# Data Complexity in the $\mathcal{EL}$ family of DLs

Adila Krisnadhi[1] and Carsten Lutz[2]

[1] Faculty of Computer Science, University of Indonesia
[2] Institute for Theoretical Computer Science, TU Dresden, Germany
adila@cs.ui.ac.id, lutz@tcs.inf.tu-dresden.de

## 1 Introduction

In recent years, lightweight description logics (DLs) have experienced increased interest because they admit highly efficient reasoning on large-scale ontologies. Most prominently, this is witnessed by the ongoing research on the DL-Lite and $\mathcal{EL}$ families of DLs, but see also [11, 14] for other examples. The main application of $\mathcal{EL}$ and its relatives is as an ontology language [5, 1, 3]. In particular, the DL $\mathcal{EL}^{++}$ proposed in [1] admits tractable reasoning while still providing sufficient expressive power to represent, for example, life-science ontologies. In contrast, the DL-Lite family of DLs is specifically tailored towards data intensive applications [8, 6, 7]. In such applications, instance checking and conjunctive query answering are the most relevant reasoning tasks and should thus be computationally cheap, preferably tractable. When determining the computational complexity of these task for a given DL, it is often realistic to consider *data complexity*, where the size of the input is measured only in terms of the ABox (which represents the data and tends to be large), but not in terms of the TBox and the query concept (which tend to be comparatively small). This is in contrast to *combined complexity*, where also the size of the TBox and query concept are taken into account.

The aim of this paper is to analyse the suitability of the $\mathcal{EL}$ family of DLs for data intensive applications. In particular, we analyse the data complexity of instance checking and conjunctive query answering in extensions of $\mathcal{EL}$. For the DL-Lite family, such an investigation has been carried out in [7], with complexities ranging from LOGSPACE-complete to coNP-complete. It follows from the results in [7] that, at least w.r.t. general TBoxes, we cannot expect the data complexity to be below PTIME for members of the $\mathcal{EL}$ family. The reason is that, in a crucial aspect, DL-Lite is even more lightweight than $\mathcal{EL}$: in contrast to the latter, the former does not allow for qualified existential (nor universal) restrictions and thus the interaction between different domain elements is very limited. When analyzing the data complexity of instance checking and conjunctive query answering in $\mathcal{EL}$ and its extensions, we therefore concentrate on mapping out the boundary of tractability.

We consider a wide range of extensions of $\mathcal{EL}$, and analyze the data complexity of the mentioned tasks with acyclic TBoxes and with general TBoxes. When selecting extensions of $\mathcal{EL}$, we focus on DLs for which instance checking has been proved *intractable* regarding combined complexity in [1]. We show that, in most of these extensions, instance checking is also intractable regarding data complexity. The notable

exceptions are $\mathcal{EL}$ extended with globally functional roles and $\mathcal{EL}$ extended with inverse roles. It is shown in [2] that instance checking in these DLs is EXPTIME-complete regarding combined complexity. On the other hand, it follows from results in [11] that instance checking is tractable regarding data complexity in $\mathcal{ELI}^f$, the extension of $\mathcal{EL}$ with both globally functional and inverse roles. In this paper, we extend this result to conjunctive query answering in $\mathcal{ELI}^f$ is still tractable regarding data complexity.

We recommend to the reader the papers [15, 16], which also analyze the complexity of conjunctive query answering in extensions of $\mathcal{EL}$. The results in these papers have been obtained independently of and in parallel to the results in the current paper.

## 2 Preliminaries

We use standard notation for the syntax and semantics of $\mathcal{EL}$ and its extensions, see [4]. The additional constructors we consider are atomic negation $\neg A$, disjunction $C \sqcup D$, sink restrictions $\forall r.\bot$, value restrictions $\forall r.C$, at-most restrictions $(\leq\, n\, r)$, at-least restrictions $(\geq\, n\, r)$, inverse roles $\exists r^-.C$, role negation $\exists \neg r.C$, role union $\exists r \cup s.C$, and transitive closure of roles $\exists r^+.C$. We denote extensions of $\mathcal{EL}$ in a canonical way, writing e.g. $\mathcal{EL}^{\forall r.\bot}$ for $\mathcal{EL}$ extended with sink restrictions and $\mathcal{EL}^{C \sqcup D}$ for $\mathcal{EL}$ extended with disjunction. Since we perform a very fine grained analysis, $\mathcal{EL}^{(\leq nr)}$ means the extension of $\mathcal{EL}$ with $(\leq n\, r)$ for some *fixed* $n$ (but not for some fixed $r$). We will also consider $\mathcal{EL}$ extended with *global* at-most restrictions: $\mathcal{EL}^{kf}$ denotes the version of $\mathcal{EL}$ obtained by reserving a set of *k-functional roles* that satisfy $|\{e \mid (d,e) \in r^{\mathcal{I}}\}| \leq k$ for all interpretations $\mathcal{I}$ and all $d \in \Delta^{\mathcal{I}}$. Instead of 1-functional roles, we will speak of functional roles as usual.

We will consider acyclic TBoxes which are defined in the usual way, and general TBoxes which are finite sets of concept inclusions $C \sqsubseteq D$. As usual when analyzing data complexity, we do not admit complex concepts in the ABox. Thus, ABoxes are sets of assertions $A(a)$ and $r(a,b)$, where $A$ is a concept name. Most of our results do not depend on the unique name assumption (UNA), which states that $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ for all distinct individual names $a, b$. Whenever they do, we will state explicitly whether the UNA is adopted or not. We write $\mathcal{A}, \mathcal{T} \models C(a)$ to denote that $a$ is an instance of $C$ w.r.t. $\mathcal{A}$ and $\mathcal{T}$ (defined in the usual way). Also, we use $\mathsf{Ind}(\mathcal{A})$ to denote the set of individual names occurring in $\mathcal{A}$.

Since conjunctive query answering is not a decision problem, we will study *conjunctive query entailment* instead. For us, a *conjunctive query* is a set $q$ of atoms $A(v)$ and $r(u,v)$, where $A$ is a concept name, $r$ a role name or an inverse role, and $u, v$ are variables. We use $\mathsf{Var}(q)$ to denote the variables used in $q$. If $\mathcal{I}$ is an interpretation and $\pi$ is a mapping from $\mathsf{Var}(q)$ to $\Delta^{\mathcal{I}}$, we write $\mathcal{I} \models^{\pi} A(v)$ if $\pi(v) \in A^{\mathcal{I}}$, $\mathcal{I} \models^{\pi} r(u,v)$ if $(\pi(u), \pi(v)) \in r^{\mathcal{I}}$, and $\mathcal{I} \models^{\pi} q$ if $\mathcal{I} \models^{\pi} \alpha$ for all $\alpha \in q$. If $\pi$ is not important, we simply write $\mathcal{I} \models q$. Finally, $\mathcal{A}, \mathcal{T} \models q$ means that for all models $\mathcal{I}$ of the ABox $\mathcal{A}$ and the TBox $\mathcal{T}$, we have $\mathcal{I} \models q$. Now, *conjunctive query entailment* is to decide given $\mathcal{A}$, $\mathcal{T}$, and $q$, whether $\mathcal{A}, \mathcal{T} \models q$. It is not hard to see that instance checking is a special case of conjunctive query entailment. Note that we do not allow individual names in conjunctive queries in place of variables. It is well-known that conjunctive query entail-

ment in which individual names are allowed in the query can be polynomially reduced to conjunctive query entailment as introduced here, see for example [9].

## 3 Lower Bounds

In [17], Schaerf proves that instance checking in $\mathcal{EL}^{\neg A}$ w.r.t. empty TBoxes is co-NP-hard regarding data complexity. He uses a reduction from a variant of SAT that he calls 2+2-SAT. Our lower bounds for extensions of $\mathcal{EL}$ are obtained by variations of Schaerf's reduction. They all apply to the case of acyclic TBoxes.

Before we start, a note on TBoxes is in order. We will usually not consider the case where there is no TBox at all because, then, ABoxes that are restricted to concept *names* are extremely inexpressive. Actually, it is not hard to show that, without TBoxes, conjunctive query containment is tractable regarding data complexity for all extensions of $\mathcal{EL}$ considered in this paper with the exception of $\mathcal{EL}^{kf}$, for which it is coNP-complete (which is proved below).

### 3.1 Basic Cases

A *2+2 clause* is of the form $(p_1 \vee p_2 \vee \neg n_1 \vee \neg n_2)$, where each of $p_1, p_2, n_1, n_2$ is a propositional letter or a truth constant 1, 0. A *2+2 formula* is a finite conjunction of 2+2 clauses. Now, 2+2-SAT is the problem of deciding whether a given 2+2 formula is satisfiable. It is shown in [17] that 2+2-SAT is NP-complete. To get started with our lower bound proofs, we repeat Schaerf's proof showing that instance checking in $\mathcal{EL}$ extended with primitive negation is co-NP-hard regarding data complexity.

Let $\varphi = c_0 \wedge \cdots \wedge c_{n-1}$ be a 2+2-formula in $m$ propositional letters $q_0, \ldots, q_{m-1}$. Let $c_i = p_{i,1} \vee p_{i,2} \vee \neg n_{i,1} \vee \neg n_{i,2}$ for all $i < n$. We use $f$, the propositional letters $q_0, \ldots, q_{m-1}$, the truth constants 1, 0, and the clauses $c_0, \ldots, c_{n-1}$ as individual names. Define the TBox $\mathcal{T}$ as $\{\overline{A} \doteq \neg A\}$ and the ABox $\mathcal{A}_\varphi$ as follows, where $c$, $p_1$, $p_2$, $n_1$, and $n_2$ are role names:

$$\begin{aligned}\mathcal{A}_\varphi := &\{A(1), \overline{A}(0), c(f, c_0), \ldots, c(f, c_{n-1})\} \\ &\cup \textstyle\bigcup_{i<n} \{p_1(c_i, p_{i,1}), p_2(c_i, p_{i,2}), n_1(c_i, n_{i,1}), n_2(c_i, n_{i,2})\}\end{aligned}$$

Models of $\mathcal{A}_\varphi$ and $\mathcal{T}$ represent truth assignments for $\varphi$ by way of setting $q_i$ to true iff $q_i \in A^{\mathcal{I}}$. Set $C := \exists c.(\exists p_1.\overline{A} \sqcap \exists p_2.\overline{A} \sqcap \exists n_1.A \sqcap \exists n_2.A)$. Intuitively, $C$ expresses that $\varphi$ is not satisfied, i.e., there is a clause in which the two positive literals and the two negative literals are all false. It is not hard to show that $\mathcal{A}_\varphi, \mathcal{T} \not\models C(f)$ iff $\varphi$ is satisfiable. Thus, instance checking in $\mathcal{EL}^{\neg A}$ w.r.t. acyclic TBoxes is co-NP-hard regarding data complexity.

This reduction can easily be adapted to $\mathcal{EL}^{\forall r.\bot}$. In all interpretations $\mathcal{I}$, $\exists r.\top$ and $\forall r.\bot$ partition the domain $\Delta^{\mathcal{I}}$ and can thus be used to simulate the concept name $A$ and its negation $\neg A$ in the original reduction. We can thus simply replace the TBox $\mathcal{T}$ with $\mathcal{T}' := \{A \doteq \exists r.\top, \overline{A} \doteq \forall r.\bot\}$.

In some extensions of $\mathcal{EL}$, we only find concepts that cover the domain, but not necessarily partition it. An example is $\mathcal{EL}^{(\leq kr)}$, $k \geq 1$, in which $\exists r.\top$ and $(\leq k\, r)$

provide a covering (for $k = 0$, observe that $(\leq k\, r)$ is equivalent to $\forall r.\bot$). Interestingly, this does not pose a problem for the reduction. In the case of $\mathcal{EL}^{(\leq kr)}$, we use the TBox $\mathcal{T} := \{A \doteq \exists r.\top, \overline{A} \doteq (\leq k\, r)\}$, and the ABox $\mathcal{A}_\varphi$ as well as the query concept $C$ remain unchanged. Let us show that $\mathcal{A}_\varphi, \mathcal{T} \not\models C(f)$ iff $\varphi$ is satisfiable. For the "if" direction, it is straightfoward to convert a truth assignment satisfying $\varphi$ into a model $\mathcal{I}$ of $\mathcal{A}_\varphi$ and $\mathcal{T}$ such that $f \notin C^\mathcal{I}$. For the "only if" direction, let $\mathcal{I}$ be a model of $\mathcal{A}_\varphi$ and $\mathcal{T}$ such that $f \notin C^\mathcal{I}$. Define a truth assignment $t$ by choosing for each propositional letter $q_i$, a truth value $t(q_i)$ such that $t(q_i) = 1$ implies $q_i^\mathcal{I} \in A$ and $t(q_i) = 0$ implies $q_i^\mathcal{I} \in \overline{A}$. Such a truth assignment exists since $A$ and $\overline{A}$ cover the domain. However, it is not necessarily unique since $A$ and $\overline{A}$ need not be disjoint. To show that $t$ satisfies $\varphi$, assume that it does not. Then there is a clause $c_i = (p_{i,1} \vee p_{i,2} \vee \neg n_{i,1} \vee \neg n_{i,2})$ that is not satisfied by $t$. By definition of $t$ and $\mathcal{A}_\varphi$, it is not hard to show that $c_i^\mathcal{I} \in (\exists p_1.\overline{A} \sqcap \exists p_2.\overline{A} \sqcap \exists n_1.A \sqcap \exists n_2.A)^\mathcal{I}$ and thus $f \in C^\mathcal{I}$, which is a contradiction.

The cases $\mathcal{EL}^{\forall r.C}$ and $\mathcal{EL}^{\exists\neg r.C}$ can be treated similarly because a covering of the domain can be achieved by choosing the concepts $\exists r.\top$ and $\forall r.X$ in the case of $\mathcal{EL}^{\forall r.C}$, and $\exists r.\top$ and $\exists\neg r.\top$ in the case of $\mathcal{EL}^{\exists\neg r.C}$. In the case, $\mathcal{EL}^{C\sqcup D}$, we use a TBox $\mathcal{T}' := \{V \doteq X \sqcup Y\}$. In all models of $\mathcal{T}'$, the extension of $V$ is covered by the concepts $X$ and $Y$. Thus, we can use the above ABox $\mathcal{A}_\varphi$, add $V(q_i)$ for all $i < m$, and use the TBox $\mathcal{T} := \mathcal{T}' \cup \{A \doteq X, \overline{A} \doteq Y\}$ and the same query concept $C$ as above. The case $\mathcal{EL}^{\exists r^+.C}$ is quite similar. In all models of the TBox $\mathcal{T}' := \{V \doteq \exists r^+.C\}$, the extension of $V$ is covered by the concepts $\exists r.C$ and $\exists r.\exists r^+.C$. Thus, we can use the same ABox and query concept as for $\mathcal{EL}^{C\sqcup D}$, together with the TBox $\mathcal{T} := \mathcal{T}' \cup \{A \doteq \exists r.C, \overline{A} \doteq \exists r.\exists r^+.C\}$.

**Theorem 1.** *For the following, instance checking w.r.t. acyclic TBoxes is co-*NP*-hard regarding data complexity: $\mathcal{EL}^{\neg A}$, $\mathcal{EL}^{\forall r.\bot}$, $\mathcal{EL}^{\forall r.C}$, $\mathcal{EL}^{\exists\neg r.C}$, $\mathcal{EL}^{C\sqcup D}$, $\mathcal{EL}^{\exists r^+.C}$, and $\mathcal{EL}^{(\leq kr)}$ for all $k \geq 0$.*

### 3.2 Cases that depend on the UNA

The results in the previous subsection are independent of whether or not the UNA is adopted. In the following, we consider some cases that depend on the (non-)UNA, starting with $\mathcal{EL}^{(\geq k\, r)}$.

In $\mathcal{EL}^{(\geq k\, r)}$, $k \geq 2$, it does not seem possible to find two concepts that a priori cover the domain and can be used to represent truth values in truth assignments. However, if we add slightly more structure to the ABox, such concepts can be found. We treat only the case $k = 3$ explicitly, but it easily generalizes to other values of $k$. Consider the ABox

$$\mathcal{A} = \{r(a, b_1), r(a, b_2), r(a, b_3), r(b_1, b_2), r(b_2, b_3), r(b_1, b_3)\}.$$

Without the UNA, there are two cases for models of $\mathcal{A}$: either two of $b_1$, $b_2$, $b_3$ identify the same domain element or they do not. In the first case, $a$ satisfies $\exists r^4.\top$, where $\exists r^4$ denotes the four-fold nesting of $\exists r$. In the second case, $a$ satisfies $(\geq 3\, r)$. It follows that we can reduce satisfiability of 2+2 formulas using a reduction very similar to the one for $\mathcal{EL}^{(\neg A)}$. The main differences are that (i) a copy of $\mathcal{A}$ is plugged in for each $q_i$, with $a$ replaced by $q_i$ and (ii) we use the TBox $\mathcal{T} := \{A \doteq \exists r^4.\top, \overline{A} \doteq (\geq 3\, r)\}$.

Unlike the previous results, this lower bound clearly depends on the fact that the UNA is not adopted. We leave it as an open problem whether instance checking in $\mathcal{EL}^{(\geq k\, r)}$ w.r.t. acyclic TBoxes is tractable if the UNA is adopted. In the following, we show that instance checking becomes coNP-hard under the UNA if we admit general TBoxes. Again, we only treat the case $k = 3$ explicitly. Define a TBox

$$\begin{aligned}\mathcal{T} := \{ \qquad & V \sqsubseteq \exists r.X \sqcap \exists r.Y \sqcap \exists r.Z \\ & (\geq 3\, r) \sqsubseteq A \\ & \exists r.(X \sqcap Y) \sqsubseteq \overline{A} \quad \exists r.(X \sqcap Z) \sqsubseteq \overline{A} \quad \exists r.(Y \sqcap Z) \sqsubseteq \overline{A}\ \}.\end{aligned}$$

In models of $\mathcal{T}$, the extension of $V$ is covered by $A$ and $\overline{A}$. Therefore, we can adapt the reduction by using the reduction ABox defined for $\mathcal{EL}^{C \sqcup D}$.

**Theorem 2.** *For $\mathcal{EL}^{(\geq k\, r)}$ with $k \geq 2$, instance checking is co-*NP*-hard in the following cases: (i) w.r.t. acyclic TBoxes and without UNA and (ii) w.r.t. general TBoxes and with (or without) UNA.*

Another case that depends on the (non-)UNA is $\mathcal{EL}^{kf}$ with $k \geq 2$. We can prove coNP-hardness provided that the UNA is not adopted. For the case $\mathcal{EL}^{1f}$, we will prove in Section 4 that instance checking (and even conjunctive query entailment) is tractable regarding data complexity, with or without the UNA. For simplicity, we only treat the case $\mathcal{EL}^{2f}$ explicitly. It is easy to generalize our argument to larger values of $k$. Like in $\mathcal{EL}^{(\geq 3r)}$, in $\mathcal{EL}^{2f}$ it does not seem possible to find two concepts that cover the domain without providing additional structure via an ABox. Set

$$\mathcal{A}' = \{r(a, b_1), r(a, b_2), r(a, b_3), r(b_1, b_2), A(b_1), A(b_2), B(b_3)\}\}.$$

where $r$ is 2-functional and thus at least two of $b_1, b_2, b_3$ have to identify the same domain element. We can distinguish two cases: either $b_3$ is identified with $b_1$ or $b_2$, then $a$ satisfies $\exists r.(A \sqcap B)$. Or $b_1$ and $b_2$ are identified, then $a$ satisfies $\exists r^3.\top$, where $\exists r^3$ denotes the three-fold nesting of $\exists r$. It follows that we can reduce satisfiability of 2+2 formulas using a reduction very similar to that for $\mathcal{EL}^{(\geq 3r)}$ above. Interestingly, we do not need a TBox at all to make this work. We take the original ABox $\mathcal{A}_\varphi$ defined for $\mathcal{EL}^{\neg A}$, add a copy of $\mathcal{A}'$ for each $q_i$ with $a$ replaced by $q_i$, and replace $A(1)$ with $\{r(1, e), A(e), B(e)\}$ and $\overline{A}(0)$ with $\{r(0, e_0), r(e_0, e_1), r(e_1, e_2)\}$. Thus, 1 satisfies $\exists r.(A \sqcap B)$ (representing true) and 0 satisfies $\exists r^3.\top$ (representing false). It remains to modify the query concept to $C' := \exists c.(\exists p_1.\exists r^3.\top \sqcap \exists p_2.\exists r^3.\top \sqcap \exists n_1.\exists r.(A \sqcap B) \sqcap \exists n_2.\exists r.(A \sqcap B))$.

With the UNA and without TBoxes, instance checking in $\mathcal{EL}^{kf}$, $k \geq 2$ is tractable regarding data complexity. The same holds for conjunctive query answering. In a nutshell, a polytime algorithm is obtained by considering the input ABox as a (complete) description of an interpretation and then checking all possible matches of the conjunctive query. A special case that has to be taken into account are inconsistent ABoxes such as those containing $\{r(a, b_1), r(a, b_2), r(a, b_3)\}$ for a 2-functional role $r$ and with the $b_i$ mutually distinct. Such inconsistencies are easily detected. If found, the algorithm returns "yes" because an inconsistent ABox entails every consequence.

If we add TBoxes, instance checking in $\mathcal{EL}^{kf}$, $k \geq 2$ becomes co-NP-hard also with the UNA. We only treat the case $k = 3$, but our arguments generalize. As in the case of $\mathcal{EL}^{2f}$ without UNA, we have to give additional structure to the ABox. Consider the TBox $\mathcal{T}' = \{V \doteq \exists r.B\}$ and the ABox

$$\mathcal{A} = \{V(a), r(a, b_1), r(a, b_2), r(a, b_3), s(a, b_1), s'(a, b_2), s'(a, b_3)\}.$$

with $r$ a 3-functional role. Then $a$ satisfies $\exists r.B$ in all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$. Because of the UNA, we can distinguish two cases: either $b_1$ satisfies $B$ or one of $b_2, b_3$ does. In the first case, $a$ satisfies $\exists s.A$ and in the second case, it satisfies $\exists s'.A$. We can now continue the reduction as in the previous cases. Start with the ABox $\mathcal{A}_\varphi$ from the reduction for $\mathcal{EL}^{\neg A}$ and add $V(q_i)$ for all $i < m$. Then use the TBox $\mathcal{T} = \mathcal{T}' \cup \{A \doteq \exists s.A, \overline{A} \doteq \exists s'.A\}$ and the original query concept $C$.

**Theorem 3.** *For $\mathcal{EL}^{kf}$ with $k \geq 2$, instance checking is*

- *tractable w.r.t. the empty TBox and with UNA;*
- *co-*NP*-hard in the following cases: (i) w.r.t. the empty TBox and without UNA, and (ii) w.r.t. acyclic TBoxes and with UNA.*

## 4 Upper Bound

We consider the extension $\mathcal{ELI}^f$ of $\mathcal{EL}$ with inverse roles and globally functional roles. If any of these two is added to $\mathcal{EL}$, instance checking w.r.t. general TBoxes becomes EXPTIME-complete regarding combined complexity [1]. However, it follows from the results on Horn-$\mathcal{SHIQ}$ in [11] that instance checking in $\mathcal{ELI}^f$ w.r.t. general TBoxes is tractable regarding data complexity. A direct proof can be found in [12]. Here, we show that even conjunctive query answering in $\mathcal{ELI}^f$ is tractable regarding data complexity.

In $\mathcal{ELI}^f$, roles and also their inverses can be declared functional using statements $\top \sqsubseteq (\leq 1\, r)$ in the TBox. For conveniently dealing with inverse roles, we use the following convention: if $r = s^-$ (with $s$ a role name), then $r^-$ denotes $s$.

As a preliminary, we assume that TBoxes and ABoxes are in a certain normal form, which we introduce next. For TBoxes, we assume that all concept inclusions are of one of the following forms, where $A$, $A_1$, $A_2$, and $B$ are concept names or $\top$ and $r$ is a role name or an inverse role:

$$\begin{array}{rrr} A \sqsubseteq B, & A \sqsubseteq \exists r.B, & \top \sqsubseteq (\leq 1\, r) \\ A_1 \sqcap A_2 \sqsubseteq B, & \exists r.A \sqsubseteq B & \end{array}$$

The normal form for ABoxes simply requires that $r(a, b) \in \mathcal{A}$ iff $r^-(b, a) \in \mathcal{A}$, for all role names $r$ and individual names $a, b$.

Let $\mathcal{A}$ be an ABox and $\mathcal{T}$ a TBox. $\mathcal{T}$ can be converted into normal form $\mathcal{T}'$ in polytime, by introducing additional concept names. See [1] for more details. Converting $\mathcal{A}$ into normal form $\mathcal{A}'$ can obviously also be done in polytime. Moreover, it is not too difficult to see that for every conjunctive query $q$ not using any of the concept names that occur in $\mathcal{T}'$ but not in $\mathcal{T}$, we have $\mathcal{A}, \mathcal{T} \models q$ iff $\mathcal{A}', \mathcal{T}' \models q$.

Another (standard) assumption that we make w.l.o.g. is that conjunctive queries are connected, i.e., for all $u, v \in \mathsf{Var}(q)$, there are atoms $r(u_0, u_1), \ldots, r(u_{n-1}, u_n) \in q$, $n \geq 0$, such that $u = u_0$ and $v = u_n$. Entailment of non-connected queries is easily (and polynomially) reduced to entailment of connected queries, see e.g. [9].

Our algorithm for conjunctive query answering in $\mathcal{ELI}^f$ is based on canonical models. To introduce canonical models, we need some preliminaries. Let $\mathcal{T}$ be a TBox and $\Gamma$ a finite set of concept names. We use

$$\mathsf{sub}_\mathcal{T}(\Gamma) := \{A \in \mathsf{N}_\mathsf{C}^\mathcal{T} \mid \bigsqcap_{A' \in \Gamma} A' \sqsubseteq_\mathcal{T} A\}$$

to denote the *closure* of $\Gamma$ under subsuming concept names w.r.t. $\mathcal{T}$. For the next definition, the reader should intuitively assume that we want to make all elements of $\Gamma$ (jointly) true at a domain element in a model of $\mathcal{T}$. If $A \in \Gamma$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$, then we say that $\Gamma$ *has* $\exists r.B$*-obligation* $O$, where

$$O = \mathsf{sub}_\mathcal{T}\big(\, \{B\} \cup \{B' \in \mathsf{N}_\mathsf{C}^\mathcal{T} \mid \exists A' \in \Gamma : \exists r^-.A' \sqsubseteq B' \in \mathcal{T}\} \cup O' \,\big),$$

and $O' = \emptyset$ if $\top \sqsubseteq (\leq 1\, r) \notin \mathcal{T}$ and $O' = \{B' \in \mathsf{N}_\mathsf{C}^\mathcal{T} \mid \exists A' \in \Gamma : A' \sqsubseteq \exists r.B' \in \mathcal{T}\}$ otherwise.

Let $\mathcal{T}$ be a TBox and $\mathcal{A}$ an ABox, both in normal form, for which we want to decide conjunctive query entailment (for a yet unspecified query $q$). To define a canonical model for $\mathcal{A}$ and $\mathcal{T}$, we have to require that $\mathcal{A}$ is *admissible* w.r.t. $\mathcal{T}$. What admissibility means depends on whether or not we make the UNA: $\mathcal{A}$ is admissible w.r.t. $\mathcal{T}$ if (i) the UNA is made and $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ or (ii) the UNA is not made and $(\top \sqsubseteq (\leq 1\, r)) \in \mathcal{T}$ implies that there are no $a, b, c \in \mathsf{Ind}(\mathcal{A})$ with $r(a, b), r(a, c) \in \mathcal{A}$ and $b \neq c$.

We define a sequence of interpretations $\mathcal{I}_0, \mathcal{I}_1, \ldots$, and the canonical model for $\mathcal{A}$ and $\mathcal{T}$ will then be the limit of this sequence. To facilitate the construction, it is helpful to use domain elements that have an internal structure. An *existential* for $\mathcal{T}$ is a concept $\exists r.A$ that occurs on the right-hand side of some inclusion in $\mathcal{T}$. A *path* $p$ for $\mathcal{T}$ is a finite (possibly empty) sequence of existentials for $\mathcal{T}$. We use $\mathsf{ex}(\mathcal{T})$ to denote the set of all existentials for $\mathcal{T}$, $\mathsf{ex}(\mathcal{T})^*$ to denote the set of all paths for $\mathcal{T}$, and $\varepsilon$ to denote the empty path. All interpretations $\mathcal{I}_i$ in the above sequence will satisfy

$$\Delta^{\mathcal{I}_i} := \{\langle a, p\rangle \mid a \in \mathsf{Ind}(\mathcal{A}) \text{ and } p \in \mathsf{ex}^*(\mathcal{T})\}$$

For convenience, we use a slightly non-standard representation of interpretations when defining the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$ and canonical interpretations: the function $\cdot^\mathcal{I}$ maps every element $d \in \Delta^\mathcal{I}$ to a set of concept names $d^\mathcal{I}$ instead of every concept name $A$ to a set of elements $A^\mathcal{I}$. It is obvious how to translate back and forth between the standard representation and this one, and we will switch freely in what follows.

To start to construction of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$, define $\mathcal{I}_0$ as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_0} &:= \{\langle a, \varepsilon\rangle \mid a \in \mathsf{Ind}(\mathcal{A})\} \\
r^{\mathcal{I}_0} &:= \{(\langle a, \varepsilon\rangle, \langle b, \varepsilon\rangle) \mid r(a, b) \in \mathcal{A}\} \\
\langle a, \varepsilon\rangle^{\mathcal{I}_0} &:= \{A \in \mathsf{N}_\mathsf{C} \mid \mathcal{A}, \mathcal{T} \models A(a)\} \\
a^{\mathcal{I}_0} &:= \langle a, \varepsilon\rangle
\end{aligned}$$

Now assume that $\mathcal{I}_i$ has already been defined. We want to construct $\mathcal{I}_{i+1}$. An element $\langle a, p\rangle \in \Delta^{\mathcal{I}_i}$ is a *leaf* in $\mathcal{I}_i$ if there is no $\alpha \in \mathsf{ex}(\mathcal{T})$ such that $\langle a, p\alpha\rangle \in \Delta^{\mathcal{I}_i}$. If it exists, select a leaf $\langle a, p\rangle$ and an $\alpha = \exists r.A \in \mathsf{ex}(\mathcal{T})$ such that $\langle a, p\rangle^{\mathcal{I}_i}$ has $\alpha$-obligation $O$ and (i) $(\top \sqsubseteq (\leq 1\ r)) \notin \mathcal{T}$ or (ii) there is no $\langle b, q\rangle \in \Delta^{\mathcal{I}_i}$ with $(\langle a, pt\rangle, \langle b, q\rangle) \in r^{\mathcal{I}_i}$. Then do the following:

- add $\langle a, p\alpha\rangle$ to $\Delta^{\mathcal{I}_i}$;
- if $r$ is a role name, add $(\langle a, p\rangle, \langle a, p\alpha\rangle)$ to $r^{\mathcal{I}_i}$;
- if $r = s^-$, add $(\langle a, p\alpha\rangle, \langle a, p\rangle)$ to $s^{\mathcal{I}_i}$;
- set $\langle a, p\alpha\rangle^{\mathcal{I}_i} := O$.

The resulting interpretation is $\mathcal{I}_{i+1}$ (and $\mathcal{I}_{i+1} = \mathcal{I}_i$ if there are no $\langle a, p\rangle$ and $\alpha$ to be selected). We assume that the selected leaf $\langle a, p\rangle$ is such that the length of $p$ is minimal, and thus all obligations are eventually satisfied.

Finally, the canonical model $\mathcal{I}$ for $\mathcal{A}$ and $\mathcal{T}$ is defined by setting $\Delta^{\mathcal{I}} := \bigcup_i \Delta^{\mathcal{I}_i}$, $A^{\mathcal{I}} := \bigcup_i A^{\mathcal{I}_i}$, $r^{\mathcal{I}} := \bigcup_i r^{\mathcal{I}_i}$, and $a^{\mathcal{I}} := a^{\mathcal{I}_0}$. A proof of the following result can be found in the full version [13].

**Lemma 1.** *The canonical model $\mathcal{I}$ for $\mathcal{T}$ and $\mathcal{A}$ is a model of $\mathcal{T}$ and of $\mathcal{A}$.*

Our aim is to prove that we can verify whether $\mathcal{A}$ and $\mathcal{T}$ entail a conjunctive query $q$ by checking whether the canonical model $\mathcal{I}$ for $\mathcal{A}$ and $\mathcal{T}$ matches $q$. Key to this result is the observation that the canonical model of $\mathcal{A}$ and $\mathcal{T}$ can be homomorphically embedded into any model of $\mathcal{A}$ and $\mathcal{T}$. We first define homomorphisms and then state the relevant lemma.

Let $\mathcal{I}$ and $\mathcal{J}$ be interpretations. A function $h : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$ is a *homomorphism* from $\mathcal{I}$ to $\mathcal{J}$ if the following holds:

1. for all individual names $a$, $h(a^{\mathcal{I}}) = a^{\mathcal{J}}$;
2. for all concept names $A$ and all $d \in \Delta^{\mathcal{I}}$, $d \in A^{\mathcal{I}}$ implies $h(d) \in A^{\mathcal{J}}$;
3. for all $d, e \in \Delta^{\mathcal{I}}$ with $(d, e) \in r^{\mathcal{I}}$, $r$ a (possibly inverse) role, $(h(d), h(e)) \in r^{\mathcal{J}}$.

**Lemma 2.** *Let $\mathcal{I}$ be the canonical model for $\mathcal{A}$ and $\mathcal{T}$, and $\mathcal{J}$ a model of $\mathcal{A}$ and $\mathcal{T}$. Then there is a homomorphism $h$ from $\mathcal{I}$ to $\mathcal{J}$.*

**Proof.** Let $\mathcal{I}$ and $\mathcal{J}$ be as in the lemma. For each interpretation $\mathcal{I}_i$ in the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$ used to construct $\mathcal{I}$, we define a homomorphism $h_i$ from $\mathcal{I}_i$ to $\mathcal{J}$. The limit of the sequence $h_0, h_1, \ldots$ is then the desired homomorphism $h$ from $\mathcal{I}$ to $\mathcal{J}$. To start, define $h_0$ by setting $h_0(\langle a, \varepsilon\rangle) := a^{\mathcal{J}}$ for all individual names $a$. Clearly, $h_0$ is a homomorphism:

- Condition 1 is satisfied by construction.
- For Condition 2, let $\langle a, \varepsilon\rangle \in A^{\mathcal{I}_0}$. Then $\mathcal{A}, \mathcal{T} \models A(a)$. Since $\mathcal{J}$ is a model of $\mathcal{A}$ and $\mathcal{T}$, $h_0(\langle a, \varepsilon\rangle) = a^{\mathcal{J}} \in A^{\mathcal{J}}$.
- For Condition 3, let $(\langle a, \varepsilon\rangle, \langle b, \varepsilon\rangle) \in r^{\mathcal{I}_0}$. Then $r(a, b) \in \mathcal{A}$ and since $\mathcal{J}$ is a model of $\mathcal{A}$ and by definition of $h_0$, we have $(h_0(\langle a, \varepsilon\rangle), h_0(\langle b, \varepsilon\rangle)) \in r^{\mathcal{J}}$.

Now assume that $h_i$ has already been defined. If $\mathcal{I}_{i+1} = \mathcal{I}_i$, then $h_{i+1} = h_i$. Otherwise, there is a unique $\langle a, p\rangle \in \Delta^{\mathcal{I}_{i+1}} \setminus \Delta^{\mathcal{I}_i}$. Let $p = q\alpha$. Then $\langle a, q\rangle \in \Delta^{\mathcal{I}_i}$, and there is an $\alpha = \exists r.B$-obligation $O$ of $\langle a, q\rangle^{\mathcal{I}_i}$ such that $\langle a, p\rangle^{\mathcal{I}_{i+1}} = \mathsf{sub}_{\mathcal{T}}(O)$. Let $A \in \langle a, q\rangle^{\mathcal{I}_i}$ such that $A \sqsubseteq \exists r.B \in \mathcal{T}$. By Condition 2 of homomorphisms, we have $d = h_i(\langle a, q\rangle) \in A^{\mathcal{J}}$. Since $A \sqsubseteq \exists r.B \in \mathcal{T}$, there is an $e \in B^{\mathcal{J}}$ with $(d, e) \in r^{\mathcal{J}}$. Define $h_{i+1}$ as the extension of $h_i$ with $h_{i+1}(\langle a, p\rangle) := e$. We prove that the three conditions of homomorphisms are preserved:

- Condition 1 is untouched by the extension.
- For Condition 2, let $\langle a, p\rangle \in A'^{\mathcal{I}_{i+1}}$. By definition of obligations, we have that $\exists r^-.\bigsqcap_{B' \in \langle a,q\rangle^{\mathcal{I}_0}} B' \sqsubseteq_{\mathcal{T}} \mathsf{sub}_{\mathcal{T}}(O)$. Since $h_i(\langle a, q\rangle) = d$ and by Condition 2 of homomorphisms, $d \in (\bigsqcap_{B' \in \langle a,q\rangle^{\mathcal{I}_0}} B')^{\mathcal{J}}$. Since $(d, e) \in r^{\mathcal{J}}$, we thus have $e \in (\bigsqcap_{B' \in \mathsf{sub}_{\mathcal{T}}(O)} B')^{\mathcal{J}}$ and it remains to remind that $A' \in \langle a, p\rangle^{\mathcal{I}_{i+1}} = \mathsf{sub}_{\mathcal{T}}(O)$.
- Condition 3 was satisfied by $\mathcal{I}_i$ and is preserved by the extension to $\mathcal{I}_{i+1}$. ❑

**Lemma 3.** *Let $\mathcal{I}$ be the canonical model for $\mathcal{A}$ and $\mathcal{T}$, and $q$ a conjunctive query. Then $\mathcal{A}, \mathcal{T} \models q$ iff $\mathcal{I} \models q$.*

**Proof.** Let $\mathcal{I}$ and $q$ be as in the lemma. If $\mathcal{I} \not\models q$, then $\mathcal{A}, \mathcal{T} \not\models q$ since, by Lemma 1, $\mathcal{I}$ is a model of $\mathcal{A}$ and $\mathcal{T}$. Now assume $\mathcal{I} \models^{\pi} q$, and let $\mathcal{J}$ be a model of $\mathcal{A}$ and $\mathcal{T}$. By Lemma 2, there is a homomorphism $h$ from $\mathcal{I}$ to $\mathcal{J}$. Define $\pi' : \mathsf{Var}(q) \to \Delta^{\mathcal{J}}$ by setting $pi'(v) := h(\pi(v))$. It is easily seen that $\mathcal{J} \models^{\pi'} q$. ❑

Thus, we can decide query entailment by looking only at the canonical model. At this point, we are faced with the problem that we cannot simply construct the canonical model $\mathcal{I}$ and check whether $\mathcal{I} \models q$ since $\mathcal{I}$ is infinite. However, we can show that if $\mathcal{I} \models q$, then $\mathcal{I} \models^{\pi} q$ for some match $\pi$ that maps all variables to elements that can be reached by travelling only a bounded number of role edges from some ABox individual. Thus, it suffices to construct a sufficiently large "initial part" of $\mathcal{I}$ and check whether it matches $q$.

To make this formal, let $n$ be the size of $\mathcal{A}$, $m$ the size of $\mathcal{T}$, and $k$ the size of $q$. In the following, we use $|p|$ to denote the length of a path $p$. The *initial canonical model* $\mathcal{I}'$ for $\mathcal{A}$ and $\mathcal{T}$ is obtained from the canonical model $\mathcal{I}$ for $\mathcal{A}$ and $\mathcal{T}$ by setting

$$\begin{aligned} \Delta^{\mathcal{I}'} &:= \{\langle a, p\rangle \mid |p| \leq 2^m + k\} \\ A^{\mathcal{I}'} &:= A^{\mathcal{I}} \cap \Delta^{\mathcal{I}'} \\ r^{\mathcal{I}'} &:= r^{\mathcal{I}} \cap (\Delta^{\mathcal{I}'} \times \Delta^{\mathcal{I}'}) \\ a^{\mathcal{I}'} &:= a^{\mathcal{I}} \end{aligned}$$

**Lemma 4.** *Let $\mathcal{I}$ be the canonical model for $\mathcal{A}$ and $\mathcal{T}$, $\mathcal{I}'$ the initial canonical model, and $q$ a conjunctive query. Then $\mathcal{I} \models q$ iff $\mathcal{I}' \models q$.*

**Proof.** Let $\mathcal{I}, \mathcal{I}'$, and $q$ be as in the lemma. It is obvious that $\mathcal{I}' \models q$ implies $\mathcal{I} \models q$. For the converse direction, let $\mathcal{I} \models^{\pi} q$. First assume that there is an $a \in \mathsf{Ind}(\mathcal{A})$ and a $v \in \mathsf{Var}(q)$ such that $\pi(q) = a^{\mathcal{I}}$. Since $q$ is connected, this means that for all $v \in \mathsf{Var}(q)$, we have $\pi(v) = \langle a, p\rangle$ such that $|p| \leq k$. It follows that $\mathcal{I}' \models^{\pi} q$.

Now assume that there are no such $a$ and $v$. Then there is an $a \in \mathsf{Ind}(\mathcal{A})$ such that for all $v \in \mathsf{Var}(q)$, we have $\pi(v) = \langle a, p\rangle$, for some $p \in \mathsf{ex}^*(\mathcal{T})$. If $\pi(v) = \langle a, p\rangle$ with $|p| \leq 2^m + k$ for all $v \in \mathsf{Var}(q)$, then $\mathcal{I}' \models^\pi q$. Otherwise, there is a $v \in \mathsf{Var}(q)$ such that $\pi(v) = \langle a, p\rangle$ with $p \in \mathsf{ex}^*(\mathcal{T})$ such that $|p| > 2^m + k$. Since $q$ is connected, this implies that for *all* $v \in \mathsf{Var}(q)$, we have $\pi(v) = \langle a, p\rangle$, for some $p \in \mathsf{ex}^*(\mathcal{T})$ with $|p| > 2^m$. Once more since $q$ is connected, there is a $v_0 \in \mathsf{Var}(q)$ such that $\pi(v_0) = \langle a, p_0\rangle$ and for all $v \in \mathsf{Var}(q)$, we have $\pi(v) = \langle a, p\rangle$ with $p_0$ a prefix of $p$.

Since $|p_0| > 2^m$, we can split $p_0$ into $p_1p_2p_3$ such that $\langle a, p_1\rangle^\mathcal{I} = \langle a, p_1p_2\rangle^\mathcal{I}$, and $p_2 \neq \varepsilon$. Now, let $\pi' : \mathsf{Var}(q) \to \Delta^\mathcal{I}$ be obtained by setting $\pi'(v) := \langle a, p_1p_3p\rangle$ if $\pi(v) = \langle a, p_1p_2p_3p\rangle$. We show the following: for all $v \in \mathsf{Var}(q)$,

1. $\pi(v)^\mathcal{I} = \pi'(v)^\mathcal{I}$;
2. $\mathcal{I} \models^{\pi'} q$.

For Point 1, let $\pi(v) = \langle a, p_1p_2p_3p\rangle$. Then $\pi(v') = \langle a, p_1p_3p\rangle$. We prove by induction on the length of $p'$ that for all prefixes $p'$ of $p_3p$, $\langle a, p_1p'\rangle^\mathcal{I} = \langle a, p_1p_2p'\rangle^\mathcal{I}$. For $p' = \varepsilon$, this is true by choice of $p_1$ and $p_2$. Now assume that the claim has already been shown for $p'$, and let $\alpha \in \mathsf{ex}(\mathcal{T})$ such that $p'\alpha$ is a prefix of $p_3p$. Since $\langle a, p_1p'\rangle^\mathcal{I} = \langle a, p_1p_2p'\rangle^\mathcal{I}$, $\langle a, p_1p'\alpha\rangle^\mathcal{I}$ is the $\alpha$-obligation of $\langle a, p_1p'\rangle^\mathcal{I}$, and $\langle a, p_1p_2p'\alpha\rangle^\mathcal{I}$ is the $\alpha$-obligation of $\langle a, p_1p_2p'\rangle^\mathcal{I}$, it is readily checked that $\langle a, p_1p'\alpha\rangle^\mathcal{I} = \langle a, p_1p_2p'\alpha\rangle^\mathcal{I}$. This finishes the proof of Point 1

For Point 2, let $A(v) \in q$. By Point 1, $\mathcal{I} \models^\pi A(v)$ implies $\mathcal{I} \models^{\pi'} A(v)$. Now let $r(u, v) \in q$. Then $(\pi(u), \pi(v)) \in r^\mathcal{I}$. By construction of $\mathcal{I}$, this implies that one of the following holds:

1. $\pi(u) = \langle a, p_1p_2p_3p\rangle$ and $\pi(v) = \langle a, p_1p_2p_3p\alpha\rangle$ for some $\alpha = \exists r.B \in \mathsf{ex}(\mathcal{T})$;
2. $\pi(u) = \langle a, p_1p_2p_3p\alpha\rangle$ and $\pi(v) = \langle a, p_1p_2p_3p\rangle$ for some $\alpha = \exists r^-.B \in \mathsf{ex}(\mathcal{T})$.

In Case 1, we have $\pi'(u) = \langle a, p_1p_3p\rangle$ and $\pi(v) = \langle a, p_1p_3p\alpha\rangle$. Again by construction of $\mathcal{I}$, this means $(\pi'(u), \pi'(v)) \in r^\mathcal{I}$. Case 2 is analogous.

When applying this construction exhaustively, we eventually obtain a $\pi^*$ such that $\pi^*(v) = \langle a, p\rangle$ with $|p| \leq 2^m + k$ for all $v \in \mathsf{Var}(q)$ ❑

The initial canonical model $\mathcal{I}'$ for $\mathcal{A}$ and $\mathcal{T}$ can be constructed in time polynomial in the size of $\mathcal{A}$. In particular, (i) $\mathcal{I}_0$ can be constructed in polytime since, due to the results of [11, 12], instance checking in $\mathcal{ELI}^f$ is tractable regarding data complexity; (ii) obligations can be computed in polytime since subsumption in $\mathcal{ELI}^f$ w.r.t. general TBoxes is decidable and the required checks are independent of the size of $\mathcal{A}$; (iii) the number of elements in the initial canonical model is bounded by $\ell := n \cdot m^{2^m+k}$ and is thus independent of the size of $\mathcal{A}$.

Our algorithm for deciding entailment of a conjunctive query $q$ by a TBox $\mathcal{T}$ and ABox $\mathcal{A}$ in normal form is as follows. If the UNA is made, we first check consistency of $\mathcal{A}$ w.r.t. $\mathcal{T}$ using one of the polytime algorithms from [11, 12]. If $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$, we answer "yes". If the UNA is not made, then we convert $\mathcal{A}$ into an ABox $\mathcal{A}'$ that is admissible w.r.t. $\mathcal{T}$, and continue working with $\mathcal{A}'$. Obviously, the conversion can be done in time polynomial in the size of $\mathcal{A}$ simply by identifying ABox individuals. Both with and without UNA, at this point we have an ABox that is admissible w.r.t. $\mathcal{T}$. The

next step is to construct the initial canonical structure $\mathcal{I}'$ for $\mathcal{T}$ and $\mathcal{A}$, and then check matches of $q$ against this structure. The latter can be done in time polynomial in the size of $\mathcal{A}$: there are at most $\ell^k$ (and thus polynomially many) mappings $\tau : \mathsf{Var}(q) \to \Delta^{\mathcal{I}'}$, and each of them can be checked for being a match in polynomial time.

**Theorem 4.** *In $\mathcal{EL}\mathcal{I}^f$, conjunctive query w.r.t. general TBoxes is in P regarding data complexity.*

A matching lower bound can be taken from [7] (which relies on the presence of general TBoxes and already applies to the instance problem), and thus we obtain P-completeness.

## 5 Summary

The results of our investigation are summarized in Table 1, and in all cases they apply both to instance checking and conjunctive query entailment. The coNP upper bounds are a consequence of the results in [9]. When the UNA is not explicitly mentioned, the results hold both with and without UNA. We point out two interesting issues. First, for all of the considered extensions we were able to show tractability regarding data complexity if and only if the logic is *convex regarding instances*, i.e., $\mathcal{A}, \mathcal{T} \models C(a)$ with $C = D_0 \sqcup \cdots \sqcup D_{n-1}$ implies $\mathcal{A}, \mathcal{T} \models D_i(a)$ for some $i < n$. It would be interesting to capture this phenomenon in a general result. And second, it is interesting to point out that subtle differences such as the UNA or local versus global functionality (for the latter, see $\mathcal{EL}^{(\leq 1r)}$ vs. $\mathcal{EL}\mathcal{I}^f$) can have an impact on tractability.

As future work, it would be interesting extend our upper bound by including more operators from the tractable description logic $\mathcal{EL}^{++}$ as proposed in [1]. For a start, it is not hard to show that conjunctive query entailment in full $\mathcal{EL}^{++}$ is undecidable due to the presence of role inclusions $r_1 \circ r_2 \sqsubseteq s$. In the following, we briefly sketch the proof, which is by reduction of the problem of deciding whether the intersection of two languages defined by given context-free grammars $G_i = (N_i, T, P_i, S_i)$, $i \in \{1, 2\}$, is empty. We assume w.l.o.g. that the set of non-terminals $N_1$ and $N_2$ are disjoint. Then define a TBox

$$\mathcal{T} := \{\top \sqsubseteq \exists r_a.\top \mid a \in T\} \cup \{r_{A_1} \circ \cdots \circ r_{A_n} \sqsubseteq r_A \mid A \to A_1 \cdots A_n \in P_1 \cup P_2\}.$$

It is not too difficult to see that $L(G_1) \cap L(G_2) \neq \emptyset$ iff the conjunctive query $S_1(u, v) \wedge S_2(u, v)$ is matched by the ABox $\{\top(a)\}$ and TBox $\mathcal{T}$.

**Acknowledgement** We are grateful to Markus Krötzsch, Meng Suntisrivaraporn, and the anonymous reviewers for valuable comments on earlier versions of this paper.

## References

1. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of the 19th Int. Joint Conf. on AI (IJCAI-05)*, pages 364–369. Morgan Kaufmann, 2005.
2. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. Submitted to a Journal. 2007
3. F. Baader, C. Lutz, and B. Suntisrivaraporn. Is tractable reasoning in extensions of the description logic $\mathcal{EL}$ useful in practice? In *Proc. of the 4th Int. WS on Methods for Modalities (M4M'05)*, 2005.

| Extensions of $\mathcal{EL}$ | w.r.t. acyclic TBoxes | w.r.t. general TBoxes |
|---|---|---|
| $\mathcal{EL}^{\neg A}$ | coNP-complete [17] | coNP-complete [17] |
| $\mathcal{EL}^{C \sqcup D}$ | coNP-complete | coNP-complete |
| $\mathcal{EL}^{\forall r.\bot}$, $\mathcal{EL}^{\forall r.C}$ | coNP-complete | coNP-complete |
| $\mathcal{EL}^{(\leq k r)}$, $r \geq 0$ | coNP-complete | coNP-complete |
| $\mathcal{EL}^{kf}$, $k \geq 2$ w/o UNA | coNP-complete (even w/o TBox) | coNP-complete |
| $\mathcal{EL}^{kf}$, $k \geq 2$ with UNA | coNP-complete (in P w/o TBox) | coNP-complete |
| $\mathcal{EL}^{(\geq k r)}$, $k \geq 2$ w/o UNA | coNP-complete | coNP-complete |
| $\mathcal{EL}^{(\geq k r)}$, $k \geq 2$ with UNA | in coNP | coNP-complete |
| $\mathcal{EL}^{\exists \neg r.C}$ | coNP-hard | coNP-hard |
| $\mathcal{EL}^{\exists r \cup s.C}$ | coNP-hard | coNP-hard |
| $\mathcal{EL}^{\exists r^+.C}$ | coNP-hard | coNP-hard |
| $\mathcal{ELI}^f$ | in P | P-complete |

**Table 1.** Complexity of instance checking and conjunctive query entailment


4. F. Baader, D. L. McGuiness, D. Nardi, and P. Patel-Schneider. *The Description Logic Handbook: Theory, implementation and applications*. Cambridge University Press, 2003.
5. S. Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In *Proc. of the 16th European Conf. on AI (ECAI-2004)*, pages 298–302. IOS Press, 2004.
6. D. Calvanese, G. D. Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. DL-lite: Tractable description logics for ontologies. In *Proc. of the 20th National Conf. on AI (AAAI'05)*, pages 602–607. AAAI Press, 2005.
7. D. Calvanese, G. D. Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Data complexity of query answering in description logics. In *Proc. of the 10th Int. Conf. on KR (KR'06)*. AAAI Press, 2006.
8. D. Calvanese, G. D. Giacomo, M. Lenzerini, R. Rosati, and G. Vetere. DL-lite: Practical reasoning for rich dls. In *Proc. of the 2004 Int. WS on DLs (DL2004)*, volume 104 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2004.
9. B. Glimm and I. Horrocks and C. Lutz and U. Sattler. Conjunctive Query Answering for the Description Logic $\mathcal{SHIQ}$. In {em Proc. of the 20th Int. Joint Conf. on AI (IJCAI-07). AAAI Press, 2007.
10. G. D. Giacomo and M. Lenzerini. Boosting the correspondence between description logics and propositional dynamic logics. In *Proc. of the 12th National Conf. on AI (AAAI'94). Volume 1*, pages 205–212. AAAI Press, 1994.
11. U. Hustadt, B. Motik, and U. Sattler. Data complexity of reasoning in very expressive description logics. In *Proc. of the 19th Int. Joint Conf. on AI (IJCAI'05)*, pages 466–471. Professional Book Center, 2005.
12. A. Krisnadhi. Data complexity of instance checking in the $\mathcal{EL}$ family of description logics. Master thesis, TU Dresden, Germany, 2007.
13. A. Krisnadhi and C. Lutz Data complexity of instance checking in the $\mathcal{EL}$ family of description logics. Available from http://lat.inf.tu-dresden.de/∼clu/papers/

14. M. Krötzsch, S. Rudolf, and P. Hitzler. On the complexity of horn description logics. In *Proc. of the 2nd WS on OWL: Experiences and Directions*, number 216 in CEUR-WS (http://ceur-ws.org/), 2006.
15. M. Krötzsch and S. Rudolf. Conjunctive Queries for $\mathcal{EL}$ with Composition of Roles. In *Proc. of the 2007 Int. WS on DLs (DL2007)*. CEUR-WS.org, 2007.
16. R. Rosati. On conjunctive query answering in $\mathcal{EL}$. In *Proc. of the 2007 Int. WS on DLs (DL2007)*. CEUR-WS.org, 2007.
17. A. Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. *Journal of Intelligent Information Systems*, 2:265–278, 1993.

## A Omitted Proof of Lemma 1

**Lemma 1.** The canonical model $\mathcal{I}$ for $\mathcal{T}$ and $\mathcal{A}$ is a model of $\mathcal{T}$ and of $\mathcal{A}$.

**Proof.** By definition of $\mathcal{I}_0$ and $\mathcal{I}$ and since $\mathcal{A}$ is in normal form, the canonical model is a model of $\mathcal{A}$. To show that it is also a model of $\mathcal{T}$, we make a case distinction according to the possible forms of concept inclusions in $\mathcal{T}$:

- $A \sqsubseteq B$ and $A_1 \sqcap A_2 \sqsubseteq B$. Satisfied since for all $\langle a, p\rangle \in \Delta^{\mathcal{I}}$, we clearly have $\langle a, p\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(\langle a, p\rangle^{\mathcal{I}})$.
- $A \sqsubseteq \exists r.B$. Let $\langle a, p\rangle \in A^{\mathcal{I}}$. There are two cases.
  - ⋆ If $p = \varepsilon$, then $A \in \langle a, \varepsilon\rangle^{\mathcal{I}_0}$ and thus $\mathcal{A}, \mathcal{T} \models A(a)$. We distinguish two subcases. First assume that $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and there is a $b \in \mathsf{Ind}(\mathcal{A})$ such that $r(a, b) \in \mathcal{A}$. Then $\mathcal{A}, \mathcal{T} \models B(b)$. By construction of $\mathcal{I}_0$, we have $(\langle a, \varepsilon\rangle, \langle b, \varepsilon\rangle) \in r^{\mathcal{I}_0} \subseteq r^{\mathcal{I}}$ and $B \in \langle b, \varepsilon\rangle^{\mathcal{I}_0}$. We thus obtain $\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$ by definition of $\mathcal{I}$ and the semantics.
    For the second subcase, assume that (i) $(\top \sqsubseteq (\leq 1\ r)) \notin \mathcal{T}$ or (ii) there is no $b \in \mathsf{Ind}(\mathcal{A})$ such that $r(a, b) \in \mathcal{A}$. If (ii) is the case, then (ii') there is no $\langle b, q\rangle \in \Delta^{\mathcal{I}_0}$ with $(\langle a, \varepsilon\rangle, \langle b, q\rangle) \in r^{\mathcal{I}_0}$. Since $A \in \langle a, p\rangle^{\mathcal{I}_0}$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$, $\langle a, p\rangle^{\mathcal{I}_0}$ has $\alpha$-obligation $O$, where $\alpha = \exists r.B$. By (i) and (ii'), there is an $i > 0$ such that $(\langle a, \varepsilon\rangle, \langle a, \alpha\rangle) \in r^{\mathcal{I}_i} \subseteq r^{\mathcal{I}}$ and $\langle a, \alpha\rangle^{\mathcal{I}_i} = \mathsf{sub}_{\mathcal{T}}(O)$. Since $B \in O$, $\langle a, \alpha\rangle \in B^{\mathcal{I}_i}$. It follows that $\langle a, \varepsilon\rangle \in (\exists r.B)^{\mathcal{I}}$.
  - ⋆ Let $p \neq \varepsilon$. Then there is an $i > 0$ such that $\langle a, p\rangle \in \Delta^{\mathcal{I}_i}$. Let $i$ be minimal with this property. There are two subcases. First, assume that $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and there is a $\langle b, q\rangle \in \Delta^{\mathcal{I}_i}$ such that $(\langle a, p\rangle, \langle b, q\rangle) \in r^{\mathcal{I}_i}$. By construction of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$ and since $p \neq \varepsilon$, this can only be the case if $a = b$ and
    1. $p = q\alpha$ or for some $\alpha = \exists r^-.B' \in \mathsf{ex}(\mathcal{T})$, or
    2. $q = p\alpha$ for some $\alpha = \exists r.B' \in \mathsf{ex}(\mathcal{T})$.

    First for Case 1. Then $(\langle a, p\rangle, \langle a, q\rangle) \in r^{\mathcal{I}_i}$, $\langle a, q\rangle^{\mathcal{I}_i}$ has $\exists r^-.B'$-obligation $O$, and $\langle a, p\rangle^{\mathcal{I}_i} = \mathsf{sub}_{\mathcal{T}}(O)$. By definition of obligations, $A \in \mathsf{sub}_{\mathcal{T}}(O)$ implies that $\bigsqcap_{X \in \langle a,q\rangle^{\mathcal{I}_i}} X \sqsubseteq_{\mathcal{T}} \exists r^-.A$. Together with $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$, we get $\bigsqcap_{X \in \langle a,q\rangle^{\mathcal{I}_i}} X \sqsubseteq_{\mathcal{T}} B$. Since $\langle a, q\rangle^{\mathcal{I}_i} = \mathsf{sub}_{\mathcal{T}}(\langle a, q\rangle^{\mathcal{I}_i})$, we thus have $B \in \langle a, q\rangle^{\mathcal{I}_i}$. By the semantics, $\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$.
    Now for Case 2. Then $(\langle a, p\rangle, \langle a, q\rangle) \in r^{\mathcal{I}_i}$, $\langle a, p\rangle^{\mathcal{I}_i}$ has $\exists r.B'$-obligation $O$, and $\langle a, q\rangle^{\mathcal{I}_i} = \mathsf{sub}_{\mathcal{T}}(O)$. Since $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$, we have $B \in O$. Thus $B \in \langle a, q\rangle^{\mathcal{I}_i}$ and, $\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$.
    For the second subcase, assume that $(\top \sqsubseteq (\leq 1\ r)) \notin \mathcal{T}$ or there is no $\langle b, q\rangle \in \Delta^{\mathcal{I}_i}$ such that $(\langle a, p\rangle, \langle b, q\rangle) \in r^{\mathcal{I}_i}$. Clearly, $\langle a, p\rangle^{\mathcal{I}_i}$ has $\alpha = \exists r.B$-obligation $O$ and $B \in O$. Thus, there is a $j > i$ such that $(\langle a, p\rangle, \langle a, p\alpha\rangle) \in r^{\mathcal{I}_j}$ and $\langle a, p\alpha\rangle \in B^{\mathcal{I}_j}$. Thus, $\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$.
- $\exists r.A \sqsubseteq B$. Let $\langle a, p\rangle \in (\exists r.A)^{\mathcal{I}}$. Then there is a $\langle b, q\rangle \in A^{\mathcal{I}}$ and such that $(\langle a, p\rangle, \langle b, q\rangle) \in r^{\mathcal{I}}$. We distinguish four cases.
  - ⋆ $p = q = \varepsilon$. Then $r(a, b) \in \mathcal{A}$ and $\mathcal{A}, \mathcal{T} \models A(b)$. Thus, $\mathcal{A}, \mathcal{T} \models B(a)$ and $a \in B^{\mathcal{I}}$ by definition of $\mathcal{I}_0$.
  - ⋆ $p = \varepsilon$, $q \neq \varepsilon$. By construction of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$, this implies $a = b$ and $q = \alpha = \exists r.B' \in \mathsf{ex}(\mathcal{T})$. Also by construction, $\langle a, \varepsilon\rangle^{\mathcal{I}}$ has $\exists r.B'$-obligation $O$,

and $\langle a, \alpha\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(O)$. Since $A \in \mathsf{sub}_{\mathcal{T}}(O)$, it follows that $\bigsqcap_{X \in \langle a,\varepsilon\rangle^{\mathcal{I}}} X \sqsubseteq_{\mathcal{T}} \exists r.A$. Together with $\exists r.A \sqsubseteq B \in \mathcal{T}$, we get $\bigsqcap_{X \in \langle a,\varepsilon\rangle^{\mathcal{I}}} X \sqsubseteq_{\mathcal{T}} B$. Thus, $B \in \langle a, \varepsilon\rangle^{\mathcal{I}}$.

- ⋆ $p \neq \varepsilon$, $q \neq \varepsilon$. There are two subcases. If $q = p\alpha$ for some $\alpha = \exists r.B' \in \mathsf{ex}(\mathcal{T})$, then we can argue analogous to the previous case. Thus, we only consider the case $p = q\alpha$ for some $\alpha = \exists r^-.B' \in \mathsf{ex}(\mathcal{T})$. In this case, $\langle a, q\rangle^{\mathcal{I}}$ has $\exists r^-.B'$-obligation $O$, and $\langle a, p\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(O)$. Since $A \in \langle a, q\rangle^{\mathcal{I}}$ and $\exists r.A \sqsubseteq B$, $B \in O$. It follows that $\langle a, p\rangle \in B^{\mathcal{I}}$.
- ⋆ $p \neq \varepsilon$, $q = \varepsilon$. By construction of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$, this implies $a = b$ and $p = \alpha = \exists r^-.B' \in \mathsf{ex}(\mathcal{T})$. Also by construction, $\langle a, \varepsilon\rangle^{\mathcal{I}}$ has $\exists r^-.B'$-obligation $O$, and $\langle a, \alpha\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(O)$. Since $A \in \langle a, \varepsilon\rangle^{\mathcal{I}}$ and $\exists r.A \sqsubseteq B \in \mathcal{T}$, we have $B \in O$. Thus, $\langle a, \alpha\rangle = \langle a, p\rangle \in B^{\mathcal{I}}$.

– $\top \sqsubseteq (\leq 1\ r)$. Since $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$, there are no $a, b, c \in \mathsf{Ind}(\mathcal{A})$ with $b \neq c$ such that for some role name $r$, $r(a, b)$ and $r(a, c)$ are in $\mathcal{A}$ and $\top \sqsubseteq (\leq 1\ r) \in \mathcal{T}$. It follows that $\mathcal{I}_0$ satisfies all $\top \sqsubseteq (\leq 1\ r) \in \mathcal{T}$. This property is clearly preserved when constructing $\mathcal{I}_i$ with $i > 0$, and thus it holds for $\mathcal{I}$.

❑