Deciding FO-Rewritability in $\mathcal{EL}$

Meghyn Bienvenu[1] and Carsten Lutz[2] and Frank Wolter[3]

[1] LRI - CNRS & Université Paris Sud, France
[2] Department of Computer Science, University of Bremen, Germany
[3] Department of Computer Science, University of Liverpool, UK

Abstract. We consider the problem of deciding, given an instance query $A(x)$, an $\mathcal{EL}$-TBox $\mathcal{T}$, and possibly an ABox signature Σ, whether $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ-ABoxes. Our main results are PSPACE-completeness for the case where Σ comprises all symbols and EXPTIME-completeness for the general case. We also show that the problem is in PTIME for classical TBoxes and that every instance query is FO-rewritable into a polynomial-size FO query relative to every (semi)-acyclic TBox (under some mild assumptions on the data).

1 Introduction

Over the last years, query answering over instance data has developed into one of the most prominent problems in description logic (DL) research. Many approaches aim at utilizing relational databases systems (RDBMSs), exploiting their mature technology, advanced optimization techniques, and the general infrastructure that those systems offer. Roughly, RDBMS-based approaches can be classified into *query rewriting approaches*, where the original query and the DL TBox are compiled into an SQL query that is passed to the RDBMS for execution [5], and *combined approaches*, where the consequences of the TBox are materialized in the data in a compact form and some query rewriting is used to ensure correct answers despite the compact representation [12, 11]. This division is by no means strict, as illustrated by the approach presented in [7] which is based on query rewriting, but also has strong similarities with combined approaches.

A fundamental difference between the query rewriting approach and the combined approach is that, in query rewriting, an exponential blowup of the query is often unavoidable [8] while the combined approach typically blows up both query and data only polynomially [12, 11]. It is thus unsurprising that query execution is more efficient in the combined approach than in the query rewriting approach, see the experiments in [11]. Depending on the application, however, there can still be good reasons to use pure query rewriting. *Ease of implementation*: Query rewriting approaches are often easier to implement as they do not involve a data completion phase. When the TBox is sufficiently small so that the exponential blowup of the query is not prohibitive or when only a prototype implementation is aimed at, it may not be worthwhile to implement a full combined approach. *Access limitations*: If the user does not have permission to modify the data in the database, materializing the consequences of the TBox in the data might simply be out of the question. This problem arises notably in information integration applications.

In this paper, we are interested in TBoxes formulated in the description logic $\mathcal{EL}$, which forms the basis of the OWL EL fragment of OWL 2 and is popular as a basic language for large-scale ontologies. In general, query rewriting approaches are not applicable to $\mathcal{EL}$ because instance query answering in this DL is PTIME-complete regarding data complexity while AC_0 data complexity marks the boundary of DLs for which the pure query rewriting approach can be made work [5]. For example, the query $A(x)$ cannot be answered by an SQL-based RDBMS in the presence of the very simple $\mathcal{EL}$-TBox $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$, 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. We say that $A(x)$ is not *FO-rewritable relative to* $\mathcal{T}$, alluding to the known equivalence of first-order (FO) formulas and SQL queries.

Of course, such an isolated example does not rule out the possibility that *some* $\mathcal{EL}$-TBoxes, including those that are used in applications, still enjoy FO-rewritability. For example, the query $A(x)$ is FO-rewritable relative to the $\mathcal{EL}$-TBox $\mathcal{T}' = \{A \sqsubseteq \exists r.A\}$: since the additional instances of A stipulated by $\mathcal{T}'$ are 'anonymous objects' (nulls in database parlance) rather than primary data objects, there is no unbounded propagation *through the data* and, in fact, we can simply drop $\mathcal{T}'$ when answering $A(x)$. Inspired by these observations, the aim of this paper is to study FO-rewritability on the level of individual TBoxes, essentially following the non-uniform approach initiated in [15]. In particular, we are interested in deciding, for a given instance query (IQ) $A(x)$ and $\mathcal{EL}$-TBox $\mathcal{T}$, whether q is FO-rewritable relative to $\mathcal{T}$. Sometimes, we additionally allow as a third input an ABox-signature Σ that restricts the symbols which can occur in the data [2, 3].

Our main result is that deciding FO-rewritability of IQs relative to *general* $\mathcal{EL}$-*TBoxes* (sets of concept inclusions $C \sqsubseteq D$) is PSPACE-complete when the ABox signature Σ is full (i.e., all symbols are allowed in the ABox) and EXPTIME-complete when Σ is given as an input. For proving these results, we establish some characterizations and properties that are of independent interest, such as the following:whenever an IQ is FO-rewritable, then it is FO-rewritable into a union of tree-shaped conjunctive queries. We also study more restricted forms of TBoxes, showing that FO-rewritability of IQs relative to *classical TBoxes* (sets of concept definitions $A \equiv C$ and concept implications $A \sqsubseteq C$ with A atomic, cycles allowed) is in PTIME, even when Σ is part of the input. For *semi-acyclic TBoxes* $\mathcal{T}$ (classical TBoxes without cycles that involve only concept definitions, but potentially with cycles that involve at least one concept inclusion), we observe that every IQ is FO-rewritable relative to $\mathcal{T}$ (for any ABox signature Σ) and that, under the mild assumption that the admitted databases have domain size at least two, even a polynomial-sized rewriting is possible. While it is not our primary aim in this first publication to actually generate FO-rewritings, we note that all our results come with effective procedures for doing this (the rewritings are of triple-exponential size in the worst case).

Although we focus on simple IQs of the form $A(x)$, all results in this paper also apply to instance queries of the form $C(x)$ with C an $\mathcal{EL}$-concept. The treatment of conjunctive queries (CQs) is left for future work. We also discuss the connection of FO-rewritability in $\mathcal{EL}$ to boundedness in datalog and in the μ-calculus. Proof details can be found in the appendix.

2 Preliminaries

We remind the reader that $\mathcal{EL}$-concepts are built up from concept names and the concept $\top$ using conjunction $C \sqcap D$ and existential restriction $\exists r.C$. When we speak of a *TBox* without further qualification, we mean a *general TBox*, i.e., a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$. Other forms of TBoxes will be introduced later as needed. 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. We use $\mathsf{Ind}(\mathcal{A})$ to denote the set of all individual names used in $\mathcal{A}$. It will sometimes be convenient to view an ABox $\mathcal{A}$ as an interpretation $\mathcal{I}_\mathcal{A}$, defined in the obvious way (see [15]).

Regarding query languages, we focus on *instance queries (IQ)*, which have the form $A(x)$ with A a concept name and x a variable. We write $\mathcal{T}, \mathcal{A} \models A(a)$ if $a^\mathcal{I} \in A^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ and call a 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 $\mathcal{T}$. To define FO-rewritability, we require first-order queries (FOQs), which are first-order formulas constructed from atoms $A(x)$, $r(x,y)$, and $x = y$. We use $\mathsf{ans}(\mathcal{I}, q)$ to denote the set of all answers to the FOQ q in the interpretation $\mathcal{I}$.

A *signature* is a set of concept and role names, which are uniformly called *symbols* in this context. A *Σ-ABox* is an ABox that uses only concept and role names from Σ. The *full* signature is the signature that contains all concept and role names.

Definition 1 (FO-rewritability). *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and Σ an ABox signature. An IQ q is* FO-rewritable relative to $\mathcal{T}$ *and Σ if there is a FOQ φ such that* $\mathsf{cert}_\mathcal{T}(\mathcal{A}, q) = \mathsf{ans}(\mathcal{I}_\mathcal{A}, \varphi)$ *for all Σ-ABoxes $\mathcal{A}$. Then φ is an* FO-rewriting of q relative to $\mathcal{T}$ and Σ.

Example 1. Recall from the introduction that $A(x)$ is not FO-rewritable relative to $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$ and the full signature. If we add $\exists r.\top \sqsubseteq A$ to $\mathcal{T}$, then $A(x)$ is FO-rewritable relative to the resulting TBox and the full signature, 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 $A(x)$. Conversely, if a query q is FO-rewritable relative to a TBox $\mathcal{T}'$ and a signature Σ, then q is FO-rewritable relative to $\mathcal{T}'$ and any $\Sigma' \subseteq \Sigma$ (take an FO-rewriting relative to $\mathcal{T}'$ and Σ and replace all atoms which involve predicates that are not in Σ' with false).

Sometimes, instance queries have the more general form $C(x)$ with C an $\mathcal{EL}$-concept. Since $C(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ whenever $A(x)$ is FO-rewritable relative to $\mathcal{T} \cup \{A \equiv C\}$ and Σ, A a fresh concept name, queries of this form are captured by the results in this paper.

3 General TBoxes – Upper Bounds

We first characterize failure of FO-rewritability of an IQ $A(x)$ relative to a TBox $\mathcal{T}$ and an ABox signature Σ in terms of the existence of certain Σ-ABoxes and then show how to decide the latter. The following result provides the starting point. An ABox is called *tree-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 $r = s$. It is called *forest-shaped* if the graph is a disjoint union of trees and $r(a,b), s(a,b) \in \mathcal{A}$ implies $r = s$. A FOQ is a *tree-UCQ* if it is a disjunction $q_1 \vee \cdots \vee q_n$ and each q_i is a conjunctive query (CQ) that is tree-shaped (defined in analogy with acyclic ABoxes) where the root is the only answer variable; see e.g. [15] for details on CQs.

Theorem 1. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, Σ an ABox signature, and $A(x)$ an IQ. Then*

1. *If $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ, then there is a tree-UCQ that is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ;*
2. *If $\varphi(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and forest-shaped Σ-ABoxes and $\varphi(x)$ is a tree-UCQ, then $\varphi(x)$ is an FO-rewriting relative to $\mathcal{T}$ and Σ;*

Of the two points in Theorem 1, Point 1 is most laborious to prove. It involves applying an Ehrenfeucht-Fraïssé game and explicitly constructing a tree-UCQ as a disjunction of certain $\mathcal{EL}$-concepts (c.f. the characterization of FO-rewritability in terms of datalog boundedness given in [15] and its proof). Point 2 can then be derived from Point 1.

For a forest-shaped ABox $\mathcal{A}$ and $k \geq 0$, we use $\mathcal{A}|_k$ to denote the restriction of $\mathcal{A}$ to depth k. The following provides the first version of the announced characterization of FO-rewritability in terms of the existence of certain ABoxes.

Theorem 2. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, Σ an ABox signature, and $A(x)$ an IQ. Then $A(x)$ is not FO-rewritable relative to $\mathcal{T}$ and Σ iff for every $k \geq 0$, there is a tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding k with root a_0 s.t. $\mathcal{T}, \mathcal{A} \models A(a_0)$ and $\mathcal{T}, \mathcal{A}|_k \not\models A(a_0)$.*

The proof of Theorem 2 builds on Point 1 of Theorem 1. Note that if $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ, then there is a $k \geq 0$ such that for all tree-shaped Σ-ABoxes $\mathcal{A}$ of depth exceeding k with root a_0, $\mathcal{T}, \mathcal{A} \models A(a_0)$ implies $\mathcal{T}, \mathcal{A}|_k \models A(a_0)$. In the proof of Theorem 2, we explicitly construct FO-rewritings which are tree-UCQs of outdegree at most $|\mathcal{T}|$ and depth at most k.

To proceed, it is convenient to work with TBoxes in *normal form*, where all CIs must be of one of the forms $A \sqsubseteq B_1$, $A \sqsubseteq \exists r.B$, $\top \sqsubseteq A$, $B_1 \sqcap B_2 \sqsubseteq A$, $\exists r.B \sqsubseteq A$ with A, B, B_1, B_2 concept names. This can be assumed without loss of generality when the ABox signature is finite.

Lemma 1. *For any $\mathcal{EL}$-TBox $\mathcal{T}$, finite ABox signature Σ, and IQ $A(x)$, there is a TBox $\mathcal{T}'$ in normal form such that for any FOQ φ, we have that $\varphi(x)$ 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.*

To exploit Theorem 2 for building a decision procedure for FO-rewritability, we impose a bound on k. The next theorem is proved using Theorem 2 and a pumping argument.

Theorem 3. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox in normal form, Σ an ABox signature, $A(x)$ an IQ, and $n = |(\mathsf{sig}(\mathcal{T}) \cup \Sigma) \cap \mathsf{N_C}|$. Then $A(x)$ is not FO-rewritable relative to $\mathcal{T}$ and Σ iff there exists a tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding 2^{2n} with root a_0 such that $\mathcal{T}, \mathcal{A} \models A(a_0)$ and $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models A(a_0)$.*

Note that, with the remark after Theorem 2, we obtain a triple exponential upper bound on the size of FO-rewritings for TBoxes in normal form. By Lemma 1, this bound transfers to unrestricted TBoxes for finite ABox signatures. The same is true for the full signature since any FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and the full signature is trivially also an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and the finite signature $\mathsf{sig}(\mathcal{T}) \cup \{A\}$, and vice versa. We conjecture that no shorter rewritings *into tree-UCQs* are possible, and that this can be shown using examples from [13, 16]. Here, we only remark that the bound in Theorem 3 cannot be significantly improved: for every $n \geq 1$, there is an $\mathcal{EL}$-TBox $\mathcal{T}$ and an IQ $A(x)$ such that $|\mathsf{sig}(\mathcal{T}) \cap \mathsf{N_C}| = n$, $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full Σ, and for all ABoxes of depth at least 2^n with root a_0, we have $\mathcal{T}, \mathcal{A} \models A(a_0)$ iff $\mathcal{T}, \mathcal{A}|_{2^n-1} \models A(a_0)$. Such a $\mathcal{T}$ can be constructed by simulating a binary counter, see Section 4 of [13]. Based on Theorem 3, we can establish the following result.

Theorem 4. *Deciding FO-rewritability of an IQ relative to an $\mathcal{EL}$-TBox and an ABox signature is in* ExpTime.

The proof utilizes non-deterministic bottom-up automata on finite, ranked trees: we construct exponential-size automata that accept precisely the ABoxes $\mathcal{A}$ from Theorem 3 and then decide their emptiness in PTime.

Somewhat unexpectedly, when Σ is full, the characterization given in Theorem 3 can be further improved from tree-shaped to linear ABoxes. To prepare for proving this, we first show that, also when considering the full signature, we can work with TBoxes is normal form. Note that the statement made by the following lemma is weaker than that made by Lemma 1 since it does not claim that the same FO-rewritings can be used (which indeed is not the case).

Lemma 2. *For any $\mathcal{EL}$-TBox $\mathcal{T}$ and IQ $A(x)$, there is a TBox $\mathcal{T}'$ in normal form such that $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full signature if and only if $A(x)$ is FO-rewritable relative to $\mathcal{T}'$ and the full signature.*

An ABox $\mathcal{A}$ is *linear* if it consists of role assertions $r_0(a_0, a_1), \ldots, r_{n-1}(a_{n-1}, a_n)$ and concept assertions $A(a)$ with $a \in \{a_0, \ldots, a_n\}$. We now improve Theorem 3 to linear ABoxes as announced.

Theorem 5. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox in normal form, $A(x)$ an IQ, and $n = |\mathsf{sig}(\mathcal{T}) \cap \mathsf{N_C}|$. Then $A(x)$ is not FO-rewritable relative to $\mathcal{T}$ (and the full Σ) iff there exists a linear ABox $\mathcal{A}$ of depth exceeding 2^{2n} with root a_0 such that $\mathcal{T}, \mathcal{A} \models A(a_0)$ and $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models A(a_0)$.*

The proof of Theorem 5 is based on the careful extraction of a linear ABox from the tree-shaped one whose existence is guaranteed by Theorem 3.

The following example shows that, when Σ is not full, tree-shaped ABoxes in Theorem 3 cannot be replaced with linear ones.

Example 2. Let
$$\mathcal{T} = \{A_i \sqsubseteq X_i,\ B_i \sqcap X_i \sqsubseteq Y_i,\ \exists r.Y_i \sqsubseteq X_i \mid i \in \{1,2\}\} \cup \{X_1 \sqcap X_2 \sqsubseteq X,\ B_1 \sqcap B_2 \sqsubseteq Z,\ \exists r.Z \sqsubseteq X\},$$
$\Sigma = \{A_1, A_2, B_1, B_2, r\}$, and take the IQ $X(x)$. The tree-shaped ABox

$$\mathcal{A} = \{r(a_0, a_{i,0}), r(a_{i,0}, a_{i,1}), \ldots, r(a_{i,2^n-1}, a_{i,2^{2n}+1}) \mid i \in \{1,2\}\} \cup \{B_i(a_{i,0}), \ldots, B_i(a_{i,2^n}), A_i(a_{i,2^{2n}+1}) \mid i \in \{1,2\}\},$$

with n as in Theorems 3 and 5, is of depth exceeding 2^{2n} and it can be verified that $\mathcal{T}, \mathcal{A} \models X(a_0)$, but $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models X(a_0)$. However, for all linear Σ-ABoxes $\mathcal{A}$, we have $\mathcal{T}, \mathcal{A} \models X(a_0)$ iff $\mathcal{T}, \mathcal{A}|_1 \models X(a_0)$.

Theorem 5 allows us to replace the non-deterministic tree automata in the proof of Theorem 4 with word automata, improving the upper bound to PSPACE.

Theorem 6. *Deciding FO-rewritability of an IQ relative to an $\mathcal{EL}$-TBox and the full ABox signature is in* PSPACE.

4 General TBoxes – Lower Bounds

We establish lower bounds that match the upper bounds from the previous section.

Theorem 7. *Deciding FO-rewritability of an IQ relative to a general $\mathcal{EL}$-TBox and an ABox signature Σ is (1)* PSPACE*-hard when Σ is full and (2)* EXPTIME*-hard when Σ is an input.*

The proof of Point 1 is by reduction of the word problem of polynomially space-bounded deterministic Turing machines (DTMs). For Point 2, we use polynomially space-bounded alternating Turing machines (ATMs). We start with the former.

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 $p(k)$ tape cells when k 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 5, non-FO-rewritability of B w.r.t. $\mathcal{T}$ is witnessed by a sequence of linear ABoxes of increasing depth. In the reduction, these 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 $aq'b$.[4] This is represented by ABoxes of the form

$$\{r(b_1, b_0), s(b_2, b_1), r(b_3, b_2), t(b_4, b_3), r(b_5, b_4), \dots, r(b_k, b_{k-1})\}$$

where additionally, the concept (a, q) is asserted for $b_0, b_4, b_8, \dots,$ b is asserted for $b_1, b_5, b_9, \dots,$ a for $b_2, b_6, b_{10}, \dots,$ and (b, q') for $b_3, b_7, b_{11}, \dots$. 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

[4] $uqv \in \Gamma^* Q \Gamma^*$ 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.

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:

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

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

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

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, that there is not more than one head position per configuration, and at least one:

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

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

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

$$a \sqsubseteq \overline{H} \text{ for all } a \in \Gamma \tag{7}$$

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

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.

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:

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

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

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.

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

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

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

Lemma 3. *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 2, it is enough to show that for every $k \geq 0$, there is a tree-shaped ABox $\mathcal{A}$ with root a_0 and of depth exceeding k such that $\mathcal{T}, \mathcal{A} \models B(a_0)$ and $\mathcal{T}, \mathcal{A}|_k \not\models B(a_0)$. Fix a k and 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 $\mathcal{A}$ (tree-shaped with outdegree one) of length $> k$. 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 $k = (2 \cdot \mathsf{step}_M + 2) \cdot p(n) + p(n)$.

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

Proof of claim. By Theorem 2, we find a tree-shaped ABox $\mathcal{A}'$ of depth exceeding $k + p(n) + 1$ with root a_0 such that $\mathcal{T}, \mathcal{A}' \models B(a_0)$ and $\mathcal{T}, \mathcal{A}'|_{k+p(n)+1} \not\models B(a_0)$. The desired ABox $\mathcal{A}$ is obtained by closing $\mathcal{A}'$ under CIs (4) to (10). Clearly, we have $\mathcal{T}, \mathcal{A} \models B(a_0)$. Now consider that 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 $\mathcal{T}, \mathcal{A}'|_{k+p(n)+1} \models \alpha$. Since $\mathcal{T}, \mathcal{A}'|_{k+p(n)+1} \not\models B(a_0)$, we thus have $\mathcal{T}, \mathcal{A}|_k \not\models B(a_0)$, 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 $\mathcal{T}^-, \mathcal{A} \models B(a_0)$ and $\mathcal{T}^-, \mathcal{A}|_k \not\models B(a_0)$. We can assume w.l.o.g. that there is an individual a on level $k+1$ of $\mathcal{A}$ such that $\mathcal{T}^-, \mathcal{A}^- \not\models B(a_0)$, 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) $\mathcal{T}^-, \mathcal{A} \models B(b_i)$;

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

Proof of claim.

(a) Follows from the fact that $\mathcal{T}^-, \mathcal{A} \models B(a_0)$, $\mathcal{T}^-, \mathcal{A}^- \not\models B(a_0)$, 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 $\mathcal{T}^-, \mathcal{A} \models X(b_i)$.

Since $\mathcal{T}^-, \mathcal{A} \models B(b_0)$ and all CIs in $\mathcal{T}^-$ are of role depth one, we have $\mathcal{T}^-, \mathcal{A}^+ \models B(b_0)$. 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, $\mathcal{T}^-, \mathcal{A}^- \cup \{B(b_i)\} \models B(a_0)$ 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 $\mathcal{T}^-, \mathcal{A}^- \not\models B(a_0)$ 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. ❏

We now come to Point 2 of Theorem 7. Let $M = (Q_\exists, Q_\forall, \Omega, \Gamma, \delta, q_0, q_{\mathsf{acc}}, q_{\mathsf{rej}})$ be an 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. 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 be 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

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

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 $P_1, \dots, P_{p(n)}$, which are not in Σ, indicate that the contents of a given tape cell have been specified, and $\mathsf{Tape} \notin \Sigma$ does the same for the whole tape:

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

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

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 $H_i, \overline{H}_i \notin \Sigma$:

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

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

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

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

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

$$H_i \sqcap A_{a,j} \sqcap \exists r_\ell.B_{b,j} \sqsubseteq E \quad \text{for all } A_{a,j} \in \mathcal{C}_j^a \text{ and } B_{b_j} \in \mathcal{C}_j^b,\; a \neq b, \tag{22}$$

$$\text{distinct } i, j \in [1, \dots, p(n)], \text{ and } \ell \in \{1, 2\}$$

where C_i^a denotes the restriction of C_i to those tuples with first component a. Errors in a computation tree imply B at the root of that tree:

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

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

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

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

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 $\ell \in \{1, 2\}$:

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

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

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:

$$(a, q_{\mathsf{acc}}, i) \sqcap \exists t.(A^0_1 \sqcap \ldots \sqcap A^0_{p(n)} \sqcap B) \sqsubseteq B \tag{29}$$

where (abusing notation) we use $A^0_1, \ldots, A^0_{p(n)}$ to denote the sequence of symbols from $\mathcal{C}$ that corresponds to the initial configuration. We also have to prevent continued travel along r 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 4. *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 2, it is enough to show that for every $k \geq 0$, there is a tree-shaped ABox $\mathcal{A}$ with root a_0 and of depth exceeding k such that $\mathcal{T}, \mathcal{A} \models B(a_0)$ and $\mathcal{T}, \mathcal{A}|_k \not\models B(a_0)$. Note that the computation tree T can be converted into a tree-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 computations at their corresponding individual names, and use the role names r and s 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 tree-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 2, we find a tree-shaped ABox $\mathcal{A}$ of depth exceeding k with root a_0 such that $\mathcal{T}, \mathcal{A} \models B(a_0)$ and $\mathcal{T}, \mathcal{A}|_k \not\models B(a_0)$. We can assume w.l.o.g. that there is an individual a on level $k+1$ of $\mathcal{A}$ such that $\mathcal{T}, \mathcal{A}^- \not\models B(a_0)$, where $\mathcal{A}^-$ is $\mathcal{A}$ with the subtree rooted at a dropped. Let $a_0, \dots, a_{k+1}$ 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 $a_0, \dots, a_p$ be the longest prefix of $a_0, \dots, a_{k+1}$ 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. $\mathcal{T}, \mathcal{A}^- \not\models E(a_i)$;
2. $\mathcal{T}, \mathcal{A} \models \mathsf{Tape}(a_i)$ 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 $\mathcal{T}, \mathcal{A}^- \not\models B(a_0)$ 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 $\mathcal{T}, \mathcal{A} \not\models C(a_0)$.

Otherwise, we obtain $\mathcal{T}, \mathcal{A}^- \models C(a_0)$ by the claim, implying $\mathcal{T}, \mathcal{A}^- \models E(a_0)$, which is a contradiction to Point 1. Since $\mathcal{T}, \mathcal{A} \models B(a_0)$ and $\mathcal{T}, \mathcal{A}^- \not\models B(a_0)$ 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 $\mathcal{T}, \mathcal{A} \models \mathsf{Tape}(a_0)$. 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 $\mathcal{T}, \mathcal{A}^- \models E(a_i)$. By Point 2 of IH and the claim, we have $\mathcal{T}, \mathcal{A}^- \models \mathsf{Tape}(a_{i-1})$. By CI (23), $\mathcal{T}, \mathcal{A}^- \models E(a_{i-1})$, 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, $\mathcal{T}, \mathcal{A}^- \not\models E(a_{i-1})$. This finishes the proof of Points 1-3.

By Point 3, the length of the configuration sequence $C_0, \dots, C_p$ 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 $a_0, \dots, a_p$, a_{p+1} must be a t-individual, but cannot be an r_ℓ-individual for any $\ell \in \{1, 2\}$. Since $\mathcal{T}, \mathcal{A} \models B(a_0)$, $\mathcal{T}, \mathcal{A}^- \not\models B(a_0)$, 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) $\mathcal{T}, \mathcal{A} \models B(a)$ and $\mathcal{T}, \mathcal{A}^- \not\models E(a)$;
(b) $\mathcal{T}, \mathcal{A} \models \mathsf{Tape}(a)$ 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 $\mathcal{T}, \mathcal{A} \models B(a_q)$. 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 $\mathcal{T}, \mathcal{A} \models B(a_i)$ 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 $\mathcal{T}, \mathcal{A} \models B(a)$. 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 $\mathcal{T}, \mathcal{A} \models \mathsf{Tape} \sqcap B(a_i)$, for $i \in \{1, 2\}$. We also have $\mathcal{T}, \mathcal{A}^- \not\models E(a_i)$ for $i \in \{1, 2\}$ since the contrary would imply $\mathcal{T}, \mathcal{A}^- \models B(a_0)$, 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 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 $\mathcal{T}, \mathcal{A} \models B(a)$ and $\mathcal{T}, \mathcal{A}^- \not\models E(a)$. By CI (30), a does thus not have any r_1- or r_2-successors in $\mathcal{A}$. Consequently, $\mathcal{T}, \mathcal{A} \models B(a)$ 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. ❏

5 Classical TBoxes

A *classical TBox* $\mathcal{T}$ is a finite set of *concept definitions* $A \equiv C$ and CIs $A \sqsubseteq C$ where A is a concept name. No concept name is allowed to occur more than once on the left hand side of a statement in $\mathcal{T}$.

Theorem 8. *Deciding FO-rewritability of an IQ relative to a classical $\mathcal{EL}$-TBox and an ABox signature is in* PTIME.

We give examples illustrating FO-rewritability in classical TBoxes and give the main idea of the proof.

Example 3. (a) The IQ $A(x)$ is not FO-rewritable relative to the classical TBox $\{A \equiv \exists r.A\}$ and the full ABox signature.

(b) The concept name A has a cyclic definition in the TBox $\mathcal{T} = \{\ A \equiv B \sqcap \exists r.A, B \sqsubseteq \exists r.A\ \}$ which often indicates non-FO-rewritability, but in this case the IQ $A(x)$ has an FO-rewriting relative to $\mathcal{T}$ and the full ABox signature, namely $\varphi(x) = A(x) \vee B(x)$.

To present the idea of the proof, we use an appropriate normal form for classical TBoxes. A concept name A is *defined in* $\mathcal{T}$ if there is a definition $A \equiv C \in \mathcal{T}$ and *primitive* otherwise; A is *non-conjunctive* in $\mathcal{T}$ if it occurs in $\mathcal{T}$, but there is no CI of the form $A \equiv B_1 \sqcap \cdots \sqcap B_n$ in $\mathcal{T}$ with $n \geq 1$ and $B_1, \ldots, B_n$ concept names. We use non-conj$(\mathcal{T})$ to denote the set of non-conjunctive concepts in $\mathcal{T}$. A classical TBox $\mathcal{T}$ is in *normal form* if it is a set of statements $A \equiv \exists r.B$ and $A \equiv B_1 \sqcap \cdots \sqcap B_n$ where $B, B_1, \ldots, B_n$ are concept names and $B_1, \ldots, B_n$ are non-conjunctive. For every classical TBox $\mathcal{T}$, one can construct in polynomial time a classical TBox $\mathcal{T}'$ in normal form that uses additional concept names such that $\mathcal{T}' \models \mathcal{T}$ and every model of $\mathcal{T}$ can be expanded to a model of $\mathcal{T}'$ [10]. It is not hard to verify that an IQ $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ if and only if it is FO-rewritable relative to $\mathcal{T}'$ and Σ, provided that A is not among the new concept names introduced during the construction of $\mathcal{T}'$. For a classical TBox $\mathcal{T}$ in normal form and a concept name A, define

$$\text{non-conj}_{\mathcal{T}}(A) = \begin{cases} \{A\} & \text{if } A \text{ is non-conjunctive in } \mathcal{T} \\ \{B_1, \ldots, B_n\} & \text{if } A \equiv B_1 \sqcap \cdots \sqcap B_n \in \mathcal{T}. \end{cases}$$

Our polytime algorithm utilizes an ABox introduced in [10, 9] in the context of conservative extensions and logical difference: given a classical TBox $\mathcal{T}$ in normal form and an ABox signature Σ, we compute in polytime a polysize Σ-ABox $\mathcal{A}_{\mathcal{T},\Sigma}$ with individual names a_B, B non-conjunctive in $\mathcal{T}$, such that for any Σ-ABox $\mathcal{A}$, individual name a in $\mathcal{A}$, and concept name A the following conditions are equivalent:

- $\mathcal{T}, \mathcal{A} \not\models A(a)$;
- there exists $B \in$ non-conj$_{\mathcal{T}}(A)$ such that $(\mathcal{A}, a)$ is simulated by $(\mathcal{A}_{\mathcal{T},\Sigma}, a_B)$ (see appendix).

It follows that to check whether $A(x)$ is FO-rewritable, instead of considering arbitrary tree-shaped Σ-ABoxes $\mathcal{A}$ and $\mathcal{A}|_k$ as in Theorem 2, it suffices to consider the tree unfolding of $\mathcal{A}_{\mathcal{T},\Sigma}$ at a_B and its restriction to depth k, for all $B \in$ non-conj$_{\mathcal{T}}(A)$. The original search problem has been reduced to the problem of analysing the tree unfolding of $\mathcal{A}_{\mathcal{T},\Sigma}$. A polytime algorithm performing that analysis is given in the appendix.

6 Semi-Acyclic TBoxes

It is easy to see that every IQ is FO-rewritable relative to every acyclic $\mathcal{EL}$-TBox and every ABox signature Σ. We observe that the same holds for semi-acyclic TBoxes, where some cycles are still allowed, and that it is possible to find rewritings of polynomial size when only databases of domain size at least two are admitted.

A *semi-acyclic TBox* is defined like a classical TBox, except that definitorial cycles are disallowed, i.e., there cannot be concept definitions $A_0 \equiv C_0, \ldots, A_{n-1} \equiv C_{n-1}$ such that A_i occurs in $C_{i+1 \,\mathsf{mod}\, n}$. Note that cycles via concept *inclusions*, such as $A \sqsubseteq \exists r.A$, are still permitted. Let $\mathcal{T}$ be a semi-acyclic TBox and Σ an ABox signature. For an $\mathcal{EL}$-concept C, we use $\mathsf{pre}_{\mathcal{T},\Sigma}(C)$ to denote the FO-formula $\bigvee_{B \in \Sigma \mid \mathcal{T} \models B \sqsubseteq C} B(x)$. For all concept names A and $\mathcal{EL}$-concepts C and D and role names r, set

$$\begin{aligned}
\varphi^{\Sigma}_{\top,\mathcal{T}}(x) &= \mathsf{true} \\
\varphi^{\Sigma}_{A,\mathcal{T}}(x) &= \mathsf{pre}_{\mathcal{T},\Sigma}(A) && \text{if } A \text{ is primitive} \\
\varphi^{\Sigma}_{A,\mathcal{T}}(x) &= \varphi^{\Sigma}_{C,\mathcal{T}}(x) && \text{if } A \equiv C \in \mathcal{T} \\
\varphi^{\Sigma}_{C \sqcap D,\mathcal{T}}(x) &= \varphi^{\Sigma}_{C,\mathcal{T}}(x) \wedge \varphi^{\Sigma}_{D,\mathcal{T}}(x) \\
\varphi^{\Sigma}_{\exists r.C,\mathcal{T}}(x) &= \mathsf{pre}_{\mathcal{T},\Sigma}(\exists r.C) \vee \exists y.(r(x,y) \wedge \varphi^{\Sigma}_{C,\mathcal{T}}[y/x]) && \text{if } r \in \Sigma \\
\varphi^{\Sigma}_{\exists r.C,\mathcal{T}}(x) &= \mathsf{pre}_{\mathcal{T},\Sigma}(\exists r.C) && \text{if } r \notin \Sigma
\end{aligned}$$

where $\varphi[x/y]$ denotes the result of first renaming all bound variables in φ so that y does not occur, and then replacing the free variable x of φ with y.

Lemma 5. *For all IQs $A(x)$, $\varphi^{\Sigma}_{A,\mathcal{T}}(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ.*

The size of $\varphi^{\Sigma}_{A,\mathcal{T}}(x)$ can clearly be exponential in the size of $\mathcal{T}$, for example when $A = A_n$ and $\mathcal{T} = \{A_i \equiv \exists r.A_{i-1} \sqcap \exists s.A_{i-1} \mid 1 \leq i \leq n\}$. To reduce $\varphi^{\Sigma}_{A,\mathcal{T}}$ to polynomial size, we can use Avigad's observation that FO supports structure sharing [1]. More precisely, let φ be a positive FOQ (such as $\varphi^{\Sigma}_{A,\mathcal{T}}$) whose subformulas include $\psi(x_1), \ldots, \psi(x_n)$. The multiple occurences of ψ can be avoided by rewriting φ to $\exists u \forall y \forall z \big(\; (\psi(y) \leftrightarrow z = u) \rightarrow \varphi' \big)$ where φ' is φ with each $\psi(x_i)$ replaced with $y = x_i \rightarrow z = u$. Intuitively, we iterate over all y and memorize whether $\psi(y)$ holds using identity of z with u. Since we need at least two different 'values' for z to make this trick work, the resulting FOQ is an FO-rewriting only on ABoxes with at least two individual names.

7 Related Work

In [15], deciding FO-rewritability is studied in the context of the expressive DL $\mathcal{ALCFI}$ and several of its fragments. In general, though, the setup in that paper is different: while we are interested in deciding FO-rewritability of a single query relative to a TBox, the results in [15] concern deciding whether, for a given TBox $\mathcal{T}$, *all* queries are FO-rewritable relative to $\mathcal{T}$. It is shown that this problem is decidable for Horn-$\mathcal{ALCFI}$-TBoxes of depth at most two and for Horn-$\mathcal{ALCF}$-TBoxes (queries are IQs or, equivalently, CQs). As a by-product of these results, a close connection between FO-rewritability of TBoxes formulated in Horn DLs and boundedness of datalog programs is observed, see e.g. [6, 18] for the latter problem. In its original formulation, the following result is established for a larger class of TBoxes, namely materializable $\mathcal{ALCFI}$-TBoxes of depth one.

Lemma 6 **([15]).** *For every (general) $\mathcal{EL}$-TBox $\mathcal{T}$ in normal form, there is a datalog program $\Pi_\mathcal{T}$ such that for every ABox signature Σ and IQ $A(x)$, the predicate A is bounded in $\Pi_\mathcal{T}$ relative to Σ-databases iff $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ.*

In [15], the program $\Pi_\mathcal{T}$ is of exponential size. Since we are only interested in $\mathcal{EL}$-TBoxes, it is easy to find a $\Pi_\mathcal{T}$ of polynomial size. More specifically, $\Pi_\mathcal{T}$ consists of

$A(x) \leftarrow \mathsf{true}$	if $\top \sqsubseteq A \in \mathcal{T}$		
$B(x) \leftarrow r(x,y), A(x)$	if $\exists r.A \sqsubseteq B \in \mathcal{T}$	$X_{\exists r.A}(x) \leftarrow B(x)$	if $B \sqsubseteq \exists r.A \in \mathcal{T}$
$B(x) \leftarrow A_1(x), A_2(x)$	if $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$		(where possibly $A_1 = A_2$)
$B(x) \leftarrow X_{\exists r.A}(x)$	if $\exists r.B_0 \sqsubseteq B \in \mathcal{T}$ and $\mathcal{T} \models A \sqsubseteq B_0$		

This allows to carry over the 2EXPTIME upper bound for predicate boundedness of connected monadic datalog programs [6] to FO-rewritability of an IQ relative to a general $\mathcal{EL}$-TBoxes and an ABox signature.[5]

Note that boundedness has been studied also in the context of the μ-calculus and monadic second order (MSO) logic [17, 4]. Here, an EXPTIME upper bound is known from [17] and it seems likely that this result can be utilized to find an alternative proof of Theorem 4. In particular, it is possible to find a *μ-calculus rewriting* φ of an IQ $A(x)$ relative to an $\mathcal{EL}$-TBox $\mathcal{T}$ and ABox signature Σ: proceeding similarly to the construction of the above datalog program $\Pi_\mathcal{T}$, we can find a μ-calculus formula $\varphi_{\mathcal{T},\Sigma}$ such that for all Σ-ABoxes $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{T}, \mathcal{A} \models A(a)$ iff $\mathcal{I}_\mathcal{A}, a \models \varphi$. When simultaneous fixpoints are admitted, φ even has polynomial size.

8 Conclusions

It would be interesting to generalize the results presented in this paper to more expressive DLs and to more expressive query languages. Regarding the former, we note that using the techniques in [14, 15] it is possible to derive a NEXPTIME upper bound for deciding FO-rewritability of IQs relative to Horn-$\mathcal{ALCI}$-TBoxes and ABox signatures. Regarding the latter, CQs are a natural choice and we believe that a mix of techniques from this paper and those in [3] might provide a good starting point. It is interesting to note that FO-rewritability of all IQ-atoms $A(x)$ in a CQ q does not imply that q is FO-rewritable and the converse fails, too.

Acknowledgements. C. Lutz was supported by the DFG SFB/TR 8 'Spatial Cognition'.

References

1. J. Avigad. Eliminating definitions and skolem functions in first-order logic. In Proc. of LICS, pages 139–146. IEEE Computer Society, 2001.
2. F. Baader, M. Bienvenu, C. Lutz, and F. Wolter. Query and predicate emptiness in description logics. In Proc. of KR. AAAI Press, 2010.

[5] Note that we explicitly fix the signature Σ of the databases over which boundedness of $\Pi_\mathcal{T}$ is considered instead of assuming that only the EDB predicates can be used in the data as in [6]; this is only for simplicity and, in fact, it is easy to adapt $\Pi_\mathcal{T}$ to the latter assumption.

3. M. Bienvenu, C. Lutz, and F. Wolter. Query containment in description logics reconsidered. In Proc. of KR, 2012. To appear.
4. A. Blumensath, M. Otto, and M. Weyer. Decidability results for the boundedness problem. Manuscript, 2012.
5. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. J. of Automated Reasoning, 39(3):385–429, 2007.
6. S. S. Cosmadakis, H. Gaifman, P. C. Kanellakis, and M. Y. Vardi. Decidable optimization problems for database logic programs. In Proc. of STOC, pages 477–490. ACM, 1988.
7. G. Gottlob and T. Schwentick. Rewriting ontological queries into small nonrecursive datalog programs. In Proc. of DL. CEUR-WS, 2011.
8. S. Kikot, R. Kontchakov, V. V. Podolskii, and M. Zakharyaschev. Exponential lower bounds and separation for query rewriting. CoRR, abs/1202.4193, 2012.
9. B. Konev, M. Ludwig, D. Walther, and F. Wolter. The logical diff for the lightweight description logic $\mathcal{EL}$. Technical report, U. of Liverpool, http://www.liv.ac.uk/~frank/publ/, 2011.
10. B. Konev, D. Walther, and F. Wolter. The logical difference problem for description logic terminologies. In Proc. of IJCAR, pages 259–274. Springer, 2008.
11. R. Kontchakov, C. Lutz, D. Toman, F. Wolter, and M. Zakharyaschev. The combined approach to query answering in DL-Lite. In Proc. of KR. AAAI Press, 2010.
12. C. Lutz, D. Toman, and F. Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In Proc. of IJCAI, pages 2070–2075. AAAI Press, 2009.
13. C. Lutz and F. Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. In J. of Symbolic Computation 45(2): 194-228, 2010.
14. C. Lutz and F. Wolter. Non-uniform data complexity of query answering in description logics. In Proc. of DL. CEUR-WS, 2011.
15. C. Lutz and F. Wolter. Non-uniform data complexity of query answering in description logics. In Proc. of KR, 2012. To appear.
16. N. Nikitina and S. Rudolph. ExpExpExplosion: uniform interpolation in general $\mathcal{EL}$ terminologies. In *Proc. of ECAI*, 2012.
17. M. Otto. Eliminating recursion in the μ-calculus. In Proc. of STACS, pages 531–540. Springer, 1999.
18. R. van der Meyden. Predicate boundedness of linear monadic datalog is in PSPACE. Int. J. Found. Comput. Sci., 11(4):591–612, 2000.

A Proofs for Section 3

Canonical Models for General $\mathcal{EL}$-TBoxes

The notion of a *canonical model* used in various places throughout Appendix A, so we recall the definition here. Let $\mathcal{T}$ be a general $\mathcal{EL}$-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:

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

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

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

It is standard to prove the following.

Lemma 7. *$\mathcal{I}_{\mathcal{T},\mathcal{A}}$ 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 $a^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} \in C^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$ iff $\mathcal{T}, \mathcal{A} \models C(a)$.*

Characterizations of (Non)-FO-Rewritability in Terms of Tree-Shaped ABoxes

Theorem 1. Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, Σ an ABox signature, and $A(x)$ an IQ. Then

1. If $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ, then there is a tree-UCQ that is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ;
2. If $\varphi(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and forest-shaped Σ-ABoxes and $\varphi(x)$ is a tree-UCQ, then $\varphi(x)$ is an FO-rewriting relative to $\mathcal{T}$ and Σ;

Proof. *Point 1.* Let $A(x)$ be FO-rewritable relative to $\mathcal{T}$ and Σ, and let $\varphi(x)$ be an FO-rewriting of quantifier rank n. Put $k = 2^n - 1$.

Note that a tree-shaped Σ-ABox $\mathcal{A}$ can be viewed as an $\mathcal{EL}$-concept $C_{\mathcal{A}}$ in an obvious way: if $\mathcal{A}$ has root a and $A_1(a), \ldots, A_m(a), r_1(a, b_1), \ldots r_\ell(a, b_\ell)$ are all assertions in $\mathcal{A}$ that involve a, then define $C_{\mathcal{A}} = A_1 \sqcap \cdots \sqcap A_m \sqcap \exists r_1.C_{\mathcal{A}|_{b_1}} \sqcap \cdots \sqcap \exists r_\ell.C_{\mathcal{A}|_{b_\ell}}$ where $\mathcal{A}|_{b_i}$ denotes the restriction of $\mathcal{A}$ to the sub-ABox rooted at b_i. Let Γ be the set of all ABoxes $\mathcal{A}|_k$ where $\mathcal{A}$ is a minimal tree-shaped Σ-ABox with root a such that $\mathcal{T}, \mathcal{A} \models A(a)$. It is not hard to show that the outdegree of any $\mathcal{A} \in \Gamma$ is bounded by $|\mathcal{T}|$, thus Γ is finite. We claim that the tree-UCQ $\psi(x) = \bigvee_{\mathcal{A} \in \Gamma} C_{\mathcal{A}}$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ (where the concepts $C_{\mathcal{A}}$ are seen as FO-formulas with one free variable in the standard way). A central observation is the following.

Claim. For all tree-shaped Σ-ABoxes $\mathcal{A}$ with root a_0, we have that $\mathcal{T}, \mathcal{A} \models A(a_0)$ implies $\mathcal{T}, \mathcal{A}|_k \models A(a_0)$.

Proof of Claim. For any $a \in \mathsf{Ind}(\mathcal{A})$, let $\mathcal{A}|_k^a$ denote the sub-ABox of $\mathcal{A}$ obtained by keeping only individual names b that are within distance at most k from a, i.e., there is a sequence $a_0, \dots, a_\ell \in \mathsf{Ind}(\mathcal{A})$ such that $\ell \leq k$ and for all $i < \ell$ there is a role name r such that $r(a_i, a_{i+1}) \in \mathcal{A}$ or $r(a_{i+1}, a_i) \in \mathcal{A}$. Let $\widehat{\mathcal{A}}$ denote the ABox that consists of n disjoint copies of $\mathcal{A}|_k^a$, for all $a \in \mathsf{Ind}(\mathcal{A})$; we assume w.l.o.g. that the individual names in $\widehat{\mathcal{A}}$ are disjoint from those in $\mathcal{A}$. Then, $\mathcal{A}_1$ is the disjoint union of $\mathcal{A}$ and $\widehat{\mathcal{A}}$ while $\mathcal{A}_2$ is the disjoint union of $\mathcal{A}|_k$ and $\widehat{\mathcal{A}}$.

One can show, using Ehrenfeucht-Fraïssé games with n rounds, that $\mathcal{I}_{\mathcal{A}_1} \models \varphi[a_0]$ iff $\mathcal{I}_{\mathcal{A}_2} \models \varphi[a_0]$. Specifically, a winning strategy for duplicator is to keep in round m the distance 2^{n-m} from elements that have already received a pebble, whenever this is possible; if spoiler selects an element b in round m, then:

- first assume that b is within distance 2^{n-m} of an element c that has received a pebble within a previous round and let c' be the other element played in that round; then duplicator selects an element b' according to the 'local 2^{n-m}-isomorphism' around c and c', picking an element that relates to c' in the same way that b relates to c;
- now assume that no element within distance 2^{n-m} of b has yet received a pebble and that b corresponds to the element b_0 in $\mathcal{A}$ (which includes the case that b *is* in the connected component $\mathcal{A}$ of $\mathcal{A}_1$ or in the connected component $\mathcal{A}|_k$ of $\mathcal{A}_2$); then duplicator replies by selecting the element that corresponds to b_0 in a copy of $\mathcal{A}|_k^{b_0}$ which has not yet received any pebble.

Since $\mathcal{T}, \mathcal{A} \models A(a_0)$ and the connected component of $\mathcal{A}_1$ that contains a_0 is $\mathcal{A}$, we must have $\mathcal{T}, \mathcal{A}_1 \models A(a_0)$ (a formal proof uses canonical models). It follows that $\mathcal{I}_{\mathcal{A}_1} \models \varphi[a_0]$ since φ is an FO-rewriting of $A(x)$ w.r.t. $\mathcal{T}$. As we have just shown, this yields $\mathcal{I}_{\mathcal{A}_2} \models \varphi[a_0]$, thus $\mathcal{T}, \mathcal{A}_2 \models A(a_0)$, which (again via canonical models) implies $\mathcal{T}, \mathcal{A}|_k \models A(a_0)$ as required. *End of proof of claim.*

To show that $\psi(x)$ is as required, let $\mathcal{A}$ be a Σ-ABox and $a \in \mathsf{Ind}(\mathcal{A})$ and first assume that $\mathcal{T}, \mathcal{A} \models A(a)$. Since every $\mathcal{EL}$-TBox is unraveling tolerant [15], we can unravel the sub-ABox of $\mathcal{A}$ rooted at a into a tree-shaped Σ-ABox $\mathcal{A}'$ with root a such that $\mathcal{T}, \mathcal{A}' \models A(a)$. Take a minimal tree-shaped Σ-ABox $\mathcal{A}'' \subseteq \mathcal{A}'$ such that $\mathcal{T}, \mathcal{A}'' \models A(a)$. Then $C_{\mathcal{A}''|_k}$ is a disjunct of $\psi(x)$ and $a \in (C_{\mathcal{A}''|_k})^{\mathcal{I}_{\mathcal{A}''}} \subseteq (C_{\mathcal{A}''|_k})^{\mathcal{I}_{\mathcal{A}'}}$. Since there is a simulation for $\mathcal{I}_{\mathcal{A}'}$ to $\mathcal{I}_{\mathcal{A}}$, we have $a \in (C_{\mathcal{A}''|_k})^{\mathcal{I}_{\mathcal{A}}}$ and thus $\mathcal{I}_{\mathcal{A}} \models \psi[a]$.

Conversely, assume that $\mathcal{I}_{\mathcal{A}} \models \psi[a]$, i.e., $a \in (C_{\mathcal{B}|k})^{\mathcal{I}_{\mathcal{A}}}$ for some $\mathcal{B}|_k \in \Gamma$. Since $\mathcal{T}, \mathcal{B} \models A(a)$ by choice of Γ, the claim yields $\mathcal{T}, \mathcal{B}|_k \models A(a)$. Since $a \in (C_{\mathcal{B}|k})^{\mathcal{I}_{\mathcal{A}}}$, there is a simulation from $\mathcal{B}|_k$ into $\mathcal{A}$, thus $\mathcal{T}, \mathcal{A} \models A(a)$ as required.

Point 2. Let $\varphi(x)$ be an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and forest-shaped Σ-ABoxes such that $\varphi(x)$ is a tree-UCQ. Let $\mathcal{A}$ be a Σ-ABox (not necessarily tree-shaped) and $a \in \mathsf{Ind}(\mathcal{A})$. Let $\mathcal{A}'$ be the unraveling of $\mathcal{A}$ into a forest-shaped Σ-ABox. Then $\mathcal{T}, \mathcal{A} \models A(a)$ iff $\mathcal{T}, \mathcal{A}' \models A(a)$ because of unraveling tolerance. Since $\varphi(x)$ is an FO-rewriting relative to $\mathcal{T}$ and forest-shaped Σ-ABoxes, we have $\mathcal{T}, \mathcal{A}' \models A(a)$ iff $\mathcal{I}_{\mathcal{A}'} \models \varphi[a]$. Since $\varphi(x)$ is a tree-UCQ and there is a simulation from $\mathcal{I}_{\mathcal{A}}$ to $\mathcal{I}_{\mathcal{A}'}$ and vice versa, $\mathcal{I}_{\mathcal{A}'} \models \varphi[a]$ iff $\mathcal{I}_{\mathcal{A}} \models \varphi[a]$. In summary, $\mathcal{T}, \mathcal{A} \models A(a)$ iff $\mathcal{I}_{\mathcal{A}} \models \varphi[a]$ as required.

❏

Theorem 2. Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, Σ an ABox signature, and $A(x)$ an IQ. Then $A(x)$ is not FO-rewritable relative to $\mathcal{T}$ and Σ iff for every $k \geq 0$, there exists a tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding k with root a_0 such that

1. $\mathcal{T}, \mathcal{A} \models A(a_0)$;
2. $\mathcal{T}, \mathcal{A}|_k \not\models A(a_0)$.

Proof. "if" Assume that $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ. By Point 1 of Theorem 1, there is an FO-rewriting $\varphi(x)$ that is a tree-UCQ. Let k be the maximum number of atoms in any disjunct of $\varphi(x)$. It is simple to show k is the required bound, i.e., there is no tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding k such that Points 1 and 2 from the theorem are satisfied. Let $\mathcal{A}$ be a tree-shaped Σ-ABox of depth exceeding k with root a_0 and $\mathcal{T}, \mathcal{A} \models A(a_0)$. We have $\mathcal{T}, \mathcal{A} \models A(a_0)$ iff $\mathcal{I}_\mathcal{A} \models \varphi[a_0]$ iff $\mathcal{I}_{\mathcal{A}|_k} \models \varphi[a_0]$ (since each disjunct in φ is a connected query that involves an answer variable and has at most k atoms) iff $\mathcal{T}, \mathcal{A}|_k \models A(a_0)$.

"only if". Assume that there is a $k \geq 0$ for which there is no tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding k such that Points 1 and 2 from the theorem are satisfied. Let Γ be the set of all tree-shaped Σ-ABoxes $\mathcal{A}$ of depth at most k such that $\mathcal{T}, \mathcal{A} \models A(a)$ where a denotes the root of $\mathcal{A}$ and $\mathcal{A}$ is minimal with Property (i). It is not hard to show that the outdegree of any $\mathcal{A} \in \Gamma$ is bounded by $|\mathcal{T}|$, thus Γ is finite. We aim to show that $\varphi(x) = \bigvee_{\mathcal{A} \in \Gamma} C_\mathcal{A}$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ (see proof of Theorem 1 for the definition of $C_\mathcal{A}$). We have to prove that for all forest-shaped Σ-ABoxes $\mathcal{A}$, $\mathsf{cert}_\mathcal{T}(\mathcal{A}, A(x)) = \mathsf{ans}(\mathcal{I}_\mathcal{A}, \varphi(x))$.

For the "$\supseteq$" direction, note that $\mathcal{I}_\mathcal{A} \models \varphi(a)$ implies $a \in (C_\mathcal{B})^{\mathcal{I}_\mathcal{A}}$ for some $\mathcal{B} \in \Gamma$ and thus $\mathcal{T}, \mathcal{A} \models C_\mathcal{B}(a)$; it is not hard to show (using canonical models) that, since $\mathcal{T}, \mathcal{B} \models A(b)$ with b root of $\mathcal{B}$, this yields $\mathcal{T}, \mathcal{A} \models A(a)$ as required. For "$\subseteq$", let $\mathcal{T}, \mathcal{A} \models A(a)$ and let $\mathcal{A}|_a$ be the sub-ABox of $\mathcal{A}$ rooted at a. If the depth of $\mathcal{A}|_a$ does not exceed k, then $C_{\mathcal{A}_a}$ is a disjunct of φ and thus $\mathcal{I}_\mathcal{A} \models \varphi(a)$ as required. Otherwise, $\mathcal{T}, \mathcal{A} \models A(a)$ and the assumption that there is no tree-shaped Σ-ABox that exceeds depth k and satisfies Points 1 and 2 from the theorem yields $\mathcal{T}, \mathcal{A}|_{a,k} \models A(a)$ where $\mathcal{A}|_{a,k}$ denotes the restriction of $\mathcal{A}_a$ to depth k. It follows that $C_{\mathcal{A}|_{a,k}}$ is a disjunct of φ. Since clearly $\mathcal{T}, \mathcal{A} \models C_{\mathcal{A}|_{a,k}}(a)$, we again have $\mathcal{I}_\mathcal{A} \models \varphi(a)$ as required. ❑

Lemma 1. For any $\mathcal{EL}$-TBox $\mathcal{T}$, ABox signature Σ that is finite or full, and IQ $A(x)$, there is a TBox $\mathcal{T}'$ in normal form such that for any FOQ φ, we have that $\varphi(x)$ 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.

Proof. Let $\mathcal{T}$, Σ, and $A(x)$ be given. Assume first that Σ is finite. We use $\mathsf{sub}(\mathcal{T})$ to denote the set of subconcepts C of (concepts that occurs 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 $\mathcal{T}$ and Σ and is distinct from 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

$$\mathcal{T}' = \bigcup_{C \sqsubseteq D \in \mathcal{T}} X_C \sqsubseteq X_D \quad \cup \quad \bigcup_{C \in \mathsf{sub}(\mathcal{T})} X_C \equiv \sigma(C).$$

After replacing CIs $A \sqsubseteq B_1 \sqcap B_2$ with $A \sqsubseteq B_1$ and $A \sqsubseteq B_2$, $\mathcal{T}'$ is of the required form. Using the fact that the concept names X_C do not occur in $\mathcal{T}$, Σ, A, it is not hard to prove that $\mathsf{cert}_{\mathcal{T}}(\mathcal{A}, A(x)) = \mathsf{cert}_{\mathcal{T}'}(\mathcal{A}, A(x))$ for all Σ-ABoxes $\mathcal{A}$. Consequently, every FO-rewriting $\varphi(x)$ of $A(x)$ relative to $\mathcal{T}$ and Σ is also a FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and Σ, and vice-versa. ❑

For every ABox $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$, we write $\mathsf{CN}^{\models}_{\mathcal{A}}(a)$ to denote the set of concept names A with $\mathcal{A}, \mathcal{T} \models A(a)$.

Theorem 3. Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox in normal form, Σ an ABox signature, $A(x)$ an IQ, and $n = |(\mathsf{sig}(\mathcal{T}) \cup \Sigma) \cap \mathsf{N_C}|$. Then $A(x)$ is not FO-rewritable relative to $\mathcal{T}$ and Σ iff there exists a tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding 2^{2n} with root a_0 such that

1. $\mathcal{T}, \mathcal{A} \models A(a_0)$;
2. $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models A(a_0)$.

Proof. Since the "only if" direction is an immediate consequence of Theorem 2, we concentrate on "if". Thus assume that there is a tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding 2^{2n} with root a_0 that satisfies Conditions 1 and 2 from the theorem. We may assume w.l.o.g. that $\mathcal{A}$ is minimal in the sense that, for every individual a , we have $\mathcal{T}, \mathcal{A}^- \not\models A(a_0)$, where $\mathcal{A}^-$ is obtained from $\mathcal{A}$ by dropping the subtree rooted at a. Let $k > 0$. We have to show that there is a tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding k with root a_0 such that Conditions 1 and 2 from Theorem 2 are satisfied. While constructing this ABox, it is convenient to assume a certain naming scheme for the individuals in $\mathsf{Ind}(\mathcal{A})$, namely that they are of the form a_w with w a word over the infinite alphabet $\mathbb{N}$, and that whenever $r(a_w, b_{w'}) \in \mathcal{A}$, then $w' = w \cdot n$ for some $n \in \mathbb{N}$. Clearly, the root node a_0 fits this scheme.

Let a_w be a node in $\mathcal{A}$ on level $2^{2n} + 1$ and let $\mathcal{A}^-$ debote the result of dropping in $\mathcal{A}$ the subtree rooted at a_w. On the path from a_0 to a, there must be at least two individuals a_u and $a_{u'}$ with the same value for $\mathsf{CN}^{\models}_{\mathcal{A}}(\cdot)$ and $\mathsf{CN}^{\models}_{\mathcal{A}^-}(\cdot)$. Assume w.l.o.g. that u is a proper prefix of u'. Let the new ABox $\mathcal{B}$ be obtained from $\mathcal{A}$ by taking the subtree rooted at a_u and plugging it in at $a_{u'}$, i.e., $\mathcal{B}$ is defined as follows where $\mathcal{X}$ is obtained from $\mathcal{A}$ by dropping the subtree rooted at $a_{u'}$ and $r_p(a_p, a_{u'})$ is the incoming edge to $a_{u'}$ in $\mathcal{A}$:

$$\begin{aligned} \mathcal{B} = \mathcal{X} \cup \{r_p(a_p, a_{u'})\} \cup \{r(a_{u'v_1}, a_{u',v_2}) \mid r(a_{uv_1}, a_{uv_2}) \in \mathcal{A}\} \cup \\ \{A(a_{u'v}) \mid A(a_{uv}) \in \mathcal{A}\}. \end{aligned}$$

Clearly, the depth of $\mathcal{B}$ exceeds $2^{2n} + 1$. Let x be such that $w = ux$ and set $w' = u'x$. Then $a_{w'}$ is the 'copy' of a_w in $\mathcal{B}$. Let $\mathcal{B}^-$ be the result of dropping in $\mathcal{B}$ the subtree rooted at $a_{w'}$. We establish the following claim.

Claim.

1. $\mathsf{CN}^{\models}_{\mathcal{B}}(a_v) = \mathsf{CN}^{\models}_{\mathcal{A}}(a_v)$ for all $a_v \in \mathsf{Ind}(\mathcal{X})$;
2. $\mathsf{CN}^{\models}_{\mathcal{B}}(a_{u'v}) = \mathsf{CN}^{\models}_{\mathcal{A}}(a_{uv})$ for all $a_{u'v} \in \mathsf{Ind}(\mathcal{B})$;
3. $\mathsf{CN}^{\models}_{\mathcal{B}^-}(a_v) = \mathsf{CN}^{\models}_{\mathcal{A}^-}(a_v)$ for all $a_v \in \mathsf{Ind}(\mathcal{X})$;
4. $\mathsf{CN}^{\models}_{\mathcal{B}^-}(a_{u'v}) = \mathsf{CN}^{\models}_{\mathcal{A}^-}(a_{uv})$ for all $a_{u'v} \in \mathsf{Ind}(\mathcal{B}^-)$.

The proof is by a straightforward induction, noting that in any tree-shaped ABox $\mathcal{Y}$, the value of $\mathsf{CN}^{\models}_{\mathcal{Y}}(a)$ for any $a \in \mathsf{Ind}(\mathcal{A})$ depends only on the values of $\mathsf{CN}^{\models}_{\mathcal{Y}}(b)$ for all successors b of a in $\mathcal{Y}$ because $\mathcal{T}$ is in normal form. Details are left to the reader.

It is an immediate consequence of Points 1 and 3 and the minimality of $\mathcal{A}$ that $\mathcal{T}, \mathcal{B} \models A(a_0)$ and $\mathcal{T}, \mathcal{B}^- \not\models A(a_0)$. Since the level of $a_{w'}$ in $\mathcal{B}$ clearly exceeds $2^{2n}+1$, this entails $\mathcal{T}, \mathcal{B}|_{2^{2n}+1} \not\models A(a_0)$, as required.

Repeating the operation used to obtain $\mathcal{B}$ from $\mathcal{A}$, we will eventually find the required ABox of depth exceeding k such that Conditions 1 and 2 from Theorem 2 are satisfied. ❏

ExpTime Decision Procedure

To develop an ExpTime procedure for deciding the existence of a tree-shaped Σ-ABox that satisfies Conditions (1) and (2) of Theorem 3, we utilize non-deterministic bottom-up automata on finite, ranked trees. Such an automaton has the form $\mathfrak{A} = (Q, \mathcal{F}, Q_f, \Theta)$ with Q a finite set of *states*, $\mathcal{F}$ a *ranked alphabet*, $Q_f \subseteq Q$ a set of *final states*, and Θ a set of *transition rules* of the form $f(q_1, \dots, q_n) \to q$, where $n \geq 0$, $f \in \mathcal{F}$ is of rank n, and $q_1, \dots, q_n, q \in Q$. Transition rules for predicates of rank 0 replace initial states.

Automata work on finite, node-labeled, ordered trees $T = (V, E, \ell)$, where V is a finite set of nodes, $E \subseteq V \times V$ is a set of edges, and ℓ is a node-labeling function that maps each node $v \in V$ with i successors to a predicate $\ell(v) \in \mathcal{F}$ of rank i. We assume an implicit total order on the successors of each node. A *run* of the automaton $\mathfrak{A}$ on T is a map $\rho : V \to Q$ such that $\rho(\varepsilon) \in Q_f$, with $\varepsilon \in V$ the root of T, and for all $v \in V$ with $\ell(v) = f$ and where v has (ordered) successors $v_1, \dots, v_n$, $n \geq 0$, we have that $f(\rho(v_1), \dots, \rho(v_n)) \to \rho(v)$ is a rule in Θ. An automaton $\mathfrak{A}$ *accepts* a tree T if there is a run of $\mathfrak{A}$ on T. We use $L(\mathfrak{A})$ to denote the set of all trees accepted by $\mathfrak{A}$. It can be computed in polynomial time whether $L(\mathfrak{A}) = \emptyset$.

Given an $\mathcal{EL}$-TBox $\mathcal{T}$ in normal form, ABox signature Σ, and IQ $A(x)$, we outline how to construct an automaton $\mathfrak{A}_{\mathcal{T},\Sigma,A} = (Q, \mathcal{F}, Q_f, \Theta)$ such that $L(\mathfrak{A}_{\mathcal{T},\Sigma,A})$ is empty if and only if there is some tree-shaped Σ-ABox satisfying the conditions of Theorem 3. The automaton $\mathfrak{A}_{\mathcal{T},\Sigma,A}$ takes as input trees whose nodes are labeled with tuples $\langle t, r_1, \dots, r_k \rangle$, where t is a set of concept names from Σ, and $r_1, \dots, r_k$ are role names from Σ. We choose k in such a way that if there exists tree-shaped Σ-ABox satisfying the conditions of Theorem 3, then there exists one in which every non-leaf individual has precisely k successors. To every tree $T = (V, E, \ell)$, we can associate the tree-shaped Σ-ABox

$$\begin{aligned}\mathcal{A}_T := &\{A(a_v) \mid v \in V \text{ and } \ell(v) = \langle t, r_1, \dots, r_k \rangle \text{ with } A \in t\} \cup \\ &\{r(a_v, a_{v_i}) \mid v_i \text{ is } i\text{-th successor of } v \text{ and } \ell(v) = \langle t, r_1, \dots, r_k \rangle \text{ with } r_i = r\}.\end{aligned}$$

The states in Q are of the form $(c_1, \sigma_1, c_2, \sigma_2)$ where $c_1, c_2 \in [0, \ldots, 2^{2n} + 1]$ (with $n = |\mathcal{T}|$) and σ_1, σ_2 are *types*, i.e. subsets of $\mathsf{N_C}$. States with $c_1 = 0$, $c_2 = 1$, $A \in \sigma_1$, and $A \notin \sigma_2$ are final.

The first state component c_1 implements a counter which is used to check whether the input tree has depth at least $2^{2n} + 1$. The second state component σ_1 is used to keep track of the atomic concepts that are entailed at the ABox individual associated with a tree node; requiring that A belong to σ_1 at the root enforces that A is entailed at the root a_0 of $\mathcal{A}_T$. The third state component c_2 serves to identify nodes which appear on the first 2^{2n} levels of T. Finally, the fourth state component σ_2 computes the set of entailed atomic concepts with respect to the ABox $\mathcal{A}_T|_{2^{2n}}$, exploiting the value of the counter c_2 in order to identify the leaves of $\mathcal{A}_T|_{2^{2n}}$. By prohibiting A from appearing in σ_2 at the root, we ensure that $\mathcal{T}, \mathcal{A}'_T \not\models A(a_0)$.

We now give the proof of Theorem 4 which is based upon the automaton construction outlined in Section 3.

Theorem 4. Deciding FO-rewritability of an IQ relative to an $\mathcal{EL}$-TBox and ABox signature is in ExpTime.

Proof. Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, Σ be an ABox signature, $A_0(x)$ be an IQ, and $n = |(\mathsf{sig}(\mathcal{T}) \cup \Sigma) \cap \mathsf{N_C}|$.

We can assume w.l.o.g. that the ABox signature Σ is finite since 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 $(\mathsf{sig}(\mathcal{T}) \cup \{A\}) \cap \Sigma$. Because of Lemma 1, we can thus also assume that $\mathcal{T}$ is in normal form. In particular, this means that every subconcept $\exists r.C$ in $\mathcal{T}$ has $C \in \mathsf{N_C}$. We will use the term *type* to denote a subset of $\mathsf{N_C}$, and *Σ-type* to denote a subset of $\Sigma \cap \mathsf{N_C}$. Given a set of $\mathcal{EL}$-concepts S, we will use $\mathsf{type}_{\mathcal{T}}(S)$ to denote the type $\{B \in \mathsf{N_C} \mid \mathcal{T} \models \bigsqcap S \sqsubseteq B\}$. We use $\mathsf{ex}(\mathcal{T})$ to denote the number of concepts of the form $\exists r.B$ that occur in $\mathcal{T}$.

Define an automaton $\mathfrak{A} = (Q, \mathcal{F}, Q_f, \Delta)$ as follows:

- $\mathcal{F} = \{\langle \tau, r_1, \ldots, r_n \rangle \mid \tau \text{ is a } \Sigma\text{-type}, r_1, \ldots, r_n \in \Sigma \cap \mathsf{N_R}, 0 \leq n \leq \mathsf{ex}(\mathcal{T})\}$ where each $\langle \tau, r_1, \ldots, r_n \rangle$ is of rank n;

- Q is the set of tuples $(c_1, \sigma_1, c_2, \sigma_2)$ such that $c_1, c_2 \in [0, \ldots, 2^{2n} + 1]$ and σ_1, σ_2 are types;

- $Q_f = \{(c_1, \sigma_1, c_2, \sigma_2) \in Q \mid c_1 = 0, c_2 = 1, A_0 \in \sigma_1, A_0 \notin \sigma_2\}$;

– Θ consists of all rules $f(q_1, \ldots, q_n) \to q$ with $f = \langle \tau, r_1, \ldots, r_n \rangle$, $q_i = (d_1^i, d_2^i, \chi_i^1, \chi_2^i)$, $q = (c_1, c_2, \sigma_1, \sigma_2)$ such that:

$$c_1 = \begin{cases} 2^{2n} + 1 & f \text{ has rank } 0 \\ 0 & \min_{1 \leq i \leq n} d_1^i = 0 \\ \min_{1 \leq i \leq n} d_1^i - 1 & \text{otherwise} \end{cases}$$

$$c_2 \in \begin{cases} \{0, \ldots, 2^{2n} + 1\} & f \text{ has rank } 0 \\ \{2^{2n}, 2^{2n} + 1\} & d_2^1 = \ldots = d_2^n = 2^n + 1 \\ \{v - 1\} & d_2^1 = \ldots = d_2^n = v \text{ with } v \leq 2^{2n} \\ \{0\} & \text{otherwise} \end{cases}$$

$$\sigma_1 = \begin{cases} \mathsf{type}_{\mathcal{T}}(\tau) & f \text{ has rank } 0 \\ \mathsf{type}_{\mathcal{T}}(\tau \cup \{\exists r.B \mid r = r_i \text{ and } B \in \chi_2^i \text{ for some } i\}) & \text{otherwise} \end{cases}$$

$$\sigma_2 = \begin{cases} \mathsf{type}_{\mathcal{T}}(\tau) & f \text{ has rank } 0 \\ \mathsf{type}_{\mathcal{T}}(\tau) & c_2 = 2^{2n} + 1 \\ \mathsf{type}_{\mathcal{T}}(\tau \cup \{\exists r.B \mid r = r_i \text{ and } B \in \chi_2^i \text{ for some } i\}) & \text{otherwise} \end{cases}$$

We briefly explain the intuition behind the automaton construction. The first state component c_1 implements a counter which is used to check whether the input tree has depth at least $2^{2n} + 1$. More precisely, the transition rules in Θ assign a c_1-value of $2^{2n} + 1$ to leaf nodes; internal nodes receive a c_1-value which is either one less than the least c_1-value of its children, or 0 if the latter is negative. Trees accepted by the automaton have $c_1 = 0$ at the root, and hence have depth at least $2^{2n} + 1$.

The second state component σ_1 is used to keep track of the atomic concepts that are entailed at the ABox individual associated with a tree node. More precisely, the transition rules will ensure that a node v is assigned a set σ_1 which contains precisely those $B \in \mathsf{N_C}$ such that $\mathcal{T}, \mathcal{A}_T \models B(a_v)$. Thus, requiring that A belong to σ_1 at the root enforces that A is entailed at the root a_0 of $\mathcal{A}_T$.

The third state component c_2 serves to identify nodes which appear on the first 2^{2n} levels of T. At leaf nodes, we guess a value between 0 and $2^{2n} + 1$. For internal nodes, the transition rules verify that a node's children all have the same c_2-value. If the value is between 1 and 2^{2n}, then this means the counter is "active", and so we decrement the counter by one; if the value is $2^{2n} + 1$, then we either stay at $2^{2n} + 1$ or switch to 2^{2n} (starting the counter); and if the value is 0, we stay at 0. Since Q_f requires that the root have $c_2 = 1$, it follows that in every run, the nodes at level $l \leq 2^{2n}$ have c_2-value $l + 1$.

Finally, the fourth state component σ_2 computes the set of entailed atomic concepts with respect to the restriction $\mathcal{A}_T|_{2^{2n}}$, exploiting the value of the counter c_2 in order to identify the leaves of $\mathcal{A}_T|_{2^{2n}}$. By prohibiting A from appearing in σ_2 at the root, we ensure that $\mathcal{T}, \mathcal{A}_T|_{2^{2n}} \not\models A(a_0)$.

Since $\mathfrak{A}$ is single-exponentially large in $|\mathcal{T}|$ and the emptiness problem can be decided in polynomial time in the size of the automaton, it remains to establish the following claim (using the arguments informally presented above) to obtain a single-exponential-time procedure for FO-rewritability of IQs.

Claim 1. $L(\mathfrak{A}) \neq \emptyset$ iff A_0 is not FO-rewritable relative to $\mathcal{T}$ and Σ.

Before we can establish Claim 1, we prove the following technical result.

Claim 2. For every Σ-ABox $\mathcal{A}$, $\mathcal{EL}$-TBox $\mathcal{T}$ in normal form, $a \in \mathsf{Ind}(\mathcal{A})$, and $A \in \mathsf{N_C}$, we have $\mathcal{T}, \mathcal{A} \models A(a)$ iff $A \in \mathsf{type}_{\mathcal{T}}(t_a)$ where $t_a = \{B \in \Sigma \mid B(a) \in \mathcal{A}\} \cup \{\exists r.B \mid B \in \mathsf{N_C}$ and there exists b such that $r(a,b) \in \mathcal{A}$ and $\mathcal{T}, \mathcal{A} \models B(b)\}$.

Since the "if" direction is straightforward, we concentrate on the "only if" direction. Let $\mathcal{A}$ be a Σ-ABox and $\mathcal{T}$ an $\mathcal{EL}$-TBox in normal form. Assume that $A \in \mathsf{N_C}$ is such that $A \notin \mathsf{type}_{\mathcal{T}}(t_a)$. We have to show that $\mathcal{T}, \mathcal{A} \not\models A(a)$. Using Lemma 7, for each $b \in \mathsf{Ind}(\mathcal{A})$ we can find an interpretation $\mathcal{I}_b$ and an element $d_{\mathcal{I}_b} \in \Delta^{\mathcal{I}_b}$ such that for all atomic concepts $B \in \mathsf{N_C}$, we have $d_{\mathcal{I}_b} \in B^{\mathcal{I}_b}$ iff $B \in \mathsf{type}_{\mathcal{T}}(t_b)$: in that lemma, simply choose the ABox $\{B(b) \mid B \in \mathsf{type}_{\mathcal{T}}(t_b)\}$. We may assume that the $\Delta^{\mathcal{I}_b}$ are mutually disjoint. Take the following union $\mathcal{I}$ of the models $\mathcal{I}_b$:

- $\Delta^{\mathcal{I}} = \bigcup_{b \in \mathsf{Ind}(\mathcal{A})} \Delta^{\mathcal{I}_b}$;
- $B^{\mathcal{I}} = \bigcup_{b \in \mathsf{Ind}(\mathcal{A})} B^{\mathcal{I}_b}$, for $B \in \mathsf{N_C}$;
- $r^{\mathcal{I}} = \bigcup_{b \in \mathsf{Ind}(\mathcal{A})} r^{\mathcal{I}_b} \cup \{(d_{\mathcal{I}_b}, d_{\mathcal{I}_c}) \mid r(b,c) \in \mathcal{A}\}$, for $r \in \mathsf{N_R}$;
- $b^{\mathcal{I}} = d_{\mathcal{I}_b}$, for $b \in \mathsf{Ind}(\mathcal{A})$.

Note that by construction, $\mathcal{I}$ is a model of $\mathcal{A}$. We now show that $\mathcal{I}$ is also a model of $\mathcal{T}$. We recall that $\mathcal{T}$ is in normal form, and so each inclusion in $\mathcal{T}$ has one of the following forms:

$$B \sqsubseteq \top \qquad B_1 \sqsubseteq B_2 \qquad B_1 \sqsubseteq \exists r.B_2 \qquad \top \sqsubseteq B \qquad B_1 \sqcap B_2 \sqsubseteq B_3 \qquad \exists r.B_1 \sqsubseteq B_2$$

where $B, B_1, B_2, B_3 \in \mathsf{N_C}$. We note that it follows from the construction of $\mathcal{I}$ that for all $B \in \mathsf{N_C}$, $b \in \mathsf{Ind}(\mathcal{A})$, and $d \in \Delta^{\mathcal{I}_b}$, we have

$$d \in B^{\mathcal{I}_b} \quad \text{iff} \quad d \in B^{\mathcal{I}}$$

Using this property and the fact that each $\mathcal{I}_b$ is a model of $\mathcal{T}$, we can immediately derive the satisfaction of inclusions in $\mathcal{T}$ of the forms $B \sqsubseteq \top$, $B_1 \sqsubseteq B_2$, $\top \sqsubseteq B$, and $B_1 \sqcap B_2 \sqsubseteq B_3$. Next take some inclusion $B_1 \sqsubseteq \exists r.B_2 \in \mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$ such that $d \in B_1^{\mathcal{I}}$. We know that d belongs to some $\mathcal{I}_b$. As $\mathcal{I}_b$ is a model of $\mathcal{T}$, it follows there is some $e \in \mathcal{I}_b$ such that $(d,e) \in r^{\mathcal{I}_b}$ and $e \in B_2^{\mathcal{I}_b}$. As $r^{\mathcal{I}_b} \subseteq r^{\mathcal{I}}$ and $B_2^{\mathcal{I}_b} \subseteq B_2^{\mathcal{I}}$, we obtain $(d,e) \in r^{\mathcal{I}}$ and $e \in B_2^{\mathcal{I}}$, so this inclusion is satisfied in $\mathcal{I}$. Finally, suppose we have an inclusion $\exists r.B_1 \sqsubseteq B_2 \in \mathcal{T}$ and $d \in (\exists r.B_1)^{\mathcal{I}}$. If d is not the root of some $\mathcal{I}_b$, then we can simply use the fact that each $\mathcal{I}_b$ is a model of $\mathcal{T}$. So suppose that $d = d_{\mathcal{I}_b}$. Then we have that $\exists r.B_1 \in t_b$, and so $B_2 \in \mathsf{type}_{\mathcal{T}}(t_b)$. It follows that $d = d_{\mathcal{I}_b} \in B_2^{\mathcal{I}_b} \subseteq B_2^{\mathcal{I}}$. We have thus shown that all inclusions in $\mathcal{T}$ hold in $\mathcal{I}$, so $\mathcal{I}$ is a model of $\mathcal{T}$. To complete the proof of the claim, we note that $d_{\mathcal{I}_a} \notin A^{\mathcal{I}_a}$, hence $a^{\mathcal{I}} = d_{\mathcal{I}_a} \notin A^{\mathcal{I}}$, yielding the desired $\mathcal{T}, \mathcal{A} \not\models A(a)$.

We now prove Claim 1, starting with the "if" direction. Suppose that A_0 is not FO-rewritable relative to $\mathcal{T}$ and Σ. It follows from Theorem 3 that there exists a tree-shaped Σ-ABox $\mathcal{A}$ of depth exceeding 2^{2n} with root a_0 such that $\mathcal{T}, \mathcal{A} \models A_0(a_0)$, and $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models A_0(a_0)$. When $r(a,b) \in \mathcal{A}$, we call b a *successor* of a in $\mathcal{A}$. By Claim 2, we can assume without loss of generality that the number of successors of each $a \in \mathsf{Ind}(\mathcal{A})$ in $\mathcal{A}$ is bounded by $\mathsf{ex}(\mathcal{T})$. Indeed, if it is not the case, choose for each $\exists r.B$ (with $B \in \mathsf{N_C}$) some $b \in \mathsf{Ind}(\mathcal{A})$ with $r(a,b) \in \mathcal{A}$ and $\mathcal{T}, \mathcal{A} \models B(b)$ (if such a b exists), and then drop all subtrees rooted at successors of a in $\mathcal{A}$ that have not been chosen. The resulting ABox still entails $A_0(a_0)$ given $\mathcal{T}$ due to Claim 2 and the fact that for all individual names a in the resulting ABox, the set $\mathsf{type}_{\mathcal{T}}(a)$ did not change. It is also clear that the operation preserves the non-entailment of $A_0(a_0)$ for the restriction of the ABox to depth 2^{2n}, which also means that the modified ABox continues to have depth exceeding 2^{2n}.

For each individual in $\mathcal{A}$, fix a total order on the successors. For $a \in \mathsf{Ind}(\mathcal{A})$, we use $\tau_{\mathcal{A}}(a)$ to denote the set $\{A \in \Sigma \mid A(a) \in \mathcal{A}\}$. Define a tree $T = (V, E, \ell)$ as follows:

- $V = \mathsf{Ind}(\mathcal{A})$;
- $E = \{(a,b) \in V \times V \mid r(a,b) \in \mathcal{A}\}$ and the order of successor in T agrees with the chosen order on successors in $\mathcal{A}$;
- $\ell(a) = \langle \tau_{\mathcal{A}}(a), r_1, \ldots, r_n \rangle$ where r_i is the (unique!) role such that $r_i(a, a_i) \in \mathcal{A}$, with a_i the i-th successor of a.

Define a mapping ρ that maps each $a \in \mathsf{Ind}(\mathcal{A})$ to the state $(c_1, c_2, \sigma_1, \sigma_2)$ defined as follows:

- letting d_{T_a} be the depth of the subtree of T based at a, we set $c_1 = (2^{2n}+1) - d_{T_a}$ if $d_{T_a} \leq 2^{2n}+1$, else $c_1 = 0$
- letting d_a be the level of a in T, we set $c_2 = d$ if $d_a \leq 2^{2n}$, else $c_2 = 2^{2n}+1$
- $\sigma_1 = \{B \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A} \models B(a)\}$
- $\sigma_2 = \{B \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A}|_{2^{2n}} \models B(a)\}$ if $c_2 \leq 2^{2n}$, else $\sigma_2 = \mathsf{type}_{\mathcal{T}}(\tau_{\mathcal{A}}(a))$

We show that ρ is a run of $\mathfrak{A}$ on T. Let $\rho(a_0) = (c_1, c_2, \sigma_1, \sigma_2)$. By definition of ρ, we have $c_2 = 0$ (since a_0 has depth 0) and $c_1 = 0$ (since $d_T > 2^{2n}$). Next note that $\mathcal{T}, \mathcal{A} \models A_0(a_0)$ and $\mathcal{T}, \mathcal{A} \models A_0(a_0)$ imply respectively that $A_0 \in \sigma_1$ and $A_0 \notin \sigma_2$. We thus have $\rho(a_0) \in Q_f$. It remains to show that ρ respects the transition rules in Θ. This can be shown by induction on the co-depth of a node. For the base case, consider some individual a with no successors in T, and let $\rho(a) = (c_1, c_2, \sigma_1, \sigma_2)$. Since a is a leaf node, we have $c_1 = 2^{2n}+1$, and $c_2 \in \{0, \ldots, 2^{2n}+1\}$ which satisfies the conditions of Θ. From Claim 2 and the fact that a is a leaf node, we have that $\mathcal{T}, \mathcal{A} \models C(a)$ iff $C \in \mathsf{type}_{\mathcal{T}}(\tau_{\mathcal{A}}(a))$. It follows that c_1 and c_2 also satisfy Θ. For the induction step, suppose that the property holds for all $b \in \mathsf{Ind}(\mathcal{A})$ which are less than k steps from some leaf node in $\mathcal{A}$, and let a be some individual which is exactly distance k from a leaf node. Suppose that $\ell(a) = \langle \tau_{\mathcal{A}}(a), r_1, \ldots, r_n \rangle$, with e_i being the i-th child of a. Let $\rho(a) = (c_1, c_2, \sigma_1, \sigma_2)$, and let $\rho(b_i) = (d_1^i, d_2^i, \chi_i^1, \chi_2^i)$ for each $1 \leq i \leq n$. Let $m = \min_{1 \leq i \leq n} d_1^i$. If $m = 0$, then this means that for one of the children b_i, the subtree rooted at b_i has depth at least $2^{2n}+1$, hence $d_1^i = 0$. It follows the same must be true of the subtree rooted at a, yielding $c_1 = 0$, which means Θ is satisfied. If instead $m = \min_{1 \leq i \leq n} d_1^i > 0$, then the maximum depths of a subtree rooted at some child b_i is g, where $g = 2^{2n}+1-m$. So the subtree rooted at a must have depth $g+1$, which means $c_1 = m-1$, again satisfying Θ. We know that c_2 is either the depth of a in T, or $2^{2n}+1$, whichever is smaller. Now if c_2 is $2^{2n}+1$, this means its children $b_1, \ldots, b_n$ all have the same depth greater than $2^{2n}+1$, hence $d_2^1 = \ldots = d_2^n = 2^{2n} + 1$. If instead $c_2 < 2^{2n} + 1$, then the child nodes all have depth $c_2 + 1$, so $d_2^1 = \ldots = d_2^n = c_2 + 1$. In both cases, Θ is respected. Next we note that $\sigma_1 = \{B \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A} \models B(a)\}$, and by Claim 2, $\mathcal{T}, \mathcal{A} \models B(a)$ if and only if $B \in \mathsf{type}_\mathcal{T}(t_a)$ where $t_a = \{D \in \Sigma \mid D(a) \in \mathcal{A}\} \cup \{\exists r.D \mid D \in \mathsf{N_C}$ and there exists b such that $r(a,b) \in \mathcal{A}$ and $\mathcal{T}, \mathcal{A} \models D(b)\}$. From the definition of ρ, we have $\chi_i^1 = \{D \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A} \models D(b_i)\}$ for each $1 \leq i \leq n$ It follows that $\sigma_1 = \mathsf{type}_\mathcal{T}(\tau_\mathcal{A}(a) \cup \{\exists r.D \mid r = r_i$ and $D \in \chi_2^i$ for some $i\}$, and hence that the conditions of Θ are satisfied w.r.t. σ_1. Finally, for σ_2, we remark that Θ is trivially satisfied when $c_2 = 2^{2n} + 1$, and otherwise, we can use a similar argument as was used for c_1 to show that $\sigma_2 = \mathsf{type}_\mathcal{T}(\tau \cup \{\exists r.D \in\mid r = r_i$ and $D \in \chi_2^i$ for some $i\})$.

For the "only if" direction of Claim 1, let $T = (V, E, \ell)$ be a tree accepted by $\mathfrak{A}$, and ρ be a run of $\mathfrak{A}$ on T. Define a tree-shaped Σ-ABox

$$\mathcal{A} := \{A(a_v) \mid v \in V \text{ and } \ell(v) = \langle t, r_1, \ldots, r_n\rangle \text{with } A \in t\} \cup$$
$$\{r(a_v, a_{v_i}) \mid v_i \text{ is } i\text{-th successor of } v \text{ and } \ell(v) = \langle t, r_1, \ldots, r_n\rangle \text{ with } r_i = r\}$$

Let a_0 be the root of $\mathcal{A}$.We want to show that $\mathcal{A}$ witnesses the non-FO-rewritability of $A_0(x)$. More precisely, we aim to prove that (i) $\mathcal{A}$ has depth at least $2^{2n} + 1$, (ii) $\mathcal{T}, \mathcal{A} \models A_0(a_0)$, and (iii) $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models A_0(a_0)$.

First note that if $\rho(a_0) = (c_1, c_2, \sigma_1, \sigma_2)$, then we must have $c_1 = 0$, $c_2 = 1$, $A_0 \in \sigma_1$, and $A_0 \notin \sigma_2$. For (i), we note that $c_1 = 0$ means that $\mathcal{A}$ has depth at least $2^{2n} + 1$, since Θ ensures that leaf nodes start with value $2^{2n} + 1$ and that values are decremented by exactly one when moving from child to parent. For (ii), we show the following claim, where $\rho(v) = (c_1^v, c_2^v, \sigma_1^v, \sigma_2^v)$.

Claim 4. For all $v \in V$ and $B \in \sigma_1^v$: $\mathcal{T}, \mathcal{A} \models B(a_v)$.

The proof is by induction on the co-depth of v. If v is a leaf and $B \in \sigma_1^v$, then the definition of Θ and $\mathcal{A}$ yields $B \in \mathsf{type}_\mathcal{T}(\{A \mid A(a_v) \in \mathcal{A}\})$. It follows that $\mathcal{T}, \mathcal{A} \models B(a_v)$. Now let v be a non-leaf with $\ell(v) = \langle t, r_1, \ldots, r_n\rangle$ and successors $v_1, \ldots, v_n$. Moreover, let $C \in \sigma_1^v$. Then from the definition of $\mathcal{A}$ and Θ, we have

$$B \in \mathsf{type}_\mathcal{T}(\{A \mid A(a_v) \in \mathcal{A}\} \cup \{\exists r.D \mid r = r_i,\ D \in \sigma_1^{v_i} \text{ for some } 1 \leq i \leq n\})$$

By the induction hypothesis, we know that $D \in \sigma_1^{v_i}$ implies $\mathcal{T}, \mathcal{A} \models D(a_{v_i})$. Thus, we also have $\mathcal{T}, \mathcal{A} \models B(a_v)$. This completes the proof of Claim 4. Using this claim, we can infer that $\mathcal{T}, \mathcal{A} \models A_0(a_0)$, as desired.

To show (iii), we start by establishing the following claim:

Claim 5. For all $v \in V$ with depth $d \leq 2^{2n}$: $c_2^v = d + 1$.

The proof of Claim 5 proceeds by induction on the depth of v. If v has depth 0 (i.e. it is the root node a_0), then we know from above that $c_2^v = 0$. Now suppose the claim holds for all nodes with depth at most k, and consider some node v with depth $k + 1 \leq 2^{2n}$. We know from the IH that v's unique parent node u is such that $c_2^u = k$. Thus, from the definition of Θ and the fact that ρ is a run, we must have $c_2^v = k + 1$. This completes the proof of Claim 5. We now use this claim to establish the following:

Claim 6. For all $v \in V$ with depth at most 2^{2n}: $\sigma_2^v = \{D \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A}|_{2^{2n}} \models D(a_v)\}$

The proof is by induction on the co-depth of a_v in $\mathcal{A}|_{2^{2n}}$. We thus have two base cases: leafs in $\mathcal{A}$ of depth at most 2^{2n} and non-leaves in $\mathcal{A}$ with precisely depth 2^{2n}. First suppose a_v is a leaf of $\mathcal{A}$ appearing at depth at most 2^{2n}. Then by definition of Θ, we will have $\sigma_2^v = \mathsf{type}_\mathcal{T}(\{A \mid A(a_v) \in \mathcal{A}\})$, and by Claim 2, we have $\sigma_2^v = \{D \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A} \models D(a_v)\}$, and hence $\sigma_2^v = \{D \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A}|_{2^{2n}} \models D(a_v)\}$. Next consider the case where a_v is a non-leaf individual in $\mathcal{A}$ with depth 2^{2n}. Then by Claim 5, we have $c_2^v = 2^{2n} + 1$, which means that $\sigma_2^v = \mathsf{type}_\mathcal{T}(\{A \mid A(a_v) \in \mathcal{A}\})$. It follows then by Claim 2 and the fact that a_v is a leaf in $\mathcal{A}|_{2^{2n}}$ with the same concept assertions as in $\mathcal{A}$ that $\sigma_2^v = \{D \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A}|_{2^{2n}} \models D(a_v)\}$. Now suppose that the claim holds for individuals whose co-depth in $\mathcal{A}|_{2^{2n}}$ is at most k, and let a_v have co-depth $k+1$. Suppose that $\ell(v) = \langle t, r_1, \ldots, r_n \rangle$ with successors $v_1, \ldots, v_n$. Using Claim 5, we can infer that $c_2^v < 2^{2n} + 1$. Then using the definitions of $\mathcal{A}$ and Θ, we have:

$$\sigma_2^v = \mathsf{type}_\mathcal{T}(\{A \mid A(a_v) \in \mathcal{A}\} \cup \{\exists r.D \mid r = r_i \text{ and } D \in \chi_2^i \text{ for some } i\})$$

By the induction hypothesis, we know that for each child a_{v_i}, we have $\sigma_2^{v_i} = \{D \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A}|_{2^{2n}} \models D(a_{v_i})\}$. It follows then from Claim 2 that $\sigma_2^v = \{D \in \mathsf{N_C} \mid \mathcal{T}, \mathcal{A}|_{2^{2n}} \models D(a_v)\}$. This completes the proof of Claim 6. We can apply this claim to the root individual a_0, using the fact that $A_0 \notin \sigma_2^{a_0}$ to infer that $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models A_0(a_0)$.

Now that we have found an ABox $\mathcal{A}$ satisfying properties (i), (ii), and (iii), we can use Theorem 3 to infer that A_0 is not FO-rewritable relative to $\mathcal{T}$ and Σ. ❑

Normal Form for Full Signature

Lemma 2. For any $\mathcal{EL}$-TBox $\mathcal{T}$ and IQ $A(x)$, there is a TBox $\mathcal{T}'$ in normal form such $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full signature if and only if $A(x)$ is FO-rewritable relative to $\mathcal{T}'$ and the full signature.

Proof. First suppose that $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full signature. Then by Theorem 1, there exists a tree-UCQ φ such that $\mathsf{cert}_\mathcal{T}(\mathcal{A}, A(x)) = \mathsf{ans}(\mathcal{I}_\mathcal{A}, \varphi)$ for every ABox $\mathcal{A}$. W.l.o.g., we can assume that φ does not use the new concept names X_D from $\mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$. We aim to modify φ to obtain a FO-rewriting φ' of $A(x)$ relative to $\mathcal{T}'$ and the full signature. Intuitively, φ' will be a UCQ consisting of all CQs which are obtained by taking a disjunct of φ and replacing some of the subqueries by concept names introduced in $\mathcal{T}'$. Formally, to every tree-shaped 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 tree-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 φ' be the disjunction of all queries q' which are definition-rewritings of some disjunct q of φ. Note that φ' is well-defined since there can be only finitely many definition-rewritings of a given CQ.

We aim to show that φ' is such that FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and the full signature. Let $\mathcal{A}'$ be an ABox. First suppose that $a \in \mathsf{cert}'_{\mathcal{T}}(\mathcal{A}', A(x))$. We define an 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:

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

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 $\mathsf{inst}(D, b)$.

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 of our assumption $a \in \mathsf{cert}_{\mathcal{T}'}(\mathcal{A}', A(x))$, we must also have $a \in \mathsf{cert}_{\mathcal{T}'}(\mathcal{A}, A(x))$. As the answers to instance queries are preserved under the normalization procedure, we must also have $a \in \mathsf{cert}_{\mathcal{T}}(\mathcal{A}, A(x))$. Using our assumption that φ is a FO-rewriting of $A(x)$ relative to $\mathcal{T}$, we obtain $\mathcal{I}_{\mathcal{A}} \models \varphi[a]$. It follows that there must exist a tree-CQ q which is a disjunct of φ such that $\mathcal{I}_{\mathcal{A}} \models q[a]$. 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 tree-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) $\alpha = B(v)$ ($B \in \mathsf{sig}(\mathcal{T}) \cup \{A\}$), $\pi(v) \in \mathsf{Ind}(\mathcal{A}')$, 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 tree-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 $B \in \mathsf{sig}(\mathcal{T}) \cup \{A\}$). 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 $(*)$. 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 the rule 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 φ'. 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 φ' such that $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}'}, q')$. We know from the construction of φ' that there exists a tree-CQ q which is a disjunct of φ 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 $\mathsf{Ind}(\mathcal{A}_k) \cap \mathsf{Ind}(\mathcal{A}_{k+1})$ 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 induction. Now since we have $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}_0}, q)$ and the query q uses no concept names X_D, it follows that the ABox $\mathcal{A}$ resulting from $\mathcal{A}_0$ by dropping all assertions of the form $X_D(b)$ also satisfies $a \in \mathsf{ans}(\mathcal{I}_{\mathcal{A}}, q)$. 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 $h_0, \ldots, h_{n-1}$, 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, and let φ' be a concrete FO-rewriting that is a UCQ. Trivially, φ' is also an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ over all Σ-ABoxes, where Σ is the full signature with all concept names X_D introduced during minimization removed. Clearly, this is still true if we remove from φ' all disjuncts that involve some concept name X_D. Call the resulting UCQ φ. Then φ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$. To see why, consider an arbitrary ABox $\mathcal{B}$, and let $\mathcal{B}^-$ be the restriction of $\mathcal{B}$ to Σ. Then we have the following chain of equivalences:

$$
\begin{aligned}
a \in \mathsf{cert}_{\mathcal{T}}(\mathcal{B}, A(x)) \quad &\Leftrightarrow \quad a \in \mathsf{cert}_{\mathcal{T}}(\mathcal{B}^-, A(x)) && (31)\\
&\Leftrightarrow \quad a \in \mathsf{ans}(\mathcal{I}_{\mathcal{B}^-}, \varphi) && (32)\\
&\Leftrightarrow \quad a \in \mathsf{ans}(\mathcal{I}_{\mathcal{B}}, \varphi) && (33)
\end{aligned}
$$

Note that the forward direction of (31) holds because any model of $(\mathcal{T}, \mathcal{B}')$ which witnesses $a \notin \mathsf{cert}_{\mathcal{T}}(\mathcal{B}', A(x))$ can be extended to a model of $(\mathcal{T}, \mathcal{B}')$ witnessing $a \notin \mathsf{cert}_{\mathcal{T}}(\mathcal{B}, A(x))$, simply by interpreting the symbols in $\mathsf{sig}(\mathcal{B}) \setminus \Sigma$ so as to satisfy the assertions in $\mathcal{B}' \setminus \mathcal{B}$. The implication from (33) to (32) uses our assumption that φ is a UCQ which only uses symbols from Σ. ❏

Characterizations of (Non)-FO-Rewritability in Terms of Linear ABoxes

We aim at proving Theorem 5. In view of Theorem 3, it is clearly sufficient to show the following.

Lemma 8. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox in normal form, $A(x)$ an IQ, and $k \geq 0$. If there is a tree-shaped ABox $\mathcal{A}$ of depth exceeding k with root a_0 such that $\mathcal{T}, \mathcal{A} \models A(a_0)$ and $\mathcal{T}, \mathcal{A}|_k \not\models A(a_0)$, then there is a linear ABox $\mathcal{A}'$ that satisfies the same properties.*

For an ABox $\mathcal{A}$ and an $a \in \mathsf{Ind}(\mathcal{A})$, we use

- $\mathsf{CN}_{\mathcal{A}}(a)$ to denote the restriction of $\mathcal{A}$ to assertions of the form $A(a)$;
- $\mathsf{CN}^{\models}_{\mathcal{A}}(a)$ to denote the set of all assertions $A(a)$ with $\mathcal{T}, \mathcal{A} \models A(a)$;
- $\mathsf{succ}_{\mathcal{A}}(a)$ to denote the set of all assertions in $\mathcal{A}$ that are of the form $r(a, b)$;
- $\mathsf{SCN}^{\models}_{\mathcal{A}}(a)$ to denote the ABox that comprises all concept assertions $A(b)$ such that $\mathcal{T}, \mathcal{A} \models A(b)$ and b is a successor of a in $\mathcal{A}$;

– $\mathsf{up}_{r,A}(a)$ to denote the ABox $\{B(a) \mid \exists r.A \sqsubseteq B \in \mathcal{T}\}$.

For a set of concept assertions $\mathcal{B}$, we write $\mathcal{T}, \mathcal{A} \models \mathcal{B}$ if $\mathcal{T}, \mathcal{A} \models A(a)$ for all $A(a) \in \mathcal{B}$. We will make use of the following straightforward properties.

Lemma 9. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox in normal form, $\mathcal{A}$ a tree-shaped ABox, and $a \in \mathsf{Ind}(\mathcal{A})$. Then for all $A \in \mathsf{N_C}$,*

$$\mathcal{T}, \mathcal{A} \models A(a) \text{ iff } \mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a) \cup \mathsf{succ}_{\mathcal{A}}(a) \cup \mathsf{SCN}^{\models}_{\mathcal{A}}(a)) \models A(a).$$

Lemma 10. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox in normal form, $\mathcal{A}$ a tree-shaped ABox, $a \in \mathsf{Ind}(\mathcal{A})$, $r(a,b) \in \mathcal{A}$. Then for all $A \in \mathsf{N_C}$, we have*

$$\mathcal{T}, \mathcal{A} \models A(a) \text{ iff } \mathcal{T}, (\mathcal{A}^{-b} \cup \bigcup_{B \in \mathsf{CN}^{\models}_{\mathcal{A}}(b)} \mathsf{up}_{r,B}(a)) \models A$$

where $\mathcal{A}^{-b}$ is $\mathcal{A}$ with $r(a,b)$ and the subtree rooted at b dropped.

We say that an ABox $\mathcal{A}'$ is a *fragment* of an ABox $\mathcal{A}$ if $\mathcal{A}' \subseteq \mathcal{A} \cup \{A(a) \mid \mathcal{T}, \mathcal{A} \models A(a)\}$. Instead of proving Lemma 8 directly, we establish the following lemma; Lemma 8 is then an immediate consequence by taking $n = 1$ and $S_1 = \{A\}$.

Lemma 11. *Let $\mathcal{A}$ be a tree-shaped ABox of depth exceeding k with root a_0 and let $S_1, \ldots, S_n$ be sets of assertions of the form $A(a_0)$ such that*

(i) $\mathcal{T}, \mathcal{A} \models S_i$ for some i with $1 \leq i \leq n$ and
(ii) $\mathcal{T}, \mathcal{A}|_k \not\models S_i$ for $1 \leq i \leq n$.

Then there is a linear ABox $\mathcal{A}'$ with root a_0 that is a fragment of $\mathcal{A}$ and satisfies (i) and (ii).

Proof. The proof is by induction on k.

Induction start, i.e., $k = 0$. Let $\mathcal{A}$, a_0, and $S_1, \ldots, S_n$ be as in the lemma. Since $\mathcal{T}, \mathcal{A} \models S_i(a_0)$ for some i and by Lemma 9, there is an $\Omega \subseteq \mathsf{SCN}^{\models}_{\mathcal{A}}(a_0)$ such that, for some S_{i_0}, we have

$$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \mathsf{succ}_{\mathcal{A}}(a_0) \cup \Omega) \models S_{i_0}. \qquad (\dagger)$$

Assume w.l.o.g. that Ω is inclusion-minimal with this property. Then Ω is non-empty: otherwise, we have $\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \mathsf{succ}_{\mathcal{A}}(a_0)) \models S_{i_0}$ which yields $\mathcal{T}, \mathsf{CN}_{\mathcal{A}}(a_0) \models S_{i_0}$ since $\mathcal{T}$ does not contain CIs of the form $\exists r.\top \sqsubseteq A$, which yields $\mathcal{T}, \mathcal{A}|_0 \models S_{i_0}$ in contradiction to Condition (ii) of the lemma. Choose an individual name a_1 that occurs in Ω and let $r_0(a_0, a_1)$ be the (unique) role assertion in $\mathcal{A}$ that connects a_0 and a_1. Let Ω_1 be the set of all assertions $A(a_1) \in \Omega$ and $\Omega^- = \Omega \setminus \Omega_{a_1}$. Define a new ABox

$$\mathcal{A}' = \mathsf{CN}_{\mathcal{A}}(a_0) \cup \{r_0(a_0, a_1)\} \cup \Omega_{a_1} \cup \bigcup_{r(a_0,b) \in \mathcal{A} \wedge A(b) \in \Omega^-} \mathsf{up}_{r,A}(a_0).$$

Note that $\mathcal{A}'$ is linear and a fragment of $\mathcal{A}$. From (†) and Lemma 10, we obtain $\mathcal{T}, \mathcal{A}' \models S_{i_0}$. To show that $\mathcal{T}, \mathcal{A}'|_0 \not\models S_i$ for all S_i, assume to the contrary that for some i,

$$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \bigcup_{r(a_0,b) \in \mathcal{A} \wedge A(b) \in \Omega^-} \mathsf{up}_{r,A}(a_0)) \models S_i.$$

By Lemma 10, we have

$$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \mathsf{succ}_{\mathcal{A}}(a_0) \cup \Omega^-) \models S_i,$$

in contradiction to the minimality of Ω.

Induction step. Assume that $k > 0$ and let $\mathcal{A}$, a_0, and $S_1, \dots, S_n$ be as in the lemma. Since $\mathcal{T}, \mathcal{A} \models S_i$ for some i and by Lemma 9, there is an $\Omega \subseteq \mathsf{SCN}^{\models}_{\mathcal{A}}(a_0) \setminus \mathsf{SCN}^{\models}_{\mathcal{A}|_k}(a_0)$ such that, for some S_i, we have

$$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \mathsf{succ}_{\mathcal{A}}(a_0) \cup \mathsf{SCN}^{\models}_{\mathcal{A}|_k}(a_0) \cup \Omega) \models S_i.$$

Since $\mathcal{T}, \mathcal{A}|_k \not\models S_i$ for all i and by Lemma 9, Ω is non-empty. Choose an individual name a_1 that occurs in Ω and let $r_0(a_0, a_1)$ be the (unique) role assertion in $\mathcal{A}$ that connects a_0 and a_1. Let $\Omega^- = \{A(b) \in \Omega \mid a \neq b\}$ and let $T_1, \dots, T_m$ be all inclusion-minimal subsets of $\mathsf{CN}^{\models}_{\mathcal{A}}(a_1) \setminus \mathsf{CN}^{\models}_{\mathcal{A}|_k}(a_1)$ such that for each T_i, there is an S_{j_i} with

$$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \mathsf{succ}_{\mathcal{A}}(a_0) \cup \mathsf{SCN}^{\models}_{\mathcal{A}|_k}(a_0) \cup \Omega^- \cup T_i(a_1)) \models S_{j_i}. \qquad (\dagger)$$

By choice of Ω, there is at least one such T_i. Finally, let $\mathcal{B}$ denote the restriction of $\mathcal{A}$ to the subtree rooted at a_1. We aim to apply the induction hypothesis (IH) with $\mathcal{B}$ playing the role of $\mathcal{A}$, a_1 the role of a_0, and $T_1, \dots, T_m$ the role of $S_1, \dots, S_n$. Indeed, we have $\mathcal{T}, \mathcal{B} \models T_i$ even for *all* i with $1 \leq i \leq m$ by choice of T_i and since $\mathsf{CN}^{\models}_{\mathcal{A}}(a_1) = \mathsf{CN}^{\models}_{\mathcal{B}}(a_1)$ and $\mathcal{T}, \mathcal{B}|_{k-1} \not\models T_i$ for $1 \leq i \leq m$ by choice of T_i and since $\mathsf{CN}^{\models}_{\mathcal{B}|_{k-1}}(a_1) \subseteq \mathsf{CN}^{\models}_{\mathcal{A}|_k}(a_1)$. Thus the IH is indeed applicable and we obtain a linear ABox $\mathcal{B}'$ that is a fragment of $\mathcal{B}$ (and thus of $\mathcal{A}$) such that

(*i*′) $\mathcal{T}, \mathcal{B}' \models T_i$ for some i with $1 \leq i \leq m$;
(*ii*′) $\mathcal{T}, \mathcal{B}'|_{k-1} \not\models T_i$ for $1 \leq i \leq m$.

Let Γ be the set of all pairs (r, A) with $r(a_0, b) \in \mathcal{A}$ and $A(b) \in \mathsf{SCN}^{\models}_{\mathcal{A}|_k}(a_0) \cup \Omega^-$. Define

$$\mathcal{A}' = \mathsf{CN}_{\mathcal{A}}(a_0) \cup \{r_0(a_0, a_1)\} \cup \mathcal{B}' \cup \bigcup_{r,A \in \Gamma} \mathsf{up}_{r,A}(a_0)$$

Note that $\mathcal{A}'$ is linear and a fragment of $\mathcal{A}$. It thus remains to show the following:

- $\mathcal{T}, \mathcal{A}' \models S_i$ for some i.
  By (*i*′), we have $\mathcal{T}, \mathcal{B}' \models T_i$ for some i. From (†), we thus obtain

  $$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \mathsf{succ}_{\mathcal{A}}(a_0) \cup \mathsf{SCN}^{\models}_{\mathcal{A}|_k}(a_0) \cup \Omega^- \cup \mathcal{B}') \models S_{j_i}.$$

  Now, Lemma 10 yields

  $$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \{r_0(a_0, a_1)\} \cup \bigcup_{r,A \in \Gamma} \mathsf{up}_{r,A}(a_0) \cup \mathcal{B}') \models S_{j_i}$$

  and it remains to note that the ABox in the above statement is $\mathcal{A}'$.

– $\mathcal{T}, \mathcal{A}'|_k \not\models S_i$ for $1 \leq i \leq n$.
Assume to the contrary that $\mathcal{T}, \mathcal{A}'|_k \models S_i$. By definition of $\mathcal{A}'$, this means that

$$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \{r_0(a_0, a_1)\} \cup \mathcal{B}'|_{k-1} \cup \bigcup_{r,A \in \Gamma} \mathsf{up}_{r,A}(a_0)) \models S_i.$$

By Lemma 10, we obtain

$$\mathcal{T}, (\mathsf{CN}_{\mathcal{A}}(a_0) \cup \mathsf{succ}_{\mathcal{A}}(a_0) \cup \mathsf{CN}^{\models}_{\mathcal{B}'|_{k-1}}(a_1) \cup \mathsf{SCN}^{\models}_{\mathcal{A}|_k}(a_0) \cup \Omega^-) \models S_i.$$

Since $\mathcal{B}'$ is a fragment of $\mathcal{A}$, we have $\mathsf{CN}^{\models}_{\mathcal{B}'|_{k-1}}(a_1) \subseteq \mathsf{CN}^{\models}_{\mathcal{A}}(a_1)$. By choice of $T_1, \ldots, T_m$, this means that $T_i \subseteq \mathsf{CN}^{\models}_{\mathcal{B}'|_{k-1}}(a_1)$ for some i with $1 \leq i \leq m$, in contradiction to (*ii*$'$).

❑

PSpace Decision Procedure

To obtain the PSPACE upper bound, we define a word automaton which accepts (words representing) linear ABoxes satisfying Conditions (1) and (2) of Theorem 5. Instead of defining a word automaton from scratch, we adapt the tree automaton construction from the EXPTIME upper bound.

Consider an $\mathcal{EL}$-TBox $\mathcal{T}$ in normal form and IQ $A(x)$, and let $\mathfrak{A} = (Q, \mathcal{F}, Q_f, \Delta)$ be the tree automaton for $\mathcal{T}, \mathsf{N_C}, A$, as defined in the proof of Theorem 4. We define another tree automaton $\mathfrak{A}^\ell = (Q, \mathcal{F}', Q_f, \Delta')$ by restricting the alphabet and transition rules:

– $\mathcal{F}' = \{\langle \tau \rangle \mid \tau$ is a $\mathsf{N_C}$-type $\} \cup \{\langle \tau, r \rangle \mid \tau$ is a $\mathsf{N_C}$-type, $r \in \mathsf{N_R}\}$
– Δ' contains all transitions from Δ using only symbols of arity 0 or 1

Note that because $\mathfrak{A}^\ell$ only uses symbols with arities at most 1, $L(\mathfrak{A}^\ell)$ consists of trees with branching factor 1, or equivalently, *words*. Every such tree/word defines a linear ABox. Also note that by construction, we have $\mathcal{F} \subseteq \mathcal{F}'$ and $\Delta' \subseteq \Delta$. This allows us to reuse the arguments from the proof of Theorem 4 to show that $L(\mathfrak{A}^\ell)$ consists of all words of length $2^{2n} + 1$ such that the associated linear ABox $\mathcal{A}$ is such that $\mathcal{T}, \mathcal{A} \models A(a_0)$ and $\mathcal{T}, \mathcal{A}|_{2^{2n}} \not\models A(a_0)$ (where a_0 is the root of $\mathcal{A}$). Thus, if $L(\mathfrak{A}^\ell)$ is non-empty, then there is a witness ABox satisfying the conditions of Theorem 5. Conversely, if there is a linear ABox satisfying the required conditions, we can assume that it has precisely length $2^{2n} + 1$, and hence $L(\mathfrak{A}^\ell) \neq \emptyset$.

The automaton $\mathfrak{A}^\ell$ has exponentially many states, so to obtain a PSPACE procedure, we must test for emptiness without actually constructing $\mathfrak{A}^\ell$. We note that given a symbol $f \in \mathcal{F}'$ and a pair of states $q, q' \in Q$, it can be decided in polytime (hence polyspace) whether $f(q) \to q'$. This means that we can decide non-emptiness by constructing on-the-fly a linear tree and associated run, by working from leaf to root and keeping only the two most recently generated symbol-state pairs in memory. A binary counter is used to count the number of pairs generated, and when we reach $2^{2n} + 1$, we stop and check whether the current pair contains a state from Q_f.

B Proofs for Section 5

We introduce some notions and results required for the proofs. A relation S between two Σ-ABoxes $\mathcal{A}_1$ and $\mathcal{A}_2$ is called a *simulation* if

- $(a, b) \in S$ and $A(a) \in \mathcal{A}_1$ imply $A(b) \in \mathcal{A}_2$;
- $(a, b) \in S$ and $r(a, a') \in \mathcal{A}_1$ imply that there exists b' with $(a', b') \in S$ and $r(b, b') \in \mathcal{A}_2$.

We say that $(\mathcal{A}_1, a_1)$ is *simulated by* $(\mathcal{A}_2, a_2)$, in symbols $(\mathcal{A}_1, a_1) \leq (\mathcal{A}_2, a_2)$, if there exists a simulation S between $\mathcal{A}_1$ and $\mathcal{A}_2$ with $(a_1, a_2) \in S$. The following lemma is readily checked.

Lemma 12. *For any two Σ-ABoxes $\mathcal{A}_1$ and $\mathcal{A}_2$, if $(\mathcal{A}_1, a_1) \leq (\mathcal{A}_2, a_2)$, then $(\mathcal{T}, \mathcal{A}_1) \models C(a_1)$ implies $(\mathcal{T}, \mathcal{A}_2) \models C(a_2)$, for all $\mathcal{EL}$-concepts C.*

We now define the ABox $\mathcal{A}_{\mathcal{T},\Sigma}$ in detail. It uses an individual name a_A, for each concept name A that is non-conjunctive in $\mathcal{T}$, and an additional individual name a_Σ. For each concept name A that is primitive in $\mathcal{T}$, set

$$\mathcal{A}(A) = \{E(a_A) \mid E \in \Sigma, \mathcal{T} \not\models E \sqsubseteq A\} \cup \{r(a_A, a_\Sigma) \mid r \in \Sigma\}$$

For each concept definition $A \equiv \exists r.B \in \mathcal{T}$, set

$$\begin{aligned}\mathcal{A}(A) = \{E(a_A) \mid E \in \Sigma, \mathcal{T} \not\models E \sqsubseteq A\} \cup \{s(a_A, a_\Sigma) \mid r \neq s \in \Sigma\} \\ \cup \{r(a_A, a_E) \mid E \in \text{non-conj}_{\mathcal{T}}(B), r \in \Sigma\}\end{aligned}$$

and let $\mathcal{A}_\Sigma = \{r(a_\Sigma, a_\Sigma) \mid r \in \Sigma\} \cup \{A(a_\Sigma) \mid A \in \Sigma\}$. Finally we set

$$\mathcal{A}_{\mathcal{T},\Sigma} = \mathcal{A}_\Sigma \cup \{\mathcal{A}(A) \mid A \text{ non-conjunctive in } \mathcal{T}\}.$$

The following lemma characterizes the ABox $\mathcal{A}_{\mathcal{T},\Sigma}$ (cf. [9]).

Lemma 13. *For every classical $\mathcal{EL}$-TBox $\mathcal{T}$ in normal form, every ABox-signature Σ, and concept name A, the following conditions are equivalent for every Σ-ABox $\mathcal{A}$ and individual name a in $\mathcal{A}$:*

- *there exists $B \in$ non-conj$_{\mathcal{T}}(A)$ such that $(\mathcal{A}, a) \leq (\mathcal{A}_{\mathcal{T},\Sigma}, a_B)$;*
- *$\mathcal{T}, \mathcal{A} \not\models A(a)$.*

We illustrate the ABox $\mathcal{A}_{\mathcal{T},\Sigma}$ using examples.

Example 4. (a) For $\mathcal{T} = \{A \equiv \exists r.A\}$ and $\Sigma = \{A, r\}$, we have $\mathcal{A}_{\mathcal{T},\Sigma} = \{r(a_A, a_A), r(a_\Sigma, a_\Sigma), A(a_\Sigma)\}$.
(b) For $\mathcal{T} = \{A \sqsubseteq \exists r.A\}$ and $\Sigma = \{A, r\}$, we have $\mathcal{A}_{\mathcal{T},\Sigma} = \{r(a_A, a_A), r(a_A, a_\Sigma), r(a_\Sigma, a_\Sigma), A(a_\Sigma)\}$.

We also require the following result about inclusions that follow from classical $\mathcal{EL}$-TBoxes [10].

Lemma 14. *Let $\mathcal{T}$ be a normal $\mathcal{EL}$ TBox, r a role name, A a primitive concept name in $\mathcal{T}$ and D an $\mathcal{EL}$-concept.*

1. *If*
$$\mathcal{T} \models \bigsqcap_{1 \leq i \leq n} A_i \sqcap \bigsqcap_{1 \leq j \leq m} \exists r_j.C_j \sqsubseteq A$$
where A_i are concept names and C_j are $\mathcal{EL}$-concepts. Then there exists A_i, $1 \leq i \leq n$, such that $\mathcal{T} \models A_i \sqsubseteq A$.
2. *Assume now*
$$\mathcal{T} \models \bigsqcap_{1 \leq i \leq n} A_i \sqcap \bigsqcap_{1 \leq j \leq m} \exists r_j.C_j \sqsubseteq \exists r.D,$$
where A_j are concept names and C_j are $\mathcal{EL}$-concepts, then
   - *there exists A_i, $1 \leq i \leq n$, such that $\mathcal{T} \models A_i \sqsubseteq \exists r.D$ or*
   - *there exists r_j such that $r_j = r$ and $\mathcal{T} \models C_j \sqsubseteq D$.*

We now describe the PTIME algorithm for deciding FO-rewritability of IQs relative to classical $\mathcal{EL}$-TBoxes in normal form. First compute the polysized ABox $\mathcal{A}_{\mathcal{T},\Sigma}$, in polytime. Define a labeling function L that assigns to every a_B, B non-conjunctive in $\mathcal{T}$, the set of concept names $L(a_B) = \{A \mid B \in \text{non-conj}_{\mathcal{T}}(A)\}$. Remove from each $L(a_B)$ all concept names A for which one of the following conditions is true:

1. there is no Σ-ABox $\mathcal{A}$ with $\mathcal{A}, \mathcal{T} \models A(a)$ (decidable in polytime [2]);
2. there is a $B' \in \text{non-conj}_{\mathcal{T}}(A)$ that is primitive in $\mathcal{T}$ and with $B'(a_B) \notin \mathcal{A}_{\mathcal{T},\Sigma}$;
3. there is a $B' \in \text{non-conj}_{\mathcal{T}}(A)$ with $B' \equiv \exists s.A' \in \mathcal{T}$ s.t. $s \notin \Sigma$ and $B'(a_B) \notin \mathcal{A}_{\mathcal{T},\Sigma}$.

Denote the resulting labeling function by L_0. Remove, recursively, from every $L_0(a_B)$ each concept name A such that

($*$) there exists $B' \in \text{non-conj}_{\mathcal{T}}(A)$ with $B'(a_B) \notin \mathcal{A}_{\mathcal{T},\Sigma}$ and $B' \equiv \exists r.A' \in \mathcal{T}$ such that there is no $B'' \in \text{non-conj}_{\mathcal{T}}(A')$ with $A' \in L_0(a_{B''})$

and denote the result by L_0^*. Clearly, L_0^* can be computed in polytime.

Lemma 15. *$A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ iff $A \notin L_0^*(a_B)$ for some $B \in$ non-conj$_{\mathcal{T}}(A)$.*

Proof. Some notation is required. For every concept name B that is non-conjunctive in $\mathcal{T}$, the unfolding of $\mathcal{A}_{\mathcal{T},\Sigma}$ at a_B up to depth k, in symbols $\mathcal{A}_B^{\leq k}$, is the following Σ-ABox: $\mathsf{Ind}(\mathcal{A}_B^{\leq k})$ consists of all paths

$$p = a_{X_0} r_1 a_{X_1} \cdots r_n a_{X_n}$$

such that $k \geq n \geq 0$, $X_0 = B$, $a_{X_i} \in \mathsf{Ind}(\mathcal{A}_{\mathcal{T},\Sigma})$ for all $0 \leq i \leq n$, and $r_{i+1}(a_{X_i}, a_{X_{i+1}}) \in \mathcal{A}_{\mathcal{T},\Sigma}$, for all $0 \leq i < n$. Now $\mathcal{A}_B^{\leq k}$ contains the assertions:

- $r(p, pra_X)$, for all $p, pra_X \in \mathsf{Ind}(\mathcal{A}_B^{\leq k})$;
- $A(p)$, for all A such that $A(\mathsf{tail}(p)) \in \mathcal{A}_{\mathcal{T},\Sigma}$ and $p \in \mathsf{Ind}(\mathcal{A}_B^{\leq k})$.

Observe that $\mathcal{T}, \mathcal{A}_B^{\leq k} \not\models A(a_B)$ whenever $\mathcal{T}, \mathcal{A}_{\mathcal{T},\Sigma} \not\models A(a_B)$ (since the function h from $\mathcal{A}_B^{\leq k}$ to $\mathcal{A}_{\mathcal{T},\Sigma}$ defined by setting $h(p) = \mathsf{tail}(p)$ is a homomorphism).

We say that $p \in \mathsf{Ind}(\mathcal{A}_B^{\leq k})$ has depth m, in symbols $\mathsf{depth}(p) = m$, if $p \in \mathsf{Ind}(\mathcal{A}_B^{\leq m}) \setminus \mathsf{Ind}(\mathcal{A}_B^{\leq m-1})$.

($\Rightarrow$) Assume that there exists $B \in \text{non-conj}_{\mathcal{T}}(A)$ with $A \in L_0^*(a_B)$. We show that A is not FO-rewritable w.r.t. $\mathcal{T}$ and Σ using Theorem 2.

Assume $k > 0$ is given. We have $\mathcal{T}, \mathcal{A}_B^{\leq k} \not\models A(a_B)$. Thus, it is sufficient to construct a Σ-ABox $\mathcal{A}_B^{k+}$ that coincides with $\mathcal{A}_B^{\leq k}$ up to depth k and such that $\mathcal{T}, \mathcal{A}_B^{k+} \models A(a_B)$. We set

$$\begin{aligned}\mathcal{A}_B^{k+} = \mathcal{A}_B^{\leq k+1} &\cup \{r(q,p) \mid \mathsf{dept}(q) = k, \mathsf{depth}(p) = k+1, r \in \Sigma\} \cup \\ &\{r(p,p) \mid \mathsf{depth}(p) = k+1, r \in \Sigma\} \cup \\ &\{A'(p) \mid \mathsf{depth}(p) = k+1, A' \in \Sigma\}\end{aligned}$$

We prove that $\mathcal{T}, \mathcal{A}_B^{k+} \models A(a_B)$. To this end, we show by induction, starting from $k+1$ downward, for every $i \leq k+1$, path p, and concept name A':

Claim 1. If $A' \in L_0^*(\mathsf{tail}(p))$ and $\mathsf{tail}(p) = B$ with $\mathsf{depth}(p) = i$, then $\mathcal{T}, \mathcal{A}_B^{k+} \models A'(p)$.

For $i = k+1$, Claim 1 follows from the condition that A' is not Σ-empty (otherwise $A' \notin L_0(\mathsf{tail}(p))$ by definition of labeling function L_0 (Condition 1)).

Assume Claim 1 is proved for $i+1$. We show Claim 1 for i.

Assume $A' \in L_0^*(\mathsf{tail}(p))$ and $\mathsf{tail}(p) = B$ with $\mathsf{depth}(p) = i$. We distinguish two cases. First assume $A' = B$. By Condition 2 for labeling function L_0, A' is not primitive. By Condition 3 for labeling function L_0, $B \equiv \exists r.E \in \mathcal{T}$ for some $r \in \Sigma$. Thus, by definition of L_0^* using elimination rule $(*)$, there exists $B'' \in \text{non-conj}_{\mathcal{T}}(E)$ with $E \in L_0^*(a_{B''})$. By IH, $\mathcal{T}, \mathcal{A}_B^{k+} \models E(pra_{B''})$. But then $\mathcal{T}, \mathcal{A}_B^{k+} \models \exists r.E(p)$ and, therefore, $\mathcal{T}, \mathcal{A}_B^{k+} \models A'(p)$, as required.

Now assume $A' \neq B$. We have $A \equiv B_1 \sqcap \cdots \sqcap B_n \in \mathcal{T}$ and $B = B_i$ for some $1 \leq i \leq n$. Let

$$X = \{B_j \mid 1 \leq j \leq n\} \setminus \{B \mid B(\mathsf{tail}(p)) \in \mathcal{A}_{\mathcal{T},\Sigma}\}$$

By Condition 2 for L_0, no member of X is primitive in $\mathcal{T}$. By Condition 3 for L_0, for all $B_j \in X$, $B_j \equiv \exists r_j.E_j \in \mathcal{T}$ for some $r_j \in \Sigma$. Thus, by $(*)$, there exist $B_j'' \in \text{conj}_{\mathcal{T}}(E_j)$ with $E_j \in L(a_{B_j''})$. By IH, $\mathcal{T}, \mathcal{A}_B^{k+} \models E_j(pra_{B_j''})$. But then $\mathcal{T}, \mathcal{A}_B^{k+} \models \exists r_j.E_j(p)$ and, therefore, $\mathcal{T}, \mathcal{A}_B^{k+} \models A'(p)$, as required. This finishes the proof of Claim 1.

For $p = a_B$ we obtain $\mathcal{T}, \mathcal{A}_B^{k+} \models A(a_B)$, as required.

($\Leftarrow$) Assume that there does not exist $B \in \text{non-conj}_{\mathcal{T}}(A)$ with $A \in L_0^*(a_B)$. We show FO-rewritability of $A(x)$ w.r.t. $\mathcal{T}$ and Σ.

We first require the following:

Claim 2. If for every $B \in \text{conj}_{\mathcal{T}}(A)$ there exists $k > 0$ such that $\mathcal{T}, \mathcal{A}_B^{k+} \not\models A(a_B)$, then $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ.

Assume $A(x)$ is not FO-rewritable relative to $\mathcal{T}$ and Σ and let $k > 0$. By Theorem 2, there exists a tree-shaped Σ-ABoxes $\mathcal{A}_k$ such that $\mathcal{T}, \mathcal{A}_k \models A(a_0)$ but $\mathcal{T}, \mathcal{A}'_k \not\models A(a_0)$ for the restriction of $\mathcal{A}'_k$ to depth k. By Lemma 13, $(\mathcal{A}'_k, a_0) \leq (\mathcal{A}_{\mathcal{T},\Sigma}, a_B)$ for some $B \in \text{non-conj}_{\mathcal{T}}(A)$. But then $(\mathcal{A}'_k, a_0) \leq (\mathcal{A}_B^{\leq k}, a_B)$ because $\mathcal{A}'_k$ has depth k, and $(\mathcal{A}_k, a_0) \leq (\mathcal{A}_B^{k+}, a_B)$, by the definition of $\mathcal{A}_B^{k+}$. We have derived a contradiction since $\mathcal{T}, \mathcal{A}_B^{k+} \models A(a_B)$ follows from Lemma 12.

Claim 3. For all A' and $B \in \text{non-conj}_{\mathcal{T}}(A')$ with $A' \not\in L_0^*(a_B)$: $\mathcal{T}, \mathcal{A}_B^{k+} \not\models A'(a_B)$ for all $k > M^2$, where M is the number of concept names in $\mathcal{T}$.

First, it follows from Lemma 14 that if $A' \not\in L_0(a_B)$ then $\mathcal{T}, \mathcal{A}_B^{1+} \not\models A'(a_B)$.

Now the proof is by induction on the number of applications of $(*)$. Assume that the claim holds for all $k > m$ for all A' and $B \in \text{non-conj}_{\mathcal{T}}(A')$ such that A' is deleted from $L_0(a_B)$ after at most m applications of $(*)$. Assume that A_0 is deleted from $L(a_{B_0})$ by the next application of $(*)$. Then there exists $B' \in \text{non-conj}_{\mathcal{T}}(A_0)$ with $B'(a_{B_0}) \not\in \mathcal{A}$ and $B' \equiv \exists r.A'' \in \mathcal{T}$ and there does not exist $B'' \in \text{non-conj}_{\mathcal{T}}(A'')$ with A'' not deleted from $L_0(a_{B''})$.

We have, by IH, $\mathcal{T}, \mathcal{A}_B^{k+} \not\models A''(a_{B''})$ for all $B'' \in \text{non-conj}_{\mathcal{T}}(A'')$, for $k > m$. But then $\mathcal{T}, \mathcal{A}_B^{k+1+} \not\models \exists r.A''(a_{B_0})$ (using Lemma 14), and so $\mathcal{T}, \mathcal{A}_B^{k+1+} \not\models B'(a_{B_0})$, for all $k > m$. Hence $\mathcal{T}, \mathcal{A}_B^{k+1+} \not\models A_0(a_{B_0})$, for all $k > m$, as required.

Claim 2 and 3 together imply FO-rewritability of $A(x)$ relative to $\mathcal{T}$ and Σ. ❑