Query-Based Entailment and Inseparability for $\mathcal{ALC}$ Ontologies (Full Version)

Elena Botoeva
Free University of Bozen-Bolzano
botoeva@inf.unibz.it

Carsten Lutz
University of Bremen
clu@uni-bremen.de

Vladislav Ryzhikov
Free University of Bozen-Bolzano
ryzhikov@inf.unibz.it

Frank Wolter
University of Liverpool
wolter@liverpool.ac.uk

Michael Zakharyaschev
Birkbeck, University of London
michael@dcs.bbk.ac.uk

Abstract

We investigate the problem whether two $\mathcal{ALC}$ knowledge bases are indistinguishable by queries over a given vocabulary. We give model-theoretic criteria and prove that this problem is undecidable for conjunctive queries (CQs) but decidable in 2EXPTIME for unions of rooted CQs. We also consider the problem whether two $\mathcal{ALC}$ TBoxes give the same answers to any query in a given vocabulary over all ABoxes, and show that for CQs this problem is undecidable, too, but becomes decidable and 2EXPTIME-complete in *Horn*-$\mathcal{ALC}$, and even EXPTIME-complete in *Horn*-$\mathcal{ALC}$ when restricted to (unions of) rooted CQs.

1 Introduction

In recent years, data access using description logic (DL) TBoxes has become one of the most important applications of DLs [28, 4], where the underlying idea is to use a TBox to specify semantics and background knowledge for the data (stored in an ABox), and thereby derive more complete query answers. A major research effort has led to the development of efficient algorithms and tools for a number of DLs ranging from DL-Lite [10, 29] via more expressive Horn DLs such as Horn-$\mathcal{ALC}$ [15, 30] to DLs with all Boolean constructors such as $\mathcal{ALC}$ [20, 32].

While query answering with DLs is now well-developed, this is much less the case for reasoning services that support ontology engineering and target query answering as an application. In ontology versioning, for example, one would like to know whether two versions of an ontology give the same answers to all queries formulated over a given vocabulary of interest, which means that the newer version can safely replace the older one [21]. Similarly, if one wants to know whether an ontology can be safely replaced by a smaller subset (module), it is the answers to all queries that should be preserved [23]. In this context, the fundamental relationship between ontologies is thus not whether they are logically equivalent (have the same models), but whether they give the same answers to any relevant query. The resulting entailment problem can be formalized in two ways, with different applications. First, given a class $\mathcal{Q}$ of queries, knowledge bases (KBs) $\mathcal{K}_1$ and

<table>
<tr><th>Queries</th><th>$\mathcal{ALC}$</th><th>Horn-$\mathcal{ALC}$ to $\mathcal{ALC}$</th><th>$\mathcal{ALC}$ to Horn-$\mathcal{ALC}$</th><th>Horn-$\mathcal{ALC}$</th></tr>
<tr><td>CQ</td><td colspan="2">undecidable</td><td rowspan="2">?</td><td rowspan="2">=EXPTIME$^{(\star)}$</td></tr>
<tr><td>UCQ</td><td colspan="2">?</td></tr>
<tr><td>rCQ</td><td colspan="2">undecidable</td><td rowspan="2">$\leq$ 2EXPTIME</td><td rowspan="2">=EXPTIME$^{(\star)}$</td></tr>
<tr><td>rUCQ</td><td colspan="2">$\leq$ 2EXPTIME</td></tr>
</table>

Figure 1: KB query entailment.

<table>
<tr><th>Queries</th><th>$\mathcal{ALC}$</th><th>Horn-$\mathcal{ALC}$ to $\mathcal{ALC}$</th><th>$\mathcal{ALC}$ to Horn-$\mathcal{ALC}$</th><th>Horn-$\mathcal{ALC}$</th></tr>
<tr><td>CQ</td><td colspan="2">undecidable</td><td rowspan="2">?</td><td rowspan="2">=2EXPTIME</td></tr>
<tr><td>UCQ</td><td colspan="2">?</td></tr>
<tr><td>rCQ</td><td colspan="2">undecidable</td><td rowspan="2">=EXPTIME</td><td rowspan="2">=EXPTIME</td></tr>
<tr><td>rUCQ</td><td colspan="2">?</td></tr>
</table>

Figure 2: TBox query entailment.

$\mathcal{K}_2$, and a signature Σ of relevant concept and role names, we say that $\mathcal{K}_1$ *Σ-$\mathcal{Q}$-entails* $\mathcal{K}_2$ if the answers to any Σ-query in $\mathcal{Q}$ over $\mathcal{K}_2$ are contained in the answers to the same query over $\mathcal{K}_1$. Further, $\mathcal{K}_1$ and $\mathcal{K}_2$ are *Σ-$\mathcal{Q}$-inseparable* if they Σ-$\mathcal{Q}$-entail each other. Note that a KB includes an ABox, and thus this notion of entailment is appropriate if the data is known and does not change frequently. Applications include data-oriented KB versioning and KB module extraction, KB forgetting [31], and knowledge exchange [1].

If the data is not known or changes frequently, it is not KBs that should be compared, but TBoxes. Given a pair $\Theta = (\Sigma_1, \Sigma_2)$ that specifies a relevant signature Σ_1 for ABoxes and Σ_2 for queries, we say that a TBox $\mathcal{T}_1$ *Θ-$\mathcal{Q}$-entails* a TBox $\mathcal{T}_2$ if, for every Σ_1-ABox $\mathcal{A}$, the KB $(\mathcal{T}_1, \mathcal{A})$ Σ_2-$\mathcal{Q}$-entails $(\mathcal{T}_2, \mathcal{A})$. $\mathcal{T}_1$ and $\mathcal{T}_2$ are *Θ-$\mathcal{Q}$-inseparable* if they Θ-$\mathcal{Q}$-entail each other. Applications include data-oriented TBox versioning, TBox modularization and TBox forgetting [23].

In this paper, we concentrate on the most important choices for $\mathcal{Q}$, conjunctive queries (CQs) and unions thereof (UCQs); we also consider the practically relevant classes of rooted CQs (rCQs) and UCQs (rUCQs), in which every variable is connected to an answer variable. So far, CQ-entailment has been studied for Horn DL KBs [9], $\mathcal{EL}$ TBoxes [26, 21], DL-Lite TBoxes [22], and also for OBDA specifications, that is, DL-Lite TBoxes with mappings [5]. No results are available for non-Horn DLs (neither in the KB nor in the TBox case) and for expressive Horn DLs in the TBox case. In particular, query entailment in non-Horn DLs has had the reputation of being a technically challenging problem.

This paper makes a first breakthrough into understanding query entailment and inseparability in these cases, with the main results summarized in Figures 1 and 2 (those marked with $(\star)$ are from [9]). Three of them came as a real surprise to us. First, it turned out that CQ- and rCQ-entailment between $\mathcal{ALC}$ KBs is undecidable, even when the first KB is formulated in *Horn*-$\mathcal{ALC}$ (in fact, $\mathcal{EL}$) and without any signature restriction. This should be contrasted with the decidability of subsumption-based entailment between $\mathcal{ALC}$ TBoxes [16] and of CQ-entailment between *Horn*-$\mathcal{ALC}$ KBs [9]. The second surprising result is that entailment between $\mathcal{ALC}$ KBs becomes decidable when CQs are replaced with rUCQs. For $\mathcal{ALC}$ TBoxes, CQ- and rCQ-entailment are undecidable as well. We obtain decidability for *Horn*-$\mathcal{ALC}$ TBoxes (where CQ- und UCQ-entailments coincide) using the fact that non-entailment is always witnessed by tree-shaped ABoxes. As another surprise, CQ-entailment of *Horn*-$\mathcal{ALC}$ TBoxes is 2EXPTIME-complete while rCQ-entailment is only EXPTIME-complete. This should be contrasted with the $\mathcal{EL}$ case, where both problems are EXPTIME-complete [26]. All upper bounds and most lower bounds hold also for inseparability in place of entailment. A model-theoretic foundation for these results is a characterization of query entailment between KBs and TBoxes in terms of (partial) homomorphisms, which, in particular, enables the use of tree automata techniques to establish the upper bounds in Figs. 1 and 2.

2 Preliminaries

Fix lists of *individual names* a_i, *concept names* A_i, and *role names* R_i, for $i < \omega$. *$\mathcal{ALC}$-concepts*, C, are defined by the grammar

$$C \ ::= \ A_i \ \mid \ \top \ \mid \ \neg C \ \mid \ C_1 \sqcap C_2 \ \mid \ \exists R_i.C.$$

We use $\bot$, $C_1 \sqcup C_2$ and $\forall R.C$ as abbreviations for $\neg\top$, $\neg(\neg C_1 \sqcap \neg C_2)$ and $\neg\exists R.\neg C$, respectively. A *concept inclusion* (CI) takes the form $C \sqsubseteq D$, where C and D are concepts. An *$\mathcal{ALC}$ TBox* is a finite set of CIs. In a *Horn-$\mathcal{ALC}$ TBox*, no concept of the form $\neg C$ occurs negatively and no $\exists R.\neg C$ occurs positively [18, 19]. An *$\mathcal{EL}$ TBox* does not contain $\neg$ at all. An *ABox*, $\mathcal{A}$, is a finite set of *assertions* of the form $A_k(a_i)$ or $R_k(a_i, a_j)$; $\mathsf{ind}(\mathcal{A})$ is the set of individual names in $\mathcal{A}$. Taken together, $\mathcal{T}$ and $\mathcal{A}$ form a *knowledge base* (KB) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$; we set $\mathsf{ind}(\mathcal{K}) = \mathsf{ind}(\mathcal{A})$.

The semantics is defined as usual based on interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ that comply with the *standard name assumption* in the sense that $a_i^{\mathcal{I}} = a_i$ [3]. We write $\mathcal{I} \models \alpha$ if an inclusion or assertion α is true in $\mathcal{I}$. If $\mathcal{I} \models \alpha$, for all $\alpha \in \mathcal{T} \cup \mathcal{A}$, then we call $\mathcal{I}$ a *model* of $\mathcal{K}$ and write $\mathcal{I} \models \mathcal{K}$. $\mathcal{K}$ is *consistent* if it has a model; we then also say that *$\mathcal{A}$ is consistent with $\mathcal{T}$*. $\mathcal{K} \models \alpha$ means that $\mathcal{I} \models \alpha$ for all $\mathcal{I} \models \mathcal{K}$.

A *conjunctive query* (CQ) $\boldsymbol{q}(\boldsymbol{x})$ is a formula $\exists \boldsymbol{y}\, \varphi(\boldsymbol{x}, \boldsymbol{y})$, where φ is a conjunction of atoms of the form $A_k(z_1)$ or $R_k(z_1, z_2)$ with z_i in $\boldsymbol{x}, \boldsymbol{y}$; the variables in $\boldsymbol{x}$ are the *answer variables* of $\boldsymbol{q}(\boldsymbol{x})$. We call $\boldsymbol{q}$ *rooted* (rCQ) if every $y \in \boldsymbol{y}$ is connected to some $x \in \boldsymbol{x}$ by a path in the graph whose nodes are the variables in $\boldsymbol{q}$ and edges are the pairs $\{u, v\}$ with $R(u, v) \in \boldsymbol{q}$, for some R. A *union of CQs* (UCQ) is a disjunction $\boldsymbol{q}(\boldsymbol{x}) = \bigvee_i \boldsymbol{q}_i(\boldsymbol{x})$ of CQs $\boldsymbol{q}_i(\boldsymbol{x})$ with the same *answer variables* $\boldsymbol{x}$; it is *rooted* (rUCQ) if all $\boldsymbol{q}_i$ are rooted.

A tuple $\boldsymbol{a}$ in $\mathsf{ind}(\mathcal{K})$ is a *certain answer to a UCQ $\boldsymbol{q}(\boldsymbol{x})$ over $\mathcal{K} = (\mathcal{T}, \mathcal{A})$* if $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for all $\mathcal{I} \models \mathcal{K}$; in this case we write $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$. If $\boldsymbol{x} = \emptyset$, the answer to $\boldsymbol{q}$ is 'yes' if $\mathcal{K} \models \boldsymbol{q}$ and 'no' otherwise. The problem of checking whether a tuple is a certain answer to a given (U)CQ over a given $\mathcal{ALC}$ KB is known to be ExpTime-complete for combined complexity [25]. The ExpTime lower bound actually holds for *Horn*-$\mathcal{ALC}$ [24].

A set $\boldsymbol{M}$ of models of a KB $\mathcal{K}$ is called *complete for* $\mathcal{K}$ if, for every UCQ $\boldsymbol{q}(\boldsymbol{x})$, we have $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$ iff $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for all $\mathcal{I} \in \boldsymbol{M}$. We call an interpretation $\mathcal{I}$ a *ditree interpretation* if the directed graph $G_\mathcal{I}$ with nodes $d \in \Delta^\mathcal{I}$ and edges $(d, e) \in R^\mathcal{I}$, for some R, is a tree and $R^\mathcal{I} \cap S^\mathcal{I} = \emptyset$, for any distinct roles R and S. $\mathcal{I}$ has *outdegree* n if $G_\mathcal{I}$ has outdegree n. A model $\mathcal{I}$ of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is *forest-shaped* if $\mathcal{I}$ is the disjoint union of ditree interpretations $\mathcal{I}_a$ with root a, for $a \in \mathsf{ind}(\mathcal{A})$, extended with all $R(a, b) \in \mathcal{A}$. The *outdegree* of $\mathcal{I}$ is the maximum outdegree of the $\mathcal{I}_a$. It is well known that the class $\boldsymbol{M}^{fo}_\mathcal{K}$ of all forest-shaped models of an $\mathcal{ALC}$ KB $\mathcal{K}$ of outdegree bounded by $|\mathcal{T}|$ is complete for $\mathcal{K}$ [25]. If $\mathcal{K}$ is a *Horn*-$\mathcal{ALC}$ KB, then a single member $\mathcal{I}_\mathcal{K}$ of $\boldsymbol{M}^{fo}_\mathcal{K}$ is complete for $\mathcal{K}$. $\mathcal{I}_\mathcal{K}$ is constructed using the standard chase procedure and called the *canonical model of* $\mathcal{K}$.

A *signature*, Σ, is a set of concept and role names. By a Σ-concept, Σ-CQ, etc. we understand any concept, CQ, etc. constructed using the names from Σ. We say that Σ is *full* if it contains all concept and role names. A model $\mathcal{I}$ of a KB $\mathcal{K}$ is Σ-*connected* if, for any $u \in \Delta^\mathcal{I} \setminus \mathsf{ind}(\mathcal{K})$, there is a path $R_1^\mathcal{I}(a, u_1), \dots, R_n^\mathcal{I}(u_n, u)$ with $a \in \mathsf{ind}(\mathcal{K})$ and the R_i in Σ.

Definition 1. Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be consistent KBs, Σ a signature, and $\mathcal{Q}$ one of CQ, rCQ, UCQ or rUCQ. We say that $\mathcal{K}_1$ Σ-$\mathcal{Q}$-*entails* $\mathcal{K}_2$ if $\mathcal{K}_2 \models \boldsymbol{q}(\boldsymbol{a})$ implies $\boldsymbol{a} \subseteq \mathsf{ind}(\mathcal{K}_1)$ and $\mathcal{K}_1 \models \boldsymbol{q}(\boldsymbol{a})$, for all Σ-$\mathcal{Q}$ $\boldsymbol{q}(\boldsymbol{x})$ and all tuples $\boldsymbol{a}$ in $\mathsf{ind}(\mathcal{K}_2)$. $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ-$\mathcal{Q}$ *inseparable* if they Σ-$\mathcal{Q}$ entail each other.

As larger classes of queries separate more KBs, Σ-UCQ inseparability implies all other inseparabilities. The following example shows that, in general, no other implications between the different notions of inseparability hold for $\mathcal{ALC}$.

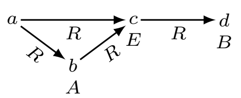
Example 2. Suppose $\mathcal{T}_0 = \emptyset$, $\mathcal{T}_0' = \{E \sqsubseteq A \sqcup B\}$ and $\Sigma_0 = \{A, B, E\}$. Let $\mathcal{A}_0 = \{E(a)\}$, $\mathcal{K}_0 = (\mathcal{T}_0, \mathcal{A}_0)$, and $\mathcal{K}_0' = (\mathcal{T}_0', \mathcal{A}_0)$. Then $\mathcal{K}_0$ and $\mathcal{K}_0'$ are Σ_0-CQ inseparable but not Σ_0-rUCQ inseparable. In fact, $\mathcal{K}_0' \models \boldsymbol{q}(a)$ and $\mathcal{K}_0 \not\models \boldsymbol{q}(a)$ for $\boldsymbol{q}(x) = A(x) \vee B(x)$.

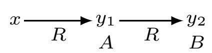
Now, let $\Sigma_1 = \{E, B\}$, $\mathcal{T}_1 = \emptyset$, and $\mathcal{T}_1' = \{E \sqsubseteq \exists R.B\}$. Let $\mathcal{A}_1 = \{E(a)\}$, $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A}_1)$, and $\mathcal{K}_1' = (\mathcal{T}_1', \mathcal{A}_1)$. Then $\mathcal{K}_1$ and $\mathcal{K}_1'$ are Σ_1-rUCQ inseparable but not Σ_1-CQ inseparable. In fact, $\mathcal{K}_1' \models \exists x B(x)$ but $\mathcal{K}_1 \not\models \exists x B(x)$.

Definition 3. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes, $\mathcal{Q}$ one of CQ, rCQ, UCQ or rUCQ, and let $\Theta = (\Sigma_1, \Sigma_2)$ be a pair of signatures. We say that $\mathcal{T}_1$ Θ-$\mathcal{Q}$ *entails* $\mathcal{T}_2$ if, for every Σ_1-ABox $\mathcal{A}$ that is consistent with both $\mathcal{T}_1$ and $\mathcal{T}_2$, the KB $(\mathcal{T}_1, \mathcal{A})$ Σ_2-$\mathcal{Q}$ entails the KB $(\mathcal{T}_2, \mathcal{A})$. $\mathcal{T}_1$ and $\mathcal{T}_2$ are Θ-$\mathcal{Q}$ *inseparable* if they Θ-$\mathcal{Q}$ entail each other. If Σ_1 is the set of all concept and role names, we say '*full ABox signature* Σ_2-$\mathcal{Q}$ *entails*' or '*full ABox signature* Σ_2-$\mathcal{Q}$ *inseparable*'.

We only consider ABoxes that are consistent with both TBoxes because the problem whether a Σ_1-ABox consistent with $\mathcal{T}_2$ is also consistent with $\mathcal{T}_1$ is well understood: it is mutually polynomially reducible with the containment problem for ontology-mediated queries with CQs of the form $\exists x A(x)$, which is NExpTime-complete for $\mathcal{ALC}$ and ExpTime-complete for *Horn*-$\mathcal{ALC}$ [6, 8].

Example 4. Consider the TBoxes $\mathcal{T}_0$ and $\mathcal{T}_0'$ from Example 2 and let $\Theta = (\Sigma, \Sigma)$ for $\Sigma = \{R, A, B, E\}$. Then $\mathcal{T}_0$ does not Θ-rCQ entail $\mathcal{T}_0'$ as $(\mathcal{T}_0', \mathcal{A}) \models \boldsymbol{q}(a)$ and $(\mathcal{T}_0, \mathcal{A}) \not\models \boldsymbol{q}(a)$ for

$\mathcal{A}$: 

$\boldsymbol{q}(x)$: 

We observe that Θ-CQ-entailment in the restricted case with $\Theta = (\Sigma, \Sigma)$ has been investigated for $\mathcal{EL}$ TBoxes by Lutz and Wolter [2010] and Konev et al. [2012].

As in the KB case, Σ-UCQ inseparability of $\mathcal{ALC}$ TBoxes implies all other types of inseparability, and Example 2 can be used to show that no other implications hold in general. The situation changes for *Horn*-$\mathcal{ALC}$ KBs and TBoxes. The following can be proved by observing that a *Horn*-$\mathcal{ALC}$ KB entails a UCQ iff it entails one of its disjuncts:

Theorem 5. *Let $\mathcal{K}_1$ be an $\mathcal{ALC}$ KB and $\mathcal{K}_2$ a* Horn-$\mathcal{ALC}$ *KB. Then $\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ iff $\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2$. The same holds for rUCQ and rCQ, and for TBox entailment.*

3 Model-Theoretic Criteria for $\mathcal{ALC}$ KBs

We now give model-theoretic criteria for Σ-entailment between KBs. The *product* $\prod \boldsymbol{\mathcal{I}}$ of a set $\boldsymbol{\mathcal{I}}$ of interpretations is defined as usual in model theory [14, page 405]. Note that, for any CQ $\boldsymbol{q}(\boldsymbol{x})$ and any tuple $\boldsymbol{a}$ of individual names, $\prod \boldsymbol{\mathcal{I}} \models \boldsymbol{q}(\boldsymbol{a})$ iff $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for each $\mathcal{I} \in \boldsymbol{\mathcal{I}}$.

Suppose $\mathcal{I}_i$ is an interpretation for a KB $\mathcal{K}_i$, $i = 1, 2$. A function $h \colon \Delta^{\mathcal{I}_2} \to \Delta^{\mathcal{I}_1}$ is called a Σ-*homomorphism* if $u \in A^{\mathcal{I}_2}$ implies $h(u) \in A^{\mathcal{I}_1}$ and $(u, v) \in R^{\mathcal{I}_2}$ implies $(h(u), h(v)) \in R^{\mathcal{I}_1}$ for all $u, v \in \Delta^{\mathcal{I}_2}$, Σ-concept names A, and Σ-role names R, and $h(a) = a$ for all $a \in \mathsf{ind}(\mathcal{K}_2)$. It is known from database theory that homomorphisms characterize CQ-containment [12]. For KB Σ-query entailment, finite partial homomorphisms are required. We say that $\mathcal{I}_2$ is $n\Sigma$-*homomorphically embeddable into* $\mathcal{I}_1$ if, for any subinterpretation $\mathcal{I}_2'$ of $\mathcal{I}_2$ with $|\Delta^{\mathcal{I}_2'}| \leq n$, there is a Σ-homomorphism from $\mathcal{I}_2'$ to $\mathcal{I}_1$. If, additionally, we require $\mathcal{I}_2'$ to be Σ-connected then $\mathcal{I}_2$ is said to be *con-*$n\Sigma$-*homomorphically embeddable into* $\mathcal{I}_1$.

Theorem 6. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be $\mathcal{ALC}$ KBs, Σ a signature, and let $\boldsymbol{M}_i$ be complete for $\mathcal{K}_i$, $i = 1, 2$.*

(1) *$\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ iff, for any $n > 0$ and $\mathcal{I}_1 \in \boldsymbol{M}_1$, there exists $\mathcal{I}_2 \in \boldsymbol{M}_2$ that is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$.*

(2) *$\mathcal{K}_1$ Σ-rUCQ entails $\mathcal{K}_2$ iff, for any $n > 0$ and $\mathcal{I}_1 \in \boldsymbol{M}_1$, there exists $\mathcal{I}_2 \in \boldsymbol{M}_2$ that is con-$n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$.*

(3) *$\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2$ iff $\prod \boldsymbol{M}_2$ is $n\Sigma$-homomorphically embeddable into $\prod \boldsymbol{M}_1$ for any $n > 0$.*

(4) *$\mathcal{K}_1$ Σ-rCQ entails $\mathcal{K}_2$ iff $\prod \boldsymbol{M}_2$ is con-$n\Sigma$-homomorphically embeddable into $\prod \boldsymbol{M}_1$ for any $n > 0$.*

Proof. We only show (1). Suppose $\mathcal{K}_2 \models \boldsymbol{q}$ but $\mathcal{K}_1 \not\models \boldsymbol{q}$. Let n be the number of variables in $\boldsymbol{q}$. Take $\mathcal{I}_1 \in \boldsymbol{M}_1$ such that $\mathcal{I}_1 \not\models \boldsymbol{q}$. Then no $\mathcal{I}_2 \in \boldsymbol{M}_2$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$. Conversely, suppose $\mathcal{I}_1 \in \boldsymbol{M}_1$ is such that, for some n, no $\mathcal{I}_2 \in \boldsymbol{M}_2$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$. We can regard any subinterpretation of any $\mathcal{I}_2 \in \boldsymbol{M}_2$ with domain of size $\leq n$ as a CQ (with answer variable corresponding to ABox individuals). The disjunction of all such CQs is entailed by $\mathcal{K}_2$ but not by $\mathcal{K}_1$. □

Note that $n\Sigma$-homomorphic embeddability cannot be replaced by Σ-homomorphic embeddability. For example, in (1), let $\mathcal{K}_1 = \mathcal{K}_2 = (\{\top \sqsubseteq \exists R.\top\}, \{A(a)\})$, $\boldsymbol{M}_1 = \{\mathcal{I}_1\}$, where $\mathcal{I}_1$ is the infinite R-chain starting with a, and let $\boldsymbol{M}_2$ contain arbitrary finite R-chains starting with a followed by an arbitrary long R-cycle. $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are both complete for $\mathcal{K}$, but there is no Σ-homomorphism from any $\mathcal{I}_2 \in \boldsymbol{M}_2$ to $\mathcal{I}_1$. In Section 5, we show that in some cases we *can* find characterizations with full Σ-homomorphisms and use them to present decision procedures for entailment.

If both $\boldsymbol{M}_i$ are finite and contain only finite interpretations, then Theorem 6 provides a decision procedure for KB entailment. This applies, for example, to KBs with acyclic classical TBoxes [3], and to KBs for which the chase terminates [17].

4 Undecidability for $\mathcal{ALC}$ KBs and TBoxes

We show that CQ and rCQ-entailment and inseparability for $\mathcal{ALC}$ KBs are undecidable—even if the signature is full and $\mathcal{K}_1$ is a Horn-$\mathcal{ALC}$ (in fact, $\mathcal{EL}$) KB. We establish the same results for TBoxes except that in the rCQ case, we leave it open whether the full ABox signature is sufficient for undecidability.

Theorem 7. (i) *The problem whether a* Horn-$\mathcal{ALC}$ *KB Σ-$\mathcal{Q}$ entails an $\mathcal{ALC}$ KB is undecidable for $\mathcal{Q} \in \{CQ, rCQ\}$.*

(ii) *Σ-$\mathcal{Q}$ inseparability between* Horn-$\mathcal{ALC}$ *and $\mathcal{ALC}$ KBs is undecidable for $\mathcal{Q} \in \{CQ, rCQ\}$.*

(iii) *Both (i) and (ii) hold for the full signature Σ.*

Proof. The proof is by reduction of the undecidable $N \times M$-*tiling problem*: given a finite set $\mathfrak{T}$ of *tile types* T with four colours *up*(T), *down*(T), *left*(T) and *right*(T), a tile type $I \in \mathfrak{T}$, and two colours W (for wall) and C (for ceiling), decide whether there exist $N, M \in \mathbb{N}$ such that the $N \times M$ grid can be tiled using $\mathfrak{T}$ in such a way that $(1, 1)$ is covered by a tile of type I; every (N, i), for $i \leq M$, is covered by a tile of type T with *right*$(T) = W$; and every (i, M), for $i \leq N$, is covered by a tile of type T with *up*$(T) = C$.

Given an instance of this problem, we first describe a KB $\mathcal{K}_2 = (\mathcal{T}_2, \{A(a)\})$ that uses (among others) 3 concept names T_k, $k = 0, 1, 2$, for each tile type $T \in \mathfrak{T}$. If a point x in a model $\mathcal{I}$ of $\mathcal{K}_2$ is in T_k and *right*$(T) =$ *left*(T'), then x has an R-successor in T_k'. Thus, branches of $\mathcal{I}$ define (possibly infinite) horizontal

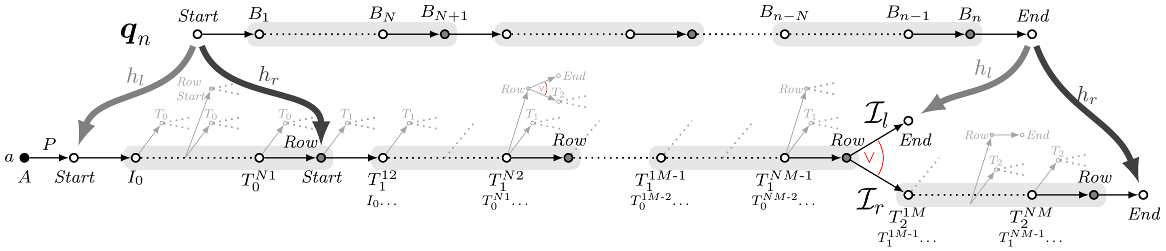


Figure 3: The structure of models $\mathcal{I}_l$ and $\mathcal{I}_r$ of $\mathcal{K}_2$, and homomorphisms $h_l\colon \boldsymbol{q}_n \to \mathcal{I}_l$ and $h_r\colon \boldsymbol{q}_n \to \mathcal{I}_r$.

rows of tilings with $\mathfrak{T}$. If a branch contains a point $y \in T_k$ with *right*$(T) = W$, then this y can be the last point in the row, which is indicated by an R-successor $z \in$ *Row* of y. In turn, z has R-successors in all $T_{(k+1) \bmod 3}$ that can be possible beginnings of the next row of tiles. To coordinate the *up* and *down* colours between the rows—which will be done by the CQs separating $\mathcal{K}_1$ and $\mathcal{K}_2$— we make every $x \in T_k$, starting from the second row, an instance of all $T'_{(k-1) \bmod 3}$ with *down*$(T) =$ *up*(T'). The row started by $z \in$ *Row* can be the last one in the tiling, in which case we require that each of its tiles T has *up*$(T) = C$. After the point in *Row* indicating the end of the final row, we add an R-successor in *End* for the end of tiling. The beginning of the first row is indicated by a P-successor in *Start* of the ABox element a, after which we add an R-successor in I_0 for the given initial tile type I; see the lowest branch in Fig. 3. To generate a tree with all possible branches described above, we only require $\mathcal{EL}$ axioms of the form $E \sqsubseteq D$ and $E \sqsubseteq \exists S.D$.

The existence of a tiling of some $N \times M$ grid for the given instance can be checked by Boolean CQs $\boldsymbol{q}_n$ that require an R-path from *Start* to *End* going through T_k- or *Row*-points:

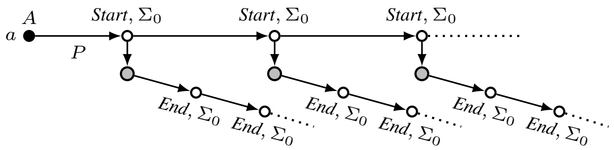
$$\exists \boldsymbol{x}\big(\mathit{Start}(x_0) \wedge \bigwedge_{i=0}^{n} R(x_i, x_{i+1}) \wedge \bigwedge_{i=1}^{n} B_i(x_i) \wedge \mathit{End}(x_{n+1})\big)$$

with $B_i \in \{\mathit{Row}\} \cup \{T_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$; see Fig. 3. The key trick is—using an axiom of the form $D \sqsubseteq E \sqcup E'$—to ensure that the *Row*-point before the final row of the tiling has *two alternative* continuations: one as described above, and the other one having just a single R-successor in *End*; see Fig. 3 where $\vee$ indicates an *or-node*. This or-node gives two models of $\mathcal{K}_2$ denoted $\mathcal{I}_l$ and $\mathcal{I}_r$ in the picture. If $\mathcal{K}_2 \models \boldsymbol{q}_n$, then $\boldsymbol{q}_n$ holds in both of them, and so there are homomorphisms $h_l\colon \boldsymbol{q}_n \to \mathcal{I}_l$ and $h_r\colon \boldsymbol{q}_n \to \mathcal{I}_r$. As $h_l(x_{n-1})$ and $h_r(x_{n-1})$ are instances of B_{n-1}, we have $B_{n-1} = T_1^{NM-1}$ in the picture, and so *up*$(T^{NM-1}) =$ *down*(T^{NM}). By repeating this argument until x_0, we see that the colours between horizontal rows match and the rows are of the same length. (For this trick to work, we have to make the first *Row*-point in every branch an instance of *Start*.) In fact, we have:

Lemma 8. *An instance of the $N \times M$-tiling problem has a positive answer iff there exists $\boldsymbol{q}_n$ such that $\mathcal{K}_2 \models \boldsymbol{q}_n$.*

It is to be noted that to construct $\mathcal{T}_2$ with the properties described above one needs quite a few auxiliary concept names.

Next, we define $\mathcal{K}_1 = (\mathcal{T}_1, \{A(a)\})$ to be the $\mathcal{EL}$ KB with the following canonical model:



where $\Sigma_0 = \{\mathit{Row}\} \cup \{T_k \mid T \in \mathfrak{T},\, k = 0, 1, 2\}$. Note that the vertical R-successors of the *Start*-points are not instances of any concept name, and so $\mathcal{K}_1$ does not satisfy any query $\boldsymbol{q}_n$. On the other hand, $\mathcal{K}_2 \models \boldsymbol{q}$ implies $\mathcal{K}_1 \models \boldsymbol{q}$, for every Σ-CQ $\boldsymbol{q}$ without a subquery of the form $\boldsymbol{q}_n$ and $\Sigma = \mathsf{sig}(\mathcal{K}_1)$.

This proves (*i*) for Σ-CQ entailment. For Σ-rCQ entailment, we slightly modify the construction, in particular, by adding $R(a, a)$ and *Row*(a) to the ABox $\{A(a)\}$, and a conjunct $R(y, x_0)$ with a free y to $\boldsymbol{q}_n$. (The loop $R(a, a)$ plays roughly the same role as the path between two *Start*-points in Fig. 3.) To prove (*ii*), we take $\mathcal{K}_2' = \mathcal{K}_2 \cup \mathcal{K}_1$ and show that $\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2$ iff $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ-CQ inseparable. Finally, we prove (*iii*) by replacing non-Σ symbols in $\mathcal{K}_2$ with complex $\mathcal{ALC}$-concepts that cannot be used in CQs and extending the TBoxes appropriately; cf. [27, Lemma 21]. □

The TBoxes from the proof above can also be used to obtain

Theorem 9. (*i*) *The problem whether a* Horn-$\mathcal{ALC}$ *TBox* Θ-$\mathcal{Q}$ *entails an* $\mathcal{ALC}$ *TBox is undecidable for* $\mathcal{Q} \in \{CQ, rCQ\}$.
(*ii*) Θ-$\mathcal{Q}$ *inseparability between* Horn-$\mathcal{ALC}$ *and* $\mathcal{ALC}$ *TBoxes is undecidable for* $\mathcal{Q} \in \{CQ, rCQ\}$.
(*iii*) *For CQs,* (*i*) *and* (*ii*) *hold for full ABox signatures and for* $\Theta = (\Sigma_1, \Sigma_2)$ *with* $\Sigma_1 = \Sigma_2$.

Observe that our undecidability proof does not work for UCQs as the UCQ composed of the two disjunctive branches shown in Fig. 3 (for non-trivial instances) distinguishes between the KBs independently from the existence of tiling. We now show that, at least for rUCQs, entailment is decidable.

5 rUCQ-Entailment for $\mathcal{ALC}$-KBs

Theorem 7 might seem to suggest that any reasonable notion of query inseparability is undecidable for $\mathcal{ALC}$ KBs. Interestingly, this is not the case: we show now that rUCQ-entailment is decidable. We first strengthen the characterization of Theorem 6 (2), and then develop a decision procedure based on tree automata. The first step replaces con-nΣ-homomorphic embeddability with con-Σ-homomorphic embeddability, where $\mathcal{I}_2$ is *con-Σ-homomorphically embeddable into* $\mathcal{I}_1$ if the maximal Σ-connected subinterpretation of $\mathcal{I}_2$ is Σ-homomorphically embeddable into $\mathcal{I}_1$.

Theorem 10. *Let* $\mathcal{K}_1$ *and* $\mathcal{K}_2$ *be* $\mathcal{ALC}$ *KBs,* Σ *a signature, and let* $\boldsymbol{M}_1$ *be complete for* $\mathcal{K}_1$*. Then* $\mathcal{K}_1$ Σ*-rUCQ entails* $\mathcal{K}_2$ *iff for any* $\mathcal{I}_1 \in \boldsymbol{M}_1$*, there exists* $\mathcal{I}_2 \models \mathcal{K}_2$ *such that* $\mathcal{I}_2$ *is con-Σ-homomorphically embeddable into* $\mathcal{I}_1$.

Proof. In view of Theorem 6 (2), it suffices to prove ($\Rightarrow$). Suppose $\mathcal{I}_1 \in \boldsymbol{M}_1$. By Theorem 6 (2), for every $n \geq 0$, we have $\mathcal{J} \in \boldsymbol{M}^{fo}_{\mathcal{K}_2}$ and a Σ-homomorphism $h_n \colon \mathcal{J}_{|\leq n} \to \mathcal{I}_1$, where $\mathcal{J}_{|\leq n}$ is the subinterpretation of $\mathcal{J}$ whose elements are connected to ABox individuals by Σ-paths of length $\leq n$. Clearly, for any $n \geq 0$, there are only finitely many non-isomorphic pairs $(\mathcal{J}_{|\leq n}, h_n)$. It can be shown that, thus, one can construct the required $\mathcal{I}_2 \in \boldsymbol{M}^{fo}_{\mathcal{K}_2}$ and con-Σ-homomorphism h as the limits of suitable chains $\mathcal{J}_{|\leq 0} \subseteq \mathcal{J}_{|\leq 1} \subseteq \cdots$ and $h_0 \subseteq h_1 \subseteq \cdots$, respectively. □

For the second step, let $\mathcal{K}_1, \mathcal{K}_2$ be $\mathcal{ALC}$-KBs and Σ a signature. We use two-way alternating automata on infinite trees (2ATAs) with a trivial acceptance condition (every run is accepting) and employ Theorem 10 for the class $\boldsymbol{M}^{fo}_{\mathcal{K}_1}$ and encode forest-shaped interpretations as labeled trees to make them accessible to 2ATAs. A *tree* is a non-empty (possibly infinite) set $T \subseteq \mathbb{N}^*$ closed under prefixes with root ε. We say that T is m*-ary* if, for every $x \in T$, the set $\{i \mid x \cdot i \in T\}$ is of cardinality m. Let Γ be an alphabet with symbols from the set

$$\{root, empty\} \cup (\mathsf{ind}(\mathcal{K}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}) \cup (\mathsf{RN}(\mathcal{T}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}),$$

where $\mathsf{CN}(\mathcal{T}_i)$ (resp. $\mathsf{RN}(\mathcal{T}_i)$) denotes the set of concept (resp. role) names in $\mathcal{T}_i$. A Γ*-labeled tree* is a pair (T, L) with T a tree and $L \colon T \to \Gamma$ a node labeling function. We represent forest-shaped models of $\mathcal{T}_1$ as m-ary Γ-labeled trees, with $m = \max(|\mathcal{T}_1|, |\mathsf{ind}(\mathcal{K}_1)|)$. The root node labeled with $root$ is not used in the representation. Each ABox individual is represented by a successor of the root labeled with a symbol from $\mathsf{ind}(\mathcal{K}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}$; non-ABox elements are represented by nodes deeper in the tree labeled with a symbol from $\mathsf{RN}(\mathcal{T}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}$. The label $empty$ is used for padding to make sure that every tree node has exactly m successors.

Now we construct three 2ATAs $\mathfrak{A}_i$, for $i = 0, 1, 2$. $\mathfrak{A}_0$ ensures that the tree is labeled in a meaningful way, e.g. that the *root* label only occurs at the root node; $\mathfrak{A}_1$ accepts Γ-labeled trees that represent a model of $\mathcal{K}_1$, and $\mathfrak{A}_2$ accepts Γ-labeled trees (T, L) which represent an interpretation $\mathcal{I}_{(T,L)}$ such that some model of $\mathcal{K}_2$ is con-Σ-homomorphically embeddable into $\mathcal{I}_{(T,L)}$. The most interesting automaton is $\mathfrak{A}_2$, which guesses a model of $\mathcal{K}_2$ along with a homomorphism to $\mathcal{I}_{(T,L)}$; in fact, both can be read off from a successful run of the automaton. The number of states of the $\mathfrak{A}_i$ is exponential in $|\mathcal{K}_1 \cup \mathcal{K}_2|$. It then remains to combine these automata into a single 2ATA $\mathfrak{A}$ such that $\mathcal{L}(\mathfrak{A}) = \mathcal{L}(\mathfrak{A}_0) \cap \mathcal{L}(\mathfrak{A}_1) \cap \overline{\mathcal{L}(\mathfrak{A}_2)}$, which is possible with only polynomial blowup, and to test (in time exponential in the number of states) whether $\mathcal{L}(\mathfrak{A}) = \emptyset$.

Theorem 11. *It is in* 2ExpTime *to decide whether an* $\mathcal{ALC}$ *KB* $\mathcal{K}_1$ Σ*-rUCQ entails an* $\mathcal{ALC}$ *KB* $\mathcal{K}_2$.

The best known lower bound is ExpTime, which is easy to establish by reduction from satisfiability.

6 (r)CQ-Entailment for (Horn-)$\mathcal{ALC}$-TBoxes

We show that CQ- and rCQ-entailment between $\mathcal{ALC}$ TBoxes becomes decidable when the second TBox is given in Horn-$\mathcal{ALC}$. In this case, entailments for CQs and UCQs and, respectively, rCQs and rUCQs coincide. We start with rCQs.

Our first observation is that if a Σ_1-ABox is a witness for non-Θ-rCQ entailment, then one can find a witness Σ_1-ABox that is tree-shaped and of bounded outdegree. Here, an ABox $\mathcal{A}$ is *tree-shaped* if the graph with nodes $\mathsf{ind}(\mathcal{A})$ and edges $\{a, b\}$ for each $R(a, b) \in \mathcal{A}$ is a tree, and $R(a, b) \in \mathcal{A}$ implies $S(a, b) \notin \mathcal{A}$ for all $S \neq R$ and $S(b, a) \notin \mathcal{A}$ for all S.

Theorem 12. *Let $\mathcal{T}_1$ be an $\mathcal{ALC}$ TBox, $\mathcal{T}_2$ a Horn-$\mathcal{ALC}$ TBox, and $\Theta = (\Sigma_1, \Sigma_2)$. Then $\mathcal{T}_1$ Θ-rCQ-entails $\mathcal{T}_2$ iff, for all tree-shaped Σ_1-ABoxes $\mathcal{A}$ of outdegree bounded by $|\mathcal{T}_2|$ and consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is con-Σ_2-homomorphically embeddable into any model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$.*

Proof. It is known that Horn-$\mathcal{ALC}$ is unravelling tolerant, that is, $(\mathcal{T}, \mathcal{A}) \models C(a)$ for a Horn-$\mathcal{ALC}$ TBox $\mathcal{T}$ and $\mathcal{EL}$-concept C iff $(\mathcal{T}, \mathcal{A}') \models C(a)$ for a finite sub-ABox $\mathcal{A}'$ of the tree-unravelling of $\mathcal{A}$ at a [27]. Thus, any witness ABox for non-entailment w.r.t. $\mathcal{EL}$-instance queries can be transformed into a tree-shaped witness ABox. The result follows by observing that if $\mathcal{T}_1$ does not Θ-rCQ-entail $\mathcal{T}_2$, then this is witnessed by an $\mathcal{EL}$-instance query and by applying Theorem 10 to the KBs. The bound on the outdegree is obtained by a careful analysis of derivations. □

For the automaton construction, let $\mathcal{T}_1$ be an $\mathcal{ALC}$ TBox, $\mathcal{T}_2$ a Horn-$\mathcal{ALC}$ TBox, and $\Theta = (\Sigma_1, \Sigma_2)$ a pair of signatures. Though Theorem 12 provides a natural characterization that is similar in spirit to Theorem 10, we first need a further analysis of con-Σ_2-homomorphic embeddability in terms of simulations whose advantage is that they are more compositional (they can be partial and are closed under union).

Let $\mathcal{I}_1, \mathcal{I}_2$ be interpretations and Σ a signature. A relation $\mathcal{S} \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is a Σ-*simulation* from $\mathcal{I}_1$ to $\mathcal{I}_2$ if (*i*) $d \in A^{\mathcal{I}_1}$ and $(d, d') \in \mathcal{S}$ imply $d' \in A^{\mathcal{I}_2}$ for all Σ-concept names A, and (*ii*) if $(d, e) \in R^{\mathcal{I}_1}$ and $(d, d') \in \mathcal{S}$ then there is a $(d', e') \in R^{\mathcal{I}_2}$ with $(e, e') \in \mathcal{S}$ for all Σ-role names R. Let $d_i \in \Delta^{\mathcal{I}_i}$, $i \in \{1, 2\}$. $(\mathcal{I}_1, d_1)$ is Σ-simulated by $(\mathcal{I}_2, d_2)$, in symbols $(\mathcal{I}_1, d_1) \leq_\Sigma (\mathcal{I}_2, d_2)$, if there exists a Σ-simulation $\mathcal{S}$ with $(d_1, d_2) \in \mathcal{S}$.

Lemma 13. *Let $\mathcal{A}$ be a Σ_1-ABox and $\mathcal{I}_1$ a model of $(\mathcal{T}_1, \mathcal{A})$. Then $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$ iff there is $a \in \mathsf{ind}(\mathcal{A})$ such that one of the following holds:*

(1) *there is a Σ_2-concept name A with $a \in A^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}} \setminus A^{\mathcal{I}_1}$;*

(2) *there is an R-successor d of a in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, for some Σ_2-role name R, such that $d \notin \mathsf{ind}(\mathcal{A})$ and, for all R-successors e of a in $\mathcal{I}_1$, we have $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d) \not\leq_{\Sigma_2} (\mathcal{I}_1, e)$.*

We use a mix of two-way alternating Büchi automata on finite trees (2ABTAs) and non-deterministic top-down automata on finite trees (NTAs). A finite tree T is m-*ary* if, for every $x \in T$, the set $\{i \mid x \cdot i \in T\}$ is of cardinality zero or exactly m. We use labeled trees to represent a tree-shaped ABox $\mathcal{A}$ and a model $\mathcal{I}_1$ such that, for some $a \in \mathsf{ind}(\mathcal{A})$, conditions (1) and (2) from Lemma 13 are satisfied, and thus $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$. To ensure that later, additional bookkeeping information is needed. Node labels are taken from the alphabet

$$\Gamma = \Gamma_0 \times 2^{\mathsf{cl}(\mathcal{T}_1)} \times 2^{\mathsf{CN}(\mathcal{T}_2)} \times \{0, 1\} \times 2^{\mathsf{sub}(\mathcal{T}_2)},$$

where Γ_0 is the set of all subsets of $\Sigma_1 \cup \{R^- \mid R \in \Sigma_1\}$ that contain at most one role (a role name R or its inverse R^-), $\mathsf{cl}(\mathcal{T}_i)$ is the set of subconcepts of (concepts in) $\mathcal{T}_i$ closed under single negation, and $\mathsf{sub}(\mathcal{T}_2)$ is the set of subconcepts of (concepts in) $\mathcal{T}_2$. For a Γ-labeled tree (T, L) and a node x from T, we use $L_i(x)$ to denote the $(i+1)$st component of $L(x)$, where $i \in \{0, \ldots, 4\}$. Intuitively, the L_0-component represents the ABox $\mathcal{A}$, the L_1-component the model $\mathcal{I}_1$, the L_2-component represents $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, and the L_3- and L_4-components help to guarantee conditions (1) and (2) from Lemma 13.

To ensure that each component $i \in \{0, \ldots, 4\}$ indeed represents what it is supposed to, we impose on it an i-*properness* condition. For example, a Γ-labeled (T, L) tree is *0-proper* if (i) $L_0(\varepsilon)$ contains no role and (ii) for every non-root node x of T, $L_0(x)$ contains a role. A 0-proper Γ-labeled tree (T, L) represents the following tree-shaped Σ_1-ABox:

$$\begin{aligned}\mathcal{A}_{(T,L)} = \; & \{A(x) \mid A \in L_0(x)\} \cup \\ & \{R(x, y) \mid R \in L_0(y),\ y \text{ is a child of } x\} \cup \\ & \{R(y, x) \mid R^- \in L_0(y),\ y \text{ is a child of } x\}.\end{aligned}$$

Due to space limitations, we skip the remaining definitions of properness and concentrate on explaining the most interesting components L_3 and L_4 of Γ-labels. The L_3-component marks a single node x in the tree, which is the individual a from Lemma 13 that satisfies conditions (1) and (2). If (1) is satisfied, we do not need the L_4-component. Otherwise, we store in that component at x a set of concepts $S = \{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$ such that $R \in \Sigma_2$ and all concepts from S are true at x in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$. This *successor set* represents the R-successor d in condition (2) of Lemma 13. We then have to make sure that, for any neighboring node y of x that represents an R-successor of x in $\mathcal{A}_{(T,L)}$, we have $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d) \not\leq_{\Sigma_2} (\mathcal{I}_1, y)$. This can again happen via a concept name or via a successor; we are done in the fomer case and use the L_4-component of y in the latter. It is important to note that we can never return to the same node in this tracing process since we only follow roles in the forward direction and the represented ABox is tree-shaped. This is crucial for achieving the EXPTIME overall complexity.

We show that $\mathcal{T}_2$ is not Θ-rCQ-entailed by $\mathcal{T}_1$ iff there is an m-ary Γ-labeled tree that is i-proper for any $i \in \{0, \ldots, 4\}$. It then remains to design a 2ABTA $\mathfrak{A}$ that accepts exactly those trees. We construct $\mathfrak{A}$ as the intersection of five automata $\mathfrak{A}_i$, $i < 5$, where each $\mathfrak{A}_i$ ensures i-properness. Some of the automata are 2ABTAs with polynomially many states while others are NTAs with exponentially many states. We mix automata models since some properness conditions (2-properness) are much easier to describe with a 2ABTA while for others (4-properness), it does not seem to be possible to construct a 2ABTA with polynomially many states. In summary, we obtain the following result.

Theorem 14. *It is* EXPTIME*-complete to decide whether an $\mathcal{ALC}$ TBox $\mathcal{T}_1$ (Σ_1, Σ_2)-rCQ entails a Horn-$\mathcal{ALC}$ TBox $\mathcal{T}_2$.*

Note that the EXPTIME lower bound holds already for entailment of $\mathcal{EL}$ TBoxes and $\Sigma_1 = \Sigma_2$ [26]. We now study the non-rooted case, starting with an analogue of Theorem 12. As expected, moving to unrestricted queries corresponds to moving to unrestricted homomorphisms.

Theorem 15. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALC}$ TBoxes and $\Theta = (\Sigma_1, \Sigma_2)$. Then $\mathcal{T}_1$ Θ-CQ entails $\mathcal{T}_2$ iff, for all tree-shaped Σ_1-ABoxes $\mathcal{A}$ of outdegree $\leq |\mathcal{T}_2|$ and consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is Σ_2-homomorphically embeddable into $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.*

The automata construction described above can largely be reused for this case. The main difference is that the two conditions in Lemma 13 need to be extended with a third one: there is an element d in the subtree of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ rooted at a that has an R-successor d_0, $R \notin \Sigma_2$, such that, for all elements e of $\mathcal{I}_1$, we have $(\mathcal{I}_2, d_0) \not\leq_{\Sigma_2} (\mathcal{I}_1, e)$. To deal with this condition, it becomes necessary to store multiple successor sets in the L_4-components instead of only a single one, which increases the overall complexity to 2EXPTIME. A matching lower bound can be proved by a (non-trivial) reduction of the word problem for exponentially bounded alternating Turing machines.

Theorem 16. *Θ-CQ entailment for Horn-$\mathcal{ALC}$ TBoxes is 2*EXPTIME*-complete. The lower bound holds for $\Theta = (\Sigma, \Sigma)$.*

7 Future Work

We have made first steps towards understanding query entailment and inseparability for KBs and TBoxes in expressive DLs. Many problems remain to be addressed. From a theoretical viewpoint, it would be of interest to solve the open problems in Figures 1 and 2, and also consider other expressive DLs such as *DL-Lite*$^{\mathcal{H}}_{bool}$ [2] or $\mathcal{ALCI}$. For example, if Theorem 10 could be generalized to UCQs (and Σ-homomorphisms), we would obtain a 2EXPTIME upper bound for UCQ-entailment between $\mathcal{ALC}$ KBs using the same technique as for rUCQs. Also, our undecidability proof goes through for *DL-Lite*$^{\mathcal{H}}_{bool}$, but the other cases remain open. From a practical viewpoint, our model-theoretic criteria for query entailment are a good starting point for developing algorithms for approximations of query entailment based on simulations. Our undecidability and complexity results also indicate that rUCQ-entailment is more amenable to practical algorithms than, say, CQ-entailment and can be used as an approximation of the latter.

Acknowledgments. This work has been supported by the EU IP project Optique, grant n. FP7-318338, ERC grant 647289 (CODA), and EPSRC UK grants EP/M012646/1 and EP/M012670/1 (iTract).

References

[1] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Exchanging OWL 2 QL knowledge bases. In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI 2013)*. AAAI Press, 2013. ISBN 978-1-57735-633-2.

[2] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The DL-Lite family and relations. *Journal of Artificial Intelligence Research (JAIR)*, 36:1–69, 2009.

[3] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[4] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *11th Reasoning Web Int. Summer School Tutorial Lectures (RW 2015)*, pages 218–307, 2015. doi: 10.1007/978-3-319-21768-0_9.

[5] Meghyn Bienvenu and Riccardo Rosati. Query-based comparison of OBDA specifications. In *Proc. of the 28th Int. Workshop on Description Logics (DL 2015)*, volume 1350. CEUR-WS, 2015.

[6] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Query containment in description logics reconsidered. In *Proc. of the 13th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2012)*, pages 221–231, 2012.

[7] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. First-order rewritability of atomic queries in horn description logics. In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI 2013)*, 2013.

[8] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through Disjunctive Datalog, CSP, and MMSNP. *ACM Trans. on Database Systems*, 39(4):33:1–33:44, 2014. doi: 10.1145/2661643.

[9] Elena Botoeva, Roman Kontchakov, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Query inseparability for description logic knowledge bases. In *Proc. of the 14th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2014)*, pages 238–247, 2014.

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

[11] Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. Answering regular path queries in expressive description logics: An automata-theoretic approach. In *Proc. of the 22nd Nat. Conf. on Artificial Intelligence (AAAI)*, pages 391–396, 2007.

[12] Ashok K. Chandra and Philip M. Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Proc. of the 9th ACM Symp. on Theory of Computing (STOC)*, pages 77–90, 1977. doi: 10.1145/800105.803397.

[13] Ashok K. Chandra, Dexter Kozen, and Larry J. Stockmeyer. Alternation. *J. ACM*, 28(1):114–133, 1981.

[14] C.C. Chang and H.J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1990.

[15] Thomas Eiter, Magdalena Ortiz, Mantas Simkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for Horn-SHIQ plus rules. In *Proc. of the 26th Nat. Conf. on Artificial Intelligence (AAAI 2012)*, pages 726–733. AAAI Press, 2012.

[16] S. Ghilardi, C. Lutz, and F. Wolter. Did I damage my ontology? A case for conservative extensions in description logics. In Patrick Doherty, John Mylopoulos, and Christopher Welty, editors, *Proc. of the 10th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2006)*, pages 187–197, 2006.

[17] Bernardo Cuenca Grau, Ian Horrocks, Markus Krötzsch, Clemens Kupke, Despoina Magka, Boris Motik, and Zhe Wang. Acyclicity notions for existential rules and their application to query answering in ontologies. *Journal of Artificial Intelligence Research (JAIR)*, 47:741–808, 2013.

[18] Ulrich Hustadt, Boris Motik, and Ulrike Sattler. Data complexity of reasoning in very expressive description logics. In *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005)*, pages 466–471, 2005.

[19] Yevgeny Kazakov. Consequence-driven reasoning for Horn SHIQ ontologies. In *Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI 2009)*, pages 2040–2045, 2009.

[20] Ilianna Kollia and Birte Glimm. Optimizing SPARQL query answering over OWL ontologies. *Journal of Artificial Intelligence Research (JAIR)*, 48:253–303, 2013. doi: 10.1613/jair.3872. URL `http://dx.doi.org/10.1613/jair.3872`.

[21] Boris Konev, Michel Ludwig, Dirk Walther, and Frank Wolter. The logical difference for the lightweight description logic EL. *Journal of Artificial Intelligence Research (JAIR)*, 44:633–708, 2012.

[22] R. Kontchakov, L. Pulina, U. Sattler, T. Schneider, P. Seimer, F. Wolter, and M. Zakharyaschev. Minimal module extraction from *DL-Lite* ontologies using QBF solvers. In *Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI 2009)*, pages 836–840, 2009.

[23] Roman Kontchakov, Frank Wolter, and Michael Zakharyaschev. Logic-based ontology comparison and module extraction, with an application to DL-Lite. *Artificial Intelligence*, 174:1093–1141, 2010.

[24] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Complexities of Horn description logics. *ACM Trans. on Computational Logic*, 14(1):2, 2013.

[25] Carsten Lutz. The complexity of conjunctive query answering in expressive description logics. In Alessandro Armando, Peter Baumgartner, and Gilles Dowek, editors, *Proc. of the 4th Int. Joint Conf. on Automated Reasoning (IJCAR)*, number 5195 in LNAI, pages 179–193. Springer, 2008.

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

[27] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of the 13th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2012)*, pages 297–307, 2012.

[28] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *J. on Data Semantics*, 10:133–173, 2008.

[29] Mariano Rodriguez-Muro, Roman Kontchakov, and Michael Zakharyaschev. Ontology-based data access: Ontop of databases. In *Proc. of the 12th Int. Semantic Web Conf. (ISWC 2013)*, pages 558–573. Springer, 2013.

[30] Despoina Trivela, Giorgos Stoilos, Alexandros Chortaras, and Giorgos B. Stamou. Optimising resolution-based rewriting algorithms for OWL ontologies. *J. of Web Semantics*, 33:30–49, 2015.

[31] Kewen Wang, Zhe Wang, Rodney W. Topor, Jeff Z. Pan, and Grigoris Antoniou. Eliminating concepts and roles from ontologies in expressive descriptive logics. *Computational Intelligence*, 30(2):205–232, 2014. doi: 10.1111/j.1467-8640.2012.00442.x.

[32] Yujiao Zhou, Bernardo Cuenca Grau, Yavor Nenov, Mark Kaminski, and Ian Horrocks. Pagoda: Pay-as-you-go ontology query answering using a datalog reasoner. *Journal of Artificial Intelligence Research (JAIR)*, 54:309–367, 2015.

A Proof of Theorem 7

A.1 Minimal models

We consider $\mathcal{ELU}_{rhs}$ TBoxes, $\mathcal{T}$, that consist of concept inclusions of the form

- $A \sqsubseteq C$,
- $A \sqsubseteq B \sqcup C$,
- $A \sqsubseteq \exists R.C$,

where A, B, C are concept names and R is a role name. We construct by induction a (possibly infinite) labelled forest $\mathfrak{O}$ with a labelling function ℓ. For each $a \in \mathsf{ind}(\mathcal{A})$, a is the root of a tree in $\mathfrak{O}$ with $A \in \ell(a)$ iff $A(a) \in \mathcal{A}$. Suppose now that σ is a node in $\mathfrak{O}$ and $A \in \ell(\sigma)$. If $A \sqsubseteq C$ is an axiom of $\mathcal{T}$ and $C \notin \ell(\sigma)$, then we add C to $\ell(\sigma)$. If $A \sqsubseteq B \sqcup C$ is an axiom of $\mathcal{T}$ and neither $B \in \ell(\sigma)$ nor $C \in \ell(\sigma)$, then we add to $\ell(\sigma)$ either B or C (but not both); in this case, we call σ an *or-node*. If $A \sqsubseteq \exists R.C$ is an axiom of $\mathcal{T}$, but the constructed part of the tree does not contain a node $\sigma \cdot w_{\exists R.C}$, then we add $\sigma \cdot w_{\exists R.C}$ as an R-successor of σ and set $\ell(\sigma \cdot w_{\exists R.C}) = \{C\}$.

Given an $\mathcal{ELU}_{rhs}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, we define a *minimal model* $\mathcal{M} = (\Delta^{\mathcal{M}}, \cdot^{\mathcal{M}})$ of $\mathcal{K}$ by taking $\Delta^{\mathcal{M}}$ to be the set of nodes in $\mathfrak{O}$, $R^{\mathcal{M}}$ to be the R-relation in $\mathfrak{O}$ together with (a, b) such that $R(a, b) \in \mathcal{A}$, and set

$$A^{\mathcal{M}} = \{\sigma \in \Delta^{\mathcal{M}} \mid A \in \ell(\sigma)\},$$

for every concept name A.

Lemma 17. *Let $\mathcal{K}$ be an $\mathcal{ELU}_{rhs}$ KB $\mathcal{K}$ and let $\boldsymbol{M}_{\mathcal{K}}$ be the set of its minimal models. Then $\boldsymbol{M}_{\mathcal{K}}$ is complete for $\mathcal{K}$.*

Proof. It suffices to show that (*i*) every minimal model is a model of $\mathcal{K}$, and (*ii*) for every model $\mathcal{I}$ of $\mathcal{K}$, there is a minimal model $\mathcal{M}$ that is homomorphically embeddable into $\mathcal{I}$. The former follows from the construction.

(*ii*) Let $\mathcal{I}$ be a model of $\mathcal{K}$. We construct by induction a set Δ and a labelling function ℓ defining a minimal model $\mathcal{M}$ and a function h such that h is a homomorphism from $\mathcal{M}$ to $\mathcal{I}$. First we set $a \in \Delta$ and $A \in \ell(a)$, for each $A(a) \in \mathcal{A}$. Suppose that $A \in \ell(a)$ for some a. If $A \sqsubseteq C$ is an axiom in $\mathcal{T}$ and $C \notin \ell(A)$, we add C to $\ell(a)$. Suppose now that $A \sqsubseteq B \sqcup C$ is an axiom in $\mathcal{T}$, and $B \notin \ell(A)$, $C \notin \ell(a)$. Since $\mathcal{I}$ is a model of $\mathcal{K}$, it must be the case that $B \in \boldsymbol{t}^{\mathcal{I}}(a)$ or $C \in \boldsymbol{t}^{\mathcal{I}}(a)$. In the former case, we add B to $\ell(a)$, in the latter case, we add C to $\ell(a)$. We now set $h(a) = a$, for each $a \in \mathsf{ind}(\mathcal{A})$. Clearly, $A \in \boldsymbol{t}^{\mathcal{I}}(h(a))$, for each $A \in \ell(a)$.

Suppose that $\sigma \in \Delta^{\mathcal{M}}$ such that $h(\sigma)$ is set, and $A \in \ell(\sigma)$. Suppose further that $A \sqsubseteq \exists R.C \in \mathcal{T}$ and $\sigma \cdot w_{\exists R.C}$ is not in Δ. Since $\mathcal{I}$ is a model of $\mathcal{K}$ and by inductive assumption $A \in \boldsymbol{t}^{\mathcal{I}}(h(\sigma))$, there exists $d \in \Delta^{\mathcal{I}}$ such that $(h(\sigma), d) \in R^{\mathcal{I}}$ and $d \in C^{\mathcal{I}}$. So we add $\sigma \cdot w_{\exists R.C}$ to Δ as successor of σ, define $\ell(\sigma \cdot w_{\exists R.C})$ similarly to the base case starting from $\{C\}$, and set $h(\sigma \cdot w_{\exists R.C}) = d$. Clearly, for each $\sigma \in \Delta$, for each $A \in \ell(\sigma)$ we have that $A \in \boldsymbol{t}^{\mathcal{I}}(h(\sigma))$.

Now the minimal model $\mathcal{M}$ is defined as $(\Delta, \cdot^{\mathcal{M}})$, where $\cdot^{\mathcal{M}}$ is defined as in the construction of minimal model. By the construction of Δ and the fact that $\mathcal{M}$ is minimal, we obtain that h is indeed a homomorphism from $\mathcal{M}$ to $\mathcal{I}$. □

A.2 Proof of Theorem 7 (*i*) and (*ii*) for CQs

A *tile type* $T = (\mathit{up}(T), \mathit{down}(T), \mathit{left}(T), \mathit{right}(T))$ consists of four colours. The following $N \times M$-*tiling problem* is known to be undecidable: given a finite set $\mathfrak{T}$ of tile types, a tile type $I \in \mathfrak{T}$ and two colours wall and $\mathit{ceiling}$, decide whether there exist $N, M \in \mathbb{N}$ such that the $N \times M$ grid can be tiled using $\mathfrak{T}$ in such a way that $(1, 1)$ is covered with a tile of type I, every (N, i), for $i \leq M$, is covered with a tile of some type T with $\mathit{right}(T) = \mathit{wall}$, and every (i, M), for $i \leq N$, is covered with a tile of some type T with $\mathit{up}(T) = \mathit{ceiling}$.

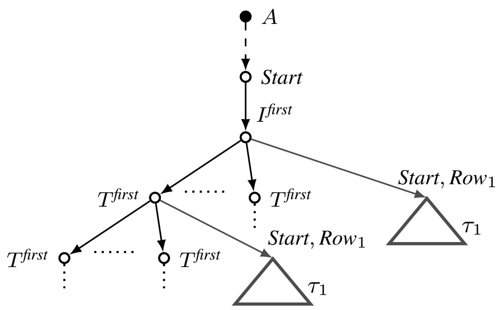
We require role names P and R, and the following concept names:

- $T^{\mathit{first}}, T_k, T_k^{\mathit{halt}}, \widehat{T}_k$ for $T \in \mathfrak{T}$, $k = 0, 1, 2$;
- $\mathit{Row}, \mathit{Row}_k, \mathit{Row}_k^{\mathit{halt}}$, for $k = 0, 1, 2$;
- A, *Start* and *End*.

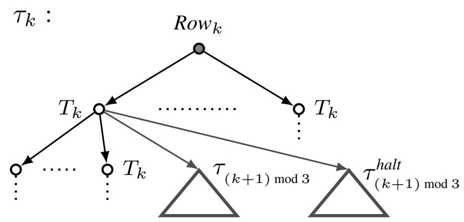
Let $\mathcal{K}_2 = (\mathcal{T}_2, \{A(a)\})$, where $\mathcal{T}_2$ contains the following axioms, for $k = 0, 1, 2$:

$$A \sqsubseteq \exists P.(\mathit{Start} \sqcap \exists R.I^{\mathit{first}}), \tag{1}$$
$$T^{\mathit{first}} \sqsubseteq \exists R.S^{\mathit{first}}, \text{ if } \mathit{right}(T) = \mathit{left}(S),\ T, S \in \mathfrak{T}, \tag{2}$$
$$T^{\mathit{first}} \sqsubseteq \exists R.(\mathit{Start} \sqcap \mathit{Row}_1), T \in \mathfrak{T}, \mathit{right}(T) = \mathit{wall}, \tag{3}$$
$$T^{\mathit{first}} \sqsubseteq \widehat{T}_0, \quad \text{for } T \in \mathfrak{T}, \tag{4}$$
$$\mathit{Row}_k \sqsubseteq \exists R.T_k, \quad \text{for } T \in \mathfrak{T}, \tag{5}$$
$$T_k \sqsubseteq \exists R.S_k, \quad \text{if } \mathit{right}(T) = \mathit{left}(S) \text{ and } T, S \in \mathfrak{T}, \tag{6}$$
$$T_k \sqsubseteq \exists R.\mathit{Row}_{(k+1) \bmod 3}, \quad \text{if } \mathit{right}(T) = \mathit{wall}, \tag{7}$$
$$T_k \sqsubseteq \exists R.\mathit{Row}^{\mathit{halt}}_{(k+1) \bmod 3}, \quad \text{if } \mathit{right}(T) = \mathit{wall}, \tag{8}$$
$$\mathit{Row}_k \sqsubseteq \mathit{Row}, \tag{9}$$
$$T_k \sqsubseteq \widehat{T}_k, \quad \text{for } T \in \mathfrak{T}, \tag{10}$$
$$T_k \sqsubseteq \widehat{S}_{(k-1) \bmod 3}, \quad \text{if } \mathit{down}(T) = \mathit{up}(S), T, S \in \mathfrak{T}, \tag{11}$$
$$\mathit{Row}^{\mathit{halt}}_k \sqsubseteq \exists R.\mathit{End} \sqcup \bigsqcap_{\mathit{up}(T) = \mathit{ceiling}} \exists R.T^{\mathit{halt}}_k, \tag{12}$$
$$T^{\mathit{halt}}_k \sqsubseteq \exists R.S^{\mathit{halt}}_k, \quad \text{if } \mathit{right}(T) = \mathit{left}(S) \text{ and } \mathit{up}(S) = \mathit{ceiling}, \tag{13}$$
$$T^{\mathit{halt}}_k \sqsubseteq \exists R.(\mathit{Row} \sqcap \exists R.\mathit{End}), \quad \text{if } \mathit{right}(T) = \mathit{wall}, \tag{14}$$
$$\mathit{Row}^{\mathit{halt}}_k \sqsubseteq \mathit{Row}, \tag{15}$$
$$T^{\mathit{halt}}_k \sqsubseteq \widehat{S}_{(k-1) \bmod 3}, \ \text{if } \mathit{down}(T) = \mathit{up}(S), T, S \in \mathfrak{T}. \tag{16}$$

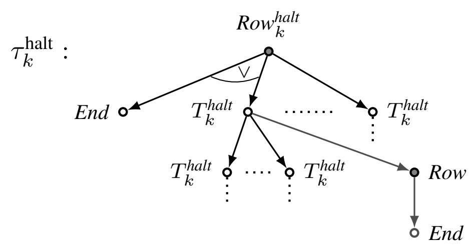
The axioms (1)-(4) produce the following tree rooted at an A-point:



The axioms (5)-(11) produce trees τ_k rooted at Row_k-points:



Finally, the axioms (12)-(16) produce trees τ^{halt}_k rooted at $\mathit{Row}^{\mathit{halt}}_k$-points:



Denote by $\boldsymbol{q}_n$ any Boolean CQ of the form

$$\exists \vec{x}\big(\mathit{Start}(x_0) \wedge \bigwedge_{i=0}^{n} R(x_i, x_{i+1}) \wedge \bigwedge_{i=1}^{n} B_i(x_i) \wedge \mathit{End}(x_{n+1})\big)$$

where $B_i \in \{\mathit{Row}\} \cup \{\widehat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$.

Lemma 18. *There exists a CQ $\boldsymbol{q}_n$ such that $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n$ iff there exist $N, M \in \mathbb{N}$ for which $\mathfrak{T}$ tiles the $N \times M$ grid as described above.*

Proof. ($\Leftarrow$) Suppose $\mathfrak{T}$ tiles the $N \times M$ grid so that a tile of type $T^{ij} \in \mathfrak{T}$ covers (i, j). Let

$$\mathit{block}_j = (\widehat{T}_k^{1,j}, \dots, \widehat{T}_k^{N,j}, \mathit{Row}),$$

for $j = 1, \dots, M-1$ and $k = (j-1) \mod 3$. Let $\boldsymbol{q}_n$ be the CQ in which the B_i follow the pattern

$$\mathit{block}_1,\ \mathit{block}_2, \dots,\ \mathit{block}_{M-1}$$

(thus, $n = (N+1) \times (M-1)$). In view of Lemma 17, we only need to prove $\mathcal{M} \models \boldsymbol{q}_n$ for each minimal model $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$. Take such an $\mathcal{M}$. We have to show that there is an R-path $x_0, \dots, x_{n+1}$ in $\mathcal{M}$ such that $x_i \in B_i^{\mathcal{M}}$ and $x_{n+1} \in \mathit{End}^{\mathcal{M}}$.

First, we construct an auxiliary R-path $y_0, \dots, y_n$. We take $y_0 \in \mathit{Row}^{\mathcal{M}}$ and $y_1 \in I_0^{\mathcal{M}}$ by (1) ($I_0 = T^{1,1}$). Then we take $y_2 \in (T^{1,1})^{\mathcal{M}}, \dots, y_{N+1} \in (T^{N,1})^{\mathcal{M}}$ by (2). We now have $\mathit{right}(T^{N,1}) = \mathit{wall}$. By (3), we obtain $y_{N+2} \in \mathit{Row}_1$. By (9), $y_{N+2} \in \mathit{Row}_1^{\mathcal{M}} \subseteq \mathit{Row}^{\mathcal{M}}$. We proceed in this way, starting with (5), till the moment we construct $y_{n-1} \in T^{N,M-1}$, for which we use (8) and (15) to obtain $y_n \in \mathit{Row}_k^{\mathit{halt}} \subseteq \mathit{Row}^{\mathcal{M}}$, for some k. Note that $T^{\mathcal{M}} \subseteq \widehat{T}^{\mathcal{M}}$ by (10).

By (12), two cases are possible now.

Case 1: there is y such that $(y_n, y) \in R^{\mathcal{M}}$ and $y \in \mathit{End}^{\mathcal{M}}$. Then we take $x_0 = y_0, \dots, x_n = y_n, x_{n+1} = y$.

Case 2: there is an object z_1 such that $(y_n, z_1) \in R^{\mathcal{M}}$ and $z_1 \in (T_k^{\mathit{halt}})^{\mathcal{M}}$, where $T = T^{1,M}$ for which $\mathit{up}(T) = \mathit{ceiling}$. We then use (13) and find objects $z_2, \dots, z_N, u, v$ such that $z_i \in (T_k^{\mathit{halt}})^{\mathcal{M}}$, where $T = T^{i,M}$, $u \in \mathit{Row}^{\mathcal{M}}$ and $v \in \mathit{End}^{\mathcal{M}}$. We take $x_0 = y_{N+1}, \dots, x_{n-N-1} = y_n$, $x_{n-N} = z_1, \dots, x_{n-1} = z_N$, and $x_n = u, x_{n+1} = v$. Note that, by (11) and (16), we have $(T^{i,j})^{\mathcal{M}} \subseteq (\widehat{T}^{i,j-1})^{\mathcal{M}}$.

($\Rightarrow$) Let $\boldsymbol{q}_n$ be such that $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n$, and so $\mathcal{M} \models \boldsymbol{q}_n$ for each $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$. Consider all the pairwise distinct pairs $(\mathcal{M}, h)$ such that $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$ and h is a homomorphism from $\boldsymbol{q}_n$ to $\mathcal{M}$. Note that $h(\boldsymbol{q}_n)$ contains an or-node σ_h (which is an instance of $\mathit{Row}_k^{\mathit{halt}}$, for some k). We call $(\mathcal{M}, h)$ and h *left* if $h(x_{n+1}) = \sigma_h \cdot w_{\exists R.\mathit{End}}$, and *right* otherwise. It is not hard to see that there exist a left $(\mathcal{M}_\ell, h_\ell)$ and a right $(\mathcal{M}_r, h_r)$ with $\sigma_{h_\ell} = \sigma_{h_r}$ (if this is not the case, we can construct $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$ such that $\mathcal{M} \not\models \boldsymbol{q}_n$).

Take $(\mathcal{M}_\ell, h_\ell)$ and $(\mathcal{M}_r, h_r)$ such that $\sigma_{h_\ell} = \sigma_{h_r} = \sigma$ and use them to construct the required tiling. Let $\sigma = a w_0 \cdots w_n$. We have $h_\ell(x_{n+1}) = \sigma \cdot w_{\exists R.\mathit{End}}$ and $h_\ell(x_n) = \sigma$. Let $h_r(x_{n+1}) = \sigma v_1 \cdots v_{m+2}$, which is an instance of *End*. Then $h_r(x_n) = \sigma v_1 \cdots v_{m+1}$, which is an instance of *Row*.

Suppose $v_m = w_{\exists R.T_2^{\mathit{halt}}}$ (other k's are treated analogously). By (14), $\mathit{right}(T) = \mathit{wall}$; by (13), $\mathit{up}(T) = \mathit{ceiling}$. Suppose $w_{n-1} = w_{\exists R.S_k}$. Then it must be that $k = 1$. By (8), $\mathit{right}(S) = \mathit{wall}$. Consider the atom $B_{n-1}(x_{n-1})$ from $\boldsymbol{q}_n$. Then both $a w_0 \cdots w_{n-1}$ and $\sigma v_1 \cdots v_m$ are instances of B_{n-1}. By (10) and (16), $B_{n-1} = \widehat{S}_1$ and $\mathit{down}(T) = \mathit{up}(S)$.

Suppose $v_{m-1} = w_{\exists R.U_2^{\mathit{halt}}}$. By (13), $\mathit{right}(U) = \mathit{left}(T)$ and $\mathit{up}(U) = \mathit{ceiling}$. Suppose $w_{n-2} = w_{\exists R.Q_1}$. By (6), $\mathit{right}(Q) = \mathit{left}(S)$. Consider the atom $B_{n-2}(x_{n-2})$ from $\boldsymbol{q}_n$. Then both $a w_0 \cdots w_{n-2}$ and $\sigma v_1 \cdots v_{m-1}$ are instances of B_{n-2}. By (10) and (16), $B_{n-2} = \widehat{Q}_1$ and $\mathit{down}(U) = \mathit{up}(Q)$.

We proceed in the same way until we reach σ and $a w_0 \cdots w_{n-N-1}$, for $N = m$, both of which are instances of $B_{n-N-1} = \mathit{Row}$. Thus have tiled the two last rows of the grid. We proceed further and tile the whole $N \times M$ grid, where $M = n/(N+1) + 1$. ☐

Note that $\mathcal{K}_2$ encodes tilings with at least 3 rows, hence, $M \geq 3$.

We now define a KB $\mathcal{K}_1 = (\mathcal{T}_1, \{A(a)\})$. Let $\Sigma_0 = \{\mathit{Row}\} \cup \{\hat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$, and let $\mathcal{T}_1$ contain

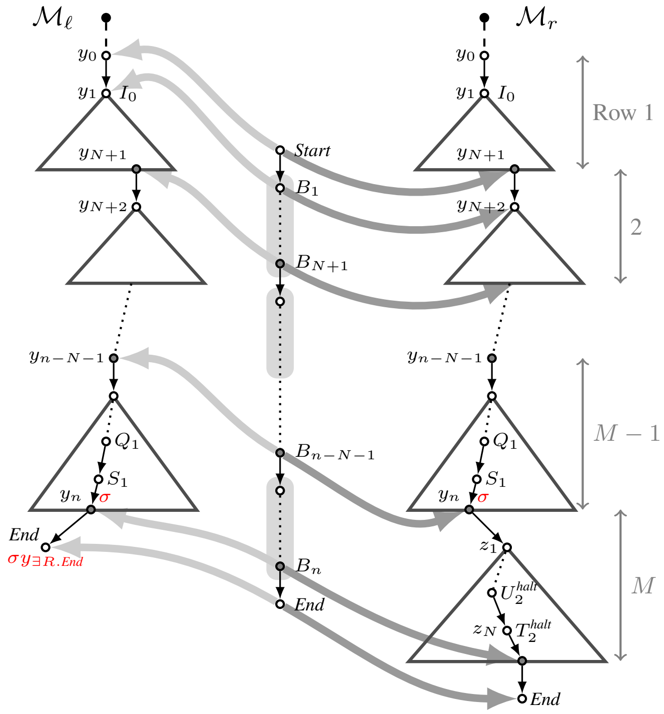


Figure 4: The two homomorphisms to two minimal models

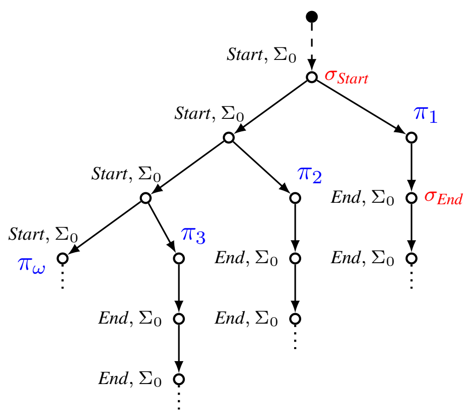
the following axioms:

$$A \sqsubseteq \exists P.D, \tag{17}$$

$$D \sqsubseteq \exists R.D \sqcap \exists R.\exists R.E \sqcap \bigsqcap_{X \in \Sigma_0} X \sqcap \mathit{Start}, \tag{18}$$

$$E \sqsubseteq \exists R.E \sqcap \bigsqcap_{X \in \Sigma_0} X \sqcap \mathit{End}. \tag{19}$$

As $\mathcal{K}_1$ is an $\mathcal{EL}$-KB, it has a canonical model $\mathcal{M}_{\mathcal{K}_1}$:



Let Σ be the signature of $\mathcal{K}_1$.

Lemma 19. *$\prod \boldsymbol{M}_{\mathcal{K}_2}$ is $n\Sigma$-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$ for any n iff there does not exist a CQ $\boldsymbol{q}_n$ such that $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n$.*

Proof. ($\Rightarrow$) Suppose $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n$ for some n. Since $\prod \boldsymbol{M}_{\mathcal{K}_2}$ is $n\Sigma$-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$, we then have $\mathcal{M}_1 \models \boldsymbol{q}_n$, which is clearly impossible because of the B_i and *End* in $\boldsymbol{q}_n$.

($\Leftarrow$) Suppose $\prod \boldsymbol{M}_{\mathcal{K}_2} \not\models \boldsymbol{q}_n$ for all CQs of the form $\boldsymbol{q}_n$. Take any subinterpretation of $\prod \boldsymbol{M}_{\mathcal{K}_2}$ whose domain contains m elements. We can regard this subinterpretation as a Boolean Σ-CQ, and so denote it by $\boldsymbol{q}$. Without loss of generality we can assume that $\boldsymbol{q}$ is connected; clearly, $\boldsymbol{q}$ is tree-shaped. We know that there is no Σ-homomorphism from $\boldsymbol{q}_n$ into $\boldsymbol{q}$ for any n; in particular, $\boldsymbol{q}$ does not have a subquery of the form $\boldsymbol{q}_n$. We have to show that $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}$.

Suppose $\boldsymbol{q}$ contains A or P, then they appear at the root of $\boldsymbol{q}$ or, respectively, in the fist edge of $\boldsymbol{q}$. By the structure of $\mathcal{K}_2$, it follows then $\boldsymbol{q}$ does not contain *End* and, therefore, can be mapped into π_ω. In what follows, we assume that $\boldsymbol{q}$ does not contain A and P.

If $\boldsymbol{q}$ does not contain *Start* atoms, or $\boldsymbol{q}$ does not contain *End* atoms, then clearly, $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}$. In the former case, $\boldsymbol{q}$ can be mapped to π_1 by sending the root of $\boldsymbol{q}$ to σ_{End}. In the latter case, $\boldsymbol{q}$ can be mapped to π_ω by sending the root of $\boldsymbol{q}$ to σ_{Start}.

Assume that $\boldsymbol{q}$ contains both *Start* and *End* atoms. If there exists a(n R-)path from a *Start* node to an *End* node in $\boldsymbol{q}$, then by the structure of $\mathcal{K}_2$, the *Start* node must be the root of $\boldsymbol{q}$. Since $\boldsymbol{q}$ does not contain a subquery of the form $\boldsymbol{q}_n$, this R-path should contain variables with the empty Σ-concept label, in which case $\boldsymbol{q}$ can be mapped into some π_i, $1 \leq i < \omega$, by mapping the root of $\boldsymbol{q}$ to σ_{Start}.

Now, assume that in $\boldsymbol{q}$ there does not exist a path from a *Start* node to an *End* node. Hence, the *Start* node is not the root of $\boldsymbol{q}$. Let $\mathcal{M}$ be a minimal model of $\mathcal{K}_2$. Then the root y_0 of $\boldsymbol{q}$ should be mapped to an element of the form $\delta \cdot w_{\exists R.T^{first}}$ in $\Delta^{\mathcal{M}}$, since there is a path from the root of $\boldsymbol{q}$ to a *Start* node. By the structure of $\mathcal{K}_2$, the general form of $\boldsymbol{q}$ should be as follows:

$$\begin{aligned} &Q_{T_0} \sqcap \exists R.(Q_{Start} \sqcap Q_{\text{noEnd}}) \sqcap \\ &\quad \exists R.(Q_{T_0} \sqcap \exists R.(Q_{Start} \sqcap Q_{\text{noEnd}}) \sqcap \\ &\qquad \exists R.(Q_{T_0} \sqcap \exists R.(Q_{Start} \sqcap Q_{\text{noEnd}}) \sqcap \\ &\qquad\qquad \cdots \sqcap \exists R.Q_{End})) \end{aligned}$$

where Q_{End} is an $\mathcal{EL}$ concept constructed using R and concepts in $\Sigma_0 \cup \{End\}$, Q_{noEnd} is an $\mathcal{EL}$ concept constructed using R and concepts in Σ_0, Q_{Start} is either an empty query or a *Start* atom, and Q_{T_0} is either an empty query or a $\widehat{T}_0$ atom. We prove that each path in $\boldsymbol{q}$ ending with an *End* node must have at least one intermediate node with the empty Σ-concept label.

For simplicity assume that $\boldsymbol{q}$ consists of two subtrees $\boldsymbol{q}_{End}$ and $\boldsymbol{q}_{Start}$, where $\boldsymbol{q}_{End}$ is a path ending with an *End* node, and $\boldsymbol{q}_{Start}$ is a tree rooted in a *Start* node. By contradiction, assume that each intermediate node in $\boldsymbol{q}_{End}$ is labeled with either some $\widehat{T}_k$ or *Row*. Since $\mathcal{K}_2 \models \boldsymbol{q}_{End}$ it follows that there is some n such that the distance between two neighbour *Row* nodes in $\boldsymbol{q}_{End}$ is n. Let $\mathcal{M}_\ell$ and $\mathcal{M}_r$ be minimal models that satisfy (12) by picking the first and the second disjunct, respectively, and identical, otherwise. Assume that $\mathcal{M}_\ell$ satisfies $\boldsymbol{q}_{End}$ by mapping y_0 to σ_l of the form $\delta \cdot w_{\exists R.T^{first}}$ and $\mathcal{M}_r$ satisfies $\boldsymbol{q}_{End}$ by mapping y_0 to σ_r of the form $\sigma_l \cdots w_{\exists R.T^{first}}$. Then the distance between σ_l and σ_r is n. Let the distance from y_0 to the first *Row* node y_m be m. Then m should be less than or equal $n-1$. Therefore, y_m should be mapped to a predecessor σ' of σ_r in $\mathcal{M}_\ell$. However, such a mapping is not a homomorphism as the Σ-label of σ' does not contain *Row* (only, a concept of the form $\widehat{T}_0$). Contradiction with the assumption that $\mathcal{K}_2 \models \boldsymbol{q}$ and that the label of y_l is non-empty.

Finally, we conclude that $\boldsymbol{q}$ can be mapped to $\mathcal{M}_1$ as follows: y_0 to σ_{Start}, $\boldsymbol{q}_{Start}$ into π_ω, and $\boldsymbol{q}_{End}$ into π_i, where the distance from y_0 to the first gap is i, for $1 \leq i < |\boldsymbol{q}|$. □

As an immediate consequence of the obtained results we have:

Theorem 7 (*i*) *The problem whether a* Horn-$\mathcal{ALC}$ *KB* Σ*-CQ entails an* $\mathcal{ALC}$ *KB is undecidable.*

Theorem 7 (*ii*) Σ*-CQ inseparability between* Horn-$\mathcal{ALC}$ *and* $\mathcal{ALC}$ *KBs is undecidable.*

Proof. Let $\mathcal{K}_2' = \mathcal{K}_2 \cup \mathcal{K}_1$. Then the following set $\boldsymbol{M}_{\mathcal{K}_2'}$ is complete for $\mathcal{K}_2'$:

$$\boldsymbol{M}_{\mathcal{K}_2'} = \{\mathcal{M} \uplus \mathcal{M}_{\mathcal{K}_1} \mid \mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}\},$$

where $\mathcal{M} \uplus \mathcal{M}_{\mathcal{K}_1}$ is the interpretation that results from merging the roots a of $\mathcal{M}$ and $\mathcal{M}_{\mathcal{K}_1}$. As before, we set $\Sigma = \mathsf{sig}(\mathcal{K}_1)$. It suffices to show that $\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2$ iff $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ-CQ inseparable.

($\Leftarrow$) follows from $\mathcal{K}_2 \models \boldsymbol{q}(\boldsymbol{a}) \;\Rightarrow\; \mathcal{K}_2' \models \boldsymbol{q}(\boldsymbol{a})$.

($\Rightarrow$) It follows from the definition that $\mathcal{K}_2'$ Σ-CQ entails $\mathcal{K}_1$. So we have to show that $\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2'$. Suppose this is not the case and there is a Σ-CQ $\boldsymbol{q}$ such that $\mathcal{K}_2' \models \boldsymbol{q}$ and $\mathcal{K}_1 \not\models \boldsymbol{q}$. We can assume $\boldsymbol{q}$ to be a *smallest connected* CQ with this property; in particular, no proper sub-CQ of $\boldsymbol{q}$ separates $\mathcal{K}_1$ and $\mathcal{K}_2'$.

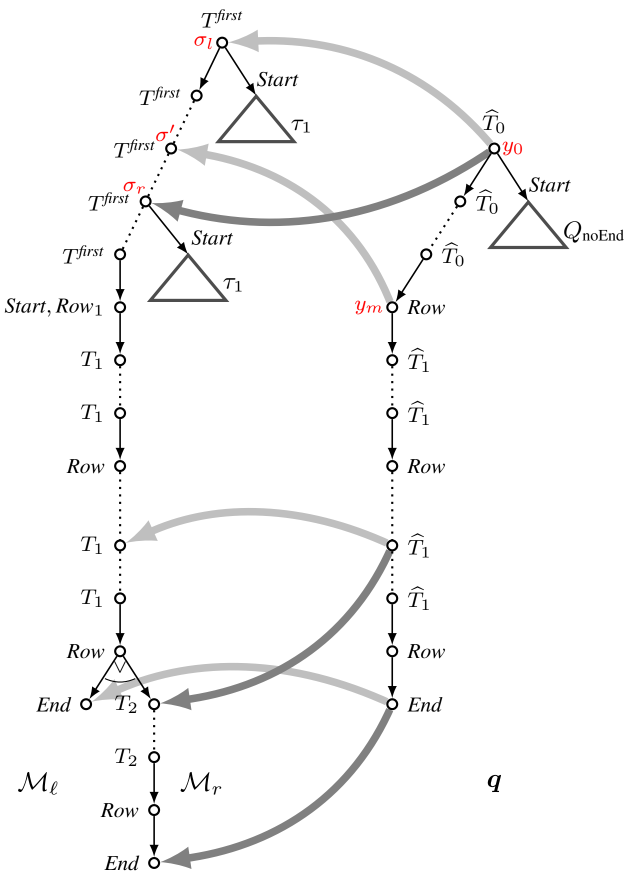


Figure 5: A query that contains both *Start* and *End* atoms must have variables with empty concept labels.

Now, we cannot have $\mathcal{K}_2 \models \boldsymbol{q}$ because this would contradict the fact that $\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2$. Then $\mathcal{K}_2 \not\models \boldsymbol{q}$, and so there is $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$ such that $\mathcal{M} \not\models \boldsymbol{q}$. On the other hand, we have $\mathcal{M} \uplus \mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}$. Take a homomorphism $h\colon \boldsymbol{q} \to \mathcal{M} \uplus \mathcal{M}_{\mathcal{K}_1}$. As $\boldsymbol{q}$ is connected, $\mathcal{M} \not\models \boldsymbol{q}$ and $\mathcal{M}_{\mathcal{K}_1} \not\models \boldsymbol{q}$, there is a variable x in $\boldsymbol{q}$ such that $h(x) = a$. For every variable x with $h(x) = a$, we remove $\exists x$ from the prefix of $\boldsymbol{q}$ if any. Denote by $\boldsymbol{q}'$ the maximal sub-CQ of $\boldsymbol{q}$ such that $h(\boldsymbol{q}') \subseteq \mathcal{M}$ (more precisely, $S(\boldsymbol{y})$ is in $\boldsymbol{q}'$ iff $h(\boldsymbol{y}) \subseteq \Delta^{\mathcal{M}}$). Clearly, $\boldsymbol{q}' \subsetneq \boldsymbol{q}$ and $\mathcal{K}_2' \models \boldsymbol{q}'$. Denote by $\boldsymbol{q}''$ the complement of $\boldsymbol{q}'$ to $\boldsymbol{q}$. Now, we either have $\mathcal{K}_1 \models \boldsymbol{q}'$ or $\mathcal{K}_1 \not\models \boldsymbol{q}'$. The latter case contradicts the choice of $\boldsymbol{q}$ because $\boldsymbol{q}'$ is its proper sub-CQ. Thus, $\mathcal{K}_1 \models \boldsymbol{q}'$, and so there is a homomorphism $h'\colon \boldsymbol{q}' \to \mathcal{M}_{\mathcal{K}_1}$ with $h'(x) = a$ for every free variable x. Define a map $g\colon \boldsymbol{q} \to \mathcal{M}_{\mathcal{K}_1}$ by taking $g(y) = h'(y)$ if y is in $\boldsymbol{q}'$ and $g(y) = h(y)$ otherwise. The map g is a homomorphism because all the variables that occur in both $\boldsymbol{q}'$ and $\boldsymbol{q}''$ are free and must be mapped by g to a. Therefore, $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}$, which is a contradiction. □

A.3 Proof of Theorem 7 (*i*) and (*ii*) for rCQs

Let

$$\mathcal{A} = \{R(a,a), \mathit{Row}(a), A(a)\} \cup \{\widehat{T}_0(a) \mid T \in \mathfrak{T}\}. \tag{20}$$

$\mathcal{T}_2$ contains the following axioms, where $k = 0, 1, 2$:

$$A \sqsubseteq \exists R.(\mathit{Row} \sqcap \exists R.I_0), \tag{21}$$
$$\mathit{Row}_k \sqsubseteq \exists R.T_k, \quad \text{for } T \in \mathfrak{T}, \tag{22}$$
$$T_k \sqsubseteq \exists R.S_k, \quad \text{if } \mathit{right}(T) = \mathit{left}(S) \text{ and } T, S \in \mathfrak{T}, \tag{23}$$
$$T_k \sqsubseteq \exists R.\mathit{Row}_{(k+1) \bmod 3}, \quad \text{if } \mathit{right}(T) = \mathit{wall}, \tag{24}$$
$$T_k \sqsubseteq \exists R.\mathit{Row}^{\mathit{halt}}_{(k+1) \bmod 3}, \quad \text{if } \mathit{right}(T) = \mathit{wall}, \tag{25}$$
$$\mathit{Row}_k \sqsubseteq \mathit{Row}, \tag{26}$$
$$T_k \sqsubseteq \widehat{T}_k, \quad \text{for } T \in \mathfrak{T}, \tag{27}$$
$$T_k \sqsubseteq \widehat{S}_{(k-1) \bmod 3}, \quad \text{if } \mathit{down}(T) = \mathit{up}(S), T, S \in \mathfrak{T}, \tag{28}$$
$$\mathit{Row}^{\mathit{halt}}_k \sqsubseteq \exists R.\mathit{End} \sqcup \bigsqcap_{\mathit{up}(T) = \mathit{ceiling}} \exists R.T^{\mathit{halt}}_k, \tag{29}$$
$$T^{\mathit{halt}}_k \sqsubseteq \exists R.S^{\mathit{halt}}_k, \quad \text{if } \mathit{right}(T) = \mathit{left}(S) \text{ and } \mathit{up}(S) = \mathit{ceiling}, \tag{30}$$
$$T^{\mathit{halt}}_k \sqsubseteq \exists R.(\mathit{Row} \sqcap \exists R.\mathit{End}), \quad \text{if } \mathit{right}(T) = \mathit{wall}, \tag{31}$$
$$\mathit{Row}^{\mathit{halt}}_k \sqsubseteq \mathit{Row}, \tag{32}$$
$$T^{\mathit{halt}}_k \sqsubseteq \widehat{S}_{(k-1) \bmod 3}, \quad \text{if } \mathit{down}(T) = \mathit{up}(S), T, S \in \mathfrak{T}. \tag{33}$$

Let $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$. Consider a CQ $\boldsymbol{q}_n(X)$ of the form

$$\exists \vec{x} \big(R(X, x_0) \land \bigwedge_{i=0}^{n} \big(R(x_i, x_{i+1}) \land B_i(x_i) \big) \land \mathit{End}(x_{l+1}) \big)$$

where $B_i \in \{\mathit{Row}\} \cup \{\widehat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$.

Lemma 20. *There exists a CQ $\boldsymbol{q}_n(X)$ such that $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n(a)$ iff there exist $N, M \in \mathbb{N}$ for which $\mathfrak{T}$ tiles the $N \times M$ grid as described above.*

Proof. ($\Leftarrow$) Suppose $\mathfrak{T}$ tiles the $N \times M$ grid under which a tile of type $T^{ij} \in \mathfrak{T}$ covers (i, j). Let

$$\mathit{block}_j = (\widehat{T}^{1,j}_k, \dots, \widehat{T}^{N,j}_k, \mathit{Row}),$$

for $j = 1, \dots, M-1$ and $k = (j-1) \mod 3$. Let $\boldsymbol{q}_n$ be the CQ in which the B_i follow the pattern

$$\mathit{Row},\ \mathit{block}_1,\ \mathit{block}_1,\ \mathit{block}_2, \dots,\ \mathit{block}_{M-1}$$

(thus, $n = (N+1) \times M + 1$). In view of Proposition 5 we only need to prove $\mathcal{M} \models \boldsymbol{q}_n$ for each minimal model $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$. Take such an $\mathcal{M}$. We have to show that there is an R-path $a, x_0, \dots, x_{n+1}$ in $\mathcal{M}$ such that $x_i \in B_i^{\mathcal{M}}$ and $x_{n+1} \in \mathit{End}^{\mathcal{M}}$.

First, we construct an auxiliary R-path $y_0, \dots, y_{n-N-1}$. We take $y_0 \in \mathit{Row}^{\mathcal{M}}$ and $y_1 \in I_0^{\mathcal{M}}$ by (21) ($I_0 = T^{1,1}$). Then we take $y_2 \in (T^{2,1})^{\mathcal{M}}, \dots, y_N \in (T^{N,1})^{\mathcal{M}}$ by (23). We now have $\mathit{right}(T^{N,1}) = \mathit{wall}$. By (24), we obtain $y_{N+1} \in \mathit{Row}_1$. By (26), $y_{N+1} \in \mathit{Row}_1^{\mathcal{M}} \subseteq \mathit{Row}^{\mathcal{M}}$. We proceed in this way, starting with (22), till the moment we construct $y_{n-1} \in T^{N,M-1}$, for which we use (25) and (32) to obtain $y_n \in \mathit{Row}_k^{\mathit{halt}} \subseteq \mathit{Row}^{\mathcal{M}}$, for some k. Note that $T^{\mathcal{M}} \subseteq \widehat{T}^{\mathcal{M}}$ by (27).

By (29), two cases are possible now.

Case 1: there is an object y such that $(y_n, y) \in R^{\mathcal{M}}$ and $y \in \mathit{End}^{\mathcal{M}}$. Then we take $x_0 = \dots = x_N = a$, $x_{N+1} = y_0, \dots, x_n = y_{n-N-1}, x_{n+1} = y$.

Case 2: there is an object z_1 such that $(y_n, z_1) \in R^{\mathcal{M}}$ and $z_1 \in (T_k^{\mathit{halt}})^{\mathcal{M}}$, where $T = T^{1,M}$ for which $\mathit{up}(T) = \mathit{ceiling}$. We then use (30) and find objects $z_2, \dots, z_N, u, v$ such that $z_i \in (T_k^{\mathit{halt}})^{\mathcal{M}}$, where $T = T^{i,M}$, $u \in \mathit{Row}^{\mathcal{M}}$ and $v \in \mathit{End}^{\mathcal{M}}$. We take $x_0 = y_0, \dots, x_{n-N-1} = y_{n-N-1}, x_{n-N} = z_1, \dots, x_{n-1} = z_N, x_n = u, x_{n+1} = v$. Note that, by (28) and (33), we have $(T^{i,j})^{\mathcal{M}} \subseteq (\widehat{T}^{i,j-1})^{\mathcal{M}}$.

($\Rightarrow$) Let $\boldsymbol{q}_n(X)$ be such that $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n(a)$, by Proposition 5 it follows $\mathcal{M} \models \boldsymbol{q}_n$ for each $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$. Consider all the pairwise distinct pairs $(\mathcal{M}, h)$ such that $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$ and h a homomorphism from $\boldsymbol{q}$ to $\mathcal{M}$. Note that $h(\boldsymbol{q})$ contains an or-node σ_h (which is an instance of $\mathit{Row}_k^{\mathit{halt}}$, for some k). We call $(\mathcal{M}, h)$ and h *left* if $h(x_{n+1}) = \sigma_h \cdot w_{\exists R.\mathit{End}}$, and *right* otherwise. It is not hard to see that there exist a left $(\mathcal{M}_\ell, h_\ell)$ and a right $(\mathcal{M}_r, h_r)$ with $\sigma_{h_\ell} = \sigma_{h_r}$ (if this is not the case, we can construct $\mathcal{M} \in \boldsymbol{M}_{\mathcal{K}_2}$ such that $\mathcal{M} \not\models \boldsymbol{q}$).

Take $(\mathcal{M}_\ell, h_\ell)$ and $(\mathcal{M}_r, h_r)$ such that $\sigma_{h_\ell} = \sigma_{h_r} = \sigma$ and use them to construct the required tiling. Let $\sigma = aw_0 \cdots w_{n'}$. We have $h_\ell(x_n) = \sigma$, $h_\ell(x_{n+1}) = \sigma \cdot w_{\exists R.\mathit{End}}$. Let $h_r(x_{n+1}) = \sigma v_1 \dots v_{m+2}$, which is an instance of *End*. Then $h_r(x_n) = \sigma v_1 \dots v_{m+1}$, which is an instance of *Row*. Suppose $v_m = w_{\exists R.T_2^{\mathit{halt}}}$ (other k's are treated analogously). By (31), $\mathit{right}(T) = \mathit{wall}$; by (30), $\mathit{up}(T) = \mathit{ceiling}$. Suppose $w_{n'-1} = w_{\exists R.S_k}$. Now, we know that $k = 1$. By (25), $\mathit{right}(S) = \mathit{wall}$. Consider the atom $B_{n-1}(x_{n-1})$ from $\boldsymbol{q}$. Then both $aw_0 \cdots w_{n'-1}$ and $\sigma v_1 \cdots v_m$ are instances of B_{n-1}. By (27) and (33), $B_{n-1} = \widehat{S}_1$ and $\mathit{down}(T) = \mathit{up}(S)$.

Suppose $v_{m-1} = w_{\exists R.U_2^{\mathit{halt}}}$. By (30), $\mathit{right}(U) = \mathit{left}(T)$ and $\mathit{up}(U) = \mathit{ceiling}$. Suppose $w_{n'-2} = w_{\exists R.Q_1}$. By (23), $\mathit{right}(Q) = \mathit{left}(S)$. Consider the atom $B_{n-2}(x_{n-2})$ from $\boldsymbol{q}$. Then both $aw_0 \cdots w_{n'-2}$ and $\sigma \cdots v_{m-1}$ are instances of B_{n-2}. By (27) and (33), $B_{n-2} = \widehat{Q}_1$ and $\mathit{down}(U) = \mathit{up}(Q)$.

We proceed in the same way until we reach σ and $aw_0 \cdots w_{n'-N-1}$, for $N = m$, both of which are instances of $B_{n-N-1} = \mathit{Row}$. Thus we have tiled the last two rows of the grid. Let us proceed in that fashion until we have reached some variable x_t, for $t \geq 0$, of $\boldsymbol{q}$ that is mapped by h_ℓ to aw_0w_1 (see Fig. 6). Note that this situation is guaranteed to occur. Indeed, $h_\ell(a) = a$, $h_\ell(x_0) \in \{a, aw_0\}$, $h_\ell(x_1) \in \{a, aw_0, aw_0w_1\}$ etc. Clearly, assuming $h_\ell(x_i) \in \{a, aw_0\}$ for all $0 \leq i \leq n+1$ produces a contradiction.

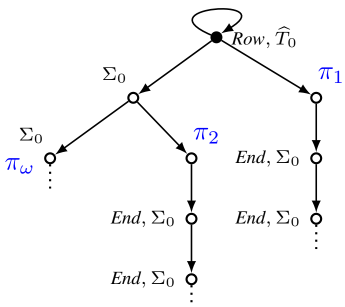
Let $h_r(x_t) = aw_0 \cdots w_s$ for some $s > 1$ and note that $s = N + 2$. By (21), it follows that aw_0w_1 is an instance of I_0 therefore $B_t = \hat{I}_0$ and, by (28), we also get that $aw_0 \cdots w_s$ is an instance of V_1 for some tile V such that $\mathit{down}(V) = \mathit{up}(I)$. Thus, we have the tiling as required since the vertical and horizontal compatibility of the tiles is ensured by the construction above and by the fact that the tile I occurs in it as the initial tile. $\square$

Let $\Sigma_0 = \{\mathit{Row}\} \cup \{\hat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$. Set $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{T}_1$ to contain the following axioms:

$$A \sqsubseteq \exists R.D \sqcap \exists R.\exists R.E \sqcap \bigsqcap_{X \in \Sigma_0} X, \tag{34}$$

$$E \sqsubseteq \exists R.E \sqcap \bigsqcap_{X \in \Sigma_0} X \sqcap \mathit{End}. \tag{35}$$

The canonical model $\mathcal{M}_{\mathcal{K}_1}$ of $\mathcal{K}_1$ is as follows:



As before, let $\Sigma = \mathsf{sig}(\mathcal{K}_1)$.

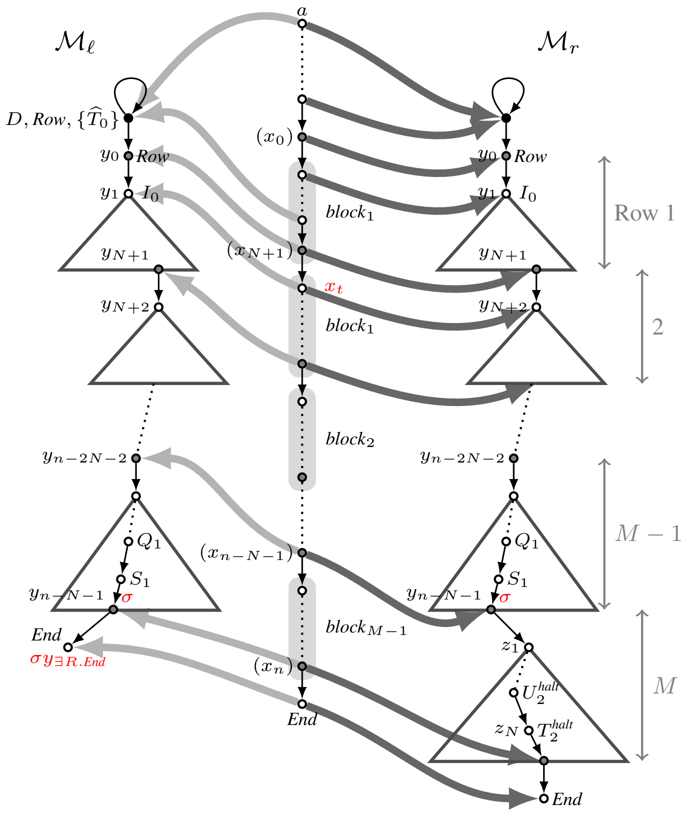


Figure 6: The two homomorphisms to two minimal models

Lemma 21. *$\prod \boldsymbol{M}_{\mathcal{K}_2}$ is $n\Sigma$-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$ for any n iff there does not exist a CQ $\boldsymbol{q}_n$ such that $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n$.*

Proof. ($\Rightarrow$) Suppose $\prod \boldsymbol{M}_{\mathcal{K}_2} \models \boldsymbol{q}_n(a)$ for some n. Since $\prod \boldsymbol{M}_{\mathcal{K}_2}$ is $n\Sigma$-homomorphically embeddable into $\mathcal{M}_{\mathcal{K}_1}$, we then have $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}_n(a)$, which is clearly impossible because of the B_i and *End* in $\boldsymbol{q}_n$.

($\Leftarrow$) Suppose $\prod \boldsymbol{M}_{\mathcal{K}_2} \not\models \boldsymbol{q}_n(a)$ for all n. Take any subinterpretation of $\prod \boldsymbol{M}_{\mathcal{K}_2}$ whose domain contains m elements. We can regard this subinterpretation as a Boolean Σ-CQ, and so denote it by $\boldsymbol{q}$. Without loss of generality we can assume that $\boldsymbol{q}$ is connected; clearly, $\boldsymbol{q}$ is either:

(i) tree shaped with a root different from a,

(ii) tree shaped rooted in a and containing a loop $R(a, a)$

We know that there is no Σ-homomorphism from $\boldsymbol{q}_n$ into $\boldsymbol{q}$ for any n; in particular, $\boldsymbol{q}$ does not have a subquery of the form $\boldsymbol{q}_n$. We have to show that $\mathcal{M}_{\mathcal{K}_1} \models \boldsymbol{q}$.

If **(i)** holds we map $\boldsymbol{q}$ to the branch π_1 in the obvious way. Suppose, **(ii)** holds. We will show how to map $\boldsymbol{q}$ starting from a. We call a variable x in $\boldsymbol{q}$ a *gap* if there exists no $A \in \Sigma$ such that $A(x)$ is in $\boldsymbol{q}$. By the condition of the lemma we know that every path ρ in $\boldsymbol{q}$ either:

(a) does not contain *End*(x), or

(b) contains *End*(x) and contains a gap y that occurs between the root a and x

For the paths ρ of type **(b)** let t_ρ be the minimal distance from the root a to a gap of the path ρ. Denote by $\mathcal{R}$ the set of all path ρ of $\boldsymbol{q}$. If all $\rho \in \mathcal{R}$ are of type **(a)** we map $\boldsymbol{q}$ on the path π_ω. Otherwise, let t_0 be the minimal number of all the t_ρ (that are defined) and $\mathcal{R}_{t_0}$ the set of paths ρ such that $t_\rho = t_0$. We map all the path of $\mathcal{R}_{t_0}$ to the path π_{t_0} of $\mathcal{M}_{\mathcal{K}_1}$. For the rest $\mathcal{R} \setminus \mathcal{R}_{t_0}$ we find again the minimal number t_1 of all the t_ρ for $\rho \in \mathcal{R} \setminus \mathcal{R}_{t_0}$ and denote by $\mathcal{R}_{t_1}$ the set of paths ρ such that $t_\rho = t_1$. Clearly, can map all the paths in $\mathcal{R}_{t_1}$ to π_{t_1}. We continue in that way for sufficiently many steps to map all the paths of $\mathcal{R}$. $\square$

We now obtain Theorem 7 (*i*) and (*ii*) for rCQs in the same way as in the previous section.

A.4 Proof of Theorem 7 (*iii*)

To prove undecidability results if separating CQs can have arbitrary symbols we modify the KBs introduced above. We follow [27] and replace the non-Σ-symbols by complex $\mathcal{ALC}$-concepts that, in contrast to concept names, cannot occur in CQs. In detail, consider a set Σ_{hide} of concept names and take a fresh concept name Z_B and fresh role names r_B and s_B for every $B \in \Sigma_{\mathsf{hide}}$. Now let for each $B \in \Sigma_{\mathsf{hide}}$

$$H_B = \forall r_B.\exists s_B.\neg Z_B$$

and set

$$\mathcal{T}_{\Sigma_{\mathsf{hide}}} = \{\top \sqsubseteq \exists r_B.\top, \top \sqsubseteq \exists s_B.Z_B \mid B \in \Sigma_{\mathsf{hide}}\}$$

Note that $\mathcal{T}_{\Sigma_{\mathsf{hide}}}$ is an $\mathcal{EL}$ TBox that generates trees with edges r_B and s_B such that the s_Z-successors satisfy Z_B. One can satisfy H_B in a certain node by introducing in addition to the s_B-successors satisfying Z_B other s_B-successors not satisfying Z_B. Those additional s_B-successors will not influence the answers to CQs. We now summarize the main properties of $\mathcal{T}_{\Sigma_{\mathsf{hide}}}$ in a formal way. For an ABox $\mathcal{A}$ and any set $p(\Sigma_{\mathsf{hide}}) = \{J_B \mid B \in \Sigma_{\mathsf{hide}}\}$ with $J_B \subseteq \mathsf{ind}(\mathcal{A})$ for all $B \in \Sigma_{\mathsf{hide}}$ construct a model $\mathcal{I}$ as follows: $\Delta^{\mathcal{I}}$ is the set of words $w = av_1 \cdots v_n$ such that $a \in \mathsf{ind}(\mathcal{A})$ and $v_i \in \{r_B, s_B, \bar{s}_B \mid B \in \Sigma_{\mathsf{hide}}\}$ where $v_i \neq \bar{s}_B$ if (i) $i > 2$ or (ii) $i = 2$ and ($a \notin J_B$ or $v_1 \neq r_B$). For all concept names A not of the form Z_B set

$$A^{\mathcal{I}} = \{a \mid A(a) \in \mathcal{A}\}$$

Let for $B \in \Sigma_{\mathsf{hide}}$:

$$Z_B^{\mathcal{I}} = \{a \mid Z_B(a) \in \mathcal{A}\} \cup \{w \mid \mathsf{tail}(w) = s_B\}$$

where $\mathsf{tail}(w)$ is the last symbol in w. For all role names R not of the form r_B or s_B set

$$R^{\mathcal{I}} = \{(a, b) \mid R(a, b) \in \mathcal{A}\}$$

Finally, let for $B \in \Sigma_{\mathsf{hide}}$:

$$\begin{aligned} r_B^{\mathcal{I}} &= \{(a, b) \mid r_B(a, b) \in \mathcal{A}\} \cup \{(w, wr_B) \mid wr_B \in \Delta^{\mathcal{I}}\} \\ s_B^{\mathcal{I}} &= \{(a, b) \mid s_B(a, b) \in \mathcal{A}\} \cup \{(w, ws_B) \mid wr_B \in \Delta^{\mathcal{I}}\} \cup \{(w, w\bar{s}_B) \mid w\bar{s}_B \in \Delta^{\mathcal{I}}\}. \end{aligned}$$

The following result summarizes the main properties of $\mathcal{I}$ [27].

Lemma 22. *The following holds for every $\mathcal{A}$ and $p(\Sigma_{\mathsf{hide}})$:*

- *$\mathcal{I}$ is a model of $\mathcal{T}_{\Sigma_{\mathsf{hide}}}$ and $\mathcal{A}$;*
- *$J_B = (H_B)^{\mathcal{I}}$ for all $B \in \Sigma_{\mathsf{hide}}$;*
- *for every CQ $q(\vec{x})$ and $\vec{a}$ in $\mathsf{ind}(\mathcal{A})$:* $\quad \mathcal{T}_{\Sigma_{\mathsf{hide}}}, \mathcal{A} \models q(\vec{a}) \quad \Leftrightarrow \quad \mathcal{I} \models q(\vec{a})$

A *hiding scheme* $\mathcal{H}$ consists of three sets of concept names, Σ_{in}, Σ_{out}, and Σ_{hide}. Let $C^{\Sigma_{\mathsf{hide}}}$ be the result of replacing in a concept C every $B \in \Sigma_{\mathsf{hide}}$ by H_B. For a given TBox $\mathcal{T}$ we denote by $\mathcal{T}^{\mathcal{H}}$ the TBox containing $\mathcal{T}_{\Sigma_{\mathsf{hide}}}$ and the following CIs:

- $A \sqsubseteq H_A$, for $A \in \Sigma_{\mathsf{in}}$;
- $C^{\Sigma_{\mathsf{hide}}} \sqsubseteq D^{\Sigma_{\mathsf{hide}}}$, for all $C \sqsubseteq D \in \mathcal{T}$;
- $H_A \sqsubseteq A$, for all $A \in \Sigma_{\mathsf{out}}$.

A TBox $\mathcal{T}$ *admits trivial models* if the singleton interpretation in which all concept and role names are interpreted by the empty set is a model of $\mathcal{T}$. We consider TBoxes that admit trivial models since for such TBoxes the nodes generated by $\mathcal{T}_{\Sigma_{\mathsf{hide}}}$ trivially satisfy $\mathcal{T}$. Oberve that the TBoxes constructed in the undecidability proofs above all admit trivial models.

Theorem 23. *The problem whether a Horn-$\mathcal{ALC}$ KB full signature-CQ entails an $\mathcal{ALC}$ KB is undecidable.*

Proof. We consider the KBs $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$ and $\Sigma = \mathsf{sig}(\mathcal{K}_1)$ constructed in the proof of Theorem 7 (*i*) for Σ-CQ-entailment.

Define a hiding scheme $\mathcal{H}$ by setting

- $\Sigma_{\mathsf{in}} = \mathsf{sig}(\mathcal{A}) = \{A\}$;
- Σ_{out} is the set of concept names in Σ;
- $\Sigma_{\mathsf{hide}} = \mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2)$.

Define new KBs as follows: $\mathcal{K}_1' = (\mathcal{T}_1 \cup \mathcal{T}_{\Sigma_{\mathsf{hide}}}, \mathcal{A})$, $\mathcal{K}_2' = (\mathcal{T}_2^{\mathcal{H}}, \mathcal{A})$. Using the facts that

- $\mathsf{sig}(\mathcal{A}) \subseteq \Sigma$;
- all role names in $\mathcal{K}_2$ are contained in Σ;
- $\mathcal{T}_1$ and $\mathcal{T}_2$ admit trivial models

it is straightforward to check that $\mathcal{K}_1$ Σ-CQ-entails $\mathcal{K}_2$ iff $\mathcal{K}_1'$ full signature CQ-entails $\mathcal{K}_2'$ □

Theorem 24. *The problem whether a* Horn-$\mathcal{ALC}$ *KB is full signature-CQ inseparable from an* $\mathcal{ALC}$ *KB is undecidable.*

Proof. We consider the KBs $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2' = \mathcal{K}_1 \cup \mathcal{K}_2$ and the signature $\Sigma = \mathsf{sig}(\mathcal{K}_1)$ constructed in the proof of Theorem 7 (*ii*) for Σ-CQ-inseparability. Assume $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$.

Consider the same hiding scheme $\mathcal{H}$ as in the proof of Theorem 23:

- $\Sigma_{\mathsf{in}} = \mathsf{sig}(\mathcal{A}) = \{A\}$;
- Σ_{out} is the set of concept names in Σ;
- $\Sigma_{\mathsf{hide}} = \mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2)$.

Define new KBs $\mathcal{K}_1^*$ and $\mathcal{K}_2^*$ as follows: $\mathcal{K}_1^* = (\mathcal{T}_1 \cup \mathcal{T}_{\Sigma_{\mathsf{hide}}}, \mathcal{A})$, $\mathcal{K}_2^* = (\mathcal{T}_1 \cup \mathcal{T}_2^{\mathcal{H}}, \mathcal{A})$. Using the facts that

- $\mathsf{sig}(\mathcal{K}_1) \subseteq \Sigma$;
- all role names in $\mathcal{K}_1 \cup \mathcal{K}_2$ are contained in Σ;
- $\mathcal{T}_1$ and $\mathcal{T}_2$ admit trivial models

it is straightforward to check that $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ-CQ-inseparable iff $\mathcal{K}_1^*$ and $\mathcal{K}_2^*$ are full signature CQ-inseparable. □

Theorem 25. *The problem whether a* Horn-$\mathcal{ALC}$ *KB full signature-rCQ entails an* $\mathcal{ALC}$ *KB is undecidable.*

Proof. We consider the KBs $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$ and $\Sigma = \mathsf{sig}(\mathcal{K}_1)$ constructed in the proof of Theorem 7 (*i*) for Σ-rCQ-entailment.

Define a hiding scheme $\mathcal{H}$ by setting

- $\Sigma_{\mathsf{in}} = \mathsf{sig}(\mathcal{A}) = \{R, Row, A\} \cup \{\widehat{T}_0 \mid T \in \mathfrak{T}\}$;
- Σ_{out} is the set of concept names in Σ;
- $\Sigma_{\mathsf{hide}} = \mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2)$.

Define new KBs as follows: $\mathcal{K}_1' = (\mathcal{T}_1 \cup \mathcal{T}_{\Sigma_{\mathsf{hide}}}, \mathcal{A})$, $\mathcal{K}_2' = (\mathcal{T}_2^{\mathcal{H}}, \mathcal{A})$. Using the facts that

- $\mathsf{sig}(\mathcal{A}) \subseteq \Sigma$;
- all role names in $\mathcal{K}_2$ are contained in Σ;
- $\mathcal{T}_1$ and $\mathcal{T}_2$ admit trivial models

it is straightforward to check that $\mathcal{K}_1$ Σ-rCQ-entails $\mathcal{K}_2$ iff $\mathcal{K}_1'$ full signature rCQ-entails $\mathcal{K}_2'$ □

The proof of the following results is now similar to the proof of Theorem 24 using the KBs constructed in the proof of Theorem 7 (*ii*) for Σ-rCQ-inseparability.

Theorem 26. *The problem whether a* Horn-$\mathcal{ALC}$ *KB is full signature-rCQ inseparable from an* $\mathcal{ALC}$ *KB is undecidable.*

B Proof of Theorem 9

B.1 Proof of Theorem 9 (*i*) and (*ii*) for CQs

We formulate the result again.

Theorem 27. *Let* $\Theta = (\Sigma_1, \Sigma_2)$.
(i) *The problem of whether a* Horn-$\mathcal{ALC}$ *TBox* Θ-*CQ-entails an* $\mathcal{ALC}$ *TBox is undecidable.*
(ii) Θ-*CQ inseparability between* Horn-$\mathcal{ALC}$ *TBoxes and* $\mathcal{ALC}$ *TBoxes is undecidable.*
(iii) Θ-*CQ inseparability between* Horn-$\mathcal{ALC}$ *TBoxes and* $\mathcal{ALC}$ *TBoxes is undecidable for* $\Sigma_1 = \Sigma_2$.

Proof. We prove (*i*). The proof of (*ii*) is similar. Let $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$ be the KBs and Σ be the signature from the proof of Theorem 7 (*i*) for Σ-CQ-entailment. Recall that $\mathcal{A} = \{A(a)\}$. Let $\Sigma_1 = \{A\}$, $\Sigma_2 = \Sigma$, and $\Theta = (\Sigma_1, \Sigma_2)$. We claim that $\mathcal{T}_1$ Θ-CQ-entails $\mathcal{T}_2$ iff $\mathcal{K}_1$ Σ-CQ-entails $\mathcal{K}_2$. Clearly, if $\mathcal{K}_1$ does not Σ-CQ-entail $\mathcal{K}_2$, then we have found a Σ_1-ABox $\mathcal{A}$ that witnesses that $\mathcal{T}_1$ does not Θ CQ-entail $\mathcal{T}_2$. Conversely, observe that all Σ_1-ABoxes $\mathcal{A}'$ are sets of assertions of the form $A(b)$ and so if any such $\mathcal{A}'$ provides a counterexample for Θ-CQ-entailment between $\mathcal{T}_1$ and $\mathcal{T}_2$, then $\mathcal{A}$ does.

We now prove (*iii*). Consider $\mathcal{K}_1$ and $\mathcal{K}_2' = (\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A})$ from the proof of Theorem 7 (*ii*) for Σ-CQ-inseparability. Now let

$$\Sigma = \{A, R, \mathit{Row}, \mathit{End}, \mathit{Start}\} \cup \{\hat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$$

Then one can show that $\mathcal{T}_1$ and $\mathcal{T}_1 \cup \mathcal{T}_2$ are (Σ, Σ)-CQ-inseparable iff $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ-CQ-inseparable. The latter is undecidable. □

B.2 Proof of Theorem 9 for full ABox signature and CQs

We now aim to extend the result above to the full ABox signature case and inseparability.

Theorem 28. (i) *The problem of whether a* Horn-$\mathcal{ALC}$ *TBox full ABox signature* Σ-*CQ-entails an* $\mathcal{ALC}$ *TBox is undecidable.*
(ii) *Full ABox signature* Σ-*CQ inseparability between* Horn-$\mathcal{ALC}$ *TBoxes and* $\mathcal{ALC}$ *TBoxes is undecidable.*

Proof. We consider the inseparability case. Let $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2' = (\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A})$ be the KBs and $\Sigma = \mathsf{sig}(\mathcal{K}_1)$ be the signature from the proof of Theorem 7 (*ii*) for Σ-CQ-inseparability between KBs. We set

$$\Sigma_0 = \{R, \mathit{Row}, \mathit{End}, \mathit{Start}\} \cup \{\hat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$$

Observe that for any signature Γ between Σ_0 and $\Sigma_0 \cup \mathsf{sig}(\mathcal{K}_1)$, the KBs $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Γ-CQ-inseparable iff they are Σ_0-CQ-inseparable. We construct TBoxes $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$ from the TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ such that full ABox signature Σ_0-CQ-inseparability between $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$ is undecidable. To this end define a hiding scheme $\mathcal{H}$ by setting

- $\Sigma_{\mathsf{in}} = \{A\}$;
- Σ_{out} is the set of concept names in Σ_0;
- $\Sigma_{\mathsf{hide}} = \mathsf{sig}(\mathcal{K}_2)$.

Define TBoxes $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$ by setting

$$\mathcal{T}_1^* = \mathcal{T}_1 \cup \mathcal{T}_{\Sigma_{\mathsf{hide}}}, \quad \mathcal{T}_2^* = \mathcal{T}_1 \cup \mathcal{T}_2^{\mathcal{H}}$$

Now one can prove that $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ-CQ-inseparable iff $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$ are full ABox signature Σ_0-CQ-inseparable. The direction from right to left is trivial as we can take the ABox $\mathcal{A}$ as a witness separating $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$ if $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ_0-CQ-separable. For the converse direction assume that an ABox $\mathcal{A}'$ Σ_0-CQ-separates $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$. As $P \not\in \Sigma_0$ one can then prove that there exists $A(b) \in \mathcal{A}'$ such that $\{A(b)\}$ is an ABox that separates $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$. But then $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ_0-CQ-separable as well. □

B.3 Proof of Theorem 9 (*i*) and (*ii*) for rCQs

We state the result again.

Theorem 29. *(i) The problem of whether a Horn-$\mathcal{ALC}$ TBox Θ-rCQ-entails an $\mathcal{ALC}$ TBox is undecidable.*
(ii) Θ-rCQ inseparability between Horn-$\mathcal{ALC}$ TBoxes and $\mathcal{ALC}$ TBoxes is undecidable.

Proof. We consider the inseparability case. Let $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2' = (\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A})$ be the KBs from the proof of Theorem 7 (*ii*) for Σ-rCQ-inseparability between KBs. Let $\Sigma_1 = \mathsf{sig}(\mathcal{A})$ and

$$\Sigma_2 = \{R, \mathit{Row}, \mathit{End}\} \cup \{\hat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$$

and let $\Theta = (\Sigma_1, \Sigma_2)$. Define $\mathcal{T}_2' = \mathcal{T}_1 \cup \mathcal{T}_2$. It is sufficient to show that $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ_2-rCQ-inseparable iff $\mathcal{T}_1$ and $\mathcal{T}_2'$ are Θ-rCQ-inseparable. The direction from right to left is trivial as we can use the ABox $\mathcal{A}$ as a witness ABox for Θ-rCQ-separability between $\mathcal{T}_1$ and $\mathcal{T}_2'$ if $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ_2-rCQ-separable. Conversely, assume there is a Σ_1-ABox $\mathcal{A}'$ which Θ-rCQ-separates $\mathcal{T}_1$ and $\mathcal{T}_2'$. Using the fact that $\mathit{End} \notin \Sigma_1$ and that

$$A \sqsubseteq \exists R.(\mathit{Row} \sqcap \exists R.I_0)$$

and

$$A \sqsubseteq \exists R.D \ \sqcap \ \exists R.\exists R.E \ \sqcap \ \bigsqcap_{X \in \Sigma_0} X$$

are the only concept inclusions in $\mathcal{T}_1 \cup \mathcal{T}_2$ that generate new R-successors from ABox individuals one can now readily show that $\mathcal{K}_1$ and $\mathcal{K}_2'$ are Σ_2-rCQ–separable. □

C Proof of Theorem 10

We give a more detailed proof of Theorem 10 which we state again for the convenience of the reader.

Theorem 10. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be $\mathcal{ALC}$ KBs, Σ a signature, and let $\boldsymbol{M}_1$ be complete for $\mathcal{K}_1$. Then $\mathcal{K}_1$ Σ-rUCQ entails $\mathcal{K}_2$ iff for any $\mathcal{I}_1 \in \boldsymbol{M}_1$, there exists $\mathcal{I}_2 \models \mathcal{K}_2$ such that $\mathcal{I}_2$ is conΣ-homomorphically embeddable into $\mathcal{I}_1$.*

Proof. In view of Theorem 6 (2), it suffices to prove ($\Rightarrow$). Suppose $\mathcal{I}_1 \in \boldsymbol{M}_1$. Since there is $\mathcal{I}_1' \in \boldsymbol{M}^{fo}_{\mathcal{K}_1}$ that maps Σ-homomorphically to $\mathcal{I}_1$, we can assume $\mathcal{I}_1 \in \boldsymbol{M}^{fo}_{\mathcal{K}_1}$. By Theorem 6 (2),

($*$) for any $n > 0$ there exists a model $\mathcal{J} \in \boldsymbol{M}^{fo}_{\mathcal{K}_2}$ that is con-$n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$.

Denote by $\mathcal{J}_{|\leq n}$ the subinterpretation of $\mathcal{J}$ whose elements are connected to ABox individuals by Σ-paths of length $\leq n$. A (Σ, n)-homomorphism h from $\mathcal{J}$ is a Σ-homomorphism with domain $\mathcal{J}_{|\leq n}$. Let Ξ_n be the class of $\mathcal{J} \in \boldsymbol{M}^{fo}_{\mathcal{K}_2}$ such that there is a (Σ, n)-homomorphism h from $\mathcal{J}$ to $\mathcal{I}_1$. By ($*$) all Ξ_n are non-empty. We may assume that for $\mathcal{I}, \mathcal{J} \in \Xi_n$ we have $\mathcal{I}_{|\leq n} = \mathcal{J}_{|\leq n}$ if $\mathcal{I}_{|\leq n}$ and $\mathcal{J}_{|\leq n}$ are isomorphic. For every $n \geq 0$, we define classes X_n of pairs $(\mathcal{J}, h)$ such that $\mathcal{J} \in \Xi_n$ and h is a (Σ, m)-homomorphism from $\mathcal{J}$ to $\mathcal{I}_1$ for some $m \geq n$ such that the following conditions hold:

- for every $m > n$ there exists $(\mathcal{J}, h) \in X_n$ such that h is a (Σ, m)-homomorphism from $\mathcal{J}$ to $\mathcal{I}_1$;
- $\mathcal{I}_{|\leq n} = \mathcal{J}_{|\leq n}$ and $h_{|\leq n} = f_{|\leq n}$ for all $(\mathcal{I}, h), (\mathcal{J}, f) \in X_n$ ($h_{|\leq n}$ denotes the restriction of h to $\mathcal{I}_{|\leq n}$).

Let X_0 be the set of all pairs $(\mathcal{J}, h)$ such that $\mathcal{J} \in \Xi_0$ and h is a (Σ, n)-homomorphism from $\mathcal{J}$ into $\mathcal{I}_1$ for some $n \geq 0$. Our assumptions directly imply that X_0 has the properties above since $h(a) = a$ holds for every homomorphism h and all ABox individuals a. Suppose that X_n has been defined. It follows from the bounded outdegree of $\mathcal{I}_1$ and all $\mathcal{J} \in \Xi_n$ that we can construct X_{n+1} with the required properties in a straightforward way.

We now define an interpretation $\mathcal{J}$ with a Σ-homomorphism h as follows:

$$\begin{aligned} \mathcal{J} &= \bigcup_{n<\omega} \{\mathcal{J}_{|\leq n} \mid \exists h \ (\mathcal{J}, h) \in X_n\} \\ h &= \bigcup_{n<\omega} \{h_{|\leq n} \mid \exists \mathcal{J} \ (\mathcal{J}, h) \in X_n\} \end{aligned}$$

It is straightforward to show that $\mathcal{J}$ is a model of $\mathcal{K}_2$ and h is a conΣ-homomorphism from $\mathcal{J}$ into $\mathcal{I}_1$, as required. □

D Proof of Theorem 11

We aim to prove that it is in 2ExpTime to decide whether an $\mathcal{ALC}$ KB $\mathcal{K}_1$ Σ-rUCQ entails an $\mathcal{ALC}$ KB $\mathcal{K}_2$.

D.1 Tree Automata Preliminaries

We introduce two-way alternating automata on infinite trees (2ATAs). Let $\mathbb{N}$ denote the *positive* integers. A *tree* is a non-empty (and potentially infinite) set $T \subseteq \mathbb{N}^*$ closed under prefixes. The node ε is the *root* of T. As a convention, we take $x \cdot 0 = x$ and $(x \cdot i) \cdot -1 = x$. Note that $\varepsilon \cdot -1$ is undefined. We say that T is m*-ary* if for every $x \in T$, the set $\{i \mid x \cdot i \in T\}$ is of cardinality exactly m. W.l.o.g., we assume that all nodes in an m-ary tree are from $\{1, \ldots, m\}^*$.

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

Definition 30 (2ATA). *A* two-way alternating automaton (2ATA) on infinite m-ary trees *is a tuple* $\mathfrak{A} = (Q, \Gamma, \delta, q_0)$ *where Q is a finite set of* states, *Γ is a finite alphabet, $\delta : Q \times \Gamma \to \mathcal{B}^+(\mathsf{tran}(\mathfrak{A}))$ is the* transition function *with* $\mathsf{tran}(\mathfrak{A}) = [m] \times Q$ *the set of* transitions *of* $\mathfrak{A}$, *and $q_0 \in Q$ is the* initial state.

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

Definition 31 (Run, Acceptance). *A* run *of a 2ATA $\mathfrak{A} = (Q, \Gamma, \delta, q_0, R)$ on an infinite Γ-labeled tree (T, L) is a $T \times Q$-labeled tree (T_r, r) such that the following conditions are satisfied:*

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

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

We will use the following results from automata theory:

Theorem 32.

1. *Given a 2ATA, we can construct in polynomial time a 2ATA that accepts the complement language;*
2. *Given a constant number of 2ATAs, we can construct in polytime a 2ATA (resp. an NTA) that accepts the intersection language;*
3. *Emptiness of 2ATAs can be checked in single exponential time in the number of states.*

D.2 Γ-labeled Trees

We aim to prove Theorem 11. To this end, fix $\mathcal{ALC}$ KBs $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A}_1)$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A}_2)$, and a signature Σ. By Theorem 10, we can decide whether $\mathcal{K}_2$ is not Σ-rUCQ entailed by $\mathcal{K}_1$ by checking that there is a model $\mathcal{I}_1 \in \boldsymbol{M}^{fo}_{\mathcal{K}_1}$ into which no model $\mathcal{I}_2$ of $\mathcal{K}_2$ is con-Σ_2-homomorphically embeddable.

In the following, we construct a 2ATA $\mathfrak{A}$ that accepts (suitable representations of) the desired models $\mathcal{I}_1$, and for deciding Σ-rUCQ entailment of $\mathcal{K}_2$ by $\mathcal{K}_1$, it then remains to check emptiness. We start with encoding forest-shaped interpretations as labeled trees. For $i \in \{1, 2\}$, we use $\mathsf{CN}(\mathcal{T}_i)$ and $\mathsf{RN}(\mathcal{T}_i)$ to denote the set of concept names and role names in $\mathcal{T}_i$, respectively. Node labels are taken from the alphabet

$$\Gamma \quad = \quad \{\mathit{root}, \mathit{empty}\} \cup (\mathsf{ind}(\mathcal{A}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}) \cup (\mathsf{RN}(\mathcal{T}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}).$$

The $\boldsymbol{M}^{fo}_{\mathcal{K}_1}$ models will be represented as m-ary Γ-labeled trees, with $m = \max(|\mathcal{T}_1|, |\mathsf{ind}(\mathcal{K}_1)|)$. The root node is not used in the representation and receives label root. Each ABox individual is represented by a successor of the root labeled with a symbol from $\mathsf{ind}(\mathcal{A}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}$; anonymous elements are represented by nodes deeper in the tree labeled with a symbol from $\mathsf{RN}(\mathcal{T}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}$. The node label $empty$ is used for padding to achieve that every tree node has exactly m successors. We call a Γ-labeled tree *proper* if it satisfies the following conditions:

- the root is labeled with $root$;
- for every $a \in \mathsf{ind}(\mathcal{A}_1)$, there is exactly one successor of the root that is labeled with a symbol from $\{a\} \times 2^{\mathsf{CN}(\mathcal{T}_1)}$; all remaining successors of the root are labeled with $empty$;
- all other nodes are labeled with a symbol from $\mathsf{RN}(\mathcal{T}_1) \times 2^{\mathsf{CN}(\mathcal{T}_1)}$ or with $empty$;
- if a node is labeled with $empty$, then so are all its successors.

A proper Γ-labeled tree (T, L) represents the following interpretation $\mathcal{I}_{(T,L)}$:

$$
\begin{aligned}
\Delta^{\mathcal{I}_{(T,L)}} &= \mathsf{ind}(\mathcal{A}_1) \cup \{x \in T \mid |x| > 1\} \\
A^{\mathcal{I}_{(T,L)}} &= \{a \mid \exists x \in T : L(x) = (a, \boldsymbol{t}) \text{ with } A \in \boldsymbol{t}\} \cup \{x \in T \mid L(x) = (R, \boldsymbol{t}) \text{ with } A \in \boldsymbol{t}\} \\
R^{\mathcal{I}_{(T,L)}} &= \{(a, b) \mid R(a, b) \in \mathcal{A}_1\} \cup \\
&\quad \{(a, ij) \mid ij \in T, L(i) = (a, \boldsymbol{t}_1), L(ij) = (R, \boldsymbol{t}_2)\} \cup \\
&\quad \{(x, xi) \mid xi \in T, L(x) = (S, \boldsymbol{t}_1), L(xi) = (R, \boldsymbol{t}_2)\}.
\end{aligned}
$$

Note that $\mathcal{I}_{(T,L)}$ satisfies all required conditions to qualify as a forest-shaped model of $\mathcal{T}_1$ (except that it need not satisfy $\mathcal{T}_1$), and that its outdegree is bounded by $|\mathcal{T}_1|$. Conversely, every forest-shaped model of $\mathcal{T}_1$ with outdegree bounded by $|\mathcal{T}_1|$ can be represented as a proper m-ary Γ-labeled tree.

D.3 The automata construction

The desired 2ATA $\mathfrak{A}$ is assembled from the following three automata:

- a 2ATA $\mathfrak{A}_0$ that accepts an m-ary Γ-labeled tree iff it is proper;
- a 2ATA $\mathfrak{A}_1$ that accepts a proper m-ary Γ-labeled tree (T, L) iff $\mathcal{I}_{(T,L)}$ is a model of $\mathcal{T}_1$;
- a 2ATA $\mathfrak{A}_2$ that accepts a proper m-ary Γ-labeled tree (T, L) iff there is a model $\mathcal{I}_2$ of $\mathcal{K}_2$ that is con-Σ-homomorphically embeddable into $\mathcal{I}_{(T,L)}$.

By what was said above, $\mathcal{K}_2$ is then Σ-rUCQ entailed by $\mathcal{K}_1$ iff $\mathcal{L}(\mathfrak{A}_0) \cap \mathcal{L}(\mathfrak{A}_1) \cap \overline{\mathcal{L}(\mathfrak{A}_2)} = \emptyset$. We thus define $\mathfrak{A}$ to be the intersection of $\mathfrak{A}_0$, $\mathfrak{A}_1$, and the complement of $\mathfrak{A}_2$.

The construction of $\mathfrak{A}_0$ is trivial, details are omitted.

The construction of $\mathfrak{A}_1$ is quite standard [11]. Let $C_{\mathcal{T}_1}$ be the negation normal form (NNF) of the concept

$$\bigsqcap_{C \sqsubseteq D \in \mathcal{T}_1} (\neg C \sqcup D)$$

and let $\mathsf{cl}(C_{\mathcal{T}_1})$ denote the set of subconcepts of $C_{\mathcal{T}_1}$, closed under single negation. Now, the 2ATA $\mathfrak{A}_1 = \langle Q, \Gamma, \delta, q_0 \rangle$ is defined by setting

$$Q = \{q_0, q_1, q_\emptyset\} \cup \{q^{a,C}, q^C, q^R, q^{\neg R} \mid a \in \mathsf{Ind}(\mathcal{A}_1), C \in \mathsf{cl}(C_{\mathcal{T}_1}), R \in \mathsf{RN}(\mathcal{T}_1)\}$$

and defining the transition function δ as follows:

$$\begin{aligned}
\delta(q_0, root) &= \bigwedge_{i=1}^{m}(i, q_1)\\
\delta(q_1, \ell) &= ((0, q_\emptyset) \vee (0, q^{C_{\mathcal{T}_1}})) \wedge \bigwedge_{i=1}^{m}(i, q_1)\\
\delta(q^{\exists R.C}, (a, U)) &= \bigvee_{i=1}^{m}((i, q^R) \wedge (i, q^C)) \vee \bigvee_{R(a,b)\in\mathcal{A}_1}(-1, q^{b,C})\\
\delta(q^{\forall R.C}, (a, U)) &= \bigwedge_{i=1}^{m}((i, q_\emptyset) \vee (i, q^{\neg R}) \vee (i, q^C)) \wedge \bigwedge_{R(a,b)\in\mathcal{A}_1}(-1, q^{b,C})\\
\delta(q^{a,C}, root) &= \bigvee_{i=1}^{m}(i, q^{a,C})\\
\delta(q^{a,C}, (a, U)) &= (0, q^C)\\
\delta(q^{\exists R.C}, (S, U)) &= \bigvee_{i=1}^{m}((i, q^R) \wedge (i, q^C))\\
\delta(q^{\forall R.C}, (S, U)) &= \bigwedge_{i=1}^{m}((i, q_\emptyset) \vee (i, q^{\neg R}) \vee (i, q^C))\\
\delta(q^{C\sqcap C'}, (x, U)) &= (0, q^C) \wedge (0, q^{C'})\\
\delta(q^{C\sqcup C'}, (x, U)) &= (0, q^C) \vee (0, q^{C'})\\
\delta(q^A, (x, U)) &= \mathsf{true},\ \text{if } A \in U\\
\delta(q^{\neg A}, (x, U)) &= \mathsf{true},\ \text{if } A \notin U\\
\delta(q^R, (R, U)) &= \mathsf{true}\\
\delta(q^{\neg R}, (S, U)) &= \mathsf{true},\ \text{if } R \neq S\\
\delta(q_\emptyset, empty) &= \mathsf{true}\\
\delta(q, \ell) &= \mathsf{false} \quad \text{for all other } q \in Q,\ \ell \in \Gamma.
\end{aligned}$$

Where x in the labels (x, U) stands for an individual a or for a role name S, and ℓ in the second transition is any label from Γ. It is standard to show that $\mathfrak{A}_1$ accepts the desired tree language.

For constructing $\mathfrak{A}_2$, we first introduce some preliminaries. We use $\mathsf{cl}(\mathcal{T}_2)$ to denote the set of subconcepts of (concepts in) $\mathcal{T}_2$, closed under single negation. For each interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ and $d \in \Delta^\mathcal{I}$, the *$\mathcal{T}_2$-type of d in $\mathcal{I}$*, denoted $\boldsymbol{t}^\mathcal{I}_{\mathcal{T}_2}(d)$, is defined as $\boldsymbol{t}^\mathcal{I}_{\mathcal{T}_2}(d) = \{C \in \mathsf{cl}(\mathcal{T}_2) \mid d \in C^\mathcal{I}\}$. A subset $\boldsymbol{t} \subseteq \mathsf{cl}(\mathcal{T}_2)$ is a *$\mathcal{T}_2$-type* if $\boldsymbol{t} = \boldsymbol{t}^\mathcal{I}_{\mathcal{T}_2}(d)$, for some model $\mathcal{I}$ of $\mathcal{T}_2$ and $d \in \Delta^\mathcal{I}$. With $\mathsf{type}(\mathcal{T}_2)$, we denote the set of all $\mathcal{T}_2$-types. Let $\boldsymbol{t}, \boldsymbol{t}' \in \mathsf{type}(\mathcal{T}_2)$. For $\exists R.C \in \boldsymbol{t}$, we say that $\boldsymbol{t}'$ is an *$\exists R.C$-witness for $\boldsymbol{t}$* if $C \in \boldsymbol{t}'$ and $\bigsqcap \boldsymbol{t} \sqcap \exists R.(\bigsqcap \boldsymbol{t}')$ is satisfiable w.r.t. $\mathcal{T}_2$. Denote by $\mathsf{succ}_{\exists R.C}(\boldsymbol{t})$ the set of all $\exists R.C$-witnesses for $\boldsymbol{t}$. A *completion of* $\mathcal{K}_2$ is a function $\tau\colon \mathsf{ind}(\mathcal{A}_2) \to \mathsf{type}(\mathcal{T}_2)$ such that, for any $a \in \mathsf{ind}(\mathcal{A}_2)$, the KB

$$\big(\mathcal{T}_2 \cup \bigcup_{a\in\mathsf{ind}(\mathcal{A}_2), C\in\tau(a)} A_a \sqsubseteq C,\ \mathcal{A} \cup \bigcup_{a\in\mathsf{ind}(\mathcal{A}_2)} A_a(a)\big)$$

is consistent, where A_a is a fresh concept name for each $a \in \mathsf{ind}(A_2)$. Denote by $\mathsf{compl}(\mathcal{K}_2)$ the set of all completions of $\mathcal{K}_2$; it can be computed in exponential time in $|\mathcal{K}_2|$.

We now construct the 2ATA $\mathfrak{A}_2$. It is easy to see that if there is an assertion $R(a, b) \in \mathcal{A}_2 \setminus \mathcal{A}_1$ with $R \in \Sigma$, then no model of $\mathcal{K}_2$ is con-Σ_2-homomorphically embeddable into a forest-shaped model of $\mathcal{K}_1$. In this case, we choose $\mathfrak{A}_2$ so that it accepts the empty language.

Now assume that there is no such assertion. It is also easy to see that any model $\mathcal{I}_2$ of $\mathcal{K}_2$ such that some $a \in \mathsf{ind}(\mathcal{K}_2) \setminus \mathsf{ind}(\mathcal{K}_1)$ occurs in the $\mathcal{I}_2$-extension of some Σ-symbol is not con-Σ_2-homomorphically embeddable into a forest-shaped model of $\mathcal{K}_1$. For this reason, we should only consider completions of $\mathcal{K}_2$ such that for all $a \in \mathsf{ind}(\mathcal{K}_2) \setminus \mathsf{ind}(\mathcal{K}_1)$, $\tau(a)$ contains no Σ-concept names and no existential restrictions $\exists R.C$ with $R \in \Sigma$. We use $\mathsf{compl}_\mathsf{ok}(\mathcal{K}_2)$ to denote the set of all such completions. Now the 2ATA $\mathfrak{A}_2 = \langle Q, \Gamma, \delta, q_0\rangle$ is defined by setting

$$Q = \{q_0\} \cup \{q^{a,\boldsymbol{t}}, q^{R,\boldsymbol{t}} \mid a \in \mathsf{ind}(\mathcal{A}_1), \boldsymbol{t} \in \mathsf{type}(\mathcal{T}_2), R \in \mathsf{RN}(\mathcal{T}_2) \cap \Sigma\}$$

and defining the transition function δ as follows:

$$\delta(q_0, root) = \bigvee_{\tau\in\mathsf{compl_{ok}}(\mathcal{K}_2)} \bigwedge_{a\in\mathsf{ind}(\mathcal{A}_2)\cap\mathsf{ind}(\mathcal{A}_1)} \bigvee_{i=1}^{m} (i, q^{a,\tau(a)})$$

$$\delta(q^{a,\boldsymbol{t}}, (a,U)) = \bigwedge_{\substack{\exists R.C\in\boldsymbol{t}\\ R\in\Sigma}} \bigvee_{\boldsymbol{s}\in\mathsf{succ}_{\exists R.C}(\boldsymbol{t})} \left(\bigvee_{i=1}^{m} (i, q^{R,\boldsymbol{s}}) \vee \bigvee_{R(a,b)\in\mathcal{A}_1} (-1, q^{b,\boldsymbol{s}}) \right)$$

$$\delta(q^{S,\boldsymbol{t}}, (S,U)) = \bigwedge_{\substack{\exists R.C\in\boldsymbol{t}\\ R\in\Sigma}} \bigvee_{\boldsymbol{s}\in\mathsf{succ}_{\exists R.C}(\boldsymbol{t})} \bigvee_{i=1}^{m} (i, q^{R,\boldsymbol{s}})$$

where the latter two transitions are subject to the conditions that every Σ-concept name in $\boldsymbol{t}$ is also in U; also put

$$\delta(q^{a,\boldsymbol{t}}, root) = \bigvee_{i=1}^{m} (i, q^{a,\boldsymbol{t}})$$

$$\delta(q,\ell) = \mathsf{false} \quad \text{for all other } q \in Q \text{ and } \ell \in \Gamma.$$

Lemma 33. *$(T,L) \in L(\mathfrak{A}_2)$ iff there is a model $\mathcal{I}_2$ of $\mathcal{K}_2$ such that $\mathcal{I}_2$ is con-Σ-homomorphically embeddable into $\mathcal{I}_{(T,L)}$.*

Proof. ($\Rightarrow$) Given an accepting run (T_r, r) for (T,L), we can construct a forest-shaped model $\mathcal{I}_2$ of $\mathcal{K}_2$ and a Σ-homomorphism h from the maximal Σ-connected sub-interpretation of $\mathcal{I}_2$ to $\mathcal{I}_{(T,L)}$. Intuitively, each node $y \in T_r$ with $r(y) = (x, q^{a,\boldsymbol{t}})$ imposes that a has type $\boldsymbol{t}$ in $\mathcal{I}_2$, and each node $y \in T_r$ with $r(y) = (x, q^{R,\boldsymbol{t}})$ imposes that $\mathcal{I}_2$ contains an element y that belongs to a tree-shaped part of $\mathcal{I}_2$, is connected to its predecessor via R, and has type $\boldsymbol{t}$. The homomorphism h is defined by choosing the identity on individual names, and setting $h(y) = a$ when $r(y) = (x, q^{a,\boldsymbol{t}})$ and $h(y) = x$ when $r(y) = (x, q^{R,\boldsymbol{t}})$.

($\Leftarrow$) Assume that there is a model $\mathcal{I}_2$ of $\mathcal{K}_2$ such that $\mathcal{I}_2$ is con-Σ-homomorphically embeddable into $\mathcal{I}_{(T,L)}$. By the proof of Theorem 6, we can assume $\mathcal{I}_2 \in \boldsymbol{M}^{fo}_{\mathcal{K}_2}$. It is now straightforward to construct an accepting run for (T,L) by using $\mathcal{I}_2$ as a guide. $\square$

It is easy to verify that the constructed automaton $\mathfrak{A}$ has only single exponentially many states. It thus remains to invoke Point 3 of Theorem 32 to obtain the upper bound in Theorem 11.

E Proof of Thereoms 12 and 15

In this section we prove the semantic characterizations given in Theorem 12 and Theorem 15. In what follows we assume that *Horn-$\mathcal{ALC}$* TBoxes are given in *normal form* with concept inclusions of the following form:

$$A \sqsubseteq B, \quad A_1 \sqcap A_2 \sqsubseteq B, \quad \exists R.A \sqsubseteq B$$

and

$$A \sqsubseteq \bot, \quad \top \sqsubseteq B, \quad A \sqsubseteq \exists R.B, \quad A \sqsubseteq \forall R.B$$

where A, B range over concept names. We define the canonical model $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ of a consistent *Horn-$\mathcal{ALC}$* KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ in normal form in the standard way using a chase procedure. Consider the following rules that are applied to ABox $\mathcal{A}$:

1. if $A(a) \in \mathcal{A}$ and $A \sqsubseteq B \in \mathcal{T}$, then add $B(a)$ to $\mathcal{A}$;
2. if $A_1(a) \in A$ and $A_2(a) \in \mathcal{A}$ and $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$, then add $B(a)$ to $\mathcal{A}$;
3. if $R(a,b) \in A$ and $A(b) \in \mathcal{A}$ and $\exists R.A \sqsubseteq B \in \mathcal{T}$, then add $B(a)$ to $\mathcal{A}$;
4. if $a \in \mathsf{ind}(\mathcal{A})$ and $\top \sqsubseteq B \in \mathcal{T}$, then add $B(a)$ to $\mathcal{A}$;
5. if $A(a) \in \mathcal{A}$ and $A \sqsubseteq \exists R.B \in \mathcal{T}$ and there are no $R(a,b), B(b) \in \mathcal{A}$, then add assertions $R(a,b), B(b)$ to $\mathcal{A}$ for a fresh b;

6. if $A(a) \in \mathcal{A}$ and $R(a,b) \in \mathcal{A}$ and $A \sqsubseteq \forall R.B \in \mathcal{T}$, then add $B(b)$ to $\mathcal{A}$.

Denote by $\mathcal{A}^c$ the (possibly infinite) ABox resulting from $\mathcal{A}$ by applying these rules exhaustively to $\mathcal{A}$. Then the canonical model $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ is the interpretation defined by $\mathcal{A}^c$.

We now come to the proof of Theorem 12.

Theorem 12. *Let $\mathcal{T}_1$ be an $\mathcal{ALC}$ TBox, $\mathcal{T}_2$ a Horn-$\mathcal{ALC}$ TBox, and $\Theta = (\Sigma_1, \Sigma_2)$. Then $\mathcal{T}_1$ Θ-rCQ-entails $\mathcal{T}_2$ iff for all tree-shaped Σ_1-ABoxes $\mathcal{A}$ of outdegree bounded by $|\mathcal{T}_2|$ and consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is con-Σ_2-homomorphically embeddable into any model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$.*

Proof. It is known that Horn-$\mathcal{ALC}$ is unravelling tolerant [27], that is, if $(\mathcal{T}, \mathcal{A}) \models C(a)$ for a Horn-$\mathcal{ALC}$ TBox $\mathcal{T}$ and $\mathcal{EL}$-concept C, then $(\mathcal{T}, \mathcal{A}') \models C(a)$ for a finite subABox $\mathcal{A}'$ of the tree-unravelling $\mathcal{A}^u$ of $\mathcal{A}$ at a. Thus, any witness ABox for non-entailment w.r.t. $\mathcal{EL}$-instance queries can be transformed into a tree-shaped witness ABox. By Theorem 10 it is therefore sufficient to prove that if $\mathcal{T}_1$ does not Θ-rCQ-entail $\mathcal{T}_2$, then this is witnessed by an $\mathcal{EL}$-instance query $C(a)$.

Claim. If $\mathcal{T}_1$ does not Θ-rCQ-entail $\mathcal{T}_2$, then there exists a Σ_1-ABox $\mathcal{A}$ and an $\mathcal{EL}$-concept C over Σ_2 such that $\mathcal{T}_2, \mathcal{A} \models C(a)$ and $\mathcal{T}_1, \mathcal{A} \not\models C(a)$ for some $a \in \mathsf{ind}(\mathcal{A})$.

Assume $\mathcal{A}$ is a Σ_1-ABox and $q(\vec{x})$ a Σ_2-rCQ such that $\mathcal{T}_2, \mathcal{A} \models q(\vec{a})$ but $\mathcal{T}_1, \mathcal{A} \not\models q(\vec{a})$.

First we show that there exists a Σ_2-CQ $q'(\vec{z})$ such that $\mathcal{T}_2, \mathcal{A} \models q'(\vec{b})$ but $\mathcal{T}_1, \mathcal{A} \not\models q'(\vec{b})$ for some $\vec{b}$ and, moreover, there exists a match π for q' in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ witnessing this such that no quantified variable in q' is mapped to $\mathsf{ind}(\mathcal{A})$. Let π be a match for $q(\vec{x})$ in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$. Assume $q(\vec{x}) = \exists \vec{y} \varphi(\vec{x}, \vec{y})$. Let $\vec{y_1}$ be the additional variables mapped by π to elements of $\mathsf{ind}(\mathcal{A})$ and let $q'(\vec{x}, \vec{y_1}) = \exists \vec{y_2} \varphi(\vec{x}, \vec{y})$, where $\vec{y_2}$ are the remaining variables in $\vec{y}$ without $\vec{y_1}$. Then $\mathcal{T}_2, \mathcal{A} \models q'(\pi(\vec{x}, \vec{y_1}))$ but $\mathcal{T}_1, \mathcal{A} \not\models q'(\pi(\vec{x}, \vec{y_1}))$. Clearly q' is as required.

We can decompose q' into

- a quantifier-free core q_0 containing all $A(x)$ and $r(x,y)$ in q' such that x, y are answer variables;
- queries $q_1(x_1), \ldots, q_n(x_n)$ that each have exactly one answer variable.

We distinguish the following cases:

- $\mathcal{T}_1, \mathcal{A} \not\models q_0(\pi(\vec{x}, \vec{y_1}))$: in this case we find a single concept name $A \in \Sigma_2$ and $a \in \mathsf{ind}(\mathcal{A})$ such that $\mathcal{T}_2, \mathcal{A} \models A(a)$ and $\mathcal{T}_1, \mathcal{A} \not\models A(a)$.
- there exists $1 \leq i \leq n$ such that $\mathcal{T}_1, \mathcal{A} \not\models q_i(\pi(x_i))$. Let C_i be the image of q_i under π (π maps all variables from q_i except x_i to elements of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ not in $\mathcal{A}$. Thus this image is tree-shaped and can be identified with an $\mathcal{EL}$-concept). Then $\mathcal{T}_2, \mathcal{A} \models C_i(\pi(x_i))$ and $\mathcal{T}_1, \mathcal{A} \not\models C_i(\pi(x_i))$.

We have thus shown that there exists a Σ_1-ABox $\mathcal{A}$ and an $\mathcal{EL}$-concept C and $a \in \mathsf{ind}(\mathcal{A})$ such that $\mathcal{T}_2, \mathcal{A} \models C(a)$ and $\mathcal{T}_1, \mathcal{A} \not\models C(a)$. ☐

Theorem 15. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALC}$ TBoxes and Σ_1, Σ_2 be signatures. Then $\mathcal{T}_1$ (Σ_1, Σ_2)-CQ (equivalently, (Σ_1, Σ_2)-UCQ) entails $\mathcal{T}_2$ iff for all tree-shaped Σ_1-ABoxes $\mathcal{A}$ of outdegree bounded by $|\mathcal{T}_2|$ that are consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is Σ_2-homomorphically embeddable into $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.*

Proof. The proof is similar to the proof of Theorem 12. Denote by $\mathcal{EL}^u$ the extension of $\mathcal{EL}$ with the universal role u. Unravelling tolerance of Horn-$\mathcal{ALC}$ implies also that if $(\mathcal{T}, \mathcal{A}) \models C(a)$ for a Horn-$\mathcal{ALC}$ TBox $\mathcal{T}$ and $\mathcal{EL}^u$-concept C, then $(\mathcal{T}, \mathcal{A}') \models C(a)$ for a finite subABox $\mathcal{A}'$ of the tree-unravelling $\mathcal{A}^u$ of $\mathcal{A}$ at a. Thus, any witness ABox for non-entailment w.r.t. $\mathcal{EL}^u$-instance queries can be transformed into a tree-shaped witness ABox. By the homomorphism criterion for Σ-CQ entailment between Horn-$\mathcal{ALC}$-KBs proved in [9] it is therefore sufficient to prove that if $\mathcal{T}_1$ does not Θ-CQ-entail $\mathcal{T}_2$, then this is witnessed by an $\mathcal{EL}^u$-instance query $C(a)$. This proof is a straightforward extension of the proof of Theorem 12. ☐

The notion of con-Σ_2-homomorphic embeddability is slightly unwieldy to use in the subsequent definitions and constructions. We therefore resort to simulations. The following lemma gives an analysis of non-con-Σ_2-homomorphic embeddability in terms of simulations that is relevant for the subsequent constructions.

Lemma 13. *Let $\mathcal{A}$ be a Σ_1-ABox and $\mathcal{I}_1$ a model of $(\mathcal{T}_1, \mathcal{A})$. Then $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$ iff there is $a \in \mathsf{ind}(\mathcal{A})$ such that one of the following holds:*

(1) *there is a Σ_2-concept name A with $a \in A^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}} \setminus A^{\mathcal{I}_1}$;*

(2) *there is an R-successor d of a in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, for some Σ_2-role name R, such that $d \notin \mathsf{ind}(\mathcal{A})$ and, for all R-successors e of a in $\mathcal{I}_1$, we have $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d) \not\leq_{\Sigma_2} (\mathcal{I}_1, e)$.*

Proof. (sketch) The "if" direction is clear by definition of homomorphisms and because of the following: if there is a Σ_2-homomorphism h from the maximal Σ-connected subinterpretation of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to $\mathcal{I}_1$, $a \in \mathsf{ind}(\mathcal{A})$, and d is an R-successor of a in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with $R \in \Sigma_2$ and $d \notin \mathsf{ind}(\mathcal{A})$, then $h(d)$ is an R-successor of a in $\mathcal{I}_1$ and h contains a Σ_2-simulation from $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d)$ to $(\mathcal{I}_1, h(d))$.

For the "only if" direction, assume that both Point 1 and Point 2 are false for all $a \in \mathsf{ind}(\mathcal{A})$. Then for every $a \in \mathsf{ind}(\mathcal{A})$, R-successor d of a in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with $R \in \Sigma_2$ and $d \notin \mathsf{ind}(\mathcal{A})$, there is an R-successor d' of a in $\mathcal{I}_1$ and a simulation $\mathcal{S}_d$ from $\mathcal{I}_1$ to $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ such that $(d, d') \in \mathcal{S}_d$. Because the subinterpretation of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ rooted at d is tree-shaped, we can assume that $\mathcal{S}_d$ is a partial function. Now consider the function h defined by setting $h(a) = a$ for all $a \in \mathsf{ind}(\mathcal{A})$ and then taking the union with all the simulations $\mathcal{S}_d$. It can be verified that h is a Σ_2-homomorphism from the maximal Σ-connected subinterpretation of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to $\mathcal{I}_1$. □

F Proof of Theorem 14

We aim to prove Theorem 14, i.e., that it is EXPTIME-complete to decide whether an $\mathcal{ALC}$ TBox $\mathcal{T}_1$ (Σ_1, Σ_2)-rCQ entails a Horn-$\mathcal{ALC}$ TBox $\mathcal{T}_2$. We are going to use automata on finite trees.

F.1 Tree Automata Preliminaries

We introduce two-way alternating Büchi automata on finite trees (2ABTAs). A finite tree T is *m-ary* if for every $x \in T$, the set $\{i \mid x \cdot i \in T\}$ is of cardinality zero or exactly m. An *infinite path* P of T is a prefix-closed set $P \subseteq T$ such that for every $i \geq 0$, there is a unique $x \in P$ with $|x| = i$.

Definition 34 (2ABTA)**.** *A* two-way alternating Büchi automaton (2ABTA) on finite m-ary trees *is a tuple $\mathfrak{A} = (Q, \Gamma, \delta, q_0, R)$ where Q is a finite set of* states, Γ *is a finite alphabet,* $\delta : Q \times \Gamma \to \mathcal{B}^+(\mathsf{tran}(\mathfrak{A}))$ *is the* transition function *with* $\mathsf{tran}(\mathfrak{A}) = ([m] \times Q) \cup \mathsf{leaf}$ *the set of* transitions *of* $\mathfrak{A}$, $q_0 \in Q$ *is the* initial state*, and* $R \subseteq Q$ *is a set of* recurring states.

Transitions have the same intuition as for 2ATAs on infinite trees. The additional transition leaf verifies that the automaton is currently at a leaf node.

Definition 35 (Run, Acceptance)**.** *A* run *of a 2ABTA $\mathfrak{A} = (Q, \Gamma, \delta, q_0, R)$ on a finite Γ-labeled tree (T, L) is a $T \times Q$-labeled tree (T_r, r) such that the following conditions are satisfied:*

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

We say that (T_r, r) is accepting *if in all infinite paths $\varepsilon = y_1 y_2 \cdots$ of T_r, the set $\{i \geq 0 \mid r(y_i) = (x, q)$ for some $q \in R\}$ is infinite. A finite Γ-labeled tree (T, L) is* accepted *by $\mathfrak{A}$ if there is an accepting run of $\mathfrak{A}$ on (T, L). We use $L(\mathfrak{A})$ to denote the set of all finite Γ-labeled tree accepted by $\mathfrak{A}$.*

Apart from 2ABTAs, we will also use nondeterministic tree automata, introduced next.

A *nondeterministic top-down tree automaton* (NTA) on finite m-ary trees is a tuple $\mathfrak{A} = (Q, \Gamma, Q^0, \delta, F)$ where Q is a finite set of *states*, Γ is a finite alphabet, $Q^0 \subseteq Q$ is the set of *initial states*, $F \subseteq Q$ is a set of *final states*, and $\delta : Q \times \Gamma \to 2^{Q^m}$ is the *transition function*.

Let (T, L) be a Γ-labeled m-ary tree. A *run* of $\mathfrak{A}$ on (T, L) is a Q-labeled m-ary tree (T, r) such that $r(\varepsilon) \in Q^0$ and for each node $x \in T$, we have $\langle r(x \cdot 1), \ldots, r(x \cdot m)\rangle \in \delta(r(x), L(x))$. The run is *accepting* if for every leaf x of T, we have $r(x) \in F$. The set of trees accepted by $\mathcal{A}$ is denoted by $L(\mathcal{A})$.

We will use the following results from automata theory:

Theorem 36.

1. *Every 2ABTA $\mathfrak{A} = (Q, \Gamma, \delta, q_0, R)$ can be converted into an equivalent NTA $\mathfrak{A}'$ whose number of states is (single) exponential in $|Q|$; the conversion needs time polynomial in the size of $\mathfrak{A}'$;*

2. *Given a constant number of 2ABTAs (resp. NTAs), we can construct in polytime a 2ABTA (resp. an NTA) that accepts the intersection language;*

3. *Emptiness of NTAs can be checked in polytime.*

F.2 Γ-labeled Trees

For the proof of Theorem 14, fix an $\mathcal{ALC}$ TBox $\mathcal{T}_1$ and a Horn-$\mathcal{ALC}$ TBox $\mathcal{T}_2$ and signatures Σ_1, Σ_2. Set $m := |\mathcal{T}_2|$. Ultimately, we aim to construct an NTA $\mathfrak{A}$ such that a tree is accepted by $\mathfrak{A}$ if and only if this tree encodes a Σ_1-ABox $\mathcal{A}$ of outdegree at most m that is consistent with both $\mathcal{T}_1$ and $\mathcal{T}_2$ and a model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$ such that the canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ of $(\mathcal{T}_2, \mathcal{A})$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$. By Theorem 12, this means that $\mathfrak{A}$ accepts the empty language if and only if $\mathcal{T}_2$ is (Σ_1, Σ_2)-rCQ entailed by $\mathcal{T}_1$. In this section, we make precise which trees should be accepted by the NTA $\mathfrak{A}$ and in the subsequent section, we construct $\mathfrak{A}$.

As before, we assume that $\mathcal{T}_1$ takes the form $\top \sqsubseteq C_{\mathcal{T}_1}$ with $C_{\mathcal{T}_1}$ in NNF and use $\mathsf{cl}(C_{\mathcal{T}_1})$ to denote the set of subconcepts of $C_{\mathcal{T}_1}$, closed under single negation. We also assume that $\mathcal{T}_2$ is in the Horn-$\mathcal{ALC}$ normal form introduced above. We use $\mathsf{CN}(\mathcal{T}_2)$ to denote the set of concept names in $\mathcal{T}_2$ and $\mathsf{sub}(\mathcal{T}_2)$ for the set of subconcepts of (concepts in) $\mathcal{T}_2$.

Let Γ_0 denote the set of all subsets of $\Sigma_1 \cup \{R^- \mid R \in \Sigma_1\}$ that contain at most one role. Automata will run on m-ary Γ-labeled trees where

$$\Gamma = \Gamma_0 \times 2^{\mathsf{cl}(\mathcal{T}_1)} \times 2^{\mathsf{CN}(\mathcal{T}_2)} \times \{0, 1\} \times 2^{\mathsf{sub}(\mathcal{T}_2)}.$$

For easier reference, in a Γ-labeled tree (T, L) and for a node x from T, we write $L_i(x)$ to denote the $i+1$st component of $L(x)$, for each $i \in \{0, \ldots, 4\}$. Intuitively, the projection of a Γ-labeled tree to

- the L_0-components of its Γ-labels represents the Σ_1-ABox $\mathcal{A}$ that witnesses non-Σ_2-query entailment of $\mathcal{T}_2$ by $\mathcal{T}_1$;
- L_1-components (partially) represents a model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$;
- L_2-components (partially) represents the canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ of $(\mathcal{T}_2, \mathcal{A})$;
- L_3-components marks the individual a in $\mathcal{A}$ such that $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, a)$ is not Σ_2-simulated by $(\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}, a)$;
- L_4-components contains bookkeeping information that helps to ensure that the afore mentioned Σ_2-simulation indeed fails.

We now make these intuitions more precise by defining certain properness conditions for Γ-labeled trees, one for each component in the labels. A Γ-labeled (T, L) tree is *0-proper* if it satisfies the following conditions:

1. for the root ε of T, $L_0(\varepsilon)$ contains no role;
2. every non-root node x of T, $L_0(x)$ contains a role.

Every 0-proper Γ-labeled tree (T, L) represents the tree-shaped Σ_1-ABox

$$\begin{aligned}\mathcal{A}_{(T,L)} = \{&A(x) \mid A \in L_0(x)\} \\ &\cup \{R(x, y) \mid R \in L_0(y), y \text{ is a child of } x\} \\ &\cup \{R(y, x) \mid R^- \in L_0(y), y \text{ is a child of } x\}.\end{aligned}$$

A Γ-labeled tree (T, L) is *1-proper* if it satisfies the following conditions for all $x_1, x_2 \in T$:

1. there is a model $\mathcal{I}$ of $\mathcal{T}_1$ and a $d \in \Delta^{\mathcal{I}}$ such that $d \in C^{\mathcal{I}}$ iff $C \in L_1(x_1)$ for all $C \in \mathsf{cl}(\mathcal{T}_1)$;
2. $A \in L_0(x_1)$ implies $A \in L_1(x_1)$;

3. if x_2 is a child of x_1 and $R \in x_2$, then $\forall R.C \in L_1(x_1)$ implies $C \in L_1(x_2)$ for all $\forall R.C \in \mathsf{cl}(\mathcal{T}_1)$;

4. if x_2 is a child of x_1 and $R^- \in x_2$, then $\forall R.C \in L_1(x_2)$ implies $C \in L_1(x_1)$ for all $\forall R.C \in \mathsf{cl}(\mathcal{T}_1)$.

A Γ-labeled tree (T, L) is *2-proper* if for every node $x \in T$,

1. $A \in L_2(x)$ iff $\mathcal{A}_{(T,L)}, \mathcal{T}_2 \models A(x)$, for all $A \in \mathsf{CN}(\mathcal{T}_2)$;

2. if $A \in L_2(x)$, then $A \sqsubseteq \bot \notin \mathcal{T}_2$.

It is *3-proper* if there is exactly one node x with $L_3(x) = 1$. For defining 4-properness, we first give some preliminaries.

Let $t \subseteq \mathsf{CN}(\mathcal{T}_2)$. Then $\mathsf{cl}_{\mathcal{T}_2(S)} = \{A \in \mathsf{CN}(\mathcal{T}_2) \mid \mathcal{T}_2 \models \bigsqcap S \sqsubseteq A\}$. Moreover, we say that $S = \{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$ is a Σ_2-*successor set for* t if there is a concept name $A' \in t$ such that $A' \sqsubseteq \exists R.A \in \mathcal{T}_2$ and $\forall R.B_1, \ldots, \forall R.B_n$ is the set of all concepts of this form such that, for some $B \in t$, we have $B \sqsubseteq \forall R.B_i \in \mathcal{T}_2$. In the following, it will sometimes be convenient to speak about the canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{S}}$ of $\mathcal{T}_2$ and a finite set of concepts C that occur on the right-hand side of a CI in $\mathcal{T}_2$. What we mean with $\mathcal{I}_{\mathcal{T}_2,\mathcal{S}}$ is the interpretation obtained from the canonical model for the TBox $\mathcal{T}_i \cup \{A_C \sqsubseteq C \mid C \in \mathcal{S}\}$ and the ABox $\{A_C(a_\varepsilon) \mid C \in \mathcal{S}\}$ in which all fresh concept names A_C are removed.

A Γ-labeled tree is *4-proper* if it satisfies the following conditions for all nodes x_1, x_2:

1. if $L_3(x_1) = 1$, then there is a Σ_2-concept name in $L_2(x_1) \setminus L_1(x_1)$ or $L_4(x_1)$ is a Σ_2-successor set for $L_2(x_1)$;

2. if $L_4(x_1) \neq \emptyset$, then there is a model $\mathcal{I}$ of $\mathcal{T}_1$ and a $d \in \Delta^{\mathcal{I}}$ such that $d \in C^{\mathcal{I}}$ iff $C \in L_1(x_1)$ for all $C \in \mathsf{cl}(\mathcal{T}_1)$ and $(\mathcal{I}_{\mathcal{T}_2,L_4(x_1)}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}, d)$;

3. if x_2 is a child of x_1, $L_0(x_2)$ contains the role name R, and $L_4(x_1) = \{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$, then there is a Σ_2-concept name in $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\}) \setminus L_1(x_2)$ or $L_4(x_2)$ is a Σ_2-successor set for $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\})$;

4. if x_2 is a child of x_1, $L_0(x_2)$ contains the role R^-, and $L_4(x_2) = \{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$, then there is a Σ_2-concept name in $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\}) \setminus L_1(x_1)$ or $L_4(x_1)$ is a Σ_2-successor set for $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\})$.

Note how 4-properness addresses Condition 2 of Lemma 13. By that condition, there is a set of simulations from certain pointed "source" interpretations to certain pointed "target" interpretations that should be avoided. In the L_4-component of Γ-labels, we store the source interpretations, represented as sets of concepts. 4-properness then ensures that there is no simulation to the relevant target interpretations.

Lemma 37. *There is an m-ary Γ-labeled tree that is i-proper for all $i \in \{0, \ldots, 4\}$ iff there is a tree-shaped Σ_1-ABox $\mathcal{A}$ of outdegree at most m that is consistent w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$ and a model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$ such that the canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ of $(\mathcal{T}_2, \mathcal{A})$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$.*

Proof. "if". Let (T, L) be an m-ary Γ-labeled tree that is i-proper for all $i \in \{0, \ldots, 4\}$. Then $\mathcal{A}_{(T,L)}$ is a tree-shaped Σ_1-ABox of outdegree at most m. Moreover, $\mathcal{A}_{(T,L)}$ is consistent w.r.t. $\mathcal{T}_2$: because of the second condition of 2-properness, the canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ of $(\mathcal{T}_2, \mathcal{A})$ is indeed a model of $\mathcal{T}_2$ and $\mathcal{A}$.

Since (T, L) is 3-proper, there is exactly one $x_0 \in T$ with $L_3(x_0) = 1$. By construction, x_0 is also an individual name in $\mathcal{A}_{(T,L)}$. To finish this direction of the proof, it suffices to construct a model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A}_{(T,L)})$ such that $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, x_0) \not\lesssim_{\Sigma_2} (\mathcal{I}_1, x_0)$. In fact, $\mathcal{I}_1$ witnesses consistency of $\mathcal{A}_{(T,L)}$ with $\mathcal{T}_1$. Moreover, by definition of simulations $\mathcal{I}_1$ must satisfy one of Points 1 and 2 of Lemma 13 with a replaced by x_0. Consequently, by that Lemma $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$.

Start with the interpretation $\mathcal{I}_0$ defined as follows:

$$
\begin{aligned}
\Delta^{\mathcal{I}_0} &= T \\
A^{\mathcal{I}_0} &= \{x \in T \mid A \in L_1(x)\} \\
R^{\mathcal{I}_0} &= \{(x_1, x_2) \mid x_2 \text{ child of } x_1 \text{ and } R \in L_0(x_2)\} \cup \\
&\quad\ \{(x_2, x_1) \mid x_2 \text{ child of } x_1 \text{ and } R^- \in L_0(x_2)\}.
\end{aligned}
$$

Then take, for each $x \in T$, a model $\mathcal{I}_x$ of $\mathcal{T}$ such that $x \in C^{\mathcal{I}_x}$ iff $C \in L_1(x)$ for all $C \in \mathsf{cl}(\mathcal{T}_1)$. Moreover, if $L_4(x) \neq \emptyset$, then choose $\mathcal{I}_x$ such that $(\mathcal{I}_{\mathcal{T}_2,L_3(x)}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}_x, x)$. These choices are possible since (T, L) is

1-proper and 4-proper. Further assume that $\Delta^{\mathcal{I}_0}$ and $\Delta^{\mathcal{I}_x}$ share only the element x. Then $\mathcal{I}_1$ is the union of $\mathcal{I}_0$ and all chosen interpretations $\mathcal{I}_x$. It is not difficult to prove that $\mathcal{I}_1$ is indeed a model of $(\mathcal{T}_1, \mathcal{A}_{(T,L)})$.

We show that $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}_{(T,L)}}, x_0) \not\leq_{\Sigma_2} (\mathcal{I}_1, x_0)$. By Point 1 of 4-properness, there is a Σ_2-concept name in $L_2(x_0) \setminus L_1(x_0)$ or $L_4(x_0)$ is a Σ_2-successor set for $L_2(x)$. In the former case, we are done. In the latter case, it suffices to show the following.

Claim. For all $x \in T$: if $L_4(x) \neq \emptyset$, then $(\mathcal{I}_{\mathcal{T}_2,L_4(x)}, a_\varepsilon) \not\leq_{\Sigma_2} (\mathcal{I}_1, x)$.

The proof of the claim is by induction on the co-depth of x in $\mathcal{A}_{(T,L)}$, which is the length n of the longest sequence of role assertions $R_1(x, x_1), \ldots, R_n(x_{n-1}, x_n)$ in $\mathcal{A}_{(T,L)}$. It uses Conditions 2 to 4 of 4-properness.

"only if". Let $\mathcal{A}$ be a Σ_1-ABox of outdegree at most m that is consistent w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$, and $\mathcal{I}_1$ a model of $(\mathcal{T}_1, \mathcal{A})$ such that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$. By duplicating successors, we can make sure that every non-leaf in $\mathcal{A}$ has exactly m successors. We can further assume w.l.o.g. that $\mathsf{ind}(\mathcal{A})$ is a prefix-closed subset of $\mathbb{N}^*$ that reflects the tree-shape of $\mathcal{A}$, that is, $R(a, b) \in \mathcal{A}$ implies $b = a \cdot c$ or $a = b \cdot c$ for some $c \in \mathbb{N}$. By Lemma 13, there is an $a_0 \in \mathsf{ind}(\mathcal{A})$ such that one of the following holds:

1. there is a Σ_2-concept name A with $a_0 \in A^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}} \setminus A^{\mathcal{I}_1}$;
2. there is an R_0-successor d_0 of a_0 in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, for some Σ_2-role name R_0, such that $d_0 \notin \mathsf{ind}(\mathcal{A})$ and for all R_0-successors d of a_0 in $\mathcal{I}_1$, we have that $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d_0) \not\leq_{\Sigma_2} (\mathcal{I}_1, d)$.

We now show how to construct from $\mathcal{A}$ a Γ-labeled tree (T, L) that is i-proper for all $i \in \{0, \ldots, 4\}$. For each $a \in \mathsf{ind}(\mathcal{A})$, let $R(a)$ be undefined if $a = \varepsilon$ and otherwise let $R(a)$ be the unique role R (i.e., role name or inverse role) such that $R(b, a) \in \mathcal{A}$ and $a = b \cdot c$ for some $c \in \mathbb{N}$. Now set

$$\begin{array}{rcl} T & = & \mathsf{ind}(\mathcal{A}) \\ L_1(x) & = & \{C \in \mathsf{cl}(\mathcal{T}_1) \mid x \in C^{\mathcal{I}_1}\} \\ L_2(x) & = & \{C \in \mathsf{CN}(\mathcal{T}_2) \mid \mathcal{A}, \mathcal{T}_2 \models C(x)\} \\ L_3(x) & = & \begin{cases} 1 & \text{if } x = a_0 \\ 0 & \text{otherwise} \end{cases} \end{array}$$

It remains to define L_4. Start with setting $L_4(x) = \emptyset$ for all x. If Point 1 above is true, we are done. If Point 2 is true, then there is a Σ_2-successor set $S = \{\exists R_0.A, \forall R_0.B_1, \ldots, \forall R_0.B_n\}$ for $L_2(a_0)$ such that the restriction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to the subtree-interpretation rooted at d_0 is the canonical model $\mathcal{I}_{\mathcal{T},\{A,B_1,\ldots,B_n\}}$. Set $L_4(a_0) = S$. We continue to modify L_4, proceeding in rounds. To keep track of the modifications that we have already done, we use a set

$$\Gamma \subseteq \mathsf{ind}(\mathcal{A}) \times (\mathsf{N_R} \cap \Sigma_2) \times \Delta^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}}$$

such that the following conditions are satisfied:

(i) if $(a, R, d) \in \Gamma$, then $L_4(a)$ has the form $\{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$ and the restriction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to the subtree-interpretation rooted at d is the canonical model $\mathcal{I}_{\mathcal{T},\{A,B_1,\ldots,B_n\}}$;

(ii) if $(a, R, d) \in \Gamma$ and d' is an R-successor of a in $\mathcal{I}_1$, then $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d) \not\leq_{\Sigma_2} (\mathcal{I}_1, d')$.

Initially, set $\Gamma = \{(a_0, R_0, d_0)\}$. In each round of the modification of L_4, iterate over all elements $(a, R, d) \in \Gamma$ that have not been processed in previous rounds. Let $L_4(a) = \{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$ and iterate over all R-successors b of a in $\mathcal{A}$. By (ii), $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d) \not\leq_{\Sigma_2} (\mathcal{I}_1, b)$. By (i), there is thus a top-level Σ_2-concept name A' in $A \sqcap B_1 \sqcap \cdots \sqcap B_n$ such that $b \notin A'^{\mathcal{I}_1}$ or there is an R'-successor d' of d in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, R' a Σ_2-role name, such that for all R'-successors d'' of b in $\mathcal{I}_1$, $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d') \not\leq_{\Sigma_2} (\mathcal{I}_1, d'')$. In the former case, do nothing. In the latter case, there is a Σ_2-successor set $S' = \{\exists R'.A', \forall R'.B'_1, \ldots, \forall R'.B'_{n'}\}$ for $\{A, B_1, \ldots, B_n\}$ such that the restriction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to the subtree-interpretation rooted at d' is the canonical model $\mathcal{I}_{\mathcal{T},\{A',B'_1,\ldots,B'_{n'}\}}$. Set $L_4(b) = S'$ and add (b, R', d') to Γ.

Since we are following only role names (but not inverse roles) during the modification of L_4 and since $\mathcal{A}$ is tree-shaped, we will never process tuples $(a_1, R_1, d_1), (a_2, R_2, d_2)$ from Γ such that $a_1 = a_2$. For any x, we might thus only redefine $L_4(x)$ from the empty set to a non-empty set, but never from one non-empty set to another. For the same reason, the definition of L_4 finishes after finitely many rounds.

It can be verified that the Γ-labeled tree (T, L) just constructed is i-proper for all $i \in \{0, \ldots, 4\}$. The most interesting point is 4-properness, which consists of four conditions. Condition 1 is satisfied by construction of L_4. Condition 2 is satisfied by $(*)$ and Conditions 3 and 4 again by construction of L_4. □

F.3 Upper Bound in Theorem 14

By Theorem 12 and Lemma 40, we can decide whether $\mathcal{T}_1$ does (Σ_1, Σ_2)-rCQ entail $\mathcal{T}_2$ by checking that there is no Γ-labeled tree that is i-proper for each $i \in \{0, \ldots, 4\}$. We do this by constructing automata $\mathcal{A}_0, \ldots, \mathcal{A}_4$ such that each $\mathcal{A}_i$ accepts exactly the Γ-labeled trees that are i-proper, then intersecting the automata and finally testing for emptiness. Some of the constructed automata are 2ABTAs while others are NTAs. Before intersecting, all 2ABTAs are converted into equivalent NTAs (which involves an exponential blowup). Emptiness of NTAs can be decided in time polynomial in the number of states. To achieve EXPTIME overall complexity, the constructed 2ABTAs should thus have at most polynomially many states while the NTAs can have at most (single) exponentially many states.It is straightforward to construct

1. an NTA $\mathfrak{A}_0$ that checks 0-properness and has constantly many states;
2. a 2ABTA $\mathfrak{A}_1$ that checks 1-properness and whose number of states is polynomial in $|\mathcal{T}_1|$ (note that Conditions 1 and 2 of 1-properness are in a sense trivial as they could also be guaranteed by removing undesired symbols from the alphabet Γ);
3. an NTA $\mathfrak{A}_3$ that checks 3-properness and has constantly many states;
4. an NTA $\mathfrak{A}_4$ that checks 4-properness and whose number of states is (single) exponential in $|\mathcal{T}_2|$ (note that Conditions 1 and 2 of 4-properness could again be ensured by refining Γ).

Details are omitted. It thus remains to construct an automaton $\mathfrak{A}_2$ that checks 2-properness. For this purpose, it is more convenient to use a 2ABTA than an NTA. In fact, the reason for mixing 2ABTAs and NTAs is that while $\mathfrak{A}_2$ is more easy to be constructed as a 2ABTA, there is no obvious way to construct $\mathfrak{A}_4$ as a 2ABTA with only polynomially many states: it seems one needs that one state is needed for every possible value of the L_4-components in Γ-labels.

The 2ABTA $\mathfrak{A}_2$ is actually the intersection of two 2ABTAs $\mathfrak{A}_{2,1}$ and $\mathfrak{A}_{2,2}$. The 2ABTA $\mathfrak{A}_{2,1}$ ensures one direction of Condition 1 of 2-properness as well as Condition 2, that is:

(i) if $\mathcal{A}_{(T,L)}$ is consistent w.r.t. $\mathcal{T}_2$, then $\mathcal{A}_{(T,L)}, \mathcal{T}_2 \models A(x)$ implies $A \in L_2(x)$ for all $x \in T$ and $A \in \mathsf{CN}(\mathcal{T}_2)$;

(ii) if $A \in L_2(x)$, then $A \sqsubseteq \bot \notin \mathcal{T}_2$.

It is simple for a 2ABTA to verify (ii), alternatively one could refine Γ. To achieve (i), it suffices to guarantee the following conditions for all $x_1, x_2 \in T$, which are essentially just the rules of a chase required for Horn-$\mathcal{ALC}$ TBoxes in normal form:

1. $A \in L_0(x_1)$ implies $A \in L_2(x_1)$;
2. if $A_1, \ldots, A_n \in L_2(x_1)$ and $\mathcal{T}_2 \models A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A$, then $A \in L_2(x_1)$;
3. if $A \in L_2(x_1)$, x_2 is a successor of x_1, $R \in L_0(x_2)$, and $A \sqsubseteq \forall R.B \in \mathcal{T}_2$, then $B \in L_2(x_2)$;
4. if $A \in L_2(x_2)$, x_2 is a successor of x_1, $R^- \in L_0(x_2)$, and $A \sqsubseteq \forall R.B \in \mathcal{T}_2$, then $B \in L_2(x_1)$;
5. if $A \in L_2(x_2)$, x_2 is a successor of x_1, $R \in L_0(x_2)$, and $\exists R.A \sqsubseteq B \in \mathcal{T}_2$, then $B \in L_2(x_1)$;
6. if $A \in L_2(x_1)$, x_2 is a successor of x_1, $R^- \in L_0(x_2)$, and $\exists R.A \sqsubseteq B \in \mathcal{T}_2$, then $B \in L_2(x_2)$

All of this is easily verified with a 2ABTA, details are again ommitted. Note that Conditions 1 and 2 can again be ensured by refining Γ.

The purpose of $\mathfrak{A}_{2,2}$ is to ensure the converse of (i). Before constructing it, it is convenient to first characterize the entailment of concept names at ABox individuals in terms of derivation trees. A $\mathcal{T}_2$-*derivation tree* for an assertion $A_0(a_0)$ in $\mathcal{A}$ with $A_0 \in \mathsf{CN}(\mathcal{T}_2)$ is a finite $\mathsf{ind}(\mathcal{A}) \times \mathsf{CN}(\mathcal{T}_2)$-labeled tree (T, V) that satisfies the following conditions:

- $V(\varepsilon) = (a_0, A_0)$;
- if $V(x) = (a, A)$ and neither $A(a) \in \mathcal{A}$ nor $\top \sqsubseteq A \in \mathcal{T}_2$, then one of the following holds:
  - x has successors $y_1, \ldots, y_n$ with $V(y_i) = (a, A_i)$ for $1 \leq i \leq n$ and $\mathcal{T}_2 \models A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A$;

- x has a single successor y with $V(y) = (b, B)$ and there is an $\exists R.B \sqsubseteq A \in \mathcal{T}_2$ such that $R(a, b) \in \mathcal{A}$;
- x has a single successor y with $V(y) = (b, B)$ and there is a $B \sqsubseteq \forall R.A \in \mathcal{T}_2$ such that $R(b, a) \in \mathcal{A}$.

Lemma 38. *If $\mathcal{A}, \mathcal{T}_2 \models A(a)$ and $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_2$, then there is a derivation tree for $A(a)$ in $\mathcal{A}$, for all assertions $A(a)$ with $A \in \mathsf{CN}(\mathcal{T}_2)$ and $a \in \mathsf{ind}(\mathcal{A})$.*

A proof of Lemma 38 is based on the chase procedure, details can be found in [7] for the extension $\mathcal{ELI}_\bot$ of Horn-$\mathcal{ALC}$.

We are now ready to construct the remaining 2ABTA $\mathfrak{A}_{2,2}$. By Lemma 38 and since $\mathfrak{A}_{2,1}$ ensures that $\mathcal{A}_{(T,L)}$ is consistent w.r.t. $\mathcal{T}_2$, it is enough for $\mathfrak{A}_{2,2}$ to verify that, for each node $x \in T$ and each concept name $A \in L_2(x)$, there is a $\mathcal{T}_2$-derivation tree for $A(x)$ in $\mathcal{A}_{(T,L)}$.

For readability, we use $\Gamma^- := \Gamma_0 \times \mathsf{CN}(\mathcal{T}_2)$ as the alphabet instead of Γ since transitions of $\mathfrak{A}_{2,2}$ only depend on the L_0- and L_2-components of Γ-labels. Let $\mathsf{rol}(\mathcal{T}_2)$ be the set of all roles R, R^- such that the role name R occurs in $\mathcal{T}_2$. Set $\mathfrak{A}_2 = (Q, \Gamma^-, \delta, q_0, R)$ with

$$Q \quad = \quad \{q_0\} \uplus \{q_A \mid A \in \mathsf{CN}(\mathcal{T}_2)\} \uplus \{q_{A,R}, q_R \mid A \in \mathsf{CN}(\mathcal{T}_2), R \in \mathsf{rol}(\mathcal{T}_2)\}$$

and $R = \emptyset$ (i.e., exactly the finite runs are accepting). For all $(\sigma_0, \sigma_2) \in \Gamma^-$, set

- $\delta(q_0, (\sigma_0, \sigma_2)) = \bigwedge_{A \in \sigma_2} (0, q_A) \wedge (\mathsf{leaf} \vee \bigwedge_{i \in 1..m} (i, q_0))$;
- $\delta(q_A, (\sigma_0, \sigma_2)) = \mathsf{true}$ whenever $A \in \sigma_1$ or $\top \sqsubseteq A \in \mathcal{T}_2$;
- $\delta(q_A, (\sigma_0, \sigma_2)) = \bigvee_{\mathcal{T}_2 \models A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A} ((0, q_{A_1}) \wedge \cdots \wedge (0, q_{A_n})) \vee$
  $\bigvee_{\exists R.B \sqsubseteq A \in \mathcal{T},\ R \in \Sigma_1} (((0, q_{R^-}) \wedge (-1, q_B)) \vee \bigvee_{i \in 1..m} (i, q_{B,R})) \vee$
  $\bigvee_{B \sqsubseteq \forall R.A \in \mathcal{T},\ R \in \Sigma_1} ((0, q_R) \wedge (-1, q_B)) \vee \bigvee_{i \in 1..m} (i, q_{B,R^-}))$

  whenever $A \notin \sigma_0$ and $\top \sqsubseteq A \notin \mathcal{T}_2$;
- $\delta(q_{A,R}, (\sigma_0, \sigma_2)) = (0, q_A)$ whenever $R \in \sigma_0$;
- $\delta(q_{A,R}, (\sigma_0, \sigma_2)) = \mathsf{false}$ whenever $R \notin \sigma_0$;
- $\delta(q_R, (\sigma_0, \sigma_2)) = \mathsf{true}$ whenever $R \in \sigma_0$;
- $\delta(q_R, (\sigma_0, \sigma_2)) = \mathsf{false}$ whenever $R \notin \sigma_0$.

Note that the finiteness of runs ensures that $\mathcal{T}_2$-derivation trees are also finite, as required.

F.4 Upper Bound in Theorem 16

The 2EXPTIME upper bound stated in Theorem 16 can be obtained by a modification of the construction given in Section F.3. We now have to build in the characterization given in Theorem 12 instead of the one from Theorem 15. There are two differences: first, the theorem refers to the canonical model $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ instead of quantifying over all models $\mathcal{I}$ of $(\mathcal{T}_1, \mathcal{A})$; and second, we need to consider Σ_2-homomorphic embeddability instead of con-Σ_2-homomorphic embeddability. The former difference can be ignored. In fact, Theorem 15 remains true if we quantify over all models $\mathcal{I}$ of $(\mathcal{T}_1, \mathcal{A})$, as in Theorem 12, because $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ is Σ_2-homomorphically embeddable into any model of $\mathcal{T}_1$ and $\mathcal{A}$. The second difference, however, does make a difference. To understand it more properly, we first give the following adaptation of Lemma 13.

Lemma 39. *Let $\mathcal{A}$ be a Σ_1-ABox and $\mathcal{I}_1$ a model of $(\mathcal{T}_1, \mathcal{A})$. Then $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not Σ_2-homomorphically embeddable into $\mathcal{I}_1$ iff there is an $a \in \mathsf{ind}(\mathcal{A})$ such that one of the following is true:*

1. *there is a Σ_2-concept name A with $a \in A^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}} \setminus A^{\mathcal{I}_1}$;*
2. *there is an R_0-successor d_0 of a in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, for some Σ_2-role name R_0, such that $d_0 \notin \mathsf{ind}(\mathcal{A})$ and for all R_0-successors d of a in $\mathcal{I}_1$, we have that $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d_0) \not\leq_{\Sigma_2} (\mathcal{I}_1, d)$.*
3. *there is an element d in the subtree of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ rooted at a (with possibly $d = a$) and d has an R_0-successor d_0, for some role name $R_0 \notin \Sigma_2$, such that for all elements e of $\mathcal{I}_1$, we have $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d_0) \not\leq_{\Sigma_2} (\mathcal{I}_1, e)$.*

The proof of Lemma 39 is very similar to that of Lemma 13, details are omitted. Note that the difference between Lemma 13 and Lemma 39 is the additional Condition 3 in the latter. This condition needs to be reflected in the definition of proper Γ-labeled trees which, in turn, requires a modification of the alphabet Γ.

An important reason for the construction in Section F.3 to yield an EXPTIME upper bound is that in the L_4-component of Γ-labels, we only need to store a single successor set instead of a set of such sets. This is not the case in the new construction (which only yields 2EXPTIME upper bound) where we let the L_4-component of Γ-labels range over $2^{2^{\mathsf{sub}(\mathcal{T}_2)}}$ instead of over $2^{\mathsf{sub}(\mathcal{T}_2)}$. We also add an L_5-component to Γ-labels, which also ranges over $2^{2^{\mathsf{sub}(\mathcal{T}_2)}}$. The notion of i-properness remains the same for $i \in \{0, 1, 2, 3\}$. We adapt the notion of 4-properness and add a notion of 5-properness.

As a preliminary, we need to define the notion of a descendant set. Let $t \subseteq \mathsf{CN}(\mathcal{T}_2)$ and define Γ to be the smallest set such that

- $t \in \Gamma$;
- if $t' \in \Gamma$, $A \in t'$, and $A' \sqsubseteq \exists R.A \in \mathcal{T}_2$, then $\{A, B_1, \ldots, B_n\} \in \Gamma$ where $B_1, \ldots, B_n$ is the set of all concept names such that, for some $B \in t'$, we have $B \sqsubseteq \forall R.B_i \in \mathcal{T}_2$.

Note that, in the above definition, R need not be from Σ_2 (nor from its complement). A subset s of $\mathsf{CN}(\mathcal{T}_2)$ is a *descendant set* for t if there is a $t' \in \Gamma$, an $A \in t'$, and an $A' \sqsubseteq \exists R.A \in \mathcal{T}_2$ with $R \notin \Sigma_2$ such that s consists of A and of all concept names B such that $B' \sqsubseteq \forall R.B \in \mathcal{T}_2$ for some $B' \in t'$.

A Γ-labeled tree (T, L) is *4-proper* if it satisfies the following conditions for all $x_1, x_2 \in T$:

1. if $L_3(x_1) = 1$, then one of the following is true:
   - there is a Σ_2-concept name in $L_2(x_1) \setminus L_1(x_1)$;
   - or $L_4(x_1)$ contains a Σ_2-successor set for $L_2(x_1)$;
   - $L_5(y)$ contains a Σ_2-descendant set for $L_2(x_1)$;
2. there is a model $\mathcal{I}$ of $\mathcal{T}_1$ and a $d \in \Delta^{\mathcal{I}}$ such that all of the following are true:
   - $d \in C^{\mathcal{I}}$ iff $C \in L_1(x_1)$ for all $C \in \mathsf{cl}(\mathcal{T}_1)$;
   - if $S \in L_4(x_1)$, then $(\mathcal{I}_{\mathcal{T}_2,S}, a_\varepsilon) \not\leq_{\Sigma_2} (\mathcal{I}, d)$;
   - if $S \in L_5(x_1)$ and $e \in \Delta^{\mathcal{I}}$, then $(\mathcal{I}_{\mathcal{T}_2,S}, a_\varepsilon) \not\leq_{\Sigma_2} (\mathcal{I}, e)$;
3. if x_2 is a child of x_1, $L_0(x_2)$ contains the role name R, and $L_4(x_1) \ni \{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$, then there is a Σ_2-concept name in $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\}) \setminus L_1(x_2)$ or $L_4(x_2)$ contains a Σ_2-successor set for $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\})$;
4. if x_2 is a child of x_1, $L_0(x_2)$ contains the role R^-, and $L_4(x_2) \ni \{\exists R.A, \forall R.B_1, \ldots, \forall R.B_n\}$, then there is a Σ_2-concept name in $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\}) \setminus L_1(x_1)$ or $L_4(x_1)$ contains a Σ_2-successor set for $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \ldots, B_n\})$.

A Γ-labeled tree (T, L) is *5-proper* if all $x \in T$ agree regarding their L_5-label.

Note how the adapted notion of 4-properness and the L_5-component of Γ-labels implements the additional third condition of Lemma 39. That condition gives rise to an additional set of simulations that have to be avoided. The (pointed) interpretations on the "source side" of these simulations are described using sets of concepts in L_5. In the pointed interpretations $(\mathcal{I}_1, e)$ on the "target side", we now have to consider all possible points e. For this reason, 5-properness distributes elements of L_5-labels to everywhere else. The simulations are then avoided via the additional third item in the second condition of 4-properness. The proof details of the following lemma are omitted.

Lemma 40. *There is an m-ary Γ-labeled tree that is i-proper for all $i \in \{0, \ldots, 5\}$ iff there is a tree-shaped Σ_1-ABox $\mathcal{A}$ of outdegree at most m that is consistent w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$ and a model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$ such that the canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ of $(\mathcal{T}_2, \mathcal{A})$ is not Σ_2-homomorphically embeddable into $\mathcal{I}_1$.*

It is now straightforward to adapt the automaton construction to the new version of 4-properness, and to add an automaton for 5-properness. The NTA for 4-properness will now have double exponentially many states because L_4- and L_5-components are sets of sets of concepts instead of sets of concepts. In fact, we could dispense NTAs altogether and use an 2ABTA that has exponentially many states. Overall, we obtain a 2EXPTIME upper bound.

F.5 2ExpTime Lower Bound

We reduce the word problem of exponentially space bounded alternating Turing machines (ATMs), see [13]. An *Alternating Turing Machine (ATM)* is of the form $M = (Q, \Sigma, \Gamma, q_0, \Delta)$. The set of *states* $Q = Q_\exists \uplus Q_\forall \uplus \{q_a\} \uplus \{q_r\}$ consists of *existential states* in $Q_\exists$, *universal states* in $Q_\forall$, an *accepting state* q_a, and a *rejecting state* q_r; Σ is the *input alphabet* and Γ the *work alphabet* containing a *blank symbol* $\square$ and satisfying $\Sigma \subseteq \Gamma$; $q_0 \in Q_\exists \cup Q_\forall$ is the *starting* state; and the *transition relation* Δ is of the form

$$\Delta \subseteq Q \times \Gamma \times Q \times \Gamma \times \{L, R\}.$$

We write $\Delta(q, \sigma)$ to denote $\{(q', \sigma', M) \mid (q, \sigma, q', \sigma', M) \in \Delta\}$ and assume w.l.o.g. that every set $\Delta(q, \sigma)$ contains exactly two elements when q is universal, and that the state q_0 is existential and cannot be reached by any transition.

A *configuration* of an ATM is a word wqw' with $w, w' \in \Gamma^*$ and $q \in Q$. The intended meaning is that the one-side infinite tape contains the word ww' with only blanks behind it, the machine is in state q, and the head is on the symbol just after w. The *successor configurations* of a configuration wqw' are defined in the usual way in terms of the transition relation Δ. A *halting configuration* (resp. *accepting configuration*) is of the form wqw' with $q \in \{q_a, q_r\}$ (resp. $q = q_a$).

A *computation tree* of an ATM M on input w is a tree whose nodes are labeled with configurations of M on w, such that the descendants of any non-leaf labeled by a universal (resp. existential) configuration include all (resp. one) of the successors of that configuration. A computation tree is *accepting* if the root is labeled with the *initial configuration* q_0w for w and all leaves with accepting configurations. An ATM M accepts input w if there is a computation tree of M on w.

There is an exponentially space bounded ATM M whose word problem is 2-ExpTime-hard and we may assume that the length of every computation path of M on $w \in \Sigma^n$ is bounded by 2^{2^n}, and all the configurations wqw' in such computation paths satisfy $|ww'| \leq 2^n$, see [13]. We may also assume w.l.o.g. that M makes at least one step on every input, and that it never reaches the last tape cell (which is both not essential for the reduction, but simplifies it).

Let w be an input to M. We aim to construct Horn-$\mathcal{ALC}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and a signature Σ such that the following are equivalent:

1. there is a tree-shaped Σ-ABox $\mathcal{A}$ such that
   (a) $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$ and
   (b) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not Σ-homomorphically embeddable into $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$;
2. M accepts w.

Note that we dropped the outdegree condition from Theorem 15. In fact, it is easy to go through the proofs of that theorem and verify that this condition is not needed; we have included it because it makes the upper bounds easier.

When dealing with an input w of length n, we represent configurations of M by a sequence of 2^n elements linked by the role name R, from now on called *configuration sequences*. These sequences are then interconnected to form a representation of the computation tree of M on w. This is illustrated in Figure 7, which shows three configuration sequences, enclosed by dashed boxes. The topmost configuration is universal, and it has two successor configurations. All solid arrows denote R-edges. We will explain later why successor configurations are separated by two consecutive edges instead of by a single one.

The above description is actually an oversimplification. In fact, every configuration sequence stores two configurations instead of only one: the current configuration and the previous configuration in the computation. We will later use the homomorphism condition (a) above to ensure that

($*$) the previous configuration stored in a configuration sequence is identical to the current configuration stored in its predecessor configuration sequence.

The actual transitions of M are then represented locally inside configuration sequences.

We next show how to use the TBox $\mathcal{T}_2$ to verify the existence of a computation tree of M on input w in the ABox, assuming ($*$). The signature Σ consists of the following symbols:

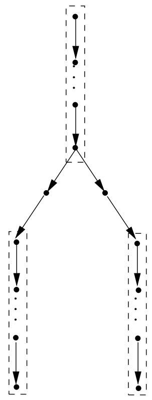


Figure 7: Configuration tree (partial)

1. concept names $A_0, \dots, A_{n-1}$ and $\overline{A}_0, \dots, \overline{A}_{n-1}$ that serve as bits in the binary representation of a number between 0 and $2^n - 1$, identifying the position of tape cells inside configuration sequences (A_0, $\overline{A}_0$ represent the lowest bit);
2. the concept names A_σ, A'_σ, $\overline{A}_\sigma$ for each $\sigma \in \Gamma$;
3. the concept names $A_{q,\sigma}$, $A'_{q,\sigma}$, $\overline{A}_{q,\sigma}$ for each $\sigma \in \Gamma$ and $q \in Q$;
4. concept names X_L, X_R that mark left and right successor configurations;
5. the role name R.

From the above list, concept names A_σ and $A_{q,\sigma}$ are used to represent the current configuration and A'_σ and $A'_{q,\sigma}$ for the previous configuration. The rôle of the concept names $\overline{A}_\sigma$ and $\overline{A}_{q,\sigma}$ will be explained later.

We start with verifying accepting configurations, in a bottom-up manner:

$$
\begin{aligned}
A_0 \sqcap \cdots \sqcap A_{n-1} \sqcap A_\sigma \sqcap A'_\sigma &\sqsubseteq V \\
A_i \sqcap \exists R.A_i \sqcap \bigsqcup_{j<i} \exists R.A_j &\sqsubseteq \mathsf{ok}_i \\
\overline{A}_i \sqcap \exists R.\overline{A}_i \sqcap \bigsqcup_{j<i} \exists R.A_j &\sqsubseteq \mathsf{ok}_i \\
A_i \sqcap \exists R.\overline{A}_i \sqcap \bigsqcap_{j<i} \exists R.\overline{A}_j &\sqsubseteq \mathsf{ok}_i \\
\overline{A}_i \sqcap \exists R.A_i \sqcap \bigsqcap_{j<i} \exists R.\overline{A}_j &\sqsubseteq \mathsf{ok}_i \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V \sqcap A_\sigma \sqcap A'_\sigma &\sqsubseteq V \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V \sqcap A_\sigma \sqcap A'_{q,\sigma'} &\sqsubseteq V_{L,\sigma} \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V \sqcap A_{q_{\mathsf{acc}},\sigma} \sqcap A'_\sigma &\sqsubseteq V_{R,q_{\mathsf{acc}}} \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V_{L,\sigma} \sqcap A_{q_{\mathsf{acc}},\sigma'} \sqcap A'_{\sigma'} &\sqsubseteq V_{L,q_{\mathsf{acc}},\sigma} \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V_{R,q_{\mathsf{acc}}} \sqcap A_\sigma \sqcap A'_{q,\sigma'} &\sqsubseteq V_{R,q_{\mathsf{acc}},\sigma} \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V_{M,q_{\mathsf{acc}},\sigma} \sqcap A_{\sigma'} \sqcap A'_{\sigma'} &\sqsubseteq V_{M,q_{\mathsf{acc}},\sigma} \\
\exists R.A_i \sqcap \exists R.\overline{A}_i &\sqsubseteq \bot
\end{aligned}
$$

where σ, σ' range over Γ, q over Q, and i over $0..n-1$. The first line starts the verification at the last tape cell, ensuring that at least one concept name A_α and one concept name A'_β is true. The following lines implement the verification of the remaining tape cells of the configuration. Lines two to five implement decrementation of a binary counter and the conjunct $\overline{A}_i$ in Lines six to eleven prevents the counter from wrapping around once it has reached zero. We use several kinds of verification markers:

- with V, we indicate that we have not yet seen the head of the TM;
- $V_{L,\sigma}$ indicates that the TM made a step to the left to reach the current configuration, writing σ;
- $V_{R,q}$ indicates that the TM made a step to the right to reach the current configuration, switching to state q;

- $V_{M,q,\sigma}$ indicates that the TM moved in direction M to reach the current configuration, switching to state q and writing σ.

In the remaining reduction, we expect that a marker of the form $V_{M,q,\sigma}$ has been derived at the first cell of the configuration. This makes sure that there is exactly one head in the current and in the previous configuration, and that the head moved exactly one step between the previous and the current position. Also note that the above CIs make sure that the tape content does not change for cells that were not under the head in the previous configuration. We exploit that M never moves its head to the right-most tape cell, simply ignoring this case in the CIs above. Note that it is not immediately clear that lines two to eleven work as intended since they can speak about different R-successors for different bits. The last line fixes this problem.

We also ensure that relevant concept names are mutually exclusive:

$$\begin{array}{rcll} A_i \sqcap \overline{A}_i & \sqsubseteq & \bot & \\ A_{\sigma_1} \sqcap A_{\sigma_2} & \sqsubseteq & \bot & \text{if } \sigma_1 \neq \sigma_2 \\ A_{\sigma_1} \sqcap A_{q_2,\sigma_2} & \sqsubseteq & \bot & \\ A_{q_1,\sigma_1} \sqcap A_{q_2,\sigma_2} & \sqsubseteq & \bot & \text{if } (q_1,\sigma_1) \neq (q_2,\sigma_2) \end{array}$$

where the i ranges over $0..n-1$, σ_1, σ_2 over Γ, and q_1, q_2 over Q. We also add the same concept inclusions for the primed versions of these concept names. The next step is to verify non-halting configurations:

$$\begin{array}{c} \exists R.\exists R.(X_L \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap (V_{M,q,\sigma} \sqcup V'_{M,q,\sigma})) \sqsubseteq L\mathsf{ok} \\ \exists R.\exists R.(X_R \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap (V_{M,q,\sigma} \sqcup V'_{M,q,\sigma})) \sqsubseteq R\mathsf{ok} \\ A_0 \sqcap \cdots \sqcap A_{n-1} \sqcap A_\sigma \sqcap A'_\sigma \sqcap L\mathsf{ok} \sqcap R\mathsf{ok} \sqsubseteq V' \\ \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_\sigma \sqcap A'_\sigma \sqsubseteq V' \\ \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_\sigma \sqcap A'_{q,\sigma'} \sqsubseteq V'_{L,\sigma} \\ \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V_{R,q} \sqcap A_\sigma \sqcap A'_{q',\sigma'} \sqsubseteq V'_{R,q,\sigma} \\ \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V'_{M,q,\sigma} \sqcap A_{\sigma'} \sqcap A'_{\sigma'} \sqsubseteq V'_{M,q,\sigma} \end{array}$$

where $\sigma, \sigma', \sigma''$ range over Γ, q and q' over Q, and i over $0..n-1$. We switch to different verification markers V', $V'_{L,\sigma}$, $V'_{R,q}$, $V'_{M,q,\sigma}$ to distinguish halting from non-halting configurations. Note that the first verification step is different for the latter: we expect to see one successor marked X_L and one marked X_R, both the first cell of an already verified (halting or non-halting) configuration. For easier construction, we require two successors also for existential configurations; they can simply be identical. The above inclusions do not yet deal with cells where the head is currently located. We need some prerequisites because when verifying these cells, we want to (locally) verify the transition relation. For this purpose, we carry the transitions implemented locally at a configuration up to its predecessor configuration:

$$\begin{array}{r} \exists R.\exists R.(X_M \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap V_{q,\sigma,M'}) \sqsubseteq S^M_{q,\sigma,M'} \\ \exists R.\exists R.(X_M \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap V'_{q,\sigma,M'}) \sqsubseteq S^M_{q,\sigma,M'} \\ \exists R.(A_\sigma \sqcap S^M_{q,\sigma',M}) \sqsubseteq S^M_{q,\sigma',M} \end{array}$$

where q ranges over Q, σ and σ' over Γ, M over $\{L, R\}$, and i over $0..n-1$. Note that markers are propagated up exactly to the head position. One issue with the above is that additional $S_{q\sigma,M}$-markers could be propagated up not from the successors that we have verified, but from surplus (unverified) successors. To prevent such undesired markers, we put

$$S^M_{q_1,\sigma_1,M_1} \sqcap S^M_{q_2,\sigma_2,M_2} \sqsubseteq \bot$$

for all $M \in \{L, R\}$ and all distinct $(q_1, \sigma_1, M_1), (q_2, \sigma_2, M_2) \in Q \times \Gamma \times \{L, R\}$. We can now implement the verification of cells under the head in non-halting configurations. Put

$$\begin{array}{rcl} \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^L_{q_2,\sigma_2,M_2} \sqcap S^R_{q_3,\sigma_3,M_3} & \sqsubseteq & V'_{R,q_1} \\ \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V'_{L,\sigma} \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^L_{q_2,\sigma_2,M_2} \sqcap S^R_{q_3,\sigma_3,M_3} & \sqsubseteq & V'_{L,q_1,\sigma} \end{array}$$

for all $(q_1, \sigma_1) \in Q \times \Gamma$ with q_1 a universal state and $\Delta(q_1, \sigma_1) = \{(q_2, \sigma_2, M_2), (q_3, \sigma_3, M_3)\}$, i from $0..n-1$, and σ from Γ; moreover, put

$$\begin{array}{rcl} \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^L_{q_2,\sigma_2,M_2} \sqcap S^R_{q_2,\sigma_2,M_2} & \sqsubseteq & V'_{R,q_1} \\ \mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V'_{L,\sigma} \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^L_{q_2,\sigma_2,M_2} \sqcap S^R_{q_2,\sigma_2,M_2} & \sqsubseteq & V'_{L,q,\sigma} \end{array}$$

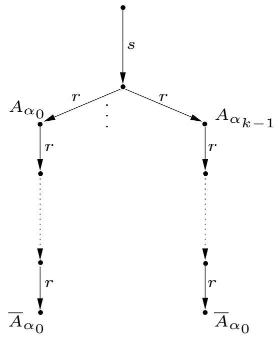


Figure 8: Tree gadget.

for all $(q_1, \sigma_1) \in Q \times \Gamma$ with q_1 an existential state, for all $(q_2, \sigma_2, M_2) \in \Delta(q_1, \sigma_1)$, all i from $0..n-1$, and all σ from Γ. It remains to verify the initial configuration. Let $w = \sigma_0 \cdots \sigma_{n-1}$, let $(C = i)$ be the conjunction over the concept names A_i, $\overline{A}_i$ that expresses i in binary for $0 \leq i < n$, and let $(C \geq n)$ be the Boolean concept over the concept names A_i, $\overline{A}_i$ which expresses that the counter value is at least n. Then put

$$
\begin{array}{c}
A_0 \sqcap \cdots \sqcap A_{n-1} \sqcap A_{\Box} \sqcap A'_\sigma \sqcap L\mathsf{ok} \sqcap R\mathsf{ok} \sqsubseteq V^I \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap (C \geq n) \sqcap \exists R.V^I \sqcap A_{\Box} \sqcap A'_\sigma \sqsubseteq V^I \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap (C = i) \sqcap \exists R.V^I \sqcap A_{\sigma_i} \sqcap A'_\sigma \sqsubseteq V^I \\
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap (C = 1) \sqcap \exists R.V^I \sqcap A_{\sigma_1} \sqcap A'_{q,\sigma'} \sqsubseteq V^I_{R,q}
\end{array}
$$

where i ranges over $2..n-1$ and σ, σ' over Γ. This verifies the initial conditions except for the left-most cell, where the head must be located (in initial state q_0) and where we must verify the transition, as in all other configurations. Recall that we assume q_0 to be an existential state. We can thus add

$$
\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap (C = 0) \sqcap \exists R.V^I_{R,q} \sqcap A_{q_0,\sigma_0} \sqcap A'_\sigma \sqcap S^L_{q,\sigma,M} \sqcap S^R_{q,\sigma,M} \quad \sqsubseteq \quad I
$$

for all $(q, \sigma, M) \in \Delta(q_0, \sigma_0)$ and $\sigma \in \Gamma$.

At this point, we have finished the verification of the computation tree, except that we have assumed but not yet established $(*)$. To achieve $(*)$, we use both $\mathcal{T}_1$ and $\mathcal{T}_2$. Let $\alpha_0, \ldots, \alpha_{k-1}$ be the elements of $\Gamma \cup (Q \times \Gamma)$. We use concept names A^ℓ_i, $\overline{A}^\ell_i$, $\ell \in \{0, \ldots, k-1\}$, to implement k additional counters. This time, we have to count up to $2^n + 1$ (because successor configuration sequences are separated by two edges), so i ranges from 0 to $m := \lceil \log(2^n + 1) \rceil$. We first add to $\mathcal{T}_2$:

$$
\begin{array}{rcl}
\exists R.I & \sqsubseteq & \exists S. \bigsqcap_{\ell < k} \exists R.(A_{\alpha_\ell} \sqcap (C^\ell = 0)) \\
\overline{A}^\ell_i & \sqsubseteq & \exists R.\top \\
A^\ell_i \sqcap \bigsqcap_{j<i} A^\ell_j & \sqsubseteq & \forall R.\overline{A}^\ell_i \\
\overline{A}^\ell_i \sqcap \bigsqcap_{j<i} A^\ell_j & \sqsubseteq & \forall R.A^\ell_i \\
A^\ell_i \sqcap \bigsqcup_{j<i} \overline{A}^\ell_j & \sqsubseteq & \forall R.A^\ell_i \\
\overline{A}^\ell_i \sqcap \bigsqcup_{j<i} \overline{A}^\ell_j & \sqsubseteq & \forall R.\overline{A}^\ell_i \\
(C^\ell = 2^n) & \sqsubseteq & \overline{A}_{\alpha_\ell}
\end{array}
$$

where ℓ ranges over $0..k-1$, i over $0..m-1$, and $(C^\ell = j)$ denotes the conjunction over A^ℓ_i, $\overline{A}^\ell_i$ which expresses that the value of the ℓ-th counter is j. We will explain shortly why we need to travel one more R-step (in the first line) after seeing I.

The above inclusions generate, after the verification of the computation tree has ended successfully, a tree in the canonical model of the input ABox and of $\mathcal{T}_2$ as shown in Figure 8. Note that the topmost edge is labeled with the role name S, which is *not* in Σ. By Condition (b) above and since, up to now, we have always only used non$-\Sigma$-symbols on the right-hand side of concept inclusions, we must not (homomorphically) find the subtree rooted at the node with the incoming S-edge *anywhere* in the canonical model of the ABox and $\mathcal{T}_1$. We use this effect which we to ensure that $(*)$ is satisfied *everywhere*. Note that, the paths in Figure 8 have length $2^n + 1$ and that we do not display the labeling with the concept names A_i^ℓ, $\overline{A}_i^\ell$. These concept names are not in Σ anyway and only serve the purpose of achieving the intended path length. Intuitively, every path in the tree represents one possible *copying defect*. The concept names of the form $\overline{A}_\alpha$ need not occur in the input ABox and stand for the disjunction over all $\overline{A}_\beta$ with $\beta \neq \alpha$. They need to be in Σ, though, because we want them to be taken into account in Σ-homomorphisms.

We next extend $\mathcal{T}_1$ as follows:

$$\begin{array}{rcl}
A_\alpha & \sqsubseteq & \overline{A}_\beta \\
\exists R.A_{\alpha_i} & \sqsubseteq & \displaystyle\bigsqcap_{\ell \in \{0,\dots,k-1\} \setminus \{i\}} \exists R.(A_{\alpha_\ell} \sqcap (C^\ell = 0)) \\
\overline{A}_i^\ell & \sqsubseteq & \exists R.\top \\
A_i^\ell \sqcap \displaystyle\bigsqcap_{j<i} A_j^\ell & \sqsubseteq & \forall R.\overline{A}_i^\ell \\
\overline{A}_i^\ell \sqcap \displaystyle\bigsqcap_{j<i} A_j^\ell & \sqsubseteq & \forall R.A_i^\ell \\
A_i^\ell \sqcap \displaystyle\bigsqcup_{j<i} \overline{A}_j^\ell & \sqsubseteq & \forall R.A_i^\ell \\
\overline{A}_i^\ell \sqcap \displaystyle\bigsqcup_{j<i} \overline{A}_j^\ell & \sqsubseteq & \forall R.\overline{A}_i^\ell \\
(C^\ell = 2^n) & \sqsubseteq & \overline{A}_{\alpha_\ell}
\end{array}$$

where ℓ ranges over $0..k-1$, i over $0..m-1$, and α, β over distinct elements of $\Gamma \cup (Q \times \Gamma)$. Note that it is not important to use the same counter concepts A_i^ℓ, $\overline{A}_i^\ell$ in $\mathcal{T}_1$ and $\mathcal{T}_2$: since they are not in Σ, one could as well use different ones. Also note that the intended behaviour of the concept names $\overline{A}_\alpha$ is implemented in the first line.

The idea for achieving $(*)$ is as follows: the tree shown in Figure 8 contains all possible copying defects, that is, all paths of length $2^n + 1$ such that A_{α_i} is true at the beginning, but some A_{α_j} with $j \neq i$ is true at the end. At each point of the computation tree where some A_{α_i} is true at an R-predecessor, the above inclusions in $\mathcal{T}_1$ generate a tree in the canonical model of the ABox and of $\mathcal{T}_1$ which is similar to that in Figure 8, except that the initial S-edge and the path representing an A_{α_i}-defect are missing. Consequently, if A_{α_i} is not properly copied to A'_{α_i} at all nodes that are $2^n + 1$ R-steps away, then we homomorphically find the tree from Figure 8 in the canonical model of the ABox and of $\mathcal{T}_1$. Consequently, not finding the tree anywhere in that model means that all copying is done correctly.

We need to avoid that the inclusions in $\mathcal{T}_1$ enable a homomorphism from the tree in Figure 8 due to an ABox where some node has two R-successors labeled with different concepts A_α, A_β:

$$\exists R.A_\alpha \sqcap \exists R.A_\beta \sqsubseteq \bot.$$

This explains why we need to separate successor configurations by two R-steps. In fact, the mid point needs not make true any of the concept names A_α and thus we are not forced to violate the above constraint when branching at the end of configuration sequences. Also note that copying the content of the first cell of the initial configuration requires traveling one more R-step after seeing I, as implemented above.

Lemma 41. *The following conditions are equivalent:*

1. *there is a tree-shaped Σ-ABox $\mathcal{A}$ such that*
   (a) *$\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$ and*
   (b) *$\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not Σ-homomorphically embeddable into $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$;*
2. *M accepts w.*

Proof. (sketch) For the direction "2 ⇒ 1", assume that M accepts w. An accepting computation tree of M on w can be represented as a Σ-ABox as detailed above alongside the construction of the TBoxes $\mathcal{T}_2$ and $\mathcal{T}_1$. The representation only uses the role name R and the concept names of the form A_i, $\overline{A}_i$, A_σ, $A_{q,\sigma}$, A'_σ, $A'_{q,\sigma}$, X_L, and X_R, but not the concept names of the form $\overline{A}_\sigma$ and $\overline{A}_{q,\sigma}$. As explained above, we need to duplicate the successor configurations of existential configurations to ensure that there is binary branching after each configuration. Also, we need to add one additional incoming R-edge to the root of the tree as explained above.

The resulting ABox $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$. Moreover, since there are no copying defects, there is no homomorphism from $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

For the direction "1 $\Rightarrow$ 2", assume that there is a tree-shaped Σ-ABox $\mathcal{A}$ that satisfies Conditions (a) and (b). Because of Condition (b), I must be true somewhere in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$: otherwise, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ does not contain anonymous elements and the identity is a homomorphism from $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$, contradicting (b). Since I is true somewhere in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ and by construction of $\mathcal{T}_2$, the ABox must contain the representation of a computation tree of M on w, except satisfaction of $(*)$. For the same reason, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ must contain a tree as shown in Figure 8. As has already been argued during the construction of $\mathcal{T}_2$ and $\mathcal{T}_1$, however, condition $(*)$ follows from the existence of such a tree in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ together with (b). □

We remark that the above reduction also yields 2ExpTime hardness for CQ entailment in $\mathcal{ELI}$. In fact, universal restrictions on the right-hand sides of concept inclusions can easily be simulated using universal roles and disjunctions on the left-hand sides can be removed with only a polynomial blowup (since there are always only two disjuncts). It thus remains to eliminate $\bot$, which only occurs non-nested on the right-hand side of concept inclusions. With the exception of the inclusions

$$S^M_{q_1,\sigma_1,M_1} \sqcap S^M_{q_2,\sigma_2,M_2} \sqsubseteq \bot,$$

this can be done as follows: include all concept inclusions with $\bot$ on the right-hand side in $\mathcal{T}_1$ instead of in $\mathcal{T}_2$; then replace $\bot$ with D and add the following concept inclusions to $\mathcal{T}_1$:

$$\begin{array}{rcl}
D & \sqsubseteq & \exists S.\bigsqcap_{\ell<k} \exists R.(A_{\alpha_\ell} \sqcap (C^\ell = 0)) \\
\overline{A}^\ell_i & \sqsubseteq & \exists R.\top \\
A^\ell_i \sqcap \bigsqcap_{j<i} A^\ell_j & \sqsubseteq & \forall R.\overline{A}^\ell_i \\
\overline{A}^\ell_i \sqcap \bigsqcap_{j<i} A^\ell_j & \sqsubseteq & \forall R.A^\ell_i \\
A^\ell_i \sqcap \bigsqcup_{j<i} \overline{A}^\ell_j & \sqsubseteq & \forall R.A^\ell_i \\
\overline{A}^\ell_i \sqcap \bigsqcup_{j<i} \overline{A}^\ell_j & \sqsubseteq & \forall R.\overline{A}^\ell_i \\
(C^\ell = 2^n) & \sqsubseteq & \overline{A}_{\alpha_\ell}
\end{array}$$

where ℓ ranges over $0..k-1$ and i over $0..m$. The effect of these additions is that any ABox which satisfies the left-hand side of a $\bot$-concept inclusion in the original $\mathcal{T}_2$ cannot satisfy Condition (b) from Lemma 41 and thus needs not be considered.

For the inclusions excluded above, a different approach needs to be taken. Instead of introducing the concept names $S^M_{q_1,\sigma_1,M_1}$, one would propagate transitions inside the V'-markers. Thus, $S^L_{q_1,\sigma_1,M_1}$, $S^R_{q_2,\sigma_2,M_2}$, and V' would be integrated into a single marker $V'_{q_1,\sigma_1,M_1,q_2,\sigma_2,M_2}$, and likewise for $V_{R,q}$. The concept inclusion excluded above can then simply be dropped.