First-Order Rewritability of Atomic Queries in Horn Description Logics

Meghyn Bienvenu
CNRS & Université Paris Sud
Orsay, France

Carsten Lutz
Fachbereich Informatik
Universität Bremen, Germany

Frank Wolter
Department of Computer Science
University of Liverpool, UK

Abstract

One of the most advanced approaches to querying data in the presence of ontologies is to make use of relational database systems, rewriting the original query and the ontology into a new query that is formulated in SQL or, equivalently, in first-order logic (FO). For ontologies written in many standard description logics (DLs), however, such FO-rewritings are not guaranteed to exist. We study FO-rewritings and their existence for a basic class of queries and for ontologies formulated in Horn DLs such as Horn-$\mathcal{SHI}$ and $\mathcal{EL}$. Our results include characterizations of the existence of FO-rewritings, tight complexity bounds for deciding whether an FO-rewriting exists (ExpTime and PSpace), and tight bounds on the (worst-case) size of FO-rewritings, when presented as a union of conjunctive queries.

1 Introduction

A prominent application of description logic (DL) ontologies is to facilitate access to data. Specifically, the ontology serves to assign a semantics to the relation symbols used in the data; it can also provide additional relation symbols that, although not explicitly occurring in the data, can be used in the query. Several approaches to querying data in the presence of ontologies utilize relational databases systems (RDBMSs), aiming to exploit their mature technology, advanced optimization techniques, and the general infrastructure that those systems offer. One of the most popular such approaches is to rewrite the original query and the DL ontology into an SQL query that is passed to the RDBMS for execution [Calvanese *et al.*, 2007; Pérez-Urbina *et al.*, 2009; Chortaras *et al.*, 2011; Gottlob *et al.*, 2011]. Based on the equivalence of first-order (FO) formulas and SQL queries, we call the rewritten query an *FO-rewriting*.

The FO-rewriting approach to ontology-based data access comes with its own set of DLs specifically designed for this purpose, the so-called DL-Lite family. To guarantee that an FO-rewriting of queries and ontologies always exists, DLs from this family are significantly restricted in expressive power. Unfortunately, this is not acceptable for all applications, in particular when the ontology is used to model the application domain in a detailed way instead of providing only more abstract, database-style constraints. Given that very expressive DLs of the $\mathcal{ALC}$ and $\mathcal{SHIQ}$ families do not admit tractable query answering (regarding data complexity), a good compromise between expressive power and computational complexity is provided by so-called Horn DLs such as $\mathcal{EL}$, $\mathcal{ELI}$, and Horn-$\mathcal{SHIQ}$. Already in the basic DLs of this kind, such as $\mathcal{EL}$, and for the simplest kind of queries, known as atomic queries (AQs) or instance queries, FO-rewritings are not guaranteed to always exist. To address this problem, other approaches to utilize RDBMSs and related database technology have been brought forward, including the combined approach [Lutz *et al.*, 2009] and rewritings into datalog [Pérez-Urbina *et al.*, 2009; Eiter *et al.*, 2012].

Depending on the application, however, there can still be good reasons to use the FO-rewriting approach for Horn DLs. First, an important feature of this approach is that it allows the ontology to be added on top of the query interface without any modifications to the underlying database. By contrast, the combined approach involves a data completion step, and thus can only be used if data manipulations are permitted. Second, a rewriting into FO queries rather than datalog programs means that one can exploit the comparatively more advanced optimization techniques available for SQL queries. For both reasons, when FO-rewritings happen to exist for the relevant queries and ontologies, the FO-rewriting approach might be very appropriate.

In this paper, we consider ontologies formulated in the Horn DLs $\mathcal{EL}$, $\mathcal{ELI}_\bot$, and Horn-$\mathcal{SHI}$. Notably, $\mathcal{EL}$ forms the basis of the OWL EL fragment of the OWL 2 web ontology language and is popular as a basic language for large-scale ontologies [Baader *et al.*, 2005]. $\mathcal{ELI}_\bot$ can be viewed as the smallest DL that contains as a fragment both $\mathcal{EL}$ and the core version of DL-Lite, and Horn-$\mathcal{SHI}$ is a generalization of $\mathcal{ELI}_\bot$ inspired by the well-known DL Horn-$\mathcal{SHIQ}$, but in contrast to the latter does not admit number restrictions [Hustadt *et al.*, 2007]. As an example for why the rewriting approach fails for these DLs, consider the AQ $A(x)$ and the $\mathcal{EL}$ ontology $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$. The query $A(x)$ cannot be rewritten into an FO-query in the presence of $\mathcal{T}$, intuitively because $\mathcal{T}$ forces the concept name A to be *propagated unboundedly* along r-chains in the data and thus the rewritten query would have to express transitive closure of r. Of course, such an isolated example does not rule out the possibility that for some AQs and some $\mathcal{EL}$ (or Horn-$\mathcal{SHI}$) ontologies, including those that are used in applications, FO-rewritings do exist. For example, $A(x)$ is FO-rewritable relative to the $\mathcal{EL}$ ontology $\mathcal{T}' = \{A \sqsubseteq \exists r.A\}$, which has a lot of similarity with the aforementioned ontology $\mathcal{T}$. In fact, $\mathcal{T}'$ can simply be ignored when answering $A(x)$ without losing any answers. Inspired by these observations, the aim of this paper is to study FO-rewritings of AQs in the presence of ontologies formulated in $\mathcal{EL}$, $\mathcal{ELI}$, $\mathcal{ELI}_\bot$, and Horn-$\mathcal{SHI}$.

We primarily study the problem to decide, given an atomic query (AQ) q, an ontology $\mathcal{T}$, and a finite set Σ of symbols that are allowed to be used in the data (ABox), whether q is FO-rewritable relative to $\mathcal{T}$ over Σ-ABoxes. Note that the restriction of the data signature is natural in many applications of ontology-based data access, cf. [Baader *et al.*, 2010; Bienvenu *et al.*, 2012b]. We show that this problem is EXPTIME-complete for ontologies formulated in Horn-$\mathcal{SHI}$, where the lower bound applies even to $\mathcal{ELI}$ ontologies and when the ABox signature is the full signature (that is, it must contain all concept and role names) rather than being an input. For ontologies formulated in $\mathcal{EL}$, the problem remains EXPTIME-complete when the ABox signature is an input (though the lower bound is more difficult to establish), but is only PSPACE-complete when the ABox signature is full.

Our analysis also yields characterizations of the existence of FO-rewritings in terms of the existence of certain tree-shaped ABoxes, which are interesting in their own right. Surprisingly, tree-shaped ABoxes can even be replaced with linear ABoxes (single role chains decorated with concept assertions) when the ontology is formulated in $\mathcal{EL}$ and the ABox signature is full. Our proofs also yield a way to effectively construct FO-rewritings when they exist. We use this observation to analyze the size of FO-rewritings, showing that they can always be represented by a union of conjunctive queries (UCQ) of at most triple exponential size, and that this bound is essentially optimal: there are families of AQs and $\mathcal{EL}$ ontologies for which FO-rewritings exist, but such that every presentation of the rewritings as a UCQ is necessarily triple-exponential in size.

Some proof details are deferred to the appendix of the long version, http://www.informatik.uni-bremen.de/∼clu/papers/.

2 Preliminaries

Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be disjoint and countably infinite sets of *concept* and *role names*. A *role* is a role name r or an *inverse role* r^-, with r a role name. A *Horn-$\mathcal{SHI}$ concept inclusion (CI)* is of the form $L \sqsubseteq R$, where L and R are concepts defined by the syntax rules

$$R, R' ::= \top \mid \bot \mid A \mid \neg A \mid R \sqcap R' \mid \neg L \sqcup R \mid \exists r.R \mid \forall r.R$$
$$L, L' ::= \top \mid \bot \mid A \mid L \sqcap L' \mid L \sqcup L' \mid \exists r.L$$

with A ranging over concept names and r over roles. In DLs, ontologies are formalized as TBoxes. A *Horn-$\mathcal{SHI}$ TBox* $\mathcal{T}$ is a finite set of Horn-$\mathcal{SHI}$ CIs, *transitivity assertions* $\mathsf{trans}(r)$, and *role inclusions (RI)* $r \sqsubseteq s$, with r and s roles. Note that different definitions of Horn-$\mathcal{SHI}$ can be found in the literature [Hustadt *et al.*, 2007; Eiter *et al.*, 2008; Kazakov, 2009]. As the original definition from [Hustadt *et al.*, 2007] based on polarity is rather technical, we prefer the above (equivalent) definition.

An $\mathcal{ELI}_\bot$ *TBox* is a finite set of inclusions of the form $L \sqsubseteq L'$ where L, L' are constructed by the rule above, but without using disjunction. An $\mathcal{ELI}_\bot$ TBox that does not use the $\bot$ concept is an $\mathcal{ELI}$ *TBox*, and an $\mathcal{ELI}$ TBox that does not use inverse roles is an $\mathcal{EL}$ *TBox*.

An *ABox* is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a, b)$ where A is a concept name, r a role name, and a, b individual names from a countably infinite set $\mathsf{N_I}$. We sometimes write $r^-(a, b)$ instead of $r(b, a)$ and use $\mathsf{Ind}(\mathcal{A})$ to denote the set of all individual names used in $\mathcal{A}$.

The semantics of DLs is given in terms of *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set (the *domain*) and $\cdot^\mathcal{I}$ is the *interpretation function*, assigning to each $A \in \mathsf{N_C}$ a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, to each $r \in \mathsf{N_R}$ a relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, and to each $a \in \mathsf{N_I}$ an element $a^\mathcal{I} \in \Delta^\mathcal{I}$ such that $a_1^\mathcal{I} \neq a_2^\mathcal{I}$ whenever $a_1 \neq a_2$ (the so-called unique name assumption). The interpretation $C^\mathcal{I} \subseteq \Delta^\mathcal{I}$ of a concept C in $\mathcal{I}$ is defined as usual, see [Baader *et al.*, 2003]. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, a transitivity assertion $\mathsf{trans}(r)$ if $r^\mathcal{I}$ is transitive, an RI $r \sqsubseteq s$ if $r^\mathcal{I} \subseteq s^\mathcal{I}$, a concept assertion $A(a)$ if $a^\mathcal{I} \in A^\mathcal{I}$, and a role assertion $r(a, b)$ if $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$. We say that $\mathcal{I}$ is a *model of* a TBox or an ABox if it satisfies all inclusions and assertions in it. An ABox $\mathcal{A}$ is *consistent* w.r.t. a TBox $\mathcal{T}$ if $\mathcal{A}$ and $\mathcal{T}$ have a common model.

An *atomic query (AQ)* takes the form $A(x)$, with A a concept name and x a variable. We write $\mathcal{A}, \mathcal{T} \models A(a)$ if $a^\mathcal{I} \in A^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$. If $\mathcal{A}, \mathcal{T} \models A(a)$ and $a \in \mathsf{Ind}(\mathcal{A})$, then a is a *certain answer* to $A(x)$ given $\mathcal{A}$ and $\mathcal{T}$. We use $\mathsf{cert}_\mathcal{T}(A(x), \mathcal{A})$ to denote the set of all certain answers to $A(x)$ given $\mathcal{A}$ and $\mathcal{T}$. A first-order query (FOQ), is a first-order formula φ constructed from atoms $A(x)$, $r(x, y)$, and $x = y$; here, concept names are viewed as unary predicates, role names as binary predicates, and predicates of other arity, function symbols, and constant symbols are not permitted. As usual, we write $\varphi(\vec{x})$ to indicate that the free variables of φ are among $\vec{x}$ and call $\vec{x}$ the *answer variables* of φ. The number of answer variables is the *arity* of φ and a FOQ φ is *Boolean* if it has arity zero. We use $\mathsf{ans}(\mathcal{I}, \varphi)$ to denote the set of all answers to the FOQ φ in the interpretation $\mathcal{I}$; that is, if φ is n-ary, then $\mathsf{ans}(\mathcal{I}, \varphi)$ contains all tuples $\vec{d} \in (\Delta^\mathcal{I})^n$ such that the FO-sentence $\varphi[\vec{d}]$ is satisfied in $\mathcal{I}$ (written $\mathcal{I} \models \varphi[\vec{d}]$). To bridge the gap between certain answers and "normal" answers, we sometimes view an ABox $\mathcal{A}$ as an interpretation $\mathcal{I}_\mathcal{A}$, defined in the obvious way; see [Lutz and Wolter, 2012].

A *signature* is a set of concept and role names, which are uniformly called *symbols* in this context. We use $\mathsf{sig}(\mathcal{T})$ to denote the set of symbols used in the TBox $\mathcal{T}$. A Σ-*ABox* is an ABox that uses only concept and role names from Σ. We speak of an *ABox signature* if the purpose of the signature is to fix the symbols permitted in ABoxes.

Definition 1 (FO-rewriting)**.** *Let $\mathcal{T}$ be a TBox and Σ an ABox signature. A FOQ $\varphi(x)$ is an* FO-rewriting of an AQ $A(x)$ relative to $\mathcal{T}$ and Σ *if* $\mathsf{cert}_\mathcal{T}(A(x), \mathcal{A}) = \mathsf{ans}(\mathcal{I}_\mathcal{A}, \varphi)$ *for all Σ-ABoxes $\mathcal{A}$. If there is such a $\varphi(x)$, then $A(x)$ is* FO-rewritable relative to $\mathcal{T}$ and Σ.

Thus, FO-rewritings reduce the computation of certain answers (which is a form of deduction) to standard query answering on structures (which is a form of model checking).

Example 2. *Recall from the introduction that $A(x)$ is not FO-rewritable relative to $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$ and the signature $\Sigma = \{r, A\}$. If we add $\exists r.\top \sqsubseteq A$ to $\mathcal{T}$, then $A(x)$ is FO-rewritable relative to the resulting TBox and Σ, and $\varphi(x) = A(x) \vee \exists y\, r(x, y)$ is an FO-rewriting. If we choose $\Sigma = \{A\}$, then $A(x)$ becomes FO-rewritable also relative to the original $\mathcal{T}$, with the trivial FO-rewriting $A(x)$.*

In some applications, the signature of the ABox is not restricted at all and thus, in principle, infinite. However, FO-rewritability of an AQ $A(x)$ relative to $\mathcal{T}$ and any (potentially infinite) signature $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$ coincides with FO-rewritability of $A(x)$ relative to $\mathcal{T}$ and the (finite) ABox signature $\mathsf{sig}(\mathcal{T}) \cap \Sigma$. In fact, any FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ is trivially also an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T}) \cap \Sigma$, and when $\varphi(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T}) \cap \Sigma$, then (i) $\varphi(x)$ is also an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ if $A \in \mathsf{sig}(\mathcal{T})$ and (ii) $\varphi(x) \vee A(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ otherwise. Consequently, we from now on restrict our attention to ABox signatures Σ with $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$ and speak of the *full signature* Σ when $\Sigma = \mathsf{sig}(\mathcal{T})$.

Atomic queries are closely related to queries of the more general form $C(x)$ with C an $\mathcal{EL}$ concept or an $\mathcal{ELI}$ concept. Note that such queries can be viewed as tree-shaped conjunctive queries where the root is the only answer variable. The results presented in this paper also capture these more general queries since $\varphi(x)$ is an FO-rewriting of $C(x)$ relative to $\mathcal{T}$ and an ABox signature Σ iff it is an FO-rewriting of $A(x)$ relative to $\mathcal{T} \cup \{A \equiv C\}$ and Σ, with A a fresh concept name.

The reasoning problem studied in this paper is as follows: given an AQ $A(x)$, a TBox $\mathcal{T}$, and an ABox signature Σ (with $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$), decide whether $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ and if this is the case, produce an FO-rewriting. We obtain different versions of this problem by varying the language in which the TBox $\mathcal{T}$ can be formulated, and by admitting a finite ABox signature Σ as input or fixing it to be the full signature.

We also consider restricted forms of FOQs for rewriting atomic queries, and restricted kinds of ABoxes. A FOQ is a *conjunctive query (CQ)* if it has the form $\exists \vec{y} \varphi(\vec{x}, \vec{y})$ with φ a conjunction of atoms; it is a *union of conjunctive queries (UCQ)* if it is a disjunction of CQs. For simplicity, we disallow equality in CQs and UCQs. If an FO-rewriting is a UCQ, we speak of a *UCQ-rewriting*. With every ABox $\mathcal{A}$, we associate the undirected graph $G_\mathcal{A}$ with nodes $\mathsf{Ind}(\mathcal{A})$ and edges $\{\{a, b\} \mid r(a, b) \in \mathcal{A} \text{ or } r(b, a) \in \mathcal{A}\}$. An ABox $\mathcal{A}$ is *acyclic* if the corresponding graph $G_\mathcal{A}$ is acyclic and $r(a, b) \in \mathcal{A}$ implies that (i) $s(a, b) \notin \mathcal{A}$ for all $s \neq r$ and (ii) $s(b, a) \notin \mathcal{A}$ for all role names s; $\mathcal{A}$ is *tree-shaped* if it is acyclic and $G_\mathcal{A}$ is connected. In tree-shaped ABoxes $\mathcal{A}$, we often distinguish one individual $\rho_\mathcal{A} \in \mathsf{Ind}(\mathcal{A})$ as the *root* of $\mathcal{A}$.

We often identify a CQ q with the set of its atoms and regard q as an ABox whose individual names are the variables of q. Then q is called *acyclic* or *tree-shaped* if the ABox corresponding to q has the same property.

3 FO-rewritability in Horn-$\mathcal{SHI}$ and $\mathcal{ELI}_\bot$

We show that deciding FO-rewritability in Horn-$\mathcal{SHI}$ and $\mathcal{ELI}_\bot$ is ExpTime-complete. To achieve this, we provide a characterization of FO-rewritability in terms of the existence of certain ABoxes, which is of independent interest. We also show how to compute FO-rewritings if they exist and give upper and lower bounds on their size.

We start by observing that it suffices to concentrate on TBoxes that are formulated in $\mathcal{ELI}_\bot$ and in *normal form*, that is, all CIs are of one of the forms

$$A \sqsubseteq \bot \quad A \sqsubseteq \exists r.B \quad \top \sqsubseteq A \quad B_1 \sqcap B_2 \sqsubseteq A \quad \exists r.B \sqsubseteq A$$

with A, B, B_1, B_2 concept names and r a role.

Theorem 3. *For every Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$ and ABox signature Σ, one can construct in polynomial time an $\mathcal{ELI}_\bot$ TBox $\mathcal{T}'$ such that for all AQs $A(x)$ with $A \notin \mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$, every FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and Σ, and vice versa.*

The proof of Theorem 3 is similar to reductions in [Hustadt *et al.*, 2007; Kazakov, 2009]. For the elimination of value restrictions $\forall r.B$, observe that the CI $A \sqsubseteq \forall r.B$ is logically equivalent to the CI $\exists r^-.A \sqsubseteq B$. In the following, we will generally work with $\mathcal{ELI}_\bot$ TBoxes and assume normal form whenever this is more convenient.

The following is a direct consequence of a result by Rossman and the fact that the class of finite pointed structures $(\mathcal{I}_\mathcal{A}, a)$, with $\mathcal{A}, \mathcal{T} \models A(a)$, is preserved under homomorphisms [Rossman, 2008].

Proposition 4. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox and Σ an ABox signature. If an AQ $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ, then there is a UCQ-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ.*

Next, we observe that ABox inconsistency plays a central role for FO-rewritability of AQs relative to $\mathcal{ELI}_\bot$ TBoxes.

Example 5. *Let $\mathcal{T} = \{\exists r.A \sqsubseteq A, A \sqcap B \sqsubseteq \bot\}$. Then $B(x)$ is not FO-rewritable since $\mathcal{A}, \mathcal{T} \models B(a)$ whenever $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$, which is the case iff in $\mathcal{A}$, there are individuals a and b such that $A(a) \in \mathcal{A}$, a is reachable from b on an r-path, and $B(b) \in \mathcal{A}$. Clearly, this condition cannot be expressed by an FO-formula. If we only admit ABoxes that are consistent w.r.t. $\mathcal{T}$, then $B(x)$ is trivially rewritable (it is a rewriting itself).*

To make precise the interplay between FO-rewritability of AQs and of ABox inconsistency, we require some further notions. We say that *ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and Σ* if there is a FOQ $\varphi()$ such that for every Σ-ABox $\mathcal{A}$, $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ iff $\mathcal{I}_\mathcal{A} \models \varphi()$. We call an AQ $A(x)$ *FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes* if there exists a FOQ $\varphi(x)$ such that $\mathsf{cert}_\mathcal{T}(q, \mathcal{A}) = \mathsf{ans}_{\mathcal{I}_\mathcal{A}}(\varphi)$ for all Σ-ABoxes $\mathcal{A}$ that are consistent w.r.t. $\mathcal{T}$. Finally, we an AQ $A(x)$ is *Σ-trivial relative to $\mathcal{T}$* if $\mathcal{A}, \mathcal{T} \models A(a)$ for all Σ-ABoxes $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$. Note that Σ-trivial AQs are FO-rewritable relative to Σ (with $x = x$ a rewriting) and that $A(x)$ is Σ-trivial relative to $\mathcal{T}$ iff $\mathcal{T} \models C \sqsubseteq A$ for all concept names $C \in \Sigma$ and all C of the form $\exists r.\top$ and $\exists r^-.\top$ with $r \in \Sigma$. Thus, it is straightforward to check Σ-triviality of AQs.

Proposition 6. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox, Σ an ABox signature, and $A(x)$ an AQ that is not Σ-trivial relative to $\mathcal{T}$. Then $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ iff*

1. *$A(x)$ is FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes, and*
2. *ABox inconsistency is FO-rewritable relative to $\mathcal{T}$, Σ.*

Proof. Assume first that Points 1 and 2 hold. Let $\varphi_1()$ be an FO-rewriting of ABox inconsistency relative to $\mathcal{T}$ and Σ, and let $\varphi_2(x)$ be an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and consistent Σ-ABoxes. Then $(\varphi_1 \wedge x = x) \vee \varphi_2$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ.

Conversely, assume that there is an FO-rewriting $\varphi(x)$ of $A(x)$ relative to $\mathcal{T}$ and Σ, and that $A(x)$ is not Σ-trivial relative to $\mathcal{T}$. Point 1 is trivial since $\varphi(x)$ is an FO-rewriting of $A(x)$ also relative to $\mathcal{T}$ and consistent Σ-ABoxes. For Point 2, Proposition 4 implies that we may assume $\varphi(x)$ to be a UCQ $q_1 \vee \cdots \vee q_n$. Let $\psi()$ be the union of all Boolean CQs p such that

(i) $p \subseteq q_i$ (p is a subset of q_i) for some $1 \leq i \leq n$;

(ii) for all Σ-ABoxes $\mathcal{A}$: if $\mathcal{I}_\mathcal{A} \models p$, then $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$.

We show that $\psi()$ is an FO-rewriting of ABox inconsistency relative to $\mathcal{T}$ and Σ. By Point (ii), $\mathcal{I}_\mathcal{A} \models \psi()$ implies that $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$. Conversely, assume that $\mathcal{A}$ is a Σ-ABox that is inconsistent w.r.t. $\mathcal{T}$. Since $A(x)$ is not Σ-trivial relative to $\mathcal{T}$, there is a Σ-ABox $\mathcal{A}_0$ and $a_0 \in \mathsf{Ind}(\mathcal{A}_0)$ such that $\mathcal{A}_0, \mathcal{T} \not\models A(a_0)$. Let $\mathsf{Ind}(\mathcal{A}) \cap \mathsf{Ind}(\mathcal{A}_0) = \emptyset$. We have $\mathcal{A}_0 \cup \mathcal{A}, \mathcal{T} \models A(a_0)$ since $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$. Hence $\mathcal{I}_{\mathcal{A}\cup\mathcal{A}_0} \models \varphi[a_0]$ and there is a q_i such that $\mathcal{I}_{\mathcal{A}\cup\mathcal{A}_0} \models q_i(a_0)$. Let π be a match of q_i in $\mathcal{I}_{\mathcal{A}\cup\mathcal{A}_0}$ and let p be the Boolean CQ that consists of all atoms in q_i whose variables are mapped by π to $\mathsf{Ind}(\mathcal{A})$. It is readily checked that p is a disjunct of $\psi()$ and so $\mathcal{I}_\mathcal{A} \models \psi()$, as required. □

Proposition 6 suggests a decomposition of the test for FO-rewritability of AQs: first check FO-rewritability of ABox inconsistency and then check FO-rewritability of the AQ relative to consistent ABoxes. In the following, we pursue this approach.

We now develop characterizations of FO-rewritability in terms of the existence of certain ABoxes. For a tree-shaped ABox $\mathcal{A}$ with distinguished root $\rho_\mathcal{A}$ and $k \geq 0$, we use $\mathcal{A}|_k$ to denote the restriction of $\mathcal{A}$ to all $a \in \mathsf{Ind}(\mathcal{A})$ with distance from $\rho_\mathcal{A}$ less than or equal to k. Moreover, $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ denotes the (acyclic) ABox obtained from $\mathcal{A}$ by removing from $\mathcal{A}$ all assertions that involve $\rho_\mathcal{A}$.

Theorem 7. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox, Σ an ABox signature, and $A(x)$ an AQ.*

1. *$A(x)$ is FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes iff there is a $k \geq 0$ such that for all tree-shaped Σ-ABoxes $\mathcal{A}$ with root $\rho_\mathcal{A}$ that are consistent w.r.t. $\mathcal{T}$: if $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$, then $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$;*
2. *ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and Σ iff there is a $k \geq 0$ such that for all tree-shaped Σ-ABox $\mathcal{A}$ with root $\rho_\mathcal{A}$: if $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ and $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ is consistent w.r.t. $\mathcal{T}$, then $\mathcal{A}|_k$ is inconsistent w.r.t. $\mathcal{T}$.*

Proof (sketch). For the "only if" direction of Point 1, let φ be an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and consistent Σ-ABoxes. By Proposition 4, we may assume φ to be a UCQ. Let k be the maximum number of atoms in any CQ in φ. It can be shown that k is the required bound, that is, for all tree-shaped Σ-ABoxes $\mathcal{A}$ with root $\rho_\mathcal{A}$ that are consistent w.r.t. $\mathcal{T}$, $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ implies $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$.

For the "if" direction, assume that k satisfies the conditions in Point 1. We consider all minimal tree-shaped Σ-ABoxes $\mathcal{A}$ of depth at most k that are consistent w.r.t. $\mathcal{T}$ and such that $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ and $\mathcal{A}|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$. View each such ABox $\mathcal{A}$ as a (tree-shaped) CQ $q_\mathcal{A}$ in the obvious way with the root $\rho_\mathcal{A}$ translated into the answer variable x, and define $\varphi(x)$ to be the UCQ obtained as the disjunction of all the CQs $q_\mathcal{A}$ (it is not hard to see that there are only finitely many such queries, up to equivalence). Then $\varphi(x)$ is a UCQ-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ on acyclic Σ-ABoxes. We use a result on 'unraveling tolerance' from [Lutz and Wolter, 2012] to show that such a $\varphi(x)$ must also be a rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ (for this argument, it is not sufficient to know that $\varphi(x)$ is a UCQ-rewriting on tree-shaped Σ-ABoxes). The proof of Point 2 is similar. □

The following examples illustrate Theorem 7.

Example 8. *(1) We use Point 1 of Theorem 7 to show that $A(x)$ is not FO-rewritable relative to $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$ and the full ABox signature. In fact, it suffices to observe that the ABoxes $\mathcal{A}_k = \{r(a_0, a_1), \ldots, r(a_k, a_{k+1}), A(a_{k+1})\}$ with $\rho_{\mathcal{A}_k} = a_0$ are consistent w.r.t. $\mathcal{T}$ and satisfy $\mathcal{A}_k, \mathcal{T} \models A(\rho_{\mathcal{A}_k})$ and $\mathcal{A}_k|_k, \mathcal{T} \not\models A(\rho_{\mathcal{A}_k})$.*

(2) In Example 5, it was claimed that ABox inconsistency is not FO-rewritable relative to $\mathcal{T} = \{\exists r.A \sqsubseteq A, A \sqcap B \sqsubseteq \bot\}$ and the full ABox signature. This is a consequence of Point 2 of Theorem 7 and the facts that the ABoxes $\mathcal{A}'_k = \mathcal{A}_k \cup \{B(a_0)\}$ are not consistent w.r.t. $\mathcal{T}$, but with $\rho_{\mathcal{A}'_k} = a_0$, both $\mathcal{A}'_k|_k$ and $\mathcal{A}'_k \setminus \{\rho_{\mathcal{A}'_k}\}$ are consistent w.r.t. $\mathcal{T}$.

(3) In Point 2 of Theorem 7, the precondition that $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ has to be consistent w.r.t. $\mathcal{T}$ cannot be dropped. To show this, let $\mathcal{T} = \{A \sqsubseteq \bot\}$. Then ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and $\Sigma = \{A, r\}$ (with rewriting $\exists x A(x)$), but $\mathcal{A}_k$ is inconsistent w.r.t. $\mathcal{T}$ and $\mathcal{A}_k|_k$ is consistent w.r.t. $\mathcal{T}$.

To exploit Theorem 7 for developing a decision procedure for FO-rewritability, we prove that the depth and outdegree of the tree-shaped ABoxes considered in that theorem can be bounded. We use $|\mathcal{T}|$ to denote the size of the TBox $\mathcal{T}$, that is, the number of symbols needed to write $\mathcal{T}$.

Theorem 9. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox in normal form, Σ an ABox signature, $A(x)$ an AQ, and $n = |\mathcal{T}|$. Then Points 1 and 2 of Theorem 7 still hold when*

1. *"there is a $k \geq 0$" is replaced with "for $k = 2^{3n^2}$" and*
2. *"tree-shaped Σ-ABox $\mathcal{A}$" is replaced with "tree-shaped Σ-ABox $\mathcal{A}$ of outdegree at most n".*

Point 1 is established by a very careful pumping argument (here, the presence of inverse roles complicates matters significantly), and Point 2 relies on a selection of the relevant individuals in tree-shaped ABoxes.

Theorem 9 immediately suggests a naïve decision procedure for deciding FO-rewritability: simply enumerate all tree-shaped Σ-ABoxes up to the relevant bounds and check that they have the required properties. To obtain better complexity, we construct a tree automaton that accepts precisely those ABoxes which violate the required properties, and use a subsequent emptiness test. Specifically, we work with alternating two-way Büchi automata on finite trees, using the two-way feature to handle inverse roles. We obtain EXPTIME upper bounds and establish matching lower bounds via a reduction from subsumption in $\mathcal{ELI}$, which is EXPTIME-hard [Baader *et al.*, 2008]. Theorem 3 lifts the upper bounds to Horn-$\mathcal{SHI}$.

Theorem 10. *The following problems are* EXPTIME*-complete, with the lower bounds already applying to $\mathcal{ELI}$ (Points 1 and 3) and $\mathcal{ELI}_\bot$ (Point 2), and to the full ABox signature:*

1. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, an ABox signature Σ, and an AQ $A(x)$, is $A(x)$ FO-rewritable relative to $\mathcal{T}$ and Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$?*
2. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$ and an ABox signature Σ, is inconsistency of Σ-ABoxes FO-rewritable relative to $\mathcal{T}$ and Σ?*
3. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, an ABox signature Σ, and an AQ $A(x)$, is $A(x)$ FO-rewritable relative to $\mathcal{T}$ and Σ?*

We now discuss the actual computation of FO-rewritings. A method for computing FO-rewritings of AQs relative to consistent ABoxes and FO-rewritings of ABox inconsistency is implicit in the proof (sketch) of Theorem 7. As explained in the proof of Proposition 6, these rewritings can be combined into FO-rewritings of AQs, without the restriction to consistent ABoxes. By Theorem 3, this can be lifted from $\mathcal{ELI}_\bot$ to Horn-$\mathcal{SHI}$. The constructed rewritings are UCQs whose disjuncts are tree-shaped CQs and whose size is at most triple exponential in the size of the TBox. Using constructions from [Lutz and Wolter, 2010; Nikitina and Rudolph, 2012], one can show that this is essentially optimal. The lower bound already applies to $\mathcal{EL}$-TBoxes and the full ABox signature.

Theorem 11.

1. *For every Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, signature Σ, and AQ $A(x)$ that is rewritable relative to $\mathcal{T}$ and Σ, one can effectively construct a UCQ-rewriting $\varphi(x)$ of size at most $2^{2^{2^{\mathcal{O}(|\mathcal{T}|^2)}}}$, in time polynomial in the size of $\varphi(x)$.*
2. *There is a family of $\mathcal{EL}$ TBoxes $\mathcal{T}_0, \mathcal{T}_1, \ldots$ such that for all $i \geq 0$, $|\mathcal{T}_i| \in \mathcal{O}(i^2)$ and $A(x)$ is FO-rewritable relative to $\mathcal{T}_i$ and the full ABox signature $\mathsf{sig}(\mathcal{T}_i)$, but the smallest UCQ-rewriting is of size at least $2^{2^{2^i}}$.*

4 FO-rewritability in $\mathcal{EL}$

We next consider FO-rewritability relative to TBoxes formulated in the popular lightweight DL $\mathcal{EL}$. Unlike for Horn-$\mathcal{SHI}$ and $\mathcal{ELI}_\bot$, we obtain different complexities depending on whether we admit the ABox signature as an input or fix it to be the full signature. Note that, in $\mathcal{EL}$, ABox consistency is not an issue as every ABox is consistent w.r.t. every TBox. The results in this section can be extended to the extension $\mathcal{EL}_\bot$ of $\mathcal{EL}$ with the $\bot$ concept by dealing with inconsistent ABoxes in essentially the same way as was done in Section 3.

Again, we can work with TBoxes in normal form.

Theorem 12. *FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes and the full ABox signature can be polynomially reduced to FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes in normal form and the full ABox signature.*

Note that Theorem 12 differs from Theorem 3 in that we are interested in FO-rewritability of $A(x)$ relative to the normalized TBox $\mathcal{T}'$ *and the full signature* $\mathsf{sig}(\mathcal{T}')$, rather than the original signature $\mathsf{sig}(\mathcal{T})$ as in Theorem 3. In fact, proving Theorem 12 requires a more careful construction.

We now show the surprising result that, when the TBox is formulated in $\mathcal{EL}$ and the ABox signature is full, the tree-shaped ABoxes from Theorem 9 can be replaced with linear ones. Formally, an ABox $\mathcal{A}$ is *linear* if it consists of role assertions $r_0(a_0, a_1), \ldots, r_{n-1}(a_{n-1}, a_n)$ with $a_i \neq a_j$ for $i \neq j$ and concept assertions $A(a)$ with $a \in \{a_0, \ldots, a_n\}$.

Theorem 13. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox in normal form, $A(x)$ an AQ, $n = |\mathcal{T}|$, and $k = 2^{3n^2}$. Then $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full ABox signature iff for all linear ABoxes $\mathcal{A}$ with root $\rho_\mathcal{A}$, $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ implies $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$.*

Proof. (sketch) In light of Theorem 9, it is sufficient to show that if there is a tree-shaped ABox $\mathcal{A}$ with root $\rho_\mathcal{A}$ such that $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ and $\mathcal{A}|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$ with $k = 2^{3n^2}$, then there is a linear ABox $\mathcal{A}'$ that satisfies the same properties. Since $\mathcal{EL}$ does not allow inverse roles, we can assume w.l.o.g. that $\mathcal{A}$ has the shape of a *directed* tree, that is, whenever $r(a, b) \in \mathcal{A}$, then b is further away from the root b than a. Note that the depth of $\mathcal{A}$ must exceed k. We can further assume that $B(b) \in \mathcal{A}$ whenever $\mathcal{A}|_k, \mathcal{T} \models B(b)$, since these assertions can be added without changing the relevant properties of $\mathcal{A}$. By replacing subtrees with the concept assertions that they entail, we can further ensure that there is only a single individual on level $k + 1$, and no individuals on any level $> k + 1$. The desired linear ABox $\mathcal{A}'$ is then defined as the restriction of $\mathcal{A}$ to assertions that involve only the individuals that appear on the unique path in $\mathcal{A}$ of length $k + 1$. Since $\mathcal{A}'|_k \subseteq \mathcal{A}|_k$, we have $\mathcal{A}'|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$. Since all assertions entailed by $\mathcal{A}|_k$ and $\mathcal{T}$ appear in $\mathcal{A}$ and all individuals in $\mathcal{A} \setminus \mathcal{A}'$ are on level at most k, we have $\mathcal{A}', \mathcal{T} \models A(\rho_\mathcal{A})$. The latter argument relies on $\mathcal{A}$ being a directed tree. $\Box$

The following example shows that, even for $\mathcal{EL}$, it is not possible to replace tree-shaped ABoxes with linear ones if we are interested in signatures other than the full signature.

Example 14. *Let*

$$\mathcal{T} = \{A_i \sqsubseteq X_i,\ B_i \sqcap X_i \sqsubseteq Y_i,\ \exists r.Y_i \sqsubseteq X_i \mid i \in \{1,2\}\} \cup \{X_1 \sqcap X_2 \sqsubseteq X,\ B_1 \sqcap B_2 \sqsubseteq Z,\ \exists r.Z \sqsubseteq X\},$$

choose $\Sigma = \{A_1, A_2, B_1, B_2, r\}$, and take the AQ $X(x)$. The tree-shaped ABox $\mathcal{A}$ composed of the assertions

$$\{r(a_0, a_{i,0}), r(a_{i,0}, a_{i,1}), ..., r(a_{i,2^{3n^2}}, a_{i,2^{3n^2}+1}) \mid i \in \{1,2\}\}$$
$$\cup \{B_i(a_{i,0}), \ldots, B_i(a_{i,2^{3n^2}+1}), A_i(a_{i,2^{3n^2}+1}) \mid i \in \{1,2\}\},$$

with n as in Theorem 9, is of depth exceeding 2^{3n^2}, and we can show that $\mathcal{A}, \mathcal{T} \models X(a_0)$, but $\mathcal{A}|_{2^{3n^2}}, \mathcal{T} \not\models X(a_0)$. However, for all linear Σ-ABoxes $\mathcal{A}$, we have $\mathcal{A}, \mathcal{T} \models X(a_0)$ iff $\mathcal{A}|_1, \mathcal{T} \models X(a_0)$: since $X, X_1, X_2 \notin \Sigma$, we can only have $\mathcal{A}, \mathcal{T} \models X(a_0)$ if there is an r-successor b of a_0 in $\mathcal{A}$ where Z is entailed, or where Y_1 and Y_2 are entailed. Since $Y_1, Y_2, Z \notin \Sigma$, this in turn can only be the case when $B_1(b), B_2(b) \in \mathcal{A}$. But then, $\mathcal{A}|_1, \mathcal{T} \models X(a_0)$.

Theorem 13 allows us to replace the alternating tree automata in the proof of Theorem 10 with alternating word automata, improving the upper bound to PSPACE.

Theorem 15. *Deciding FO-rewritability of an AQ relative to an $\mathcal{EL}$ TBox and the full ABox signature is in* PSPACE.

We establish matching lower bounds.

Theorem 16. *Deciding FO-rewritability of an AQ relative to an $\mathcal{EL}$ TBox and an ABox signature Σ is (1)* PSPACE*-hard when Σ is full and (2)* EXPTIME*-hard when Σ is an input.*

Point 1 is proved by a reduction from the word problem for polynomially space-bounded deterministic Turing machines. For Point 2, we use polynomially space-bounded alternating Turing machines. As both reductions are lengthy and somewhat subtle, we present instead a proof of coNP-hardness, which illustrates some general ideas that are also used in the other reductions.

We reduce propositional tautology to FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes and the full signature. Let ϑ be a propositional formula in negation normal form with variables $p_1, \ldots, p_n$, and let $\mathsf{sub}(\vartheta)$ be the set of subformulas of ϑ. Define a TBox $\mathcal{T}$ with the CIs:

$$\begin{array}{rcll}
\exists r.(L_i \sqcap V_{i,i}) & \sqsubseteq & L_{i-1} & V \in \{T,F\}, 1 \leq i \leq n \\
\exists r.V_{i,j} & \sqsubseteq & V_{i,j-1} & V \in \{T,F\}, 1 \leq j \leq i \leq n \\
T_{i,0} & \sqsubseteq & A_{p_i} & p_i \in \mathsf{sub}(\vartheta) \\
F_{i,0} & \sqsubseteq & A_{\neg p_i} & \neg p_i \in \mathsf{sub}(\vartheta) \\
A_\varphi \sqcap A_\psi & \sqsubseteq & A_{\varphi \wedge \psi} & \varphi \wedge \psi \in \mathsf{sub}(\vartheta) \\
A_\rho & \sqsubseteq & A_{\varphi \vee \psi} & \rho \in \{\varphi, \psi\}, \varphi \vee \psi \in \mathsf{sub}(\vartheta) \\
A_\vartheta & \sqsubseteq & L_0 & \\
\exists r.L_0 & \sqsubseteq & L_n &
\end{array}$$

Lemma 17. *ϑ is a tautology iff $L_0(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full signature.*

Proof. First assume that ϑ is not a tautology. Then there is a truth assignment t such that $t \not\models \vartheta$. Let $k = 2^{3|\mathcal{T}|^2}$ and define a linear ABox $\mathcal{A}$ as the union of

$$\{\, r(a_{1,0}, a_{1,1}), \ldots, r(a_{1,n}, a_{2,0}), \ldots, r(a_{k,n-1}, a_{k,n})\}$$
$$\{r(a_{k,n}, a), L_0(a)\,\}$$
$$\{F_{i,j}(a_{\ell,j}) \mid 1 \leq i \leq n,\ 0 \leq j \leq i,\ 1 \leq \ell \leq k,\ t(p_i) = \mathsf{f}\}$$
$$\{T_{i,j}(a_{\ell,j}) \mid 1 \leq i \leq n,\ 0 \leq j \leq i,\ 1 \leq \ell \leq k,\ t(p_i) = \mathsf{t}\}.$$

Starting from the assertion $L_0(a)$, one can derive $L_n(a_{k,n})$, then $L_{n-1}(a_{k,n-1})$, and so on, until one obtains $L_0(a_{1,0})$. Note that the generation of $L_0(a_{1,0})$ cannot be 'shortcut' using the inclusion $A_\vartheta \sqsubseteq L_0$: since $t \not\models \vartheta$, A_ϑ is not derived anywhere on the chain. We thus have $\mathcal{A}, \mathcal{T} \models L_0(a_{1,0})$, but $\mathcal{A}|_k, \mathcal{T} \not\models L_0(a_{1,0})$. By Theorem 13, L_0 is not FO-rewritable relative to $\mathcal{T}$ and the full ABox signature.

Now assume that ϑ is a tautology and that $\mathcal{A}, \mathcal{T} \models L_0(a_0)$ with $\mathcal{A}$ linear and $r(a_0, a_1), \ldots, r(a_{m-1}, a_m)$ the role assertions in $\mathcal{A}$. Assume to the contrary of what is to be shown that $\mathcal{A}|_k, \mathcal{T} \not\models L_0(a_0)$, with k as above. Then $m > k$. By analyzing $\mathcal{T}$, it can be verified that we must have assertions $V_{i,i}(a_i) \in \mathcal{A}$ with $V \in \{T, F\}$, for $1 \leq i \leq n$. These assertions represent a truth assignment t. Since ϑ is valid, we have $t \models \vartheta$. Again analyzing $\mathcal{T}$, this can be used to show that $\mathcal{A}|_k, \mathcal{T} \models L_0(a_0)$, a contradiction. □

5 Related Work

As observed in [Lutz and Wolter, 2011], there is a close connection between FO-rewritability of AQs relative to DL TBoxes and the *boundedness problem* for datalog programs. In fact, the known 2EXPTIME upper bound for predicate boundedness of connected monadic datalog programs [Cosmadakis *et al.*, 1988] can be used to obtain a 3EXPTIME upper bound for FO-rewritability of an AQ relative to an $\mathcal{ELI}$ TBox; via our Proposition 6, this can be extended to $\mathcal{ELI}_\bot$ TBoxes. Boundedness was studied also in the context of the μ-calculus, for which it is EXPTIME-complete [Otto, 1999], and for monadic second order logic [Blumensath *et al.*, 2009]. A different approach to FO-rewritability is suggested in [Bienvenu *et al.*, 2013], based on a connection between query answering in DLs and constraint satisfaction problems (CSPs). This approach is different in spirit from ours and tailored towards expressive DLs of the $\mathcal{ALC}$ family. However, it also yields a NEXPTIME upper bound for FO-rewritability relative to Horn-$\mathcal{SHI}$ TBoxes. Finally, it is shown in [Bienvenu *et al.*, 2012a] that FO-rewritability of AQs relative to a restricted form of $\mathcal{EL}$ TBoxes called *classical TBoxes* is PTIME-complete; that work also analyzes *acyclic TBoxes*, for which FO-rewritings always exist.

6 Future Work

It would be interesting to generalize our approach both regarding the query language and the ontology language covered. Regarding the latter, it would be particularly interesting to generalize our results from Horn-$\mathcal{SHI}$ to Horn-$\mathcal{SHIQ}$, which we conjecture to be possible using slight extensions of the techniques introduced in this paper. Regarding the query language, it would be interesting to analyze FO-rewriting of conjunctive queries. We believe that a mix of techniques from this paper and those in [Bienvenu *et al.*, 2012b] might provide a good starting point. Finally, existing ontologies should be investigated regarding FO-rewritability. Important questions are: How many atomic queries are FO-rewritable w.r.t. natural ABox signatures? How difficult is it to find FO-rewritings if they exist, and how large are they? Interesting ontologies to consider are GALEN and non-acyclic versions of NCI.

Acknowledgements. M. Bienvenu was supported by ANR project PAGODA (ANR-12-JS02-007-01). C. Lutz was supported by the DFG SFB/TR 8 'Spatial Cognition'.

References

[Baader *et al.*, 2003] Franz Baader, Deborah L. McGuiness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.

[Baader *et al.*, 2005] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of IJCAI*, pages 364–369, 2005.

[Baader *et al.*, 2008] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope further (long version). In *Proc. of OWLED workshop*, 2008.

[Baader *et al.*, 2010] Franz Baader, Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Query and predicate emptiness in description logics. In *Proc. of KR*, pages 192–202, 2010.

[Bienvenu *et al.*, 2012a] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Deciding FO-rewritability in $\mathcal{EL}$. In *Proc. of DL workshop*, 2012.

[Bienvenu *et al.*, 2012b] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Query containment in description logics reconsidered. In *Proc. of KR*, 2012.

[Bienvenu *et al.*, 2013] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive datalog, csp, and mmsnp. In *Proc. of PODS*, 2013.

[Blumensath *et al.*, 2009] Achim Blumensath, Martin Otto, and Mark Weyer. Boundedness of monadic second-order formulae over finite words. In *Proc. of ICALP*, pages 67–78, 2009.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. of Automated Reasoning*, 39(3):385–429, 2007.

[Chortaras *et al.*, 2011] Alexandros Chortaras, Despoina Trivela, and Giorgos B. Stamou. Optimized query rewriting for OWL 2 QL. In *Proc. of CADE*, pages 192–206, 2011.

[Cosmadakis *et al.*, 1988] Stavros S. Cosmadakis, Haim Gaifman, Paris C. Kanellakis, and Moshe Y. Vardi. Decidable optimization problems for database logic programs (preliminary report). In *Proc. of STOC*, pages 477–490, 1988.

[Eiter *et al.*, 2008] Thomas Eiter, Georg Gottlob, Magdalena Ortiz, and Mantas Simkus. Query answering in the description logic Horn-$\mathcal{SHIQ}$. In *Proc. of JELIA*, pages 166–179, 2008.

[Eiter *et al.*, 2012] Thomas Eiter, Magdalena Ortiz, Mantas Simkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for Horn-$\mathcal{SHIQ}$ plus rules. In *Proc. of AAAI*, 2012.

[Gottlob *et al.*, 2011] Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological queries: Rewriting and optimization. In *Proc. of ICDE*, pages 2–13, 2011.

[Hustadt *et al.*, 2007] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Reasoning in description logics by a reduction to disjunctive datalog. *J. of Automated Reasoning*, 39(3):351–384, 2007.

[Kazakov, 2009] Yevgeny Kazakov. Consequence-driven reasoning for Horn-$\mathcal{SHIQ}$ ontologies. In Craig Boutilier, editor, *Proc. of IJCAI*, pages 2040–2045, 2009.

[Lutz and Wolter, 2010] Carsten Lutz and Frank Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. of Symbolic Computation*, 45(2):194–228, 2010.

[Lutz and Wolter, 2011] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of DL workshop*, 2011.

[Lutz and Wolter, 2012] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of KR*, 2012.

[Lutz *et al.*, 2009] Carsten Lutz, David Toman, and Frank Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *Proc. of IJCAI*, pages 2070–2075, 2009.

[Nikitina and Rudolph, 2012] Nadeschda Nikitina and Sebastian Rudolph. Expexpexplosion: Uniform interpolation in general $\mathcal{EL}$ terminologies. In *Proc. of ECAI*, pages 618–623, 2012.

[Otto, 1999] Martin Otto. Eliminating recursion in the μ-calculus. In *Proc. of STACS*, pages 531–540, 1999.

[Pérez-Urbina *et al.*, 2009] Héctor Pérez-Urbina, Ian Horrocks, and Boris Motik. Efficient query answering for OWL 2. In *Proc. of ISWC*, pages 489–504, 2009.

[Rossman, 2008] Benjamin Rossman. Homomorphism preservation theorems. *J. of the ACM*, 55(3), 2008.

[Serre, 2006] Olivier Serre. Parity games played on transition graphs of one-counter processes. In *Proc. of FoSSaCS*, pages 337–351, 2006.

[Vardi, 1998] Moshe Y. Vardi. Reasoning about the past with two-way automata. In *Proc. of ICALP*, pages 628–641, 1998.

A Proofs for Section 3

A.1 Proof of Theorem 3.

Theorem 3 *For every Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$ and ABox signature Σ, one can construct in polynomial time an $\mathcal{ELI}_\bot$ TBox $\mathcal{T}'$ such that for all AQs $A(x)$ with $A \notin \mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$, every FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and Σ, and vice versa.*

The proof is similar to reductions provided in [Hustadt *et al.*, 2007; Kazakov, 2009]. Nevertheless, because [Kazakov, 2009] considers reductions preserving subsumption only and because both [Hustadt *et al.*, 2007] and [Kazakov, 2009] do not reduce to $\mathcal{ELI}_\bot$ TBoxes, we give a detailed proof.

First, we provide a reduction to Horn-$\mathcal{SHI}$ TBoxes whose concept inclusions are in normal form (but which can still contain role inclusions and transitivity assertions).

Lemma 18. *For every Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$ and ABox signature Σ, one can construct in polynomial time a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}'$ whose CIs form an $\mathcal{ELI}_\bot$ TBox in normal form such that for all AQs $A(x)$ with $A \notin \mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$, every FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and Σ, and vice versa.*

Proof. The following rules can be used to rewrite $\mathcal{T}$ into an $\mathcal{ELI}_\bot$ TBox in normal form (all freshly introduced concept names are not in $\mathsf{sig}(\mathcal{T}) \cup \Sigma \cup \{A\}$):

- If L is of the form $L_1 \sqcap L_2$ and R is not a concept name, then take a fresh concept name A and replace $L \sqsubseteq R$ by $L \sqsubseteq A$ and $A \sqsubseteq R$. If R is a concept name, and either L_1 or L_2 are not concept names, then take fresh concept names A_1, A_2 and replace $L \sqsubseteq R$ by $L_1 \sqsubseteq A_1$, $L_2 \sqsubseteq A_2$ and $A_1 \sqcap A_2 \sqsubseteq R$;
- If L is of the form $L_1 \sqcup L_2$ and R is a concept name, then replace $L \sqsubseteq R$ by $L_1 \sqsubseteq R$ and $L_2 \sqsubseteq R$. Otherwise take a fresh concept name A and replace $L \sqsubseteq R$ by $L \sqsubseteq A$ and $A \sqsubseteq R$;
- If L is of the form $\exists r.L'$ and L' is not a concept name, then take a fresh concept name A' and replace $L \sqsubseteq R$ by $L' \sqsubseteq A'$ and $\exists r.A' \sqsubseteq R$;
- If R is of the form $\neg A$, then replace $L \sqsubseteq R$ by $L \sqcap A \sqsubseteq \bot$;
- If R is of the form $R_1 \sqcap R_2$ and L is not a concept name, then take a fresh concept name A and replace $L \sqsubseteq R$ by $L \sqsubseteq A$ and $A \sqsubseteq R$. Otherwise take fresh concept names A_1, A_2 and replace $L \sqsubseteq R$ by $L \sqsubseteq A_1$, $L \sqsubseteq A_2$, $A_1 \sqsubseteq R_1$, and $A_2 \sqsubseteq R_2$;
- If R is of the form $\neg L' \sqcup R'$, then replace $L \sqsubseteq R$ by $L \sqcap L' \sqsubseteq R'$;
- If R is of the form $\exists r.R'$ and R' is not a concept name, then take a fresh concept name A' and replace $L \sqsubseteq R$ by $L \sqsubseteq \exists r.A'$ and $A' \sqsubseteq R'$;
- If R is of the form $\forall r.R'$, then replace $L \sqsubseteq R$ by $\exists r^-.L \sqsubseteq R$.

The resulting TBox $\mathcal{T}'$ is as required. In particular, for every Σ-ABox $\mathcal{A}$ and model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}'$, we have that $\mathcal{I}$ is also a model of $\mathcal{T}$; conversely, every model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ can be extended to a model of $\mathcal{T}$ by appropriately interpreting the fresh concept names. Consequently, we have $\mathsf{cert}_{\mathcal{T}}(A(x), \mathcal{A}) = \mathsf{cert}_{\mathcal{T}'}(A(x), \mathcal{A})$ and thus, every FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and Σ, and vice versa. $\square$

Now we show how transitivity assertions and role inclusions can be eliminated. Let $\mathcal{T}$ be a Horn-$\mathcal{SHI}$ TBox whose CIs are $\mathcal{ELI}_\bot$-inclusions in normal form. Add, for any two distinct roles r, s with $\mathcal{T} \models r \sqsubseteq s$ the inclusions $\exists r.B \sqsubseteq \exists s.B$ to $\mathcal{T}$ for any concept name $B \in \mathsf{sig}(\mathcal{T})$ and add for all roles s, r (with $s = r$ not excluded) with $\mathcal{T} \models \mathsf{trans}(r)$ and $\mathcal{T} \models s \sqsubseteq r$, the inclusion $\exists s.\exists r.B \sqsubseteq \exists r.B$ to $\mathcal{T}$ for any concept name $B \in \mathsf{sig}(\mathcal{T})$. Call the resulting TBox without role inclusions and without transitivity statements $\mathcal{T}'$. Clearly $\mathcal{T} \models \mathcal{T}'$. Conversely, we have the following fact:

Lemma 19. *Let $\mathcal{I}$ be a model of $\mathcal{T}'$. Then there exists a model $\mathcal{J}$ of $\mathcal{T}$ such that $\Delta^{\mathcal{I}} = \Delta^{\mathcal{J}}$, $B^{\mathcal{I}} = B^{\mathcal{J}}$ for all concept names B, and $r^{\mathcal{J}} \supseteq r^{\mathcal{I}}$ for all role names r.*

Proof. $\mathcal{J}$ is obtained from $\mathcal{I}$ by applying the following two rules recursively:

- if $(d, d') \in r^{\mathcal{I}}$ and $\mathcal{T} \models r \sqsubseteq s$, then update $\mathcal{I}$ by adding the pair (d, d') to $s^{\mathcal{I}}$;
- if $(d, d'), (d', d'') \in r^{\mathcal{I}}$, and $\mathsf{trans}(r) \in \mathcal{T}$, then update $\mathcal{I}$ by adding the pair (d, d'') to $\mathcal{I}$.

Now one can show by induction over rule applications: if $\mathcal{I}$ satisfies $\mathcal{T}'$ and $\mathcal{I}'$ is the result of applying any of the two rules above to $\mathcal{I}$, then $\mathcal{I}'$ satisfies $\mathcal{T}'$. To this end, it is sufficient to show:

1. for every $\exists r.B$ with B a concept name in $\mathcal{T}$ and every $d \in \Delta^{\mathcal{I}}$: $d \in (\exists r.B)^{\mathcal{I}}$ iff $d \in (\exists r.B)^{\mathcal{I}'}$;
2. all inclusions $\exists s.\exists r.B \sqsubseteq \exists r.B$ that have been added to $\mathcal{T}$ hold in $\mathcal{I}'$ if they hold in $\mathcal{I}$.

The proof is straightforward and omitted. $\square$

It follows from Lemma 19 that for any Σ-ABox $\mathcal{A}$, $a \in \mathsf{Ind}(\mathcal{A})$, and AQ $A(x)$: $\mathcal{A}, \mathcal{T} \models A(a)$ iff $\mathcal{A}, \mathcal{T}' \models A(a)$.

To prove Theorem 3, it now remains to replace the fresh CIs of $\mathcal{T}'$ that are not in normal form by inclusions in normal form in the same way as in the proof of Lemma 18.

A.2 Proof of Theorem 7

Theorem 7 *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox, Σ an ABox signature, and $A(x)$ an AQ.*

1. *$A(x)$ is FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes iff there is a $k \geq 0$ such that for all tree-shaped Σ-ABoxes $\mathcal{A}$ with root $\rho_{\mathcal{A}}$ that are consistent w.r.t. $\mathcal{T}$: if $\mathcal{A}, \mathcal{T} \models A(\rho_{\mathcal{A}})$, then $\mathcal{A}|_k, \mathcal{T} \models A(\rho_{\mathcal{A}})$;*
2. *ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and Σ iff there is a $k \geq 0$ such that for all tree-shaped Σ-ABox $\mathcal{A}$ with root $\rho_{\mathcal{A}}$: if $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ and $\mathcal{A} \setminus \{\rho_{\mathcal{A}}\}$ is consistent w.r.t. $\mathcal{T}$, then $\mathcal{A}|_k$ is inconsistent w.r.t. $\mathcal{T}$.*

Proof. We start with Point 1. For the "only if" direction, assume that $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes, and let φ be a concrete FO-rewriting which, by Proposition 4, we can assume to be a UCQ. Let k be the maximum number of atoms in any CQ in φ. We show that k is the required bound: let $\mathcal{A}$ be a tree-shaped Σ-ABox that is consistent w.r.t. $\mathcal{T}$ with root $\rho_\mathcal{A}$ and such that $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$. We show that $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$. We have $\mathcal{I}_\mathcal{A} \models \varphi[\rho_\mathcal{A}]$. Let $\mathcal{A}'$ be obtained from $\mathcal{A}$ by taking the disjoint union of $\mathcal{A}|_k$ and $\mathcal{A}$, with the individual names of $\mathcal{A}$ renamed. Since $\mathcal{I}_\mathcal{A} \models \varphi[\rho_\mathcal{A}]$ and the number of atoms in each CQ of φ is bounded by k, we have $\mathcal{I}_{\mathcal{A}'} \models \varphi[\rho_\mathcal{A}]$ (note that adding the renamed version of $\mathcal{A}$ is necessary since the CQs in φ need not be connected). Consequently, $\mathcal{A}', \mathcal{T} \models A(\rho_\mathcal{A})$, which implies $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$ since $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$.

For the "if" direction, we first establish the following claim.

Claim 1. If there is a UCQ-rewriting of $A(x)$ relative to $\mathcal{T}$ and acyclic Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$, then there is a UCQ-rewriting of $A(x)$ relative to $\mathcal{T}$ and arbitrary Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$.

For the proof of the claim, assume that $\varphi(x) = q_1 \vee \cdots \vee q_n$ is a UCQ-rewriting of $A(x)$ relative to $\mathcal{T}$ and acyclic Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$. We can w.l.o.g. assume all CQs q_i in $\varphi(x)$ to be acyclic. To see this, note first that, when a CQ q_i has a match in an acyclic interpretation $\mathcal{I}_\mathcal{A}$, then this match gives rise to an acyclic CQ q_i' that is obtained from q_i by identifying variables and also has a match in $\mathcal{I}_\mathcal{A}$. As a consequence, we can replace each CQ q_i in $\varphi(x)$ with the disjunction of all acyclic CQs that can be obtained from q_i by identifying variables, and the resulting $\varphi(x)$ will still be an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and acyclic Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$.

We show that, then, $\varphi(x)$ is a rewriting of $A(x)$ relative to $\mathcal{T}$ and arbitrary Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$. Let $\mathcal{A}$ be a Σ-ABox that is consistent w.r.t. $\mathcal{T}$. If $\mathcal{I}_\mathcal{A} \models \varphi[a]$, then $\mathcal{I}_\mathcal{A} \models q_i[a]$ for some i. Let $\mathcal{A}_{q_i}$ be the acyclic Σ-ABox that corresponds to the CQ q_i, that is, each variable y of q_i gives rise to an individual name a_y in $\mathcal{A}_{q_i}$. Assume first that x is a free variable in q_i. We trivially have $\mathcal{A}_{q_i}, \mathcal{T} \models A[a_x]$ and since the match of q_i in $\mathcal{A}$ provides a homomorphism h from $\mathcal{A}_{q_i}$ to $\mathcal{A}$ with $h(a_x) = a$, this yields $\mathcal{A}, \mathcal{T} \models A[a]$, as required. Assume now that x is not a free variable in q_i. Since q_i has a match in $\mathcal{A}_{q_i}$ it follows that $\mathcal{A}_{q_i}, \mathcal{T} \models A[a_y]$ for all individuals a_y and, moreover, $\mathcal{A}_{q_i} \cup \mathcal{A}' \models A(a')$ for any acyclic Σ-Abox $\mathcal{A}'$ whose individuals are disjoint from $\mathcal{A}_{q_i}$ and any individual a' in $\mathcal{A}'$. But then either $\mathcal{T} \models \top \sqsubseteq A$ or $\mathcal{A}_{q_i}$ is not consistent w.r.t. $\mathcal{T}$. In the first case, Claim 1 is trivial. In the second case $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$ and we have a contradiction to the assumption that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$.

Conversely, assume that $\mathcal{A}, \mathcal{T} \models A(a)$. It was observed in [Lutz and Wolter, 2012] that every $\mathcal{ELI}_\bot$ TBox is unraveling tolerant, which roughly means that $\mathcal{A}, \mathcal{T} \models A(a)$ iff $\mathcal{A}^*, \mathcal{T} \models A(a)$, with $\mathcal{A}^*$ the unraveling of $\mathcal{A}$ into a forest. Thus, we can unravel the ABox $\mathcal{A}$ into a (possibly infinite) acyclic Σ-ABox $\mathcal{A}^*$ with $\mathcal{A}^*, \mathcal{T} \models A(a)$. By compactness of first-order logic, there is a finite subset $\mathcal{A}'$ of $\mathcal{A}^*$ such that $\mathcal{A}', \mathcal{T} \models A(a)$. Note that $\mathcal{A}'$ is acyclic (but will typically not be tree-shaped). Hence $\mathcal{I}_{\mathcal{A}'} \models q[a]$. By definition of unraveling, there is a homomorphims h from $\mathcal{I}_{\mathcal{A}'}$ to $\mathcal{I}_\mathcal{A}$ with $h(a) = a$. Hence $\mathcal{I}_\mathcal{A} \models q[a]$, as required. This finishes the proof of the claim.

Now assume there is a bound $k \geq 0$ for which there does not exist a tree-shaped Σ-ABox $\mathcal{A}$ with root $\rho_\mathcal{A}$ such that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$, $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$. We consider the set Γ of all minimal tree-shaped Σ-ABoxes $\mathcal{A}$ of depth at most k that are consistent w.r.t. $\mathcal{T}$ and such that $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ and $\mathcal{A}|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$. It is not hard to show that Γ is finite. Each ABox $\mathcal{A} \in \Gamma$ gives rise to a corresponding CQ $q_\mathcal{A}(x)$, where each individual name in $\mathcal{A}$ is viewed as a variable, and the root $\rho_\mathcal{A}$ corresponds to the answer variable x. We aim to show that

$$\varphi(x) = \bigvee_{\mathcal{A} \in \Gamma} q_\mathcal{A}(x)$$

is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and consistent Σ-ABoxes. By Claim 1, it is sufficient to show that $\mathsf{cert}_\mathcal{T}(A(x), \mathcal{A}) = \mathsf{ans}(\mathcal{I}_\mathcal{A}, \varphi(x))$ for all acyclic Σ-ABoxes $\mathcal{A}$ that are consistent w.r.t. $\mathcal{T}$.

"$\supseteq$". Assume that $\mathcal{I}_\mathcal{A} \models \varphi[a]$. Then $\mathcal{I}_\mathcal{A} \models q_\mathcal{B}[a]$ for some $\mathcal{B} \in \Gamma$. Consequently, there is a homomorphism h from $\mathcal{B}$ to $\mathcal{A}$ with $h(\rho_\mathcal{B}) = a$. By definition of Γ, we have $\mathcal{B}, \mathcal{T} \models A(\rho_\mathcal{B})$. Thus, $\mathcal{A}, \mathcal{T} \models A(a)$ as required.

"$\subseteq$". Assume that $\mathcal{A}, \mathcal{T} \models A(a)$ and let $\mathcal{A}' \subseteq \mathcal{A}$ be minimal such that $\mathcal{A}', \mathcal{T} \models A(a)$. Clearly, $\mathcal{A}'$ is acyclic. Since $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$, it can even be shown that $\mathcal{A}'$ is tree-shaped. Thus, $\mathcal{A}' \in \Gamma$ (modulo isomorphism) and therefore $\mathcal{I}_\mathcal{A} \models \varphi[a]$.

We come to Point 2. For the "only if" direction, assume that ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and Σ. Let $\varphi()$ be a concrete rewriting of inconsistency relative to $\mathcal{T}$ and Σ. One can argue as for Proposition 4 that we can w.l.o.g. assume $\varphi()$ to be a UCQ. It can be shown that for every ABox $\mathcal{A}$ that is inconsistent w.r.t. $\mathcal{T}$, there is a connected ABox $\mathcal{A}' \subseteq \mathcal{A}$ that is inconsistent w.r.t. $\mathcal{T}$. Consequently, we can further assume that each disjunct of $\varphi()$ is connected. Let k be the maximum number of atoms in any disjunct of $\varphi()$. Then k is the required bound: let $\mathcal{A}$ be a tree-shaped Σ-ABox with root $\rho_\mathcal{A}$ such that $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ is consistent w.r.t. $\mathcal{T}$ and $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$. Hence $\mathcal{I}_\mathcal{A} \models \varphi()$, but $\mathcal{I}_{\mathcal{A} \setminus \{\rho_\mathcal{A}\}} \not\models \varphi()$. By the former, there is a match of some disjunct q of $\varphi()$ in $\mathcal{I}_\mathcal{A}$. By the latter, $\rho_\mathcal{A}$ is in the range of this match. Since q is connected and has at most k atoms, this yields $\mathcal{I}_{\mathcal{A}|_k} \models \varphi()$ and thus $\mathcal{A}|_k$ is not consistent w.r.t. $\mathcal{T}$.

For the "if" direction, we first establish the following claim.

Claim 2. If there is a UCQ-rewriting of ABox inconsistency relative to $\mathcal{T}$ and acyclic Σ-ABoxes, then there is a UCQ-rewriting of ABox inconsistency relative to $\mathcal{T}$ and arbitrary Σ-ABoxes.

For the proof of the claim, assume that $\varphi = q_1 \vee \cdots \vee q_n$ is a UCQ-rewriting of ABox inconsistency relative to $\mathcal{T}$ and acyclic Σ-ABoxes. As in the proof of Claim 1 we can w.l.o.g.

assume all CQs q_i in φ to be acyclic. Note that no q_i contains any free variable.

We show that φ is a rewriting of ABox inconsistency relative to $\mathcal{T}$ and arbitrary Σ-ABoxes. Let $\mathcal{A}$ be a Σ-ABox. If $\mathcal{I}_\mathcal{A} \models \varphi$, then $\mathcal{I}_\mathcal{A} \models q_i$ for some i. Let $\mathcal{A}_{q_i}$ be the acyclic Σ-ABox that corresponds to the CQ q_i. Then $\mathcal{A}_{q_i}$ is inconsistent w.r.t. $\mathcal{T}$. The match of q_i in $\mathcal{A}$ provides a homomorphism from $\mathcal{A}_{q_i}$ to $\mathcal{A}$ and so $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$.

Conversely, assume that $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ As in the proof of Claim 1 we obtain an unraveling $\mathcal{A}^*$ of $\mathcal{A}$ into a forest that is inconsistent w.r.t. $\mathcal{T}$. By compactness of first-order logic, there is a finite subset $\mathcal{A}'$ of $\mathcal{A}^*$ such that $\mathcal{A}'$ is inconsistent w.r.t. $\mathcal{T}$. $\mathcal{A}'$ is acyclic. Hence $\mathcal{I}_{\mathcal{A}'} \models \varphi$. By definition of unraveling, there is a homomorphims from $\mathcal{I}_{\mathcal{A}'}$ to $\mathcal{I}_\mathcal{A}$. Hence $\mathcal{I}_\mathcal{A} \models \varphi$, as required. This finishes the proof of the claim.

Now assume that there is a bound $k \geq 0$ for which there does not exist a tree-shaped Σ-ABox $\mathcal{A}$ with root $\rho_\mathcal{A}$ such that $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ and $\mathcal{A}|_k$ are consistent w.r.t. $\mathcal{T}$ and $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$. Let Γ' be the set of all minimal tree-shaped Σ-ABoxes $\mathcal{A}$ of depth at most k such that $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$. It is not hard to show that Γ' is finite. Let

$$\varphi() = \bigvee_{\mathcal{A} \in \Gamma'} \exists x. q_\mathcal{A}(x).$$

We show that $\varphi()$ is an FO-rewriting of inconsistency w.r.t. $\mathcal{T}$ and Σ. By Claim 2, it is sufficient to show that for every acyclic Σ-ABox $\mathcal{A}$, we have that $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ iff $\mathcal{I}_\mathcal{A} \models \varphi()$.

"if". Assume that $\mathcal{I}_\mathcal{A} \models \varphi()$. Then $\mathcal{I}_\mathcal{A} \models q_\mathcal{B}[a]$ for some $\mathcal{B} \in \Gamma'$. Consequently, there is a homomorphism h from $\mathcal{B}$ to $\mathcal{A}$. By definition of Γ', this implies that $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$.

"only if". Assume that $\mathcal{A}$ is an acyclic Σ-ABox that is not consistent w.r.t. $\mathcal{T}$. Let $\mathcal{A}' \subseteq \mathcal{A}$ be a minimal subset of $\mathcal{A}$ that is not consistent w.r.t. $\mathcal{T}$. It can be shown that $\mathcal{A}'$ is tree-shaped, thus $\mathcal{A}' \in \Gamma'$ (modulo isomorphism) and so $\mathcal{I}_\mathcal{A} \models \varphi()$. □

A.3 Proof of Theorem 9

For the pumping argument, we require some preparation. A *generalized ABox* is an ABox that can contain assertions of the form $\bot(a)$. Any ABox containing such an assertion is of course inconsistent w.r.t. any TBox. For a generalized ABox $\mathcal{A}$ and individual a, we set

$$\mathcal{A}|_a = \{A(a) \mid A(a) \in \mathcal{A}, A \in \mathsf{N_C} \cup \{\bot\}\}.$$

Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox in normal form and $\mathcal{A}$ a Σ-ABox. We define a sequence of generalized ABoxes $\mathcal{A}_0, \mathcal{A}_1, \dots$ by setting $\mathcal{A}_0 = \mathcal{A}$ and defining $\mathcal{A}_{i+1}$ to be $\mathcal{A}_i$ extended as follows:

(i) for each $\exists r.B \sqsubseteq A \in \mathcal{T}$ such that $r(a,b), B(b) \in \mathcal{A}_i$ and $A(a) \notin \mathcal{A}_i$, add $A(a)$;

(ii) for each $\exists r^-.B \sqsubseteq A \in \mathcal{T}$ such that $r(b,a), B(b) \in \mathcal{A}_i$ and $A(a) \notin \mathcal{A}_i$, add $A(a)$;

(iii) for each $A(a)$ such that $\mathcal{A}_i|_a, \mathcal{T} \models A(a)$ and $A(a) \notin \mathcal{A}_i$, add $A(a)$.

(iv) if $\mathcal{A}_i|_a$ is not consistent w.r.t. $\mathcal{T}$, then add $A(b)$ for any $A \in \Sigma \cup \mathsf{sig}(\mathcal{T}) \cup \{\bot\}$ and any $b \in \mathsf{Ind}(\mathcal{A})$.

The *$\mathcal{T}$-completion* $\mathcal{A}^c_\mathcal{T}$ of $\mathcal{A}$ is the limit of the sequence $\mathcal{A}_0, \mathcal{A}_1, \dots$.

Lemma 20. *For all $A(a)$ with $A \in \Sigma \cup \mathsf{sig}(\mathcal{T}) \cup \{\bot\}$ and $a \in \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{A}, \mathcal{T} \models A(a)$ iff $A(a) \in \mathcal{A}^c_\mathcal{T}$. Moreover, $\bot(a) \in \mathcal{A}^c_\mathcal{T}$ iff $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$.*

Proof. "if". It can be proved by induction on i that $A(a) \in \mathcal{A}_i$ implies $\mathcal{A}, \mathcal{T} \models A(a)$.

"only if". Build a model of $\mathcal{I}$ and $\mathcal{A}^c_\mathcal{T}$ that makes true precisely the concept assertions in $\mathcal{A}^c_\mathcal{T}$ by plugging in tree models that witness existential restrictions. □

For a set $\mathcal{X}$ of concepts and an individual u, we set $\mathcal{X}(u) = \{C(u) \mid C \in \mathcal{X}\}$. Let $\mathcal{A}$ be a tree-shaped Σ-ABox with root $\rho_\mathcal{A}$ and let $u \in \mathsf{Ind}(\mathcal{A})$. Define

$$\mathsf{AT}^{\vdash}_\mathcal{A}(u) := \{A \in \mathsf{N_C} \cup \{\bot\} \mid A(u) \in \mathcal{A}^c_\mathcal{T}\}.$$

Let $\mathcal{A}^\downarrow_u$ denote the subtree of $\mathcal{A}$ rooted at u and let $\mathcal{A}^\uparrow_u$ be the ABox obtained from $\mathcal{A}$ by dropping $\mathcal{A}^\downarrow_u$ from $\mathcal{A}$ except for u itself. Define the *transfer sequence* $\mathcal{X}_0, \mathcal{X}_1, \dots$ *of* $(\mathcal{A}, u)$ *w.r.t.* $\mathcal{T}$ by induction as follows:

- $\mathcal{X}_0 = \mathsf{AT}^{\vdash}_{\mathcal{A}^0}(u)$, where $\mathcal{A}^0 = \mathcal{A}^\uparrow_u$;
- $\mathcal{X}_1 = \mathsf{AT}^{\vdash}_{\mathcal{A}^1}(u)$, where $\mathcal{A}^1 = \mathcal{A}^\downarrow_u \cup \mathcal{X}_0(u)$;
- $\mathcal{X}_{2i+2} = \mathsf{AT}^{\vdash}_{\mathcal{A}^{2i+2}}(u)$, where $\mathcal{A}^{2i+2} = \mathcal{A}^{2i} \cup \mathcal{X}_{2i+1}(u)$, for $i \geq 0$;
- $\mathcal{X}_{2i+1} = \mathsf{AT}^{\vdash}_{\mathcal{A}^{2i+1}}(u)$, where $\mathcal{A}^{2i+1} = \mathcal{A}^{2i-1} \cup \mathcal{X}_{2i}(u)$, for $i \geq 1$.

The sequence of ABoxes $\mathcal{A}^0, \mathcal{A}^1 \dots$ defined above is called the *ABox transfer sequence for* $(\mathcal{A}, u)$ *w.r.t.* $\mathcal{T}$.

Lemma 21. *Let $n = |\mathsf{sig}(\mathcal{T}) \cap \mathsf{N_C}| + 1$. Then $\mathcal{X}_n = \mathcal{X}_m$ for all $m > n$ and $\mathcal{A}^{n-1} \cup \mathcal{A}^n = \mathcal{A}^c_\mathcal{T}$.*

Proof. By definition, $\mathcal{X}_m \subseteq \mathcal{X}_{m+1}$, for all $m > 0$. Moreover, if $\mathcal{X}_{m+1} = \mathcal{X}_m$ for some $m > 0$ then, by Lemma 20,

- all $\mathcal{A}^{(m+1)+2i}$, $i \geq 0$, coincide;
- all $\mathcal{A}^{(m+2)+2i}$, $i \geq 0$, coincide.

It follows that $\mathcal{X}_{m'} = \mathcal{X}_m$ for all $m' > m$. □

Lemma 22. *Let $\mathcal{A}$ and $\mathcal{B}$ be tree-shaped Σ-ABoxes with $a \in \mathsf{Ind}(\mathcal{A})$ and $b \in \mathsf{Ind}(\mathcal{B})$ such that*

- *$\{A \mid A(a) \in \mathcal{A}\} = \{B \mid B(b) \in \mathcal{B}\}$;*
- *the transfer sequence of $(\mathcal{A}, a)$ w.r.t. $\mathcal{T}$ coincides with the transfer sequence of $(\mathcal{B}, b)$ w.r.t. $\mathcal{T}$ and is given by $\mathcal{X}_0, \dots$.*

Denote by $\mathcal{C}$ the ABox obtained from $\mathcal{A}$ by replacing the subtree $\mathcal{A}^\downarrow_a$ by $\mathcal{B}^\downarrow_b$. Then

- *$\mathcal{X}_0, \dots$ is also the transfer sequence of $(\mathcal{C}, b)$ w.r.t. $\mathcal{T}$.*
- *Given the ABox transfer sequences $\mathcal{A}^0, \dots$ and $\mathcal{B}^0, \dots$ of $(\mathcal{A}, a)$ and $(\mathcal{B}, b)$ w.r.t. $\mathcal{T}$, respectively, the ABox transfer sequence $\mathcal{C}^0, \dots$ of $(\mathcal{C}, b)$ w.r.t. $\mathcal{T}$ is given by setting $\mathcal{C}^{2i} = \mathcal{A}^{2i}$ and $\mathcal{C}^{2i+1} = \mathcal{B}^{2i+1}$, for $i \geq 0$.*

Proof. Straightforward using Lemma 20. □

Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$-TBox, Σ an ABox signature, $A(x)$ an AQ, and $k \geq 0$. Then a tree-shaped Σ-ABox $\mathcal{A}$ is a *k-entailment witness* for $\mathcal{T}$, Σ, and $A(x)$ if $\mathcal{A}$ is a tree-shaped Σ-ABox consistent w.r.t. $\mathcal{T}$, $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$, and $\mathcal{A}|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$. $\mathcal{A}$ is a *k-inconsistency witness* for $\mathcal{T}$, Σ, and $A(x)$ if $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ and $\mathcal{A}|_k$ are consistent w.r.t. $\mathcal{T}$, but $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$ (the two points of Theorem 7).

The following lemma is a straightforward reformulation of Part 1 of Theorem 9.

Lemma 23. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox, Σ an ABox signature with $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$, $A(x)$ an AQ, and $n = |\mathcal{T}|$.*

1. *$A(x)$ is not FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes iff there exists a k_0-entailment witness for $\mathcal{T}$, Σ, and $A(x)$ for $k_0 = 2^{3n^2}$.*
2. *ABox inconsistency is not FO-rewritable relative to $\mathcal{T}$ and Σ iff there exists a k_0-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$ for $k_0 = 2^{3n^2}$.*

Proof. (1) The direction ($\Rightarrow$) follows from Theorem 7. Conversely, assume that there is a k_0-entailment witness for $\mathcal{T}$, Σ, and $A(x)$. We show that for every $k > k_0$ there exists a k-entailment witness for $\mathcal{T}$, Σ, and $A(x)$. Then non FO-rewritability of $A(x)$ relative to $\mathcal{T}$ and consistent Σ-ABoxes follows from Point 1 of Theorem 7.

Assume $\mathcal{A}$ is a k-entailment witness for $\mathcal{T}$, Σ, and $A(x)$ for some $k \geq k_0$. It is sufficient to construct a tree-shaped Σ-ABox $\mathcal{A}'$ which is a k'-entailment witness for $\mathcal{T}$, Σ, and $A(x)$ for some $k' > k$. We may assume w.l.o.g. that $\mathcal{A}$ is minimal in the sense that, for every individual a we have $\mathcal{A}_a^-, \mathcal{T} \not\models A(\rho_\mathcal{A})$, where $\mathcal{A}_a^-$ is obtained from $\mathcal{A}$ by dropping the subtree rooted at a (including a).

Let w be a leaf node in $\mathcal{A}$ of maximal distance from $\rho_\mathcal{A}$. Then the distance of w from $\rho_\mathcal{A}$ is at least $k+1$. Since by Lemma 21 the number of transfer sequences w.r.t. $\mathcal{T}$ does not exceed 2^{n^2}, on the path from $\rho_\mathcal{A}$ to w there must be at least two individuals u_1 and u_2 for which $\{B \mid B(u_1) \in \mathcal{A}\} = \{B \mid B(u_2) \in \mathcal{A}\}$ and

- the transfer sequences of $(\mathcal{A}, u_1)$ and $(\mathcal{A}, u_2)$ w.r.t. $\mathcal{T}$ coincide;
- the transfer sequences of $(\mathcal{A}|_k, u_1)$ and $(\mathcal{A}|_k, u_2)$ w.r.t. $\mathcal{T}$ coincide.

We may assume that u_1 is between $\rho_\mathcal{A}$ and u_2. Let $\mathcal{A}'$ be the ABox obtained from $\mathcal{A}$ by replacing $\mathcal{A}_{u_2}^\downarrow$ by $\mathcal{A}_{u_1}^\downarrow$ in $\mathcal{A}$. By renaming nodes in $\mathcal{A}_{u_1}^\downarrow$, we can assume that the root of the subtree $\mathcal{A}_{u_1}^\downarrow$ of $\mathcal{A}'$ is denoted by u_2. It follows from Lemma 22 that $\mathcal{A}'$ is consistent w.r.t. $\mathcal{T}$, that $\mathcal{A}', \mathcal{T} \models A(\rho_\mathcal{A})$, and that $\mathcal{A}'', \mathcal{T} \not\models A(\rho_\mathcal{A})$ for the ABox $\mathcal{A}''$ obtained from $\mathcal{A}'$ by removing the subtree rooted at w. Clearly $\mathcal{A}'' \supseteq \mathcal{A}'|_{k'}$ for $k' = k+1$. Thus, $\mathcal{A}'$ is a k'-entailment witness for $\mathcal{T}$, Σ, and $A(x)$ for some $k' > k$, as required.

(2) The proof is similar to the proof of (1). Again, the direction ($\Rightarrow$) follows from Theorem 7.

Conversely, assume that there is a k_0-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$. We show that for every $k > k_0$ there exists a k-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$. Then non FO-rewritability of ABox inconsistency follows from Point 2 of Theorem 7.

Assume $\mathcal{A}$ is a k-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$ for some $k \geq k_0$. It is sufficient to construct a tree-shaped Σ-ABox $\mathcal{A}'$ which is a k'-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$ for some $k' > k$. We may assume w.l.o.g. that $\mathcal{A}$ is minimal in the sense that, for every individual a we have $\mathcal{A}_a^-$ is consistent w.r.t. $\mathcal{T}$.

Let w be a leaf node in $\mathcal{A}$ of maximal distance from $\rho_\mathcal{A}$. Then the distance of w from $\rho_\mathcal{A}$ is at least $k+1$. Since by Lemma 21 the number of transfer sequences w.r.t. $\mathcal{T}$ does not exceed 2^{n^2}, on the path from $\rho_\mathcal{A}$ to w there must be at least two individuals u_1 and u_2 for which $\{B \mid B(u_1) \in \mathcal{A}\} = \{B \mid B(u_2) \in \mathcal{A}\}$ and

- the transfer sequences of $(\mathcal{A}, u_1)$ and $(\mathcal{A}, u_2)$ w.r.t. $\mathcal{T}$ coincide;
- the transfer sequences of $(\mathcal{A}|_k, u_1)$ and $(\mathcal{A}|_k, u_2)$ w.r.t. $\mathcal{T}$ coincide;
- the transfer sequences of $(\mathcal{A} \setminus \{\rho_\mathcal{A}\}, u_1)$ and $(\mathcal{A} \setminus \{\rho_\mathcal{A}\}, u_2)$ w.r.t. $\mathcal{T}$ coincide.

We proceed in the same way as above: we may assume that u_1 is between $\rho_\mathcal{A}$ and of u_2. Let $\mathcal{A}'$ be the ABox obtained from $\mathcal{A}$ by replacing $\mathcal{A}_{u_2}^\downarrow$ by $\mathcal{A}_{u_1}^\downarrow$ in $\mathcal{A}$. By renaming nodes in $\mathcal{A}_{u_1}^\downarrow$, we can assume that the root of the subtree $\mathcal{A}_{u_1}^\downarrow$ of $\mathcal{A}'$ is denoted by u_2. It follows from Lemma 22 that $\mathcal{A}'$ is not consistent w.r.t. $\mathcal{T}$, that $\mathcal{A}' \setminus \{\rho'_\mathcal{A}\}$ is consistent w.r.t. $\mathcal{T}$, and that $\mathcal{A}''$ is consistent w.r.t. $\mathcal{T}$ for the ABox $\mathcal{A}''$ obtained from $\mathcal{A}'$ by removing the subtree rooted at w. Again $\mathcal{A}'' \supseteq \mathcal{A}'|_{k'}$ for $k' = k+1$. Thus, $\mathcal{A}'$ is a k'-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$ for some $k' > k$, as required. □

The following proposition implies Part 2 of Theorem 9.

Proposition 24. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox in normal form, $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$ an ABox signature, $A(x)$ an AQ, and $k \geq 0$. Then*

1. *there exists a k-entailment witness for $\mathcal{T}$, Σ, and $A(x)$ iff there exists such a witness of outdegree at most $|\mathcal{T}|$;*
2. *there is a k-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$ iff there is such a witness of outdegree at most $|\mathcal{T}|$.*

Proof sketch. For Point 1, let $\mathcal{A}$ be a k-entailment witness for $\mathcal{T}$, Σ, and $A(x)$. We have to show that there is also such a witness of outdegree at most $|\mathcal{T}|$. This is done by marking certain individuals in $\mathcal{A}$ in an appropriate way, then drop all individuals that are not marked, and finally showing the the resulting ABox is as required.

Let $\mathcal{A} = \mathcal{A}_0, \mathcal{A}_1, \ldots$ be the sequence of ABoxes with limit $\mathcal{A}_\mathcal{T}^c$ defined in Appendix A. Recall that, by Lemma 20, $\mathcal{A}, \mathcal{T} \models B(a)$ iff $B(a) \in \mathcal{A}_\mathcal{T}^c$ for all assertions $B(a)$. Therefore, each $B(a)$ with $\mathcal{A}, \mathcal{T} \models B(a)$ is associated with a number $\mu(B(a))$ which is minimal such that $B(a) \in \mathcal{A}_{\mu(B(a))}$ holds.

Now, the marking of individuals in $\mathcal{A}$ is as follows: for all $a_1, a_2 \in \mathsf{Ind}(\mathcal{A})$ and $\exists r.A_2 \sqsubseteq A_1$ such that $r(a_1, a_2) \in \mathcal{A}$ with a_2 a successor of a_1 in the tree-shaped ABox $\mathcal{A}$ and

$\mathcal{A}, \mathcal{T} \models A_2(a_2)$, mark an individual a_2 with the stated properties and such that $\mu(A_2(a_2))$ is minimal among all individuals a_2 satisfying these properties. Also mark the root $\rho_\mathcal{A}$ of $\mathcal{A}$.

Now let $\mathcal{A}'$ be the restriction of $\mathcal{A}$ to all individuals a such that, on the unique path from $\rho_\mathcal{A}$ to a, there are only marked individuals. Let $\mathcal{A}' = \mathcal{A}'_0, \mathcal{A}'_1, \dots$ be the sequence of ABoxes with limit $\mathcal{A}'^c_\mathcal{T}$ defined in Appendix A, but now starting with $\mathcal{A}'$. It can be proved by induction on i that for all $a \in \mathsf{Ind}(\mathcal{A}')$, concept names B, and $i \geq 0$, we have $B(a) \in \mathcal{A}_i$ iff $B(a) \in \mathcal{A}'_i$ and $B(a) \in \mathcal{A}_i|_k$ iff $B(a) \in \mathcal{A}'_i|_k$. Since the outdegree of $\mathcal{A}'$ is bounded by $|\mathcal{T}|$ (assuming that only a minimal number of nodes is marked), $\mathcal{A}'$ is the desired k-entailment witness for $\mathcal{T}$, Σ, and $A(x)$ of outdegree at most $|\mathcal{T}|$.

Point 2 is proved using essentially the same argument. The only difference is that, instead of marking the root, one now has to mark all nodes from the root to some chosen individual a such that there is an $A(a) \in \mathcal{A}^c_\mathcal{T}$ with $A \sqsubseteq \bot \in \mathcal{T}$; otherwise, the constucted $\mathcal{A}'$ might not be inconsistent as desired. $\square$

A.4 Proof of Theorem 10

We introduce two-way alternating Büchi automata on finite trees (TWABAs). Let $\mathbb{N}$ denote the *positive* integers. A *tree* is a non-empty (and potentially infinite) set $T \subseteq \mathbb{N}^*$ closed under prefixes. The node ε is the *root* of T. As a convention, we take $x \cdot 0 = x$ and $(x \cdot c) \cdot -1 = x$. Note that $\varepsilon \cdot -1$ is undefined. We say that T is *m-ary* if for every $x \in T$, the set $\{i \mid x \cdot i \in T\}$ is of cardinality at most m. W.l.o.g., we assume that all nodes in an m-ary tree are from $\{1, \dots, m\}^*$. An *infinite path* P of T is a prefix-closed set $P \subseteq T$ such that for every $i \geq 0$, there is a unique $x \in P$ with $|x| = i$.

We use $[m]$ to denote the set $\{-1, 0, \dots, m\}$ and for any set X, let $\mathcal{B}^+(X)$ denote the set of all positive Boolean formulas over X, i.e., formulas built using conjunction and disjunction over the elements of X used as propositional variables, and where the special formulas true and false are allowed as well. For an alphabet Γ, a *Γ-labeled tree* is a pair (T, V) with T a tree and $V : T \to \Gamma$ a node labeling function.

Definition 25 (TWABA). *A* two-way alternating Büchi automaton (TWABA) on finite m-ary trees *is a tuple* $\mathfrak{A} = (Q, \Gamma, \delta, q_0, R)$ *where Q is a finite set of* states, *Γ is a finite alphabet, $\delta : Q \times \Gamma \to \mathcal{B}^+(\mathsf{tran}(\mathfrak{A}))$ is the* transition function *with $\mathsf{tran}(\mathfrak{A}) = [m] \times Q$ the set of* transitions *of $\mathfrak{A}$, $q_0 \in Q$ is the* initial state, *and $R \subseteq Q$ is a set of* recurring states.

Intuitively, a transition (i, q) with $i > 0$ means that a copy of the automaton in state q is sent to the i-th successor of the current node, which is then required to exist. Similarly, $(0, q)$ means that the automaton stays at the current node and switches to state q, and $(-1, q)$ indicates moving to the predecessor of the current node.

Definition 26 (Run, Acceptance). *A* run *of a TWABA $\mathfrak{A} = (Q, \Gamma, \delta, q_0, R)$ on a finite Γ-labeled tree (T, V) is a $T \times Q$-labeled tree (T_r, r) such that the following conditions are satisfied:*

1. $r(\varepsilon) = (\varepsilon, q_0)$
2. *if $y \in T_r$, $r(y) = (x, q)$, and $\delta(q, V(x)) = \varphi$, then there is a (possibly empty) set $S = \{(c_1, q_1), \dots, (c_n, q_n)\} \subseteq \mathsf{tran}(\mathfrak{A})$ such that S satisfies φ and for $1 \leq i \leq n$, $x \cdot c_i$ is defined and a node in T, and there is a $y \cdot i \in T_r$ such that $r(y \cdot i) = (x \cdot c_i, q_i)$.*

We say that (T_r, r) is accepting *if in all infinite paths $\varepsilon = y_1 y_2 \cdots$ of T_r, the set $\{i \geq 0 \mid r(y_i) = (x, q)$ for some $q \in R\}$ is infinite.*

A finite Γ-labeled tree (T, V) is accepted *by $\mathfrak{A}$ if there is an accepting run of $\mathfrak{A}$ on (T, V). We use $L(\mathfrak{A})$ to denote the set of all finite Γ-labeled tree accepted by $\mathfrak{A}$.*

It is known (and easy to see) that TWABAs are closed under complementation, i.e., for any TWABA $\mathfrak{A}$ over finite Γ-labeled m-ary trees, there is a TWABA $\overline{\mathfrak{A}}$ over finite Γ-labeled m-ary trees such that $L(\overline{\mathfrak{A}})$ is the set of those finite Γ-labeled m-ary trees (T, V) such that $(T, V) \notin L(\mathfrak{A})$. It is also known that TWABAs are closed under union and intersection, and that their emptiness problem is ExpTime-complete.

Theorem 27 ([Vardi, 1998])**.** *Given a TWABA $\mathfrak{A}$, it is* ExpTime*-complete to decide whether $L(\mathfrak{A}) = \emptyset$.*

Let Σ be an ABox signature, let $\mathsf{rol}(\Sigma)$ denote the set of (possibly inverse) roles over role names from Σ, and let Γ_Σ be the alphabet that consists of all sets $\sigma \in 2^{\Sigma \cup \mathsf{rol}(\Sigma)}$ which contain exactly one element of $\mathsf{rol}(\Sigma)$. Each Γ_Σ-labeled finite tree (T, V) *represents* the Σ-ABox $\mathcal{A}$

$$\{A(a_x) \mid x \in T \text{ and } A \in V(x)\} \cup$$
$$\{r(a_x, a_{x \cdot c}) \mid x \cdot c \in T, r \text{ a role name, and } r \in V(x \cdot c)\} \cup$$
$$\{r(a_{x \cdot c}, a_x) \mid x \cdot c \in T, r \text{ a role name, and } r^- \in V(x \cdot c)\}$$

with root a_ε (the role in the label of the root node is ignored). Clearly, every Σ-ABox except the empty ABox is represented by some finite Γ_Σ-labeled tree. Note that these translations preserve the outdegree. As the next step, we aim to prove the following.

Proposition 28. *For every $\mathcal{ELI}_\bot$ TBox $\mathcal{T}$ in normal form, ABox signature Σ, AQ $A(x)$, and $m > 0$, there is*

1. *a TWABA $\mathfrak{A}_{\mathcal{T},\Sigma,A,m}$ over finite m-ary Γ_Σ-trees such that $(T, V) \in L(\mathfrak{A}_{\mathcal{T},\Sigma,A,m})$ iff (T, V) represents a tree-shaped Σ-ABox $\mathcal{A}$ such that $\mathcal{A}, \mathcal{T} \models A(a_\varepsilon)$;*
2. *a TWABA $\mathfrak{A}_{\mathcal{T},\Sigma,\bot,m}$ over m-ary Γ_Σ-trees such that $(T, V) \in L(\mathfrak{A}_{\mathcal{T},\Sigma,A,m})$ iff (T, V) represents a tree-shaped Σ-ABox $\mathcal{A}$ such that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$.*

Moreover, the mentioned automata can be constructed in time exponential in the size of $\mathcal{T}$, Σ, and m, and the number of states is $\mathcal{O}(|\mathcal{T}|)$ in both cases.

To prove Proposition 28, it is convenient to first characterize entailment of AQs in terms of derivation trees. Fix an $\mathcal{ELI}_\bot$ TBox $\mathcal{T}$ in normal form and an ABox $\mathcal{A}$. Let $\mathcal{T}_\exists$ be the restriction of $\mathcal{T}$ to CIs of the form $\exists r.A \sqsubseteq B$, with r a (potentially inverse) role. A *derivation tree* for an assertion $A_0(a_0)$ in $\mathcal{A}$ with $A_0 \in \mathsf{N_C} \cup \{\bot\}$ is a finite $\mathsf{Ind}(\mathcal{A}) \times (\mathsf{N_C} \cup \{\bot\})$-labeled tree (T, V) that satisfies the following conditions:

- $V(\varepsilon) = (a_0, A_0)$;
- if $x \neq \varepsilon$, then $V(x)$ is not of the form $(a, \bot)$;

- if $V(x) = (a, A)$ and $A(a) \in \mathcal{A}$ or $\top \sqsubseteq A \in \mathcal{T}$, then x is a leaf;
- if $V(x) = (a, A)$ and neither $A(a) \notin \mathcal{A}$ nor $\top \sqsubseteq A \in \mathcal{T}$, then one of the following holds:
  - x has successors $y_1, \ldots, y_k$, $k \geq 1$ with $V(y_i) = (a, A_i)$ for $1 \leq i \leq k$ and $\mathcal{T} \models A_1 \sqcap \cdots \sqcap A_k \sqsubseteq A$;
  - x has a single successor y with $V(y) = (b, B)$ and there is an $\exists r.B \sqsubseteq A \in \mathcal{T}_\exists$ (with r possibly an inverse role) such that $r(a, b) \in \mathcal{A}$.

We call a TBox $\mathcal{T}$ *satisfiable* if it has a model. The main property of derivation trees is the following.

Lemma 29.

1. *$\mathcal{A}, \mathcal{T} \models A(a)$ iff $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ or there is a derivation tree for $A(a)$ in $\mathcal{A}$, for all assertions $A(a)$ with $A \in \mathsf{N_C}$ and $a \in \mathsf{Ind}(\mathcal{A})$;*
2. *$\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ iff $\mathcal{T}$ is unsatisfiable or there is a derivation tree for $\bot(a)$ in $\mathcal{A}$, for some $a \in \mathsf{Ind}(\mathcal{A})$.*

Proof. "if". Can be proved straightforwardly by induction on the depth of derivation trees.

"only if". For Point 1, assume that $\mathcal{A}, \mathcal{T} \models (a)$. If $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$, we are done. Thus assume that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$. Consider the sequence of ABoxes $\mathcal{A}_0, \mathcal{A}_1, \ldots$ with limit $\mathcal{A}^c_\mathcal{T}$ defined in Appendix A. When $\mathcal{A}, \mathcal{T} \models A(a)$, then $A(a) \in \mathcal{A}^c_\mathcal{T}$ by Lemma 20. It is thus sufficient to produce, for every assertion $\alpha = A(a) \in \mathcal{A}^c_\mathcal{T}$, a derivation tree $(V_\alpha, E_\alpha, \ell_\alpha)$ for α in $\mathcal{A}$. We proceed by induction over the sequence $\mathcal{A}_0, \mathcal{A}_1, \ldots$ as follows:

- if $A(a) \in \mathcal{A}_0 = \mathcal{A}$, then $V_\alpha = \{x\}$, $E_\alpha = \emptyset$, and $\ell_\alpha(x) = (a, A)$.
- if $A(a)$ was added to $\mathcal{A}_i$ because of rule (i) for some $i > 0$, then there is an $\exists r.B \sqsubseteq A \in \mathcal{T}$ and there are $r(a, b), B(b) \in \mathcal{A}_{i-1}$. Let $\beta = B(b)$. The derivation tree $(V_\alpha, E_\alpha, \ell_\alpha)$ is constructed by taking the derivation tree $(V_\beta, E_\beta, \ell_\beta)$ for β in $\mathcal{A}$ (which exists by induction hypothesis), adding a fresh root x with $V(x) = (a, A)$, and adding (x, y) to E_α, with y the root of $(V_\beta, E_\beta, \ell_\beta)$.
- if $A(a)$ was added to $\mathcal{A}_i$ because of rule (ii) for some $i > 0$, then there is an $\exists r^-.B \sqsubseteq A \in \mathcal{T}$ and there are $r(b, a), B(b) \in \mathcal{A}_{i-1}$. We can proceed as in the previous case.
- if $A(a)$ was added to $\mathcal{A}_i$ because of rule (iii) for some $i > 0$, then $\mathcal{A}_{i-1}|_a, \mathcal{T} \models A(a)$. Let $\beta_0, \ldots, \beta_k$ be all assertions of the form $B(a)$ in $\mathcal{A}_{i-1}$. The derivation tree $(V_\alpha, E_\alpha, \ell_\alpha)$ is constructed by taking the derivation trees $(V_{\beta_i}, E_{\beta_i}, \ell_{\beta_i})$, $i \leq k$ (which exists by induction hypothesis), adding a fresh root x with $\ell(x) = (a, A)$, and adding (x, y_i) to E_α, with y_i the root of $(V_{\beta_i}, E_{\beta_i}, \ell_{\beta_i})$, for all $i \leq k$.
- rule (iv) is not applied since $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$.

Now for Point 2. Since $\mathcal{T}$ is in normal form, $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ iff (i) $\mathcal{T}$ is unsatisfiable or (ii) there are an $a \in \mathsf{Ind}(\mathcal{A})$ and concept names $A_1, \ldots, A_k$ such that $\mathcal{A}, \mathcal{T} \models A_i(a)$ and $\mathcal{T} \models A_1 \sqcap \cdots \sqcap A_k \sqsubseteq \bot$. By Point 1, (ii) is the case iff there are an $a \in \mathsf{Ind}(\mathcal{A})$ and concept names $A_1, \ldots, A_k$ such that there is a derivation tree of $A_i(a)$ in $\mathcal{A}$ and $\mathcal{T} \models A_1 \sqcap \cdots \sqcap A_k \sqsubseteq \bot$. By definition derivation trees, it is easy to see that this is the case if there is an $a \in \mathsf{Ind}(\mathcal{A})$ such that there is a derivation tree of $\bot(a)$ in $\mathcal{A}$. $\square$

We are now ready to prove Proposition 28.

Proof of Proposition 28. We start with Point 2. If $\mathcal{T}$ is unsatisfiable, we simply choose as $\mathfrak{A}_{\mathcal{T},\Sigma,\bot,k}$ a TWABA that accepts the empty language. Otherwise, let $\mathsf{CN}(\mathcal{T})$ denote the set of all concept names in $\mathcal{T}$ and set $\mathfrak{A} = (Q, \Gamma_\Sigma, \delta, q_0, R)$ with

$$\begin{aligned} Q &= \{q_0\} \uplus \{q_A \mid A \in \mathsf{CN}(\mathcal{T}) \cup \{\bot\}\} \uplus \\ &\quad\ \{q_{A,r} \mid A \in \mathsf{CN}(\mathcal{T}) \cup \{\bot\}, r \in \mathsf{rol}(\Sigma)\} \end{aligned}$$

and $R = \emptyset$ (i.e., exactly the finite runs are accepting); define the transition function δ by setting

- $\delta(q_0, \sigma) = q_\bot \vee \bigvee_{i \in 1..m} [i]q_0$ for all $\sigma \in \Gamma_\Sigma$;
- $\delta(q_A, \sigma) = \mathsf{true}$ whenever $A \in \sigma$ or $\top \sqsubseteq A \in \mathcal{T}$;
- $\delta(q_A, \sigma) = \displaystyle\bigvee_{\mathcal{T} \models A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A} ([0]q_{A_1} \wedge \cdots \wedge [0]q_{A_n}) \vee \bigvee_{\exists r.B \sqsubseteq A \in \mathcal{T},\ r \in \mathsf{rol}(\Sigma)} ([-1]q_{B,r^-} \vee \bigvee_{i \in 1..m} [i]q_{B,r})$

  whenever $A \in \mathsf{CN}(\mathcal{T}) \cup \{\bot\}$ and $A \notin \sigma$;
- $\delta(q_{A,r}, \sigma) = [0]q_A$ whenever $r \in \sigma$;
- $\delta(q_{A,r}, \sigma) = \mathsf{false}$ whenever $r \notin \sigma$.

It is not hard to show that $\mathfrak{A}$ accepts a finite m-ary Γ_Σ-tree (T, V) which represents a tree-shaped Σ-ABox $\mathcal{A}$ iff there is a derivation tree for $\bot(a)$ in $\mathcal{A}$, for some $a \in \mathsf{Ind}(\mathcal{A})$. By Lemma 29 and since $\mathcal{T}$ is satisfiable, this is the case iff $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$. To obtain the desired TWABA $\mathfrak{A}_{\mathcal{T},\Sigma,\bot,m}$, it thus remains to take the complement of $\mathfrak{A}$.

For Point 1, we can construct a TWABA $\mathfrak{A}$ that accepts a finite m-ary Γ_Σ-tree (T, V) which represents a tree-shaped Σ-ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$ iff there is a derivation tree for $A(a_\varepsilon)$ in $\mathcal{A}$, analogously to what was done above. The desired automaton $\mathfrak{A}_{\mathcal{T},\Sigma,A,m}$ is then obtained by taking the union of $\mathfrak{A}$ and $\mathfrak{A}_{\mathcal{T},\Sigma,\bot,m}$. $\square$

To proceed, we endow the representation of tree-shaped Σ-ABoxes as finite Γ_Σ-trees with an additional component that explicitly states the depth of nodes in the tree, up to some bound. Let $k > 0$ be such a bound. We use Γ^k_Σ to denote the alphabet that consists of all pairs (σ, ℓ) with $\sigma \in \Gamma_\Sigma$ and $\ell \leq k$. A Γ^k_Σ-labeled tree is *valid* if the second components of node labels correctly represent node levels in the following sense: each node x on level $\ell \leq k$ satisfies $V(x) = (\sigma, \ell)$ for some σ, and each node on level $\ell > k$ satisfies $V(x) = (\sigma, k + 1)$ for some σ. Every finite Γ^k_Σ-labeled tree represents a tree-shaped Σ-ABox (also if it is not valid), namely the same ABox that is represented by the projection of the tree to a Γ_Σ-tree. Clearly, every Σ-ABox (except the empty one) is represented by some finite valid Γ^k_Σ-labeled tree. The outdegree is preserved.

Theorem 30. *For every $\mathcal{ELI}_\bot$ TBox in normal form, ABox signature Σ, AQ $A(x)$, and all $k, m > 0$, there is*

1. *a TWABA $\mathfrak{A}$ over m-ary Γ_Σ^{k+1}-trees such that $(T, V) \in L(\mathfrak{A})$ iff (T, V) is valid and represents a tree-shaped Σ-ABox $\mathcal{A}$ that is a k-entailment witness for $\mathcal{T}$, Σ, and $A(x)$;*
2. *a TWABA $\mathfrak{A}$ over m-ary Γ_Σ^{k+1}-trees such that $(T, V) \in L(\mathfrak{A})$ iff (T, V) is valid and represents a tree-shaped Σ-ABox $\mathcal{A}$ that is a k-inconsistency witness for $\mathcal{T}$, Σ, and $A(x)$.*

Moreover, the automata can be constructed in time exponential in the size of $\mathcal{T}$ and Σ and m, and in time polynomial in k; the number of states is $\mathcal{O}(|\mathcal{T}| + \mathsf{log}(k))$ in both cases.

Proof. We only provide a sketch of Point 1. The construction for Point 2 is similar. The desired TWABA $\mathfrak{A}$ is constructed as the intersection of the following TWABAs:

- $\mathfrak{A}_1$ accepts iff the input is valid;
- $\mathfrak{A}_2$ accepts iff the input represents an ABox $\mathcal{A}$ with $\mathcal{A}, \mathcal{T} \models A(a_\varepsilon)$;
- $\mathfrak{A}_3$ accepts iff the input represents an ABox that is consistent w.r.t. $\mathcal{T}$;
- $\mathfrak{A}_4$ accepts iff the input represents an ABox $\mathcal{A}$ with $\mathcal{A}|_k, \mathcal{T} \not\models A(a_\varepsilon)$.

The TWABA $\mathfrak{A}_1$ ensures via its initial state that the root node has a label of the form $(\sigma, 0)$. At every node of the tree, it first guesses (via a disjunction) whether the maximum counter value is reached. If this is the case, then it is verified that the current node has a label of the form $(\sigma, k+1)$, and that the same is true for all successors. If the maximum counter value is not reached, then $\mathfrak{A}_1$ guesses for every i-th bit whether the bit is currently odd or even, and whether it is toggled or not when moving to successors. The guesses are then verified in the current node and in the successor. In particular, the guess to toggle a bit requires that, at the current node, all less significant bits must be one and the guess to not toggle a bit requires that, at the current node, there is a less significant bit that is zero. Checking whether the i-th bit has value b amounts to switching to a state $q_{i,b}$ such that $\delta((\sigma, \ell), q_{i,b}) = \mathsf{false}$ if the i-th bit of the binary representation of ℓ is b, and $\delta((\sigma, \ell), q_{i,b}) = \mathsf{true}$ afterwards.

The TWABA $\mathfrak{A}_2$ is obtained by taking the automaton from Point 1 of Proposition 28 and modifying it so that it runs on Γ_Σ^k-labeled trees instead of on Γ_Σ-labeled ones, simply ignoring the additional componant of node labels. The TWABA $\mathfrak{A}_3$ is obtained in the same way from Point 2 of Proposition 28.

The TWABA $\mathfrak{A}_4$ is obtained by taking the complement of the automaton from Point 1 of Proposition 28, modifying it so that it runs on Γ_Σ^k-labeled trees, and further modifying it to ignore all nodes that have the maximum value $k+1$ in the second component. $\square$

We now come to the actual proof of Theorem 10. We separate here the upper bound (Theorem 31) and the lower bound (Theorem 32).

Theorem 31. *The following problems are in* EXPTIME*:*

1. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, an ABox signature Σ, and an AQ $A(x)$, is $A(x)$ FO-rewritable relative to $\mathcal{T}$ and Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$?*
2. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$ and an ABox signature Σ, is inconsistency of Σ-ABoxes FO-rewritable relative to $\mathcal{T}$ and Σ?*
3. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, an ABox signature Σ, and an AQ $A(x)$, is $A(x)$ FO-rewritable relative to $\mathcal{T}$ and Σ?*

Proof. By Theorem 3, it suffices to consider $\mathcal{ELI}_\bot$ TBoxes in normal form. First for Point 1 of Theorem 31. Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox, Σ an ABox signature, and $A(x)$ an AQ. Set $m = |\mathcal{T}|$ and $k = 2^{3|\mathcal{T}|^2}$, and let $\mathfrak{A}$ be the automaton from Point 1 of Theorem 30. By Theorem 9, $L(\mathfrak{A}) = \emptyset$ iff $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$. As a consequence of the "moreover" statement in Theorem 30, the automaton $\mathfrak{A}$ can be constructed in time exponential in $|\mathcal{T}|$ and has a number of states that is polynomial in $|\mathcal{T}|$. By Theorem 27, it can thus be checked in EXPTIME whether $L(\mathfrak{A}) = \emptyset$.

The same statements are true when "Point 1" is replaced with "Point 2" in all mentioned theorems. Finally, Point 3 of Theorem 31 is a consequence of Points 1 and 2 of the same theorem, together with Proposition 6. $\square$

Theorem 32. *The three decision problems in Theorem 10 are* EXPTIME*-hard. This holds even for $\mathcal{ELI}$ TBoxes and the full ABox signature.*

Proof. The proof employs the EXPTIME-hardness proof for subsumption in $\mathcal{ELI}$ given in [Baader *et al.*, 2008]. Recall that the *subsumption problem in $\mathcal{ELI}$* is the problem to decide whether a CI $C \sqsubseteq D$ is entailed by an $\mathcal{ELI}$ TBox $\mathcal{T}$, in symbols $\mathcal{T} \models C \sqsubseteq D$. EXPTIME-hardness of the subsumption problem in $\mathcal{ELI}$ is shown in [Baader *et al.*, 2008] by defining a class $\mathcal{X}$ of $\mathcal{ELI}$ TBoxes $\mathcal{T}$ and concept names Init and Good such that

- The following problem is EXPTIME-hard: given $\mathcal{T} \in \mathcal{X}$, does $\mathcal{T} \models \mathsf{Init} \sqsubseteq \mathsf{Good}$ hold?
- The concept names Init and Good occur in $\mathcal{T}$ in three CIs: (i) a CI $\mathsf{Init} \sqsubseteq C$ with neither Good nor Init in C, and (ii) two CIs of the form $\mathsf{Init} \sqcap E \sqsubseteq \mathsf{Good}$, where E is a concept name.

We prove the EXPTIME-hardness results by reduction of the subsumption problem $\mathcal{T} \models \mathsf{Init} \sqsubseteq \mathsf{Good}$ for $\mathcal{T} \in \mathcal{X}$.

We first prove EXPTIME-hardness of the problem to decide whether ABox consistency is FO-rewritable relative to an $\mathcal{ELI}_\bot$ TBox and the full ABox signature. Let $\mathcal{T} \in \mathcal{X}$. Let M and X be fresh concept names and r a fresh role name, and let $\mathcal{T}'$ be obtained from $\mathcal{T}$ by (i) replacing the CIs $\mathsf{Init} \sqcap E \sqsubseteq \mathsf{Good}$ by $\mathsf{Init} \sqcap E \sqsubseteq \bot$, and (ii) adding the CIs $\exists r.X \sqsubseteq X$, $X \sqcap M \sqsubseteq \bot$, and $M \sqsubseteq \mathsf{Init}$.

Claim. $\mathcal{T} \models \mathsf{Init} \sqsubseteq \mathsf{Good}$ iff ABox consistency is FO-rewritable relative to $\mathcal{T}'$ and the full ABox signature.

"if". Assume that $\mathcal{T} \not\models \mathsf{Init} \sqsubseteq \mathsf{Good}$. Then ABoxes that consist of longer and longer r-chains with an X at the end and an M at the beginning witness non-FO-rewritability of ABox consistency. In detail, let for $k > 0$

$$\mathcal{A}_k = \{r(a_0, a_1), \ldots, r(a_k, a_{k+1}), M(a_0), X(a_{k+1})\}$$

Let $\rho_{\mathcal{A}_k} = a_0$. Then $\mathcal{A}_k$ is not consistent w.r.t. $\mathcal{T}'$ but $\mathcal{A}_k \setminus \{\rho_{\mathcal{A}_k}\}$ and $\mathcal{A}_k|_k$ are consistent w.r.t. $\mathcal{T}'$. Hence, by Theorem 7, ABox consistency is not FO-rewritable relative to $\mathcal{T}'$ and the full ABox signature. To see that $\mathcal{A}_k|_k$ is consistent w.r.t. $\mathcal{T}'$ take a model $\mathcal{I}$ of $\mathcal{T}$ with $d \in \mathsf{Init}^{\mathcal{I}}$ and $d \notin \mathsf{Good}^{\mathcal{I}}$. Since $\mathcal{T} \in \mathcal{X}$, we may assume that $\mathsf{Init}^{\mathcal{I}} = \{d\}$. Take copies $\mathcal{I}_0, \ldots, \mathcal{I}_k$ of $\mathcal{I}$ with domains $\Delta^{\mathcal{I}_i} = \Delta^{\mathcal{I}} \times \{i\}$ and interpret a_i as (d, i), r as the set of pairs $((d, i), (d, i + 1))$, for $i < k$, and M as $\{(d, 0)\}$. The resulting interpretation, $\mathcal{J}$, is a model of $\mathcal{A}_k|_k$ and $\mathcal{T}$. $\mathcal{J}$ is a model of $\mathcal{T}'$ since $d \notin E^{\mathcal{I}}$ because $d \notin \mathsf{Good}^{\mathcal{I}}$ and $\mathcal{I}$ is a model of $\mathcal{T}$.

"only if". Assume that $\mathcal{T} \models \mathsf{Init} \sqsubseteq \mathsf{Good}$. Then the FO-rewriting of ABox consistency w.r.t. $\mathcal{T}'$ is $\neg(\exists x.M(x) \vee \exists x.\mathsf{Init}(x))$.

Recall that FO-rewritability of ABox consistency can be reduced to FO-rewritability of AQs by introducing the AQ $A(x)$ for a fresh concept name A. It follows that FO-rewritability of AQs relative $\mathcal{ELI}_\bot$ TBoxes and the full ABox signature is EXPTIME-hard as well.

Finally, we show EXPTIME-hardness of the problem to decide FO-rewritability relative to $\mathcal{ELI}$ TBoxes and consistent ABoxes over the full ABox signature. The reduction is similar to the reduction above. Let $\mathcal{T} \in \mathcal{X}$. Let M and X be fresh concept names and r a fresh role name, and let $\mathcal{T}'$ be obtained from $\mathcal{T}$ by adding the CIs $\exists r.X \sqsubseteq X$, $X \sqcap M \sqsubseteq \mathsf{Good}$, and $M \sqsubseteq \mathsf{Init}$ to $\mathcal{T}$.

Claim. $\mathcal{T} \models \mathsf{Init} \sqsubseteq \mathsf{Good}$ iff $\mathsf{Good}(x)$ is FO-rewritable relative to $\mathcal{T}'$ and consistent ABoxes over the full ABox signature.

"if". Assume that $\mathcal{T} \not\models \mathsf{Init} \sqsubseteq \mathsf{Good}$. Then ABoxes that consist of longer and longer r-chains with an X at the end and an M at the beginning witness non-FO-rewritability of $\mathsf{Good}(x)$. In detail, let for $k > 0$ again

$$\mathcal{A}_k = \{r(a_0, a_1), \ldots, r(a_k, a_{k+1}), M(a_0), X(a_{k+1})\}$$

and $\rho_{\mathcal{A}_k} = a_0$. Then $\mathcal{A}_k$ is consistent w.r.t. $\mathcal{T}'$ but $\mathcal{A}_k, \mathcal{T}' \models \mathsf{Good}(\rho_{\mathcal{A}_k})$ and $\mathcal{A}_k|_k, \mathcal{T}' \not\models \mathsf{Good}(\rho_{\mathcal{A}_k})$. Hence, by Theorem 7, $\mathsf{Good}(x)$ is not FO-rewritable relative to $\mathcal{T}'$ and consistent ABoxes over the full ABox signature.

"only if". Assume that $\mathcal{T} \models \mathsf{Init} \sqsubseteq \mathsf{Good}$. Then the FO-rewriting of $\mathsf{Good}(x)$ relative to $\mathcal{T}'$ is $\mathsf{Good}(x) \vee M(x) \vee \mathsf{Init}(x)$. $\square$

A.5 Proof of Theorem 11

Theorem 11.

1. For every Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, signature Σ, and AQ $A(x)$ that is rewritable relative to $\mathcal{T}$ and Σ, one can effectively construct a UCQ-rewriting $\varphi(x)$ of size at most $2^{2^{2^{\mathcal{O}(|\mathcal{T}|^2)}}}$, in time polynomial in the size of $\varphi(x)$.
2. There is a family of $\mathcal{EL}$ TBoxes $\mathcal{T}_0, \mathcal{T}_1, \ldots$ such that for all $i \geq 0$, $|\mathcal{T}_i| \in \mathcal{O}(i^2)$ and $A(x)$ is FO-rewritable relative to $\mathcal{T}_i$ and the full ABox signature $\mathsf{sig}(\mathcal{T}_i)$, but the smallest UCQ-rewriting is of size at least $2^{2^{2^i}}$.

Proof. We start with Point 1. By Theorem 3, it suffices to consider $\mathcal{ELI}_\bot$-TBoxes. Thus let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$-TBox, Σ a signature, and $A(x)$ an AQ that is rewritable relative to $\mathcal{T}$ and Σ. The proof of Theorem 7 yields a way to effectively construct an FO-rewriting $\varphi(x)$ of $A(x)$ relative to $\mathcal{T}$ and consistent Σ-ABoxes, and an FO-rewriting $\psi(x)$ of ABox inconsistency relative to $\mathcal{T}$ and Σ. Both $\varphi(x)$ and $\psi(x)$ are UCQs that consist only of tree-shaped CQs and use only symbols from $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$. In fact, these queries are CQ-representation of tree-shaped Σ-ABoxes that are inconsistent w.r.t. $\mathcal{T}$ and do not exceed the depth bound k from the Theorem. By Point 1 of Theorem 9, it suffices to generate only tree-shaped CQs of depth at most 2^{3n^2} in the rewritings, with $n = |\mathcal{T}|$. Moreover, using a selection argument as in the proof of Proposition 24, one can show that every tree-shaped Σ-ABox $\mathcal{A}$ that is inconsistent w.r.t. $\mathcal{T}$ contains a tree-shaped Σ-ABox that is also inconsistent w.r.t. $\mathcal{T}$ and has outdegree at most n. Thus, the outdegree of the tree-shaped CQs generated in the rewritings need also not be larger than n. It is not hard to verify that, up to equivalence, the number of CQs of the described form is $2^{2^{2^{\mathcal{O}(n^2)}}}$. We obtain the same bound for the size of UCQs that consist only of CQs of this form. It remains to note that $\varphi(x) \vee \psi(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ, see the proof of Proposition 6.

For Point 2, we utilize a construction from [Nikitina and Rudolph, 2012], itself inspired by a similar construction in [Lutz and Wolter, 2010]. For every $n > 0$, let $\mathcal{T}_n$ be the following $\mathcal{EL}$ TBox:

$$\begin{aligned}
A_1 &\sqsubseteq \overline{X_0} \sqcap \ldots \sqcap \overline{X_n} \\
A_2 &\sqsubseteq \overline{X_0} \sqcap \ldots \sqcap \overline{X_n} \\
\sqcap_{\sigma\in\{r,s\}} \exists\sigma.(\overline{X_i} \sqcap X_0 \sqcap \ldots \sqcap X_{i-1}) &\sqsubseteq X_i && i \leq n \\
\sqcap_{\sigma\in\{r,s\}} \exists\sigma.(X_i \sqcap X_0 \sqcap \ldots \sqcap X_{i-1}) &\sqsubseteq \overline{X_i} && i \leq n \\
\sqcap_{\sigma\in\{r,s\}} \exists\sigma.(\overline{X_i} \sqcap \overline{X_j}) &\sqsubseteq \overline{X_i} && j < i \leq n \\
\sqcap_{\sigma\in\{r,s\}} \exists\sigma.(X_i \sqcap \overline{X_j}) &\sqsubseteq X_i && j < i \leq n \\
X_0 \sqcap \ldots \sqcap X_n &\sqsubseteq B
\end{aligned}$$

Note that B is not FO-rewritable relative to $\mathcal{T}_n$. Indeed, let $k = 2^{3|\mathcal{T}|^2}$ and consider the linear ABox

$$\begin{aligned}
\mathcal{A} =& \{X_i(a_0) \mid 1 \leq i < n\} \cup \{r(a_i, a_{i+1}) \mid 0 \leq i \leq k\} \cup \\
& \{X_1(a_i) \mid 1 \leq i \leq k+1\} \cup \{X_n(a_{k+1})\}
\end{aligned}$$

Then $\mathcal{A}, \mathcal{T}_n \models B(a_0)$ but $\mathcal{A}|_k, \mathcal{T} \not\models B(a_0)$. To force the

FO-rewritability of B, we add the following CIs:

$$\begin{array}{rl}
\exists\sigma.\top \sqsubseteq \overline{Y_0} \sqcap \ldots \sqcap \overline{Y_n} & \sigma \in \{r,s\} \\
\exists\sigma.(\overline{Y_i} \sqcap Y_0 \sqcap \ldots \sqcap Y_{i-1}) \sqsubseteq Y_i & \sigma \in \{r,s\}, i \leq n \\
\exists\sigma.(Y_i \sqcap Y_0 \sqcap \ldots \sqcap Y_{i-1}) \sqsubseteq \overline{Y_i} & \sigma \in \{r,s\}, i \leq n \\
\exists\sigma.(\overline{Y_i} \sqcap \overline{Y_j}) \sqsubseteq \overline{Y_i} & \sigma \in \{r,s\}, j < i \leq n \\
\exists\sigma.(Y_i \sqcap \overline{Y_j}) \sqsubseteq Y_i & \sigma \in \{r,s\}, j < i \leq n \\
Y_0 \sqcap \ldots \sqcap Y_n \sqsubseteq B &
\end{array}$$

Call the resulting TBox $\mathcal{T}_n'$ (notice that $|\mathcal{T}_n'| \in O(n^2)$). The inclusions in $\mathcal{T}_n' \setminus \mathcal{T}_n$ ensure that the concept B is entailed at an individual whenever that individual has an outgoing path of length 2^{n+1}. As a consequence, we have that for every linear ABox $\mathcal{A}$ with root $\rho_\mathcal{A}$, $\mathcal{A}, \mathcal{T}_n' \models B(\rho_\mathcal{A})$ implies $\mathcal{A}|_{2^{n+1}}, \mathcal{T}_n' \models B(\rho_\mathcal{A})$. It follows then by Theorem 13 that $B(x)$ is FO-rewritable relative to $\mathcal{T}_n'$ and the full signature $\mathsf{sig}(\mathcal{T}_n')$. Let $\varphi(x)$ be any UCQ-rewriting of $B(x)$ relative to $\mathcal{T}_n'$ and $\mathsf{sig}(\mathcal{T}_n')$. Because $\mathcal{T}_n'$ is an $\mathcal{EL}$ TBox, we can assume w.l.o.g. that every disjunct $q(x)$ of $\varphi(x)$ is a connected CQ, since otherwise we could simply drop any connected component of $q(x)$ which does not contain the answer variable x.

To establish Point 2, it is sufficient to show that the UCQ $\varphi(x)$ has at least $2^{2^{2^n}}$ disjuncts. To this end, define inductively sets $\mathcal{C}_i$ of concepts by setting $\mathcal{C}_0 = \{A_1, A_2\}$ and $\mathcal{C}_{i+1} = \{\exists r.C_1 \sqcap \exists s.C_2 \mid C_1, C_2 \in \mathcal{C}_i\}$, for every $0 \leq i < 2^{n+1} - 2$. Let $\mathcal{C} = \mathcal{C}^{2^{n+1}-1}$. It is shown in [Nikitina and Rudolph, 2012] that:

(a) $\mathcal{C}$ contains $2^{2^{2^{n+1}-1}}$ distinct concepts, and

(b) each concept $C \in \mathcal{C}$ is such that $\mathcal{T}_n \models C \sqsubseteq B$, and hence $\mathcal{T}_n' \models C \sqsubseteq B$

It is also not hard to see that the concepts in $\mathcal{C}$ cannot be weakened while still implying B. Formally:

(c) if $C \in \mathcal{C}$, then there is no C' such that $\models C \sqsubseteq C'$, $\not\models C \sqsubseteq C'$, and $\mathcal{T}_n \models C' \sqsubseteq B$.

where the notation $\models G \sqsubseteq H$ is used to denote subsumption w.r.t. the empty TBox.

For every $C \in \mathcal{C}$, we let $\mathcal{A}_C$ be a tree-shaped ABox with root a_0 which instantiates the concept C, defined in the obvious way (cf. the proof of Theorem 12). One can show that for every $\mathcal{EL}$ concept D, and every $\mathcal{EL}$ TBox $\mathcal{T}$, we have

$$\mathcal{A}_C, \mathcal{T} \models D(a_0) \quad \text{iff} \quad \mathcal{T} \models C \sqsubseteq D \qquad (*)$$

In particular, by properties (b) and $(*)$, we must have $\mathcal{A}_C, \mathcal{T}_n' \models B(a_0)$. It follows that there is some disjunct $q(x)$ of $\varphi(x)$ such that $a_0 \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}_C}, q)$. Let π be a match for $q(x)$ in $\mathcal{I}_{\mathcal{A}_C}$ such that $\pi(x) = a_0$, and let $q'(x)$ be the CQ obtained by merging variables in $q(x)$ whenever they have the same image under π. We remark that since $\mathcal{I}_{\mathcal{A}_C}$ has the shape of a directed tree rooted at a_0, the CQ $q'(x)$ must take the form of a directed tree rooted at x. This means that we can "roll up" the query q' into a concept $D_{q'}$, such that for every ABox $\mathcal{B}$ and $b \in \mathsf{Ind}(\mathcal{B})$, we have $\mathcal{B} \models D_{q'}(b)$ iff $b \in \mathsf{ans}(\mathcal{I}_\mathcal{B}, q')$. Since $a_0 \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}_C}, q')$, we must have $\mathcal{A}_C \models D_{q'}(a_0)$. Applying $(*)$ with $\mathcal{T}$ the empty TBox, we obtain $\models C \sqsubseteq D_{q'}$. On the other hand, we have that for every ABox $\mathcal{B}$ and $b \in \mathsf{Ind}(\mathcal{B})$, $\mathcal{B} \models D_{q'}(b)$ implies $\mathcal{T}_n', \mathcal{B} \models B(b)$. It follows that $\mathcal{T}_n' \models D_{q'} \sqsubseteq B$. Using property (c), we obtain $\models D_{q'} \sqsubseteq C$, and hence $\models C \equiv D_{q'}$.

We have thus shown that for every concept $C \in \mathcal{C}$, there is some disjunct q of $\varphi(x)$ and some tree-shaped CQ q' obtained by merging variables in q such that $\models D_{q'} \equiv C$, where $D_{q'}$ is the concept corresponding to q'. If every concept C is associated with a different disjunct q, then we are done. Indeed, since $\mathcal{C}$ contains $2^{2^{2^{n+1}-1}}$ pairwise non-equivalent concepts, and $2^{2^{2^{n+1}-1}} > 2^{2^{2^n}}$, this would imply that there are at least $2^{2^{2^n}}$ disjuncts in $\varphi(x)$. To complete the argument, suppose for a contradiction that there are distinct concepts $C_1, C_2 \in \mathcal{C}$, a disjunct q of $\varphi(x)$, and tree-shaped CQs q_1, q_2 obtained by merging variables in q such that $\models D_{q_1} \equiv C_1$ and $\models D_{q_2} \equiv C_2$ (with D_{q_1} and D_{q_2} the concepts associated with q_1, q_2). Since C_1 and C_2 are different, there must exist some sequence $w_1 \ldots w_{2^{n+1}-1}$ of roles from $\{r, s\}$ such that C_1 and C_2 differ on the concept name (A_1 or A_2) appearing in the scope of the chain of existential quantifiers $w_1 \ldots w_{2^{n+1}-1}$. Say that $\models C_1 \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_1$ and $\models C_2 \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_2$. It follows that we must have $\models D_{q_1} \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_1$ and $\models D_{q_2} \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_2$. The first subsumption means that the query $q_1(x)$ contains a $w_1 \ldots w_{2^{n+1}-1}$-path from x to some y with $A_1(y) \in q_1$. Now let ν_1 be such that q_1 is the result of applying ν_1 to the variables in q, and let $z_1, \ldots, z_\ell$ be the variables in q such that $\nu_1(z) = y$. Since q is connected, for every variable z_i, there is a path p_{z_i} from x to z_i. It follows that in q_1, there is a path p_{z_i} from x to y, for every $1 \leq i \leq \ell$. However, since q_1 is tree-shaped, there is a unique path to every variable, and so all paths p_{z_i} must be the same and equal to $w_1 \ldots w_{2^{n+1}-1}$. The presence of atom $A_1(y)$ in q_1 implies that $A_1(z_i) \in q$ for some $1 \leq i \leq \ell$. It follows that q contains a $w_1 \ldots w_{2^{n+1}-1}$-path from x to some z with $A_1(z) \in q$. By applying the same argument to q_2, we can show that q contains a $w_1 \ldots w_{2^{n+1}-1}$-path from x to some z' with $A_2(z') \in q$. These two paths must also be present in the queries q_1 and q_2, which implies that $\models D_{q_1} \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_2$ and $\models D_{q_2} \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_1$. We thus obtain $\models C_1 \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_2$ and $\models C_2 \sqsubseteq \exists w_1 \exists w_2 \ldots \exists w_{2^{n+1}-1}.A_1$. This gives us the desired contradiction since the definition of the concepts in $\mathcal{C}$ forbids A_1 and A_2 from appearing in the scope of the same chain of existential restrictions in the same concept. □

B Proofs for Section 4

The notion of a *canonical model* is used in the proof of Theorem 12, so we recall the definition here. Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $\mathcal{A}$ an ABox. For $a \in \mathsf{Ind}(\mathcal{A})$, a *path* for $\mathcal{A}$ and $\mathcal{T}$ is a finite sequence $a\, r_1\, C_1\, r_2 C_2 \cdots r_n\, C_n$, $n \geq 0$, where the C_i are concepts that occur in $\mathcal{T}$ (potentially as a subconcept) and the r_i are roles such that the following conditions are satisfied:

- $a \in \mathsf{Ind}(\mathcal{A})$,
- $\mathcal{A}, \mathcal{T} \models \exists r_1.C_1(a)$ if $n \geq 1$,
- $\mathcal{T} \models C_i \sqsubseteq \exists r_{i+1}.C_{i+1}$ for $1 \leq i < n$.

The domain $\Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T}}}$ of the *canonical model* $\mathcal{I}_{\mathcal{A},\mathcal{T}}$ for $\mathcal{T}$ and $\mathcal{A}$ is the set of all paths for $\mathcal{A}$ and $\mathcal{T}$. If $p \in \Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} \setminus \mathsf{Ind}(\mathcal{A})$, then $\mathsf{tail}(p)$ denotes the last concept C_n in p. Set

$$\begin{array}{rcl} A^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} & := & \{a \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{A},\mathcal{T} \models A(a)\} \cup \\ & & \{p \in \Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} \setminus \mathsf{Ind}(\mathcal{A}) \mid \mathcal{T} \models \mathsf{tail}(p) \sqsubseteq A\} \\ r^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} & := & \{(a,b) \mid r(a,b) \in \mathcal{A}\} \cup \\ & & \{(p,q) \in \Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} \times \Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} \mid \\ & & \quad q = p \cdot r\, C \text{ for some } C\} \\ a^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} & := & a \text{ for all } a \in \mathsf{Ind}(\mathcal{A}) \end{array}$$

It is standard to prove the following.

Lemma 33. *$\mathcal{I}_{\mathcal{A},\mathcal{T}}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ such that:*

- *for any $a \in \mathsf{Ind}(\mathcal{A})$ and $\mathcal{EL}$ concept C, we have $a^{\mathcal{I}_{\mathcal{A},\mathcal{T}}} \in C^{\mathcal{I}_{\mathcal{A},\mathcal{T}}}$ iff $\mathcal{A},\mathcal{T} \models C(a)$,*
- *for any CQ $q(\vec{x})$ of arity k and k-tuple $\vec{a} \in \mathsf{Ind}(\mathcal{A})$, we have $\mathsf{ans}(\mathcal{I}_{\mathcal{A},\mathcal{T}}, q(\vec{x})) = \mathsf{cert}_{\mathcal{T}}(q(\vec{x}), \mathcal{A})$.*

We use the same normal form for $\mathcal{EL}$ TBoxes that was introduced in the context of $\mathcal{ELI}_\bot$. Every $\mathcal{EL}$ TBox can be put into normal form as follows. Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, Σ be an ABox signature, and $A(x)$ be an AQ. We use $\mathsf{sub}(\mathcal{T})$ to denote the set of subconcepts C of (concepts that occur in) $\mathcal{T}$. For every $C \in \mathsf{sub}(\mathcal{T})$ that is neither a concept name nor $\top$, introduce a concept name X_C that does not occur in $\mathcal{T}$ nor in Σ and is distinct from the concept name A. Set

$$\sigma(C) = \begin{cases} C & \text{if } C \in \mathsf{N_C} \cup \{\top\} \\ X_{D_1} \sqcap X_{D_2} & \text{if } C = D_1 \sqcap D_2 \\ \exists r.X_D & \text{if } C = \exists r.D \end{cases}$$

and define $\mathcal{T}'$ as

$$\bigcup_{C \sqsubseteq D \in \mathcal{T}} X_C \sqsubseteq X_D \;\cup \bigcup_{C \in \mathsf{sub}(\mathcal{T})} \{X_C \sqsubseteq \sigma(C), \sigma(C) \sqsubseteq X_C\}.$$

After replacing CIs $A \sqsubseteq B_1 \sqcap B_2$ with $A \sqsubseteq B_1$ and $A \sqsubseteq B_2$, and then replacing CIs $A \sqsubseteq B$ by $A \sqcap A \sqsubseteq B$, $\mathcal{T}'$ is of the required form.

The following lemma resumes some properties of the normalized TBoxes obtained via this procedure.

Lemma 34.

1. *For any FOQ φ and ABox signature Σ with $\Sigma \cap \mathcal{T}' \subseteq \mathcal{T}$, $\varphi(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ iff $\varphi(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and Σ.*
2. *For every concept name $B \in \mathsf{sig}(\mathcal{T})$ and individual a, $\mathcal{A},\mathcal{T} \models B(a)$ iff $\mathcal{A},\mathcal{T}' \models B(a)$*

Because $\mathcal{EL}$ does not allow inverse roles, it will prove relevant to consider ABoxes and queries which have the form of *directed trees*. Formally, an ABox $\mathcal{A}$ is called *dtree-shaped* if the directed graph $(\mathsf{Ind}(\mathcal{A}), E = \{(a,b) \mid r(a,b) \in \mathcal{A}\})$ is a tree and $r(a,b), s(a,b) \in \mathcal{A}$ implies $r = s$. A CQ is a *dtree-CQ* if it is dtree-shaped and its root is the only answer variable; a *dtree-UCQ* takes the form $q_1 \vee \cdots \vee q_n$ where each q_i is dtree-CQ rooted at the same answer variable.

The following lemma is easy to show.

Lemma 35. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, $\mathcal{A}$ be a tree-shaped ABox, and $a \in \mathsf{Ind}(\mathcal{A})$. Then $\mathcal{T},\mathcal{A} \models A(a)$ if and only if $\mathcal{T},\mathcal{A}_a \models A(a)$, where $\mathcal{A}_a$ is the restriction of $\mathcal{A}$ to those individuals which are reachable from a in the directed graph $(\mathsf{Ind}(\mathcal{A}), E = \{(a,b) \mid r(a,b) \in \mathcal{A}\})$.*

By combining the preceding lemma and the proof of Point 1 of Theorem 7, we obtain:

Lemma 36. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, Σ an ABox signature, and $A(x)$ an AQ. If $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ, then there is a dtree-UCQ that is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ.*

B.1 Proof of Theorem 12

Theorem 12. *FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes and the full ABox signature can be polynomially reduced to FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes in normal form and the full ABox signature.*

Proof. Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, $A(x)$ be an AQ, and $\mathcal{T}'$ be obtained by applying the normalization procedure to $\mathcal{T}$ described above. We show that $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full signature $\mathsf{sig}(\mathcal{T})$ if and only if $A(x)$ is FO-rewritable relative to $\mathcal{T}'$ and the full signature $\mathsf{sig}(\mathcal{T}')$. Since $\mathcal{T}'$ can be computed from $\mathcal{T}$ in polynomial time, this yields the desired reduction.

For the first direction, suppose that $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full signature $\mathsf{sig}(\mathcal{T})$. Then by Lemma 36, there exists a dtree-UCQ $\varphi(x)$ such that $\mathsf{cert}_{\mathcal{T}}(A(x), \mathcal{A}) = \mathsf{ans}(\mathcal{I}_{\mathcal{A}}, \varphi)$ for every Σ-ABox $\mathcal{A}$, where $\Sigma = \mathsf{sig}(\mathcal{T})$. We can assume w.l.o.g. that $\varphi(x)$ only uses symbols from Σ. We aim to modify $\varphi(x)$ to obtain a FO-rewriting $\varphi'(x)$ of $A(x)$ relative to $\mathcal{T}'$ and the full signature $\mathsf{sig}(\mathcal{T}')$. Intuitively, $\varphi'(x)$ will be a UCQ consisting of all CQs which are obtained by taking a disjunct of $\varphi(x)$ and replacing some of the subqueries by concept names introduced in $\mathcal{T}'$. Formally, to every dtree-CQ q and variable $v \in \mathsf{vars}(q)$, we associate an $\mathcal{EL}$ concept $C_{q,v}$ as follows:

$$C_{q,v} = \bigsqcap_{A(v) \in q} A \sqcap \bigsqcap_{r(v,v') \in q} \exists r.C_{q,v'}.$$

We say that q' is obtained from a dtree-CQ q by *definition replacement* if there exist a concept name $X_D \in \mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$ and variable $v \in \mathsf{vars}(q)$ such that q' is the result of applying the following operations to q:

1. add the atom $X_D(v)$
2. for every atom $A(v) \in q$ such that $\mathcal{T}' \models X_D \sqsubseteq A$, remove $A(v)$ from q
3. for every term v' such that $r(v,v') \in q$ and $\mathcal{T}' \models X_D \sqsubseteq \exists r.C_{q,v'}$, remove all atoms involving v' or its descendants in q

and at least one atom of q is removed during this process. We call q' a *definition-rewriting* of q if q' can be obtained from q by zero or more definition replacements. Let $\varphi'(x)$ be the disjunction of all queries q' which are definition-rewritings of some disjunct q of $\varphi(x)$. Note that $\varphi'(x)$ is well-defined since there can be only finitely many definition-rewritings of a given CQ.

We aim to show that $\varphi'(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and the full signature $\mathsf{sig}(\mathcal{T}')$. To this end, let $\mathcal{A}'$ be a $\mathsf{sig}(\mathcal{T}')$-ABox. First suppose that $a \in \mathsf{cert}_{\mathcal{T}'}(A(x), \mathcal{A}')$. We define a $\mathsf{sig}(\mathcal{T})$-ABox which intuitively corresponds to replacing each assertion $X_D(b)$ with $X_D \notin \mathsf{sig}(\mathcal{T})$ by the instantiation of D at b. Formally, to every assertion $X_D(b) \in \mathcal{A}'$, we can associate the set of assertions $\mathsf{inst}(D, b)$ defined as follows:

$$
\begin{array}{ll}
\mathsf{inst}(D, b) = \{A(b)\} & \text{if } D = A \\
\mathsf{inst}(D, b) = \mathsf{inst}(E_1, b) \cup \mathsf{inst}(E_2, b) & \text{if } D = E_1 \sqcap E_2 \\
\mathsf{inst}(D, b) = \{r(b, c)\} \cup \mathsf{inst}(E, c) & \text{if } C = \exists r.E \\
 & [\text{where } c \text{ is a fresh} \\
 & \ \text{individual name}]
\end{array}
$$

We require that if c is a fresh individual name used in $\mathsf{inst}(D, b)$, then c does not appear in $\mathsf{inst}(D', b')$, for $(D', b') \neq (D, b)$. It is easy to see that the definition of $\mathsf{inst}(D, b)$ guarantees that $\mathsf{inst}(D, b) \models D(b)$. The ABox $\mathcal{A}$ is then obtained from $\mathcal{A}'$ by replacing each assertion $X_D(b) \in \mathcal{A}'$ by the set of assertions $\mathsf{inst}(D, b)$. Note that $\mathcal{A}$ is a $\mathsf{sig}(\mathcal{T})$-ABox, since D is a $\mathsf{sig}(\mathcal{T})$-concept whenever $X_D \in \mathsf{sig}(\mathcal{T}')$.

We next observe that by construction $X_D(b) \in \mathcal{A}'$ implies $\mathcal{A} \models D(b)$. As $\mathcal{T}' \models X_D \equiv D$, we obtain $\mathcal{A}, \mathcal{T}' \models X_D(b)$, and hence $\mathcal{A}, \mathcal{T}' \models \mathcal{A}'$. Because $a \in \mathsf{cert}_{\mathcal{T}'}(A(x), \mathcal{A}')$, we must also have $a \in \mathsf{cert}_{\mathcal{T}'}(A(x), \mathcal{A})$. As the answers to AQs are preserved under the normalization procedure (cf. point 2 of Lemma 34), we must also have $a \in \mathsf{cert}_{\mathcal{T}}(A(x), \mathcal{A})$. Using our assumption that $\varphi(x)$ is a FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T})$, we obtain $a \in \mathsf{ans}(\mathcal{I}_\mathcal{A}, \varphi)$. It follows that there must exist a dtree-CQ q which is a disjunct of φ such that $a \in \mathsf{ans}(\mathcal{I}_\mathcal{A}, q)$. Take some match π for q in $\mathcal{I}_\mathcal{A}$ with $\pi(x) = a$. We define inductively a sequence $q_0, q_1, q_2 \ldots,$ of dtree-CQs by setting $q_0 = q$ and letting q_{i+1} be obtained by applying the following rule to q_i:

($*$) select an atom $\alpha \in q_i$ such that either (a) $\alpha = B(v)$, $B \in \mathsf{sig}(\mathcal{T})$, $\pi(v) \in \mathsf{Ind}(\mathcal{A}')$, and $B(\pi(v)) \notin \mathcal{A}'$, or (b) $\alpha = r(v, v')$, $\pi(v) \in \mathsf{Ind}(\mathcal{A}')$, and $\pi(v') \notin \mathsf{Ind}(\mathcal{A}')$. Let D be such that $\pi(\alpha) \in \mathsf{inst}(D, \pi(v))$. Add the atom $X_D(v)$ and remove (i) all atoms $B(v)$ such that $\mathcal{T}' \models X_D \sqsubseteq B$, and (ii) all atoms involving variable u or its descendants, whenever $s(v, u) \in q_i$ and $\mathcal{T}' \models X_D \sqsubseteq \exists s.C_{q_i,u}$.

Note that each q_{i+1} is a definition replacement of q_i. Indeed, by the choice of α, we are guaranteed to remove either $B(v)$ (case (a)) or $r(v, v')$ (case (b)). It follows that after a finite number of steps, we obtain a dtree-CQ for which neither rule is applicable. Call this query q', and let π' be the restriction of π to the variables in q'. We aim to show that π' is a match for q' in $\mathcal{I}_{\mathcal{A}'}$. Let α be an atom in q'. There are three cases to consider.

- Case 1: $\alpha = B(v)$, where $B \in \mathsf{sig}(\mathcal{T})$. First note that $B(v) \in q$, since the rule (*) only adds concept assertions involving the new concept names from $\mathcal{T}'$. Next note that $\pi(v) \in \mathsf{inds}(\mathcal{A}')$, since otherwise $B(v)$ would have been removed when treating a role assertion involving v or an ancestor of v. Likewise, we must have $B(\pi(v)) \in \mathcal{A}'$, since otherwise the rule ($*$) would be applicable with $\alpha = B(v)$, a contradiction. As $\pi'(v) = \pi(v)$, we obtain $B(\pi'(v)) \in \mathcal{A}'$.
- Case 2: $\alpha = X_D(v)$, where $X_D \in \mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$. Then $X_D(v)$ must have been added by an application of rule ($*$). Then it must be the case that $\mathsf{inst}(D, \pi(v)) \subseteq \mathcal{A}$, and hence that $X_D(\pi(v)) \in \mathcal{A}'$ (by the construction of $\mathcal{A}$). Since $\pi'(v) = \pi(v)$, we have $X_D(\pi'(v)) \in \mathcal{A}'$.
- Case 3: $\alpha = r(v, v')$. Since the rule ($*$) never adds any role atoms, we must have $r(v, v') \in q$. If $v \notin \mathsf{Ind}(\mathcal{A}')$, then $r(v, v')$ would have been removed while applying ($*$) to some ancestor of v. Hence, $v \in \mathsf{Ind}(\mathcal{A}')$. The rule ($*$) is not applicable to q', so we must also have $v' \in \mathsf{Ind}(\mathcal{A}')$. As π is a match for q, we have $r(\pi(v), \pi(v')) \in \mathcal{A}$. As $\mathcal{A}$ has the same role assertions as $\mathcal{A}'$ when restricted to individuals in $\mathcal{A}'$, we obtain $r(\pi(v), \pi(v')) \in \mathcal{A}'$, and thus $r(\pi'(v), \pi'(v')) \in \mathcal{A}'$.

We have thus found a match π' for q' in $\mathcal{I}_{\mathcal{A}'}$ with $\pi'(x) = a$. As q' is a definition-rewriting of $q \in \varphi$, it appears as a disjunct of $\varphi'(x)$. It follows that $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}'}, \varphi')$.

Next suppose that $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}'}, \varphi')$. Then there exists a disjunct q' of $\varphi'(x)$ such that $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}'}, q')$. We know from the construction of $\varphi'(x)$ that there exists a dtree-CQ q which is a disjunct of $\varphi(x)$ such that q' is a definition-rewriting of q. Let $q_0 = q, q_1, \ldots, q_n = q'$ be the sequence of definition replacements taking q to q'. We create a sequence of ABoxes $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_n = \mathcal{A}'$ such that (a) for every $0 \leq i \leq n$, $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}_i}, q_i)$, and (b) for every $0 \leq i < n$, there is a homomorphism from $\mathcal{I}_{\mathcal{A}_i,\mathcal{T}'}$ to $\mathcal{I}_{\mathcal{A}_{i+1},\mathcal{T}'}$. The base case is $i = n$, in which case (a) is immediate, and (b) is inapplicable. For the induction step, suppose that statement (a) holds for all $k + 1 \leq i \leq n$ and (b) holds for $k + 1 \leq i < n$. We know that q_{k+1} is obtained from q_k by definition replacement, so there is a unique atom $X_D(v)$ which appears in q_{k+1} but not q_k. Define the ABox $\mathcal{A}_k$ as follows:

$$
\begin{aligned}
\mathcal{A}_k = \mathcal{A}_{k+1} &\setminus \{X_D(\pi_{k+1}(v))\} \\
&\cup \{B(\pi_{k+1}(v)) \mid B(v) \in q_k, \mathcal{T}' \models X_D \sqsubseteq B\} \\
&\cup \bigcup_{r(v,v') \in q_k, \mathcal{T}' \models X_D \sqsubseteq \exists r.C_{q_k,v'}} \mathsf{inst}(\exists r.C_{q_k,v'}, \pi_{k+1}(v))
\end{aligned}
$$

By the induction hypothesis, we have that $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}_{k+1}}, q_{k+1})$, and so there is a match π_{k+1} for q_{k+1} in $\mathcal{I}_{\mathcal{A}_{k+1}}$. We use π_{k+1} to define a function π_k as follows:

- if v appears both in q_k and q_{k+1}, then $\pi_k(v) = \pi_{k+1}(v)$
- if $r(v, v') \in q_k$, v appears in q_{k+1}, but v' does not appear in q_{k+1}, then π_k maps v' and its descendants to the corresponding individuals in $\mathsf{inst}(\exists r.C_{q_k,v'}, \pi_{k+1}(v))$ (defined in the obvious way)

It is easily verified that π_k defines a match for q_k in $\mathcal{I}_{\mathcal{A}_k}$, so point (a) is satisfied. For (b), it is sufficient to exhibit a homomorphism from $\mathcal{I}_{\mathcal{A}_k}$ to $\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}$. For every individual $b \in \mathsf{Ind}(\mathcal{A}_k) \cap \mathsf{Ind}(\mathcal{A}_{k+1})$, we set $h(b) = b$. This ensures that if $B(b) \in \mathcal{A}_k \cap \mathcal{A}_{k+1}$, then $h(b) \in B^{\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}}$ and likewise for role assertions in $\mathcal{A}_k \cap \mathcal{A}_{k+1}$. Also note that if $B(\pi_{k+1}(v)) \in \mathcal{A}_k \setminus \mathcal{A}_{k+1}$, then $\mathcal{T}' \models X_D \sqsubseteq B$. Since π_{k+1}

is a match for q_{k+1} in $\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}$ and $X_D(v) \in q_{k+1}$, we must have $\pi_{k+1}(v) \in X_D^{\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}}$, hence $\pi_{k+1}(v) \in B^{\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}}$. The remaining individuals $b \in \mathsf{Ind}(\mathcal{A}_k) \setminus \mathsf{Ind}(\mathcal{A}_{k+1})$ belong to the union of the sets $\mathsf{inst}(\exists r.C_{q_k,v'}, \pi_{k+1}(v))$. Since $\pi_{k+1}(v) \in X_D^{\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}}$ and $\mathcal{T}' \models X_D \sqsubseteq \exists r.C_{q_k,v'}$, we must have $\pi_{k+1}(v) \in \exists r.C_{q_k,v'}^{\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}}$. We can thus extend h to the individuals in $\mathsf{Ind}(\mathcal{A}_k) \setminus \mathsf{Ind}(\mathcal{A}_{k+1})$ so as to satisfy all assertions in the sets $\mathsf{inst}(\exists r.C_{q_k,v'}, \pi_{k+1}(v))$, thereby obtaining the desired homomorphism from $\mathcal{I}_{\mathcal{A}_k}$ to $\mathcal{I}_{\mathcal{A}_{k+1},\mathcal{T}'}$ and completing our inductive argument. Now since $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}_0}, q)$ and the query q uses only symbols from $\Sigma = \mathsf{sig}(\mathcal{T})$, it follows that the Σ-reduct $\mathcal{A}$ of $\mathcal{A}_0$ is such that $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}}, q)$. We thus have $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}}, \varphi)$, hence $\mathcal{A}, \mathcal{T} \models A(a)$ and $\mathcal{A}_0, \mathcal{T} \models A(a)$. Since $\mathcal{T}' \models \mathcal{T}$, we must also have $\mathcal{A}_0, \mathcal{T}' \models A(a)$, or equivalently, $a \in A^{\mathcal{I}_{\mathcal{A}_0,\mathcal{T}'}}$. By composing the homomorphisms $h_0, \dots, h_{n-1}$, we obtain a homomorphism from $\mathcal{I}_{\mathcal{A}_0,\mathcal{T}'}$ to $\mathcal{I}_{\mathcal{A}_n,\mathcal{T}'} = \mathcal{I}_{\mathcal{A}',\mathcal{T}'}$. It follows that $a \in A^{\mathcal{I}_{\mathcal{A}',\mathcal{T}'}}$, hence $\mathcal{A}', \mathcal{T}' \models A(a)$.

For the other direction, suppose that $A(x)$ is FO-rewritable relative to $\mathcal{T}'$ and the full signature $\mathsf{sig}(\mathcal{T}')$. Then there exists a UCQ $\varphi'(x)$ such that $\mathsf{cert}_{\mathcal{T}}(A(x), \mathcal{A}') = \mathsf{ans}(\mathcal{I}'_{\mathcal{A}}, \varphi)$ for every $\mathsf{sig}(\mathcal{T}')$-ABox $\mathcal{A}'$. Let $\varphi(x)$ be the UCQ obtained from $\varphi'(x)$ by removing all CQs in the disjunction which use symbols outside $\mathsf{sig}(\mathcal{T})$. Then for every $\mathsf{sig}(\mathcal{T})$-ABox $\mathcal{A}$, we have the following equivalences:

$$\begin{aligned} a \in \mathsf{cert}_{\mathcal{T}}(A(x), \mathcal{A}) &\Leftrightarrow \quad a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}}, \varphi') \\ &\Leftrightarrow \quad a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}}, \varphi) \end{aligned}$$

It follows that $\varphi(x)$ is a FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and the full signature $\mathsf{sig}(\mathcal{T})$. □

B.2 Proof of Theorem 13

We aim at proving Theorem 13. In view of Theorem 9, it is clearly sufficient to show the following.

Lemma 37. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox in normal form, $A(x)$ be an AQ, and $k = 2^{3n^2}$. If there is a tree-shaped ABox $\mathcal{A}$ with root a_0 such that $\mathcal{A}, \mathcal{T} \models A(a_0)$ and $\mathcal{A}|_k, \mathcal{T} \not\models A(a_0)$, then there is a linear ABox $\mathcal{A}'$ with the same properties.*

Proof. Let $\mathcal{A}$ be a tree-shaped ABox with root a_0 such that $\mathcal{A}, \mathcal{T} \models A(a_0)$ and $\mathcal{A}|_k, \mathcal{T} \not\models A(a_0)$, where $k = 2^{3n^2}$. Because of Lemma 35, we may assume that $\mathcal{A}$ is dtree-shaped.

We first show that we can w.l.o.g. assume that in $\mathcal{A}$, there is only a single individual on level $k+1$, and no individuals on any level $> k+1$. To this end, let $b_0, \dots, b_n$ be the individuals in $\mathcal{A}$ on level $k+1$. Moreover, let $\mathcal{A}_i$ be the ABox obtained from $\mathcal{A}$ by dropping all subtree-ABoxes rooted at $b_0, \dots, b_i$ and let ℓ be smallest such that $\mathcal{A}_\ell, \mathcal{T} \models A(a_0)$. Note that $\ell < n$ since $\mathcal{A}_\ell = \mathcal{A}|_k$. The ABox $\mathcal{B}$ is obtained from $\mathcal{A}$ as follows:

1. whenever $\exists r.A \sqsubseteq B \in \mathcal{T}$, $\mathcal{A}, \mathcal{T} \models A(b_i)$ with $i > \ell$, and c_i is the predecessor of b_i in $\mathcal{A}$, then add the assertion $B(c_i)$;
2. remove the subtrees rooted at $b_0, \dots, b_{\ell-1}, b_{\ell+1}, \dots, b_n$.

We have $\mathcal{B}, \mathcal{T} \models A(a_0)$ and $\mathcal{B}|_k, \mathcal{T} \not\models A(a_0)$. The former is a consequence of $\mathcal{A}_\ell, \mathcal{T} \models A(a_0)$, the latter a consequence of $\mathcal{A}_{\ell+1}, \mathcal{T} \not\models A(a_0)$. Moreover, $\mathcal{B}$ has only a single individual on level $k+1$, which is b_ℓ. Manipulate $\mathcal{B}$ further by adding $B(b_\ell)$ whenever $\exists r.A \sqsubseteq B \in \mathcal{T}$ and $\mathcal{B}, \mathcal{T} \models A(c)$ for some individual c with $r(b_\ell, c) \in \mathcal{B}$, and then removing all individuals on level $> k+1$. We can now use $\mathcal{B}$ instead of $\mathcal{A}$.

We can also assume w.l.o.g. that when $\mathcal{A}|_k, \mathcal{T} \models B(b)$ for any assertion $B(b)$, then $B(b) \in \mathcal{A}$: if this is not the case, simply add the missing assertions. Now, let $a_0, \dots, a_{k+1}$ be the only path in $\mathcal{A}$ of length $k+1$, and let $\mathcal{A}'$ be the restriction of $\mathcal{A}$ to assertions that involve only the individuals $a_0, \dots, a_{k+1}$. Clearly, $\mathcal{A}$ is linear and since $\mathcal{A}'|_k \subseteq \mathcal{A}|_k$, we have $\mathcal{A}'|_k, \mathcal{T} \not\models A(a_0)$. It thus remains to show that $\mathcal{A}', \mathcal{T} \models A(a_0)$. Since $\mathcal{A}, \mathcal{T} \models A(a_0)$ and $\mathcal{T}$ is in normal form, it clearly suffices to argue that if $r(a_i, b) \in \mathcal{A}$ with $i \leq k$, $\mathcal{A}, \mathcal{T} \models A(b)$, and $\exists r.A \sqsubseteq B \in \mathcal{T}$, then $\mathcal{A}', \mathcal{T} \models B(a_i)$. This is straightforward: if $\mathcal{A}, \mathcal{T} \models A(b)$, then $\mathcal{A}|_k, \mathcal{T} \models A(b)$ since the subtree rooted at b is identical in $\mathcal{A}$ and $\mathcal{A}|_k$. Thus, $\mathcal{A}|_k, \mathcal{T} \models B(a_i)$ and, by assumption, we have $B(a_i) \in \mathcal{A}$, thus $B(a_i) \in \mathcal{A}'$. □

B.3 Proof of Theorem 15

Theorem 15. *Deciding FO-rewritability of an AQ relative to an $\mathcal{EL}$ TBox and the full ABox signature is in* PSPACE.

Proof. Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, Σ the full ABox signature, and $A(x)$ an AQ. Set $m = 1$ and $k = 2^{3|\mathcal{T}|^2}$, and let $\mathfrak{A}$ be the automaton from Point 1 of Theorem 30. By Theorem 13, $L(\mathfrak{A}) = \emptyset$ iff $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$. Observe that, since $m = 1$, $\mathfrak{A}$ is an alternating automaton on finite words, rather than on finite trees. It is well-known that the emptiness problem for such automata is PSPACE-complete, see for example [Serre, 2006]. To obtain an overall PSPACE decision procedure, it is additionally necessary to construct the automaton $\mathfrak{A}$ on the fly while checking its emptiness, which is standard. □

B.4 Proof of Theorem 16

We first present the reduction for Point 1 of Theorem 16. Let $M = (Q, \Omega, \Gamma, \delta, q_0, q_{\mathsf{acc}}, q_{\mathsf{rej}})$ be a DTM that solves a PSPACE-complete problem and $p(\cdot)$ its polynomial space bound. To simplify technicalities, we w.l.o.g. make the following assumptions about M. We assume that, when started in *any* (not necessarily initial) configuration C, then the computation of M terminates and uses at most $p(n)$ tape cells when n is the number of tape cells that are non-blank in C. We also assume that M always terminates with the head on the right-most tape cell, that it never attempts to move left on the left-most end of the tape, and that there are no transitions defined for q_{acc} and q_{rej}. Let $x \in \Omega^*$ be an input to M of length n. Our aim is to construct a TBox $\mathcal{T}$ and select a concept name B such that B is *not* FO-rewritable relative to $\mathcal{T}$ and the full signature iff M accepts x.

By Theorem 7, non-FO-rewritability of B w.r.t. $\mathcal{T}$ is witnessed by a sequence of tree-shaped ABoxes of increasing depth. In the reduction, such witness ABoxes take the form of longer and longer chains representing the computation of M on x, repeated over and over again. Specifically, the tape contents, the current state, and the head position are represented using the elements of $\Gamma \cup (\Gamma \times Q)$ as concept names. Each ABox element represents one tape cell of one configuration, the role name r is used to move between consecutive tape cells, the role name s is used to move between successor configurations inside the same computation, and the role name t is used to separate computations. To illustrate, suppose the computation of M on $x = ab$ consists of the two configurations qab and $aq'b$.[1] This is represented by ABoxes of the form

$$\{r(b_1,b_0), s(b_2,b_1), r(b_3,b_2), t(b_4,b_3), r(b_5,b_4), \ldots, r(b_m,b_{m-1})\}$$

where additionally, the concept (a,q) is asserted for $b_0, b_4, b_8, \ldots,$ b is asserted for $b_1, b_5, b_9, \ldots,$ a for $b_2, b_6, b_{10}, \ldots,$ and (b,q') for $b_3, b_7, b_{11}, \ldots.$ If M accepts x, then B is propagated backwards along these chains (from b_0 to b_1 etc.) unboundedly far, starting from a single explicit occurrence of B asserted for b_0. To ensure that the chain in the ABox properly represents the computation of M on x, we will make sure that B is already implied by a subchain of bounded length when there is a defect in the computation, and thus the unbounded propagation of B gets disrupted resulting in FO-rewritability of B relative to $\mathcal{T}$.

The following CI in $\mathcal{T}$ results in backwards propagation of B provided that every ABox individual is labeled with at least one symbol from $\Gamma \cup (\Gamma \times Q)$. It also makes sure that t-transitions occur exactly after the accepting state was reached:

$$\exists r.(A \sqcap B) \sqsubseteq B \quad \text{for all } A \in \Gamma \cup (\Gamma \times (Q \setminus \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\})) \quad (1)$$

$$\exists s.(A \sqcap B) \sqsubseteq B \quad \text{for all } A \in \Gamma \cup (\Gamma \times (Q \setminus \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\})) \quad (2)$$

$$\exists t.(A \sqcap B) \sqsubseteq B \quad \text{for all } A \in \Gamma \times \{q_{\mathsf{acc}}\}. \quad (3)$$

There are many properties of witness ABoxes that need to be taken care of. We start with enforcing that every tape cell has a unique label:

$$A \sqcap A' \sqsubseteq B \quad \text{for all distinct } A, A' \in \Gamma \cup (\Gamma \times Q) \quad (4)$$

We next enforce that there is not more than one head position per configuration:

$$(a,q) \sqsubseteq H \quad \text{for all } (a,q) \in \Gamma \times Q \quad (5)$$

$$\exists r^i.H \sqcap \exists r^j H \sqsubseteq B \quad \text{for } i < j < p(n) \quad (6)$$

and that there is at least one head position per configuration:

$$a \sqsubseteq \overline{H} \quad \text{for all } a \in \Gamma \quad (7)$$

$$\overline{H} \sqcap \exists r.\overline{H} \sqcap \cdots \sqcap \exists r^{p(n)-1}.\overline{H} \sqsubseteq B \quad (8)$$

where H is a concept name indicating that the head is on the current cell and $\overline{H}$ indicating that this is not the case. Note that, whenever one of the desired properties is violated in an ABox, then B is implied by a subchain of length at most $p(n)$, thus its unbounded propagation is disrupted.

[1] $uqv \in \Gamma^* Q \Gamma^*$ means that M is in state q, the tape left of the head is labeled with u, and starting from the head position, the remaining tape is labeled with v.

For technical reasons, we also want to ensure that configurations have length exactly $p(n)$ (with the possible exception of the first configuration, which can be shorter), again via disruption of propagation:

$$\exists r^{p(n)}.\top \sqsubseteq B \quad (9)$$

$$\exists S.\exists r^i.\exists S'.\top \sqsubseteq B \quad \text{for all } i < p(n)-1 \text{ and } S, S' \in \{s,t\} \quad (10)$$

We now enforce that the transition relation is respected and that the content of tape cells which are not under the head does not change. Let forbid denote the set of all tuples (A_1, A_2, A_3, A) with $A_i \in \Gamma \cup (\Gamma \times Q)$ such that whenever three consecutive tape cells in a configuration c are labeled with A_1, A_2, A_3, then in the successor configuration c' of c, the tape cell corresponding to the middle cell *cannot* be labeled with A.

$$A \sqcap \exists r^i.\exists s.\exists r^{p(n)-i-2}.(A_3 \sqcap \exists r.(A_2 \sqcap \exists r.A_1)) \sqsubseteq B \quad (11)$$

for all $0 \leq i < p(n)$ and $(A_1, A_2, A_3, A) \in \mathsf{forbid}$.

It remains to set up the initial configuration. Recall that witness ABoxes consist of repeated computations of M, which ideally we would all like to start in the initial configuration for input x. It does not seem possible to enforce this for the first computation in the ABox, so we live with this computation starting in some unknown configuration. Then, we utilize the final states q_{acc} and q_{rej} to enforce that all computations in the ABox except the first one must start with the initial configuration for x. Let $A_0^{(0)}, \ldots, A_{p(n)-1}^{(0)}$ be the concept names that describe the initial configuration, i.e., when the input x is $x_0 \cdots x_{n-1}$, then $A_0^{(0)} = (x_0, q_0)$, $A_i^{(0)} = x_i$ for $1 \leq i < n$, and $A_i^{(0)} = x_i$ is the blank symbol for $n \leq i < p(n)$. Now put

$$\exists r^i.\exists t.\top \sqsubseteq A_i^{(0)} \text{ for all } 0 \leq i < p(n). \quad (12)$$

The following lemma establishes the correctness of our reduction.

Lemma 38. *B is* not *FO-rewritable relative to $\mathcal{T}$ and the full signature iff M accepts x.*

Proof. "if". (sketch) Assume that M accepts x. By Theorem 9, it is enough to show that there is a tree-shaped ABox $\mathcal{A}$ with root a_0 such that $\mathcal{A}, \mathcal{T} \models B(a_0)$ and $\mathcal{A}|_k, \mathcal{T} \not\models B(a_0)$, where $k = 2^{3|\mathcal{T}|^2}$. Let $C_1, \ldots, C_m$ be a sequence of configurations of length $p(n)$ obtained by sufficiently often repeating the accepting computation of M on x so that $|C_1| + \cdots + |C_m| > k$. We can convert $C_1, \ldots, C_m$ into the desired witness ABox $\mathcal{A}$ in a straightforward way: introduce one individual name for each tape cell in each configuration, use the role name r to connect cells within the same configuration, the role name s to connect configurations, the role name t to connect computations, and the concept names from $\Gamma \cup (\Gamma \times Q)$ to indicate the tape inscription, current state, and head position. We obtain a linear ABox $\mathcal{A}$ of depth $> k$. Add $B(a)$ with a the only leaf of $\mathcal{A}$. It can be verified that $\mathcal{A}$ is as required.

"only if". Assume that B is not FO-rewritable relative to $\mathcal{T}$. Let step_M be the maximum number of steps M makes starting from any configuration of length $p(n)$ before entering a final state, and let $= (2 \cdot \mathsf{step}_M + 2) \cdot p(n) + p(n)$.

Claim 1. There is a dtree-shaped ABox $\mathcal{A}$ with root a_0 such that $\mathcal{A}$ is closed under applications of CIs (4) to (10), $\mathcal{A}, \mathcal{T} \models B(a_0)$, and $\mathcal{A}|_k, \mathcal{T} \not\models B(a_0)$.

Proof of claim. By Theorem 9, non-FO-rewritability of B relative to $\mathcal{T}$ implies the existence of a tree-shaped ABox $\mathcal{A}'$ with root a_0 such that $\mathcal{A}', \mathcal{T} \models B(a_0)$ and $\mathcal{A}'|_{k+p(n)+1}, \mathcal{T} \not\models B(a_0)$. By Lemma 35, we can assume that $\mathcal{A}'$ is dtree-shaped. The desired ABox $\mathcal{A}$ is obtained by closing $\mathcal{A}'$ under CIs (4) to (10). Clearly, we have $\mathcal{A}, \mathcal{T} \models B(a_0)$. Now consider the ABoxes $\mathcal{A}|_k$ and $\mathcal{A}'|_k$. Since the CIs (4) to (10) are non-recursive and of role depth at most $p(n) + 1$, all atoms $\alpha \in \mathcal{A}|_k \setminus \mathcal{A}'|_k$ are such that $\mathcal{A}'|_{k+p(n)+1}, \mathcal{T} \models \alpha$. Since $\mathcal{A}'|_{k+p(n)+1}, \mathcal{T} \not\models B(a_0)$, we thus have $\mathcal{A}|_k, \mathcal{T} \not\models B(a_0)$, as required. This finishes the proof of Claim 1.

Let $\mathcal{T}^-$ be the restriction of $\mathcal{T}$ to CIs (1) to (3). Since $\mathcal{A}$ and thus also $\mathcal{A}|_k$ is closed under applications of CIs (4) to (10), all CIs in $\mathcal{T}^-$ are of the form $C \sqsubseteq B$, and B does not occur on the left-hand sides of CIs (4) to (10), we have $\mathcal{A}, \mathcal{T}^- \models B(a_0)$ and $\mathcal{A}|_k, \mathcal{T}^- \not\models B(a_0)$. We can assume w.l.o.g. that there is an individual a on level $k + 1$ of $\mathcal{A}$ such that $\mathcal{A}^-, \mathcal{T}^- \not\models B(a_0)$, where $\mathcal{A}^-$ is $\mathcal{A}$ with the subtree rooted at a dropped. Let $b_0, \ldots, b_{k+1}$ be the (backwards) path in $\mathcal{A}$ from a to a_0.

Claim 2. For $1 \leq i \leq k + 1$, we have

(a) $\mathcal{A}, \mathcal{T}^- \models B(b_i)$;

(b) $\mathcal{A}^-, \mathcal{T}^- \cup \{B(b_i)\} \models B(a_0)$.

Proof of claim.

(a) Follows from the fact that $\mathcal{A}, \mathcal{T}^- \models B(a_0)$, $\mathcal{A}^-, \mathcal{T}^- \not\models B(a_0)$, and that all CIs in $\mathcal{T}^-$ are of the form $C \sqsubseteq B$ with C of role depth one.

(b) Fix a b_i with $1 \leq i \leq k+1$. Let $\mathcal{A}^+$ be the ABox obtained from $\mathcal{A}$ by

- dropping all subtrees rooted at successors of b_i and
- adding all concept assertions $X(b_i)$ with $\mathcal{A}, \mathcal{T}^- \models X(b_i)$.

Since $\mathcal{A}, \mathcal{T}^- \models B(b_0)$ and all CIs in $\mathcal{T}^-$ are of role depth one, we have $\mathcal{A}^+, \mathcal{T}^- \models B(b_0)$. We have $\mathcal{A}^+ \subseteq \mathcal{A}^- \cup \{B(b_i)\}$ as all CIs in $\mathcal{T}^-$ are of the form $C \sqsubseteq B$. Thus, $\mathcal{A}^-, \mathcal{T}^- \cup \{B(b_i)\} \models B(a_0)$ and the proof of Claim 2 is finished.

For $1 \leq i \leq k + 1$ and $R \in \{r, s, t\}$, we say that b_i is an *R-individual* if $R(b_i, b_{i-1}) \in \mathcal{A}$. Let o be smallest index i such that b_i is an s-individual or t-individual. By CI (9), Point (b) of Claim 2, and since $\mathcal{A}^-, \mathcal{T}^- \not\models B(a_0)$ we have $o \leq p(n)$. Similarly, by CIs (9) and (10) we can split the chain $b_o, \ldots, b_{k+1}$ into consecutive subchains of length precisely $p(n)$ such that the first individual in each subchain is an s-individual or t-individual and all others are r-individuals. By CIs (4) to (8), each such subchain represents a unique configuration of M of length $p(n)$. We thus obtain a sequence of configurations $C_1, \ldots, C_\ell$ with $\ell > 2 \cdot \mathsf{step}_M + 1$. By CIs (11), for all C_i, C_{i+1} where C_{i+1} starts with an s-individual, C_{i+1} must be a successor configuration of C_i. Since M terminates after at most step_M steps starting from any configuration, there is a C_i with $i < \mathsf{step}_M$ such that C_i is a final configuration. By CI (2) and q_{acc} and q_{rej} are excluded in Point (a) of Claim 2, C_{i+1} must start with a t-individual. By CI (12), C_{i+1} must be the initial configuration of M on input x. The sequence $C_{i+1}, \ldots, C_\ell$ is still of length exceeding $\mathsf{step}_M + 1$. It follows that an initial piece of this sequence represents the computation of M on x, say $C_{i+1}, \ldots, C_j$ with $j < \ell$. As above, we can argue that C_{j+1} must start with a t-individual. Moreover, since q_{rej} is excluded in CI (3), C_j must be an accepting configuration, thus the computation of M on x is accepting. □

Now we turn to Point 2 of Theorem 16, which is proven by a reduction from the word problem for polynomially space-bounded alternating Turing machines (ATMs). Let $M = (Q_\exists, Q_\forall, \Omega, \Gamma, \delta, q_0, q_{\mathsf{acc}}, q_{\mathsf{rej}})$ be a polynomially space-bounded ATM that solves an EXPTIME-complete problem. We assume $q_{\mathsf{acc}}, q_{\mathsf{rej}} \notin Q_\exists \cup Q_\forall$, and thus no transitions are defined for $q_{\mathsf{acc}}, q_{\mathsf{rej}}$. We may also assume w.l.o.g. that both for existential and universal states, there are exactly two transitions. Each transition has the form (q, a, m) with $m \in \{-1, +1\}$, i.e., the Turing machine cannot make a transition without moving its head. As in the previous reduction, we assume that, when started in *any* (not necessarily initial) configuration, the computation of M terminates after at most exponentially many steps. Let $x \in \Omega^*$ be an input of length n to M. We construct a TBox $\mathcal{T}$ and signature Σ such that a selected concept name $B \notin \Sigma$ is not FO-rewritable relative to $\mathcal{T}$ and Σ iff M accepts x. The construction differs in some crucial aspects from the PSPACE one given before:

(i) witness ABoxes are tree-shaped and represent repeated computation trees rather than repeated linear DTM computations;

(ii) an individual represents a whole configuration rather than only one tape cell;

(iii) the computation will proceed forward along role edges rather than backward.

In Point (i), "repeated computation trees" means that one copy of the tree is repeatedly appended to at least one leaf of another copy of the tree. The concept name B then propagates bottom-up through these repeated trees.

We use concept names from the set

$$\mathcal{C} = (\Gamma \times [1, \ldots, p(n)]) \cup (\Gamma \times Q \times [1, \ldots, p(n)]) \subseteq \Sigma$$

to specify contents of the tape cells, the head position and the current state. For easy reference, we use $\mathcal{C}_i$ to denote the restriction of $\mathcal{C}$ to all tuples with last component i. Auxilliary concept names $P_1, \ldots, P_{p(n)}$, which are not in Σ, indicate that the contents of a given tape cell have been specified:

$$A \sqsubseteq P_i \quad \text{for all } A \in \mathcal{C}_i, 1 \leq i \leq p(n) \tag{13}$$

The concept $\mathsf{Tape} \notin \Sigma$ does the same for the whole tape:

$$P_1 \sqcap \cdots \sqcap P_n \sqsubseteq \mathsf{Tape} \tag{14}$$

We use the role name r_1 to link successor configurations of existential restrictions and first successor configurations of universal configurations, and r_2 to link second successor configurations of universal configurations. The following CIs make sure that (i) whenever an individual describes a configuration, then every tape cell is labeled with some symbol in this configuration and (ii) the appropriate successors are present: for all $(a, q, i) \in \mathcal{C}$ with $q \in Q_\exists$, put

$$\mathsf{Tape} \sqcap (a, q, i) \sqcap \exists r_1.(\mathsf{Tape} \sqcap B) \sqsubseteq B. \tag{15}$$

For all $(a, q, i) \in \mathcal{C}$ with $q \in Q_\forall$, put

$$\mathsf{Tape} \sqcap (a, q, i) \sqcap \exists r_1.(\mathsf{Tape} \sqcap B) \sqcap \exists r_2.(\mathsf{Tape} \sqcap B) \sqsubseteq B \tag{16}$$

By disrupting the propagation of B, we can ensure that every cell is labeled with at most one symbol, that there is exactly one head and state, and that symbols that are not under the head do not change when the TM makes a transition. Technically, this is achieved with the help of a concept name $E \notin \Sigma$, which signals an error. We also use auxiliary concept names $H_i, \overline{H}_i \notin \Sigma$. The axioms to enforce at most one symbol per cell and exactly one head per configuration are as follows:

$$A \sqcap A' \sqsubseteq E \quad \text{for all distinct } A, A' \in \mathcal{C}_i, 1 \leq i \leq p(n) \tag{17}$$

$$(a, q, i) \sqsubseteq H_i \quad \text{for all } (a, q, i) \in \mathcal{C}_i, 1 \leq i \leq p(n) \tag{18}$$

$$(a, i) \sqsubseteq \overline{H}_i \quad \text{for all } (a, i) \in \mathcal{C}_i, 1 \leq i \leq p(n) \tag{19}$$

$$H_i \sqcap H_j \sqsubseteq E \quad \text{for } 1 \leq i < j \leq p(n) \tag{20}$$

$$\overline{H}_1 \sqcap \overline{H}_2 \sqcap \cdots \sqcap \overline{H}_{p(n)} \sqsubseteq E \tag{21}$$

Let C_i^a denote the restriction of C_i to those tuples with first component a. To ensure that the symbol under the head does not change, we include the following axiom

$$H_i \sqcap A_{a,j} \sqcap \exists r_\ell.B_{b,j} \sqsubseteq E \tag{22}$$

for every $A_{a,j} \in \mathcal{C}_j^a$ and $B_{b_j} \in \mathcal{C}_j^b$ with $a \neq b$, all distinct pairs $i, j \in [1, \ldots, p(n)]$, and $\ell \in \{1, 2\}$. Errors in a computation tree imply B at the root of that tree:

$$\mathsf{Tape} \sqcap \exists r_\ell.E \sqsubseteq E \quad \text{for } \ell \in \{1, 2\} \tag{23}$$

$$E \sqsubseteq B \tag{24}$$

To ensure that the transition relation is respected, we use the following CIs: for all $(a, q, i) \in \mathcal{C}$ with $q \in Q_\exists$ and $\delta(q, a) = \{(q_1, a_1, m_1), (q_2, a_2, m_2)\}$, and all tuples $(a', q', i'), (a'', i) \in \mathcal{C}$ with $(q', i') \neq (q_\ell, i + m_\ell)$ or $a'' \neq a_\ell$ for all $\ell \in \{1, 2\}$, put

$$(a, q, i) \sqcap \exists r_1.((a'', i) \sqcap (a', q', i')) \quad \sqsubseteq \quad E \tag{25}$$

For all $(a, q, i) \in \mathcal{C}$ with $q \in Q_\forall$ and $\delta(q, a) = \{(q_1, a_1, m_1), (q_2, a_2, m_2)\}$, put for $\ell \in \{1, 2\}$:

$$(a, q, i) \sqcap \exists r_\ell.(a', q', i') \sqsubseteq E \quad \text{for all } (a', q', i') \in \mathcal{C} \text{ with } (q', i') \neq (q_\ell, i + m_\ell) \tag{26}$$

$$(a, q, i) \sqcap \exists r_\ell.(a', i) \sqsubseteq E \quad \text{for all } a' \in \Gamma \text{ with } a' \neq a_\ell \tag{27}$$

We did not yet introduce a way to start the propagation of B (except errors). This is achieved via an additional concept name $\mathsf{Start} \in \Sigma$. To ensure that the represented computation is accepting, Start implies B only at accepting final configurations of the TM:

$$\mathsf{Start} \sqcap (a, q_{\mathsf{acc}}, i) \sqsubseteq B \tag{28}$$

Since the depth of computation trees is bounded, we did not yet achieve unbounded propagation. The following CI allows the propagation of B along multiple computation trees plugged together in the way described above. It also sets up the initial configuration in all computation trees except the topmost one. Note that we use a different role name t for plugging trees together:

$$(a, q_{\mathsf{acc}}, i) \sqcap \exists t.(A_1^0 \sqcap \ldots \sqcap A_{p(n)}^0 \sqcap B) \sqsubseteq B \tag{29}$$

where (abusing notation) we use $A_1^0, \ldots, A_{p(n)}^0$ to denote the sequence of symbols from $\mathcal{C}$ that corresponds to the initial configuration. We also have to prevent continued travel along r_1 and r_2 when the computation has stopped: for all $(a, q, i) \in \mathcal{C}$ with $q \in \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\}$, put

$$(a, q, i) \sqcap \exists r_\ell.\top \sqsubseteq E \quad \text{for } \ell \in \{1, 2\} \text{ and } q \in \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\}. \tag{30}$$

Lemma 39. *B is* not *FO-rewritable relative to $\mathcal{T}$ and Σ iff M accepts x.*

Proof. "if". (sketch) Assume that M accepts x. Then there is an accepting computation tree T of M on x. By Theorem 9, it is enough to show that there is a dtree-shaped ABox $\mathcal{A}$ with root a_0 such that $\mathcal{A}, \mathcal{T} \models B(a_0)$ and $\mathcal{A}|_k, \mathcal{T} \not\models B(a_0)$, where $k = 2^{3|\mathcal{T}|^2}$. Note that the computation tree T can be converted into a dtree-shaped ABox $\mathcal{A}_T$ in a straightforward way: introduce one individual name for each configuration, use the concept names from $\mathcal{C}$ to describe the actual configurations at their corresponding individual names, and use the role names r_1 and r_2 to connect configurations in the intended way. By repeatedly appending copies of the ABox $\mathcal{A}_T$ to leaves of this ABox using the role name t, generate a dtree-shaped ABox $\mathcal{A}$ of depth exceeding k (it is enough to start with one copy of $\mathcal{A}_T$, append a second copy at a single leaf of the first copy, a third copy at a single leaf of the second copy, and so on). Finally, add the concept name Start to all leaves of $\mathcal{A}$. It can be verified that $\mathcal{A}$ is as required.

"only if". Assume that B is not FO-rewritable relative to $\mathcal{T}$, let step_M be the maximum length of a path in a computation tree of M (starting at any configuration, not necessarily an initial one), and let $k = 2 \cdot \mathsf{step}_M + 1$. By Theorem 7 and Lemma 35, we find a dtree-shaped ABox $\mathcal{A}$ of depth exceeding k and with root a_0 such that $\mathcal{A}, \mathcal{T} \models B(a_0)$ and $\mathcal{A}|_k, \mathcal{T} \not\models B(a_0)$. We can assume w.l.o.g. that there is an individual a on level $k + 1$ of $\mathcal{A}$ such that $\mathcal{A}^-, \mathcal{T} \not\models B(a_0)$, where $\mathcal{A}^-$ is $\mathcal{A}$ with the subtree rooted at a dropped. Let $a_0, \ldots, a_{k+1}$ be the path in $\mathcal{A}$ from a_0 to a, and call an a_i with $i > 0$ on this path an *R-individual* if $R(a_{i-1}, a_i) \in \mathcal{A}$, for $R \in \{r_1, r_2, t\}$. Let $a_0, \ldots, a_p$ be the longest prefix of $a_0, \ldots, a_{k+1}$ that does not contain any t-individuals. In what follows, a configuration C' is a *1-successor configuration* of C when C is existential and C' is a successor configuration or C is universal and C' is the first successor configuration;

C' is a 2-*successor configuration* of C when C is universal and C' is the second successor configuration. We show by induction on i that

1. $\mathcal{A}^-, \mathcal{T} \not\models E(a_i)$;
2. $\mathcal{A}, \mathcal{T} \models \mathsf{Tape}(a_i)$ and the assertions $A(a_i) \in \mathcal{A}$ with $A \in \mathcal{C}$ represent a proper configuration C_i of M;
3. if a_i is an r_ℓ-individual, $\ell \in \{1, 2\}$, then C_i is an ℓ-successor configuration of C_{i-1}.

for every $i \leq p$.

For the induction start ($i = 0$), Point 1 is true since $\mathcal{A}^-, \mathcal{T} \not\models B(a_0)$ and by CI (24), and Point 3 is vacuously true. For Point 2, first note that

($*$) for each right-hand side C of the CIs (17), (20), (21), (22), we have $\mathcal{A}, \mathcal{T} \not\models C(a_0)$.

Otherwise, we obtain $\mathcal{A}^-, \mathcal{T} \models C(a_0)$ by the claim, implying $\mathcal{A}^-, \mathcal{T} \models E(a_0)$, which is a contradiction to Point 1. Since $\mathcal{A}, \mathcal{T} \models B(a_0)$ and $\mathcal{A}^-, \mathcal{T} \not\models B(a_0)$ and all concepts in $\mathcal{T}$ are of role depth at most one, some concept name must be 'propagated up' from a_1 to a_0, which can only be due to one of the CIs (15), (16), or (23). Since all these CIs have Tape on their left-hand side, we have $\mathcal{A}, \mathcal{T} \models \mathsf{Tape}(a_0)$. By (13) and (14) and since $\mathsf{Tape} \notin \Sigma$, for $1 \leq i \leq p(n)$ we have $A(a_i) \in \mathcal{A}$ for some $A \in \mathcal{C}_i$. By ($*$), these assertions indeed represent a proper configuration C_0.

For the induction step, we start with Point 1. Assume to the contrary that $\mathcal{A}^-, \mathcal{T} \models E(a_i)$. By Point 2 of IH and the claim, we have $\mathcal{A}^-, \mathcal{T} \models \mathsf{Tape}(a_{i-1})$. By CI (23), $\mathcal{A}^-, \mathcal{T} \models E(a_{i-1})$, in contradiction to Point 1 of IH. The proof of Point 2 is exactly as in the induction start. It remains to deal with Point 3, which is a consequence of CIs (25)-(27) and the fact that, by Point 1 of IH, $\mathcal{A}^-, \mathcal{T} \not\models E(a_{i-1})$. This finishes the proof of Points 1-3.

By Point 3, the length of the configuration sequence $C_0, \ldots, C_p$ is bounded by $\mathsf{step}_M + 1$, and so p is bounded by step_M. Since $k > 2 \cdot \mathsf{step}_M + 1$, we have $k > p$ and the individual a_{p+1} exists. By choice of $a_0, \ldots, a_p$, a_{p+1} must be a t-individual, but cannot be an r_ℓ-individual for any $\ell \in \{1, 2\}$. Since $\mathcal{A}, \mathcal{T} \models B(a_0)$, $\mathcal{A}^-, \mathcal{T} \not\models B(a_0)$, and all concepts in $\mathcal{T}$ are of role depth one, some concept name must be 'propagated up' from a_{p+1} to a_p. Since CI (29) is the only CI referring to the role name t, this CI must be used in the propagation. By the left-hand side of CI (29), the set $\{A \mid A(a_{p+1}) \in \mathcal{A}\}$ includes all concept names that represent the initial configuration for x.

We can now select a set of individial names $I \subseteq \mathsf{Ind}(\mathcal{A})$ such that the restriction $\mathcal{A}|_I$ of $\mathcal{A}$ to those assertions that refer only to individuals in I is tree-shaped, rooted at a_{p+1}, and satisfies the following conditions, for all nodes $a \in I$:

(a) $\mathcal{A}, \mathcal{T} \models B(a)$ and $\mathcal{A}^-, \mathcal{T} \not\models E(a)$;

(b) $\mathcal{A}, \mathcal{T} \models \mathsf{Tape}(a)$ and the assertions $A(a) \in \mathcal{A}$ with $A \in \mathcal{C}$ represent a proper configuration C_a of M;

(c) if C_a is an existential configuration, then a has a single successor b that is an r_1-individual;

(d) if C_a is a universal configuration, then a has two successors b_1, b_2 in $\mathcal{B}$ with b_1 an r_1-individual and b_2 an r_2-individual;

(e) if $r_\ell(a, b) \in \mathcal{B}$, then C_b is an ℓ-successor configuration of C_a, $\ell \in \{1, 2\}$.

Let $a_{p+1}, \ldots, a_q$ be the shortest prefix of $a_{p+1}, \ldots, a_k$ that consists only of r_ℓ-individuals, for some $\ell \in \{1, 2\}$. We start the selection of individual names with setting $I := \{a_{p+1}, \ldots, a_q\}$. We can argue as in the analysis of the chain $a_0, \ldots, a_p$ above that $a_{p+1}, \ldots, a_q$ satisfies Points 1 to 3, for $p + 1 \leq i \leq q$. Thus, Points (b) and (e) from above are also satisfied, and so is the second part of Point (a). Note that $q \leq 2 \cdot \mathsf{step}_M$, and thus the individual a_{q+1} exists and is a t-individual, but not an r_ℓ-individual for any ℓ. A concept name X must be propagated up from a_{q+1} to a_q which must be due to CI (29). Thus, X must actually be B and we have $\mathcal{A}, \mathcal{T} \models B(a_q)$. An analysis of the CIs in $\mathcal{T}$ reveals that the upwards propagation of B from a_{q+1} to a_q cannot result in any other concept name than B being propagated further up to $a_{q-1}, \ldots, a_0$. Since we know that some concept name *is* propagated up along this path, we can derive that $\mathcal{A}, \mathcal{T} \models B(a_i)$ for all $i \leq q$. Thus, the first part of Point (a) is satisfied.

To also satisfy Points (c) and (d), we may have to select additional individual names to be included in I. During this extension of I, we will always maintain Properties (a), (b), and (e). We only treat the case of universal configurations, and leave existential ones to the reader. Assume that there is some $a \in I$ such that C_a is a universal configuration. By (a), we have $\mathcal{A}, \mathcal{T} \models B(a)$. Since $B \notin \Sigma$, this must be due to some CI. Since C_a is universal, this CI must be CI (16), and thus we find an r_i-successor a_i of a in $\mathcal{A}$ with $\mathcal{A}, \mathcal{T} \models \mathsf{Tape} \sqcap B(a_i)$, for $i \in \{1, 2\}$. We also have $\mathcal{A}^-, \mathcal{T} \not\models E(a_i)$ for $i \in \{1, 2\}$ since the contrary would imply $\mathcal{A}^-, \mathcal{T} \models B(a_0)$, and thus Point (a) is satisfied for a_1 and a_2. We can argue as before that (b) and (e) are also satisfied. This finishes the definition of I. Note that the depth of the resulting ABox $\mathcal{A}|_I$ is bounded by $\mathsf{step_M}$.

Since a_{p+1} makes true all concept names that represent the initial configuration for x, Point (b) ensures that $C_{a_{p+1}}$ is the initial configuration for x. Thus $\mathcal{A}|_I$ represents the computation of M on x and it remains to show that this computation is accepting. To this end, consider a leaf a of $\mathcal{A}|_I$. Then C_a is a final configuration, i.e., the state is q_{acc} or q_{rej}. By Point (a) and CI (24), we have $\mathcal{A}, \mathcal{T} \models B(a)$ and $\mathcal{A}^-, \mathcal{T} \not\models E(a)$. By CI (30), a does thus not have any r_1- or r_2-successors in $\mathcal{A}$. Consequently, $\mathcal{A}, \mathcal{T} \models B(a)$ is due to CI (28) or (29). In both cases, we have that $A(a) \in \mathcal{A}$ for some A of the form (a, q_{acc}, i), thus by Point (b), C_a is an accepting final configuration. $\square$