Finally, an equivalence class $\overline{d}$ is characterized by: $|d| \mod 2|q| + 3$, that is one equivalence class among $2|q| + 3$ possible classes; $w^d_{|q|+1,d}$, that is a word over an alphabet with at most $|\mathcal{T}|$ symbols and a length at most $|q| + 1$; and $\chi_{|q|+1,d}$, that is a function from words over an alphabet with at most $|\mathcal{T}|$ symbols and length at most $2|q| + 1$ to a set with size at most $|\Delta^*|+2^{|\mathsf{sig}(\mathcal{T})|} + 1$. Therefore, the amount of possibly different equivalence classes, that is $|\Delta^{\mathcal{J}}|$, is at most $(2|q| + 3) \times |\mathcal{T}|^{|q|+2} \times (|\Delta^*|+2^{|\mathsf{sig}(\mathcal{T})|} + 1)^{|\mathcal{T}|^{2|q|+3}}$. Recall Lemma 1 allows to assume $|\Delta^*| \leq |\mathsf{Ind}| + (|\mathsf{Ind}(\mathcal{A})| + 3\,|\mathcal{T}|\,2^{|\mathcal{T}|})^{|q|}|q|$, we have the claimed bounds for the size of $\mathcal{J}$, which concludes the proof of Theorem 2. □

A.4 Proofs for Section 3.4 (Matching Lower Bounds)

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

Proof. To complete the proof sketch, we need to prove the following claim: $N + 1$ is a certain answer to q' over $\mathcal{K}'$ iff $\mathcal{K}$ entails q.

First assume that $N + 1$ is certain answer to q' over $\mathcal{K}'$, and consider a model $\mathcal{I}$ of $\mathcal{K}$. Add aux and all the associated facts from $\mathcal{A}' \setminus \mathcal{A}$ to obtain a model $\mathcal{I}'$ of $\mathcal{K}'$. Observe that $\mathcal{I}'$ must contain at least N matches: the disjuncts q_k and the queries q_{P}^+ and q_{P}^- all have a match sending all variables to aux, and each q_{A} has n matches due to $\mathcal{A}$, plus one more sending $\exists z_{\mathrm{A}}$ to aux. Since $N + 1$ is a certain answer, there must exists some additional match for q' in $\mathcal{I}'$. As $\mathcal{I}$ is a model of $\mathcal{K}$, it interprets each $A \in \Sigma$ as $\{\mathsf{a} \mid \mathrm{A}(\mathsf{a}) \in \mathcal{A}\}$, so there are no further matches for q_{A}. Next note that since aux is disconnected from the rest of $\mathcal{I}'$, there is no extra match for each $q_{\mathrm{P}}^{\pm}$. The only possibility then is that must be an extra match for one of the q_k, aside from the one mapping all variables aux. Since q_k is connected, this extra match is fully contained in $\Delta^{\mathcal{I}'} \setminus \{\mathsf{aux}\}$. Hence, $\mathcal{I}$ contains a match for q_k. We may thus conclude that $\mathcal{K}$ entails q.

For the other direction, suppose that $\mathcal{K}$ entails q, and consider a model $\mathcal{I}'$ of $\mathcal{K}'$. There are at least N trivial matches for q' in $\mathcal{I}'$. If there is an extra match for one of the q_{A} or one of the $q_{\mathrm{P}}^{\pm}$, then we are done. Otherwise, removing aux from $\mathcal{I}'$ yields a model $\mathcal{I}$ of $\mathcal{K}$. Since $\mathbf{a}_{\emptyset}$ is the certain answer of q over $\mathcal{K}$, there must be a match for one of the q_k in $\mathcal{I}$. It yields a new match for q_k in $\mathcal{I}'$ and concludes. □

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

Proof. The proof is by reduction from the exponential grid tiling problem EXPTIL. We recall that an instance of this problem consists of a set $\mathcal{C}$ of colors, two relations $\mathcal{H}, \mathcal{V} \subseteq \mathcal{C} \times \mathcal{C}$ that give the horizontal and vertical tiling conditions, and a number n. The task is to decide whether there exists a valid $(\mathcal{H}, \mathcal{V})$-tiling of an $2^n \times 2^n$ grid, i.e., a mapping $\tau : \{0, \ldots, 2^n - 1\} \times \{0, \ldots, 2^n - 1\} \mapsto \mathcal{C}$ such that $(\tau(i,j), \tau(i+1, j)) \in \mathcal{H}$ for every $0 \leq i < 2^n - 1$ and $(\tau(i,j), \tau(i, j+1)) \in \mathcal{V}$ for every $0 \leq j < 2^n - 1$. In what follows, we consider an instance $(n, \mathcal{C}, \mathcal{H}, \mathcal{V})$ of the EXPTIL problem.

To be able to test for the existence of a tiling of a $2^n \times 2^n$ grid, we must start by ensuring we can find such a grid in each model. Furthermore, we will need to detect horizontal and vertical adjacency in this grid, it is thus appropriate to use horizontal/vertical coordinates. To ensure a polynomial reduction, we need to use a binary encoding of these coordinates. We start from an initial element a and use TBox axioms to generate all possible coordinates of the horizontal coordinates:

$$\mathrm{A}(\mathsf{a}) \qquad \mathrm{A} \sqsubseteq \exists \mathrm{R_{h,n-1,b}} \qquad \exists \mathrm{R^-_{h,i,b}} \sqsubseteq \exists \mathrm{R_{h,i-1,b'}} \qquad \begin{pmatrix} i = 1, \ldots n \\ b, b' \in \{0,1\} \end{pmatrix}$$

We proceed similarly with the vertical coordinates, until we generate all possible pairs of coordinates:

$$\exists \mathrm{R^-_{h,0,b}} \sqsubseteq \exists \mathrm{R_{v,n-1,b'}} \qquad \exists \mathrm{R^-_{v,i,b}} \sqsubseteq \exists \mathrm{R_{v,i-1,b'}} \qquad \begin{pmatrix} i = 1, \ldots n \\ b, b' \in \{0,1\} \end{pmatrix}$$

The preceding axioms will generate a binary tree of height $2n$ in the canonical model, whose leaves represent all possible grid positions. We use the following axiom to assign a color to each of the points representing a grid position:

$$\exists \mathrm{R^-_{v,0,b}} \sqsubseteq \exists \mathrm{HasCol} \qquad (b \in \{0,1\})$$

To help us compare positions, we will include the following TBox axioms:

$$\exists \mathrm{R^-_{d,i,b}} \sqsubseteq \exists \mathrm{HasBit_{d,j}} \qquad \begin{pmatrix} 0 \leq i < j \leq n-1 \\ b \in \{0,1\} \\ d \in \{h, v\} \end{pmatrix}$$

and:

$$\exists \mathrm{R^-_{v,i,b}} \sqsubseteq \exists \mathrm{HasBit_{h,j}} \qquad \begin{pmatrix} 0 \leq i,j \leq n-1 \\ b \in \{0,1\} \end{pmatrix}$$

To keep track of elements used as color or bits, we also add:

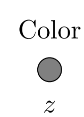
$$\exists \mathrm{HasCol^-} \sqsubseteq \mathrm{Color} \qquad \exists \mathrm{HasBit^-_{d,i}} \sqsubseteq \mathrm{Bit} \qquad \begin{pmatrix} 0 \leq i \leq n-1 \\ d \in \{h, v\} \end{pmatrix}$$

This completes our description of the TBox. We will finish our description of the ABox later in the proof, but it will be useful to know that it will contain an ABox individual c for every color $c \in \mathcal{C}$ and two ABox individuals (one, zero) to represent bits.

Let us now define the query q. In what follows, we build q step by step, providing several subqueries. For the sake of readability, we omit subscript/superscripts that would allow to decide which variable occurs in which subquery. The reason is simple: *in what follows, no variable is shared by different subqueries.*

To keep track of the colors used in a candidate tiling, we use the following subquery:

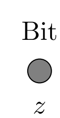
$$q_{\mathrm{Color}} := \exists z\, \mathrm{Color}(z)$$



The query q_{Color}

We also need to detect if other bits than the intended ones (one, zero) are being used to satisfy the right hand sides $\exists \mathrm{HasBit}_{d,i}$. For this purpose, we introduce the following subquery:

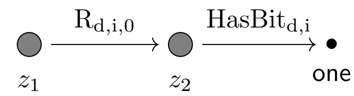
$$q_{\mathrm{Bit}} := \exists z\, \mathrm{Bit}(z)$$



The query q_{Bit}

To detect if the i^{th} bit of the coordinate in direction d is one when it should be zero:

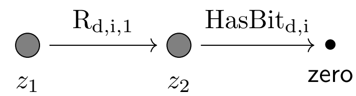
$$q_{d,i,\mathsf{one}} := \exists z_1 \exists z_2 \, \mathrm{R_{d,i,0}}(z_1, z_2) \wedge \mathrm{HasBit_{d,i}}(z_2, \mathsf{one}) \qquad \begin{pmatrix} 0 \le i \le n-1 \\ d \in \{h, v\} \end{pmatrix}$$



The query $q_{d,i,\mathsf{one}}$

And the other way around:

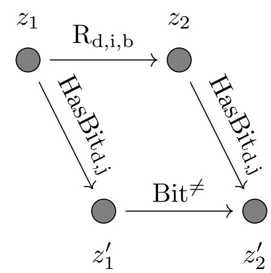
$$q_{d,i,\mathsf{zero}} := \exists z_1 \exists z_2 \, \mathrm{R_{d,i,1}}(z_1, z_2) \wedge \mathrm{HasBit_{d,i}}(z_2, \mathsf{zero}) \qquad \begin{pmatrix} 0 \le i \le n-1 \\ d \in \{h, v\} \end{pmatrix}$$



The query $q_{d,i,\mathsf{zero}}$

To detect if the j^{th} bit of the coordinate in direction d isn't carried from the i^{th} level to the next:

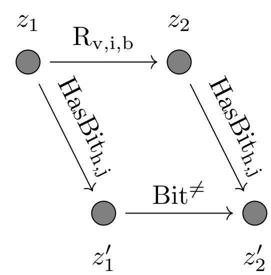
$$\begin{aligned} q_{d,i,b,j} := \exists z_1 \, \exists z_2 \, \exists z_1' \, \exists z_2' \, \mathrm{R_{d,i,b}}(z_1, z_2) \, \wedge \, \mathrm{HasBit_{d,j}}(z_1, z_1') \\ \wedge \, \mathrm{HasBit_{d,j}}(z_2, z_2') \, \wedge \, \mathrm{Bit}^{\neq}(z_1', z_2') \end{aligned} \qquad \begin{pmatrix} 0 \le i < j \le n-1 \\ b \in \{0, 1\} \\ d \in \{h, v\} \end{pmatrix}$$



The query $q_{d,i,b,j}$

To detect if the j^{th} bit of the horizontal coordinate isn't carried through the i^{th} vertical level:

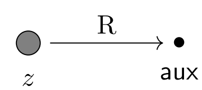
$$\begin{aligned} q_{i,b,j} := \exists z_1 \, \exists z_2 \, \exists z_1' \, \exists z_2' \, \mathrm{R_{v,i,b}}(z_1, z_2) \, \wedge \, \mathrm{HasBit_{h,j}}(z_1, z_1') \\ \wedge \, \mathrm{HasBit_{h,j}}(z_2, z_2') \, \wedge \, \mathrm{Bit}^{\neq}(z_1', z_2') \end{aligned} \qquad \begin{pmatrix} 0 \le i, j \le n-1 \\ b \in \{0, 1\} \end{pmatrix}$$



The query $q_{i,b,j}$

To detect if part of the model is collapsing on the auxiliary individual:

$$q_{\mathsf{aux},\mathrm{R}} := \exists z \, \mathrm{R}(z, \mathsf{aux}) \qquad (\mathrm{R} = \mathrm{R_{d,i,b}}, \mathrm{HasBit_{d,i}}, \mathrm{HasCol})$$



The query $q_{\mathsf{aux},\mathrm{R}}$

We next discuss the parts of the query that are used to check the tiling conditions. To detect adjacency, we remark that two grid positions $(h_1, v_1), (h_2, v_2) \in \{0, \ldots, 2^n - 1\} \times \{0, \ldots, 2^n - 1\}$ are vertically adjacent iff:

- $h_1 = h_2$, so the binary encodings of h_1 and h_2 are the same;
- $v_2 = v_1 + 1$, so the binary encodings of v_2 and v_1 are the same until, at some point, v_2 ends with $1 \cdot 0^k$ while v_1 ends with $0 \cdot 1^k$.

To detect a violation of the vertical tiling condition (i.e. two vertically adjacent tiles with colors c and c' such that $(c, c') \notin \mathcal{V}$), we need n queries, one for each possible position where the bit from the vertical coordinates differ. For each $1 \le k \le n$, we create a subquery $q^{\mathcal{V},(c,c'),k}$ defined as follows.

$$\begin{aligned} q^{\mathcal{V},(c,c'),k} = \; & \exists z_l \, \exists z_r \, \exists z_{h,0} \ldots \exists z_{h,n-1} \, \exists z_{v,k+1} \ldots \exists z_{v,n-1} \\ & \bigwedge_{i=0}^{n-1} (\mathrm{HasBit_{h,i}}(z_l, z_{h,i}) \wedge \mathrm{HasBit_{h,i}}(z_r, z_{h,i})) \, \wedge \bigwedge_{i=k+1}^{n-1} (\mathrm{HasBit_{v,i}}(z_l, z_{v,i}) \wedge \mathrm{HasBit_{v,i}}(z_r, z_{v,i})) \\ & \wedge \, \mathrm{HasBit_{v,k}}(z_l, \mathsf{zero}) \wedge \mathrm{HasBit_{v,k}}(z_r, \mathsf{one}) \, \wedge \bigwedge_{i=0}^{k-1} (\mathrm{HasBit_{v,i}}(z_l, \mathsf{one}) \wedge \mathrm{HasBit_{v,i}}(z_r, \mathsf{zero})) \\ & \wedge \, \mathrm{HasCol}(z_l, \mathsf{c}) \wedge \mathrm{HasCol}(z_r, \mathsf{c}') \end{aligned}$$

We can similarly define a set of subqueries $q^{\mathcal{H},(c,c'),k}$ that detect violations of the horizontal tiling conditions (see *e.g.* Figure 4).

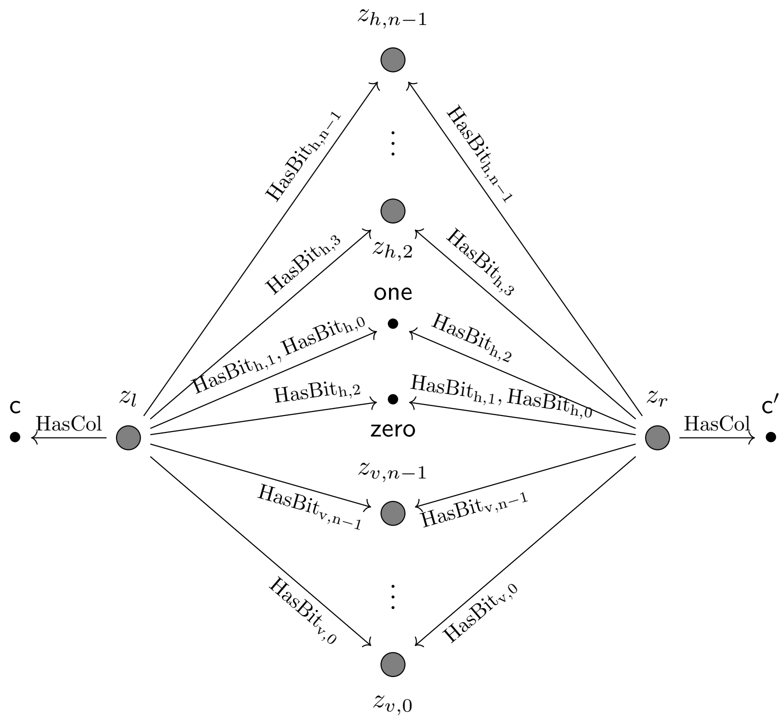


Figure 4: The query $q^{\mathcal{H},(c,c'),2}$

Finally, we let q be the conjunction of the all of the preceding subqueries.

We can now define the ABox, which introduces individuals for the intended colors and bits and a further individual d that serves to ensure that all parts of the query can be matched:

$$\begin{aligned}\mathcal{A} = &\{\mathrm{Root}(\mathsf{a}), \mathrm{Bit}(\mathsf{zero}), \mathrm{Bit}(\mathsf{one}), \mathrm{Bit}^{\neq}(\mathsf{zero}, \mathsf{one}), \mathrm{Bit}^{\neq}(\mathsf{one}, \mathsf{zero})\}\\ &\cup \{\mathrm{Color}(\mathsf{c}) \mid c \in \mathcal{C}\}\\ &\cup \{\mathrm{Root}(\mathsf{aux}), \mathrm{Bit}(\mathsf{aux}), \mathrm{Color}(\mathsf{aux}), \mathrm{Bit}^{\neq}(\mathsf{aux}, \mathsf{aux}), \mathrm{HasCol}(\mathsf{aux}, \mathsf{aux})\}\\ &\cup \{\mathrm{R_{d,i,b}}(\mathsf{aux}, \mathsf{aux}) \mid d \in \{h, v\}, i \in \{0, \dots n-1\}, b \in \{0, 1\}\}\\ &\cup \{\mathrm{HasBit_{d,i}}(\mathsf{aux}, \mathsf{aux}) \mid d \in \{h, v\}, i \in \{0, \dots n-1\}\}\end{aligned}$$

Let $p = |\mathcal{C}|$, and let $\mathcal{K}$ be the KB with the preceding TBox and ABox. To complete the proof, it suffices to establish the following claim:

Claim $[3p + 4, +\infty]$ is a certain answer for q over $\mathcal{K} \iff (n, \mathcal{C}, \mathcal{H}, \mathcal{V}) \notin \mathsf{EXPTIL}$.

First observe that there are always at least $3(p + 1)$ c-matches given by: $p + 1$ mappings for q_{Color} (on each color-individual c and on aux), times 3 mappings for q_{Bit} (on zero, one and aux), times 1 mapping for each other subquery (collapse on aux).

$(\Rightarrow)$ Assume $[3p+4, +\infty]$ is a certain answer, and take some candidate tiling $\tau : \{0, \dots 2^n - 1\} \times \{0, \dots 2^n - 1\} \to \{\mathsf{c} \mid c \in \mathcal{C}\}$. Let $\mathcal{I}_\tau$ be the model of $\mathcal{K}$ that is obtained from $\mathcal{C}_\mathcal{K}$ as follows:

- $\Delta^{\mathcal{I}_\tau}$ contains all elements from $\Delta^{\mathcal{C}_\mathcal{K}}$ except those anonymous elements whose last symbol is HasCol or $\mathrm{HasBit}_{d,i}$ (i.e. witnesses for axioms involving $\exists\mathrm{HasCol}$ or $\exists\mathrm{HasBit}_{d,i}$);
- the roles HasCol and $\mathrm{HasBit}_{d,i}$ are interpreted as follows:

$$\mathrm{HasBit}_{d,i}^{\mathcal{I}_\tau} := \{(\mathsf{aux}, \mathsf{aux})\} \cup \{(\mathsf{a}w\mathrm{R}_{d,i,0}w', \mathsf{zero}) \mid \mathsf{a}w\mathrm{R}_{d,i,0}w' \in \Delta^{\mathcal{I}_\tau}\} \cup \{(\mathsf{a}w\mathrm{R}_{d,i,1}w', \mathsf{one}) \mid \mathsf{a}w\mathrm{R}_{d,i,1}w' \in \Delta^{\mathcal{I}_\tau}\}$$

$$\begin{aligned}\mathrm{HasCol}^{\mathcal{I}_\tau} := \{(\mathsf{aux}, \mathsf{aux})\} \cup \{(&\mathsf{a}\mathrm{R_{h,n-1,h_{n-1}}} \dots \mathrm{R_{h,0,h_0}} \mathrm{R_{v,n-1,v_{n-1}}} \dots \mathrm{R_{v,0,v_0}}, \tau(h_{n-1} \dots h_0, v_{n-1} \dots v_0))\\ &\mid h_{n-1}, \dots h_0, v_{n-1}, \dots v_0 \in \{0, 1\}\}\end{aligned}$$

where by a slight abuse of notation, we use $\tau(h_{n-1}\dots h_0, v_{n-1}\dots v_0)$ to mean $\tau(h,v)$, with h and v the numbers whose binary encodings are $h_{n-1}\dots h_0$ and $v_{n-1}\dots v_0$ respectively;

- the remaining roles are interpreted exactly as in $\mathcal{C}_\mathcal{K}$.

Recall our assumption that there is an additional c-match π for q in $\mathcal{I}_\tau$. It is easily verified that the additional match can only result from one of the queries $q^{h,(c,c'),k}$ or $q^{v,(c,c'),k}$. From the definition of $\mathcal{I}_\tau$, this implies that there are two horizontally (or vertically) adjacent tiles, which positions are encoded on $\pi(z_l)$ and $\pi(z_r)$ by the endpoints of their respective roles $\mathrm{HasBit}_{d,i}$, whose respective colors c and c' violate either $\mathcal{H}$ or $\mathcal{V}$. Thus τ is not an $(\mathcal{H},\mathcal{V})$-tiling. As this construction holds for any possible tiling τ, we infer that $(n,\mathcal{C},\mathcal{H},\mathcal{V}) \notin \mathsf{EXPTIL}$.

$(\Leftarrow)$ Assume $(n,\mathcal{C},\mathcal{H},\mathcal{V}) \notin \mathsf{EXPTIL}$, and take some model $\mathcal{I}$ of $\mathcal{K}$. There is a homomorphism $f : \mathcal{C}_\mathcal{K} \to \mathcal{I}$. If there exists $\mathsf{a}w \in \Delta^{\mathcal{C}_\mathcal{K}}$ such that $f(\mathsf{a}w) = \mathsf{aux}$, then there exists a new c-match for the subquery $q_{\mathsf{aux},\mathrm{R}}$, where R is the last letter of the shortest prefix w' of w such that $f(\mathsf{a}w') = \mathsf{aux}$. Otherwise, we define $\tau : \{0,\dots,2^n-1\} \times \{0,\dots 2^n-1\} \to \Delta^\mathcal{I}$ as follows: $\tau(h_{n-1}\dots h_0, v_{n-1}\dots v_0) := f(\mathsf{a}\mathrm{R}_{h,n-1,h_{n-1}}\dots \mathrm{R}_{h,0,h_0}\mathrm{R}_{v,n-1,v_{n-1}}\dots \mathrm{R}_{v,0,v_0}\mathrm{HasCol})$ (again slightly abusing notation by working with binary encodings of numbers). There are five cases to consider:

- If there exists $(h_{n-1}\dots h_0, v_{n-1}\dots v_0)$ such that $\tau(h_{n-1}\dots h_0, v_{n-1}\dots v_0) \notin \{\mathsf{c} \mid c \in \mathcal{C}\}$, then this provides a new c-match of q in $\mathcal{I}$ in which the subquery q_{Color} is mapped as $z \mapsto \tau(h_{n-1}\dots h_0, v_{n-1}\dots v_0)$.
- Otherwise, suppose there exists an element that is in the range of Bit that is not zero nor one, then this also provides a new c-match of q, in which the subquery q_{Bit} is mapped on this element.
- Otherwise, suppose there exists an inconsistent choice of bit, that is $\mathsf{a}w\mathrm{R}_{d,i,0}$ and $f(\mathsf{a}w\mathrm{R}_{d,i,0}\mathrm{HasBit}_{d,i}) = \mathsf{one}$ (respectively: $\mathsf{a}w\mathrm{R}_{d,i,1}$ and $f(\mathsf{a}w\mathrm{R}_{d,i,0}\mathrm{HasBit}_{d,i}) = \mathsf{zero}$), then it provides a new c-match for the subquery $q_{d,i,\mathsf{one}}$ (resp: $q_{d,i,\mathsf{zero}}$).
- Otherwise, suppose there exists an non-propagated coordinate, that is $\mathsf{a}w\mathrm{R}_{d,i,b}$ such that $f(\mathsf{a}w\mathrm{HasBit}_{d',k}) \neq f(\mathsf{a}w\mathrm{R}_{d,i,b}\mathrm{HasBit}_{d',k})$, then it provides a new c-match either for the subquery $q_{d,i,b,j}$ or for the subquery $q_{i,b,j}$.
- Else, since $(n,\mathcal{C},\mathcal{H},\mathcal{V}) \notin \mathsf{EXPTIL}$, there exist two adjacent positions with coordinates $p := (h_{n-1}\dots h_0, v_{n-1}\dots v_0)$ and $p' := (h'_{n-1}\dots h'_0, v'_{n-1}\dots v'_0)$ such that $(\tau(p),\tau(p')) \in (\mathcal{C}\times\mathcal{C}) \setminus \mathcal{D}$, for $\mathcal{D}$ either $\mathcal{H}$ or $\mathcal{V}$. Letting k be the bit from which the encoding of the non-$\mathcal{D}$ coordinate differs, we obtain a new c-match for q, in which the subquery $q^{\mathcal{D},(\tau(p),\tau(p')),k}$ is satisfied by mapping z_l to $f(\mathsf{a}\mathrm{R}_{h,n-1,h_{n-1}}\dots \mathrm{R}_{h,0,h_0}\mathrm{R}_{v,n-1,v_{n-1}}\dots \mathrm{R}_{v,0,v_0})$ and z_r to $\mathsf{a}\mathrm{R}_{h,n-1,h'_{n-1}}\dots \mathrm{R}_{h,0,h'_0}\mathrm{R}_{v,n-1,v'_{n-1}}\dots \mathrm{R}_{v,0,v'_0})$ (or the converse).

In every case, there is an additional c-match for q. We thus obtain that $[p+1,+\infty]$ is a certain answer to q over $\mathcal{K}$.

□

B Proofs for Section 4 (Cardinality Queries)

B.1 Proofs for Section 4.1 (Results for $\mathcal{EL}$ and its Extensions)

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

Proof. An instance of Succinct-3COL consists of a Boolean circuit $\mathcal{C}$ with $2n$ input gates. The graph $\mathcal{G}_\mathcal{C}$ encoded by $\mathcal{C}$ has 2^n vertices, identified by binary encodings on n bits. Two vertices u and v, with respective binary encodings $u_1\dots u_n$ and $v_1\dots v_n$, are adjacent in $\mathcal{G}_\mathcal{C}$ iff $\mathcal{C}$ returns True when given as input $u_1\dots u_n$ on its first n gates and $v_1\dots v_n$ on the second half. The problem of deciding if $\mathcal{G}_\mathcal{C}$ is 3-colorable has been proven to be NEXP-complete in (Papadimitriou and Yannakakis 1986).

Let $\mathcal{C}$ be an instance of Succinct-3COL, and denote $2n$ its amount of input gates. We start by generating an exponential tree, henceforth referred to as the *reference tree*, to assign a color to each vertex, that is a binary identifier (k ranges from 1 to n):

$$\mathrm{U_0}(\mathsf{a}) \quad \begin{array}{llll} \mathrm{U_{k-1}} \sqsubseteq \exists \mathrm{R.A^0_k} & \mathrm{A^0_k} \sqsubseteq \mathrm{U_k} & \exists \mathrm{R^-.A^0_k} \sqsubseteq \mathrm{A^0_k} & \\ \mathrm{U_{k-1}} \sqsubseteq \exists \mathrm{R.A^1_k} & \mathrm{A^1_k} \sqsubseteq \mathrm{U_k} & \exists \mathrm{R^-.A^1_k} \sqsubseteq \mathrm{A^1_k} & \end{array} \quad \mathrm{A^0_k} \sqcap \mathrm{A^1_k} \sqsubseteq \bot$$

At the end of a branch, we ask for a color to be chosen among three provided options. The color can actually be chosen elsewhere, but at the cost of a new c-match for our query q_{Goal}.

$$\mathrm{U_n} \sqsubseteq \exists \mathrm{HasCol.Color} \quad \mathrm{Color} \sqsubseteq \mathrm{Goal} \quad \begin{array}{l} \mathrm{Color}(\mathsf{c_1}) \\ \mathrm{Color}(\mathsf{c_2}) \\ \mathrm{Color}(\mathsf{c_3}) \end{array}$$

We now generate all possible pairs of vertex identifiers, starting from the first identifier (k ranges from 1 to n):

$$\mathrm{V_0}(\mathsf{b}) \quad \begin{array}{lll} \mathrm{V_{k-1}} \sqsubseteq \exists \mathrm{R.B^0_k} & \mathrm{B^0_k} \sqsubseteq \mathrm{V_k} & \exists \mathrm{R^-.B^0_k} \sqsubseteq \mathrm{B^0_k} \\ \mathrm{V_{k-1}} \sqsubseteq \exists \mathrm{R.B^1_k} & \mathrm{B^1_k} \sqsubseteq \mathrm{V_k} & \exists \mathrm{R^-.B^1_k} \sqsubseteq \mathrm{B^1_k} \end{array} \quad \mathrm{B^0_k} \sqcap \mathrm{B^1_k} \sqsubseteq \bot$$

and followed by the second identifier (k ranges from 1 to n):

$$\mathrm{V_n} \sqsubseteq \mathrm{W_0} \qquad \begin{array}{l} \mathrm{W_{k-1}} \sqsubseteq \exists \mathrm{R.C^0_k} \\ \mathrm{W_{k-1}} \sqsubseteq \exists \mathrm{R.C^1_k} \end{array} \qquad \begin{array}{l} \mathrm{C^0_k} \sqsubseteq \mathrm{W_k} \\ \mathrm{C^1_k} \sqsubseteq \mathrm{W_k} \end{array} \qquad \begin{array}{l} \exists \mathrm{R^-.C^0_k} \sqsubseteq \mathrm{C^0_k} \\ \exists \mathrm{R^-.C^1_k} \sqsubseteq \mathrm{C^1_k} \end{array} \qquad \mathrm{C^0_k} \sqcap \mathrm{C^1_k} \sqsubseteq \bot$$

At the end of a branch, we ask for each node to be connected to the two corresponding nodes from the reference tree.

$$\begin{array}{l} \mathrm{W_n} \sqsubseteq \exists \mathrm{FstCol.Goal} \\ \mathrm{W_n} \sqsubseteq \exists \mathrm{SndCol.Goal} \end{array} \qquad \begin{array}{l} \exists \mathrm{FstCol^-.B^0_k} \sqsubseteq \mathrm{A^0_k} \\ \exists \mathrm{FstCol^-.B^1_k} \sqsubseteq \mathrm{A^1_k} \end{array} \qquad \begin{array}{l} \exists \mathrm{SndCol^-.C^0_k} \sqsubseteq \mathrm{A^0_k} \\ \exists \mathrm{SndCol^-.C^1_k} \sqsubseteq \mathrm{A^1_k} \end{array} \qquad \mathrm{U_n} \sqsubseteq \mathrm{Goal} \quad \mathrm{U_n} \sqcap \mathrm{Color} \sqsubseteq \bot$$

Notice axioms $\mathrm{U_n} \sqsubseteq \mathrm{Goal}$ and $\mathrm{U_n} \sqcap \mathrm{Color} \sqsubseteq \bot$ act as an incentive to reuse elements from the reference tree, otherwise it would come at the cost of a new c-match for our query q_{Goal}. We also note that, at this point, there are always at least $2^n + 3$ matches in every model given by the three possible colors $\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3$ and the 2^n instance of $\mathrm{U_n}$, which must all be disjoint. Finally, we import the chosen colors from the reference tree with the following assertions and axioms:

$$\begin{array}{lll} \mathrm{Col_1}(\mathsf{c}_1) & \exists \mathrm{FstCol.}(\exists \mathrm{HasCol.Col_1}) \sqsubseteq \mathrm{Col^{fst}_1} & \exists \mathrm{SndCol.}(\exists \mathrm{HasCol.Col_1}) \sqsubseteq \mathrm{Col^{snd}_1} \\ \mathrm{Col_2}(\mathsf{c}_2) & \exists \mathrm{FstCol.}(\exists \mathrm{HasCol.Col_2}) \sqsubseteq \mathrm{Col^{fst}_2} & \exists \mathrm{SndCol.}(\exists \mathrm{HasCol.Col_2}) \sqsubseteq \mathrm{Col^{snd}_2} \\ \mathrm{Col_3}(\mathsf{c}_3) & \exists \mathrm{FstCol.}(\exists \mathrm{HasCol.Col_2}) \sqsubseteq \mathrm{Col^{fst}_3} & \exists \mathrm{SndCol.}(\exists \mathrm{HasCol.Col_3}) \sqsubseteq \mathrm{Col^{snd}_3} \end{array}$$

It remains to evaluate the circuit to test adjacency for each pair of vertices identifiers. This is handled by the TBox in the following fashion. For the first n input gates g_k^{fst} introduce the axioms:

$$\mathrm{B^0_k} \sqsubseteq \mathrm{IsFalse_{g^{fst}_k}} \qquad \mathrm{B^1_k} \sqsubseteq \mathrm{IsTrue_{g^{fst}_k}} \qquad (k = 1, \ldots, n)$$

and for the remaining n input gates g_k^{snd} introduce the axioms:

$$\mathrm{C^0_k} \sqsubseteq \mathrm{IsFalse_{g^{snd}_k}} \qquad \mathrm{C^1_k} \sqsubseteq \mathrm{IsTrue_{g^{snd}_k}} \qquad (k = 1, \ldots, n).$$

For each negation gate g with parent gate g_0, we introduce the two axioms:

$$\mathrm{IsFalse_{g_0}} \sqsubseteq \mathrm{IsTrue_g} \qquad \mathrm{IsTrue_{g_0}} \sqsubseteq \mathrm{IsFalse_g}.$$

For each conjunctive gate g with parent gates g_1 and g_2, introduce the three axioms:

$$\mathrm{IsTrue_{g_1}} \sqcap \mathrm{IsTrue_{g_2}} \sqsubseteq \mathrm{IsTrue_g} \qquad \begin{array}{l} \mathrm{IsFalse_{g_1}} \sqsubseteq \mathrm{IsFalse_g} \\ \mathrm{IsFalse_{g_2}} \sqsubseteq \mathrm{IsFalse_g}. \end{array}$$

For each disjunctive gate g with parent gates g_1 and g_2, introduce the three axioms:

$$\begin{array}{l} \mathrm{IsTrue_{g_1}} \sqsubseteq \mathrm{IsTrue_g} \\ \mathrm{IsTrue_{g_2}} \sqsubseteq \mathrm{IsTrue_g} \end{array} \qquad \mathrm{IsFalse_{g_1}} \sqcap \mathrm{IsFalse_{g_2}} \sqsubseteq \mathrm{IsFalse_g}.$$

Finally, to detect monochromatic edges, consider the three axioms where g_{out} denotes the output gate of $\mathcal{C}$:

$$\begin{array}{l} \mathrm{IsTrue_{g_{out}}} \sqcap \mathrm{Col^{fst}_1} \sqcap \mathrm{Col^{snd}_1} \sqsubseteq \mathrm{Goal} \\ \mathrm{IsTrue_{g_{out}}} \sqcap \mathrm{Col^{fst}_2} \sqcap \mathrm{Col^{snd}_2} \sqsubseteq \mathrm{Goal} \\ \mathrm{IsTrue_{g_{out}}} \sqcap \mathrm{Col^{fst}_3} \sqcap \mathrm{Col^{snd}_3} \sqsubseteq \mathrm{Goal} \end{array}$$

To ensure this case indeed creates a new match for q_{Goal} we make sure that it cannot be an already existing match with the two negative concept inclusions:

$$\mathrm{W_n} \sqcap \mathrm{Color} \sqsubseteq \bot \qquad \mathrm{U_n} \sqcap \mathrm{W_n} \sqsubseteq \bot$$

Claim: $\mathcal{C} \notin$ Succinct-3COL iff $2^n + 4$ is a certain answer for q_{Goal} over $\mathcal{K}$.

For readability, we omit the concepts associated with the evaluation of the circuit when considering elements of $\mathcal{C}_\mathcal{K}$.

($\Longrightarrow$). Assume $\mathcal{C} \notin$ Succinct-3COL and consider a model $\mathcal{I}$ of $\mathcal{K}$. There exists an homomorphism from the canonical model of $\mathcal{K}$ to this $\mathcal{I}$, say we choose one such $f : \mathcal{C}_\mathcal{K} \to \mathcal{I}$.

If any of the $f(\mathsf{a} \ldots \mathrm{R}.\{\mathrm{U_n}, \mathrm{Goal}, \mathrm{A^{a_1}_1}, \ldots \mathrm{A^{a_n}_n}\}\mathrm{HasCol}.\{\mathrm{Color}, \mathrm{Goal}\}) \notin \{\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3\}$, then it provides a new c-match for q_{Goal} and we are done.

Otherwise, denote by τ the coloring induced by the reference tree in $\mathcal{I}$, defined by setting $\tau(a_0 \ldots a_{n-1}) := f(\mathsf{a} \ldots \mathrm{R}.\{\mathrm{U_n}, \mathrm{Goal}, \mathrm{A^{a_1}_1}, \ldots \mathrm{A^{a_n}_n}\}\mathrm{HasCol}.\{\mathrm{Color}, \mathrm{Goal}\})$. Since $\mathcal{C} \notin$ Succinct-3COL and τ only uses the 3 colors c_1, c_2 and c_3, there must exists a monochromatic edge $\{u, v\}$. Denote by $b_0, \ldots, b_{n-1}$ the identifier of u, by $c_0, \ldots, c_{n-1}$ the identifier of v, and by k the number of the shared color $\mathsf{c_k}$. Since u and v are adjacent, the concept $\mathrm{IsTrue_{g_{out}}}$ is satisfied on the element $e := f(\mathsf{b} \ldots \mathrm{R}.\{\mathrm{W_n}, \mathrm{B^{b_1}_1}, \ldots \mathrm{B^{b_n}_n}, \mathrm{C^{c_1}_1}, \ldots \mathrm{C^{c_n}_n}\}$ of $\mathcal{I}$.

If the element $f(\mathsf{b}\ldots\mathrm{R}.\{\mathrm{W_n},\mathrm{B_1^{b_1}},\ldots\mathrm{B_n^{b_n}},\mathrm{C_1^{c_1}},\ldots\mathrm{C_n^{c_n}}\}\mathrm{FstCol}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{b_1}},\ldots,\mathrm{A_n^{b_n}}\})$ (notice the first vertex identifier is converted into an identifier in the reference tree) is not equal to the corresponding element from the reference tree, that is $f(\mathsf{a}\ldots\mathrm{R}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{b_1}},\ldots\mathrm{A_n^{b_n}}\})$, then it yields a new c-match and we are done.

Otherwise, axiom $\exists\mathrm{FstCol}.(\exists\mathrm{HasCol}.\mathrm{Col_k})\sqsubseteq\mathrm{Col_k^{fst}}$ ensures $\mathrm{C_k^{fst}}$ holds on e. Similarly, we obtain that either $f(\mathsf{b}\ldots\mathrm{R}.\{\mathrm{W_n},\mathrm{B_1^{b_1}},\ldots\mathrm{B_n^{b_n}},\mathrm{C_1^{c_1}},\ldots\mathrm{C_n^{c_n}}\}\mathrm{SndCol}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{c_1}},\ldots,\mathrm{A_n^{c_n}}\})$ yields a new c-match and we are done or that $\mathrm{C_k^{snd}}$ holds on e. In the latter case, axiom $\mathrm{IsTrue_{g_{out}}}\sqcap\mathrm{Col_k^{fst}}\sqcap\mathrm{Col_k^{snd}}\sqsubseteq\mathrm{Goal}$ triggers a new match on e.

In all cases, we exhibit an additional c-match, which proves 2^n+4 is a certain answer for q_{Goal} over $\mathcal{K}$.

($\Longleftarrow$). Assume $\mathcal{C}\in$ Succinct-3COL and pick a 3-coloring τ of the underlying graph of $\mathcal{C}$, using as colors c_1, c_2 and c_3. From the canonical model of $\mathcal{K}$, identify each element $\mathsf{a}\ldots\mathrm{R}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{a_1}},\ldots\mathrm{A_n^{a_n}}\}\mathrm{HasCol}.\{\mathrm{Color},\mathrm{Goal}\}$ with the individual $\tau(a_0\ldots a_{n-1})$.

Also identify each element $\mathsf{b}\ldots\mathrm{R}.\{\mathrm{W_n},\mathrm{B_1^{b_1}},\ldots\mathrm{B_n^{b_n}},\mathrm{C_1^{c_1}},\ldots\mathrm{C_n^{c_n}}\}\mathrm{FstCol}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{b_1}},\ldots,\mathrm{A_n^{b_n}}\}$ with the element $\mathsf{a}\ldots\mathrm{R}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{b_1}},\ldots\mathrm{A_n^{b_n}}\}$.

Similarly, identify each element $\mathsf{b}\ldots\mathrm{R}.\{\mathrm{W_n},\mathrm{B_1^{b_1}},\ldots\mathrm{B_n^{b_n}},\mathrm{C_1^{c_1}},\ldots\mathrm{C_n^{c_n}}\}\mathrm{SndCol}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{c_1}},\ldots,\mathrm{A_n^{c_n}}\}$ with the element $\mathsf{a}\ldots\mathrm{R}.\{\mathrm{U_n},\mathrm{Goal},\mathrm{A_1^{c_1}},\ldots\mathrm{A_n^{c_n}}\}$.

Saturate the obtained interpretation to obtain a model $\mathcal{I}_\tau$ of $\mathcal{K}$. Because τ is a 3-coloring, there is no monochromatic edge, hence it can be verified that $\mathcal{I}_\tau$ has exactly the 2^n+3 original c-matches. This provides a model of $\mathcal{K}$ with less than 2^n+4 c-matches for q_{Goal}, ensuring 2^n+4 is not a certain answer for q_{Goal} over $\mathcal{K}$.

□

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. Let $\mathcal{L}$ be a sublogic of $\mathcal{ELHI}_\bot$ for which every satisfiable KB admits a polynomial-sized model. Then proceeding similarly to Lemma 1, we can exhibit a polynomial p such that for every satisfiable KB $\mathcal{K}=(\mathcal{T},\mathcal{A})$ and cardinality query q, there exists a model of $\mathcal{K}$ having at most $p(|\mathcal{K}|)$ matches to q.

With this in mind, let us fix a satisfiable KB $\mathcal{K}=(\mathcal{T},\mathcal{A})$ and a role cardinality query $q_{\mathrm{Goal}}=\exists z_1,z_2\mathrm{Goal}(z_1,z_2)$, and let $n_\mathcal{K}=p(|\mathcal{K}|)$. Consider a set of individual names $D\subseteq\mathsf{N_I}$ of size $2n_\mathcal{K}+|\mathsf{Ind}(\mathcal{A})|$ and containing $\mathsf{Ind}(\mathcal{A})$. For each subset $S\subseteq D\times D$, we check whether the following KB with closed predicates is satisfiable (note that Goal is the only a closed predicate):

$$\mathcal{K}_S:=(\mathcal{T},\{\mathrm{Goal}\},\mathcal{A}\cup\{\mathrm{Goal}(\mathsf{a},\mathsf{b})\mid(\mathsf{a},\mathsf{b})\in S\})$$

If such a KB is satisfiable with Goal a closed predicate, it provides a model of $\mathcal{K}$ with precisely $|S|$ matches. Conversely, if there exists a model $\mathcal{I}$ of $\mathcal{K}$ with $n\leq n_\mathcal{K}$ matches, there exists a subset $S\subseteq D\times D$ such that $\mathcal{K}_S$ is satisfiable: pick S as the pairs $(\varphi(\mathsf{a}),\varphi(\mathsf{b}))\in\mathrm{Goal}^\mathcal{I}$, where φ is an injection from the subset of $\Delta^\mathcal{I}$ appearing in matches of q_{Goal} to D which is the identity on $\mathsf{Ind}(\mathcal{A})$.

By Theorem 7 of (Ngo, Ortiz, and Šimkus 2016), this check can be performed in exponential time in $\mathcal{K}_S$, which is of polynomial size w.r.t. $\mathcal{K}$. Moreover, we consider exponentially many $\mathcal{K}_{S,n}$, hence enumerating them and checking all of them takes single-exponential time, which concludes the proof. □

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

Proof. Let $\mathcal{K}$ be a satisfiable $\mathcal{ELH}_\bot$ KB, which we may suppose w.l.o.g. to be in normal form, and consider the following interpretation $\mathcal{I}_\mathcal{K}$ (a variation on the one defined in (Lutz, Toman, and Wolter 2009) for ${\mathcal{ELH}_\bot}^{dr}$ without negative role inclusions):

$$\begin{aligned}\Delta^{\mathcal{I}_\mathcal{K}}=&\mathsf{Ind}(\mathcal{A})\cup\{x_{\mathrm{R.B}}\mid\mathrm{A}\sqsubseteq\exists\mathrm{R.B}\in\mathcal{T}\text{ and }\mathcal{T}\not\models B\sqsubseteq\bot\}\\ \mathrm{A}^{\mathcal{I}_\mathcal{K}}=&\{\mathsf{a}\mid\mathcal{K}\models\mathrm{A}(\mathsf{a})\}\cup\{x_{\mathrm{R.B}}\mid\mathcal{T}\models\mathrm{B}\sqsubseteq\mathrm{A}\}\\ \mathrm{P}^{\mathcal{I}_\mathcal{K}}=&\{(\mathsf{a},\mathsf{b})\mid\mathcal{K}\models\mathrm{P}(\mathsf{a},\mathsf{b})\}\cup\{(\mathsf{a},x_{\mathrm{R.B}})\mid\mathcal{K}\models\exists\mathrm{R.B}(\mathsf{a}),\mathcal{T}\models\mathrm{R}\sqsubseteq\mathrm{P}\}\cup\\ &\{(x_{\mathrm{R_1.B_1}},x_{\mathrm{R_2.B_2}})\mid\mathcal{T}\models\mathrm{B_1}\sqsubseteq\exists\mathrm{R_2.B_2},\mathcal{T}\models\mathrm{R_2}\sqsubseteq\mathrm{P}\}\end{aligned}$$

Note that $|\Delta^{\mathcal{I}_\mathcal{K}}|\leq|\mathcal{K}|$, so we only need to show that $\mathcal{I}_\mathcal{K}$ is a model of $\mathcal{K}$. It is not hard to see that $\mathcal{I}_\mathcal{K}$ satisfies ABox assertions of $\mathcal{A}$ and all concept axioms and positive role inclusions from $\mathcal{T}$. Suppose that $\mathcal{T}$ contains a negative role inclusion $\mathrm{T_1}\sqcap\mathrm{T_2}\sqsubseteq\bot$ and there is a pair $(u,v)\in{\mathrm{T_1}}^{\mathcal{I}_\mathcal{K}}\cap{\mathrm{T_2}}^{\mathcal{I}_\mathcal{K}}$. We cannot have $u,v\in\mathsf{Ind}(\mathcal{A})$, since this would imply that $\mathcal{K}$ is unsatisfiable. If $(u,v)=(\mathsf{a},x_{\mathrm{R.B}})$, then $\mathcal{K}\models\exists\mathrm{R.B}(\mathsf{a})$, $\mathcal{T}\models\mathrm{R}\sqsubseteq\mathrm{T_1}$, and $\mathcal{T}\models\mathrm{R}\sqsubseteq\mathrm{T_2}$, which again means $\mathcal{K}$ is unsatisfiable. Finally suppose that we have $(u,v)=(x_{\mathrm{R_1.B_1}},x_{\mathrm{R_2.B_2}})$. Then $\mathcal{T}\models\mathrm{B_1}\sqsubseteq\exists\mathrm{R_2.B_2}$, $\mathcal{T}\models\mathcal{T}\models\mathrm{R_2}\sqsubseteq\mathrm{T_1}$, and $\mathcal{T}\models\mathrm{R_2}\sqsubseteq\mathrm{T_2}$. But that would mean that $\mathcal{T}\models\mathrm{B_1}\sqsubseteq\bot$, contradicting the definition of $\Delta^{\mathcal{I}_\mathcal{K}}$. We thus conclude that $\mathcal{I}_\mathcal{K}$ is indeed a model of $\mathcal{K}$. □

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

Proof. The proof proceeds by reduction from the problem of deciding if an $\mathcal{EL}$ KB with closed predicates is satisfiable, known to be EXP-hard from (Ngo, Ortiz, and Šimkus 2016). As noticed by the authors in the conclusion of (Ngo, Ortiz, and Šimkus 2016), we point out that their reduction (Propositions 4 and 5) produces a KB $(\mathcal{T}, \Sigma, \mathcal{A})$ such that the set of closed predicates Σ only contains concept names. Therefore, we assume w.l.o.g. that our starting KB $\mathcal{K} := (\mathcal{T}, \Sigma, \mathcal{A})$ also satisfies this property. We also assume that $\mathcal{T}$ is in normal form where every concept inclusion has one of the following restricted shapes:

$$\top \sqsubseteq \mathrm{A} \qquad \mathrm{A} \sqcap \mathrm{B} \sqsubseteq \mathrm{C} \qquad \mathrm{A} \sqsubseteq \exists \mathrm{R}.\mathrm{B} \qquad \exists \mathrm{R}.\mathrm{A} \sqsubseteq \mathrm{B} \qquad \text{with } \mathrm{A}, \mathrm{B}, \mathrm{C} \in \mathsf{N_C}, \mathrm{R} \in \mathsf{N_R}.$$

We will need to consider two fresh new concept names Goal and $\mathrm{Aux}_\top$, a fresh new role name $\mathrm{R_B}$ for each closed concept name $\mathrm{B} \in \Sigma$, and a fresh individual aux. The concept Goal will be our query predicate and aims to capture excessive uses of the closed predicates.

To capture such uses on non-individual elements, we consider the axiom $\mathrm{B} \sqsubseteq \mathrm{Goal}$ for each $\mathrm{B} \in \Sigma$. Therefore, we also consider the assertion $\mathrm{Goal}(\mathsf{a})$ for each a such that there exists $\mathrm{B}(\mathsf{a}) \in \mathcal{A}$ with $\mathrm{B} \in \Sigma$. To prevent such an assertion $\mathrm{Goal}(\mathsf{a})$ to "hide" the use of a by a closed concept B such that $\mathrm{B}(\mathsf{a}) \notin \mathcal{A}$, we introduce the axiom $\exists \mathrm{R_B}.\mathrm{B} \sqsubseteq \mathrm{Goal}$ for each $\mathrm{B} \in \Sigma$ and the assertion $\mathrm{R_B}(\mathsf{aux}, \mathsf{a})$ for each $\mathsf{a} \in \mathsf{Ind}(\mathcal{A})$ and each $\mathrm{B} \in \Sigma$ such that $\mathrm{B}(\mathsf{a}) \notin \mathcal{A}$.

Adding such a new individual aux may cause axioms with shape $\top \sqsubseteq \mathrm{A}$ from $\mathcal{T}$ to trigger on aux and mess around. To prevent this, we replace each axiom $\top \sqsubseteq \mathrm{A}$ from $\mathcal{T}$ by $\mathrm{Aux}_\top \sqsubseteq \mathrm{A}$, we also add the axiom $\mathrm{A} \sqsubseteq \mathrm{Aux}_\top$ for each $\mathrm{A} \in \mathsf{sig}(\mathcal{T})$ and the assertion $\mathrm{Aux}_\top(\mathsf{a})$ for each $\mathsf{a} \in \mathsf{Ind}(\mathcal{A})$.

To summarize, we built $\mathcal{T}'$ and $\mathcal{A}'$ as:

$$\begin{aligned}
\mathcal{T}' := (&\mathcal{T} \setminus \{\top \sqsubseteq \mathrm{A} \mid \top \sqsubseteq \mathrm{A} \in \mathcal{T}\}) \\
&\cup \{\mathrm{Aux}_\top \sqsubseteq \mathrm{A} \mid \top \sqsubseteq \mathrm{A} \in \mathcal{T}\} \\
&\cup \{\mathrm{A} \sqsubseteq \mathrm{Aux}_\top \mid \mathrm{A} \in \mathsf{sig}(\mathcal{T})\} \\
&\cup \{\mathrm{B} \sqsubseteq \mathrm{Goal} \mid \mathrm{B} \in \Sigma\} \\
&\cup \{\exists \mathrm{R_B}.\mathrm{B} \sqsubseteq \mathrm{Goal} \mid \mathrm{B} \in \Sigma\}
\end{aligned}
\quad \text{and} \quad
\begin{aligned}
\mathcal{A}' := \mathcal{A} &\cup \{\mathrm{Aux}(\mathsf{a}) \mid \mathsf{a} \in \mathsf{Ind}(\mathcal{A})\} \\
&\cup \{\mathrm{R_B}(\mathsf{aux}, \mathsf{a}) \mid \mathrm{B}(\mathsf{a}) \notin \mathcal{A}, \mathrm{B} \in \Sigma\} \\
&\cup \{\mathrm{Goal}(\mathsf{a}) \mid \mathrm{B}(\mathsf{a}) \in \mathcal{A}, \mathrm{B} \in \Sigma\}
\end{aligned}$$

Set $n := |\{\mathrm{Goal}(\mathsf{a}) \mid \mathrm{B}(\mathsf{a}) \in \mathcal{A}, \mathrm{B} \in \Sigma\}|$ the amount of ABox matches for q_{Goal} in $(\mathcal{T}', \mathcal{A}')$. We now claim: $(\mathcal{T}, \Sigma, \mathcal{A})$ is satisfiable iff $n+1$ is not a certain answer for q_{Goal} over $(\mathcal{T}', \mathcal{A}')$.

($\Rightarrow$): Assume $(\mathcal{T}, \Sigma, \mathcal{A})$ is satisfiable and let $\mathcal{I}$ be one of its model. We build an interpretation $\mathcal{I}'$ of $(\mathcal{T}', \mathcal{A}')$ with domain $\Delta^{\mathcal{I}'} := \Delta^{\mathcal{I}} \cup \{\mathsf{aux}\}$ as follows:

$$\begin{array}{rcll}
\mathrm{A}^{\mathcal{I}'} & := & \mathrm{A}^{\mathcal{I}} & (\mathrm{A} \in \mathsf{sig}(\mathcal{T})) \\
\mathrm{Goal}^{\mathcal{I}'} & := & \{\mathrm{Goal}(\mathsf{a}) \mid \mathrm{B}(\mathsf{a}) \in \mathcal{A}, \mathrm{B} \in \Sigma\} & \\
\mathrm{Aux}_\top{}^{\mathcal{I}'} & := & \Delta^{\mathcal{I}} & \\
\mathrm{P}^{\mathcal{I}'} & := & \mathrm{P}^{\mathcal{I}} & (\mathrm{P} \in \mathsf{sig}(\mathcal{T})) \\
\mathrm{R_B}{}^{\mathcal{I}'} & := & \{\mathrm{R_B}(\mathsf{aux}, \mathsf{a}) \mid \mathrm{B}(\mathsf{a}) \notin \mathcal{A}, \mathrm{B} \in \Sigma\} &
\end{array}$$

Clearly, $\mathcal{I}'$ has exactly n matches for q_{Goal}. We verify it is a model of $(\mathcal{T}', \mathcal{A}')$, concluding this part of the proof as $\mathcal{I}'$ is a counter-model for $n+1$. All axioms from $\mathcal{T}$ are trivially satisfied as interpretations of concept and roles names from $\mathsf{sig}(\mathcal{T})$ are preserved (recall those with shape $\top \sqsubseteq \mathrm{A}$ have been removed!). Assertions in $\mathcal{A}'$ are also trivially satisfied, either by definition We check the other axioms in case:

$\mathrm{Aux}_\top \sqsubseteq \mathrm{A}$ $(\top \sqsubseteq \mathrm{A} \in \mathcal{T})$. Using $\mathcal{I}$ being a model of $\mathcal{T}$, we obtain: $\mathrm{Aux}_\top{}^{\mathcal{I}'} = \Delta^{\mathcal{I}} = \top^{\mathcal{I}} \subseteq \mathrm{A}^{\mathcal{I}} = \mathrm{A}^{\mathcal{I}'}$.

$\mathrm{A} \sqsubseteq \mathrm{Aux}_\top$ $(\mathrm{A} \in \mathsf{sig}(\mathcal{T}))$. Trivial: $\mathrm{A}^{\mathcal{I}'} = \mathrm{A}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} = \mathrm{Aux}_\top{}^{\mathcal{I}'}$.

$\mathrm{B} \sqsubseteq \mathrm{Goal}$ $(\mathrm{B} \in \Sigma)$. Let $e \in \mathrm{B}^{\mathcal{I}'}$. We have $\mathrm{B} \in \Sigma \subseteq \mathsf{sig}(\mathcal{T})$, hence by definition $e \in \mathrm{B}^{\mathcal{I}}$. Since $\mathrm{B} \in \Sigma$ and $\mathcal{I}$ is a model of $\mathcal{K}$, it follows that $\mathrm{B}(e) \in \mathcal{A}$. Hence, by definition: $e \in \mathrm{Goal}^{\mathcal{I}'}$.

$\exists \mathrm{R_B}.\mathrm{B} \sqsubseteq \mathrm{Goal}$ $(\mathrm{B} \in \Sigma)$. Let $e \in (\exists \mathrm{R_B}.\mathrm{B})^{\mathcal{I}'}$. We hence have an individual $\mathsf{a} \in \mathrm{B}^{\mathcal{I}'}$ such that $\mathrm{B}(\mathsf{a}) \notin \mathcal{A}$ (from the definition of $\mathrm{R_B}{}^{\mathcal{I}'}$). From the definition of $\mathrm{B}^{\mathcal{I}'}$, we obtain $\mathsf{a} \in \mathrm{B}^{\mathcal{I}}$, which implies, as $\mathcal{I}$ is a model of $\mathcal{K}$, that $\mathrm{B}(\mathsf{a}) \in \mathcal{A}$. Contradiction, hence $(\exists \mathrm{R_B}.\mathrm{B})^{\mathcal{I}'} = \emptyset$ and the axiom is trivially satisfied.

($\Leftarrow$): Assume $n+1$ is not a certain answer, that is we have a counter-model $\mathcal{I}$ (in which matches are exactly the n ABox matches). Consider the interpretation $\mathcal{I}'$ obtained by restricting $\mathcal{I}$ to the domain $\Delta^{\mathcal{I}'} := (\mathrm{Aux}_\top)^{\mathcal{I}}$.

Axioms from $\mathcal{A}$ are clearly satisfied in $\mathcal{I}'$ as $\mathcal{A} \subseteq \mathcal{A}'$. We verify that axioms from $\mathcal{T}$ also hold:

$\top \sqsubseteq \mathrm{A}$. In particular $\mathrm{Aux}_\top \sqsubseteq \mathrm{A} \in \mathcal{T}'$. From $\mathcal{I}$ being a model of $\mathcal{T}'$, we have $\mathrm{Aux}_\top{}^{\mathcal{I}} \subseteq \mathrm{A}^{\mathcal{I}}$. Therefore $\mathrm{A}^{\mathcal{I}'} = \mathrm{Aux}_\top{}^{\mathcal{I}} \cap \mathrm{A}^{\mathcal{I}} = \mathrm{A}^{\mathcal{I}}$, which yields: $\top^{\mathcal{I}'} = \mathrm{Aux}_\top{}^{\mathcal{I}} \subseteq \mathrm{A}^{\mathcal{I}} = \mathrm{A}^{\mathcal{I}'}$.

$\mathrm{A} \sqcap \mathrm{B} \sqsubseteq \mathrm{C}$. In particular $\mathrm{A} \sqcap \mathrm{B} \sqsubseteq \mathrm{C} \in \mathcal{T}'$. Using $\mathcal{I}'$ being a model of $\mathcal{T}'$, we obtain: $(\mathrm{A} \sqcap \mathrm{B})^{\mathcal{I}'} = \mathrm{A}^{\mathcal{I}} \cap \mathrm{B}^{\mathcal{I}} \cap \Delta^{\mathcal{I}'} \subseteq \mathrm{C}^{\mathcal{I}} \cap \Delta^{\mathcal{I}'} = \mathrm{C}^{\mathcal{I}'}$.

$\exists R.A \sqsubseteq B$. In particular $\exists R.A \sqsubseteq B \in \mathcal{T}'$. First notice that $(\exists R.A)^{\mathcal{I}'} \subseteq (\exists R.A)^{\mathcal{I}}$ since $R^{\mathcal{I}'} \subseteq R^{\mathcal{I}}$ and $A^{\mathcal{I}'} \subseteq A^{\mathcal{I}}$. Using $\mathcal{I}'$ being a model of $\mathcal{T}'$, we now obtain: $(\exists R.A)^{\mathcal{I}'} \subseteq (\exists R.A)^{\mathcal{I}} \cap \Delta^{\mathcal{I}'} \subseteq B^{\mathcal{I}} \cap \Delta^{\mathcal{I}'} = B^{\mathcal{I}'}$.

$A \sqsubseteq \exists R.B$. In particular both $A \sqsubseteq \exists R.B$ and $B \sqsubseteq \mathrm{Aux}_{\top}$ are in $\mathcal{T}'$. Let $e \in A^{\mathcal{I}'}$. In particular, $e \in A^{\mathcal{I}}$. Since $\mathcal{I}$ is a model of $\mathcal{T}'$, we have some $(e, e') \in R^{\mathcal{I}}$ with $e' \in B^{\mathcal{I}}$. Still from $\mathcal{I}$ being a model of $\mathcal{T}'$, we also have $e' \in \mathrm{Aux}_{\top}{}^{\mathcal{I}}$, and therefore $b \in \Delta^{\mathcal{I}'}$. Hence $(e, e') \in R^{\mathcal{I}} \cap \Delta^{\mathcal{I}'}$ and $e' \in B^{\mathcal{I}} \cap \Delta^{\mathcal{I}'}$, yielding $e \in (\exists R.B)^{\mathcal{I}'}$.

We now verify that no closed concept has been violated, which will conclude the proof. Let $e \in B^{\mathcal{I}'}$ for some closed concept $B \in \Sigma$. In particular we have both $B \sqsubseteq \mathrm{Goal}$ and $\exists R_B.B \sqsubseteq \mathrm{Goal}$ in $\mathcal{T}'$. By definition of $B^{\mathcal{I}'}$ and from $\mathcal{I}$ being a model of $\mathcal{T}'$, we obtain $e \in B^{\mathcal{I}} \subseteq \mathrm{Goal}^{\mathcal{I}}$.

From $\mathcal{I}$ being a counter-model for $n+1$, we know that $\mathrm{Goal}^{\mathcal{I}} = \{\mathrm{Goal}(\mathsf{a}) \mid B(\mathsf{a}) \in \mathcal{A}, B \in \Sigma\}$. In particular $\mathsf{aux} \notin \mathrm{Goal}^{\mathcal{I}}$. But since $\mathcal{I}$ is a model of $\mathcal{T}'$, it ensures that $\mathsf{aux} \notin (\exists R_B.B)^{\mathcal{I}}$. Recall we have $R_B(\mathsf{aux}, \mathsf{b}) \in \mathcal{A}'$ for all individuals b such that $B(\mathsf{b}) \notin \mathcal{A}$, and therefore $\mathsf{b} \notin B^{\mathcal{I}}$ for such individuals. Necessarily, it gives $B(e) \in \mathcal{A}$.

□

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

Proof. We reduce the complement of the graph 3-colorability problem to answering the $\mathcal{EL}$ OMQ $(q, \mathcal{T})$, with $q = \exists z\ B(z)$ and $\mathcal{T}$ containing $A \sqsubseteq \exists R.B$ and $\exists R.C_k \sqcap \exists E.(\exists R.C_k) \sqsubseteq B$ for $k \in \{1, 2, 3\}$.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph, and consider now the ABox given by:

$$\mathcal{A} := \{A(\mathsf{v}) \mid v \in \mathcal{V}\} \cup \{E(\mathsf{v}_1, \mathsf{v}_2) \mid \{v_1, v_2\} \in \mathcal{E}\} \cup \{C_1(\mathsf{c}_1), C_2(\mathsf{c}_2), C_3(\mathsf{c}_3), B(\mathsf{c}_1), B(\mathsf{c}_2), B(\mathsf{c}_3)\}$$

Set $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. Observe that there are 3 ABox matches: $\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3$. We claim:

$$[4, +\infty] \text{ is a certain answer of } q \text{ w.r.t. } \mathcal{K} \quad \Leftrightarrow \quad \mathcal{G} \notin \mathsf{3COL}$$

($\Leftarrow$). Assume $\mathcal{G} \notin \mathsf{3COL}$. Let $\mathcal{I}$ be a model of $\mathcal{K}$ and $f : \mathcal{C}_{\mathcal{K}} \to \mathcal{I}$ a homomorphism. We are interested in the image of elements $\mathsf{v} \cdot R.B$, with $v \in \mathcal{V}$, whose existence in $\Delta^{\mathcal{C}_{\mathcal{K}}}$ is ensured by axiom $A \sqsubseteq \exists R.B$. If there exists $v \in \mathcal{V}$ such that $f(\mathsf{v} \cdot R.B) \notin \{\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3\}$, then $f(\mathsf{v} \cdot R.B)$ provides a new match. Otherwise, define the colouring induced by $\mathcal{I}$ as $\rho_{\mathcal{I}}(v) = f(\mathsf{v} \cdot R.B) \in \{\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3\}$. Since $\mathcal{G} \notin \mathsf{3COL}$, there exists an edge $\{v_1, v_2\} \in \mathcal{E}$ with both vertices having the same colour c_k for some $k \in \{1, 2, 3\}$. For the corresponding individuals v_1 and v_2, the axiom $\exists R.C_k \sqcap \exists E.(\exists R.C_k) \sqsubseteq B$ triggers and provides two new matches: v_1 and v_2. In all cases $[4, +\infty]$ is a certain answer of q w.r.t. $\mathcal{K}$.

($\Rightarrow$). Assume $\mathcal{G} \in \mathsf{3COL}$. Consider a 3-colouring $\rho : \mathcal{V} \to \{\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3\}$. Consider the interpretation $\mathcal{I}_\rho$ obtained from $\mathcal{K}$ in which we add facts $R(\mathsf{v}, \rho(v))$ for each $v \in \mathcal{V}$, complying with the axiom $A \sqsubseteq \exists R.B$. By definition of ρ, there is no monochromatic edge, which ensures the three other axioms don't trigger on individuals v. This interpretation $\mathcal{I}_\rho$ is hence a model. It only has 3 matches, hence $[4, +\infty]$ is not a certain of q w.r.t. $\mathcal{K}$. □

B.2 Proofs for Section 4.2 (Results for DL-Lite)

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

Proof. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an oriented graph and s, t two vertices from $\mathcal{V}$. For each vertex $v \in \mathcal{V}$, we introduce a concept name V. Consider the KB given by $\mathcal{A} := \{S(\mathsf{a})\}$ and $\mathcal{T} := \{V_1 \sqsubseteq V_2 \mid (v_1, v_2) \in \mathcal{E}\}$. We are interested in the concept cardinality query $q_T := \exists z\ T(z)$.

It is now straightforward that 1 is a certain answer to q_T over $(\mathcal{T}, \mathcal{A})$ iff t is reachable from s in $\mathcal{G}$. □

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

Proof. Consider the role cardinality query $\exists z_1, z_2 P(z_1, z_2)$, and define the sets $\mathcal{D}^+_{\mathcal{K}} = \{\mathsf{a} \mid \mathsf{a}P \in \Delta^{\mathcal{C}_{\mathcal{K}}}\}$ and $\mathcal{D}^-_{\mathcal{K}} = \{\mathsf{a} \mid \mathsf{a}P^- \in \Delta^{\mathcal{C}_{\mathcal{K}}}\}$ of positive and negative demanding individuals. We assume w.l.o.g. that $|\mathcal{D}^+_{\mathcal{K}}| \leq |\mathcal{D}^-_{\mathcal{K}}|$. Let $\mathsf{p} : \mathcal{D}^+_{\mathcal{K}} \to \mathcal{D}^-_{\mathcal{K}}$ be an injection.

We partition the generated roles (*i.e.*, the roles such that there is $wT \in \Delta^{\mathcal{C}_{\mathcal{K}}}$) in four categories:

1. $\mathcal{T} \models \exists T^- \sqsubseteq \exists P$ and $\mathcal{T} \models \exists T^- \sqsubseteq \exists P^-$
2. $\mathcal{T} \models \exists T^- \sqsubseteq \exists P$ and $\mathcal{T} \not\models \exists T^- \sqsubseteq \exists P^-$
3. $\mathcal{T} \not\models \exists T^- \sqsubseteq \exists P$ and $\mathcal{T} \models \exists T^- \sqsubseteq \exists P^-$
4. $\mathcal{T} \not\models \exists T^- \sqsubseteq \exists P$ and $\mathcal{T} \not\models \exists T^- \sqsubseteq \exists P^-$

The roles in the first three cases are called demanding, and we need to consider which P-edges can be used for them.

We use the term *non-paired critical individual* to designate an individual belonging to $\mathcal{D}^+_{\mathcal{K}} \cup \mathcal{D}^-_{\mathcal{K}}$ but not to the domain of p. We then define what constitutes a solution to a demanding role:

- A solution to a case-1 demanding role is either a non-paired critical individual, or an individual a such that $\mathcal{A}, \mathcal{T} \models \exists x \mathrm{P}(\mathsf{a}, x)$ and $\mathcal{A}, \mathcal{T} \models \exists x \mathrm{P}(x, \mathsf{a})$.
- A solution to a case-2 demanding role is either a non-paired critical individual, or an individual a such that $\mathcal{A}, \mathcal{T} \models \exists x \mathrm{P}(\mathsf{a}, x)$.
- A solution to a case-3 demanding role is either a non-paired critical individual, or an individual a such that $\mathcal{A}, \mathcal{T} \models \exists x \mathrm{P}(x, \mathsf{a})$.

If a demanding role T has a solution, we let $\mathsf{sol}(\mathrm{T})$ be (an arbitrarily chosen) solution.

If all demanding roles have a solution, then the optimal number of matches is $n_{\mathcal{A}} + \mathsf{max}(|\mathcal{D}^+_{\mathcal{K}}|, |\mathcal{D}^-_{\mathcal{K}}|)$, as witnessed by the model $f(\mathcal{C}_{\mathcal{K}})$, which is the image of $\mathcal{C}_{\mathcal{K}}$ under the following *partial* function f:

- $f(\mathsf{a}) = \mathsf{a}$;
- $f(\mathsf{a}\mathrm{P}) = \mathsf{p}(\mathsf{a})$;
- $f(\mathsf{a}\mathrm{P}^-) = \mathsf{p}^{-1}(\mathsf{a})$ if defined, a otherwise;
- $f(w\mathrm{T}) = \mathsf{sol}(\mathrm{T})$ if T is neither P nor P^- and is demanding;
- $f(w\mathrm{T}) = w\mathrm{T}$ if $w\mathrm{T}$ contains no occurrence of P nor of P^- and T is not demanding.

Note that f is not defined on elements from $\mathcal{C}_{\mathcal{K}}$ with shape $\mathsf{a}w\mathrm{T}\mathrm{P}w'$ or $\mathsf{a}w\mathrm{T}\mathrm{P}^-w'$, where w is a possibly-empty word that contains neither P nor P^- and w' is a possibly-empty word. In the case of $\mathsf{a}w\mathrm{T}\mathrm{P}w'$ (the case of $\mathsf{a}w\mathrm{T}\mathrm{P}^-w'$ is similar), notice that $\mathsf{a}w\mathrm{T}$ is sent to an element $\mathsf{sol}(\mathrm{T})$, such that $\mathcal{K} \models \exists x \ \mathrm{P}(\mathsf{sol}(\mathrm{T}), x)$ by definition of a solution. Therefore the images of elements $\mathsf{a}w\mathrm{T}\mathrm{P}w'$ don't need to be specified to ensure modelhood, as the corresponding facts are already consequences of the P-edge (a, b) (if there exists b such that $(\mathsf{sol}(\mathrm{T}), \mathsf{b}) \in \mathcal{A}$) or of the P-edge $(\mathsf{a}, f(\mathsf{a}\mathrm{P}))$ (if no such b exists).

It can therefore be verified that $f(\mathcal{C}_{\mathcal{K}})$ is a model having exactly $n_{\mathcal{A}} + \mathsf{max}(|\mathcal{D}^+_{\mathcal{K}}|, |\mathcal{D}^-_{\mathcal{K}}|)$ matches.

If there is at least one demanding role that does not have a solution, then the optimal number of matches is $m_{\mathcal{A}} + \mathsf{max}(|\mathcal{D}^+_{\mathcal{K}}|, |\mathcal{D}^-_{\mathcal{K}}|) + 1$, as witnessed by the following model (which we describe by an ABox):

$$\begin{aligned}
&\mathcal{A} \cup \{\mathrm{A}(\mathsf{a}) \mid \mathcal{A}, \mathcal{T} \models \mathrm{A}\} \\
&\quad \cup \{\mathrm{P}(\mathsf{a}, \mathsf{p}(\mathsf{a})) \mid \mathsf{a} \in \mathcal{D}^+_{\mathcal{K}}\} \\
&\quad \cup \{\mathrm{R}(a, \star) \mid \mathrm{R} \neq \mathrm{P} \wedge \mathsf{a}\mathrm{R} \in \Delta^{\mathcal{C}_{\mathcal{K}}}\} \\
&\quad \cup \{\mathrm{R}(\star, \star) \mid \mathrm{R} \in \mathsf{N_R}\} \cup \{\mathrm{A}(\star) \mid \mathrm{A} \in \mathsf{N_C}\}
\end{aligned}$$

The above interpretation is indeed a model, because all elements are paired and disjointness is not expressible in DL-Lite$_{\mathsf{pos}}$. Moreover, its number of matches is $m_{\mathcal{A}} + \mathsf{max}(|\mathcal{D}^+_{\mathcal{K}}|, |\mathcal{D}^-_{\mathcal{K}}|) + 1$. This is optimal as there are at least $m_{\mathcal{A}} + |\mathcal{D}^+_{\mathcal{K}}|$ matches in any model and that there exists T a demanding role having no solution. Indeed, if T is in cases 2 or 3, there cannot be any P-edge in the ABox nor paired elements (as it would provide a solution for T), and 1 is thus the optimal as any model contains at least 1 match given by the image of the pair $(w\mathrm{T}, w\mathrm{T}\mathrm{P})$ from the canonical model (T in case 2) or of the pair $(w\mathrm{T}, w\mathrm{T}\mathrm{P}^-)$ (in case 3). Otherwise T belongs to case 1, still without a solution, which means that no individual has both an ingoing and outgoing P. Therefore, in any model, at least one of the image of the pairs $(w\mathrm{T}, w\mathrm{T}\mathrm{P})$ and $(w\mathrm{T}, w\mathrm{T}\mathrm{P}^-)$ (both exist in the canonical model, for the same w!) provides an additional match.

Note that each condition can be checked in non-deterministic logarithmic space. The number of optimal matches is thus also computable within the same bound, as the comparison with the input integer. This shows that role cardinality answering lies in NL. ☐

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

Proof. Let q_{P} be a role cardinality query. As DL-Lite$^{\mathcal{H}}_{\mathsf{core}}$ knowledge bases admits model of polynomial size in combined complexity, and that the query is atomic, there are at most polynomially many guaranteed matches. To check if $[m, +\infty]$ is a certain answer we do the following:

- if m is too big with respect to the polynomial bound, we reject
- otherwise, we guess an instance $\mathcal{A}'$ containing $\mathcal{A}$ and additionnal matches (up to m) for q_{P}. One then check whether $(\mathcal{A}', \mathcal{T}, \{\mathrm{P}\})$ is a satisfiable knowledge base with closed predicate. According to the proof of Theorem 3 of (Ngo, Ortiz, and Šimkus 2016), if it is the case, it has a model of polynomial size. We guess it, and this provides a counterexample to $[m, +\infty]$ being a certain answer.

☐

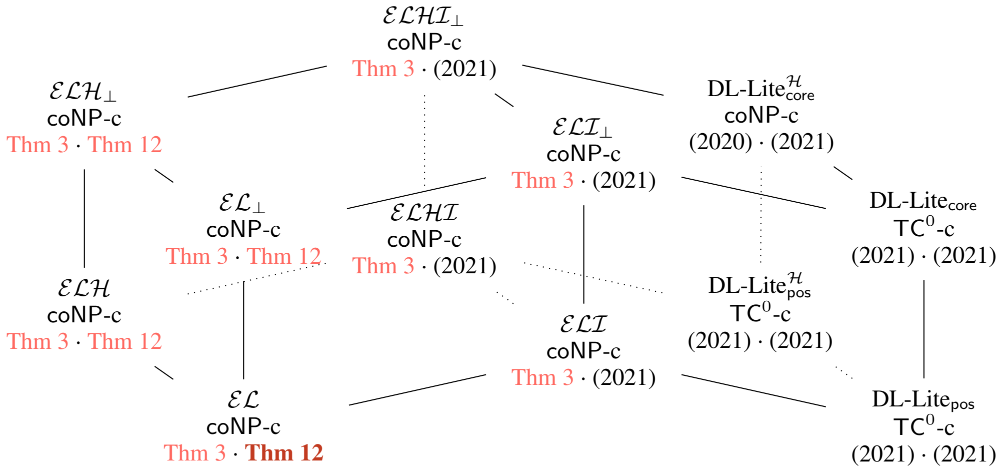


Figure 5: Concept cardinality answering: worst-case data complexity. Proofs of the upper · lower bounds are indicated. (2020) refers to (Bienvenu, Manière, and Thomazo 2020) and (2021) to (Bienvenu, Manière, and Thomazo 2021).

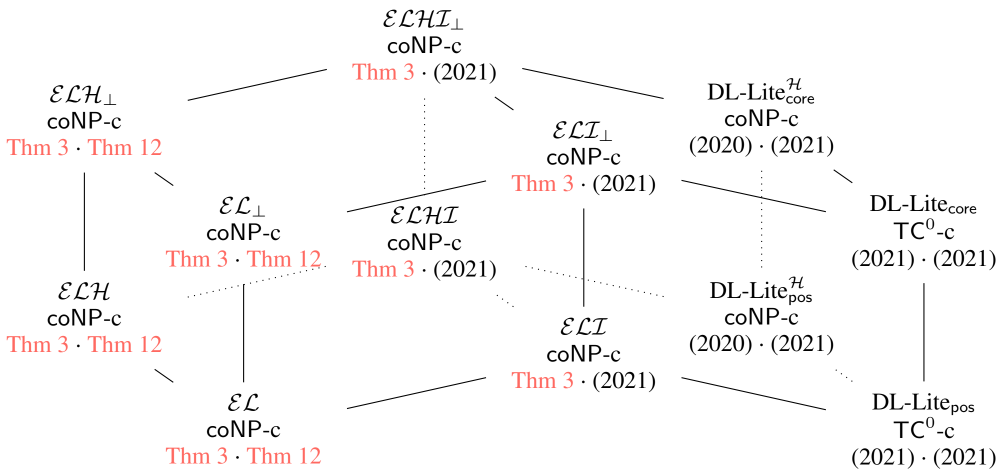


Figure 6: Role cardinality answering: worst-case data complexity. Proofs of the upper · lower bounds are indicated. (2020) refers to (Bienvenu, Manière, and Thomazo 2020) and (2021) to (Bienvenu, Manière, and Thomazo 2021).

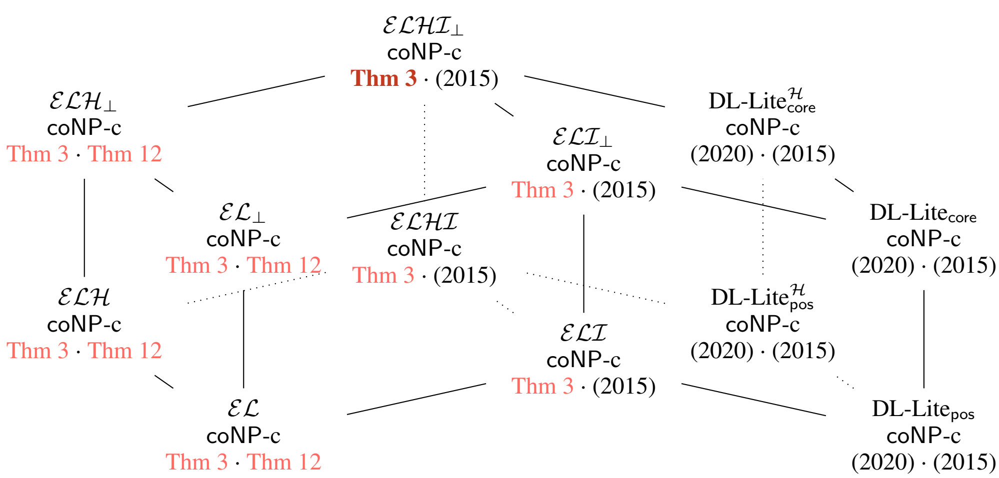


Figure 7: CCQ cardinality answering: worst-case data complexity. Proofs of the upper · lower bounds are indicated. (2015) refers to (Kostylev and Reutter 2015) and (2020) to (Bienvenu, Manière, and Thomazo 2020).

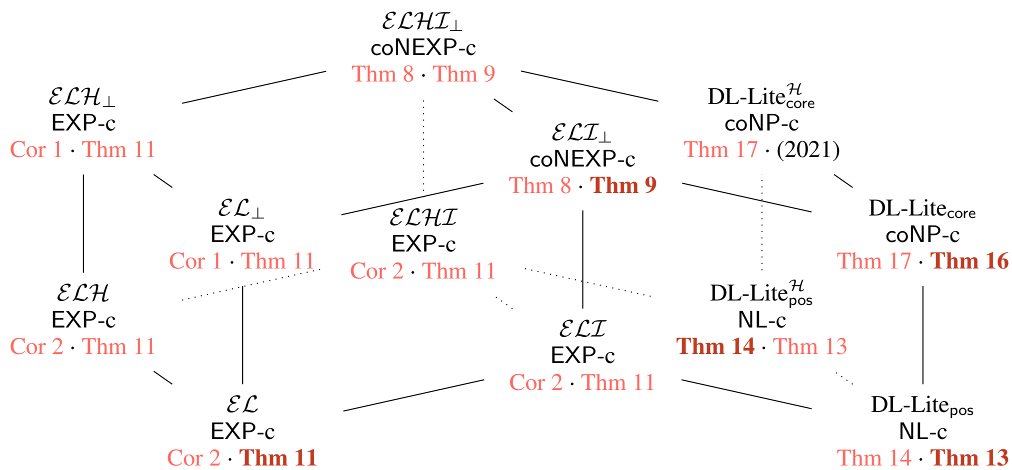


Figure 8: Concept cardinality answering: worst-case combined complexity. Proofs of the upper · lower bounds are indicated. (2021) refers to (Bienvenu, Manière, and Thomazo 2021).

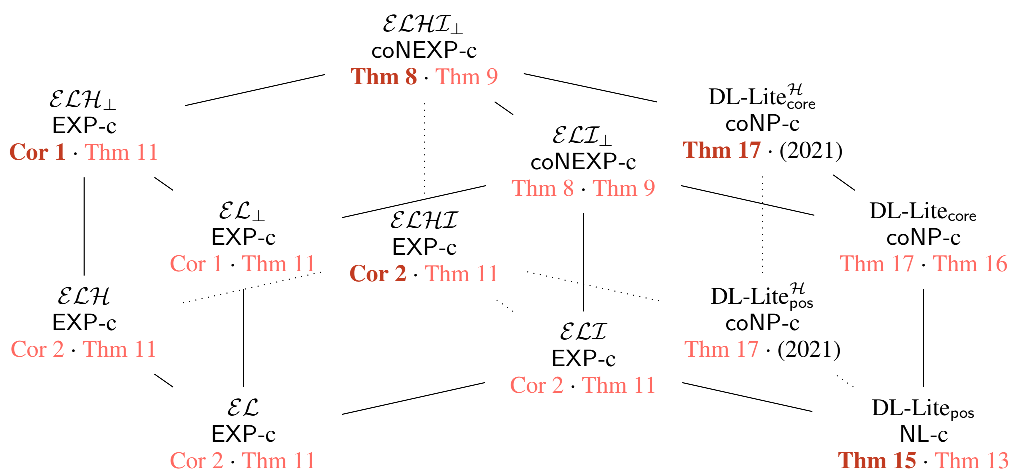


Figure 9: Role cardinality answering: worst-case combined complexity. Proofs of the upper · lower bounds are indicated. (2021) refers to (Bienvenu, Manière, and Thomazo 2021).

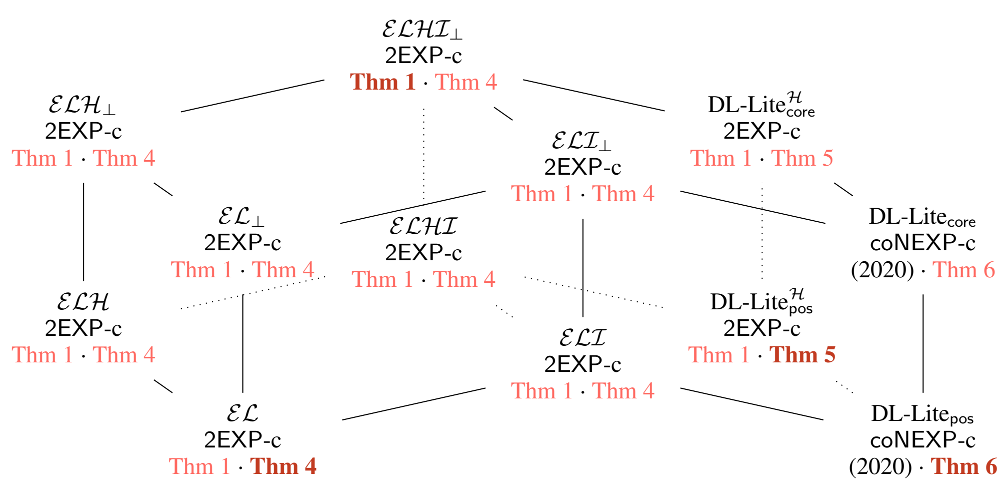


Figure 10: CCQ answering: worst-case data complexity. Proofs of the upper · lower bounds are indicated. (2020) refers to (Bienvenu, Manière, and Thomazo 2020).