Data Complexity in the $\mathcal{EL}$ family of Description Logics

Adila Krisnadhi[1] and Carsten Lutz[2]

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

Abstract. We study the data complexity of instance checking and conjunctive query answering in the $\mathcal{EL}$ family of description logics, with a particular emphasis on the boundary of tractability. We identify a large number of intractable extensions of $\mathcal{EL}$, but also show that in $\mathcal{ELI}^f$, the extension of $\mathcal{EL}$ with inverse roles and global functionality, conjunctive query answering is tractable regarding data complexity. In contrast, already instance checking in $\mathcal{EL}$ extended with only inverse roles or global functionality is EXPTIME-complete regarding combined complexity.

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 (see also [12, 16] for other examples). The main application of $\mathcal{EL}$ and its relatives is as an ontology language [6, 2, 4]. In particular, the DL $\mathcal{EL}^{++}$ proposed in [2] 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 applications with a massive amount of instance data [9, 7, 8, 1]. In such applications, instance checking and conjunctive query answering are the most relevant reasoning services and should thus be computationally cheap, preferably tractable. When determining the computational complexity of these tasks 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 instance data), but not in terms of the TBox (which corresponds to the schema) and the query, as the latter both tend to be small compared to the former. This is in contrast to *combined complexity*, where also the size of the TBox and query are taken into account.

The aim of this paper is to study the $\mathcal{EL}$ family of DLs in the light of data intensive applications. To this end, we analyze 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 e.g. in [8, 1], with complexities ranging from LOGSPACE-complete to coNP-complete. It follows from the results in [8] that we cannot expect the data complexity to be below PTIME for members of the $\mathcal{EL}$ family (at least in the presence of so-called general TBoxes, i.e., sets of GCIs). The reason is that, in a crucial aspect, DL-Lite is even more lightweight than $\mathcal{EL}$: in contrast to $\mathcal{EL}$, DL-Lite does not

Name	Syntax	Semantics
top	$\top$	$\Delta^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
atomic negation	$\neg A$	$\Delta^{\mathcal{I}} \setminus A^{\mathcal{I}}$
disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
sink restriction	$\forall r.\bot$	$\{x \mid \neg\exists y : (x,y) \in r^{\mathcal{I}}\}$
value restriction	$\forall r.C$	$\{x \mid \forall y : (x,y) \in r^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}}\}$
at-least restriction	$(\geqslant k\ r)$	$\{x \mid \#\{y \in \Delta^{\mathcal{I}} \mid (x,y) \in r^{\mathcal{I}}\} \geq k\}$
at-most restriction	$(\leqslant k\ r)$	$\{x \mid \#\{y \in \Delta^{\mathcal{I}} \mid (x,y) \in r^{\mathcal{I}}\} \leq k\}$
inverse roles	$\exists r^{-}.C$	$\{x \mid \exists y : (y,x) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
role negation	$\exists \neg r.C$	$\{x \mid \exists y \in \Delta^{\mathcal{I}} : (x,y) \notin r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
role union	$\exists r \cup s.C$	$\{x \mid \exists y \in \Delta^{\mathcal{I}} : (x,y) \in r^{\mathcal{I}} \cup s^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
transitive closure	$\exists r^{+}.C$	$\{x \mid \exists y \in \Delta^{\mathcal{I}} : (x,y) \in (r^{\mathcal{I}})^{+} \wedge y \in C^{\mathcal{I}}\}$

Table 1. Syntax and semantics of relevant DL constructors.

allow for qualified existential (neither 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 [2]. 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 [3] that instance checking in these DLs is EXPTIME-complete regarding combined complexity. On the other hand, it follows from results in [12] 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 $\mathcal{ELI}^f$, and show that this problem is still tractable regarding data complexity.

2 Preliminaries

In DLs, *concepts* are inductively defined with the help of a set of *constructors*, starting with a set $\mathsf{N_C}$ of *concept names* and a set $\mathsf{N_R}$ of *role names*. In $\mathcal{EL}$, concepts are formed using the three topmost constructors in Table 1. There and in general, we use r and s to denote role names, A and B to denote concept names, and C, D to denote concepts. The additional constructors shown in Table 1 give rise to extensions of $\mathcal{EL}$. We use

canonical names to refer to such extensions, 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, $\mathcal{EL}^{(\leq k r)}$ means the extension of $\mathcal{EL}$ with $(\leq k\, r)$ for some *fixed* $k \geq 0$ (but not for some fixed r).

In DLs, TBoxes are used to represent general knowledge about an application domain, and thus play the role of an ontology. We introduce two different forms of TBoxes. An *acyclic TBox* $\mathcal{T}$ is a finite set of concept equations $A \doteq C$ such that the left-hand sides are unique and there are no cycles, i.e., if $\{A_0 \doteq C_0, \ldots, A_{n-1} \doteq C_{n-1}\} \subseteq \mathcal{T}$ then for some $i \leq n$, A_i does not occur in C_{i+1} where $A_n := A_0$ and $C_n := C_0$. A *general TBox* is a finite set of concept inclusions $C \sqsubseteq D$ (often called *GCIs*). Every concept equation $A \doteq C$ can be written as two inclusions $A \sqsubseteq C$ and $C \sqsubseteq A$, and thus general TBoxes subsume acyclic ones. ABoxes are used to represent instance data. Let $\mathsf{N_I}$ be a set of *individual names*. An *ABox* is a finite set of expressions $A(a)$ and $r(a, b)$, where a and b are from $\mathsf{N_I}$ (here and in what follows). Observe that we disallow complex concepts in the ABox, as usual when studying data complexity.

The semantics of $\mathcal{EL}$ and its extensions is defined in terms of *interpretations* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. The *domain* $\Delta^{\mathcal{I}}$ is a non-empty set and the *interpretation function* $\cdot^{\mathcal{I}}$ maps each concept name $A \in \mathsf{N_C}$ to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$, each role name $r \in \mathsf{N_R}$ to a binary relation $r^{\mathcal{I}}$ on $\Delta^{\mathcal{I}}$, and each individual name $a \in \mathsf{N_I}$ to a domain element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. The extension of $\cdot^{\mathcal{I}}$ to complex concepts is inductively defined as shown in the third column of Table 1, where $\#S$ denotes the cardinality of the set S. An interpretation $\mathcal{I}$ *satisfies* an equation $A \doteq C$ iff $A^{\mathcal{I}} = C^{\mathcal{I}}$, an inclusion $C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, an assertion $C(a)$ iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$, and an assertion $r(a, b)$ if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$. It is a *model* of a TBox $\mathcal{T}$ (ABox $\mathcal{A}$) if it satsfies all equations/inclusions in $\mathcal{T}$ (assertions in $\mathcal{A}$).

We will also consider $\mathcal{EL}^{kf}$, the extension of $\mathcal{EL}$ with k-functional roles, i.e., roles for which every domain element can have at most k successors. In $\mathcal{EL}^{kf}$, there are no additional concept constructors that may be used to build up complex concepts. Instead, a new kind of expression $\top \sqsubseteq (\leq k\ r)$ is allowed in the TBox. These expressions can be understood as *global* at-most restrictions, in contrast to the local at-most restrictions shown in Table 1. An interpretation $\mathcal{I}$ satisfies $\top \sqsubseteq (\leq k\ r)$ if $|\{e \mid (d, e) \in r^{\mathcal{I}}\}| \leq k$ for all $d \in \Delta^{\mathcal{I}}$. Instead of 1-functional roles, we will speak of functional roles as usual.

The two main inference problems considered in this paper are instance checking and conjunctive query entailment. An individual name a is an *instance* of a concept C w.r.t. an ABox $\mathcal{A}$ and a TBox $\mathcal{T}$ (written $\mathcal{A}, \mathcal{T} \models C(a)$) iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ in all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$. The instance problem is to decide, given a, C, $\mathcal{A}$ and $\mathcal{T}$, whether $\mathcal{A}, \mathcal{T} \models C(a)$.

Conjunctive query entailment is the decision problem corresponding to conjunctive query answering, which is a search problem. A *conjunctive query* is a set q of atoms $C(v)$ and $r(u, v)$, where u, v are variables. We use $\mathsf{Var}(q)$ to denote the variables used in q. If $\mathcal{I}$ is an interpretation and π is a mapping from $\mathsf{Var}(q)$ to $\Delta^{\mathcal{I}}$, we write $\mathcal{I} \models^{\pi} C(v)$ if $\pi(v) \in C^{\mathcal{I}}$, $\mathcal{I} \models^{\pi} r(u, v)$ if $(\pi(u), \pi(v)) \in r^{\mathcal{I}}$, $\mathcal{I} \models^{\pi} q$ if $\mathcal{I} \models^{\pi} \alpha$ for all $\alpha \in q$, and $\mathcal{I} \models q$ if $\mathcal{I} \models^{\pi} q$ for some π. 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, in $\mathcal{EL}$, instance checking is a special case of conjunctive query entailment, as every $\mathcal{EL}$-concept C can be converted into a tree-shaped query.

Note that we do not partition the variables in a conjunctive query into answer variables and existentially quantified variables as usual. Since we are dealing with query entailment instead of query answering, this distinction is meaningless. Also observe that we do not allow individual names in conjunctive queries in place of variables. It is well-known that individual names in the query can be simulated by concept names with only a linear blowup of the input, see for example [10] for details.

The last preliminary worth mentioning is the *unique name assumption (UNA)*, which requires that for all $a, b \in \mathsf{N_I}$ with $a \neq b$, we have $a^{\mathcal{I}} \neq b^{\mathcal{I}}$. Most of our results do not depend on the UNA. Whenever they do, we will state explicitly whether the UNA is adopted or not.

3 Lower Bounds

We show that, in almost all extensions of $\mathcal{EL}$ introduced in Section 2, instance checking is co-NP-hard regarding data complexity. All our lower bounds assume only acyclic TBoxes.

For the sake of completeness, we note that the case where there is no TBox is not very interesting: because only concept *names* are admitted in the ABox, the additional concept constructors can then only occur in the query (which is a concept in the case of instance checking and a conjunctive query otherwise). In most cases (such as $\mathcal{EL}^{(\neg)}$ and $\mathcal{EL}^{\forall r.C}$), this means that no query which contains the additional constructor is entailed by any ABox. Thus, there is a trivial reduction to query answering in basic $\mathcal{EL}$. In other cases such as $\mathcal{EL}^{C \sqcup D}$, it is easily shown that conjunctive query containment is tractable regarding data complexity. A notable exception is $\mathcal{EL}^{kf}$, $k \geq 2$, for which instance checking is coNP-complete already without TBoxes (as is proved below).

3.1 Basic Cases

In [19], Schaerf proves that instance checking in $\mathcal{EL}^{\neg A}$ 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. For this reason, we start with repeating the original reduction of Schaerf. Before we go into detail, a remark on $\mathcal{EL}^{\neg A}$ is in order. In this extension of $\mathcal{EL}$, the application of negation is restricted to concept names. However, full negation can easily be recovered using acyclic TBoxes: instead of writing $\neg C$, we may write $\neg A$ and add a concept equation $A \doteq C$, with A a fresh concept name. Thus, we restrict the use of negation even further, namely to concept names that do not occur on the left-hand side of any concept equation in the (acyclic) TBox. As we shall see shortly, the TBoxes required for our lower bound are actually of very simple form.

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 [19] that 2+2-SAT is NP-complete.

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

$q_0, \dots, q_{m-1}$, the truth constants $1, 0$, and the clauses $c_0, \dots, c_{n-1}$ 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:

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

It should be obvious that $\mathcal{A}_\varphi$ is a straightforward representation of φ. Models of $\mathcal{A}_\varphi$ and $\mathcal{T}$ represent truth assignments for φ by way of setting q_i to true if $q_i \in A^\mathcal{I}$ and to false if $q_i \in \overline{A}^\mathcal{I}$. Since $\mathcal{I}$ is a model of $\mathcal{T}$, this truth assignment is well-defined. 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 φ 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 the following.

Lemma 1 (Schaerf). *$\mathcal{A}_\varphi, \mathcal{T} \not\models C(f)$ iff φ 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

Lemma 2. *$\mathcal{A}_\varphi, \mathcal{T} \not\models C(f)$ iff φ is satisfiable.*

Proof. "if". This direction is as in the proof of Lemma 1. Let t be a truth assignment satisfying φ. Define an interpretation $\mathcal{I}$ as follows:

$$\begin{aligned}\Delta^\mathcal{I} &:= \{f, c_0, \dots, c_{n-1}, q_0, \dots, q_{m-1}, 0, 1, d\} \\ c^\mathcal{I} &:= \{(f, c_0), \dots, (f, c_{n-1})\} \\ p_j^\mathcal{I} &:= \{(c_0, p_{0,j}), \dots, (c_{n-1}, p_{n-1,j})\} \\ n_j^\mathcal{I} &:= \{(c_0, n_{0,j}), \dots, (c_{n-1}, n_{n-1,j})\} \\ A^\mathcal{I} &:= \{1\} \cup \{q_i \mid i < m \text{ and } t(q_i) = 1\} \\ \overline{A}^\mathcal{I} &:= \Delta^\mathcal{I} \setminus A^\mathcal{I} \\ r^\mathcal{I} &:= \{(e, d) \mid e \in A^\mathcal{I}\}\end{aligned}$$

All individual names are interpreted as themselves. It is not hard to verify that $\mathcal{I}$ is a model of $\mathcal{A}_\varphi$ and $\mathcal{T}$, and that $f \notin C^\mathcal{I}$.

"only if". Here we need to deal with the non-disjointness of $\exists r.\top$ and $(\leq k\, r)$. 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 φ, 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, $p_{i,1}, p_{i,2} \in \overline{A}^{\mathcal{I}}$ and $n_{i,1}, n_{i,2} \in A^{\mathcal{I}}$. Thus $c_i^{\mathcal{I}} \in (\exists p_1.\overline{A} \sqcap \exists p_2.\overline{A} \sqcap \exists n_1.A \sqcap \exists n_2.A)^{\mathcal{I}}$ and we get $f \in C^{\mathcal{I}}$, 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 $\mathcal{EL}^{(\leq k r)}$ for all $k \geq 0$.*

For $\mathcal{EL}^{\forall r.\bot}$, $\mathcal{EL}^{\forall r.C}$, and $\mathcal{EL}^{C \sqcup D}$, co-NP-hardness of conjunctive query containment w.r.t. general TBoxes has been established in [8]. It seems likely that the proofs (which are not given in detail) actually apply to instance checking and acyclic TBoxes.

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 as long as $k \geq 2$. Consider the following auxiliary ABox, also shown in Figure 1.

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

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 $\mathcal{EL}^{(\leq k\, r)}$. 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 $\mathcal{T} := \{A \doteq \exists r^4.\top, \overline{A} \doteq (\geq 3\, r)\}$.

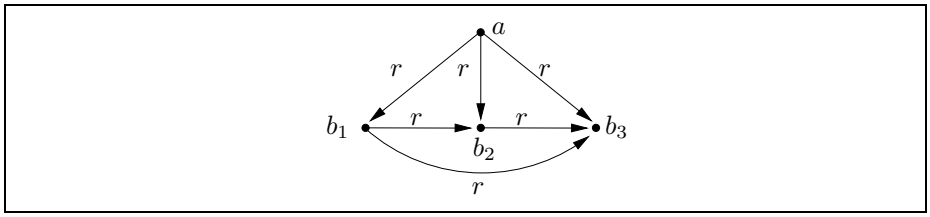

Fig. 1. Auxiliary ABox $\mathcal{A}$ for $\mathcal{EL}^{(\geq 3\, r)}$ without UNA.

Unlike the previous results, this lower bound clearly depends on the fact that the UNA is not adopted. If the UNA is adopted, we can prove the same result using a different auxiliary ABox. Again, we only treat the case $k = 3$, which easily generalizes. Let

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

and consider the TBox $\mathcal{T}' = \{V \doteq \exists r.B\}$. In every model $\mathcal{I}$ of $\mathcal{A}'$ and $\mathcal{T}'$, there is a $d \in B^{\mathcal{I}}$ such that $(a^{\mathcal{I}}, d) \in r^{\mathcal{I}}$. We can distinguish two cases: if $d = b_i$ for some $i \in \{1, 2\}$, then a satisfies $\exists r.(A \sqcap B)$. Otherwise, a satisfies $(\geq 3\ r)$. We can now continue the reduction as in the previous cases. Start with the ABox $\mathcal{A}_\varphi$ from the reduction for $\mathcal{EL}^{\neg A}$, add $V(q_i)$ for all $i < m$ and a copy of $\mathcal{A}'$ for each q_i, with a replaced by q_i. Then use the TBox $\mathcal{T} = \mathcal{T}' \cup \{A \doteq \exists r.(A \sqcap B), \overline{A} \doteq (\geq 3\, r)\}$ and the original query concept C. Observe that this reduction does not work without the UNA.

Theorem 2. *For $\mathcal{EL}^{(\geq k\, r)}$ with $k \geq 2$, instance checking w.r.t. acyclic TBoxes is co-*NP*-hard regarding data complexity, both with and without the UNA.*

Another case that depends on the (non-)UNA is $\mathcal{EL}^{kf}$ with $k \geq 2$. We start with proving 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. A graphical representation is given in Figure 2. Regarding models of $\mathcal{A}''$, 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. Observe that 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

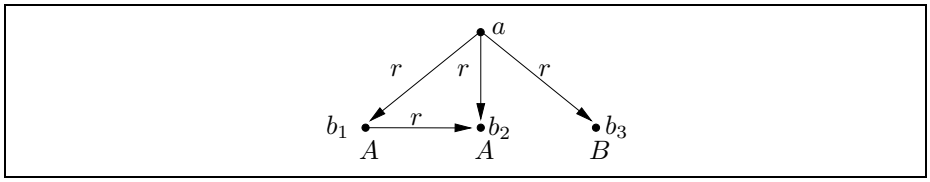

Fig. 2. Auxiliary ABox $\mathcal{A}''$ for $\mathcal{EL}^{2f}$ without UNA.

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 acyclic 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 $\exists s.B$ and in the second case, it satisfies $\exists s'.B$. We can then continue the reduction as in the previous cases.

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

The only remaining extensions of $\mathcal{EL}$ introduced in Section 2 are $\mathcal{EL}^{\exists r^-.C}$ and $\mathcal{EL}^{1f}$. For both of them, instance checking w.r.t. general TBoxes is EXPTIME-complete regarding combined complexity [2]. In this section, we consider the union $\mathcal{ELI}^f$ of $\mathcal{EL}^{\exists r^-.C}$ and $\mathcal{EL}^{1f}$, i.e., the extension of $\mathcal{EL}$ with both inverse roles and globally functional roles. It follows from the results on Horn-$\mathcal{SHIQ}$ in [12] that instance checking in $\mathcal{ELI}^f$ w.r.t. general TBoxes is tractable regarding data complexity. A direct proof

can be found in [14]. Here, we show that even conjunctive query answering in $\mathcal{ELI}^f$ is tractable regarding data complexity.

An *inverse role* is an expression r^- with r a role name. The interpretation of an inverse role is $(r^-)^{\mathcal{I}} = \{(e,d) \mid (d,e) \in r^{\mathcal{I}}\}$. 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. Observe that w.l.o.g., we do not admit inverse roles in the ABox and the query.

As a preliminary, we assume that TBoxes are in a normal form, i.e., 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:

$$A \sqsubseteq B, \qquad A \sqsubseteq \exists r.B, \qquad \top \sqsubseteq (\leq 1\, r)$$
$$A_1 \sqcap A_2 \sqsubseteq B, \qquad \exists r.A \sqsubseteq B$$

Let $\mathcal{T}$ be a TBox. $\mathcal{T}$ can be converted into normal form $\mathcal{T}'$ in polytime, by introducing additional concept names. See [2] for more details. Moreover, it is not too difficult to see that for every ABox $\mathcal{A}$ and 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 $\mathcal{A}', \mathcal{T}' \models q$.

Two other (standard) assumptions that we make w.l.o.g. is that (i) in all atoms $C(v)$ in a conjunctive query q, C is a concept name; and (ii) 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 $v = u_n$. It is easy to achieve (i) by replacing $C(v)$ with $A(v)$ and adding $A \doteq C$ to the TBox, with A a fresh concept name. Regarding (ii), it is well-known that entailment of non-connected queries can easily (and polynomially) be reduced to entailment of connected queries: if q is a non-connected query, then $\mathcal{A}, \mathcal{T} \models q$ iff $\mathcal{A}, \mathcal{T} \models q'$ for all connected components q' of q; see e.g. [10].

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 Γ a finite set of concept names. We use $\mathsf{N}_\mathsf{C}^\mathcal{T}$ to denote the set of all concept names occurring in $\mathcal{T}$ and "$\sqsubseteq_\mathcal{T}$" to denote subsumption w.r.t. $\mathcal{T}$, i.e., $C \sqsubseteq_\mathcal{T} D$ iff $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{T}$. We write

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

to denote the *closure* of Γ 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 Γ (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 Γ *has* $\exists r.B$*-obligation* O, where

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

with $O' = \emptyset$ 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 in normal form and $\mathcal{A}$ an ABox, for which we want to decide conjunctive query entailment (for a yet unspecified query q). We use $\mathsf{Ind}(\mathcal{A})$ to denote

the set of individual names occurring in $\mathcal{A}$. 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 $b \neq c$. As will be discussed later, admissibility can be ensured by an easy (polytime) preprocessing step.

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 ε to denote the empty path. All interpretations $\mathcal{I}_i$ in the above sequence will satisfy

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

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

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

Now assume that $\mathcal{I}_i$ has already been defined. We want to construct $\mathcal{I}_{i+1}$. If it exists, select a $\langle a, p\rangle \in \Delta^{\mathcal{I}_i}$ and an $\alpha = \exists r.A \in \mathsf{ex}(\mathcal{T})$ such that $\langle a, p\rangle^{\mathcal{I}_i}$ has α-obligation O, and (i) $(\top \sqsubseteq (\leq 1\ r)) \notin \mathcal{T}$ and $\langle a, p\alpha\rangle \notin \Delta^{\mathcal{I}_i}$ or (ii) there is no $\langle b, p'\rangle \in \Delta^{\mathcal{I}_i}$ with $(\langle a, p\rangle, \langle b, p'\rangle) \in r^{\mathcal{I}_i}$. 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 $\langle a, p\alpha\rangle^{\mathcal{I}_i} := \mathsf{sub}_{\mathcal{T}}(O)$.

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 α to be selected). We assume that the selected $\langle a, p\rangle$ is such that the length of p is minimal, and thus all obligations are eventually satisfied. To ensure that the constructed canonical model is unique, we also assume that the set $\mathsf{ex}(\mathcal{T})$ is well-ordered and the selected α is minimal for the node $\langle a, p\rangle$.

A proof of the following result can be found in the appendix.

Lemma 3. *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 (maybe inverse) roles r and $d, e \in \Delta^{\mathcal{I}}$, $(d, e) \in r^{\mathcal{I}}$ implies $(h(d), h(e)) \in r^{\mathcal{J}}$.

Lemma 4. *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 $\mathcal{I}_0, \mathcal{I}_1, \dots$ used to construct $\mathcal{I}$, we define a homomorphism h_i from $\mathcal{I}_i$ to $\mathcal{J}$. The limit of the sequence $h_0, h_1, \dots$ 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 $\langle a, p\alpha\rangle \in \Delta^{\mathcal{I}_{i+1}} \setminus \Delta^{\mathcal{I}_i}$. Then $\langle a, p\rangle \in \Delta^{\mathcal{I}_i}$, and $\langle a, p\rangle^{\mathcal{I}_i}$ has $\alpha = \exists r.B$-obligation O such that $\langle a, p\alpha\rangle^{\mathcal{I}_{i+1}} = \mathsf{sub}_{\mathcal{T}}(O)$. Let $A \in \langle a, p\rangle^{\mathcal{I}_i}$ such that $A \sqsubseteq \exists r.B \in \mathcal{T}$. By Condition 2 of homomorphisms, we have $d = h_i(\langle a, p\rangle) \in A^{\mathcal{J}}$. 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 $h_{i+1}(\langle a, p\alpha\rangle) := e$. We prove that the three conditions of homomorphisms are preserved:

- Condition 1 is untouched by the extension.
- Now for Condition 2. Since $\langle a, p\alpha\rangle^{\mathcal{I}_{i+1}} = \mathsf{sub}_{\mathcal{T}}(O)$ and $\mathcal{J}$ is a model of $\mathcal{T}$, it suffices to show that for all $B' \in O$, we have $e \in B'^{\mathcal{J}}$. Let $B' \in O$. By definition of O, we can distinguish three cases.
  First, let $B' = B$. Then we are done by choice of e.
  Second, let there be an $A' \in \langle a, p\rangle^{\mathcal{I}_i}$ such that $\exists r^-.A' \sqsubseteq B' \in \mathcal{T}$. Since h_i satisfies Condition 2 of homomorphisms, we have $d \in A'^{\mathcal{J}}$. Since $\mathcal{J}$ is a model of $\mathcal{T}$ and $(d, e) \in r^{\mathcal{J}}$, it follows that $e \in B'^{\mathcal{J}}$.
  The third case is that $\top \sqsubseteq (\leq 1\, r) \in \mathcal{T}$ and there is an $A' \in \langle a, p\rangle^{\mathcal{I}_i}$ such that $A' \sqsubseteq \exists r.B' \in \mathcal{T}$. It is similar to the previous case.
- Condition 3 was satisfied by $\mathcal{I}_i$ and is clearly preserved by the extension to $\mathcal{I}_{i+1}$. ❑

Lemma 5. *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, and n, m, and k as above. If $\mathcal{I} \not\models q$, then $\mathcal{A}, \mathcal{T} \not\models q$ since, by Lemma 3, $\mathcal{I}$ 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 4, 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 π 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 6. *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 $\pi(v) = a^{\mathcal{I}}$. 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. Again since q is connected, this means that 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 $|p_0| > 2^m$ and the number of distinct labels $d^{\mathcal{I}}$, $d \in \Delta^{\mathcal{I}}$, is bounded by 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 $\pi(v) = \langle a, p_1p_2p_3p\rangle$. In the full version of the proof given in the appendix, we show that $\mathcal{I} \models^{\pi'} q$. Moreover, for each $v \in \mathsf{Var}(q)$ with $\pi(v) = \langle a, p\rangle$ and $\pi'(v) = \langle a, p'\rangle$, we have that the length of p' is strictly smaller than that of p. It follows that we can repeat the described construction to

construct a new match from an existing one only a finite number of times. We ultimately end up with a π^* such that $\mathcal{I} \not\models^{\pi^*} q$ and for all $v \in \mathsf{Var}(q)$, $\pi^*(v) = \langle a, p\rangle$ with $|p| \leq 2^m + k$. ❑

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 [12, 14], 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 $\mathcal{T}$ in normal form and an ABox $\mathcal{A}$ 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 [12, 14]. 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 ℓ^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. We thus obtain a time bound for our algorithm of $p(n^k \cdot m^{k \cdot 2^m + k^2})$, with $p()$ a polynomial. This bound is clearly polynomial in n

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

We conjecture that the time bound can be improved to $\mathcal{O}((n + 2^m)^k)$ (only single-exponential in m) by a more refined approach to canonical models. Basically, the idea is to work with the filtration of the canonical model instead of with the initial part.

A matching lower bound can be taken from [8] (which relies on the presence of general TBoxes and already applies to the instance problem), and thus we obtain P-completeness.

5 Summary and Outlook

The results of our investigation are summarized in Table 2. In all cases the lower bounds apply to instance checking and the upper bounds to conjunctive query entailment. The co-NP upper bounds are a consequence of the results in [10]. 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

Extensions of $\mathcal{EL}$	w.r.t. acyclic TBoxes	w.r.t. general TBoxes
$\mathcal{EL}^{\neg A}$	coNP-complete	coNP-complete
$\mathcal{EL}^{C \sqcup D}$	coNP-complete	coNP-complete
$\mathcal{EL}^{\forall r.\bot}$, $\mathcal{EL}^{\forall r.C}$	coNP-complete	coNP-complete
$\mathcal{EL}^{(\leq k r)}$, $k \geq 0$	coNP-complete	coNP-complete
$\mathcal{EL}^{kf}$ w/o UNA, $k \geq 2$	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
$\mathcal{EL}^{(\geq k r)}$, $k \geq 2$	coNP-complete	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 2. Complexity of instance checking and conjunctive query entailment

versus global functionality (for the latter, see $\mathcal{EL}^{(\leq 1 r)}$ vs. $\mathcal{ELI}^f$) can have an impact on tractability.

As future work, it would be interesting to extend our upper bound by including more operators from the tractable description logic $\mathcal{EL}^{++}$ as proposed in [2]. 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 $r_1 \circ \cdots \circ r_n \sqsubseteq s$. 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 entailed by the ABox $\{\top(a)\}$ and TBox $\mathcal{T}$.

We have learned recently that the same undecidability result has been shown independently and in parallel in the workshop papers [17, 18]. For people interested in the complexity of conjunctive querying entailment in the $\mathcal{EL}$ family of DLs, both papers are recommended reading. In particular, the algorithms for query answering presented there seem more suitable for implementation than the brute-force canonical model approach pursued in Section 4. We have also learned that our undecidability result is very similar to a number of undecidability results for subsumption in extensions of $\mathcal{EL}$ proved in [13].

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

References

1. A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. DL-Lite in the light of first-order logic. In *Proc. of the 22nd Conf. on AI (AAAI-07)*. AAAI Press, 2007
2. 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.
3. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. Submitted to a Journal. 2007
4. 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.
5. F. Baader, D. L. McGuiness, D. Nardi, and P. Patel-Schneider. *The Description Logic Handbook: Theory, implementation and applications*. Cambridge University Press, 2003.
6. 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.
7. 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.
8. 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.
9. 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.
10. B. Glimm and I. Horrocks and C. Lutz and U. Sattler. Conjunctive Query Answering for the Description Logic $\mathcal{SHIQ}$. In *Proc. of the 20th Int. Joint Conf. on AI (IJCAI-07)*. AAAI Press, 2007.
11. 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.
12. 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.
13. Y. Kazakov. Saturation-based decision procedures for extensions of the guarded fragment, *PhD thesis*, University of Saarland, 2005.
14. A. Krisnadhi. Data complexity of instance checking in the $\mathcal{EL}$ family of description logics. Master thesis, TU Dresden, Germany, 2007.
15. 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/
16. M. Krötzsch, S. Rudolph, 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.
17. M. Krötzsch and S. Rudolph. Conjunctive Queries for $\mathcal{EL}$ with Composition of Roles. In *Proc. of the 2007 Int. WS on DLs (DL2007)*. CEUR-WS.org, 2007.
18. R. Rosati. On conjunctive query answering in $\mathcal{EL}$. In *Proc. of the 2007 Int. WS on DLs (DL2007)*. CEUR-WS.org, 2007.
19. 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 Proofs

Lemma 3. 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 $\mathcal{I}$, 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 $\langle a, p\rangle \in A^{\mathcal{I}}$. This together with $A \sqsubseteq \exists r.B \in \mathcal{T}$ means that $\langle a, p\rangle^{\mathcal{I}_0}$ has α-obligation O, where $\alpha = \exists r.B$. Clearly, $B \in O$. 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 three subcases.
    For the first subcase, assume that $(\top \sqsubseteq (\leq 1\ r)) \notin \mathcal{T}$. By construction, 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 $\langle a, \varepsilon\rangle \in (\exists r.B)^{\mathcal{I}}$.
    For the second subcase, 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 third subcase, assume that $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and there is no $b \in \mathsf{Ind}(\mathcal{A})$ such that $r(a, b) \in \mathcal{A}$. By construction, there is a $\beta = \exists r.B' \in \mathsf{ex}(\mathcal{T})$ such that $\langle a, p\rangle^{\mathcal{I}_0}$ has β-obligation O' and there is an $i > 0$ such that $(\langle a, \varepsilon\rangle, \langle a, \beta\rangle) \in r^{\mathcal{I}_i} \subseteq r^{\mathcal{I}}$ and $\langle a, \beta\rangle^{\mathcal{I}_i} = \mathsf{sub}_{\mathcal{T}}(O')$. Since $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$, $O = O'$. Since $B \in O$, $\langle a, \beta\rangle \in B^{\mathcal{I}_i}$. 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 again three subcases.
    First, assume that $(\top \sqsubseteq (\leq 1\ r)) \notin \mathcal{T}$. Then, there is a $j > i$ with $(\langle a, p\rangle, \langle a, p\alpha\rangle) \in r^{\mathcal{I}_j}$ and $\langle a, p\alpha\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(O)$. Since $B \in O$,$\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$.
    Second, assume that $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and there is no $\langle b, p'\rangle \in \Delta^{\mathcal{I}_i}$ such that $(\langle a, p\rangle, \langle b, p'\rangle) \in r^{\mathcal{I}_i}$. By construction, there is a $\beta = \exists r.B' \in \mathsf{ex}(\mathcal{T})$ such that $\langle a, p\rangle^{\mathcal{I}_i}$ has β-obligation O' and there is a $j > i$ such that $(\langle a, p\rangle, \langle a, p\beta\rangle) \in r^{\mathcal{I}_j}$ and $\langle a, p\beta\rangle^{\mathcal{I}_j} = \mathsf{sub}_{\mathcal{T}}(O')$. Since $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$, $O = O'$. Since $B \in O$, $\langle a, p\beta\rangle \in B^{\mathcal{I}_i}$. It follows that $\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$.
    Last, assume that $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and there is a $\langle b, p'\rangle \in \Delta^{\mathcal{I}_i}$ such that $(\langle a, p\rangle, \langle b, p'\rangle) \in r^{\mathcal{I}_i}$. 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. $p = p'\alpha$ or for some $\beta = \exists r^-.B' \in \mathsf{ex}(\mathcal{T})$, or
    2. $p' = p\alpha$ for some $\beta = \exists r.B' \in \mathsf{ex}(\mathcal{T})$.
    First for Case 1. Then $\langle a, p'\rangle^{\mathcal{I}_i}$ has β-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 $\bigsqcap_{X \in \langle a, p'\rangle^{\mathcal{I}_i}} X \sqsubseteq_{\mathcal{T}} \exists r^-.A$. Together with $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$, we get $\bigsqcap_{X \in \langle a, p'\rangle^{\mathcal{I}_i}} X \sqsubseteq_{\mathcal{T}} B$. Since $\langle a, p'\rangle^{\mathcal{I}_i} = \mathsf{sub}_{\mathcal{T}}(\langle a, p'\rangle^{\mathcal{I}_i})$, we thus have $B \in \langle a, p'\rangle^{\mathcal{I}_i}$. By the semantics, $\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$.
    Now for Case 2. Then $\langle a, p\rangle^{\mathcal{I}_i}$ has β-obligation O', and $\langle a, p'\rangle^{\mathcal{I}_i} = \mathsf{sub}_{\mathcal{T}}(O')$. Since $(\top \sqsubseteq (\leq 1\ r)) \in \mathcal{T}$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$, we have $B \in O$. Thus $B \in \langle a, p'\rangle^{\mathcal{I}_i}$ and, $\langle a, p\rangle \in (\exists r.B)^{\mathcal{I}}$.

– $\exists r.A \sqsubseteq B$. Let $\langle a, p\rangle \in (\exists r.A)^{\mathcal{I}}$. Then there is a $\langle b, p'\rangle \in A^{\mathcal{I}}$ and such that $(\langle a, p\rangle, \langle b, p'\rangle) \in r^{\mathcal{I}}$. We distinguish four cases.
  ⋆ $p = p' = \varepsilon$. 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 $\mathcal{I}_0$.
  ⋆ $p = \varepsilon$, $p' \neq \varepsilon$. 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 $A \in \mathsf{sub}_{\mathcal{T}}(O)$, 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, $B \in \langle a, \varepsilon\rangle^{\mathcal{I}}$.
  ⋆ $p \neq \varepsilon$, $p' \neq \varepsilon$. There are two subcases. If $p' = p\alpha$ 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 $p = p'\alpha$ for some $\alpha = \exists r^-.B' \in \mathsf{ex}(\mathcal{T})$. In this case, $\langle a, p'\rangle^{\mathcal{I}}$ has $\exists r^-.B'$-obligation O, and $\langle a, p\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(O)$. Since $A \in \langle a, p'\rangle^{\mathcal{I}}$ and $\exists r.A \sqsubseteq B$, $B \in O$. It follows that $\langle a, p\rangle \in B^{\mathcal{I}}$.
  ⋆ $p \neq \varepsilon$, $p' = \varepsilon$. 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 admissible 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 $\mathcal{I}$.

❑

Lemma 6. 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. (Full Version) 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. Again since q is connected, this means that 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 $|p_0| > 2^m$ and the number of distinct labels $d^{\mathcal{I}}$, $d \in \Delta^{\mathcal{I}}$, is bounded by 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. $\pi'(v) \in \Delta^{\mathcal{I}}$ and $\pi(v)^{\mathcal{I}} = \pi'(v)^{\mathcal{I}}$;
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,

a) $\langle a, p_1p'\rangle \in \Delta^{\mathcal{I}}$ and
b) $\langle a, p_1p'\rangle^{\mathcal{I}} = \langle a, p_1p_2p'\rangle^{\mathcal{I}}$.

For $p' = \varepsilon$, Point a) is true since $\langle a, p_0\rangle \in \Delta^{\mathcal{I}}$ and by construction of $\mathcal{I}$, $\langle a, p''\rangle \in \Delta^{\mathcal{I}}$ for all prefixes p'' of p_0, including $p'' = p_1$. Moreover, Point b) 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. Then p_1p_2p' has α-obligation O and $\langle a, p_1p_2p'\alpha\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(O)$. By induction hypothesis, $\langle a, p_1p'\rangle^{\mathcal{I}} = \langle a, p_1p_2p'\rangle^{\mathcal{I}}$. It follows that p_1p' also has α-obligation O (here we exploit the well-order on $\mathsf{ex}(\mathcal{T})$. By construction of $\mathcal{I}$, we thus have $\langle a, p_1p'\alpha\rangle \in \Delta^{\mathcal{I}}$ and $\langle a, p_1p'\alpha\rangle^{\mathcal{I}} = \mathsf{sub}_{\mathcal{T}}(O)$. The former proves Point a) and the latter Point b). 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.

We have thus proved Point 2, i.e. $\mathcal{I} \models^{\pi'} q$. Moreover, for each $v \in \mathsf{Var}(q)$ with $\pi(v) = \langle a, p\rangle$ and $\pi'(v) = \langle a, p'\rangle$, we obviously have that the length of p' is strictly smaller than that of p. It follows that we can repeat the described construction to construct a new match from an existing one only a finite number of times. We ultimately end up with a π^* such that $\mathcal{I} \models^{\pi^*} q$ and for all $v \in \mathsf{Var}(q)$, $\pi^*(v) = \langle a, p\rangle$ with $|p| \leq 2^m + k$. ❑