Artificial Intelligence 272 (2019) 1–51

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Query inseparability for $\mathcal{ALC}$ ontologies

Elena Botoeva [a,b], Carsten Lutz [c], Vladislav Ryzhikov [d], Frank Wolter [e,*], Michael Zakharyaschev [d]

[a] KRDB Research Centre, Free University of Bozen-Bolzano, Italy
[b] Department of Computing, Imperial College London, UK
[c] Fachbereich Informatik, University of Bremen, Germany
[d] Department of Computer Science and Information Systems, Birkbeck, University of London, UK
[e] Department of Computer Science, University of Liverpool, UK

ARTICLE INFO

Article history:
Received 6 July 2017
Received in revised form 3 June 2018
Accepted 24 September 2018
Available online 30 January 2019

Keywords:
Description logic
Knowledge base
Conjunctive query
Query inseparability
Computational complexity
Tree automaton

ABSTRACT

We investigate the problem whether two $\mathcal{ALC}$ ontologies are indistinguishable (or inseparable) by means of queries in a given signature, which is fundamental for ontology engineering tasks such as ontology versioning, modularisation, update, and forgetting. We consider both knowledge base (KB) and TBox inseparability. For KBs, we give model-theoretic criteria in terms of (finite partial) homomorphisms and products and prove that this problem is undecidable for conjunctive queries (CQs), but 2ExpTime-complete for unions of CQs (UCQs). The same results hold if (U)CQs are replaced by rooted (U)CQs, where every variable is connected to an answer variable. We also show that inseparability by CQs is still undecidable if one KB is given in the lightweight DL $\mathcal{EL}$ and if no restrictions are imposed on the signature of the CQs. We also consider the problem whether two $\mathcal{ALC}$ TBoxes give the same answers to any query over any ABox in a given signature and show that, for CQs, this problem is undecidable, too. We then develop model-theoretic criteria for *Horn*$\mathcal{ALC}$ TBoxes and show using tree automata that, in contrast, inseparability becomes decidable and 2ExpTime-complete, even ExpTime-complete 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 (see, e.g., [1–3] and references therein), 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 answers to queries. A major research effort has led to the development of efficient querying algorithms and tools for a number of DLs ranging from DL-Lite [4–6] via more expressive Horn DLs such as Horn$\mathcal{ALC}$ [7,8] to DLs with full Boolean constructors including $\mathcal{ALC}$ and extensions such as $\mathcal{SHIQ}$ [9,10].

While query answering with DLs is now well-developed, this is much less the case for reasoning services that support ontology engineering when ontologies are used to query data. Important ontology engineering tasks include ontology versioning [11–15], ontology modularisation [16–20], ontology revision and update [21–24], and forgetting in ontolo-

* Corresponding author.
E-mail addresses: botoeva@inf.unibz.it, e.botoeva@imperial.ac.uk (E. Botoeva), clu@informatik.uni-bremen.de (C. Lutz), vlad@dcs.bbk.ac.uk (V. Ryzhikov), wolter@liverpool.ac.uk (F. Wolter), michael@dcs.bbk.ac.uk (M. Zakharyaschev).

https://doi.org/10.1016/j.artint.2018.09.003

gies [25–31]. A fundamental reasoning problem in all these tasks is to *compare* two ontologies. For example, in ontology versioning, the user is interested in comparing two versions of an ontology and understanding the relevant difference between them. In ontology modularisation, the relevant consequences of the full ontology should be preserved when it is replaced by a module. In ontology revision and update, one typically minimises the relevant difference between the updated or revised ontology and the original ontology while taking into account new knowledge. In ontology forgetting, one constructs a new ontology, which is indistinguishable from the original ontology with respect to a signature of interest. The relevant consequences that should be considered when comparing two ontologies depend on the application. In the context of querying data via ontologies, it is natural to consider the answers the ontologies give to queries. Then, in ontology versioning, the relevant difference between two versions of an ontology is based on the queries that receive distinct answers with respect to the ontology versions. In ontology modularisation, it is the answers to queries that should be preserved when a module is extracted from an ontology. In ontology update or revision, the difference between the answers to queries over the updated or revised ontology and the original one should be minimised when constructing update or revision operators. Similarly, in forgetting, it is the answers to queries which should be preserved under appropriate forgetting operators. Thus, in the context of query answering, the fundamental relationship between ontologies is not whether they are logically equivalent (have the same models), but whether they give the same answers to any relevant query. To illustrate, consider the following simple TBox

$$\mathcal{T} = \{Book \sqsubseteq \exists author.\neg Book\}$$

saying that every book has an author who is not a book. Clearly, $\mathcal{T}$ is not logically equivalent to the TBox

$$\mathcal{T}' = \{Book \sqsubseteq \exists author.\top\},$$

which only states that every book has an author. However, if one takes as the query language the popular classes of conjunctive queries (CQs) or unions of CQs (UCQs), then no matter what the data is, every query will have the same answers independently of whether one uses $\mathcal{T}$ or $\mathcal{T}'$. Intuitively, the reason is that the 'positive' information given by $\mathcal{T}$ coincides with the 'positive' information given by $\mathcal{T}'$. If the main purpose of the ontology is answering UCQs, it is thus more important to know that $\mathcal{T}$ can be safely replaced by $\mathcal{T}'$ without affecting the answers to UCQs than to establish that $\mathcal{T}$ and $\mathcal{T}'$ are not logically equivalent.

In most ontology engineering applications for ontology-based data access, the relevant class $\mathcal{Q}$ of queries can be further restricted to those given in a finite signature of relevant concept and role names. For example, to establish that a subset $\mathcal{M}$ of an ontology $\mathcal{O}$ is a module of $\mathcal{O}$, one should not require that $\mathcal{M}$ and $\mathcal{O}$ give the same answers to all queries in $\mathcal{Q}$, but only to those that are in the signature of $\mathcal{M}$. Similarly, in the versioning context, often only the answers to queries in $\mathcal{Q}$ given in a small signature containing a fraction of the concept and role names of the ontology are relevant for the application, and so for the difference that should be presented to a user.

The resulting entailment problem can be formalised in two ways. Recall that, in DL, a knowledge base (KB) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. Now, given a class $\mathcal{Q}$ of queries, KBs $\mathcal{K}_1$ and $\mathcal{K}_2$, 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. Since a KB includes an ABox, this notion of entailment is appropriate if the data is known while the ontology engineering task is completed and does not change frequently. This is the case for many real-world ontologies, which not only provide a conceptual model of the domain of interest, but also introduce the individuals relevant for the domain and their properties. In addition to versioning, modularisation, revision, update, and forgetting, applications of Σ-KB entailment and Σ-KB inseparability also include knowledge exchange [32–34], where a user wants to transform a KB $\mathcal{K}_1$ given in a signature Σ_1 to a KB $\mathcal{K}_2$ in a new signature Σ_2 connected to Σ_1 using a mapping $\mathcal{M}$, also known as an ontology alignment or ontology matching [35]. The condition that the target KB $\mathcal{K}_2$ is a sound and complete representation of $\mathcal{K}_1$ under $\mathcal{M}$ with respect to the answers to a class $\mathcal{Q}$ of relevant queries can then be formulated as the condition that $\mathcal{K}_1 \cup \mathcal{M}$ and $\mathcal{K}_2$ are Σ_2-$\mathcal{Q}$ inseparable [34]. The following simple example illustrates the notion of KB inseparability.

Example 1. Suppose we are given the KBs $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$, where

$$\mathcal{T}_1 = \{Lecturer \sqsubseteq \forall teaches.(Undergraduate \sqcup Graduate)\}, \qquad \mathcal{T}_2 = \emptyset,$$

$$\mathcal{A} = \{Lecturer(a),\ teaches(a, b)\}.$$

Then $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ-CQ inseparable, for any signature Σ. However, they are not Σ-UCQ inseparable for the signature Σ containing the concept names *Undergraduate* and *Graduate*. To see this, consider the Σ-UCQ

$$\boldsymbol{q}(x) = Undergraduate(x) \vee Graduate(x).$$

Clearly, b is an answer to $\boldsymbol{q}(x)$ over $\mathcal{K}_1$, but not over $\mathcal{K}_2$.

Table 1
KB query inseparability.

Queries	$\mathcal{ALC}$ and $\mathcal{ALC}$	$\mathcal{ALC}$ and $\mathcal{EL}$
CQ and rCQ	undecidable	undecidable
UCQ and rUCQ	2ExpTime-complete	in 2ExpTime

Table 2
TBox query inseparability.

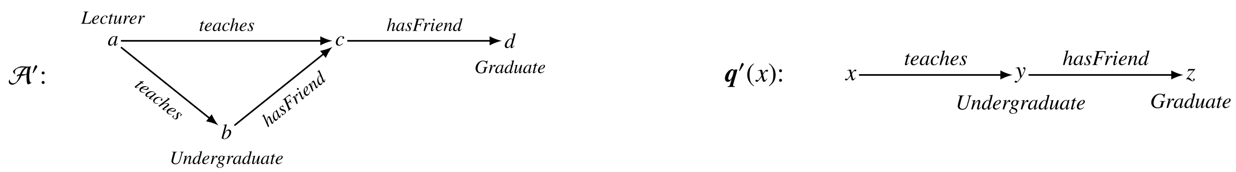
Queries	$\mathcal{ALC}$ and $\mathcal{ALC}$	$\mathcal{ALC}$ and $\mathcal{EL}$	Horn$\mathcal{ALC}$ and Horn$\mathcal{ALC}$
CQs	undecidable	undecidable	2ExpTime-complete
rCQs	undecidable	undecidable	ExpTime-complete

KB entailment and inseparability are appropriate if the data is known and does not change frequently. If, however, the data is not known or tends to change, 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 a relevant signature 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 $(\mathcal{T}_2, \mathcal{A})$. TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ are Θ-$\mathcal{Q}$ *inseparable* if they Θ-$\mathcal{Q}$ entail each other.

Example 2. Consider again the TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ from Example 1. Clearly, $\mathcal{T}_1$ and $\mathcal{T}_2$ are not (Σ_0, Σ_1)-UCQ inseparable for $\Sigma_0 = \{\mathit{Lecturer}, \mathit{teaches}\}$ and $\Sigma_1 = \{\mathit{Undergraduate}, \mathit{Graduate}\}$ as we have seen a Σ_0-ABox $\mathcal{A}$ for which $(\mathcal{T}_1, \mathcal{A})$ and $(\mathcal{T}_2, \mathcal{A})$ are not Σ_1-UCQ inseparable. Notice, however, that $\mathcal{T}_1$ and $\mathcal{T}_2$ are both (Σ_0, Σ_0)-UCQ and (Σ_1, Σ_1)-UCQ inseparable. On the other hand, it is not difficult to see that $\mathcal{T}_1$ and $\mathcal{T}_2$ are (Σ_0, Σ_1)-CQ inseparable. The situation changes drastically if the ABox can contain additional role names, for instance *hasFriend*. Indeed, suppose $\Sigma_2 = \Sigma_0 \cup \Sigma_1 \cup \{\mathit{hasFriend}\}$. Then $\mathcal{T}_1$ and $\mathcal{T}_2$ are (Σ_2, Σ_2)-CQ separable by the ABox $\mathcal{A}'$ shown in the picture below and the CQ

$$\boldsymbol{q}'(x) = \exists y \exists z\, \big(\mathit{teaches}(x, y) \land \mathit{Undergraduate}(y) \land \mathit{hasFriend}(y, z) \land \mathit{Graduate}(z)\big)$$

since a is returned as an answer to $\boldsymbol{q}'(x)$ over $(\mathcal{T}_1, \mathcal{A}')$ but not over $(\mathcal{T}_2, \mathcal{A}')$. (This example is a variant of the well-known [36, Example 4.2.5].)



In this paper, we investigate entailment and inseparability for KBs and TBoxes and for queries that are CQs or UCQs. In practice, the majority of queries are *rooted* in the sense that every variable is connected to an answer variable. We therefore also consider the classes of rooted CQs (rCQs) and UCQs (rUCQs). So far, query entailment and inseparability have been studied for Horn DL KBs [37], $\mathcal{EL}$ TBoxes [38,15], DL-Lite TBoxes [39], and also for OBDA specifications, that is, DL-Lite TBoxes with mappings [40]; for a recent survey see [41]. No results are yet 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. Here, we make first steps towards understanding query entailment and inseparability in these cases. To begin with, we give model-theoretic characterisations of these notions for $\mathcal{ALC}$ and Horn$\mathcal{ALC}$ in terms of (finite partial) homomorphisms and products of interpretations. The obtained characterisations together with various types of automata are then used to investigate the computational complexity of deciding query entailment and inseparability. Our main results on KB and TBox inseparabilities are summarised in Tables 1 and 2, respectively:

Three of these results came as a real surprise to us. First, it turned out that CQ and rCQ inseparability between $\mathcal{ALC}$ KBs is undecidable, even if one of the KBs is formulated in the lightweight DL $\mathcal{EL}$ and without any signature restriction. This should be contrasted with the decidability of subsumption-based entailment between $\mathcal{ALC}$ TBoxes [42] (and even theories in guarded fragments of FO [43]) and of CQ entailment between Horn$\mathcal{ALC}$ KBs [37]. The second surprising result is that inseparability between $\mathcal{ALC}$ KBs becomes decidable when CQs are replaced with UCQs or rUCQs. In fact, we show that inseparability is 2ExpTime-complete for both UCQs and rUCQs. An even more fine-grained picture is obtained by considering entailment instead of inseparability. It turns out that (r)CQ entailment of Horn$\mathcal{ALC}$ KBs by $\mathcal{ALC}$ KBs coincides with (r)UCQ entailment of Horn$\mathcal{ALC}$ KBs by $\mathcal{ALC}$ KBs and is 2ExpTime-complete, but that in contrast (r)CQ entailment of $\mathcal{ALC}$ KBs by Horn$\mathcal{ALC}$ KBs is undecidable.

For $\mathcal{ALC}$ TBoxes, CQ and rCQ entailment as well as CQ and rCQ inseparability are undecidable as well. We obtain decidability for Horn$\mathcal{ALC}$ TBoxes (where CQ and UCQ entailments coincide) using the fact that non-entailment is always witnessed by tree-shaped ABoxes. As another surprise, CQ inseparability of *Horn*$\mathcal{ALC}$ TBoxes is 2ExpTime-complete while rCQ-entailment is only ExpTime-complete. This applies to CQ entailment and rCQ entailment as well. This result should be contrasted with the $\mathcal{EL}$ case, where both problems are ExpTime-complete [38]. Table 2 does not contain any results in the UCQ case, as the decidability of UCQ entailment and inseparability between $\mathcal{ALC}$ TBoxes remains open.

We now discuss the structure and contributions of this paper in more detail. Section 2 defines the DLs we are interested in, which range from $\mathcal{EL}$ to *Horn*$\mathcal{ALC}$ and $\mathcal{ALC}$. It also introduces query answering for DL KBs and provides basic completeness results and homomorphism characterisations for query answering. Section 3 defines query entailment and inseparability between DL KBs. It provides illustrating examples and characterises UCQ entailment in terms of finite partial homomorphisms between models of KBs. To characterise CQ entailment, products of KB models are also required. The difference between the characterisations will play a crucial role in our algorithmic analysis of entailment. In some important cases later on in the paper, finite partial homomorphisms are replaced by full homomorphisms using, for example, automata-theoretic techniques and, in particular, Rabin's result that any tree automaton that accepts some tree accepts already a regular tree. This move from finite partial homomorphisms to full homomorphisms is non-trivial and crucial for our decision procedures.

In Section 4, we prove the undecidability of (r)CQ entailment of an $\mathcal{ALC}$ KB by an $\mathcal{EL}$ KB using a reduction of an undecidable tiling problem. The direction is important, as we prove later that (r)CQ entailment of an $\mathcal{EL}$ KB by an $\mathcal{ALC}$ KB is decidable (in 2ExpTime). We also prove undecidability of CQ inseparability between $\mathcal{EL}$ and $\mathcal{ALC}$ KBs. The model-theoretic characterisation of (r)CQ entailment via products and finite homomorphisms is crucial for these proofs. We then use a 'hiding technique' replacing concept names by complex concepts to extend the undecidability results to the full signature. Thus, for example, even without any restriction on the signature it is undecidable whether two $\mathcal{ALC}$ KBs are (r)CQ inseparable.

In Section 5, we first show that, in the (r)UCQ case, partial homomorphisms can be replaced by full homomorphisms in the model-theoretic characterisation of rUCQ entailment between $\mathcal{ALC}$ KBs if one considers regular tree-shaped models of the KBs. This result is then used to encode the UCQ entailment problem into an emptiness problem for two-way alternating parity automata on infinite trees (2APTAs). Using results from automata theory we then obtain a 2ExpTime upper bound for (r)UCQ entailment between $\mathcal{ALC}$ KBs and a characterisation of (r)UCQ entailment with full homomorphisms that does not require the restriction to regular tree-shaped models. We prove that the 2ExpTime upper bound is tight by a reduction of the word problem for alternating Turing machines. Finally, we show using the hiding technique that the 2ExpTime lower bounds still hold without restrictions on the signature.

In Section 6, we introduce query entailment and inseparability between TBoxes and prove that the undecidability results for (r)CQ entailment and (r)CQ inseparability can be lifted from KBs to TBoxes. In this case, however, undecidability without any restrictions regarding the signatures remains open. In Section 7, we develop model-theoretic criteria for (r)CQ entailment of *Horn*$\mathcal{ALC}$ TBoxes by $\mathcal{ALC}$ TBoxes. The crucial observation is that it suffices to consider tree-shaped ABoxes when searching for counterexamples to (r)CQ entailment between TBoxes. This allows us to use, in Section 8, automata on trees to decide (r)CQ entailment.

In Section 8, we first prove an ExpTime upper bound for rCQ entailment of *Horn*$\mathcal{ALC}$ TBoxes by $\mathcal{ALC}$ TBoxes via an encoding into emptiness problems for a mix of two-way alternating Büchi automata and non-deterministic top-down tree automata on finite trees (that represent tree-shaped ABoxes). As satisfiability of *Horn*$\mathcal{ALC}$ TBoxes is ExpTime-hard already, this bound is tight. We then consider arbitrary (not necessarily rooted) CQs and extend the previous encoding into emptiness problems for tree automata to this case, thereby obtaining a 2ExpTime upper bound. Here, it is non-trivial to show that this bound is tight. We use a reduction of alternating Turing machines to prove the corresponding 2ExpTime lower bound (also for CQ inseparability).

We conclude in Section 9 by discussing open problems. A small number of proofs that follow ideas presented in the main paper are deferred to the appendix. An extended abstract with initial results that led to this paper was presented at IJCAI 2016 [44].

2. Preliminaries

In DL, knowledge is represented by means of concepts and roles that are defined inductively starting from a countably infinite set $\mathsf{N_C}$ of *concept names* and a countably-infinite set $\mathsf{N_R}$ of *role names*, and using a set of concept and role constructors [45]. Different sets of concept and role constructors give rise to different DLs.

We begin by introducing the description logic $\mathcal{ALC}$. The concept constructors available in $\mathcal{ALC}$ are shown in Table 3, where R is a role name and C, D are concepts. A concept built using these constructors is called an *$\mathcal{ALC}$-concept*. $\mathcal{ALC}$ does not have any role constructors. An *$\mathcal{ALC}$ TBox* is a finite set of *$\mathcal{ALC}$ concept inclusions* (CIs) of the form $C \sqsubseteq D$ and *$\mathcal{ALC}$ concept equivalences* (CEs) $C \equiv D$. (A CE $C \equiv D$ will be regarded as an abbreviation for the two CIs $C \sqsubseteq D$ and $D \sqsubseteq C$.) The *size* $|\mathcal{T}|$ of a TBox $\mathcal{T}$ is the number of occurrences of symbols in $\mathcal{T}$.

The semantics of TBoxes is given by *interpretations* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where the *domain* $\Delta^{\mathcal{I}}$ is a non-empty set and the *interpretation function* $\cdot^{\mathcal{I}}$ maps each concept name $A \in \mathsf{N_C}$ to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$, and each role name $R \in \mathsf{N_R}$ to a binary relation $R^{\mathcal{I}}$ on $\Delta^{\mathcal{I}}$. The extension of $\cdot^{\mathcal{I}}$ to arbitrary concepts is defined inductively as shown in the third column of Table 3. We say that an interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and that $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ if $\mathcal{I}$ satisfies all the CIs in $\mathcal{T}$. A TBox is *consistent* (or *satisfiable*) if it has a model. A concept C is *satisfiable with respect to* $\mathcal{T}$ if there exists a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^{\mathcal{I}} \neq \emptyset$. A concept C is *subsumed by a concept D with respect to* $\mathcal{T}$ ($\mathcal{T} \models C \sqsubseteq D$, in symbols) if every model

Table 3
Syntax and semantics of $\mathcal{ALC}$.

Name	Syntax	Semantics
top concept	$\top$	$\Delta^{\mathcal{I}}$
bottom concept	$\bot$	$\emptyset$
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
existential restriction	$\exists R.C$	$\{ d \in \Delta^{\mathcal{I}} \mid \exists e \in C^{\mathcal{I}}\, (d,e) \in R^{\mathcal{I}} \}$
universal restriction	$\forall R.C$	$\{ d \in \Delta^{\mathcal{I}} \mid \forall e \in \Delta^{\mathcal{I}} \big((d,e) \in R^{\mathcal{I}} \to e \in C^{\mathcal{I}}\big) \}$

$\mathcal{I}$ of $\mathcal{T}$ satisfies the CI $C \sqsubseteq D$. For TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, we write $\mathcal{T}_1 \models \mathcal{T}_2$ and say that $\mathcal{T}_1$ *entails* $\mathcal{T}_2$ if $\mathcal{T}_1 \models \alpha$ for all $\alpha \in \mathcal{T}_2$. TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ are *logically equivalent* if they have the same models. This is the case if and only if $\mathcal{T}_1$ entails $\mathcal{T}_2$, and vice versa.

We next define two syntactic fragments of $\mathcal{ALC}$ for which query answering (see below) is tractable in data complexity. The fragment of $\mathcal{ALC}$ obtained by disallowing the constructors $\bot$, $\neg$, $\sqcup$ and $\forall$ is known as $\mathcal{EL}$. Thus, $\mathcal{EL}$ concepts are constructed using $\top$, $\sqcap$ and $\exists$ only [46]. A more expressive fragment with tractable query answering is Horn$\mathcal{ALC}$. Following [47,48], we say, inductively, that a concept C occurs positively in C itself and, if C occurs positively (negatively) in C', then

- C occurs positively (respectively, negatively) in $C' \sqcup D$, $C' \sqcap D$, $\exists R.C'$, $\forall R.C'$, $D \sqsubseteq C'$, and
- C occurs negatively (respectively, positively) in $\neg C'$ and $C' \sqsubseteq D$.

Now, we call an $\mathcal{ALC}$ TBox $\mathcal{T}$ *Horn* if no concept of the form $C \sqcup D$ occurs positively in $\mathcal{T}$, and no concept of the form $\neg C$ or $\forall R.C$ occurs negatively in $\mathcal{T}$. In the DL *Horn*$\mathcal{ALC}$, only Horn TBoxes are allowed.

In DL, data is represented in the form of ABoxes. To introduce ABoxes, we fix a countably-infinite set $\mathsf{N_I}$ of *individual names*, which correspond to individual constants in first-order logic. An *assertion* is an expression of the form $A(a)$ or $R(a,b)$, where A is a concept name, R a role name, and a, b individual names. An *ABox* $\mathcal{A}$ is a finite set of assertions. We call the pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ of a TBox $\mathcal{T}$ in a DL $\mathcal{L}$ and an ABox $\mathcal{A}$ an $\mathcal{L}$ *knowledge base* (KB, for short). By $\mathsf{ind}(\mathcal{A})$ and $\mathsf{ind}(\mathcal{K})$, we denote the set of individual names in $\mathcal{A}$ and $\mathcal{K}$, respectively.

To interpret ABoxes $\mathcal{A}$, we consider interpretations $\mathcal{I}$ that map all individual names $a \in \mathsf{ind}(\mathcal{A})$ to elements $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ in such a way that $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if $a \neq b$ (thus, we adopt the *unique name assumption*). It is to be noted that we do not assume all the individual names from $\mathsf{N_I}$ to be interpreted in $\mathcal{I}$. Sometimes, we make the *standard name assumption*, that is, set $a^{\mathcal{I}} = a$, for all the relevant a. Both assumptions are without loss of generality as it is well known, and easy to check, that in $\mathcal{ALC}$ the certain answers to (unions of) conjunctive queries, as defined below, do not depend on the unique name assumption. We say that $\mathcal{I}$ *satisfies* assertions $A(a)$ and $R(a,b)$ if $a^{\mathcal{I}} \in A^{\mathcal{I}}$ and, respectively, $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$. It is a *model* of an ABox $\mathcal{A}$ if it satisfies all the assertions in $\mathcal{A}$, and it is a *model* of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if it is a model of both $\mathcal{T}$ and $\mathcal{A}$. We say that $\mathcal{K}$ is *consistent* (or *satisfiable*) if it has a model. We apply the TBox terminology introduced above to KBs as well. For example, KBs $\mathcal{K}_1$ and $\mathcal{K}_2$ are *logically equivalent* if they have the same models (or, equivalently, entail each other).

We next introduce query answering over KBs, starting with conjunctive queries [49–51]. An *atom* takes the form $A(x)$ or $R(x, y)$, where x, y are from a set of *individual variables* $\mathsf{N_V}$, A is a concept name, and R a role name. A *conjunctive query* (or CQ) is an expression of the form $\boldsymbol{q}(\boldsymbol{x}) = \exists \boldsymbol{y}\, \varphi(\boldsymbol{x}, \boldsymbol{y})$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are disjoint sequences of variables and φ is a conjunction of atoms that only contain variables from $\boldsymbol{x} \cup \boldsymbol{y}$—we (ab)use set-theoretic notation for sequences where convenient. We often write $A(x) \in \boldsymbol{q}$ and $R(x, y) \in \boldsymbol{q}$ to indicate that $A(x)$ and $R(x, y)$ are conjuncts of φ. We call a CQ $\boldsymbol{q}(\boldsymbol{x}) = \exists \boldsymbol{y}\, \varphi(\boldsymbol{x}, \boldsymbol{y})$ *rooted* (or an rCQ) if every $y \in \boldsymbol{y}$ is connected to some $x \in \boldsymbol{x}$ by a path in the undirected 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 the $\boldsymbol{q}_i$ are rooted. If the sequence $\boldsymbol{x}$ is empty, $\boldsymbol{q}(\boldsymbol{x})$ is called a *Boolean* CQ or UCQ. Observe that no Boolean query is rooted.

Example 3. The CQ $\boldsymbol{q}(x_1, x_2) = \exists y_1 \exists y_2 (R(x_1, y_1) \wedge S(x_2, y_2))$ is an rCQ but $\boldsymbol{q}(x_1) = \exists x_2 \exists y_1 \exists y_2 (R(x_1, y_1) \wedge S(x_2, y_2))$ is not an rCQ.

Given a UCQ $\boldsymbol{q}(\boldsymbol{x}) = \bigvee_i \boldsymbol{q}_i(\boldsymbol{x})$ with $\boldsymbol{x} = x_1, \dots, x_k$ and a KB $\mathcal{K}$, a sequence $\boldsymbol{a} = a_1, \dots, a_k$ of individual names from $\mathcal{K}$ is called a *certain answer to* $\boldsymbol{q}(\boldsymbol{x})$ *over* $\mathcal{K}$ if, for every model $\mathcal{I}$ of $\mathcal{K}$, there exist a CQ $\boldsymbol{q}_i$ in $\boldsymbol{q}$ and a map (*homomorphism*) h of its variables to $\Delta^{\mathcal{I}}$ such that $h(x_j) = a_j^{\mathcal{I}}$, for $1 \le j \le k$, $A(z) \in \boldsymbol{q}_i$ implies $h(z) \in A^{\mathcal{I}}$, and $R(z, z') \in \boldsymbol{q}_i$ implies $(h(z), h(z')) \in R^{\mathcal{I}}$. If this is the case, we write $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$. For a Boolean UCQ $\boldsymbol{q}$, we say that the certain answer to $\boldsymbol{q}$ over $\mathcal{K}$ is 'yes' if $\mathcal{K} \models \boldsymbol{q}$ and 'no' otherwise. *CQ* or *UCQ answering* means to decide—given a CQ or UCQ $\boldsymbol{q}(\boldsymbol{x})$, a KB $\mathcal{K}$ and a tuple $\boldsymbol{a}$ from $\mathsf{ind}(\mathcal{K})$—whether $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$.

Example 4. To see that a is a certain answer to the CQ $\boldsymbol{q}'(x)$ over the KB $\mathcal{K} = (\mathcal{T}_1, \mathcal{A}')$ from Example 2, we observe that, by the axiom of $\mathcal{T}_1$, we have $c \in \mathit{Undergraduate}^{\mathcal{I}}$ or $c \in \mathit{Graduate}^{\mathcal{I}}$ in any model $\mathcal{I}$ of $\mathcal{K}$. In the former case, the map h_1 with

$h_1(x) = a$, $h_1(y) = c$ and $h_1(z) = d$ is a homomorphism from $\boldsymbol{q}'$ to $\mathcal{I}$, while in the latter one, h_2 with $h_2(x) = a$, $h_2(y) = b$ and $h_2(z) = c$ is such a homomorphism.

A *signature*, Σ, is a finite set of concept and role names. The *signature* $\mathsf{sig}(C)$ of a concept C is the set of concept and role names that occur in C, and likewise for TBoxes $\mathcal{T}$, CIs $C \sqsubseteq D$, assertions $R(a, b)$ and $A(a)$, ABoxes $\mathcal{A}$, KBs $\mathcal{K}$, UCQs $\boldsymbol{q}$. Note that individual names are not in any signature and, in particular, not in the signature of an assertion, ABox or KB. We are often interested in concepts, TBoxes, KBs, and ABoxes formulated using a specific signature Σ, in which case we use the terms Σ-*concept*, Σ-*TBox*, Σ-*KB*, etc. When dealing with Σ-KBs, it mostly suffices to consider Σ-*interpretations* $\mathcal{I}$ where $X^{\mathcal{I}} = \emptyset$ for all concept and role names $X \notin \Sigma$. A Σ-*model* of a KB is a Σ-interpretation that is a model of the KB. The Σ-*reduct* $\mathcal{J}$ of an interpretation $\mathcal{I}$ is obtained from $\mathcal{I}$ by setting $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$, $A^{\mathcal{J}} = A^{\mathcal{I}}$ for all concept names $A \in \Sigma$, $R^{\mathcal{J}} = R^{\mathcal{I}}$ for all role names $R \in \Sigma$, and $A^{\mathcal{J}} = R^{\mathcal{J}} = \emptyset$ for all remaining concept names A and role names R.

To compute the certain answers to queries over a KB $\mathcal{K}$, it is convenient to work with a 'small' subset $\boldsymbol{M}$ of $\mathsf{sig}(\mathcal{K})$-models of $\mathcal{K}$ that is *complete for* $\mathcal{K}$ in the sense that, for any UCQ $\boldsymbol{q}(\boldsymbol{x})$ and any $\boldsymbol{a} \subseteq \mathsf{ind}(\mathcal{K})$, 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 shall frequently use the following characterisation of complete sets of models based on (partial) homomorphisms.

Suppose $\mathcal{I}$ and $\mathcal{J}$ are interpretations and Σ a signature. A function $h \colon \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$ is called a Σ-*homomorphism* if $u \in A^{\mathcal{I}}$ implies $h(u) \in A^{\mathcal{J}}$ and $(u, v) \in R^{\mathcal{I}}$ implies $(h(u), h(v)) \in R^{\mathcal{J}}$, for all $u, v \in \Delta^{\mathcal{I}}$, Σ-concept names A, and Σ-role names R. If Σ is the set of all concept and role names, then h is called simply a *homomorphism*. We say that h *preserves a set* N *of individual names* if $h(a^{\mathcal{I}}) = a^{\mathcal{J}}$, for all $a \in N$ that are defined in $\mathcal{I}$. It is known from database theory that homomorphisms characterise CQ-containment [52]. To characterise completeness for KBs, we require finite partial homomorphisms. An interpretation $\mathcal{I}$ is a *subinterpretation* of an interpretation $\mathcal{J}$ (*induced by* a set Δ) if $\Delta = \Delta^{\mathcal{I}} \subseteq \Delta^{\mathcal{J}}$, $A^{\mathcal{I}} = A^{\mathcal{J}} \cap \Delta^{\mathcal{I}}$ for all concept names A, $R^{\mathcal{I}} = R^{\mathcal{J}} \cap (\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}})$ for all role names R, and the interpretation $a^{\mathcal{I}}$ of an individual name a is defined exactly if $a^{\mathcal{J}} \in \Delta^{\mathcal{I}}$, in which case $a^{\mathcal{I}} = a^{\mathcal{J}}$. For a natural number n, we say that an interpretation $\mathcal{I}$ is $n\Sigma$-*homomorphically embeddable into an interpretation* $\mathcal{J}$ if, for any subinterpretation $\mathcal{I}'$ of $\mathcal{I}$ with $|\Delta^{\mathcal{I}'}| \leq n$, there is a Σ-homomorphism from $\mathcal{I}'$ to $\mathcal{J}$. If Σ is the set of all concept and role names, then we omit Σ and speak about n-*homomorphic embeddability*. If we require all Σ-homomorphisms to preserve a set N of individual names, then we speak about $n\Sigma$-*homomorphic embeddability preserving* N.

Example 5. Let $\mathcal{I}$ and $\mathcal{J}$ be interpretations whose domain is the set $\mathbb{N}$ of natural numbers and, for any $n, m \in \mathbb{N}$, we have $(n, m) \in R^{\mathcal{I}}$ if $m = n + 1$, and $(n, m) \in R^{\mathcal{J}}$ if $n = m + 1$. Then, for all $n \geq 0$, $\mathcal{I}$ is n-homomorphically embeddable into $\mathcal{J}$, but $\mathcal{I}$ is not homomorphically embeddable into $\mathcal{J}$. Now, let $a^{\mathcal{I}} = 0$, $a^{\mathcal{J}} = m$, and $N = \{a\}$. Then $\mathcal{I}$ is $(m + 1)$-homomorphically embeddable into $\mathcal{J}$ preserving N, but $\mathcal{I}$ is not $(m + 2)$-homomorphically embeddable into $\mathcal{J}$ preserving N.

Proposition 6. *A set* $\boldsymbol{M}$ *of* $\mathsf{sig}(\mathcal{K})$*-models of an* $\mathcal{ALC}$ *KB* $\mathcal{K}$ *is complete for* $\mathcal{K}$ *iff, for any model* $\mathcal{J}$ *of* $\mathcal{K}$ *and any* $n > 0$*, there is* $\mathcal{I} \in \boldsymbol{M}$ *such that* $\mathcal{I}$ *is* n*-homomorphically embeddable into* $\mathcal{J}$ *preserving* $\mathsf{ind}(\mathcal{K})$.

Proof. Let $\Sigma = \mathsf{sig}(\mathcal{K})$ and let $\boldsymbol{M}$ be a class of Σ-models of $\mathcal{K}$. Suppose first that $\boldsymbol{M}$ is not complete for $\mathcal{K}$. Then there exist a UCQ $\boldsymbol{q}(\boldsymbol{x})$ and a tuple $\boldsymbol{a}$ from $\mathsf{ind}(\mathcal{K})$ such that $\mathcal{K} \not\models \boldsymbol{q}(\boldsymbol{a})$ but $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for all $\mathcal{I} \in \boldsymbol{M}$. Let $\mathcal{J}$ be a model of $\mathcal{K}$ such that $\mathcal{J} \not\models \boldsymbol{q}(\boldsymbol{a})$ and let n be the number of variables in $\boldsymbol{q}(\boldsymbol{x})$. For every $\mathcal{I} \in \boldsymbol{M}$, there exists a subinterpretation $\mathcal{I}'$ of $\mathcal{I}$ with $|\Delta^{\mathcal{I}'}| \leq n$ and $\mathcal{I}' \models \boldsymbol{q}(\boldsymbol{a})$. No such $\mathcal{I}'$ is homomorphically embeddable into $\mathcal{J}$ preserving $\boldsymbol{a}$, and so no $\mathcal{I} \in \boldsymbol{M}$ is n-homomorphically embeddable into $\mathcal{J}$ preserving $\mathsf{ind}(\mathcal{K})$.

Conversely, suppose there exists a model $\mathcal{J}$ of $\mathcal{K}$ and $n > 0$ such that no $\mathcal{I} \in \boldsymbol{M}$ is n-homomorphically embeddable into $\mathcal{J}$ preserving $\mathsf{ind}(\mathcal{K})$. Let $\mathsf{ind}(\mathcal{K}) = \{a_1, \dots, a_k\}$. For every finite Σ-interpretation $\mathcal{I}$ with domain $\{u_1, \dots, u_m\}$ such that $m \geq k$ and $a_i = u_i$ $(1 \leq i \leq k)$, we define the *canonical CQ* $\boldsymbol{q}_{\mathcal{I}}$ by taking

$$\boldsymbol{q}_{\mathcal{I}}(x_1, \dots, x_k) \;=\; \exists x_{k+1} \cdots \exists x_m \Big(\bigwedge_{u_i \in A^{\mathcal{I}}, A \in \Sigma} A(x_i) \wedge \bigwedge_{(u_i, u_j) \in R^{\mathcal{I}}, R \in \Sigma} R(x_i, x_j) \Big).$$

Then there exists a homomorphism from $\mathcal{I}$ to $\mathcal{J}$ preserving $\mathsf{ind}(\mathcal{K})$ iff $\mathcal{J} \models q_{\mathcal{I}}(a_1, \dots, a_k)$. Now pick for any $\mathcal{I} \in \boldsymbol{M}$ a subinterpretation $\mathcal{I}'$ of $\mathcal{I}$ with $\Delta^{\mathcal{I}'} \supseteq \mathsf{ind}(\mathcal{K})$ and $|\Delta^{\mathcal{I}'} \setminus \mathsf{ind}(\mathcal{K})| \leq n$ such that $\mathcal{I}'$ is not homomorphically embeddable into $\mathcal{J}$ preserving $\mathsf{ind}(\mathcal{K})$. Let $\boldsymbol{q}(x_1, \dots, x_k)$ be the disjunction of all canonical CQs $\boldsymbol{q}_{\mathcal{I}'}(x_1, \dots, x_k)$ determined by these $\mathcal{I}'$. Then $\mathcal{J} \not\models \boldsymbol{q}(a_1, \dots, a_k)$, and so $\mathcal{K} \not\models \boldsymbol{q}(a_1, \dots, a_k)$, but $\mathcal{I} \models \boldsymbol{q}(a_1, \dots, a_k)$, for all $\mathcal{I} \in \boldsymbol{M}$. □

Observe that, in the characterisation of Proposition 6, one cannot replace n-homomorphic embeddability by homomorphic embeddability as shown by the following example.

Example 7. Let $\mathcal{K} = (\{\top \sqsubseteq \exists R.\top\}, \{A(a)\})$. Then the class $\mathcal{M}$ of all interpretations that consist of a finite R-chain starting with $A(a)$ and followed by an R-cycle (of arbitrary length) is complete for $\mathcal{K}$. However, there is no homomorphism from any member of $\mathcal{M}$ into the model of $\mathcal{K}$ that consists of an infinite R-chain starting from $A(a)$.

We call an interpretation $\mathcal{I}$ a *ditree interpretation* if the directed graph $G_\mathcal{I}$ defined by taking

$$G_\mathcal{I} = (\Delta^\mathcal{I}, \{(d,e) \mid (d,e) \in \bigcup_{R \in \mathsf{N_R}} R^\mathcal{I}\})$$

is a directed tree and $R^\mathcal{I} \cap S^\mathcal{I} = \emptyset$, for any distinct role names R and S. $\mathcal{I}$ has *outdegree* n if $G_\mathcal{I}$ has outdegree n. A model $\mathcal{I}$ of $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ 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 $R(a,b) \in \mathcal{A}$. In this case, the *outdegree* of $\mathcal{I}$ is the maximum outdegree of the interpretations $\mathcal{I}_a$, for $a \in \mathsf{ind}(\mathcal{A})$. Denote by $\boldsymbol{M}_\mathcal{K}^{bo}$ the class of all forest-shaped $\mathsf{sig}(\mathcal{K})$-models of $\mathcal{K}$ of outdegree $\leq |\mathcal{T}|$. The following completeness result is well known [53] (the first part is shown in the proof of Proposition 9):

Proposition 8. *$\boldsymbol{M}_\mathcal{K}^{bo}$ is complete for any $\mathcal{ALC}$ KB $\mathcal{K}$. If $\mathcal{K}$ is a Horn$\mathcal{ALC}$ KB, then there is a single member $\mathcal{I}_\mathcal{K}$ of $\boldsymbol{M}_\mathcal{K}^{bo}$ that is complete for $\mathcal{K}$.*

The model $\mathcal{I}_\mathcal{K}$ mentioned in Proposition 8 is constructed using the standard chase procedure and called the *canonical model of* $\mathcal{K}$. Proposition 8 can be strengthened further. Call a subinterpretation $\mathcal{I}$ of a ditree interpretation $\mathcal{J}$ a *rooted subinterpretation of* $\mathcal{J}$ if there exists $u \in \Delta^\mathcal{J}$ such that the domain $\Delta^\mathcal{I}$ of $\mathcal{I}$ is the set of all $u' \in \Delta^\mathcal{J}$ for which there is a path $u_0, \dots, u_n \in \Delta^\mathcal{J}$ with $u_0 = u$, $u_n = u'$ and $(u_i, u_{i+1}) \in R_i^\mathcal{I}$ $(i < n)$, for some role name R_i. Call a ditree interpretation $\mathcal{I}$ *regular* if it has, up to isomorphism, only finitely many rooted subinterpretations. A forest-shaped model $\mathcal{I}$ of a KB $\mathcal{K}$ is *regular* if the ditree interpretations $\mathcal{I}_a$, $a \in \mathsf{ind}(\mathcal{K})$, are regular. Denote by $\boldsymbol{M}_\mathcal{K}^{reg}$ the class of all regular forest-shaped $\mathsf{sig}(\mathcal{K})$-models of $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ of outdegree bounded by $|\mathcal{T}|$.

Proposition 9. *$\boldsymbol{M}_\mathcal{K}^{reg}$ is complete for any $\mathcal{ALC}$ KB $\mathcal{K}$.*

Proof. Suppose $\mathcal{K}$ is an $\mathcal{ALC}$ KB and $\mathcal{K} \not\models \boldsymbol{q}(\boldsymbol{a})$, for some UCQ $\boldsymbol{q}(\boldsymbol{x})$. As shown in [53], there exists a consistent KB $\mathcal{K}' = (\mathcal{T}', \mathcal{A}')$ with $\mathcal{T}' \supseteq \mathcal{T}$, $\mathcal{A}' \supseteq \mathcal{A}$, and $\mathsf{ind}(\mathcal{A}') = \mathsf{ind}(\mathcal{A})$ such that $\mathcal{I} \not\models \boldsymbol{q}(\boldsymbol{a})$, for every model $\mathcal{I}$ of $\mathcal{K}'$ (called a *spoiler for* $\boldsymbol{q}$ *and* $\mathcal{K}$ in [53] and constructed by carefully analysing all possible homomorphism from $\boldsymbol{q}$ to models of $\mathcal{K}$ and 'spoiling' all of them by suitable KB extensions). We construct a regular model $\mathcal{J}'$ of $\mathcal{K}'$ as follows. Let $\mathcal{I}'$ be a model of $\mathcal{K}'$. We may assume that $\mathcal{T}'$ does not use the constructor $\forall r.C$. Denote by $\mathsf{cl}(\mathcal{T}')$ the set of subconcepts of concepts in $\mathcal{T}'$ closed under single negation. For $d \in \Delta^{\mathcal{I}'}$, the *$\mathcal{T}'$-type of d in $\mathcal{I}'$*, denoted $\boldsymbol{t}_{\mathcal{T}'}^{\mathcal{I}'}(d)$, is defined as $\boldsymbol{t}_{\mathcal{T}'}^{\mathcal{I}'}(d) = \{C \in \mathsf{cl}(\mathcal{T}') \mid d \in C^{\mathcal{I}'}\}$. A subset $\boldsymbol{t} \subseteq \mathsf{cl}(\mathcal{T}')$ is a *$\mathcal{T}'$-type* if $\boldsymbol{t} = \boldsymbol{t}_{\mathcal{T}'}^{\mathcal{I}}(d)$, for some model $\mathcal{I}$ of $\mathcal{T}'$ and $d \in \Delta^\mathcal{I}$. We denote the set of all $\mathcal{T}'$-types by $\mathsf{type}(\mathcal{T}')$. Let $\boldsymbol{t}, \boldsymbol{t}' \in \mathsf{type}(\mathcal{T}')$. 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 the concept $\bigsqcap \boldsymbol{t} \sqcap \exists R.(\bigsqcap \boldsymbol{t}')$ is satisfiable with respect to $\mathcal{T}'$. Denote by $\mathsf{succ}_{\exists R.C}(\boldsymbol{t})$ the set of all $\exists R.C$-witnesses for $\boldsymbol{t}$. Now choose, for any $\mathcal{T}'$-type $\boldsymbol{t}$ and $\exists R.C$ such that $\mathsf{succ}_{\exists R.C}(\boldsymbol{t}) \neq \emptyset$, a single type $s_{\exists R.C}(\boldsymbol{t}) \in \mathsf{succ}_{\exists R.C}(\boldsymbol{t})$. We construct the model $\mathcal{J}'$ of $\mathcal{K}'$ as follows. The domain $\Delta^{\mathcal{J}'}$ is the set of words

$$aR_1\boldsymbol{t}_1 \cdots R_n\boldsymbol{t}_n,$$

where $a \in \mathsf{ind}(\mathcal{K}')$ and, for $\boldsymbol{t}_0 = \boldsymbol{t}_{\mathcal{T}'}^{\mathcal{I}'}(a)$ and $i < n$, $\boldsymbol{t}_{i+1} = s_{\exists R_{i+1}.C}(\boldsymbol{t}_i)$ for some $\exists R_{i+1}.C \in \boldsymbol{t}_i$. Set $aR_1\boldsymbol{t}_1 \cdots R_n\boldsymbol{t}_n \in A^{\mathcal{J}'}$ if $n = 0$ and $A \in \boldsymbol{t}_{\mathcal{T}'}^{\mathcal{I}'}(a)$ or $n > 0$ and $A \in \boldsymbol{t}_n$. Finally, set $(aR_1\boldsymbol{t}_1 \cdots R_n\boldsymbol{t}_n, bS_1\boldsymbol{t}'_1 \cdots S_m\boldsymbol{t}'_m) \in R^{\mathcal{J}'}$ iff $n = m = 0$ and $R(a,b) \in \mathcal{A}$ or $0 < m = n+1$, $S_m = R$ and $aR_1\boldsymbol{t}_1 \cdots \boldsymbol{t}_n = bS_1\boldsymbol{t}'_1 \cdots \boldsymbol{t}'_{m-1}$. One can easily show that $\mathcal{J}'$ is a regular model of $\mathcal{K}'$. Hence $\mathcal{J}' \not\models \boldsymbol{q}(\boldsymbol{a})$. The outdegree of $\mathcal{J}'$ is bounded by $|\mathcal{T}'|$ but possibly not by $|\mathcal{T}|$, and so it remains to modify $\mathcal{J}'$ in such a way that its outdegree is bounded by $|\mathcal{T}|$. To this end, we remove from $\mathcal{J}'$ all R-successors (together with the subtrees they root) $aR_1\boldsymbol{t}_1 \cdots R_n\boldsymbol{t}_nR\boldsymbol{t}$ of all $aR_1\boldsymbol{t}_1 \cdots R_n\boldsymbol{t}_n \in \Delta^{\mathcal{J}'}$ such that $\boldsymbol{t} \neq s_{\exists R.C}(\boldsymbol{t}_n)$ for any $\exists R.C \in \mathsf{cl}(\mathcal{T})$. By the construction, the resulting interpretation $\mathcal{J}$ is still regular, it is a model of $\mathcal{K}$ (since $\mathcal{T}' \supseteq \mathcal{T}$), its outdegree is bounded by $|\mathcal{T}|$, and $\mathcal{J} \not\models \boldsymbol{q}(\boldsymbol{a})$ since $\mathcal{J}' \not\models \boldsymbol{q}(\boldsymbol{a})$. □

Example 10. Consider the KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ with $\mathcal{T} = \{A \sqcup B \sqsubseteq \exists R.(A \sqcup B)\}$ and $\mathcal{A} = \{A(a)\}$. The following class of regular models $\mathcal{I}$ is complete for $\mathcal{K}$. The domain of $\mathcal{I}$ is the natural numbers with $a^\mathcal{I} = 0 \in A^\mathcal{I}$, $(i,j) \in R^\mathcal{I}$ if $j = i+1$, for all natural numbers i and j, and there are $k, n, m \geq 0$ such that $A^\mathcal{I}$ and $B^\mathcal{I}$ are mutually disjoint, cover the initial segment $\{1, \dots, k\}$ and, on the remainder $\{k+1, \dots\}$, they are interpreted by alternating between n consecutive nodes in $A^\mathcal{I}$ and m consecutive nodes in $B^\mathcal{I}$. Then $\mathcal{I}$ is regular since the number of non-isomorphic rooted subinterpretations of $\mathcal{I}$ with root $r > k$ is $\leq n + m$ (the number of non-isomorphic rooted subinterpretations of $\mathcal{I}$ with root $r \leq k$ is clearly bounded by $k+1$).

In the undecidability proofs of Section 4, we do not use the full expressive power of $\mathcal{ALC}$ but work with a small fragment denoted $\mathcal{ELU}_{rhs}$. An *$\mathcal{ELU}_{rhs}$ TBox* $\mathcal{T}$ consists of CIs of the form

- $A \sqsubseteq C$,
- $A \sqsubseteq C \sqcup D$,

where A is a concept name and C, D are $\mathcal{EL}$-concepts. Given an $\mathcal{ELU}_{rhs}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, 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 $A \sqsubseteq C \sqcup D$ is an axiom of $\mathcal{T}$ and neither $C \in \ell(\sigma)$ nor $D \in \ell(\sigma)$, then we add to $\ell(\sigma)$ either C or D (but not both); in this case, we call σ an *or-node*. If $C \sqcap D \in \ell(\sigma)$, then we add both C and D to $\ell(\sigma)$ provided that they are not there yet. Finally, if $\exists R.C \in \ell(\sigma)$ and the constructed part of the tree does not contain a node of the form $\sigma \cdot w_{\exists R.C}$, then we add $\sigma \cdot w_{\exists R.C}$ as an R-successor of σ and set $\ell(\sigma \cdot w_{\exists R.C}) = \{C\}$. Now we define a *minimal model* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ of $\mathcal{K}$ by taking $\Delta^{\mathcal{I}}$ to be the set of nodes in $\mathfrak{O}$, $a^{\mathcal{I}} = a$ for $a \in \mathsf{ind}(\mathcal{A})$, $R^{\mathcal{I}}$ to be the R-relation in $\mathfrak{O}$ together with (a, b) such that $R(a, b) \in \mathcal{A}$, and $A^{\mathcal{I}} = \{ \sigma \in \Delta^{\mathcal{I}} \mid A \in \ell(\sigma) \}$, for every concept name A. It follows from the construction that $\mathcal{I}$ is a model of $\mathcal{K}$.

Lemma 11. *For any $\mathcal{ELU}_{rhs}$ KB $\mathcal{K}$, the set $\boldsymbol{M}_{\mathcal{K}}$ of its minimal models is complete for $\mathcal{K}$.*

Proof. By Proposition 6, it suffices to show that, for every model $\mathcal{J}$ of $\mathcal{K}$, there is a minimal model $\mathcal{I}$ that is homomorphically embeddable into $\mathcal{J}$ preserving $\mathsf{ind}(\mathcal{K})$. Suppose a model $\mathcal{J}$ of $\mathcal{K}$ is given. We can now inductively construct a set Δ, a labelling function ℓ defining a minimal model $\mathcal{I}$, and a homomorphism h from $\mathcal{I}$ to $\mathcal{J}$ such that $h(\sigma) \in C^{\mathcal{J}}$, for each $C \in \ell(\sigma)$ and $\sigma \in \Delta$. The model $\mathcal{J}$ is used as a guide. For instance, let $\sigma \in \Delta$ such that $h(\sigma)$ is set. Suppose that $A \in \ell(\sigma)$, $A \sqsubseteq C \sqcup D$ is an axiom in $\mathcal{T}$, and $C \notin \ell(\sigma)$, $D \notin \ell(\sigma)$. Since $\mathcal{J}$ is a model of $\mathcal{K}$, it must be the case that $h(\sigma)^{\mathcal{J}} \in C^{\mathcal{J}}$ or $h(\sigma)^{\mathcal{J}} \in D^{\mathcal{J}}$. In the former case, we add C to $\ell(\sigma)$, in the latter case, we add D to $\ell(\sigma)$. Suppose further that $\sigma \cdot w_{\exists R.C}$ is in Δ and $h(\sigma \cdot w_{\exists R.C})$ is not set. Since $\mathcal{J}$ is a model of $\mathcal{K}$ and by inductive assumption $h(\sigma) \in (\exists R.C)^{\mathcal{J}}$, there exists $d \in \Delta^{\mathcal{J}}$ such that $(h(\sigma), d) \in R^{\mathcal{J}}$ and $d \in C^{\mathcal{J}}$. So we set $h(\sigma \cdot w_{\exists R.C}) = d$.

Now we take the minimal model $\mathcal{I} = (\Delta, \cdot^{\mathcal{I}})$, where $\cdot^{\mathcal{I}}$ is defined according to the labelling function ℓ. By the construction of Δ and the fact that $\mathcal{I}$ is minimal, we obtain that h is indeed a homomorphism from $\mathcal{I}$ to $\mathcal{J}$. $\square$

3. Model-theoretic criteria for query entailment and inseparability between knowledge bases

In this section, we first define the central notions of query entailment and inseparability between KBs for CQs and UCQs as well as their restrictions to rooted queries. Then we give model-theoretic characterisations of these notions based on products of interpretations and (partial) homomorphisms.

Definition 12. 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 $\mathsf{ind}(\mathcal{K}_2)$. We say that $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ-$\mathcal{Q}$ *inseparable* if they Σ-$\mathcal{Q}$ entail each other. If Σ is the set of all concept and role names, we say '*full signature $\mathcal{Q}$-entails*' or '*full signature $\mathcal{Q}$-inseparable*'.

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

Example 13. 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 (and so also Σ_0-rCQ inseparable) but not Σ_0-rUCQ inseparable (and so also not Σ_0-UCQ inseparable). The former claim can be proved using the model-theoretic criterion given in Theorem 17 below, and the latter one follows from $\mathcal{K}_0' \models \boldsymbol{q}(a)$ 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 (and so also Σ_1-rCQ inseparable) but not Σ_1-CQ inseparable. The former claim can be proved using the model-theoretic criterion of Theorem 17 and the latter one follows from the observation that $\mathcal{K}_1' \models \exists x B(x)$ but $\mathcal{K}_1 \not\models \exists x B(x)$.

The situation changes for *Horn$\mathcal{ALC}$* KBs. The following can be easily proved by observing (using Proposition 8) that the certain answers to a UCQ over a *Horn$\mathcal{ALC}$* KB $\mathcal{K}$ coincide with the certain answers to its disjuncts over $\mathcal{K}$:

Proposition 14. *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.*

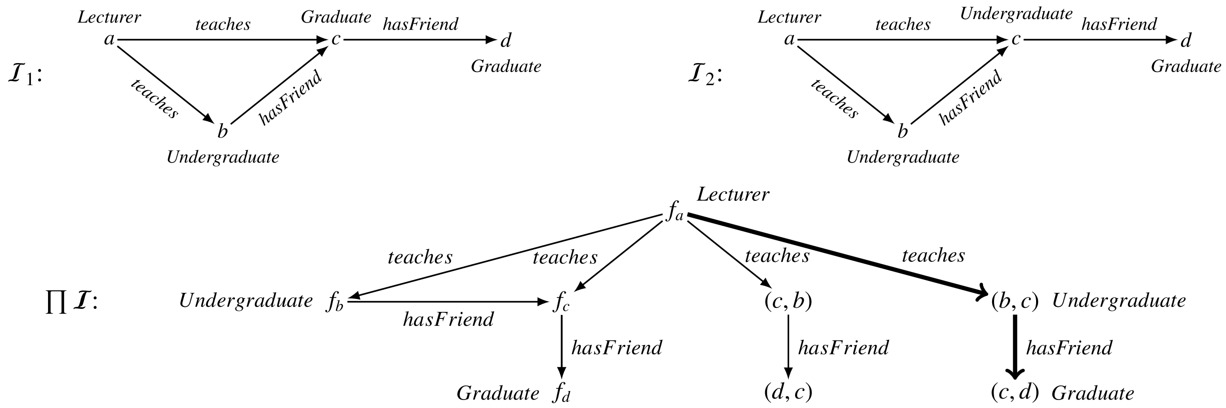
Now we give model-theoretic criteria of Σ-query entailment between KBs. As usual in model theory [54, page 405], we define the *product* $\prod \boldsymbol{\mathcal{I}}$ of a family $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_i \mid i \in I\}$ of interpretations by taking

$$\Delta^{\prod \boldsymbol{\mathcal{I}}} = \{f : I \to \bigcup_{i \in I} \Delta^{\mathcal{I}_i} \mid \forall i \in I \; f(i) \in \Delta^{\mathcal{I}_i}\},$$

$$A^{\prod \boldsymbol{\mathcal{I}}} = \{f \mid \forall i \in I\ f(i) \in A^{\mathcal{I}_i}\},$$
$$R^{\prod \boldsymbol{\mathcal{I}}} = \{(f, g) \mid \forall i \in I\ (f(i), g(i)) \in R^{\mathcal{I}_i}\},$$
$$a^{\prod \boldsymbol{\mathcal{I}}} = f_a, \quad \text{where } f_a(i) = a^{\mathcal{I}_i} \text{ for all } i \in I.$$

Proposition 15 *([54]). 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 all $\mathcal{I} \in \boldsymbol{\mathcal{I}}$.*

Example 16. The KB $\mathcal{K} = (\mathcal{T}_1, \mathcal{A}')$ from Example 2 has two minimal models: $\mathcal{I}_1$ that agrees with $\mathcal{A}'$ on a, b, d and has $c \in \mathit{Undergraduate}^{\mathcal{I}_2}$, and $\mathcal{I}_2$ that also agrees with $\mathcal{A}'$ on a, b, d but has $c \in \mathit{Graduate}^{\mathcal{I}_1}$ (cf. Example 4). By Lemma 11, the set $\boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \mathcal{I}_2\}$ is complete for $\mathcal{K}$. The picture below[1] shows the 'interesting' part of $\prod \boldsymbol{\mathcal{I}}$. Clearly, $\prod \boldsymbol{\mathcal{I}} \models \boldsymbol{q}'(a)$, where $\boldsymbol{q}'$ is the CQ from Example 2. It follows that $\mathcal{K} \models \boldsymbol{q}'(a)$.



We characterise Σ-query entailment in terms of products and $n\Sigma$-homomorphic embeddability. To also capture rooted queries, we first introduce the corresponding refinement of Σ-homomorphic and, respectively, $n\Sigma$-homomorphic embeddability. A *Σ-path ρ from u to v* in an interpretation $\mathcal{I}$ is a sequence $u_0, \dots, u_n \in \Delta^{\mathcal{I}}$ such that $u_0 = u$, $u_n = v$, and there are $R_0, \dots, R_{n-1} \in \Sigma$ with $(u_i, u_{i+1}) \in R_i^{\mathcal{I}}$, for $0 \le i < n$. For a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ and model $\mathcal{I}$ of $\mathcal{K}$, we say that $u \in \Delta^{\mathcal{I}}$ is *Σ-connected to $\mathcal{A}$ in $\mathcal{I}$* if there exist $a \in \mathsf{ind}(\mathcal{K})$ and a Σ-path from $a^{\mathcal{I}}$ to u in $\mathcal{I}$. The subinterpretation $\mathcal{I}^{con}$ of $\mathcal{I}$ induced by the set of all $u \in \Delta^{\mathcal{I}}$ that are Σ-connected to $\mathcal{A}$ in $\mathcal{I}$ is called the *Σ-component of $\mathcal{I}$ with respect to $\mathcal{K}$*. Let $\mathcal{I}_1$ be a model of $\mathcal{K}_1$ and $\mathcal{I}_2$ a model of $\mathcal{K}_2$. We say that $\mathcal{I}_2$ is *con-Σ-homomorphically embeddable into $\mathcal{I}_1$* if the Σ-component $\mathcal{I}_2^{con}$ of $\mathcal{I}_2$ with respect to $\mathcal{K}_2$ is Σ-homomorphically embeddable into $\mathcal{I}_1$; and we say that $\mathcal{I}_2$ is *con-$n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$* if the Σ-component $\mathcal{I}_2^{con}$ of $\mathcal{I}_2$ with respect to $\mathcal{K}_2$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$.

Theorem 17. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be $\mathcal{ALC}$ KBs, Σ a signature, and let $\boldsymbol{M}_i = \{\mathcal{I}_j \mid j \in I_i\}$ 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 $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$.
(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 $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$.
(3) *$\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2$ iff $\prod \boldsymbol{M}_2$ is $n\Sigma$-homomorphically embeddable into $\prod \boldsymbol{M}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$ *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 $\prod \boldsymbol{M}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$ *for any $n > 0$.*

Proof. (1) Suppose $\mathcal{K}_2 \models \boldsymbol{q}(\boldsymbol{a})$ but $\mathcal{K}_1 \not\models \boldsymbol{q}(\boldsymbol{a})$, for a Σ-UCQ $\boldsymbol{q}$ and $\boldsymbol{a}$ in $\mathsf{ind}(\mathcal{K}_1)$. Let n be the number of variables in $\boldsymbol{q}$. Take $\mathcal{I}_1 \in \boldsymbol{M}_1$ such that $\mathcal{I}_1 \not\models \boldsymbol{q}(\boldsymbol{a})$. Then no $\mathcal{I}_2 \in \boldsymbol{M}_2$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$ since this would imply $\mathcal{I}_2 \not\models \boldsymbol{q}(\boldsymbol{a})$. Conversely, suppose $\mathcal{I}_1 \in \boldsymbol{M}_1$ is such that, for some $n > 0$, no $\mathcal{I}_2 \in \boldsymbol{M}_2$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. Fix such an $n > 0$ and take for every $\mathcal{I}_2 \in \boldsymbol{M}_2$ a subinterpretation $\mathcal{I}_2'$ of $\mathcal{I}_2$ with domain of size $\le n$ such that $\mathcal{I}_2'$ is not Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(K_2)$. Recall from the proof of Proposition 6 that we can regard the Σ-reduct of any such $\mathcal{I}_2'$ as a Σ-CQ (with the answer variables corresponding to the ABox individuals). The disjunction of all these CQs (up to isomorphisms) is entailed by $\mathcal{K}_2$ but not by $\mathcal{K}_1$. The proof of (2) is similar.

(3) Suppose $\mathcal{K}_2 \models \boldsymbol{q}(\boldsymbol{a})$ but $\mathcal{K}_1 \not\models \boldsymbol{q}(\boldsymbol{a})$, for a Σ-CQ $\boldsymbol{q}$ and $\boldsymbol{a}$ in $\mathsf{ind}(\mathcal{K}_1)$. By Proposition 15, $\prod \boldsymbol{M}_2 \models \boldsymbol{q}(\boldsymbol{a})$ but $\prod \boldsymbol{M}_1 \not\models \boldsymbol{q}(\boldsymbol{a})$. Let n be the number of variables in $\boldsymbol{q}$. Then $\prod \boldsymbol{M}_2$ is not $n\Sigma$-homomorphically embeddable into $\prod \boldsymbol{M}_1$ preserving

[1] As usual in model theory, we write (b, c) for f with $f: 1 \mapsto b$ and $f: 2 \mapsto c$, and similarly for (c, b), (c, d) and (d, c).

$\mathsf{ind}(\mathcal{K}_2)$ since this would imply $\prod \boldsymbol{M}_1 \models \boldsymbol{q}(\boldsymbol{a})$. Conversely, suppose that, for some $n > 0$, $\prod \boldsymbol{M}_2$ is not $n\Sigma$-homomorphically embeddable into $\prod \boldsymbol{M}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. Let $\mathcal{I}$ be the subinterpretation of $\prod \boldsymbol{M}_2$ with domain of size $\leq n$ which cannot be Σ-homomorphically embedded in $\prod \boldsymbol{M}_1$ preserving $\mathsf{ind}(\mathcal{K}_2) \cap \{a \mid a^{\prod \boldsymbol{M}_2} \in \Delta^{\mathcal{I}}\}$. We can regard the Σ-reduct of $\mathcal{I}$ as a Σ-CQ which is entailed by $\mathcal{K}_2$ but not by $\mathcal{K}_1$ (by Proposition 15). The proof of (4) is similar. □

Example 7 can be used to show that, in Theorem 17, $n\Sigma$-homomorphic embeddability cannot be replaced by Σ-homomorphic embeddability. In Section 5, however, we show that in some cases we *can* find characterisations 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 17 provides a decision procedure for KB entailment. This applies, for example, to KBs with acyclic classical TBoxes [45], and to KBs for which the chase terminates [55].

4. Undecidability of (r)CQ-entailment and inseparability for $\mathcal{ALC}$ KBs

The aim of this section is to show that CQ and rCQ-entailment and inseparability for $\mathcal{ALC}$ KBs are undecidable. We begin by proving that it is undecidable whether an $\mathcal{EL}$ KB Σ-CQ entails an $\mathcal{ALC}$ KB. A straightforward modification of the KBs constructed in that proof is then used to prove that Σ-CQ inseparability between $\mathcal{EL}$ and $\mathcal{ALC}$ KBs is undecidable as well. It is to be noted that, as shown in Section 5, both Σ-UCQ and Σ-rUCQ entailments between $\mathcal{ALC}$ KBs are decidable, which means, by Proposition 14, that checking whether an $\mathcal{ALC}$ KB Σ-(r)CQ entails an $\mathcal{EL}$ KB is decidable. We then consider rooted CQs and prove that Σ-rCQ entailment and inseparability between $\mathcal{EL}$ and $\mathcal{ALC}$ KBs are still undecidable. (In fact, the undecidability proof for rCQs implies the undecidability results for CQs, but is somewhat trickier.) The signature Σ used in these undecidability proofs is a proper subset of the signatures of the KBs involved. In the final part of this section, we prove that one can modify the KBs in such a way that all the results stated above hold for full signature CQ and rCQ entailment and inseparability.

4.1. Undecidability of CQ-entailment and inseparability with respect to a signature Σ

Our undecidability proofs are by reduction of the undecidable *rectangle 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 $left(T) = right(T')$ if (i, j) is covered by a tile of type T and $(i+1, j)$ is covered by a tile of type T', and $1 \leq i < N$, $1 \leq j \leq M$; $up(T) = down(T')$ if (i, j) is covered by a tile of type T and $(i, j+1)$ is covered by a tile of type T', and $1 \leq i \leq N$, $1 \leq j < M$; $(1, 1)$ 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 $up(T) = C$. (The reader can easily show that this problem is undecidable by reduction of the halting problem for Turing machines; cf. [56].) If an instance $\mathfrak{T}$ of the rectangle tiling problem has a positive solution, we say that $\mathfrak{T}$ *admits tiling*.

Given such an instance $\mathfrak{T}$, we construct an $\mathcal{EL}$ TBox $\mathcal{T}^1_{\mathsf{CQ}}$, an $\mathcal{ALC}$ TBox $\mathcal{T}^2_{\mathsf{CQ}}$, an ABox $\mathcal{A}_{\mathsf{CQ}}$, and a signature Σ_{CQ} such that, for the KBs $\mathcal{K}^1_{\mathsf{CQ}} = (\mathcal{T}^1_{\mathsf{CQ}}, \mathcal{A}_{\mathsf{CQ}})$ and $\mathcal{K}^2_{\mathsf{CQ}} = (\mathcal{T}^2_{\mathsf{CQ}}, \mathcal{A}_{\mathsf{CQ}})$, the following conditions are equivalent:

- $\mathcal{K}^1_{\mathsf{CQ}}$ Σ_{CQ}-CQ entails $\mathcal{K}^2_{\mathsf{CQ}}$;
- the instance $\mathfrak{T}$ does not admit tiling.

The ABox $\mathcal{A}_{\mathsf{CQ}}$ does not depend on $\mathfrak{T}$ and is defined by setting $\mathcal{A}_{\mathsf{CQ}} = \{A(a)\}$. The TBox $\mathcal{T}^2_{\mathsf{CQ}}$ uses a role name R to encode a grid by putting one row of the grid after the other starting with the lower left corner of the grid. It also uses the following concept names:

- T^{first}, for each tile type $T \in \mathfrak{T}$, to encode the first row of a tiling;
- T_k, for $T \in \mathfrak{T}$ and $k = 0, 1, 2$, to encode intermediate rows, with three copies of each $T \in \mathfrak{T}$ needed to ensure the vertical matching conditions between rows;
- T^{halt}_k, for $T \in \mathfrak{T}$ and $k = 0, 1, 2$, to encode the last row;
- $\widehat{T}_k$, for $T \in \mathfrak{T}$ and $k = 0, 1, 2$.

Of all these concept names, only the $\widehat{T}_k$ are in the signature Σ_{CQ} of the entailment problem we construct. Thus, the T^{first}, T^{halt}_k, and T_k are auxiliary concept names used to generate tilings, while the $\widehat{T}_k$ make the tilings 'visible' to relevant CQs.

The TBox $\mathcal{T}^2_{\mathsf{CQ}}$ uses the concept names *Start* and *End* as markers for the start and end of a tiling. Both concept names are in Σ_{CQ}. To mark the end of rows, $\mathcal{T}^2_{\mathsf{CQ}}$ employs the concept names Row_k and Row^{halt}_k, for $k = 0, 1, 2$, where the Row^{halt}_k indicate the last row. Similarly to the encoding of tile types above, the concept names Row_k and Row^{halt}_k are auxiliary concept names used to construct tilings. Three copies are needed to ensure the vertical matching condition. In addition, we use a concept name $Row \in \Sigma_{\mathsf{CQ}}$ that marks the end of rows and is visible to separating CQs.

The role name R generating the grid is in Σ_{CQ}. An additional concept name A and role name P link the individual a in $\mathcal{A}_{CQ}$ to the first row of the tiling. The encoding does not depend on whether A, P are in Σ_{CQ}, but it will be useful later, when we consider full signature CQ-entailment, to include them in Σ_{CQ}.

Before writing up the axioms of $\mathcal{T}^2_{CQ}$, we explain how they generate all possible tilings. We ensure that if a point x in a model $\mathcal{I}$ of $\mathcal{K}^2_{CQ}$ is in $\widehat{T}_k$ and $right(T) = left(S)$, then x has an R-successor in $\widehat{S}_k$. Thus, branches of $\mathcal{I}$ define (possibly infinite) horizontal rows of tilings with $\mathfrak{T}$. If a branch contains a point $y \in \widehat{T}_k$ 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 $\widehat{T}_{(k+1) \bmod 3}$ 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 $\mathcal{K}^1_{CQ}$ and $\mathcal{K}^2_{CQ}$—we make every $x \in \widehat{T}_k$, starting from the second row, an instance of all $\widehat{S}_{(k-1) \bmod 3}$ with $down(T) = up(S)$. 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^{first} for the given initial tile type I.

The TBox $\mathcal{T}^2_{CQ}$ contains the following CIs, for $k = 0, 1, 2$:

$$A \sqsubseteq \exists P.(Start \sqcap \exists R.I^{first}), \tag{1}$$

$$T^{first} \sqsubseteq \exists R.S^{first}, \quad \text{if } right(T) = left(S) \text{ and } T, S \in \mathfrak{T}, \tag{2}$$

$$T^{first} \sqsubseteq \exists R.(Start \sqcap Row_1), \quad \text{if } right(T) = W \text{ and } T \in \mathfrak{T}, \tag{3}$$

$$T^{first} \sqsubseteq \widehat{T}_0, \quad \text{for } T \in \mathfrak{T}, \tag{4}$$

$$Row_k \sqsubseteq \exists R.T_k, \quad \text{for } T \in \mathfrak{T}, \tag{5}$$

$$T_k \sqsubseteq \exists R.S_k, \quad \text{if } right(T) = left(S) \text{ and } T, S \in \mathfrak{T}, \tag{6}$$

$$T_k \sqsubseteq \exists R.Row_{(k+1) \bmod 3}, \quad \text{if } right(T) = W \text{ and } T \in \mathfrak{T}, \tag{7}$$

$$T_k \sqsubseteq \exists R.Row^{halt}_{(k+1) \bmod 3}, \quad \text{if } right(T) = W \text{ and } T \in \mathfrak{T}, \tag{8}$$

$$Row_k \sqsubseteq Row, \tag{9}$$

$$T_k \sqsubseteq \widehat{T}_k, \quad \text{for } T \in \mathfrak{T}, \tag{10}$$

$$T_k \sqsubseteq \widehat{S}_{(k-1) \bmod 3}, \quad \text{if } down(T) = up(S) \text{ and } T, S \in \mathfrak{T}, \tag{11}$$

$$Row^{halt}_k \sqsubseteq \exists R.End \sqcup \bigsqcap_{up(T)=C,\, T\in\mathfrak{T}} \exists R.T^{halt}_k, \tag{12}$$

$$T^{halt}_k \sqsubseteq \exists R.S^{halt}_k, \quad \text{if } right(T) = left(S), up(S) = C \text{ and } T, S \in \mathfrak{T}, \tag{13}$$

$$T^{halt}_k \sqsubseteq \exists R.(Row \sqcap \exists R.End), \quad \text{if } right(T) = W \text{ and } T \in \mathfrak{T}, \tag{14}$$

$$Row^{halt}_k \sqsubseteq Row, \tag{15}$$

$$T^{halt}_k \sqsubseteq \widehat{S}_{(k-1) \bmod 3}, \quad \text{if } down(T) = up(S) \text{ and } T, S \in \mathfrak{T}. \tag{16}$$

The KB $\mathcal{T}^2_{CQ}$ is an $\mathcal{ELU}_{rhs}$ KB, with (12) being the only CIs with $\sqcup$. Throughout the proof, we work with the set $\boldsymbol{M}_{\mathcal{K}^2_{CQ}}$ of minimal models of $\mathcal{K}^2_{CQ}$ and use the notation introduced in the construction of minimal models. In figures, $\vee$ indicates an *or-node*. We now comment on the role of the CIs in $\mathcal{T}^2_{CQ}$.

- The CIs (1)–(3) produce all possible first rows whose ends are indicated by points in *Start* and Row_1; see Fig. 1(a), where τ_1 denotes trees described below. The CI (4) ensures that the tiling of the first row is visible in Σ_{CQ} using the concept names $\widehat{T}_0$. Note that *Row* is visible in Σ_{CQ} due to (9).
- The CIs (5)–(8) produce all possible intermediate rows starting with points in Row_k and ending by points in $Row_{(k+1) \bmod 3}$ or $Row^{halt}_{(k+1) \bmod 3}$; see Fig. 1(b), where τ_k is the tree with root in Row_k and τ^{halt}_k the tree with root in Row^{halt}_k as described below. The CIs (9)–(11) ensure that the tilings of the intermediate rows as well as *Row* are visible in Σ_{CQ}. Note that, for each intermediate row, there exists k such that the current row is encoded using $\widehat{T}_k$ and the matching previous row using $\widehat{T}_{(k-1) \bmod 3}$.
- The CIs (12)–(14) produce all possible final rows starting with points in Row^{halt}_k. The role of the disjunction is explained below; see Fig. 1(c). Finally, the axioms (15)–(16) make *Row* and the matching previous row visible in Σ_{CQ}. Note that the last row itself is not visible in Σ_{CQ}.

The existence of a tiling of some $N \times M$ grid for the given instance $\mathfrak{T}$ can be checked by Boolean CQs $\boldsymbol{q}_n$, for $n \geq 1$, that require an R-path from *Start* to *End* going through $\widehat{T}_k$- or *Row*-points:

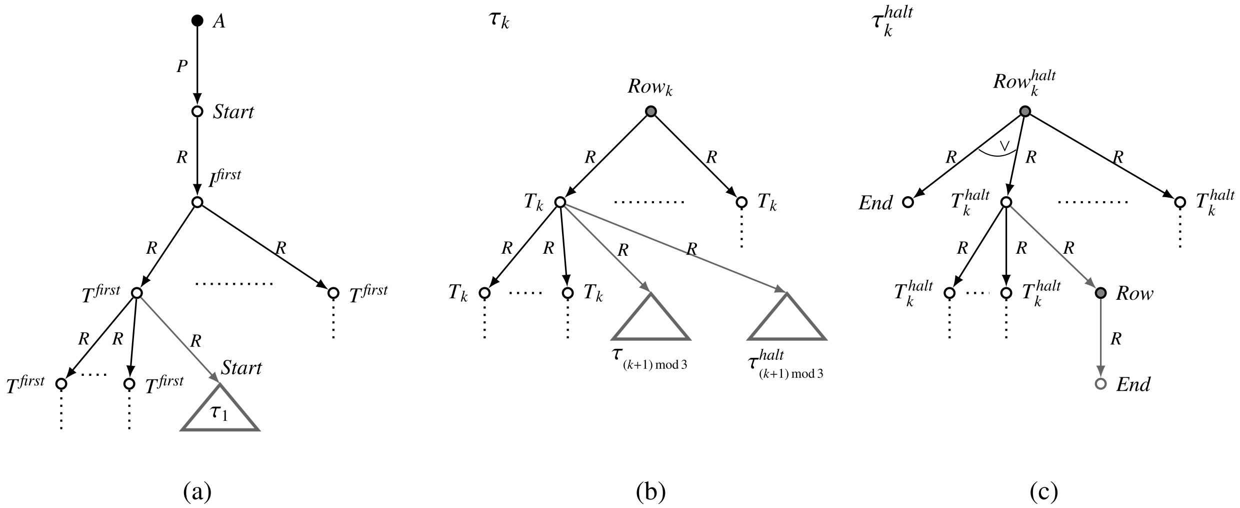


Fig. 1. The paths in the minimal models generated by the axioms of $\mathcal{T}^2_{\mathsf{CQ}}$.

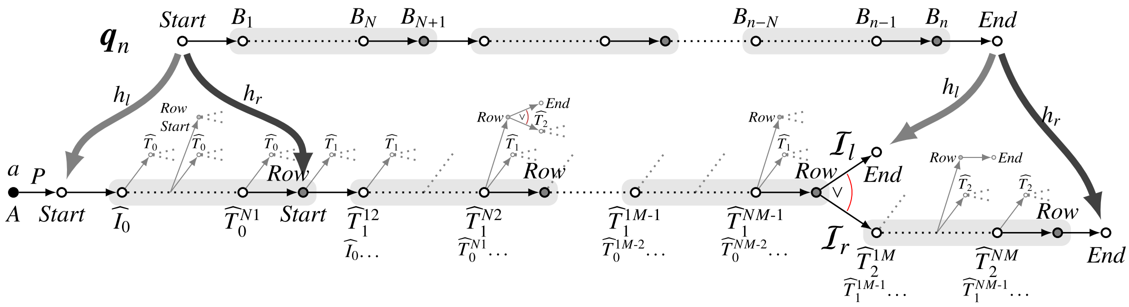


Fig. 2. The structure of the models $\mathcal{I}_l$ and $\mathcal{I}_r$ of $\mathcal{K}_2$, and homomorphisms $h_l\colon \boldsymbol{q}_n \to \mathcal{I}_l$ and $h_r\colon \boldsymbol{q}_n \to \mathcal{I}_r$.

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

where $B_i \in \{Row\} \cup \{\widehat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$. The $\boldsymbol{q}_n$ will serve as the separating Σ_{CQ}-CQs if $\mathfrak{T}$ admits a tiling (in fact, if $\mathfrak{T}$ admits a tiling of some $N \times M$ grid, then q_n is a separating Σ_{CQ}-CQ for $n = (N+1) \times (M-1)$). We illustrate the relationship between $\boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ and the CQs $\boldsymbol{q}_n$ in Fig. 2: the lower part of the figure shows two interpretations, $\mathcal{I}_l$ and $\mathcal{I}_r$, from $\boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ (we only mention the extensions of concept names in Σ_{CQ}). The two interpretations coincide up to the *Row*-point before the final row of the tiling. Then, because of the axiom (12), they realise *two alternative* continuations: one as described above, and the other one having just a single R-successor in *End*. In the picture, we show a situation where row m coincides with the row depicted below row $m+1$ (that satisfies the vertical tiling conditions with row $m+1$). For example, the first row $\widehat{I}_0 \cdots \widehat{T}_0^{N1}$ coincides with the row depicted below the second row (after the second *Start*). This is no accident and is enforced by the query $\boldsymbol{q}_n$ that is depicted in the upper part of the figure. If $\mathcal{K}^2_{\mathsf{CQ}} \models \boldsymbol{q}_n$, then $\boldsymbol{q}_n$ holds in both $\mathcal{I}_l$ and $\mathcal{I}_r$, 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 $B_{n-1} = \widehat{T}_1^{NM-1}$ in the figure, 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. Note that for this to work, we have to make both the P-successor of a and the first *Row*-point an instance of *Start*. We now formalise the observations above by proving the following:

Lemma 18. *The instance $\mathfrak{T}$ admits a rectangle tiling iff there exists $\boldsymbol{q}_n$ such that $\mathcal{K}^2_{\mathsf{CQ}} \models \boldsymbol{q}_n$.*

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

$$block_j = (\widehat{T}_k^{1,j}, \ldots, \widehat{T}_k^{N,j}, Row),$$

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

$block_1$, $block_2, \dots,$ $block_{M-1}$

(thus, $n = (N+1)\times(M-1)$). In view of Lemma 11, we only need to prove that $\mathcal{I} \models \boldsymbol{q}_n$, for each model $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$. Take such an $\mathcal{I}$. We have to show that there is an R-path $x_0, \dots, x_{n+1}$ in $\mathcal{I}$ such that $x_0 \in Start^{\mathcal{I}}$, $x_i \in B_i^{\mathcal{I}}$ for $1 \le i \le n$, and $x_{n+1} \in End^{\mathcal{I}}$.

First, we construct an auxiliary R-path $y_0, \dots, y_n$. We take $y_0 \in Start^{\mathcal{I}}$ and $y_1 \in I_0{}^{\mathcal{I}}$ by (1) ($I = T^{1,1}$). Then we take $y_2 \in (T_0^{2,1})^{\mathcal{I}}, \dots, y_N \in (T_0^{N,1})^{\mathcal{I}}$ by (2). We now have $right(T^{N,1}) = W$. By (3), we obtain $y_{N+1} \in Row_1^{\mathcal{I}} \cap Start^{\mathcal{I}}$. By (9), $y_{N+1} \in Row_1^{\mathcal{I}} \subseteq Row^{\mathcal{I}}$. We proceed in this way, starting with (5), till the moment we construct $y_{n-1} \in (T_k^{N,M-1})^{\mathcal{I}}$ with $right(T^{N,M-1}) = W$, for which we use (8) and (15) to obtain $y_n \in Row_k^{halt} \subseteq Row^{\mathcal{I}}$, for some k. Note that $T_k{}^{\mathcal{I}} \subseteq \widehat{T}_k{}^{\mathcal{I}}$ by (10), for a tile type T.

By (12), two cases are possible now:

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

Case 2: there is z_1 such that $(y_n, z_1) \in R^{\mathcal{I}}$ and $z_1 \in (T_k^{halt})^{\mathcal{I}}$, where $T = T^{1,M}$ and $up(T) = C$. We then use (13) and find a sequence $z_2, \dots, z_N, u, v$ such that $z_i \in (T_k^{halt})^{\mathcal{I}}$, where $T = T^{i,M}$, $u \in Row^{\mathcal{I}}$ and $v \in End^{\mathcal{I}}$. So 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 $(T_k^{i,j})^{\mathcal{I}} \subseteq (\widehat{T}^{i,j-1}_{(k-1) \bmod 3})^{\mathcal{I}}$.

($\Leftarrow$) Let $\boldsymbol{q}_n$ be such that $\mathcal{K}^2_{\mathsf{CQ}} \models \boldsymbol{q}_n$. Then $\mathcal{I} \models \boldsymbol{q}_n$, for each $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$. Consider all the pairwise distinct pairs $(\mathcal{I}, h)$ such that $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ and h is a homomorphism from $\boldsymbol{q}_n$ to $\mathcal{I}$. Note that $h(\boldsymbol{q}_n)$ contains an or-node σ_h (which is an instance of Row_k^{halt}, for some k). We call $(\mathcal{I}, h)$ and h *left* if $h(x_{n+1}) = \sigma_h \cdot w_{\exists R.End}$, and *right* otherwise. It is not hard to see that there exist a left $(\mathcal{I}_l, h_l)$ and a right $(\mathcal{I}_r, h_r)$ with $\sigma_{h_l} = \sigma_{h_r}$ (if this is not the case, we can construct $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ with $\mathcal{I} \not\models \boldsymbol{q}_n$ by choosing at every or-node σ the left (right) branch if there is no left (respectively, right) homomorphism h from $\boldsymbol{q}_n$ such that $h(x_n) = \sigma$).

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

Suppose $v_m = w_{\exists R.T_2^{halt}}$ (other ks are treated analogously). By (14), $right(T) = W$; by (13), $up(T) = C$. Suppose $w_{n-1} = w_{\exists R.S_k}$. Then $k = 1$. By (8), $right(S) = W$. By considering the atom $B_{n-1}(x_{n-1})$ in $\boldsymbol{q}_n$, we obtain that 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 $down(T) = up(S)$.

Suppose $v_{m-1} = w_{\exists R.U_2^{halt}}$. By (13), $right(U) = left(T)$ and $up(U) = C$. Suppose $w_{n-2} = w_{\exists R.Q_1}$. By (6), we have $right(Q) = left(S)$. By considering $B_{n-2}(x_{n-2})$ in $\boldsymbol{q}_n$, we obtain that 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 $down(U) = up(Q)$.

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 $B_{n-N-1} = Row$. Thus, we have tiled the two last rows of the grid. We proceed further and tile the whole $N \times M$ grid, where $M = n/(N+1) + 1$. $\square$

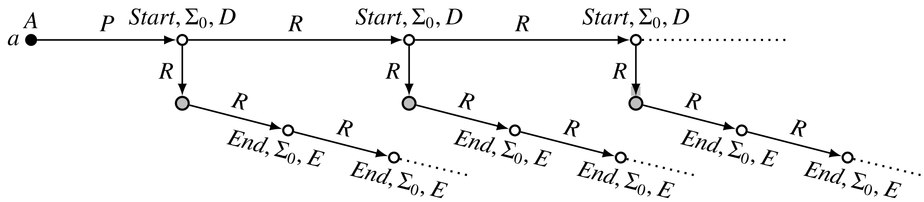
Next, we define the $\mathcal{EL}$-KB $\mathcal{K}^1_{\mathsf{CQ}} = (\mathcal{T}^1_{\mathsf{CQ}}, \mathcal{A}_{\mathsf{CQ}})$. Let $\Sigma_0 = \{Row\} \cup \{\widehat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$, and let $\mathcal{T}^1_{\mathsf{CQ}}$ contain the following CIs:

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

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

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

As $\mathcal{K}^1_{\mathsf{CQ}}$ is an $\mathcal{EL}$-KB, it has a canonical model $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}}$:



Note that the vertical R-successors of the $Start$-points are not instances of any concept name, and so $\mathcal{K}^1_{\mathsf{CQ}}$ does not satisfy any CQ $\boldsymbol{q}_n$. Now let $\Sigma_{\mathsf{CQ}} = \mathsf{sig}(\mathcal{K}^1_{\mathsf{CQ}})$. We show that $\mathcal{K}^2_{\mathsf{CQ}} \models \boldsymbol{q}$ implies $\mathcal{K}^1_{\mathsf{CQ}} \models \boldsymbol{q}$, for every Σ_{CQ}-CQ $\boldsymbol{q}$ without a subquery of the form $\boldsymbol{q}_n$.

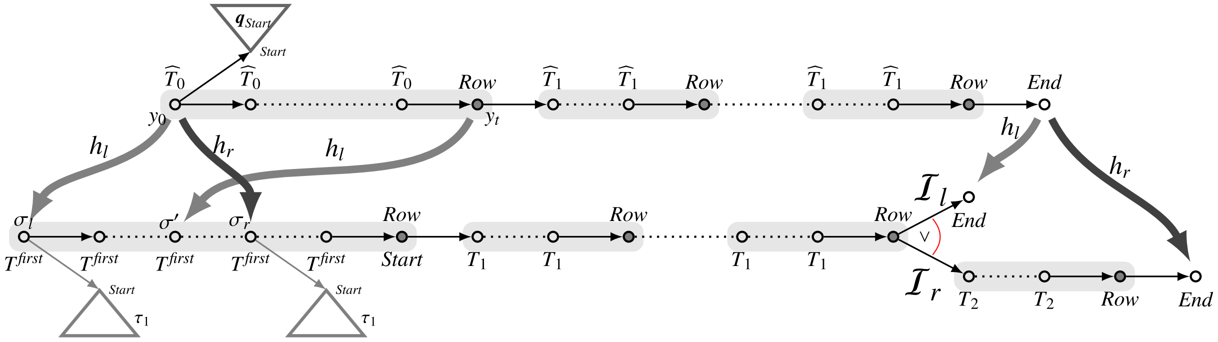


Fig. 3. A query that contains both *Start* and *End* atoms must have variables with empty concept labels.

Lemma 19. *$\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ is $n\Sigma_{\mathsf{CQ}}$-homomorphically embeddable into $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}}$ preserving $\{a\}$, for all $n \geq 1$, iff $\mathcal{K}^2_{\mathsf{CQ}} \not\models \boldsymbol{q}_m$, for all $m \geq 1$.*

Proof. ($\Rightarrow$) Suppose $\mathcal{K}^2_{\mathsf{CQ}} \models \boldsymbol{q}_m$ for some m. Then $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}} \models \boldsymbol{q}_m$. By assumption, $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ is $m\Sigma_{\mathsf{CQ}}$-homomorphically embeddable into $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}}$ preserving $\{a\}$, and so we have $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}} \models \boldsymbol{q}_m$, which is clearly impossible because none of the paths of $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}}$ contains the full sequence of symbols mentioned in $\boldsymbol{q}_m$.

($\Leftarrow$) Suppose $\mathcal{K}^2_{\mathsf{CQ}} \not\models \boldsymbol{q}_m$ for all m. Then $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}} \not\models \boldsymbol{q}_m$ for all m. Take any subinterpretation of $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ whose domain contains n elements. Recall from the proof of Proposition 6 that we can regard the Σ_{CQ}-reduct of this subinterpretation as a Boolean Σ_{CQ}-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_{CQ}-homomorphism from $\boldsymbol{q}_m$ into $\boldsymbol{q}$ for any m; in particular, $\boldsymbol{q}$ does not have a subquery of the form $\boldsymbol{q}_m$. We have to show that $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}} \models \boldsymbol{q}$.

If $\boldsymbol{q}$ contains A or P, then they appear at the root of $\boldsymbol{q}$ or, respectively, in the first edge of $\boldsymbol{q}$. By the structure of $\mathcal{K}_2$, the product $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$ does not contain a path from A to *End*, so $\boldsymbol{q}$ does not contain *End* and, therefore, can be mapped into $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}}$. In what follows, we assume that $\boldsymbol{q}$ does not contain A and P (note that D and E also do not occur in $\boldsymbol{q}$).

If $\boldsymbol{q}$ does not contain *Start* atoms or $\boldsymbol{q}$ does not contain *End* atoms, then clearly, $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}} \models \boldsymbol{q}$.

Suppose $\boldsymbol{q}$ contains both *Start* and *End* atoms. If there exists an R-path from a *Start* node to an *End* node in $\boldsymbol{q}$ then, by the structure of $\mathcal{K}^2_{\mathsf{CQ}}$, the *End* node is a leaf of $\boldsymbol{q}$ (as *End* nodes are always leaves in the models from $\boldsymbol{M}_{\mathcal{K}^2_{\mathsf{CQ}}}$) and the *Start* node is the root of $\boldsymbol{q}$ (as there are minimal models $\mathcal{I}_l$ and $\mathcal{I}_r$ in Fig. 2, in which the first *Start* node has no R-predecessor). Since $\boldsymbol{q}$ does not contain a subquery of the form $\boldsymbol{q}_m$, this R-path should contain variables with the empty Σ_{CQ}-concept label, in which case $\boldsymbol{q}$ can be mapped into $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}}$ by sending the root of $\boldsymbol{q}$ to the P-successor of a and the rest of the query so as to map a variable with the empty Σ_{CQ}-concept label to the vertical R-successor of a *Start* node.

Now, suppose that $\boldsymbol{q}$ does not contain a (directed) path from a *Start* node to an *End* node. Then the *Start* node is not the root of $\boldsymbol{q}$. We denote by $\boldsymbol{q}_{Start}$ the subtree of $\boldsymbol{q}$ generated by this node (see Fig. 3), and by $\boldsymbol{q}_{End}$ the path from the root y_0 of $\boldsymbol{q}$ to the *End* node. By the structure of $\mathcal{K}^2_{\mathsf{CQ}}$ shown in Fig. 1(a), the projection of y_0 onto every minimal model of $\mathcal{K}^2_{\mathsf{CQ}}$ is of the form $\delta \cdot w_{\exists R.T^{first}}$. We prove that $\boldsymbol{q}_{End}$ must have at least one intermediate node with the empty Σ_{CQ}-concept label. Indeed, suppose to the contrary that each intermediate variable x in $\boldsymbol{q}_{End}$ appears in an atom of the form $B(x)$, for $B \in \{\widehat{T}_k \mid k = 0, 1, 2\} \cup \{Row\}$. Since $\mathcal{K}^2_{\mathsf{CQ}} \models \boldsymbol{q}_{End}$, it follows that there is some k such that the distance between two neighbour *Row* nodes in $\boldsymbol{q}_{End}$ is k. Let $\mathcal{I}_l$ and $\mathcal{I}_r$ be the minimal models that satisfy (12) by picking the first and the second disjunct, respectively, and identical, otherwise (see Fig. 3). Suppose that $\mathcal{I}_l$ satisfies $\boldsymbol{q}_{End}$ by mapping y_0 to σ_l of the form $\delta \cdot w_{\exists R.T^{first}}$ and $\mathcal{I}_r$ 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 k. Let t be the distance from y_0 to the first *Row* node y_t. If $t \leq k$, then y_t should be mapped to σ' that is a predecessor of σ_r in $\mathcal{I}_l$ or σ_r itself. However, such a map is not possible as the Σ_{CQ}-label of σ' does not contain *Row* (only a concept of the form $\widehat{T}_0$), and we get a contradiction. In the case $t > k$, the argument is similar; one needs to observe that the structure of $\mathcal{K}^2_{\mathsf{CQ}}$ (in particular, (4), (7), (10)) makes it impossible to map $y_0, \ldots, y_t$ onto the common part of I_l and I_r in such a way that $h_r(y_i) = h_l(y_i)\sigma$ with $|\sigma| = k$. Thus, we conclude that $\boldsymbol{q}$ can be homomorphically mapped to $\mathcal{I}_{\mathcal{K}^1_{\mathsf{CQ}}}$ as follows: y_0 goes to $aw_{\exists P.D}$, $\boldsymbol{q}_{Start}$ to the infinite path of *Start* nodes, and $\boldsymbol{q}_{End}$ so as to map a variable with the empty Σ_{CQ}-concept label to the vertical successor of a *Start* node. $\square$

As an immediate consequence of Lemmas 18 and 19 and the characterisation of Σ-CQ-entailment given in Theorem 17 (3), we obtain:

Theorem 20. *The problem whether an $\mathcal{EL}$ KB Σ-CQ entails an $\mathcal{ALC}$ KB is undecidable.*

We now modify the KBs constructed in the proof of Theorem 20 to show undecidability of Σ-CQ-inseparability.

Theorem 21. *Σ-CQ inseparability between $\mathcal{EL}$ and $\mathcal{ALC}$ KBs is undecidable.*

Proof. We set $\mathcal{K}_2 = \mathcal{K}^2_{\mathrm{CQ}} \cup \mathcal{K}^1_{\mathrm{CQ}}$ and show that the following conditions are equivalent:

(1) $\mathcal{K}^1_{\mathrm{CQ}}$ Σ_{CQ}-CQ entails $\mathcal{K}^2_{\mathrm{CQ}}$;
(2) $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ are Σ_{CQ}-CQ inseparable.

Let $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ be the canonical model of $\mathcal{K}^1_{\mathrm{CQ}}$ and $\boldsymbol{M}_{\mathcal{K}^2_{\mathrm{CQ}}}$ the set of minimal models of $\mathcal{K}^2_{\mathrm{CQ}}$. One can easily show that the following set $\boldsymbol{M}_{\mathcal{K}_2}$ is complete for $\mathcal{K}_2$:

$$\boldsymbol{M}_{\mathcal{K}_2} = \{ \mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}} \mid \mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{CQ}}} \},$$

where $\mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ is the interpretation that results from merging the roots a of $\mathcal{I}$ and $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$. Now, the implication (2) $\Rightarrow$ (1) is trivial. For the converse direction, suppose $\mathcal{K}^1_{\mathrm{CQ}}$ Σ_{CQ}-CQ entails $\mathcal{K}^2_{\mathrm{CQ}}$. It follows that $\mathcal{K}_2$ Σ_{CQ}-CQ entails $\mathcal{K}^1_{\mathrm{CQ}}$. So it remains to show that $\mathcal{K}^1_{\mathrm{CQ}}$ Σ_{CQ}-CQ entails $\mathcal{K}_2$. Suppose this is not the case and there is a Σ_{CQ}-CQ $\boldsymbol{q}$ such that $\mathcal{K}_2 \models \boldsymbol{q}$ and $\mathcal{K}^1_{\mathrm{CQ}} \not\models \boldsymbol{q}$. We can assume $\boldsymbol{q}$ to be a *smallest connected* CQ with this property; in particular, no proper sub-CQ of $\boldsymbol{q}$ separates $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$. Now, we cannot have $\mathcal{K}^2_{\mathrm{CQ}} \models \boldsymbol{q}$ because this would contradict the fact that $\mathcal{K}^1_{\mathrm{CQ}}$ Σ_{CQ}-CQ entails $\mathcal{K}^2_{\mathrm{CQ}}$. Then $\mathcal{K}^2_{\mathrm{CQ}} \not\models \boldsymbol{q}$, and so there is $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{CQ}}}$ such that $\mathcal{I} \not\models \boldsymbol{q}$. On the other hand, we have $\mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}} \models \boldsymbol{q}$. Take a homomorphism $h\colon \boldsymbol{q} \to \mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$. As $\boldsymbol{q}$ is connected, $\mathcal{I} \not\models \boldsymbol{q}$ and $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}} \not\models \boldsymbol{q}$, 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 $h(\boldsymbol{q}') \subseteq \mathcal{I}$ (more precisely, $S(\boldsymbol{y}) \in \boldsymbol{q}$ is in $\boldsymbol{q}'$ iff $h(\boldsymbol{y}) \subseteq \Delta^{\mathcal{I}}$). Clearly, $\boldsymbol{q}' \subsetneq \boldsymbol{q}$ and $\mathcal{K}_2 \models \boldsymbol{q}'$. Denote by $\boldsymbol{q}''$ the complement of $\boldsymbol{q}'$ to $\boldsymbol{q}$. Obviously, $h(\boldsymbol{q}'') \subseteq \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$. Now, we either have $\mathcal{K}^1_{\mathrm{CQ}} \models \boldsymbol{q}'$ or $\mathcal{K}^1_{\mathrm{CQ}} \not\models \boldsymbol{q}'$. The latter case contradicts the choice of $\boldsymbol{q}$ because $\boldsymbol{q}'$ is a proper sub-CQ of $\boldsymbol{q}$. Thus, $\mathcal{K}^1_{\mathrm{CQ}} \models \boldsymbol{q}'$, and so there is a homomorphism $h'\colon \boldsymbol{q}' \to \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ with $h'(x) = a$, for every free variable x. Define a map $g\colon \boldsymbol{q} \to \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ 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, $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}} \models \boldsymbol{q}$, which is a contradiction. $\square$

Observe that our undecidability proof does not work for UCQs as the UCQ composed of the two disjunctive branches shown in Fig. 2 (for non-trivial instances) distinguishes between the KBs independently of the existence of a tiling. In Section 5, we show that, for UCQs, entailment is decidable.

4.2. Undecidability of rCQ-entailment and inseparability with respect to a signature Σ

It is not difficult to see that the KBs $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}^2_{\mathrm{CQ}}$ constructed in the undecidability proof for CQ-entailment cannot be used to prove undecidability of rCQ-entailment. In fact, $\mathcal{K}^1_{\mathrm{CQ}}$ Σ_{CQ}-rCQ entails $\mathcal{K}^2_{\mathrm{CQ}}$, for any instance of the rectangle tiling problem. We now sketch how the KBs defined above can be modified to show that rCQ-entailment and inseparability are indeed undecidable. Detailed proofs are given in the appendix.

Theorem 22. (i) *The problem whether an $\mathcal{EL}$ KB Σ-rCQ entails an $\mathcal{ALC}$ KB is undecidable.*
(ii) *Σ-rCQ inseparability between $\mathcal{EL}$ and $\mathcal{ALC}$ KBs is undecidable.*

Proof. For (i), we do not use the role name P but add $R(a, a)$ and $Row(a)$ to the ABox $\{A(a)\}$. The CQs $\boldsymbol{q}_n$ are modified by adding a conjunct $R(y, x_0)$ with answer variable y to $\boldsymbol{q}_n$. In more detail, suppose that an instance $\mathfrak{T}$ of the rectangle tiling problem is given. Let

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

let $\mathcal{T}^2_{\mathrm{rCQ}}$ contain the CIs (5)–(16) of $\mathcal{T}^2_{\mathrm{CQ}}$ as well as

$$A \sqsubseteq \exists R.(Row \sqcap \exists R.I_0), \tag{21}$$

and let $\mathcal{K}^2_{\mathrm{rCQ}} = (\mathcal{T}^2_{\mathrm{rCQ}}, \mathcal{A}_{\mathrm{rCQ}})$. Note that the loop $R(a, a)$ in $\mathcal{A}_{\mathrm{rCQ}}$ plays roughly the same role as the path between two *Start*-points in the previous construction (see Fig. 2). The existence of a tiling can now be checked by the rCQs

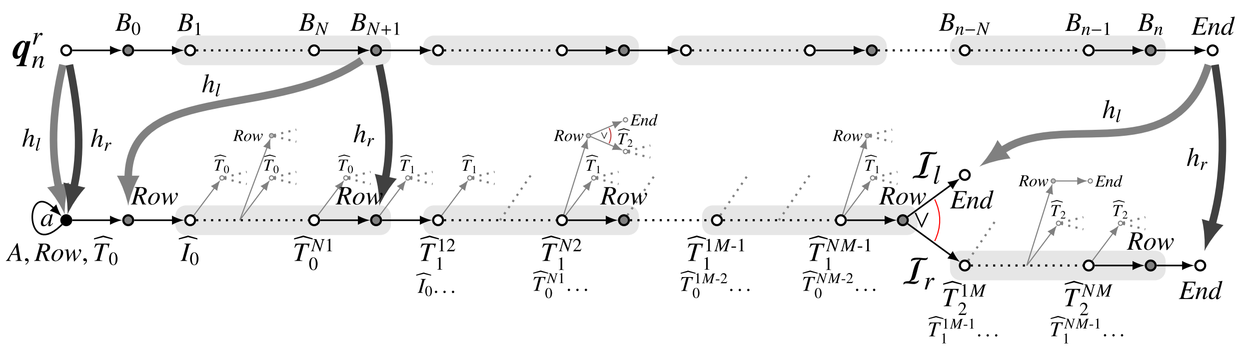


Fig. 4. The structure of models $\mathcal{I}_l$ and $\mathcal{I}_r$ of $\mathcal{K}_2$, and homomorphisms $h_l: \boldsymbol{q}_n^r \to \mathcal{I}_l$ and $h_r: \boldsymbol{q}_n^r \to \mathcal{I}_r$.

$$\boldsymbol{q}_n^r(y) \;=\; \exists \boldsymbol{x}\,\Big(R(y, x_0) \land \bigwedge_{i=0}^{n}\big(R(x_i, x_{i+1}) \land B_i(x_i)\big) \land End(x_{n+1})\Big),$$

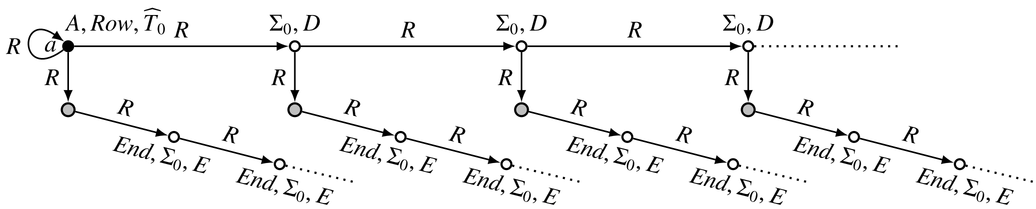
where $B_i \in \{Row\} \cup \{\widehat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$, for which we have an analogue of Lemma 18 for $\mathcal{K}^2_{\mathrm{rCQ}}$. The structure of the two homomorphisms is shown in Fig. 4. Note that the CQ encodes the first row two times. Now, we take $\mathcal{K}^1_{\mathrm{rCQ}} = (\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A}_{\mathrm{rCQ}})$, where $\mathcal{T}^1_{\mathrm{rCQ}}$ contains the following CIs (recall that we set $\Sigma_0 = \{Row\} \cup \{\widehat{T}_k \mid T \in \mathfrak{T}, k = 0, 1, 2\}$):

$$A \sqsubseteq \exists R.D \sqcap \exists R.\exists R.E, \tag{22}$$

$$D \sqsubseteq \exists R.D \sqcap \exists R.\exists R.E \sqcap \bigsqcap_{X \in \Sigma_0} X, \tag{23}$$

$$E \sqsubseteq \exists R.E \sqcap \bigsqcap_{X \in \Sigma_0} X \sqcap End. \tag{24}$$

The canonical model $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$ of $\mathcal{K}^1_{\mathrm{rCQ}}$ is shown below:



We set $\Sigma_{\mathrm{rCQ}} = \mathsf{sig}(\mathcal{K}^1_{\mathrm{rCQ}})$. Again, one can show Lemma 19 for $\mathcal{K}^1_{\mathrm{rCQ}}$ and $\mathcal{K}^2_{\mathrm{rCQ}}$. The proof of (*ii*) is similar to the non-rooted case and given in the appendix. □

4.3. Undecidability of (r)CQ-entailment and inseparability for full signature

The KBs used in the undecidability proofs above trivially do not Σ-CQ-entail each other for the *full signature* Σ. For example, the answer to the CQ $\exists y \exists z\,(P(a, y) \land R(y, z) \land I^{\mathit{first}}(z))$ is 'yes' over $\mathcal{K}^2_{\mathrm{CQ}}$ and 'no' over $\mathcal{K}^1_{\mathrm{CQ}}$. To establish undecidability results for separating CQs with arbitrary symbols, we modify the KBs constructed above. We follow [57] and replace the non-Σ-symbols by complex $\mathcal{ALC}$-concepts that, in contrast to concept names, cannot occur in CQs. Let Γ be a set of concept names. For any $B \in \Gamma$, let Z_B be a fresh concept name and let R_B and S_B be fresh role names. The *abstraction of* B is the $\mathcal{ALC}$-concept

$$H_B \;=\; \forall R_B.\exists S_B.\neg Z_B.$$

The Γ-*abstraction* $C^{\uparrow\Gamma}$ of a (possibly compound) concept C is obtained from C by replacing every $B \in \Gamma$ with H_B. The Γ-*abstraction* $\mathcal{T}^{\uparrow\Gamma}$ of a TBox $\mathcal{T}$ is obtained from $\mathcal{T}$ by replacing all concepts in $\mathcal{T}$ with their Γ-abstractions. We associate with Γ an auxiliary TBox

$$\mathcal{T}^{\exists}_{\Gamma} \;=\; \{\, \top \sqsubseteq \exists R_B.\top,\; \top \sqsubseteq \exists S_B.Z_B \mid B \in \Gamma \,\}$$

and call $\mathcal{T}^{\uparrow\Gamma} \cup \mathcal{T}^{\exists}_{\Gamma}$ the *enriched* Γ-*abstraction* of $\mathcal{T}$ for Γ. In what follows, we are going to replace TBoxes $\mathcal{T}$ with their enriched Γ-abstractions. We say that a TBox $\mathcal{T}$ *admits trivial models* if any interpretation $\mathcal{I}$ with $X^{\mathcal{I}} = \emptyset$, for any concept or role name X, is a model of $\mathcal{T}$. The TBoxes used in the undecidability proofs above admit trivial models.

Theorem 23. *Suppose $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$ are $\mathcal{ALC}$ KBs and Σ a signature such that $\mathsf{sig}(\mathcal{A}) \subseteq \Sigma$, $\Gamma = \mathsf{sig}(\mathcal{T}_1 \cup \mathcal{T}_2) \setminus \Sigma$ contains no role names, and $\mathcal{T}_1$ and $\mathcal{T}_2$ admit trivial models. Let $\mathcal{K}_i^{\uparrow\Gamma} = (\mathcal{T}_i^{\uparrow\Gamma} \cup \mathcal{T}_\Gamma^{\exists}, \mathcal{A})$, for $i = 1, 2$. Then the following conditions are equivalent:*

(1) $\mathcal{K}_1$ *Σ-(r)CQ entails* $\mathcal{K}_2$;
(2) $\mathcal{K}_1^{\uparrow\Gamma}$ *full signature (r)CQ entails* $\mathcal{K}_2^{\uparrow\Gamma}$.

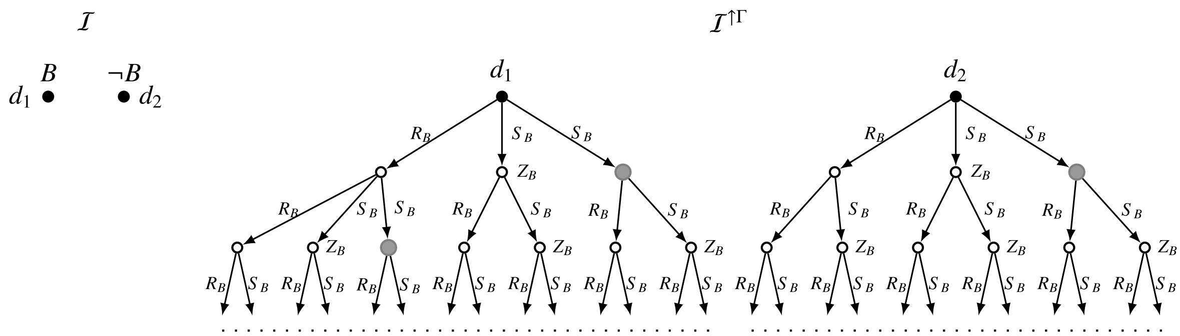
Proof. We start by defining the Γ-*abstraction* $\mathcal{I}^{\uparrow\Gamma}$ and the Γ-*instantiation* $\mathcal{I}^{\downarrow\Gamma}$ of an interpretation $\mathcal{I}$. The latter is defined in the same way as $\mathcal{I}$ except that $B^{\mathcal{I}^{\downarrow\Gamma}} = {H_B}^{\mathcal{I}}$, for all $B \in \Gamma$. It is straightforward to show the following.

Fact 1. For all $\mathcal{ALC}$ concepts D over the signature $\mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2)$ and all $d \in \Delta^{\mathcal{I}}$, we have $d \in D^{\mathcal{I}^{\downarrow\Gamma}}$ iff $d \in (D^{\uparrow\Gamma})^{\mathcal{I}}$. In particular, if $\mathcal{I}$ is a model of $\mathcal{K}_i^{\uparrow\Gamma}$, then $\mathcal{I}^{\downarrow\Gamma}$ is a model of $\mathcal{K}_i$, for $i = 1, 2$.

We now define the interpretation $\mathcal{I}^{\uparrow\Gamma}$. The domain $\Delta^{\mathcal{I}^{\uparrow\Gamma}}$ of $\mathcal{I}^{\uparrow\Gamma}$ is the set of words $w = dv_1 \cdots v_n$ such that $d \in \Delta^{\mathcal{I}}$ and $v_i \in \{R_B, S_B, \bar{S}_B \mid B \in \Gamma\}$, where $v_i \neq \bar{S}_B$ if either (*i*) $i > 2$ or (*ii*) $i = 2$ and $d \notin B^{\mathcal{I}}$ or $v_1 \neq R_B$. Then

$$\begin{aligned}
A^{\mathcal{I}^{\uparrow\Gamma}} &= A^{\mathcal{I}}, \text{ for all concept names } A \in \mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2) \setminus \Gamma;\\
B^{\mathcal{I}^{\uparrow\Gamma}} &= \emptyset, \text{ for all concept names } B \in \Gamma;\\
{Z_B}^{\mathcal{I}^{\uparrow\Gamma}} &= {Z_B}^{\mathcal{I}} \cup \{w \mid \mathsf{tail}(w) = S_B\}, \text{ for all concept names } B \in \Gamma;\\
S^{\mathcal{I}^{\uparrow\Gamma}} &= S^{\mathcal{I}}, \text{ for all role names } S \notin \{R_B, S_B \mid B \in \Gamma\};\\
{R_B}^{\mathcal{I}^{\uparrow\Gamma}} &= {R_B}^{\mathcal{I}} \cup \{(w, wR_B) \mid wR_B \in \Delta^{\mathcal{I}^{\uparrow\Gamma}}\}, \text{ for all concept names } B \in \Gamma;\\
{S_B}^{\mathcal{I}^{\uparrow\Gamma}} &= {S_B}^{\mathcal{I}} \cup \{(w, wS_B) \mid wS_B \in \Delta^{\mathcal{I}^{\uparrow\Gamma}}\} \cup \{(w, w\bar{S}_B) \mid w\bar{S}_B \in \Delta^{\mathcal{I}^{\uparrow\Gamma}}\}, \text{ for all concept names } B \in \Gamma.
\end{aligned}$$

By the construction of $\mathcal{I}^{\uparrow\Gamma}$, we have ${H_B}^{\mathcal{I}^{\uparrow\Gamma}} = B^{\mathcal{I}}$, for all concept names $B \in \Gamma$. For the interpretation $\mathcal{I}$ below consisting of two elements d_1 and d_2 with $d_1 \in B^{\mathcal{I}}$ and $d_2 \in (\neg B)^{\mathcal{I}}$ and $\Gamma = \{B\}$, the Γ-abstraction $\mathcal{I}^{\uparrow\Gamma}$ can be depicted as follows, where the grey points ● correspond to the words of the form $w\bar{S}_B$:



Fact 2. For all $\mathcal{ALC}$ concepts D over the signature $\mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2)$ and all $d \in \Delta^{\mathcal{I}}$, we have $d \in (D^{\uparrow\Gamma})^{\mathcal{I}^{\uparrow\Gamma}}$ iff $d \in D^{\mathcal{I}}$. Moreover, if $\mathcal{I}$ is a model of $\mathcal{K}_i$, then $\mathcal{I}^{\uparrow\Gamma}$ is a model of $\mathcal{K}_i^{\uparrow\Gamma}$, for $i = 1, 2$.

Proof of Fact 2. For the 'moreover'-part, observe that, for $C \sqsubseteq D \in \mathcal{T}_i$ and $d \in \Delta^{\mathcal{I}}$, we have that $d \in (C^{\uparrow\Gamma})^{\mathcal{I}^{\uparrow\Gamma}}$ implies $d \in (D^{\uparrow\Gamma})^{\mathcal{I}^{\uparrow\Gamma}}$ by the first part of Fact 2. For $d \in \Delta^{\mathcal{I}^{\uparrow\Gamma}} \setminus \Delta^{\mathcal{I}}$, observe that $d \notin H_B^{\mathcal{I}^{\uparrow\Gamma}}$ for any $B \in \Gamma$, $d \notin A^{\mathcal{I}^{\uparrow\Gamma}}$ and any concept name $A \in \mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2)$, and $(d, d') \notin R^{\mathcal{I}^{\uparrow\Gamma}}$ for any d' and role name $R \in \mathsf{sig}(\mathcal{K}_1 \cup \mathcal{K}_2)$. Thus, if $C \sqsubseteq D \in \mathcal{T}_i$ and $d \in C^{\mathcal{I}^{\uparrow\Gamma}}$ then it follows from the condition that $\mathcal{T}_i$ admits trivial models that $d \in D^{\mathcal{I}^{\uparrow\Gamma}}$. Thus $\mathcal{I}^{\uparrow\Gamma}$ is a model of $\mathcal{T}_i^{\uparrow\Gamma}$. Since $\mathcal{I}^{\uparrow\Gamma}$ is a model of $\mathcal{T}_\Gamma^{\exists}$ by construction, it follows that $\mathcal{I}^{\uparrow\Gamma}$ is a model of $\mathcal{T}_i^{\uparrow\Gamma} \cup \mathcal{T}_\Gamma^{\exists}$.

We collect further basic properties of the interpretations $\mathcal{I}^{\uparrow\Gamma}$ and $\mathcal{I}^{\downarrow\Gamma}$. In the formulation and proofs of Facts 3–6 below, the homomorphisms are always constructed in such a way that individual names are preserved. For simplicity, we do not state this explicitly.

Fact 3. Let $\mathcal{I}, \mathcal{J}$ be interpretations and $n > 0$. If $\mathcal{I}$ is n-homomorphically embeddable into $\mathcal{J}$, then $\mathcal{I}^{\uparrow\Gamma}$ is n-homomorphically embeddable into $\mathcal{J}^{\uparrow\Gamma}$.

Proof of Fact 3. Suppose $n > 0$ and $\mathcal{I}$ is n-homomorphically embeddable into $\mathcal{J}$. Let $\mathcal{I}'$ be a subinterpretation of $\mathcal{I}^{\uparrow\Gamma}$ with $|\Delta^{\mathcal{I}'}| \leq n$. For the subinterpretation $\mathcal{I}''$ of $\mathcal{I}$ induced by $\Delta_0 = \Delta^{\mathcal{I}} \cap \Delta^{\mathcal{I}'}$, there exists a homomorphism h_0 from $\mathcal{I}''$ to $\mathcal{J}$. We extend h_0 to a homomorphism h from $\mathcal{I}'$ to $\mathcal{J}^{\uparrow\Gamma}$ inductively as follows. Suppose $d \in \Delta^{\mathcal{I}'} \setminus \Delta^{\mathcal{I}}$ and $h(d)$ has not yet been defined, but there is no R_B or S_B-predecessor of d in $\mathcal{I}^{\uparrow\Gamma}$ for which $h(d)$ has not been defined. We distinguish three cases (which are mutually exclusive by the construction of $\mathcal{I}^{\uparrow\Gamma}$). If (*i*) $h(d')$ has been defined for an R_B-predecessor d' of d in $\mathcal{I}'$, then choose an R_B-successor e of $h(d')$ in $\mathcal{J}^{\uparrow\Gamma}$ and set $h(d) = e$. Observe that such an R_B-successor exists by the construction of $\mathcal{J}^{\uparrow\Gamma}$. If (*ii*) $h(d')$ has been defined for an S_B-predecessor d' of d in $\mathcal{I}'$, then choose an S_B-successor e of $h(d')$ in $\mathcal{J}^{\uparrow\Gamma}$ such that $e \in {Z_B}^{\mathcal{J}^{\uparrow\Gamma}}$ and set $h(d) = e$. Again such an S_B-successor exists by the construction of $\mathcal{J}^{\uparrow\Gamma}$. (*iii*) There does not exist any R_B or S_B-predecessor of d in $\mathcal{I}'$ for which h has been defined. In this case, choose $h(d)$ arbitrarily in $\mathcal{J}^{\uparrow\Gamma}$ such that if $d \in {Z_B}^{\mathcal{I}^{\uparrow\Gamma}}$, then $h(d) \in {Z_B}^{\mathcal{J}^{\uparrow\Gamma}}$. Such a d exists since ${Z_B}^{\mathcal{J}^{\uparrow\Gamma}} \neq \emptyset$. The resulting map is a homomorphism from $\mathcal{I}'$ to $\mathcal{J}^{\uparrow\Gamma}$.

Fact 4. Let $\mathcal{I}$ be a model of $\mathcal{K}^{\uparrow\Gamma}$, for $\mathcal{K} \in \{\mathcal{K}_1, \mathcal{K}_2\}$. Then $(\mathcal{I}^{\downarrow\Gamma})^{\uparrow\Gamma}$ is homomorphically embeddable into $\mathcal{I}$.

Proof of Fact 4. Let h_0 be the identity mapping from $\mathcal{I}^{\downarrow\Gamma}$ to $\mathcal{I}$ (observe that $\Delta^{\mathcal{I}^{\downarrow\Gamma}} = \Delta^{\mathcal{I}}$). One can now extend h_0 to a homomorphism h from $(\mathcal{I}^{\downarrow\Gamma})^{\uparrow\Gamma}$ to $\mathcal{I}$ in the same way as in the construction of h in the proof of Fact 3 above.

Fact 5. Let $\mathcal{K} \in \{\mathcal{K}_1, \mathcal{K}_2\}$. If $\boldsymbol{M}$ is complete for $\mathcal{K}$, then $\{\mathcal{I}^{\uparrow\Gamma} \mid \mathcal{I} \in \boldsymbol{M}\}$ is complete for $\mathcal{K}^{\uparrow\Gamma}$.

Proof of Fact 5. Suppose $\mathcal{J}$ is a model of $\mathcal{K}^{\uparrow\Gamma}$. By Proposition 6, it suffices to show that, for any $n > 0$, there is $\mathcal{I} \in \boldsymbol{M}$ such that $\mathcal{I}^{\uparrow\Gamma}$ is n-homomorphically embeddable into $\mathcal{J}$. Fix $n > 0$ and consider the interpretation $\mathcal{J}^{\downarrow\Gamma}$. By Fact 1, $\mathcal{J}^{\downarrow\Gamma}$ is a model of $\mathcal{K}$ and so there exists a model $\mathcal{I}$ of $\mathcal{K}$ such that $\mathcal{I}$ is n-homomorphically embeddable into $\mathcal{J}^{\downarrow\Gamma}$. But then, by Fact 3, $\mathcal{I}^{\uparrow\Gamma}$ is n-homomorphically embeddable into $(\mathcal{J}^{\downarrow\Gamma})^{\uparrow\Gamma}$ which, by Fact 4, itself is homomorphically embeddable into $\mathcal{J}$. Thus, $\mathcal{I}^{\uparrow\Gamma}$ is n-homomorphically embeddable into $\mathcal{J}$. By Fact 2, $\mathcal{I}^{\uparrow\Gamma}$ is a model of $\mathcal{K}^{\uparrow\Gamma}$.

Fact 6. Let $\boldsymbol{M}_i$ be families of interpretations with $X^{\mathcal{I}} = \emptyset$, for all $\mathcal{I} \in \boldsymbol{M}_i$ and all concept and role names X with $X \notin \mathsf{sig}(\mathcal{K}_i)$, $i = 1, 2$. Then the following conditions are equivalent:

(a) $\prod \boldsymbol{M}_2$ is $n\Sigma$-homomorphically embeddable into $\prod \boldsymbol{M}_1$, for all $n > 0$;
(b) $\prod\{\mathcal{I}^{\uparrow\Gamma} \mid \mathcal{I} \in \boldsymbol{M}_2\}$ is n-homomorphically embeddable into $\prod\{\mathcal{I}^{\uparrow\Gamma} \mid \mathcal{I} \in \boldsymbol{M}_1\}$, for all $n > 0$.

Proof of Fact 6. Suppose $\boldsymbol{M}_1 = \{\mathcal{I}_i \mid i \in I\}$ and $\boldsymbol{M}_2 = \{\mathcal{J}_j \mid j \in J\}$.

Assume first that (a) holds and let $\mathcal{J}$ is a subinterpretation of $\prod\{\mathcal{J}_j^{\uparrow\Gamma} \mid j \in J\}$ with $|\Delta^{\mathcal{J}}| \leq n$. We have to construct a homomorphism from $\mathcal{J}$ to $\prod\{\mathcal{I}_i^{\uparrow\Gamma} \mid i \in I\}$. There is a Σ-homomorphism h_0 from the subinterpretation $\mathcal{J}'$ of $\prod \boldsymbol{M}_2$ induced by $\Delta^{\mathcal{J}} \cap \Delta^{\prod \boldsymbol{M}_2}$ to $\prod \boldsymbol{M}_1$. By definition, h_0 is a homomorphism from the subinterpretation $\mathcal{J}''$ of $\prod\{\mathcal{J}_j^{\uparrow\Gamma} \mid j \in J\}$ induced by $\Delta^{\mathcal{J}} \cap \Delta^{\prod \boldsymbol{M}_2}$ to $\prod\{\mathcal{I}_i^{\uparrow\Gamma} \mid i \in I\}$ (the only difference between $\mathcal{J}'$ and $\mathcal{J}''$ is that $B^{\mathcal{J}''} = \emptyset$ for all $B \in \Gamma$). Following the proof of Fact 3, one can now expand h_0 to a homomorphism h from $\mathcal{J}$ to $\prod\{\mathcal{I}_i^{\uparrow\Gamma} \mid i \in I\}$.

Conversely, assume that (b) holds and assume that $\mathcal{J}$ is a subinterpretation of $\prod \boldsymbol{M}_2$ with $|\Delta^{\mathcal{J}}| \leq n$. We have to construct a Σ-homomorphism from $\mathcal{J}$ to $\prod \boldsymbol{M}_1$. There is a Σ-homomorphism h_0 from the subinterpretation $\mathcal{J}'$ of $\prod\{\mathcal{J}_j^{\uparrow\Gamma} \mid j \in J\}$ induced by $\Delta^{\mathcal{J}}$ to $\prod\{\mathcal{I}_i^{\uparrow\Gamma} \mid i \in I\}$. To obtain from h_0 the required Σ-homomorphism h, we have to re-define $h_0(d)$ for any d with $h_0(d) \in \Delta^{\prod\{\mathcal{I}_i^{\uparrow\Gamma} \mid i \in I\}} \setminus \Delta^{\prod \boldsymbol{M}_1}$. Consider such a d. Observe that $h_0(d) \notin B^{\prod\{\mathcal{I}_i^{\uparrow\Gamma} \mid i \in I\}}$ for any concept name $B \in \Sigma$ and $h_0(d)$ is not in the range or domain of any $R^{\prod\{\mathcal{I}_i^{\uparrow\Gamma} \mid i \in I\}}$ for any role name $R \in \Sigma$. But then, since h_0 is a Σ-homomorphism, $d \notin B^{\mathcal{J}}$ for any concept name $B \in \Sigma$ and d is not in the range or domain of $R^{\mathcal{J}}$ for any role name $R \in \Sigma$. Thus, we can choose $h(d)$ arbitrarily in $\Delta^{\prod \boldsymbol{M}_1}$ and obtain the required Σ-homomorphism.

For CQs, Theorem 23 now follows directly from Theorem 17 (3) and Facts 5 and 6. Note that we can consider sets $\boldsymbol{M}_i$ of interpretations that are complete for $\mathcal{K}_i$ such that $X^{\mathcal{I}} = \emptyset$, for all $\mathcal{I} \in \boldsymbol{M}_i$ and all concept and role names X with $X \notin \mathsf{sig}(K_i)$, $i = 1, 2$. For rCQs, we use Theorem 17 (4). □

Now, to prove undecidability of full signature (r)CQ entailment and inseparability, we apply Theorem 23 to the KBs constructed in the proofs of Theorems 20, 21 and 22. Note that the KBs $(\mathcal{K}^1_{\mathrm{CQ}})^{\uparrow\Gamma}$ with $\Gamma = \mathsf{sig}(\mathcal{K}^1_{\mathrm{CQ}} \cup \mathcal{K}^2_{\mathrm{CQ}}) \setminus \Sigma_{\mathrm{CQ}}$ and $(\mathcal{K}^1_{\mathrm{rCQ}})^{\uparrow\Gamma}$ with $\Gamma = \mathsf{sig}(\mathcal{K}^1_{\mathrm{rCQ}} \cup \mathcal{K}^2_{\mathrm{rCQ}}) \setminus \Sigma_{\mathrm{rCQ}}$ are still $\mathcal{EL}$-KBs since $\Sigma_{\mathrm{CQ}} = \mathsf{sig}(\mathcal{K}^1_{\mathrm{CQ}})$ and $\Sigma_{\mathrm{rCQ}} = \mathsf{sig}(\mathcal{K}^1_{\mathrm{rCQ}})$.

Theorem 24. (*i*) *The problem whether an $\mathcal{EL}$ KB full signature-(r)CQ entails an $\mathcal{ALC}$ KB is undecidable.*
(*ii*) *Full signature-(r)CQ inseparability between $\mathcal{EL}$ and $\mathcal{ALC}$ KBs is undecidable.*

5. Decidability of (r)UCQ-entailment and inseparability for $\mathcal{ALC}$ KBs

We show that, in sharp contrast to the case of (r)CQs, Σ-(r)UCQ-entailment and inseparability of $\mathcal{ALC}$ KBs are decidable and 2ExpTime-complete. We start by proving a new model-theoretic criterion for Σ-(r)UCQ entailment that replaces finite partial Σ-homomorphisms by Σ-homomorphisms and uses the class of regular forest-shaped models for the entailing KB $\mathcal{K}_1$ and the class of forest-shaped models for the entailed KB $\mathcal{K}_2$. We then encode this characterisation into an emptiness problem for two-way alternating parity automata on infinite trees (2APTAs) by constructing a 2APTA that accepts (representations of) forest-shaped models of the entailing KB into which there is no Σ-homomorphism from any forest-shaped model of the entailed KB. Rabin's result that such an automaton accepts a regular model iff it accepts any model will then yield the desired 2ExpTime upper bound for (r)UCQ-entailment. Matching lower bounds are proved by a reduction of the word problem for exponentially space bounded alternating Turing machines. Finally, we show that the same tight complexity bounds still hold in the full signature case.

5.1. Model-theoretic characterisation of (r)UCQ-entailment based on regular models

We show that finite partial homomorphisms can be replaced by homomorphisms in the characterisation of Σ-(r)UCQ entailment between $\mathcal{ALC}$-KBs given in Theorem 17 if one considers regular forest-shaped models of the entailing KB $\mathcal{K}_1$ and forest-shaped models of the entailed KB $\mathcal{K}_2$. Recall that, by Proposition 9, the class $\boldsymbol{M}^{reg}_{\mathcal{K}}$ of regular forest-shaped models of outdegree $\leq |\mathcal{T}|$ is complete for any $\mathcal{ALC}$-KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. We also show that if Σ contains all role names in the entailed KB, then Σ-rUCQ entailment coincides with Σ-UCQ entailment. This allows us to transfer our 2ExpTime lower bound from the non-rooted to the rooted case.

Theorem 25. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be $\mathcal{ALC}$-KBs and Σ a signature.*

(1) *$\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ iff, for any $\mathcal{I}_1 \in \boldsymbol{M}^{reg}_{\mathcal{K}_1}$, there exists $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$.
(2) *$\mathcal{K}_1$ Σ-rUCQ entails $\mathcal{K}_2$ iff, for any $\mathcal{I}_1 \in \boldsymbol{M}^{reg}_{\mathcal{K}_1}$, there exists $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is con-Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$.

Proof. We only prove (1) as the proof of (2) is similar. The direction ($\Leftarrow$) follows from Theorem 17 and the facts that $\boldsymbol{M}^{reg}_{\mathcal{K}_1}$ and $\boldsymbol{M}^{bo}_{\mathcal{K}_2}$ are complete for $\mathcal{K}_1$ and $\mathcal{K}_2$, respectively (Propositions 8 and 9). To show ($\Rightarrow$), suppose that $\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ and let $\mathcal{I}_1 \in \boldsymbol{M}^{reg}_{\mathcal{K}_1}$. We construct $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ and a Σ-homomorphism h from $\mathcal{I}_2$ to $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. By Theorem 17 (1) and Propositions 8 and 9, we have

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

Denote by $\mathcal{J}_{|\leq n}$ the subinterpretation of an interpretation $\mathcal{J} \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ induced by the domain elements of $\mathcal{J}$ connected to ABox individuals in $\mathsf{ind}(\mathcal{K}_2)$ by paths of role names (possibly not in Σ) of length $\leq n$. A (Σ, n)*-homomorphism h from $\mathcal{J}$ to $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$ is a Σ-homomorphism preserving $\mathsf{ind}(\mathcal{K}_2)$ whose domain is a finite subinterpretation of $\mathcal{J}$ that contains $\mathcal{J}_{|\leq n}$. Let Ξ_n be the class of pairs $(\mathcal{J}, h)$ with $\mathcal{J} \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ and h a (Σ, n)-homomorphism from $\mathcal{J}$ to $\mathcal{I}_1$. By ($*$), all Ξ_n are non-empty. We may assume that for $(\mathcal{I}, h), (\mathcal{J}, f) \in \bigcup_{n\geq 0} \Xi_n$, if $\mathcal{I}_{|\leq n}$ and $\mathcal{J}_{|\leq n}$ are isomorphic then $\mathcal{I}_{|\leq n} = \mathcal{J}_{|\leq n}$, for all $n \geq 0$. We define classes $\Theta_n \subseteq \bigcup_{m\geq n} \Xi_m$, $n \geq 0$, with $\Theta_0 \supseteq \Theta_1 \supseteq \cdots$ such that the following conditions hold:

(a) $\Theta_n \cap \Xi_m \neq \emptyset$ for all $m \geq n$;
(b) $\mathcal{I}_{|\leq n} = \mathcal{J}_{|\leq n}$ and $h_{|\leq n} = f_{|\leq n}$ for all $(\mathcal{I}, h), (\mathcal{J}, f) \in \Theta_n$ (here and below, $h_{|\leq n}$ denotes the restriction of h to $\mathcal{I}_{|\leq n}$).

Let Θ_0 be the set of all pairs $(\mathcal{J}, h)$ such that $(\mathcal{J}, h) \in \Xi_0$. Our assumptions imply that Θ_0 has the properties (a) and (b) because $h(a^{\mathcal{J}}) = a^{\mathcal{I}}$ holds for every Σ-homomorphism h preserving $\mathsf{ind}(\mathcal{K}_2)$ and all ABox individuals $a \in \mathsf{ind}(\mathcal{K}_2)$. Suppose now that Θ_n is defined and satisfies (a) and (b). Define an equivalence relation $\sim$ on $\Theta_n \cap (\bigcup_{m\geq n+1} \Xi_m)$ by setting $(\mathcal{I}, h) \sim (\mathcal{J}, f)$ if $\mathcal{I}_{|\leq n+1} = \mathcal{J}_{|\leq n+1}$ and, for all $x \in \Delta^{\mathcal{J}_{|\leq n+1}} \setminus \Delta^{\mathcal{J}_{|\leq n}}$, the following holds: $h(x)$ and $f(x)$ are always roots of isomorphic ditree subinterpretations of $\mathcal{I}_1$ and if, in addition, either $h(x) \in \mathsf{ind}(\mathcal{K}_1)$ or $f(x) \in \mathsf{ind}(\mathcal{K}_1)$, or there is a $y \in \Delta^{\mathcal{J}_{|\leq n}}$ such that x is an R-successor of y in $\mathcal{J}_{|\leq n+1}$, for some role name $R \in \Sigma$, then $h(x) = f(x)$. By the finite outdegree and regularity of $\mathcal{I}_1$, the properties (a) and (b) of Θ_n, and the finite outdegree of all $\mathcal{J}$ such that $(\mathcal{J}, h) \in \Xi_n$, the number of equivalence classes for $\sim$ is finite. Hence there exists an equivalence class Θ satisfying (a). Clearly, we can modify the (Σ, n)-homomorphisms h, f in the pairs $(\mathcal{I}, h), (\mathcal{J}, f) \in \Theta$ in such a way that $h(x) = f(x)$ holds for all $x \in \Delta^{\mathcal{J}_{|\leq n+1}} \setminus \Delta^{\mathcal{J}_{|\leq n}}$ while preserving the remaining properties of Θ. The resulting set of pairs satisfies (a) and (b).

We define an interpretation $\mathcal{I}_2$ as the union of all $\mathcal{J}_{|\leq n}$ such that there exists $(\mathcal{J}, h) \in \Theta_n$, $n \geq 0$:

$$\Delta^{\mathcal{I}_2} = \bigcup_{n\geq 0} \{ \Delta^{\mathcal{J}_{|\leq n}} \mid \exists h\, (\mathcal{J}, h) \in \Theta_n \};$$

$$A^{\mathcal{I}_2} = \bigcup_{n\geq 0} \{ A^{\mathcal{J}_{|\leq n}} \mid \exists h\, (\mathcal{J}, h) \in \Theta_n \}, \text{ for all concept names } A;$$

$$R^{\mathcal{I}_2} = \bigcup_{n\geq 0} \{ R^{\mathcal{J}_{|\leq n}} \mid \exists h\, (\mathcal{J}, h) \in \Theta_n \}, \text{ for all role names } R.$$

Using Conditions (a) and (b) and the fact that the sequence $\Theta_0, \Theta_1, \cdots$ is decreasing, it is straightforward to show that $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$. Define a function h from $\mathcal{I}_2$ to $\mathcal{I}_1$ by setting

$$h = \bigcup_{n\geq 0} \{ h_{|\leq n} \mid \exists \mathcal{J}\, (\mathcal{J}, h) \in \Theta_n \}.$$

It follows from Condition (b) that h is well defined. It is readily checked that h is a Σ-homomorphism from $\mathcal{I}_2$ to $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. □

Lemma 26. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be $\mathcal{ALC}$-KBs and Σ a signature containing all role names in $\mathsf{sig}(\mathcal{K}_2)$. Then $\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ iff $\mathcal{K}_1$ Σ-rUCQ entails $\mathcal{K}_2$.*

Proof. Suppose $\mathcal{K}_1$ Σ-rUCQ entails $\mathcal{K}_1$. By Theorem 25, it suffices to prove that, for any $\mathcal{I}_1 \in \boldsymbol{M}^{reg}_{\mathcal{K}_1}$, there exists $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. By Theorem 25, we know that, for any $\mathcal{I}_1 \in \boldsymbol{M}^{reg}_{\mathcal{K}_1}$, there exists $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is con-Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. Moreover, as Σ contains the role names in $\mathsf{sig}(\mathcal{K}_2)$, we may assume that every $u \in \Delta^{\mathcal{I}_2}$ is Σ-connected to the ABox $\mathcal{A}_2$ of $\mathcal{K}_2$. But then $\mathcal{I}_2$ is con-Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$ iff it is Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$, as required. □

5.2. 2ExpTime *upper bound for (r)UCQ-entailment with respect to signature* Σ

We use the model-theoretic criterion of Theorem 25 to prove a 2ExpTime upper bound for (r)UCQ-entailment between $\mathcal{ALC}$-KBs with respect to a signature Σ. We focus on the non-rooted case and then discuss the modifications required for the rooted one. Let $\mathcal{K}_1$, $\mathcal{K}_2$ be $\mathcal{ALC}$-KBs and Σ a signature. We aim to check if there is a model $\mathcal{I}_1 \in \boldsymbol{M}^{reg}_{\mathcal{K}_1}$ into which no model $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_1}$ is Σ-homomorphically embeddable. In the following, we construct an automaton $\mathfrak{A}$ that accepts (a suitable representation of) the desired models $\mathcal{I}_1$. It then remains to check whether the language $\mathcal{L}(\mathfrak{A})$ accepted by $\mathfrak{A}$ is non-empty. Note that $\mathcal{L}(\mathfrak{A})$ contains also non-regular models, but a well-known result by Rabin [58] implies that, if $\mathcal{L}(\mathfrak{A})$ is non-empty, then it contains a regular model, which is sufficient for our purposes.

We use two-way alternating parity automata on infinite trees (2APTAs) and encode forest-shaped interpretations as labelled trees to make them inputs to 2APTAs. Let $\mathbb{N}$ denote the *positive* integers. A *tree* is a non-empty (possibly infinite) set $T \subseteq \mathbb{N}^*$ closed under prefixes. The node ε is the *root* of T. As a convention, for $x \in \mathbb{N}^*$, 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. Without loss of generality, we assume that all nodes in an m-ary tree are from $\{1, \dots, m\}^*$.

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

Definition 27. A *two-way alternating parity automaton* (*2APTA*) *on infinite m-ary trees* is a tuple $\mathfrak{A} = (Q, \Gamma, \delta, q_0, c)$, where Q is a finite set of *states*, Γ a finite alphabet, $\delta\colon Q \times \Gamma \to \mathcal{B}^+(\mathsf{tran}(\mathfrak{A}))$ the *transition function* with the set of *transitions* $\mathsf{tran}(\mathfrak{A}) = [m] \times Q$, $q_0 \in Q$ the *initial state*, and $c : Q \to \mathbb{N}$ is the *acceptance condition*.

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 28. A *run* of a 2APTA $\mathfrak{A} = (Q, \Gamma, \delta, q_0, c)$ on an infinite Γ-labelled tree (T, L) is a $T \times Q$-labelled tree (T_r, r) such that the following conditions are satisfied:

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

We say that (T_r, r) is *accepting* if in all infinite paths $y_1 y_2 \cdots$ of T_r, $\min(\{c(q) \mid r(y_i) = (x, q)$ for infinitely many $i\})$ is even. An infinite Γ-labelled tree (T, L) is *accepted* by $\mathfrak{A}$ if there is an accepting run of $\mathfrak{A}$ on (T, L). We use $\mathcal{L}(\mathfrak{A})$ to denote the set of all infinite Γ-labelled trees accepted by $\mathfrak{A}$.

We require the following results from automata theory:

Theorem 29 *([58,59]).*

1. *Given a 2APTA $\mathfrak{A}$, one can construct in polynomial time a 2APTA $\mathfrak{B}$ with $L(\mathfrak{B}) = \overline{L(\mathfrak{A})}$.*
2. *Given a constant number of 2APTAs $\mathfrak{A}_1, \dots, \mathfrak{A}_c$, one can construct in polynomial time a 2APTA $\mathfrak{A}$ such that $L(\mathfrak{A}) = L(\mathfrak{A}_1) \cap \dots \cap L(\mathfrak{A}_c)$.*

3. *Emptiness of 2APTAs can be decided in time single exponential in the number of states.*
4. *For any 2APTA $\mathfrak{A}$, $\mathcal{L}(\mathfrak{A}) \neq \emptyset$ implies that $\mathcal{L}(\mathfrak{A})$ contains a regular tree.*

Now, let Γ be the 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 $\mathsf{CN}(\mathcal{T}_i)$ (respectively, $\mathsf{RN}(\mathcal{T}_i)$) denotes the set of concept (respectively, role) names in $\mathcal{T}_i$. We represent forest-shaped models of $\mathcal{T}_1$ as m-ary Γ-labelled trees, with $m = \max(|\mathcal{T}_1|, |\mathsf{ind}(\mathcal{K}_1)|)$. The root node labelled with *root* is not used in the representation. Each ABox individual is represented by a successor of the root labelled 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 labelled 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.

We call a Γ-labelled tree *proper* if it satisfies the following conditions:

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

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

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

Note that $\mathcal{I}_{(T,L)}$ is a forest-shaped interpretation of outdegree at most $|\mathcal{T}_1|$ that satisfies all required conditions to qualify as a forest-shaped model of $\mathcal{T}_1$ except that it need not satisfy $\mathcal{T}_1$. In addition, the interpretation $\mathcal{I}_{(T,L)}$ is regular iff the tree (T, L) is regular (has, up to isomorphisms, only finitely many rooted subtrees). Conversely, every model $\mathcal{I} \in \boldsymbol{M}^{bo}_{\mathcal{K}_1}$ can be represented as a proper m-ary Γ-labelled tree. Note that the assertions from $\mathcal{A}_1$ are not explicitly represented in (T, L), but readded in the construction of $\mathcal{I}_{(T,L)}$.

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

- a 2APTA $\mathfrak{A}_0$ that accepts an m-ary Γ-labelled tree iff it is proper;
- a 2APTA $\mathfrak{A}_1$ that accepts a proper m-ary Γ-labelled tree (T, L) iff $\mathcal{I}_{(T,L)}$ is a model of $\mathcal{T}_1$;
- a 2APTA $\mathfrak{A}_2$ that accepts a proper m-ary Γ-labelled tree (T, L) iff there is a model $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is Σ-homomorphically embeddable into $\mathcal{I}_{(T,L)}$ preserving $\mathsf{ind}(\mathcal{K}_2)$.

The following result shows that we would achieve our goal once we have constructed $\mathfrak{A}_0$, $\mathfrak{A}_1$, and $\mathfrak{A}_2$ and then define $\mathfrak{A}$ in such a way that $\mathcal{L}(\mathfrak{A}) = \mathcal{L}(\mathfrak{A}_0) \cap \mathcal{L}(\mathfrak{A}_1) \cap \overline{\mathcal{L}(\mathfrak{A}_2)}$.

Lemma 30. *The following conditions are equivalent:*

(1) $\mathcal{L}(\mathfrak{A}_0) \cap \mathcal{L}(\mathfrak{A}_1) \cap \overline{\mathcal{L}(\mathfrak{A}_2)} = \emptyset$,
(2) *for each model $\mathcal{I}_1 \in \boldsymbol{M}^{bo}_{\mathcal{K}_1}$, there exists a model $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$,
(3) *for each regular model $\mathcal{I}_1 \in \boldsymbol{M}^{bo}_{\mathcal{K}_1}$, there exists a model $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$,
(4) *$\mathcal{K}_1$ Σ-UCQ-entails $\mathcal{K}_2$.*

Proof. (1) ⇔ (2) follows from the properties of $\mathfrak{A}_0$, $\mathfrak{A}_1$, $\mathfrak{A}_2$; (1) ⇔ (3) follows from the properties of $\mathfrak{A}_0$, $\mathfrak{A}_1$, $\mathfrak{A}_2$, and Rabin's Theorem [58]; and (3) ⇔ (4) is Theorem 25. □

The construction of $\mathfrak{A}_0$ is trivial and left to the reader. The construction of $\mathfrak{A}_1$ is quite standard [51]. 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 2APTA $\mathfrak{A}_1 = (Q, \Gamma, \delta, q_0, c)$ 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:

$$\begin{aligned}
\delta(q_0, root) &= \bigwedge_{i=1}^{m} (i, q_1),\\
\delta(q_1, \ell) &= ((0, q_\emptyset) \vee (0, q^{C_{\mathcal{T}_1}})) \wedge \bigwedge_{i=1}^{m} (i, q_1),\\
\delta(q^{\exists R.C}, (a, U)) &= \bigvee_{i=1}^{m} ((i, q^R) \wedge (i, q^C)) \vee \bigvee_{R(a,b)\in\mathcal{A}_1} (-1, q^{b,C}),\\
\delta(q^{\forall R.C}, (a, U)) &= \bigwedge_{i=1}^{m} ((i, q_\emptyset) \vee (i, q^{\neg R}) \vee (i, q^C)) \wedge \bigwedge_{R(a,b)\in\mathcal{A}_1} (-1, q^{b,C}),\\
\delta(q^{\exists R.C}, (S, U)) &= \bigvee_{i=1}^{m} ((i, q^R) \wedge (i, q^C)),\\
\delta(q^{\forall R.C}, (S, U)) &= \bigwedge_{i=1}^{m} ((i, q_\emptyset) \vee (i, q^{\neg R}) \vee (i, q^C)),
\end{aligned}$$

$$\begin{aligned}
\delta(q^{C\sqcap C'}, (x, U)) &= (0, q^C) \wedge (0, q^{C'}),\\
\delta(q^{C\sqcup C'}, (x, U)) &= (0, q^C) \vee (0, q^{C'}),\\
\delta(q^{a,C}, root) &= \bigvee_{i=1}^{m} (i, q^{a,C}),\\
\delta(q^{a,C}, (a, U)) &= (0, q^C),\\
\delta(q^A, (x, U)) &= \mathsf{true},\ \text{if } A \in U,\\
\delta(q^{\neg A}, (x, U)) &= \mathsf{true},\ \text{if } A \notin U,\\
\delta(q^R, (R, U)) &= \mathsf{true},\\
\delta(q^{\neg R}, (S, U)) &= \mathsf{true},\ \text{if } R \neq S,\\
\delta(q_\emptyset, empty) &= \mathsf{true},\\
\delta(q, \ell) &= \mathsf{false}, \quad \text{for all other } q \in Q,\ \ell \in \Gamma.
\end{aligned}$$

Here 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. The acceptance condition c is trivial ($c(q) = 0$ for all $q \in Q$). It is standard to show that $\mathfrak{A}_1$ accepts the desired tree language.

To construct $\mathfrak{A}_2$, we use the notation introduced in the proof of Proposition 9. Note that the set $\mathsf{type}(\mathcal{T}_2)$ of $\mathcal{T}_2$-types can be computed in time single exponential in $|\mathcal{K}_2|$. A *completion of* $\mathcal{K}_2$ is a function $\tau : \mathsf{ind}(\mathcal{A}_2) \to \mathsf{type}(\mathcal{T}_2)$ 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 $a \in \mathsf{ind}(\mathcal{A}_2)$. Denote by $\mathsf{compl}(\mathcal{K}_2)$ the set of all completions of $\mathcal{K}_2$; it can be computed in time single exponential in $|\mathcal{K}_2|$.

We now construct the 2APTA $\mathfrak{A}_2$. 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 Σ-homomorphically embeddable into a forest-shaped model of $\mathcal{K}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. In this case, we choose $\mathfrak{A}_2$ so that it accepts the empty language. Suppose there is no such assertion. It is 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 $S^{\mathcal{I}_2}$, for some symbol $S \in \Sigma$, is not Σ-homomorphically embeddable into a forest-shaped model of $\mathcal{K}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$. 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. We define the 2APTA $\mathfrak{A}_2 = (Q, \Gamma, \delta, q_0, c)$ by setting

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

and defining the transition function δ as follows:

$$\begin{aligned}
\delta(q_0, root) &= \bigvee_{\tau\in\mathsf{compl}_{\mathsf{ok}}(\mathcal{K}_2)} \bigwedge_{a\in\mathsf{ind}(\mathcal{A}_2)\cap\mathsf{ind}(\mathcal{A}_1)} \bigvee_{i=1}^{m} (i, q^{a,\tau(a)}),\\
\delta(q^{a,\boldsymbol{t}}, (a, U)) &= \bigwedge_{\substack{\exists R.C\in\boldsymbol{t}\\ R\in\Sigma}} \bigvee_{\boldsymbol{s}\in\mathsf{succ}_{\exists R.C}(\boldsymbol{t})} \Big(\bigvee_{i=1}^{m} (i, q^{R,\boldsymbol{s}}) \vee \bigvee_{R(a,b)\in\mathcal{A}_1} (-1, q^{b,\boldsymbol{s}})\Big) \wedge \bigwedge_{\substack{\exists R.C\in\boldsymbol{t}\\ R\notin\Sigma}} \bigvee_{\boldsymbol{s}\in\mathsf{succ}_{\exists R.C}(\boldsymbol{t})} (0, f^{\boldsymbol{s}}),\\
\delta(q^{S,\boldsymbol{t}}, (S, U)) &= \bigwedge_{\substack{\exists R.C\in\boldsymbol{t}\\ R\in\Sigma}} \bigvee_{\boldsymbol{s}\in\mathsf{succ}_{\exists R.C}(\boldsymbol{t})} \bigvee_{i=1}^{m} (i, q^{R,\boldsymbol{s}}) \wedge \bigwedge_{\substack{\exists R.C\in\boldsymbol{t}\\ R\notin\Sigma}} \bigvee_{\boldsymbol{s}\in\mathsf{succ}_{\exists R.C}(\boldsymbol{t})} (0, f^{\boldsymbol{s}}),
\end{aligned}$$

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

$$\delta(f^{\boldsymbol{t}},(v,U)) = (0,q^{v,\boldsymbol{t}}) \vee \bigvee_{i=1}^{m}(i,f^{\boldsymbol{t}}) \vee (-1,f^{\boldsymbol{t}}),$$
$$\delta(f^{\boldsymbol{t}},root) = \bigvee_{i=1}^{m}(i,f^{\boldsymbol{t}}),$$
$$\delta(q^{a,\boldsymbol{t}},root) = \bigvee_{i=1}^{m}(i,q^{a,\boldsymbol{t}}),$$
$$\delta(q,\ell) = \mathsf{false}, \quad \text{for all other } q \in Q \text{ and } \ell \in \Gamma,$$

where v is an individual a or a role name S. Note that the states $f^{\boldsymbol{t}}$ are used to find non-deterministically the homomorphic image of Σ-disconnected successors in the tree. Finally, we set $c(q) = 0$ for $q \in \{q_0, q^{a,\boldsymbol{t}}, q^{R,\boldsymbol{t}}\}$ and $c(f^{\boldsymbol{t}}) = 1$.

Lemma 31. *$(T, L) \in \mathcal{L}(\mathfrak{A}_2)$ iff there is a model $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ such that $\mathcal{I}_2$ is Σ-homomorphically embeddable into $\mathcal{I}_{(T,L)}$ preserving* $\mathsf{ind}(\mathcal{K}_2)$.

Proof. ($\Rightarrow$) Given an accepting run (T_r, r) for (T, L), we can construct a model $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ and a Σ-homomorphism h from $\mathcal{I}_2$ to $\mathcal{I}_{(T,L)}$. Intuitively, the type $\boldsymbol{t}$ of a in $\mathcal{I}_2$ is given by the child y_a of ε in T_r with $r(y_a) = (x_a, q^{a,\boldsymbol{t}})$, and the tree-shaped part of $\mathcal{I}_2$ is defined inductively as follows. If an element d of $\mathcal{I}_2$ has type $\boldsymbol{t}$ and $y_d \in T_r$, then for each $\exists R.C \in \boldsymbol{t}$ such that $R \in \Sigma$, d has an R-successor d' whose type $\boldsymbol{s} \in \mathsf{succ}_{\exists R.C}(\boldsymbol{t})$ is determined by a child $y_{d'}$ of y_d in T_r with $r(y_{d'}) = (x_{d'}, q^{v,\boldsymbol{s}})$, for some v. Moreover, for each $\exists R.C \in \boldsymbol{t}$ such that $R \notin \Sigma$, d has an R-successor d' whose type $\boldsymbol{s} \in \mathsf{succ}_{\exists R.C}(\boldsymbol{t})$ is determined by the descendants $y_1, \dots, y_n, y_{d'}$ of y_d in T_r, $n \geq 1$, with $r(y_i) = (x_i, f^{\boldsymbol{s}})$, $1 \leq i \leq n$, and $r(y_{d'}) = (x_{d'}, q^{v,\boldsymbol{s}})$ for some v. The homomorphism h is defined by taking the identity on individual names, and setting $h(d) = a$ if $r(y_d) = (x_d, q^{a,\boldsymbol{t}})$, and $h(d) = x_d$ if $r(y_d) = (x_d, q^{R,\boldsymbol{t}})$. Observe that due to the accepting condition for which $c(f^{\boldsymbol{t}}) = 1$, the automaton cannot remain forever in the states $f^{\boldsymbol{t}}$, and so has to eventually find the homomorphic image of Σ-disconnected successors in the tree.

($\Leftarrow$) Suppose there is a model $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ such that $\mathcal{I}_2$ is Σ-homomorphically embeddable into $\mathcal{I}_{(T,L)}$ preserving $\mathsf{ind}(\mathcal{K}_2)$. It is straightforward to construct an accepting run for (T, L) by using $\mathcal{I}_2$ as a guide. □

The constructed automaton $\mathfrak{A}$ has only single exponentially many states. Thus, by Theorem 29, checking its emptiness can be done in 2ExpTime.

Theorem 32. *The problem whether an $\mathcal{ALC}$ KB Σ-UCQ entails an $\mathcal{ALC}$ KB is decidable in 2ExpTime.*

We now briefly discuss the modifications needed in the automata construction to obtain the same upper bound for Σ-rUCQ entailment. In the rooted case, we modify the automaton $\mathfrak{A}_2$ in such way that it does not attempt to construct a Σ-homomorphism when reaching Σ-disconnected successors. Thus, the set Q of states of $\mathfrak{A}_2$ does not contain $f^{\boldsymbol{t}}$, and the transition function is simplified accordingly. In particular, in the definition of the transitions $\delta(q^{x,\boldsymbol{t}}, (x, U))$, for $x \in \{a, S\}$, the second set of conjunctions for $\exists R.C \in \boldsymbol{t}$ and $R \notin \Sigma$ is omitted.

Theorem 33. *The problem whether an $\mathcal{ALC}$ KB Σ-rUCQ entails an $\mathcal{ALC}$ KB is decidable in 2ExpTime.*

Our characterisation of Σ-(r)UCQ entailment using automata also allows us to formulate Theorem 25 without the restriction to regular interpretations. For UCQs, this is a consequence of Lemma 30 and, for rUCQs, one can prove an analogous lemma.

Theorem 34. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be $\mathcal{ALC}$ KBs and Σ a signature.*

(1) *$\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ iff, for any $\mathcal{I}_1 \in \boldsymbol{M}^{bo}_{\mathcal{K}_1}$, there exists $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$.
(2) *$\mathcal{K}_1$ Σ-rUCQ entails $\mathcal{K}_2$ iff, for any $\mathcal{I}_1 \in \boldsymbol{M}^{bo}_{\mathcal{K}_1}$, there exists $\mathcal{I}_2 \in \boldsymbol{M}^{bo}_{\mathcal{K}_2}$ that is con-Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{K}_2)$.

5.3. 2ExpTime *lower bound for (r)UCQ-entailment and inseparability with respect to a signature*

We first show a 2ExpTime lower bound for Σ-UCQ entailment between $\mathcal{ALC}$ KBs by giving a polynomial reduction of the word problem for exponentially space bounded alternating Turing machines. Using Lemma 26, we obtain the same lower bound for rUCQs. We then modify the KBs from the entailment case to obtain 2ExpTime lower bounds for Σ-(r)UCQ inseparability.

An *alternating Turing machine* (ATM) is a quintuple of the form $M = (Q, \Gamma_I, \Gamma, q_0, \Delta)$, where 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; Γ_I is the *input alphabet* and $\Gamma \supseteq \Gamma_I$ the *work alphabet* containing a *blank symbol* $\square$; $q_0 \in Q_\exists \cup Q_\forall$ is the *starting* state; and the *transition relation* Δ is of the form

$$\Delta \subseteq (Q \setminus \{q_a, q_r\}) \times \Gamma \times Q \times \Gamma \times \{-1, +1\}.$$

We write $\Delta(q,\sigma)$ to denote $\{(q',\sigma',m) \mid (q,\sigma,q',\sigma',m) \in \Delta\}$ and assume without loss of generality that every set $\Delta(q,\sigma)$ contains exactly two elements. A *configuration* of M is a word wqw' with $w, w' \in \Gamma^*$ and $q \in Q$. The intended meaning is that the tape contains the word ww', 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* is of the form wqw' with $q \in \{q_a, q_r\}$. A configuration wqw' is *accepting* if it is a halting configuration and $q = q_a$ or $q \in Q_\forall$ and all of its successor configurations are accepting or $q \in Q_\exists$ and there is an accepting successor configuration. M *accepts* input w if the *initial configuration* $q_0 w$ is accepting. There is an exponentially space bounded ATM M whose word problem is 2ExpTime-hard.

Theorem 35. *The problem whether an $\mathcal{ALC}$ KB $\mathcal{K}_1$ Σ-(r)UCQ entails an $\mathcal{ALC}$ KB $\mathcal{K}_2$ is 2ExpTime-hard.*

Proof. We only consider the non-rooted case; the rooted case follows using Lemma 26 since the signature Σ in our proof contains all the role names used in the entailed KB $\mathcal{K}_2$. The proof is by reduction of the word problem for exponentially space bounded ATMs. Let $M = (Q, \Gamma_I, \Gamma, q_0, \Delta)$ be such an ATM. We may assume without loss of generality that

- the length of every (path in a) computation of M on $w \in {\Gamma_I}^n$ is bounded by 2^{2^n};
- all the configurations wqw' in such computations satisfy $|ww'| \leq 2^n$, see [60];
- M never attempts to move left of the tape cell on which the head was located in the initial configuration;
- the two transitions contained in $\Delta(q,\sigma)$ are ordered and use $\delta_0(q,\sigma)$ and $\delta_1(q,\sigma)$ to denote the first and second transition in $\Delta(q,\sigma)$, respectively;
- the existential and universal states strictly alternate: any transition from an existential state leads to a universal state, and vice versa;
- $q_0 \in Q_\exists$;
- any run of M on every input stops either in q_a or q_r.

Let $w \in {\Gamma_I}^n$ be an input to M. We construct $\mathcal{ALC}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and a signature Σ such that M accepts w iff there is a model $\mathcal{I}_1$ of $\mathcal{K}_1 = (\mathcal{T}_1, \{A(a)\})$ such that no model of $\mathcal{K}_2 = (\mathcal{T}_2, \{A(a)\})$ is Σ-homomorphically embeddable into $\mathcal{I}_1$. In our construction, the models of $\mathcal{K}_1$ encode all possible sequences of configurations of M starting from the initial one and forming a full binary tree. Hence, most of the models do not correspond to correct runs of M. The branches of the models stop at the accepting and rejecting states. On the other hand, the models of $\mathcal{K}_2$ encode all possible local defects (such as invalid configurations or incorrect executions of the transition function), after the first step of the machine, or after the second step, and so on, or detect valid (hence without local defects) but rejecting runs. Then, if there exists a finite model $\mathcal{I}_1$ of $\mathcal{K}_1$ such that no model of $\mathcal{K}_2$ is Σ-homomorphically embeddable into $\mathcal{I}_1$ preserving $\{a\}$, we have that $\mathcal{I}_1$ represents a valid accepting computation of M.

The signature Σ contains the following symbols:

- the concept name A;
- the concept names $A_0, \ldots, A_{n-1}, \overline{A}_0, \ldots, \overline{A}_{n-1}$ that serve as bits in the binary representation of a number between 0 and $2^n - 1$, identifying the position of tape cells inside configurations (A_0, $\overline{A}_0$ represent the lowest bit);
- the concept names A_σ, for $\sigma \in \Gamma$;
- the concept names $A_{q,\sigma}$, for $\sigma \in \Gamma$ and $q \in Q$;
- the concept names X_0, X_1 to distinguish the two successor configurations;
- the role names R, S; R is used to connect the successor configurations, whereas S is used to connect the root of each configuration with symbols that occur in the cells of it.

Also, we make use of the following auxiliary symbols that are not in Σ:

- B_i, $\overline{B}_i$, B_σ, $B_{q,\sigma}$; G_i, $\overline{G}_i$, G_σ, $G_{q,\sigma}$; C_σ, $C_{q,\sigma}$, for $\sigma \in \Gamma$, $q \in Q$, and $0 \leq i \leq n-1$,
- L_i^ℓ, D_{trans}^ℓ, for $\ell \in \{0,1\}$ and $0 \leq i \leq n-1$,
- K_0, K, $Stop$, Y, D, $\overline{D}$, D_{conf}, D_{trans}, D_{rej}, $D_{rej}^\exists$, $D_{rej}^\forall$, $Counter_m$ for $m \in \{-1, 0, +1\}$, E_B, E_G.

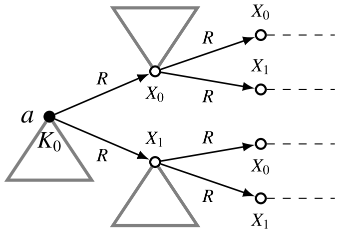


Fig. 5. The structure of the models of $\mathcal{K}_1$.

TBox $\mathcal{T}_1$. Each model of $\mathcal{K}_1$ encodes a binary tree of configurations of M. Thus, $\mathcal{T}_1$ contains the axioms:

$$\begin{aligned}
A &\sqsubseteq \exists R.(X_0 \sqcap K) \sqcap \exists R.(X_1 \sqcap K),\\
(X_0 \sqcup X_1) \sqcap \neg \mathit{Stop} &\sqsubseteq \exists R.(X_0 \sqcap K) \sqcap \exists R.(X_1 \sqcap K),\\
K &\sqsubseteq \exists S.(L_0^0 \sqcap \overline{A}_0) \sqcap \exists S.(L_0^1 \sqcap A_0),\\
L_i^\ell &\sqsubseteq \exists S.(L_{i+1}^0 \sqcap \overline{A}_{i+1}) \sqcap \exists S.(L_{i+1}^1 \sqcap A_{i+1}), \quad \text{for } 0 \le i \le n-2,\ \ell \in \{0,1\},\\
L_{n-1}^\ell &\sqsubseteq \bigsqcup_{\sigma\in\Gamma} (A_\sigma \sqcup \bigsqcup_{q\in Q} A_{q,\sigma}),\\
A_{\sigma_1} \sqcap A_{\sigma_2} &\sqsubseteq \bot, \quad \text{for } \sigma_1 \neq \sigma_2,\\
A_{\sigma_1} \sqcap A_{q_2,\sigma_2} &\sqsubseteq \bot,\\
A_{q_1,\sigma_1} \sqcap A_{q_2,\sigma_2} &\sqsubseteq \bot, \quad \text{for } (q_1,\sigma_1) \neq (q_2,\sigma_2),\\
A_i &\sqsubseteq \forall S.A_i, \qquad \overline{A}_i \sqsubseteq \forall S.\overline{A}_i,\\
\exists S^n.A_{q_a,\sigma} &\sqsubseteq \mathit{Stop}, \qquad \exists S^n.A_{q_r,\sigma} \sqsubseteq \mathit{Stop},
\end{aligned}$$

where $\exists S^n.A$ is an abbreviation for the concept $\exists S.\exists S.\ldots.\exists S.A$ (S occurs n times). The models of $\mathcal{K}_1$ look as in Fig. 5, where the grey triangles are the trees encoding configurations rooted at K except for the initial configuration. These trees are binary trees of depth n, where each leaf represents a tape cell. For $w = \sigma_1 \cdots \sigma_n$, the initial configuration is encoded at a by the following $\mathcal{T}_1$-axioms:

$$\begin{aligned}
A &\sqsubseteq \exists S.(L_0^0 \sqcap \overline{A}_0 \sqcap K_0) \sqcap \exists S.(L_0^1 \sqcap A_0 \sqcap K_0),\\
K_0 &\sqsubseteq \forall S.K_0,\\
K_0 \sqcap (\mathsf{val}_A = 0) &\sqsubseteq A_{q_0,\sigma_1},\\
K_0 \sqcap (\mathsf{val}_A = i) &\sqsubseteq A_{\sigma_{i+1}}, \ \text{for } 1 \le i \le n-1,\\
K_0 \sqcap (\mathsf{val}_A \ge n) &\sqsubseteq A_{\square},
\end{aligned}$$

where $(\mathsf{val}_A = j)$ is the conjunction over $A_i, \overline{A}_i$ expressing the fact that the value of the A-counter is j, for $j \le 2^n - 1$.

TBox $\mathcal{T}_2$. Each model of $\mathcal{K}_2$ encodes (at least) one of four possible defects:

- invalid configuration defect D_{conf};
- transition defect D_{trans} encoding errors in executing the transition function;
- copying defect D_{copy} encoding errors in copying a symbol not under the head;
- a rejecting run defect D_{rej}.

The first three defects are used to filter out sequences of configurations that do not correspond to valid runs of M. These defects are 'local', and so they are connected to a via paths. Instead, the last defect is used to detect valid rejecting runs of M, so it is 'global' and is represented by a tree. Thus, $\mathcal{T}_2$ contains the following axioms:

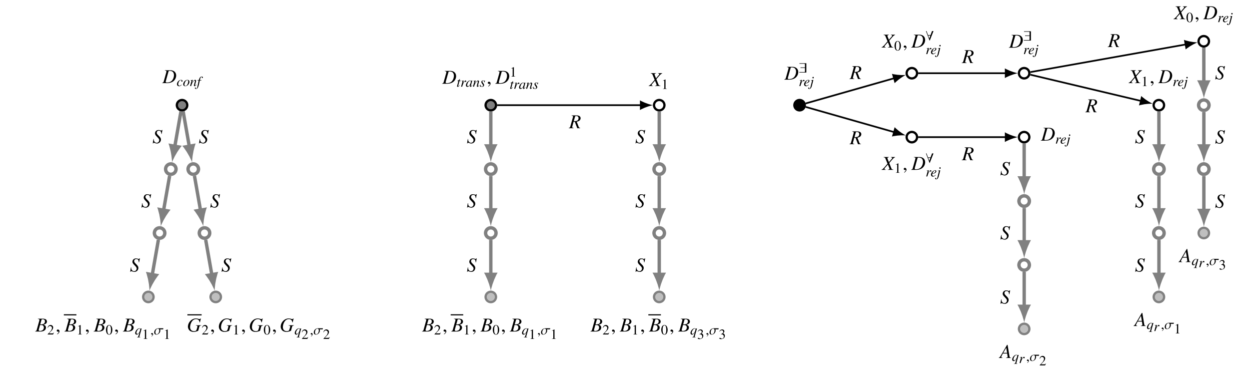
$$\begin{aligned}
A &\sqsubseteq \exists R.(X_0 \sqcap Y) \sqcup \exists R.(X_1 \sqcap Y) \sqcup D_{rej}^{\exists}, \qquad Y \sqsubseteq D \sqcup \overline{D}, \qquad D \sqcap \overline{D} \sqsubseteq \bot,\\
Y \sqcap \overline{D} &\sqsubseteq \exists R.(X_0 \sqcap Y) \sqcup \exists R.(X_1 \sqcap Y), \qquad D \sqsubseteq D_{conf} \sqcup D_{trans} \sqcup D_{copy}.
\end{aligned}$$

We now describe each of the defects separately, using the following abbreviations:

$$\mathsf{pos}^B = (\overline{B}_0 \sqcup B_0) \sqcap \cdots \sqcap (\overline{B}_{n-1} \sqcup B_{n-1}), \qquad \mathsf{symbol}^B = \bigsqcup_{\sigma\in\Gamma} B_\sigma, \qquad \mathsf{state}^B = \bigsqcup_{q\in Q,\ \sigma\in\Gamma} B_{q,\sigma}.$$

The abbreviations pos^G, symbol^G and state^G are defined analogously using concept names G_i, $\overline{G}_i$, $G_{q,\sigma}$, and G_σ.

Invalid configuration defect. D_{conf} is the simplest 'local' defect that encodes incorrect configurations, that is, configurations with at least two heads on the tape. It guesses the first position of the head, the symbol under it and the state by means of the concepts pos^B and state^B, and similarly, it guesses the second position using the corresponding concepts with the superscript G. This information is stored in the symbols transparent to Σ (B_x, $\overline{B}_x$ and G_x, $\overline{G}_x$).



(a) An invalid configuration defect. (b) A transition defect. (c) A rejecting run defect.

Fig. 6. Models of defects.

$$D_{conf} \sqsubseteq \mathsf{pos}^B \sqcap \mathsf{state}^B \sqcap \exists S^n.E_B \sqcap \mathsf{pos}^G \sqcap \mathsf{state}^G \sqcap \exists S^n.E_G \sqcap (\mathsf{val}_B \neq \mathsf{val}_G),$$

where $(\mathsf{val}_B \neq \mathsf{val}_G)$ stands for $(B_0 \sqcap \overline{G}_0) \sqcup (G_0 \sqcap \overline{B}_0) \sqcup \dots \sqcup (B_{n-1} \sqcap \overline{G}_{n-1}) \sqcup (G_{n-1} \sqcap \overline{B}_{n-1})$ and ensures that the position encoded using B-symbols is different from the position encoded using G-symbols.

All the symbols B_x and $\overline{B}_x$, and G_x and $\overline{G}_x$ are propagated down the S-successors, and at the concepts E_B and E_G they are copied into the Σ-symbols A_x and $\overline{A}_x$:

$$B_x \sqsubseteq \forall S.B_x,\ G_x \sqsubseteq \forall S.G_x,\ E_B \sqcap B_x \sqsubseteq A_x,\ E_G \sqcap G_x \sqsubseteq A_x,\ \text{ for } x \in \{0,\dots,n-1\} \cup \{(q,\sigma), \sigma \mid q \in Q, \sigma \in \Gamma\},$$
$$\overline{B}_i \sqsubseteq \forall S.\overline{B}_i,\ \overline{G}_i \sqsubseteq \forall S.\overline{G}_i,\ E_B \sqcap \overline{B}_i \sqsubseteq \overline{A}_i,\ E_G \sqcap \overline{G}_i \sqsubseteq \overline{A}_i,\ \text{ for } i \in \{0,\dots,n-1\}. \tag{25}$$

A (partial) model of an invalid configuration defect is shown in Fig. 6(a), for $n = 3$.

Transition defect. Given a (correct) configuration, D_{trans} encodes defects in a following configuration coming from an incorrect execution of the transition function. It is also a 'local' defect, but it operates on two consecutive configurations. It guesses the position of the head, the symbol under it and the state by means of the concepts pos^B and state^B, and also guesses which of the two transitions is violated:

$$D_{trans} \sqsubseteq \mathsf{pos}^B \sqcap \mathsf{state}^B \sqcap \exists S^n.E_B \sqcap (D^0_{trans} \sqcup D^1_{trans}).$$

Then, given the current state and the symbol under the head, the transition defect guesses the result of an incorrect execution of the transition function. The defective value at the successor configuration is stored in symbols C_x, while the relative position of the defect is stored in $Counter_m$, for $m \in \{-1, 0, +1\}$. Thus, for $\delta_\ell(q,\sigma) = (q_\ell, \sigma_\ell, m_\ell)$, $\ell \in \{0,1\}$, $m_\ell \in \{-1,+1\}$, we have

$$D^\ell_{trans} \sqsubseteq \exists R.(X^\ell \sqcap \exists S^n.E_B),$$
$$B_{q,\sigma} \sqcap D^\ell_{trans} \sqsubseteq (Counter_0 \sqcap \bigsqcup_{\sigma' \in \Gamma \setminus \{\sigma_\ell\}} C_{\sigma'}) \sqcup (Counter_{m_\ell} \sqcap \bigsqcup_{\sigma' \in \Gamma} (C_{\sigma'} \sqcup \bigsqcup_{q' \in Q \setminus \{q_\ell\}} C_{q',\sigma'})).$$

The position of the defect is passed/updated along the R-successor as follows:

$$\begin{aligned}
Counter_{+1} \sqcap \overline{B}_k \sqcap B_{k-1} \sqcap \dots \sqcap B_0 &\sqsubseteq \forall R.(B_k \sqcap \overline{B}_{k-1} \sqcap \dots \sqcap \overline{B}_0), && \text{for } n > k \geq 0,\\
Counter_{+1} \sqcap B \sqcap \overline{B}_k &\sqsubseteq \forall R.B, && \text{for } B \in \{B_j, \overline{B}_j \mid n > j > k\},\\
Counter_{-1} \sqcap B_k \sqcap \overline{B}_{k-1} \sqcap \dots \sqcap \overline{B}_0 &\sqsubseteq \forall R.(\overline{B}_k \sqcap B_{k-1} \sqcap \dots \sqcap B_0), && \text{for } n > k \geq 0,\\
Counter_{-1} \sqcap B \sqcap B_k &\sqsubseteq \forall R.B, && \text{for } B \in \{B_j, \overline{B}_j \mid n > j > k\},\\
Counter_0 \sqcap B &\sqsubseteq \forall R.B, && \text{for } B \in \{B_i, \overline{B}_i \mid 0 \leq i \leq n-1\}.
\end{aligned} \tag{26}$$

The defect is copied via R as follows:

$$C_x \sqsubseteq \forall R.B_x, \quad x \in \{(q,\sigma), \sigma \mid q \in Q, \sigma \in \Gamma\}. \tag{27}$$

Then the symbols B_x and $\overline{B}_x$ that have been copied via R are propagated down the S-successors, and copied at E_B into the Σ-symbols A_x and $\overline{A}_x$ using (25). A model of a transition defect is shown in Fig. 6(b), for $n = 3$ and $\delta_1(q_1, \sigma_1) = (q_2, \sigma_2, +1)$.

Copying defect. Similarly to the transition defect, the copying defect concerns two consecutive configurations and encodes errors in copying symbols that are not under the head. So it guesses a position of the head, a symbol under it, and a state by means of the concepts pos^G and state^G, and a position different from the position of the head and a symbol at this position by means of the concepts pos^B and symbol^B:

$$D_{copy} \sqsubseteq \mathsf{pos}^G \sqcap \mathsf{state}^G \sqcap \exists S^n.E_G \sqcap \mathsf{pos}^B \sqcap \mathsf{symbol}^B \sqcap \exists S^n.E_B \sqcap \exists R.\exists S^n.E_B \sqcap (\mathsf{val}_B \neq \mathsf{val}_G).$$

Then it chooses a new (incorrect) symbol (possibly with a state) at the B-position in the subsequent configuration:

$$B_\sigma \sqcap D_{copy} \sqsubseteq Counter_0 \sqcap \bigsqcup_{\sigma' \in \Gamma,\ \sigma' \neq \sigma} (C_{\sigma'} \sqcup \bigsqcup_{q \in Q} C_{q,\sigma'}).$$

Using (26) and (27), the incorrect value and its position are copied via R, and then propagated via the S-successors and copied at E_B to A-symbols using (25).

Rejecting run defect. The rejecting run defect detects when M does not accept w. It is done by checking the negation of the accepting condition. So this defect is a tree starting at A where every node at even distance from the root ($D^{\exists}_{rej}$) has two successors (recall that $q_0 \in Q_\exists$), every node at odd distance from the root ($D^{\forall}_{rej}$) has one successor, and the leaves are 'labelled' by rejecting states:

$$D^{\exists}_{rej} \sqsubseteq \bigsqcap_{\ell \in \{0,1\}} \exists R.(X_\ell \sqcap (D^{\forall}_{rej} \sqcup D_{rej})),$$
$$D^{\forall}_{rej} \sqsubseteq \exists R.(D^{\exists}_{rej} \sqcup D_{rej}),$$
$$D_{rej} \sqsubseteq \bigsqcup_{\sigma \in \Gamma} \exists S^n.A_{q_r,\sigma}.$$

A (partial) model of a rejecting defect is shown in Fig. 6(c).

Now we sketch a proof that M accepts w iff $\mathcal{K}_1$ does not Σ-UCQ-entail $\mathcal{K}_2$.

($\Rightarrow$) Suppose M accepts w. Then there is a model $\mathcal{I}_1$ of $\mathcal{K}_1$ such that

- it has no local defects, that is, it has only valid configurations, and at each step the transition function is executed correctly and all symbols not affected by the head are copied correctly;
- it contains a subtree representing an accepting computation of M on w.

Note that the former means that $\mathcal{I}_1$ is finite as we assumed that any run of M on every input stops either in q_a or q_r. So the models of $\mathcal{K}_2$ that are infinite paths or trees not 'realising' any defect (such models never actually pick D or D_{rej} to satisfy disjunction) will not be Σ-homomorphically embeddable into $\mathcal{I}_1$. Moreover, the latter implies that the models of $\mathcal{K}_2$ encoding rejecting run defect will not be Σ-homomorphically embeddable into $\mathcal{I}_1$ either. So no model of $\mathcal{K}_2$ is Σ-homomorphically embeddable into $\mathcal{I}_1$, and hence $\mathcal{K}_1$ does not Σ-UCQ entail $\mathcal{K}_2$.

($\Leftarrow$) Suppose $\mathcal{K}_1$ does not Σ-UCQ entail $\mathcal{K}_2$. Then there exists a model $\mathcal{I}_1$ of $\mathcal{K}_1$ such that no model $\mathcal{I}_2$ of $\mathcal{K}_2$ is Σ-homomorphically embeddable into $\mathcal{I}_1$. It follows that:

- parts of $\mathcal{I}_1$ in grey triangles (see Fig. 5) represent configurations with at most one head, because of the models $\mathcal{I}_2$ of $\mathcal{K}_2$ that detect invalid configurations;
- for every non-final configuration in $\mathcal{I}_1$ as explained above and for each of its two successor configurations, there are neither transition nor copying defects, because of the models of $\mathcal{I}_2$ that detect such defects;
- it is not the case that the tree of configurations represented by $\mathcal{I}_1$ witnesses that M does not accept w, because of the models $\mathcal{I}_2$ that detect such cases.

We thus conclude that $\mathcal{I}_1$ contains a valid accepting computation. □

We now modify the KBs in the proof above to obtain the following:

Theorem 36. *Σ-(r)UCQ inseparability between $\mathcal{ALC}$ KBs is 2ExpTime-hard.*

Proof. We only deal with the non-rooted case; the rooted case follows using Lemma 26. Consider the KBs $\mathcal{K}_i$, $i = 1, 2$, and the signature Σ from the proof of Theorem 35. We construct (in LogSpace) a KB $\mathcal{K}''_2$ such that $\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ iff $\mathcal{K}_1$ and $\mathcal{K}''_2$ are Σ-UCQ inseparable. This provides us with the desired lower bound for Σ-UCQ inseparability. Let $\mathcal{T}^i_i$ be a copy of $\mathcal{T}_i$ in which all concept names $X \in \mathsf{sig}(\mathcal{T}_i) \setminus \{A\}$ are replaced by fresh symbols X^i, and let $\mathcal{T}'_i$ be the extension of $\mathcal{T}^i_i$ with $X^i \sqsubseteq X$, for all concept names $X \in \Sigma \setminus \{A\}$. We set $\mathcal{K}'_i = (\mathcal{T}'_i, \{A(a)\})$, $i = 1, 2$, and let $\mathcal{K}''_2 = (\mathcal{T}'_1 \cup \mathcal{T}'_2, \{A(a)\})$. Observe that $\mathcal{K}'_i$ and $\mathcal{K}_i$ are Σ-UCQ inseparable, for $i = 1, 2$. We prove that $\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$ iff $\mathcal{K}'_1$ and $\mathcal{K}''_2$ are Σ-UCQ inseparable. The implication ($\Leftarrow$) is straightforward.

Conversely, suppose $\mathcal{K}_1$ Σ-UCQ entails $\mathcal{K}_2$. Clearly, $\mathcal{K}''_2$ Σ-UCQ entails $\mathcal{K}'_1$, and thus it remains to prove that $\mathcal{K}'_1$ Σ-UCQ entails $\mathcal{K}''_2$. For $i = 1, 2$, we consider the class $\boldsymbol{M}_i$ of models $\mathcal{I} \in \boldsymbol{M}^{bo}_{\mathcal{K}'_i}$ such that $A^{\mathcal{I}} = \{a\}$, if $a \in X^{\mathcal{I}}$ for a concept name X, then $X \in \{{D^0_{rej}}', A\}$, and $X^{\mathcal{I}} = \emptyset$, for all concept names $X \notin \mathsf{sig}(\mathcal{K}'_i)$. It follows from the construction of $\mathcal{K}_i$ that $\boldsymbol{M}_i$ is complete for $\mathcal{K}'_i$. Let

$$\boldsymbol{M} = \{\mathcal{I}_1 \uplus \mathcal{I}_2 \mid \mathcal{I}_i \in \boldsymbol{M}_i, i = 1, 2\},$$

where $\mathcal{I}_1 \uplus \mathcal{I}_2$ is the interpretation that results from merging the root a of $\mathcal{I}_1$ and $\mathcal{I}_2$. We first show that $\boldsymbol{M}$ is complete for $\mathcal{K}_2''$. The interpretations $\mathcal{I} \in \boldsymbol{M}$ are models of $\mathcal{K}_2''$ since, for all axioms $C \sqsubseteq D \in \mathcal{T}_i'$, either $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}_i} \setminus \{a\}$ or $C \in \{D^{0}_{rej}{}', A, \exists S^n.A_{q_a,\sigma}, \exists S^n.A_{q_r,\sigma}\}$ and D is either a concept name or of the form $\exists R.C'$ or $\exists S.C'$. To see that $\boldsymbol{M}$ is complete for $\mathcal{K}_2''$, let $\mathcal{J}$ be a model of $\mathcal{K}_2''$ and $n \geq 1$. It suffices to show that there exists $\mathcal{I} \in \boldsymbol{M}$ that is n-homomorphically embeddable into $\mathcal{J}$ preserving $\{a\}$ (Proposition 6). But since $\mathcal{J}$ is a model of $\mathcal{K}_i'$, there are models $\mathcal{I}_i \in \boldsymbol{M}_i$ such that $\mathcal{I}_i$ is n-homomorphically embeddable into $\mathcal{J}$ preserving $\{a\}$, $i = 1, 2$ (Proposition 6). By taking the union of the two partial witness homomorphisms from $\mathcal{I}_1$ and $\mathcal{I}_2$, one can show that $\mathcal{I}_1 \uplus \mathcal{I}_2$ is n-homomorphically embeddable into $\mathcal{J}$ preserving $\{a\}$, as required.

We now use Theorem 17 (1) to prove that $\mathcal{K}_1'$ Σ-UCQ entails $\mathcal{K}_2''$. Let $\mathcal{I}_1 \in \boldsymbol{M}_1$ and $n \geq 1$. It suffices to find $\mathcal{J} \in \boldsymbol{M}$ that is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$ preserving $\{a\}$. But since $\mathcal{K}_1'$ Σ-UCQ-entails $\mathcal{K}_2'$, there exists $\mathcal{I}_2 \in \boldsymbol{M}_2$ such that $\mathcal{I}_2$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$ preserving $\{a\}$. By combining $n\Sigma$-homomorphisms from $\mathcal{I}_2$ with the identity mapping from $\mathcal{I}_1$, it is now straightforward to show that the model $\mathcal{I}_1 \uplus \mathcal{I}_2 \in \boldsymbol{M}$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$ preserving $\{a\}$, as required. □

The following theorem summarises the results obtained so far.

Theorem 37. *Σ-(r)UCQ inseparability and Σ-(r)UCQ-entailment between $\mathcal{ALC}$ KBs are both 2ExpTime-complete.*

5.4. (r)UCQ-entailment and inseparability with full signature

We extend the 2ExpTime lower bound from Σ-(r)UCQ entailment and inseparability to full signature (r)UCQ entailment and inseparability. To this end we prove a UCQ-variant of Theorem 23:

Theorem 38. *Let $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$ be $\mathcal{ALC}$ KBs and Σ a signature such that $\mathsf{sig}(\mathcal{A}) \subseteq \Sigma$ and $\Gamma = \mathsf{sig}(\mathcal{T}_1 \cup \mathcal{T}_2) \setminus \Sigma$ contains no role names. Suppose $\mathcal{T}_1$ and $\mathcal{T}_2$ admit trivial models. Let $\mathcal{K}_i^{\uparrow\Gamma} = (\mathcal{T}_i^{\uparrow\Gamma} \cup \mathcal{T}_\Gamma^{\exists}, \mathcal{A})$, for $i = 1, 2$. Then the following conditions are equivalent:*

(1) *$\mathcal{K}_1$ Σ-(r)UCQ entails $\mathcal{K}_2$;*
(2) *$\mathcal{K}_1^{\uparrow\Gamma}$ full signature (r)UCQ entails $\mathcal{K}_2^{\uparrow\Gamma}$.*

Proof. We use and modify the proof of Theorem 23. Let $\boldsymbol{M}_i$ be complete for $\mathcal{K}_i$, $i = 1, 2$. We may assume that $X^{\mathcal{I}} = \emptyset$ for all concept and role names $X \notin \mathsf{sig}(\mathcal{K}_i)$ and $\mathcal{I} \in \boldsymbol{M}_i$, $i = 1, 2$. By Fact 5 of the proof of Theorem 23, $\{\mathcal{I}^{\uparrow\Gamma} \mid \mathcal{I} \in \boldsymbol{M}_i\}$ is complete for $\mathcal{K}_i^{\uparrow\Gamma}$. Thus, by Theorem 17, it suffices to prove that $\mathcal{I}_2$ is $n\Sigma$-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{K}_2)$ iff $\mathcal{I}_2^{\uparrow\Gamma}$ is n-homomorphically embeddable into $\mathcal{I}_1^{\uparrow\Gamma}$ preserving $\mathsf{ind}(\mathcal{K}_2)$, for any $n > 0$, $\mathcal{I}_1 \in \boldsymbol{M}_1$ and $\mathcal{I}_2 \in \boldsymbol{M}_2$. This can be done in the same way as in the proof of Fact 6. □

The following complexity result now follows from the observation that the KBs and signature Σ used in the proof of Theorem 36 satisfy the conditions of Theorem 38: Σ contains the signature of the ABox and all role names of the KBs, and the TBoxes admit trivial models.

Theorem 39. *Full signature (r)UCQ inseparability and entailment between $\mathcal{ALC}$ KBs are both 2ExpTime-complete.*

6. Query entailment and inseparability for $\mathcal{ALC}$ TBoxes

In this section, we introduce query entailment and inseparability between TBoxes. Two TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ are query inseparable for a class $\mathcal{Q}$ of queries if, for all ABoxes $\mathcal{A}$ that are consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, queries from $\mathcal{Q}$ have the same certain answers over the KBs $(\mathcal{T}_1, \mathcal{A})$ and $(\mathcal{T}_2, \mathcal{A})$. The TBox $\mathcal{T}_1$ $\mathcal{Q}$-entails $\mathcal{T}_2$ if, for any such $\mathcal{A}$, the certain answers to queries from $\mathcal{Q}$ over $(\mathcal{T}_2, \mathcal{A})$ are contained in the certain answers over $(\mathcal{T}_1, \mathcal{A})$. As in the KB case, we consider the restriction of CQs and UCQs to a signature Σ of relevant symbols and their restrictions to rooted queries. In applications, it is also natural to restrict the signature of the ABox which might be different from the signature of the relevant queries.

Definition 40. 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*'.

In the definition of Θ-$\mathcal{Q}$ entailment, we only consider ABoxes that are consistent with both TBoxes. The reason is that the complexity of the problem of deciding whether every Σ-ABox consistent with a TBox $\mathcal{T}_1$ is also consistent with a TBox $\mathcal{T}_2$ is already well understood and is dominated by the Θ-$\mathcal{Q}$-entailment problem as defined above. More precisely, we say that a TBox $\mathcal{T}_1$ $\Sigma_\perp$*-entails* a TBox $\mathcal{T}_2$ if all Σ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ are consistent with $\mathcal{T}_2$. $\Sigma_\perp$-entailment is closely related to the containment problem between ontology-mediated queries, which we define next [61–63]. For a query $\boldsymbol{q}$, TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, and a signature Σ, we say that $(\mathcal{T}_1, \boldsymbol{q})$ *is contained in* $(\mathcal{T}_2, \boldsymbol{q})$ *for* Σ and write $(\mathcal{T}_1, \boldsymbol{q}) \subseteq_\Sigma (T_2, \boldsymbol{q})$ if, for every Σ-ABox $\mathcal{A}$, the certain answers to $\boldsymbol{q}$ over $(\mathcal{T}_1, \mathcal{A})$ are contained in the certain answers to $\boldsymbol{q}$ over $(\mathcal{T}_2, \mathcal{A})$. We note that the authors of [61,63] demand that the Σ-ABoxes considered in the definition of containment are consistent with both TBoxes, but the complexity results for deciding containment do not depend on this condition. The *containment problem* for a description logic $\mathcal{L}$ relative to a class $\mathcal{Q}$ of queries is to decide, for TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ in $\mathcal{L}$, signature Σ, and query $\boldsymbol{q} \in \mathcal{Q}$, whether $(\mathcal{T}_1, \boldsymbol{q}) \subseteq_\Sigma (T_2, \boldsymbol{q})$. Thus, in contrast to Θ-$\mathcal{Q}$-entailment, an instance of the containment problem does not quantify over all $\boldsymbol{q} \in \mathcal{Q}$ but takes the queries $\boldsymbol{q} \in \mathcal{Q}$ as inputs to the decision problem. It is known [61–63] that the containment problem is

- NExpTime-complete for $\mathcal{ALC}$ TBoxes and CQs of the form $\exists x A(x)$;
- ExpTime-complete for *Horn*$\mathcal{ALC}$ TBoxes and CQs of the form $\exists x A(x)$.

It is straightforward to show that the containment problem for a DL $\mathcal{L}$ and CQs of the form $\exists x A(x)$ is mutually polynomially reducible with the problem to decide $\Sigma_\perp$-entailment between $\mathcal{L}$ TBoxes. For a polynomial reduction of $\Sigma_\perp$-entailment to containment, observe that $\mathcal{T}_1$ $\Sigma_\perp$-entails $\mathcal{T}_2$ iff $(\mathcal{T}_2, \exists x A(x)) \subseteq_\Sigma (\mathcal{T}_1, \exists x A(x))$ for $A \notin \mathsf{sig}(\mathcal{T}_1 \cup \mathcal{T}_2)$. For a polynomial reduction of containment to $\Sigma_\perp$-entailment, assume that $\mathcal{T}_1, \mathcal{T}_2, \Sigma$, and A are given. Let $\mathcal{T}_i' = \mathcal{T}_i \cup \{A \sqsubseteq \perp\}$. Then $(\mathcal{T}_1, \exists x A(x)) \subseteq_\Sigma (\mathcal{T}_2, \exists x A(x))$ iff $\mathcal{T}_2'$ $\Sigma_\perp$-entails $\mathcal{T}_1'$. We obtain the following result.

Theorem 41. *The problem whether an $\mathcal{ALC}$ TBox $\Sigma_\perp$-entails an $\mathcal{ALC}$ TBox is* NExpTime*-complete. For Horn$\mathcal{ALC}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, this problem is* ExpTime*-complete.*

It follows, in particular, that our complexity upper bounds for Θ-CQ-entailment still hold if one admits ABoxes that are not consistent with the TBoxes.

As in the KB case, Θ-UCQ inseparability of $\mathcal{ALC}$ TBoxes implies all other types of inseparability, and Example 13 can be used to show that no other implications hold in general. The situation is different for *Horn*$\mathcal{ALC}$ TBoxes. In fact, the following result follows directly from Proposition 14:

Proposition 42. *For any $\mathcal{ALC}$ TBox $\mathcal{T}_1$ and Horn$\mathcal{ALC}$ TBox $\mathcal{T}_2$, $\mathcal{T}_1$ Θ-(r)UCQ entails $\mathcal{T}_2$ iff $\mathcal{T}_1$ Θ-(r)CQ entails $\mathcal{T}_2$.*

We now show that Θ-(r)CQ entailment and inseparability are undecidable for $\mathcal{ALC}$ TBoxes. In fact, we show that Θ-(r)CQ inseparability is undecidable even if one of the TBoxes is given in $\mathcal{EL}$ and that Θ-(r)CQ entailment is undecidable even if the entailing TBox $\mathcal{T}_1$ is in $\mathcal{EL}$. The proofs re-use the TBoxes constructed in the undecidability proofs for KBs in Theorems 20 and 22. We also show that, for CQs, these problems are still undecidable in the full ABox signature case or if one assumes that the signatures for the ABoxes and CQs coincide. It remains open whether rCQ-entailment or inseparability are still undecidable in those cases.

Theorem 43. (*i*) *The problem whether an $\mathcal{EL}$ TBox Θ-$\mathcal{Q}$ entails an $\mathcal{ALC}$ TBox is undecidable for $\mathcal{Q} \in \{CQ, rCQ\}$.*

(*ii*) *Θ-$\mathcal{Q}$ inseparability between $\mathcal{EL}$ 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 $\Sigma_1 = \Sigma_2$.*

Proof. Here, we focus on the CQs; the proofs for rCQs are given in the appendix. We use the KBs $\mathcal{K}^1_{\mathrm{CQ}} = (\mathcal{T}^1_{\mathrm{CQ}}, \mathcal{A}_{\mathrm{CQ}})$ and $\mathcal{K}^2_{\mathrm{CQ}} = (\mathcal{T}^2_{\mathrm{CQ}}, \mathcal{A}_{\mathrm{CQ}})$ and the signature $\Sigma_{\mathrm{CQ}} = \mathsf{sig}(\mathcal{K}^1_{\mathrm{CQ}})$ from the proof of Theorem 20. Recall that it is undecidable whether $\mathcal{K}^1_{\mathrm{CQ}}$ Σ_{CQ}-CQ entails $\mathcal{K}^2_{\mathrm{CQ}}$. Also recall that, for $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A}_{\mathrm{CQ}})$ with $\mathcal{T}_2 = \mathcal{T}^1_{\mathrm{CQ}} \cup \mathcal{T}^2_{\mathrm{CQ}}$, it is undecidable whether $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ are Σ_{CQ}-CQ inseparable (Theorem 21).

(*i*) Let $\Sigma_1 = \{A\}$, $\Sigma_2 = \Sigma_{\mathrm{CQ}}$, and $\Theta = (\Sigma_1, \Sigma_2)$. We show that $\mathcal{T}^1_{\mathrm{CQ}}$ Θ-CQ-entails $\mathcal{T}^2_{\mathrm{CQ}}$ iff $\mathcal{K}^1_{\mathrm{CQ}}$ Σ_{CQ}-CQ-entails $\mathcal{K}^2_{\mathrm{CQ}}$. Recall that $\mathcal{A}_{\mathrm{CQ}} = \{A(a)\}$. Thus, if $\mathcal{K}^1_{\mathrm{CQ}}$ does not Σ_{CQ}-CQ entail $\mathcal{K}^2_{\mathrm{CQ}}$, then we have found a Σ_1-ABox witnessing that $\mathcal{T}^1_{\mathrm{CQ}}$ does not Θ-CQ entail $\mathcal{T}^2_{\mathrm{CQ}}$. Conversely, observe that Σ_1-ABoxes $\mathcal{A}$ are sets of the form $\{A(b) \mid b \in I\}$, with I a finite set of individual names. Thus, if there exists a Σ_1-ABox $\mathcal{A}$ such that $(\mathcal{T}^1_{\mathrm{CQ}}, \mathcal{A})$ does not Σ_{CQ}-CQ entail $(\mathcal{T}^2_{\mathrm{CQ}}, \mathcal{A})$, then $(\mathcal{T}^1_{\mathrm{CQ}}, \mathcal{A}_{\mathrm{CQ}})$ does not Σ_{CQ}-CQ entail $(\mathcal{T}^2_{\mathrm{CQ}}, \mathcal{A}_{\mathrm{CQ}})$ either.

(*ii*) Set again $\Theta = (\Sigma_1, \Sigma_2)$, for $\Sigma_1 = \{A\}$ and $\Sigma_2 = \Sigma_{\mathrm{CQ}}$. In exactly the same way as in (*i*) one can show that $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ are Σ_{CQ}-inseparable iff $\mathcal{T}^1_{\mathrm{CQ}}$ and $\mathcal{T}_2$ are Θ-CQ inseparable.

(*iii*) We first show undecidability of full ABox signature Σ-CQ inseparability. The undecidability of full ABox signature Σ-CQ entailment follows directly from our proof. We employ the abstraction technique from Theorem 23 for $\Gamma = \mathsf{sig}(\mathcal{T}_2) \setminus \Sigma_{\mathrm{CQ}}$. Let $\mathcal{T}_1' = \mathcal{T}^1_{\mathrm{CQ}} \cup \mathcal{T}^\exists_\Gamma$, $\mathcal{T}_2' = \mathcal{T}_2^{\uparrow\Gamma} \cup \mathcal{T}^\exists_\Gamma$ and $\Sigma = \Sigma_{\mathrm{CQ}} \setminus \{P\}$. We aim to prove that the following conditions are equivalent:

(1) $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ are Σ-CQ inseparable;
(2) $\mathcal{T}'_1$ and $\mathcal{T}'_2$ are full ABox signature Σ-CQ inseparable.

Observe that undecidability of full ABox signature CQ-inseparability of TBoxes of the form $\mathcal{T}'_1$ and $\mathcal{T}'_2$ follows since the proof of Theorems 20 and 21 shows that the role name P is not needed to CQ-separate the KBs $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ (if they are Σ_{CQ}-CQ separable). Thus, it is undecidable whether $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ are Σ-CQ inseparable.

The implication (2) $\Rightarrow$ (1) is straightforward: if $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ are not Σ-CQ inseparable, then the ABox $\mathcal{A}_{\mathrm{CQ}}$ witnesses that $\mathcal{T}'_1$ and $\mathcal{T}'_2$ are not full ABox signature Σ-CQ inseparable. Conversely, suppose $\mathcal{T}'_1$ and $\mathcal{T}'_2$ are not full ABox signature Σ-CQ inseparable. Then there exists an ABox $\mathcal{A}$ such that $(\mathcal{T}'_1, \mathcal{A})$ and $(\mathcal{T}'_2, \mathcal{A})$ are not Σ-CQ inseparable. The canonical model $\mathcal{I}_1$ of the $\mathcal{EL}$ KB $(\mathcal{T}'_1, \mathcal{A})$ can be constructed as follows:

- for any $A(b) \in \mathcal{A}$, take a copy of the canonical model $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ and hook it to b by identifying a in $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ with b;
- for any $D(b) \in \mathcal{A}$, take a copy of the subinterpretation of the canonical model $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ rooted at the P-successor of a and hook it to b by identifying the P-successor of a with b;
- for any $E(b) \in \mathcal{A}$, take a copy of the (unique up to isomorphism) subinterpretation of the canonical model $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ rooted at an E-node and hook it to b by identifying the E-node with b;
- to satisfy $\mathcal{T}^{\exists}_{\Gamma}$, let $\mathcal{J}$ be the singleton interpretation with $X^{\mathcal{J}} = \emptyset$ for all concept and role names X; we hook to any element u of the interpretation constructed so far a copy of $\mathcal{J}^{\uparrow\Gamma}$ by identifying the root of $\mathcal{J}^{\uparrow\Gamma}$ with u (see the proof of Theorem 23 for the construction and properties of $\mathcal{J}^{\uparrow\Gamma}$).

Let $\boldsymbol{M}$ be the class of interpretations obtained from $\mathcal{I}_1$ by adding to any b with $A(b) \in \mathcal{A}$ a P-successor b' to which one hooks the subinterpretation rooted in the P-successor of a in an interpretation from $\{\mathcal{I}^{\uparrow\Gamma} \mid \mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{CQ}}}\}$. One can show that $\boldsymbol{M}$ is complete for the KB $(\mathcal{T}'_2, \mathcal{A})$. To this end, first recall from the proof of Theorem 21 that for the canonical model $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ of $\mathcal{K}^1_{\mathrm{CQ}}$, the set $\boldsymbol{M}_{\mathcal{K}_2} = \{\, \mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}} \mid \mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{CQ}}} \,\}$ (where $\mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$ is the interpretation that results from merging the roots a of $\mathcal{I}$ and $\mathcal{I}_{\mathcal{K}^1_{\mathrm{CQ}}}$) is complete for $\mathcal{K}_2$. By Theorem 23 (Fact 5), $\{\mathcal{I}^{\uparrow\Gamma} \mid \mathcal{I} \in \boldsymbol{M}_{\mathcal{K}_2}\}$ is complete for $\mathcal{K}_2^{\uparrow\Gamma}$. Now completeness of $\boldsymbol{M}$ for $(\mathcal{T}'_2, \mathcal{A})$ follows directly from the fact that every $\mathcal{I} \in \boldsymbol{M}$ is a model of $(\mathcal{T}'_2, \mathcal{A})$. Next, observe that $P \notin \Sigma$ and that two KBs are Σ-CQ inseparable iff they are Σ-CQ inseparable for connected Σ-CQs. Thus, the only Σ-components of interpretations in $\boldsymbol{M}$ that could distinguish Σ-CQs true in $\boldsymbol{M}$ from Σ-CQs true in $\mathcal{I}_1$ are the interpretations $\{\mathcal{I}^{\uparrow\Gamma} \mid \mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{CQ}}}\}$. It follows that if $(\mathcal{T}'_1, \mathcal{A})$ and $(\mathcal{T}'_2, \mathcal{A})$ are not Σ-CQ inseparable, then $(\mathcal{K}^1_{\mathrm{CQ}})^{\uparrow\Gamma}$ and $\mathcal{K}_2^{\uparrow\Gamma}$ are not Σ-CQ inseparable either. But then, by the proof of Theorem 24, $\mathcal{K}^1_{\mathrm{CQ}}$ and $\mathcal{K}_2$ are not Σ-CQ inseparable, as required.

To show undecidability of Θ-CQ inseparability and entailment for $\Theta = (\Sigma_1, \Sigma_2)$ with $\Sigma_1 = \Sigma_2$, we re-use the undecidability proof for the full ABox signature case. Set $\Theta = (\Sigma, \Sigma)$. Then the proof above shows that $\mathcal{T}'_1$ and $\mathcal{T}'_2$ are Θ-CQ inseparable iff they are full ABox signature Σ-CQ inseparable since one can always choose the ABox $\mathcal{A}_{\mathrm{CQ}}$ as a witness for CQ-inseparability if $\mathcal{T}'_1$ and $\mathcal{T}'_2$ are full ABox signature Σ-CQ inseparable. □

7. Model-theoretic criteria for query entailment of *Horn$\mathcal{ALC}$* TBoxes by $\mathcal{ALC}$ TBoxes

We have seen that Θ-(r)CQ entailment of an $\mathcal{ALC}$ TBox $\mathcal{T}_2$ by an $\mathcal{EL}$ TBox $\mathcal{T}_1$ is undecidable. We now investigate the converse direction, with drastically different results (which even hold if $\mathcal{EL}$ TBoxes are replaced by *Horn$\mathcal{ALC}$* TBoxes). Thus, in this section, we give model-theoretic criteria for Θ-(r)CQ entailment of a *Horn$\mathcal{ALC}$* TBox $\mathcal{T}_2$ by an $\mathcal{ALC}$ TBox $\mathcal{T}_1$. In the next section, we use these criteria to prove tight complexity bounds for deciding Θ-(r)CQ entailment and inseparability. Recall that, by Proposition 42, our model-theoretic criteria and complexity results also apply to Θ-(r)UCQ entailment.

We assume that *Horn$\mathcal{ALC}$* TBoxes are given in *normal form* where concept inclusions look as follows:

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

and A, B are concept names. It is standard (see, e.g., [64, Proposition 28]) to show the following reduction of Θ-(r)CQ entailment for arbitrary *Horn$\mathcal{ALC}$* TBoxes to *Horn$\mathcal{ALC}$* TBoxes in normal form.

Proposition 44. *For any Horn$\mathcal{ALC}$ TBox $\mathcal{T}_2$ and any pair Θ of signatures, one can construct in polynomial time a Horn$\mathcal{ALC}$ TBox $\mathcal{T}'_2$ in normal form such that an $\mathcal{ALC}$ TBox $\mathcal{T}_1$ Θ-(r)CQ entails $\mathcal{T}_2$ iff $\mathcal{T}_1$ Θ-(r)CQ entails $\mathcal{T}'_2$.*

Our model-theoretic criteria are based on two crucial observations. First, to characterise Θ-(r)CQ entailment between *Horn$\mathcal{ALC}$* TBoxes and $\mathcal{ALC}$ TBoxes, it suffices to consider a very restricted class of acyclic (r)CQs that corresponds exactly to queries constructed using $\mathcal{EL}$ concepts. Second, it suffices to consider ABoxes that are tree-shaped rather than arbitrary ABoxes when searching for witnesses for non-Θ-(r)CQ entailment. We begin by introducing the relevant classes of CQs and rCQs. A *rooted $\mathcal{EL}$ query* takes the form $C(x)$, where C is an $\mathcal{EL}$ concept. The set of rooted $\mathcal{EL}$ queries is denoted by rELQ.

Given a KB $\mathcal{K}$, $a \in \mathsf{ind}(\mathcal{K})$, and an rELQ $C(x)$ we say that *a is a certain answer to $C(x)$ over $\mathcal{K}$* if $a^{\mathcal{I}} \in C^{\mathcal{I}}$, for every model $\mathcal{I}$ of $\mathcal{K}$. Note that rELQs can be regarded as acyclic CQs with one answer variable. A *Boolean $\mathcal{EL}$ query* takes the form $\exists x C(x)$, where C is an $\mathcal{EL}$ concept. The set of rooted and Boolean $\mathcal{EL}$ queries is denoted by ELQ. Given a KB $\mathcal{K}$ and a Boolean $\mathcal{EL}$ query $\exists x C(x)$, we say that $\mathcal{K}$ *entails* $\exists x C(x)$ if $C^{\mathcal{I}} \neq \emptyset$, for every model $\mathcal{I}$ of $\mathcal{K}$. Boolean $\mathcal{EL}$ queries can be regarded as Boolean acyclic CQs. In what follows we use the same notation for (r)ELQs as for (r)CQs. For TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and a pair $\Theta = (\Sigma_1, \Sigma_2)$ of signatures, we say that $\mathcal{T}_1$ Θ-(r)ELQ *entails* $\mathcal{T}_2$ if, for every Σ_1 ABox $\mathcal{A}$ that is consistent with both $\mathcal{T}_1$ and $\mathcal{T}_2$, and every Σ_2-(r)ELQ $\boldsymbol{q}(a)$ with $a \in \mathsf{ind}(\mathcal{A})$, whenever $(\mathcal{T}_2, \mathcal{A}) \models \boldsymbol{q}(a)$ then $(\mathcal{T}_1, \mathcal{A}) \models \boldsymbol{q}(a)$.

Proposition 45. *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. Then $\mathcal{T}_1$ Θ-(r)CQ entails $\mathcal{T}_2$ iff $\mathcal{T}_1$ Θ-(r)ELQ entails $\mathcal{T}_2$.*

Proof. Suppose $\mathcal{A}$ is a Σ_1-ABox and $(\mathcal{T}_2, \mathcal{A}) \models \boldsymbol{q}(\boldsymbol{a})$ but $(\mathcal{T}_1, \mathcal{A}) \not\models \boldsymbol{q}(\boldsymbol{a})$ for a Σ_2-CQ $\boldsymbol{q}$. As $(\mathcal{T}_2, \mathcal{A}) \models \boldsymbol{q}(\boldsymbol{a})$, there is a homomorphism $h \colon \boldsymbol{q} \to \mathcal{I}_{(\mathcal{T}_2,\mathcal{A})}$. Let $\mathcal{I}$ be the Σ_2-reduct of the subinterpretation of $\mathcal{I}_{(\mathcal{T}_2,\mathcal{A})}$ induced by the image of $\boldsymbol{q}$ under h. Then $\mathcal{I}$ is the disjoint union of

- ditree interpretations $\mathcal{I}_a$ attached to $a \in \mathsf{ind}(\mathcal{A}) \cap \Delta^{\mathcal{I}}$ such that $\mathsf{ind}(\mathcal{A}) \cap \Delta^{\mathcal{I}_a} = \{a\}$, and
- ditree interpretations $\mathcal{J}$ with $\mathsf{ind}(\mathcal{A}) \cap \Delta^{\mathcal{J}} = \emptyset$ (there exists no such $\mathcal{J}$ if $\boldsymbol{q}$ is an rCQ),

and, additionally, pairs (a, b) in $R^{\mathcal{I}}$ for $a, b \in \mathsf{ind}(\mathcal{A}) \cap \Delta^{\mathcal{I}}$, $R \in \Sigma_1$, and $R(a, b) \in \mathcal{A}$. Thus, if $\boldsymbol{q}$ is an rCQ then there exists $\mathcal{I}_a$ such that the canonical CQ $\boldsymbol{q}_{\mathcal{I}_a}(x)$ determined by $\mathcal{I}_a$ is an rELQ (see the proof of Proposition 6) and $(\mathcal{T}_2, \mathcal{A}) \models \boldsymbol{q}_{\mathcal{I}_a}(a)$ but $(\mathcal{T}_1, \mathcal{A}) \not\models \boldsymbol{q}_{\mathcal{I}_a}(a)$, as required. If $\boldsymbol{q}$ is not an rCQ and no such $\mathcal{I}_a$ exists, then there exists $\mathcal{J}$ such that the canonical CQ $\boldsymbol{q}_{\mathcal{J}}$ determined by $\mathcal{J}$ is a Boolean $\mathcal{EL}$ query and $(\mathcal{T}_2, \mathcal{A}) \models \boldsymbol{q}_{\mathcal{J}}$ but $(\mathcal{T}_1, \mathcal{A}) \not\models \boldsymbol{q}_{\mathcal{J}}$. $\square$

An ABox $\mathcal{A}$ is called a *tree ABox* if the undirected graph

$$G_{\mathcal{A}} = \big(\mathsf{ind}(\mathcal{A}), \big\{\{a, b\} \mid R(a, b) \in \mathcal{A}\big\}\big)$$

is an undirected tree and $R(a, b) \in \mathcal{A}$ implies $R(b, a) \notin \mathcal{A}$ and $S(a, b) \notin \mathcal{A}$, for $S \neq R$. The *outdegree* of $\mathcal{A}$ is defined as the outdegree of $G_{\mathcal{A}}$.

Theorem 46. *Let $\mathcal{T}_1$ be an $\mathcal{ALC}$ TBox, $\mathcal{T}_2$ a Horn$\mathcal{ALC}$ TBox, and $\Theta = (\Sigma_1, \Sigma_2)$. Then*

(1) *$\mathcal{T}_1$ Θ-rCQ-entails $\mathcal{T}_2$ iff, for any tree Σ_1-ABox $\mathcal{A}$ of outdegree bounded by $|\mathcal{T}_2|$ and consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, and any model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$, $\mathcal{I}_{(\mathcal{T}_2,\mathcal{A})}$ is con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{A})$.
(2) *$\mathcal{T}_1$ Θ-CQ-entails $\mathcal{T}_2$ iff, for any tree Σ_1-ABox $\mathcal{A}$ of outdegree bounded by $|\mathcal{T}_2|$ and consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, and any model $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$, $\mathcal{I}_{(\mathcal{T}_2,\mathcal{A})}$ is Σ_2-homomorphically embeddable into $\mathcal{I}_1$ preserving* $\mathsf{ind}(\mathcal{A})$.

Proof. (1) The direction from left to right follows from Theorem 34 and Proposition 14. Conversely, suppose $\mathcal{T}_1$ does not Θ-rCQ-entail $\mathcal{T}_2$. By Proposition 45, there are a Σ_1-ABox $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, a Σ_2-rELQ $C(x)$, and $a \in \mathsf{ind}(\mathcal{A})$ such that $(\mathcal{T}_2, \mathcal{A}) \models C(a)$ and $(\mathcal{T}_1, \mathcal{A}) \not\models C(a)$. It is shown in [64] (proof of Proposition 30)[2] that there exist a tree Σ_1-ABox $\mathcal{A}'$ with outdegree bounded by $|\mathcal{T}_2|$ and $(\mathcal{T}_2, \mathcal{A}') \models C(a)$, and an ABox homomorphism[3] h from $\mathcal{A}'$ to $\mathcal{A}$ with $h(a) = a$. It follows from Proposition 63 in the appendix that $\mathcal{A}'$ is consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$ and that $(\mathcal{T}_1, \mathcal{A}') \not\models C(a)$. Let $\mathcal{I}_1$ be a model of $(\mathcal{T}_1, \mathcal{A}')$ such that $\mathcal{I}_1 \not\models C(a)$. We know that $\mathcal{I}_{(\mathcal{T}_2,\mathcal{A}')} \models C(a)$. Thus, $\mathcal{I}_{(\mathcal{T}_2,\mathcal{A}')}$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$ preserving $\mathsf{ind}(\mathcal{A}')$, as required. (2) is proved similarly using ELQs instead of rELQs and Σ_2-homomorphisms instead of con-Σ_2-homomorphisms. $\square$

The notion of (con-)Σ-CQ homomorphic embeddability used in Theorem 46 is slightly unwieldy to use in the subsequent definitions and automata constructions. We therefore resort to simulations whose advantage is that they are compositional (they can be partial and are closed under unions). 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 $(d_1, d_2) \in \mathcal{S}$. Observe that every Σ-homomorphism from $\mathcal{I}_1$ to $\mathcal{I}_2$ is a Σ-simulation. Conversely, if $\mathcal{I}_1$ is a ditree interpretation and $(\mathcal{I}_1, d_1) \leq_\Sigma (\mathcal{I}_2, d_2)$, then one can construct a Σ-homomorphism h from $\mathcal{I}_1$ to $\mathcal{I}_2$ with $h(d_1) = d_2$.

---

[2] The proof of Proposition 30 in [64] shows this for $\mathcal{ELIF}_\bot$ TBoxes. Observe that we can regard every *Horn$\mathcal{ALC}$* TBox in normal form as an $\mathcal{ELI}_\bot$ TBox by replacing $A \sqsubseteq \forall R.B$ by $\exists R^-.A \sqsubseteq B$.

[3] ABox homomorphisms are defined before Proposition 63 in the appendix.

Lemma 47. *Let Σ_1 and Σ_2 be signatures, $\mathcal{A}$ a Σ_1-ABox, and $\mathcal{I}_1$ a model of $(\mathcal{T}_1, \mathcal{A})$. Then*
(i) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ 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 $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d) \not\leq_{\Sigma_2} (\mathcal{I}_1, e)$.*

(ii) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not Σ_2-homomorphically embeddable into $\mathcal{I}_1$ if there is $a \in \mathsf{ind}(\mathcal{A})$ such that (1) or (2) or (3) holds, where

(3) *there is an element d in the subinterpretation 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 $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d_0) \not\leq_{\Sigma_2} (\mathcal{I}_1, e)$, for all elements e of $\mathcal{I}_1$.*

Proof. We only prove (*ii*) as (*i*) is a direct consequence of our proof. Clearly, if there exists $a \in \mathsf{ind}(\mathcal{A})$ such that (1) or (2) or (3) holds for a, then there does not exist a Σ-homomorphism from $\mathcal{I}_1$ to $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ preserving $\{a\} \subseteq \mathsf{ind}(\mathcal{A})$.

Conversely, suppose none of (1), (2) or (3) holds for any $a \in \mathsf{ind}(\mathcal{A})$. Then, for any $a \in \mathsf{ind}(\mathcal{A})$, 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 Σ_2-simulation $\mathcal{S}_d$ from $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to $\mathcal{I}_1$ such that $(d, d') \in \mathcal{S}_d$. As the subinterpretation of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ rooted at d is a ditree interpretation, we can assume that $\mathcal{S}_d$ is a partial function. Also, for every d_0 in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with $d_0 \notin \mathsf{ind}(\mathcal{A})$ that has an R_0-predecessor in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with $R_0 \notin \Sigma_2$, we find an e in $\mathcal{I}_1$ such that there is a Σ_2-simulation $\mathcal{S}_{d_0}$ between $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ and $\mathcal{I}_1$ with $(d_0, e) \in \mathcal{S}_{d_0}$. As the subinterpretation of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ rooted at d_0 is ditree interpretation, we can assume that $\mathcal{S}_{d_0}$ 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 $\mathcal{S}_d$ and $\mathcal{S}_{d_0}$. It can be verified that h is a Σ_2-homomorphism from $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to $\mathcal{I}_1$. □

8. Decidability of query entailment of *Horn$\mathcal{ALC}$* TBoxes by $\mathcal{ALC}$ TBoxes

We show that the problem whether an $\mathcal{ALC}$ TBox Θ-CQ entails a *Horn$\mathcal{ALC}$* TBox is in 2ExpTime, and that the complexity drops to ExpTime in the case of rooted CQs. Using the fact that satisfiability of *Horn$\mathcal{ALC}$* TBoxes is ExpTime-hard, it is straightforward to prove a matching ExpTime lower bound even for the full ABox signature case and (Σ, Σ)-rCQ entailment and inseparability between *Horn$\mathcal{ALC}$* TBoxes. Proving a matching lower bound for the non-rooted case is more involved. Using a reduction of exponentially space bounded alternating Turing machines, we show that (Σ, Σ)-CQ inseparability between the empty TBox and *Horn$\mathcal{ALC}$* TBoxes is 2ExpTime-hard. It follows that both (Σ, Σ)-CQ inseparability and (Σ, Σ)-CQ entailment between *Horn$\mathcal{ALC}$* TBoxes are 2ExpTime-hard. The problem whether the 2ExpTime upper bound is tight in the full ABox signature case remains open.

8.1. ExpTime *upper bound for Θ-rCQ-entailment of Horn$\mathcal{ALC}$ TBoxes by $\mathcal{ALC}$ TBoxes*

Our aim is to establish the following:

Theorem 48. *Θ-rCQ inseparability between Horn$\mathcal{ALC}$ TBoxes and Θ-rCQ entailment of a Horn$\mathcal{ALC}$ TBox by an $\mathcal{ALC}$ TBox are both* ExpTime*-complete. The* ExpTime *lower bound holds already for Θ of the form (Σ, Σ) and the full ABox signature case.*

The lower bounds can be proved in a straightforward way using the fact that satisfiability of *Horn$\mathcal{ALC}$* TBoxes is ExpTime-hard. Note that ExpTime-hardness of (Σ, Σ)-rCQ inseparability is also inherited from [38], where this bound is shown for $\mathcal{EL}$ TBoxes. It thus remains to prove the upper bound.

We use a mix of two-way alternating Büchi automata (2ABTAs) and non-deterministic top-down tree automata (NTAs), both on *finite* trees (in contrast to Section 5.2). A finite tree T is *m-ary* if, for any $x \in T$, the set $\{i \mid x \cdot i \in T\}$ is of cardinality zero or exactly m. 2ABTAs on finite trees are defined exactly like 2APTAs on infinite trees except that

- the acceptance condition now takes the form $F \subseteq Q$ and a run is accepting if, for every infinite path $y_1 y_2 \cdots$, the set $\{i \mid r(y_i) = (x, q) \text{ with } q \in F\}$ is infinite;
- we allow a special transition leaf and add to the definition of a run r the condition that, for any node y of the input tree T, $r(y) = (x, \mathsf{leaf})$ implies that x is a leaf in T.

Note that runs can still be infinite.

Definition 49. 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 a finite alphabet, $Q_0 \subseteq Q$ a set of *initial states*, $\delta \colon Q \times \Gamma \to 2^{Q^m}$ a *transition function*, and $F \subseteq Q$ is a set of *final states*. Let (T, L) be a Γ-labelled m-ary tree. A *run* of $\mathfrak{A}$ on (T, L) is a Q-labelled m-ary tree (T, r) such that $r(\varepsilon) \in Q_0$ and $\langle r(x \cdot 1), \ldots, r(x \cdot m)\rangle \in \delta(r(x), L(x))$, for each node $x \in T$. The run is *accepting* if $r(x) \in F$, for every leaf x of T. The set of trees accepted by $\mathcal{A}$ is denoted by $L(\mathcal{A})$.

We use the following results from automata theory [59,65,66].

Theorem 50.

1. *Every 2ABTA $\mathfrak{A} = (Q, \Gamma, \delta, q_0, F)$ 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 (respectively, NTAs) $\mathfrak{A}_1, \dots, \mathfrak{A}_c$, one can construct in polynomial time a 2ABTA (respectively, an NTA) $\mathfrak{A}$ such that $L(\mathfrak{A}) = L(\mathfrak{A}_1) \cap \dots \cap L(\mathfrak{A}_c)$;*
3. *Emptiness of NTAs $\mathfrak{A} = (Q, \Gamma, Q_0, \delta, F)$ can be decided in polynomial time.*

Before proceeding further, we give a concrete definition of the canonical model for *Horn*$\mathcal{ALC}$ KBs that was mentioned in Proposition 8, tailored towards the constructions used in the rest of this section. Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a *Horn*$\mathcal{ALC}$ KB with $\mathcal{T}$ in normal form. We use $\mathsf{CN}(\mathcal{T})$ to denote the set of concept names in $\mathcal{T}$. For any $a \in \mathsf{ind}(\mathcal{A})$, we use $\mathsf{tp}_{\mathcal{K}}(a)$ to denote the set $\{A \in \mathsf{CN}(\mathcal{T}) \mid \mathcal{K} \models A(a)\}$. For $t \subseteq \mathsf{CN}(\mathcal{T})$, set $\mathsf{cl}_{\mathcal{T}}(t) = \{A \in \mathsf{CN}(\mathcal{T}) \mid \mathcal{T} \models \bigsqcap t \sqsubseteq A\}$. A set $S = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$ is a *successor set for* t if there is a concept name $A' \in t$ such that $A' \sqsubseteq \exists R.A \in \mathcal{T}$ and $\forall R.B_1, \dots, \forall R.B_n$ is the set of all concepts of this form such that, for some $B \in t$, we have $B \sqsubseteq \forall R.B_i \in \mathcal{T}$. Later on, we shall call S a Σ_2-*successor set* if $R \in \Sigma_2$. We use $S^{\downarrow}$ to denote the set $\{A, B_1, \dots, B_n\}$. A *path for* $\mathcal{K}$ is a sequence $aS_1 \cdots S_n$ such that $a \in \mathsf{ind}(\mathcal{A})$, S_1 is a successor set for $\mathsf{tp}_{\mathcal{K}}(a)$, and S_{i+1} is a successor set for $\mathsf{cl}_{\mathcal{T}}(S_i^{\downarrow})$, for $1 \le i < n$. Now, the *canonical model* $\mathcal{I}_{\mathcal{K}}$ *of* $\mathcal{K}$ is defined as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_{\mathcal{K}}} &= \mathsf{ind}(\mathcal{A}) \cup \{aS_1 \cdots S_n \mid aS_1 \cdots S_n \text{ path for } \mathcal{K}\},\\
A^{\mathcal{I}_{\mathcal{K}}} &= \{a \mid A \in \mathsf{tp}_{\mathcal{K}}(a)\} \cup \{aS_1 \cdots S_n \mid n \ge 1 \text{ and } A \in \mathsf{cl}_{\mathcal{T}}(S_n^{\downarrow})\},\\
R^{\mathcal{I}_{\mathcal{K}}} &= \{(a,b) \mid R(a,b) \in \mathcal{A}\} \cup \{(aS_1 \cdots S_{n-1}, aS_1 \cdots S_n) \mid R \text{ is the role name in } S_n\}.
\end{aligned}$$

The following result is standard:

Lemma 51. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a Horn$\mathcal{ALC}$ KB in normal form. Then $\mathcal{I}_{\mathcal{K}}$ is a model of $\mathcal{K}$ iff $\mathcal{K}$ is consistent iff there is no $a \in \mathsf{ind}(\mathcal{A})$ with $\mathcal{T} \models \mathsf{tp}_{\mathcal{K}}(a) \sqsubseteq \bot$.*

We now establish the upper bound in Theorem 48. Let $\mathcal{T}_1$ be an $\mathcal{ALC}$ TBox, $\mathcal{T}_2$ a *Horn*$\mathcal{ALC}$ TBox, and Σ_1, Σ_2 signatures. Set $m = |\mathcal{T}_2|$. We aim to construct an NTA $\mathfrak{A}$ such that a tree is accepted by $\mathfrak{A}$ iff this tree encodes a tree Σ_1-ABox $\mathcal{A}$ of outdegree at most m that is consistent with both $\mathcal{T}_1$ and $\mathcal{T}_2$ and a (part of 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 46, this means that $\mathfrak{A}$ accepts the empty language iff $\mathcal{T}_2$ is (Σ_1, Σ_2)-rCQ entailed by $\mathcal{T}_1$. To ensure that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not con-Σ_2-homomorphically embeddable into $\mathcal{I}_1$, we use the characterisation provided by Lemma 47. We first make precise which trees should be accepted by the NTA $\mathfrak{A}$ and then show how to construct $\mathfrak{A}$.

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 normal form and use $\mathsf{sub}(\mathcal{T}_2)$ 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, where a *role* is a role name R or its *inverse* R^-. Automata will run on m-ary Γ-labelled 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 a Γ-labelled tree (T, L) and a node x from T, we write $L_i(x)$ to denote the $i+1$st component of $L(x)$, for each $i \in \{0, \dots, 4\}$. Informally, the projection of a Γ-labelled tree to the

- L_0-components represents the tree 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 mark the individual a in $\mathcal{A}$ from Lemma 47;
- L_4-components contains bookkeeping information that helps to ensure that the individual marked by the L_3-component indeed satisfies one of the two conditions from Lemma 47.

By 'partial' we mean that the restriction of the respective model to individuals in $\mathcal{A}$ is represented whereas its 'anonymous' part is not. We now make these intuitions more precise by defining certain properness conditions for Γ-labelled trees, one for each component in the labels, which make sure that each component can indeed be meaningfully interpreted to represent what it is supposed to. A Γ-labelled tree (T, L) is *0-proper* if it satisfies the following conditions:

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

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

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

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

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 $C \in L_1(x)$ for all $C \in \mathsf{cl}(\mathcal{T}_1)$;
2. $A \in L_0(x)$ implies $A \in L_1(x)$;
3. if y is a child of x and $R \in L_0(y)$, then $\forall R.C \in L_1(x)$ implies $C \in L_1(y)$ for all $\forall R.C \in \mathsf{cl}(\mathcal{T}_1)$;
4. if y is a child of x and $R^- \in L_0(y)$, then $\forall R.C \in L_1(y)$ implies $C \in L_1(x)$ for all $\forall R.C \in \mathsf{cl}(\mathcal{T}_1)$.

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

1. $L_2(x) = \mathsf{tp}_{\mathcal{T}_2, \mathcal{A}_{(T,L)}}(x)$;
2. $\mathcal{T}_2 \not\models \bigsqcap L_2(x) \sqsubseteq \bot$.

It is *3-proper* if there is exactly one node x with $L_3(x) = 1$.

The canonical model $\mathcal{I}_{\mathcal{T}_2, \mathcal{S}}$ of $\mathcal{T}_2$ *and a finite set* $\mathcal{S} \subseteq \mathsf{sub}(\mathcal{T}_2)$ is the interpretation obtained from the canonical model of the KB that consists of the TBox $\mathcal{T}_2 \cup \{A_C \sqsubseteq C \mid C \in \mathcal{S}\}$ and the ABox $\{A_C(a_\varepsilon) \mid C \in \mathcal{S}\}$, with all fresh concept names A_C removed. A Γ-labelled tree (T, L) is *4-proper* if the following conditions hold, for $x_1, x_2 \in T$:

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 $L_4(x_1) = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$, 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 $(\mathcal{I}_{\mathcal{T}_2, \{A, B_1, \dots, B_n\}}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}, e)$ for all $(d, e) \in R^{\mathcal{I}}$;
3. if x_2 is a child of x_1, $L_0(x_2)$ contains the role name R, and $L_4(x_1) = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$, then there is a Σ_2-concept name in $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \dots, B_n\}) \setminus L_1(x_2)$ or $L_4(x_2)$ is a Σ_2-successor set for $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \dots, B_n\})$;
4. if x_2 is a child of x_1, $L_0(x_2)$ contains the role R^-, and $L_4(x_2) = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$, then there is a Σ_2-concept name in $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \dots, B_n\}) \setminus L_1(x_1)$ or $L_4(x_1)$ is a Σ_2-successor set for $\mathsf{cl}_{\mathcal{T}_2}(\{A, B_1, \dots, B_n\})$.

For $L_4(x) = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$, this expresses the obligation that $(\mathcal{I}_{\mathcal{T}_2, \{A, B_1, \dots, B_n\}}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}, e)$, for $(d, e) \in R^{\mathcal{I}}$, where $\mathcal{I}$ is the interpretation that is (partly) represented by the L_1-components of the labels in (T, L); see the proof of Lemma 52 for a precise definition of $\mathcal{I}$. With this in mind, note how 4-properness addresses (1) and (2) of Lemma 47. In fact, Condition 1 of 4-properness decides whether (1) or (2) is satisfied. If (2) is satisfied, which says that there is an R-successor d of x_1 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 x_1 in $\mathcal{I}$, we have $(\mathcal{I}_{\mathcal{T}_2, \mathcal{A}}, d) \not\lesssim_{\Sigma_2} (\mathcal{I}, e)$, then the role name R and the element d are represented by the successor set stored in $L_4(x_1)$. In fact, that element is $d = x_1 L_4(x_1)$, see the definition of canonical models. The remaining conditions of 4-properness implement the obligations represented by the L_4-components of node labels.

Lemma 52. *There is an m-ary Γ-labelled tree that is i-proper for all $i \in \{0, \dots, 4\}$ iff there are a tree Σ_1-ABox $\mathcal{A}$ of outdegree at most m that is consistent with $\mathcal{T}_1$ 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. (⇒) Let (T, L) be an m-ary Γ-labelled tree that is i-proper for all $i \in \{0, \dots, 4\}$. Then $\mathcal{A}_{(T,L)}$ is a tree Σ_1-ABox of outdegree at most m. Moreover, $\mathcal{A}_{(T,L)}$ is consistent with $\mathcal{T}_2$, by 2-properness and Lemma 51.

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, such an $\mathcal{I}_1$ witnesses consistency of $\mathcal{A}_{(T,L)}$ with $\mathcal{T}_1$ and, moreover, by the definition of simulations, $\mathcal{I}_1$ must satisfy one of (1) or (2) of Lemma 47 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 $\mathcal{I}_1$.

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

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

Then take, for each $x \in T$, a model $\mathcal{I}_x$ of $\mathcal{T}_1$ such that $x \in C^{\mathcal{I}_x}$ iff $C \in L_1(x)$ for all $C \in \mathsf{cl}(\mathcal{T}_1)$, which exists by Condition 1 of 1-properness. Moreover, if $L_4(x) = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$, then choose $\mathcal{I}_x$ such that $(\mathcal{I}_{\mathcal{T}_2, \{A, B_1, \dots, B_n\}}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}_x, y)$ for all $(x, y) \in R^{\mathcal{I}_x}$, which is possible by Condition 2 of 4-properness. Further, suppose $\Delta^{\mathcal{I}_0}$ 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 straightforward to prove that $\mathcal{I}_1$ is indeed a model of $(\mathcal{T}_1, \mathcal{A}_{(T,L)})$.

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

Claim. For all $x \in T$, if $L_4(x) = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$, then $(\mathcal{I}_{\mathcal{T}_2,\{A,B_1,\dots,B_n\}}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}_1, y)$ for all $(x, y) \in R^{\mathcal{I}_1}$.

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 $R_1(x, x_1), \dots, R_n(x_{n-1}, x_n)$ in $\mathcal{A}_{(T,L)}$. It uses Conditions 2 to 4 of 4-properness.

($\Leftarrow$) Let $\mathcal{A}$ be a tree Σ_1-ABox of outdegree at most m that is consistent with $\mathcal{T}_1$ 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 without loss of generality 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 47, 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 $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d_0) \not\lesssim_{\Sigma_2} (\mathcal{I}_1, d)$.

We now show how to construct from $\mathcal{A}$ a Γ-labelled tree (T, L) that is i-proper for all $i \in \{0, \dots, 4\}$. For each $a \in \mathsf{ind}(\mathcal{A})$, set $\mathsf{R}(a) = \emptyset$ if $a = \varepsilon$, and otherwise set $\mathsf{R}(a) = \{R\}$ if $R(b, a) \in \mathcal{A}$ and $a = b \cdot c$, for some $c \in \mathbb{N}$, and $\mathsf{R}(a) = \{R^-\}$ if $R(a, b) \in \mathcal{A}$ and $a = b \cdot c$, for some $c \in \mathbb{N}$. Now set

$$\begin{aligned} T &= \mathsf{ind}(\mathcal{A}), \\ L_0(x) &= \{A \mid A(x) \in \mathcal{A}\} \cup \{\mathsf{R}(x)\}, \\ L_1(x) &= \{C \in \mathsf{cl}(\mathcal{T}_1) \mid x \in C^{\mathcal{I}_1}\}, \\ L_2(x) &= \mathsf{tp}_{\mathcal{T}_2,\mathcal{A}}(x), \\ L_3(x) &= \begin{cases} 1 & \text{if } x = a_0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

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

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

such that the following conditions are satisfied:

(i) if $(a, R, d) \in \Omega$, then $L_4(a)$ has the form $\{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$ and the restriction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to the subtree-interpretation rooted at d is the canonical model $\mathcal{I}_{\mathcal{T}_2,\{A,B_1,\dots,B_n\}}$;
(ii) if $(a, R, d) \in \Omega$ and d' is an R-successor of a in $\mathcal{I}_1$, then $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d) \not\lesssim_{\Sigma_2} (\mathcal{I}_1, d')$.

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

Since we are only following role names (but not inverse roles) during the modification of L_4 and since $\mathcal{A}$ is tree-shaped, we shall never process tuples $(a_1, R_1, d_1), (a_2, R_2, d_2)$ from Ω 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-labelled tree (T, L) just constructed is i-proper for all $i \in \{0, \dots, 4\}$. The most interesting point is 4-properness, which consists of four conditions. Condition 1 is satisfied by the construction of L_4. Condition 2 is satisfied by (ii), and Conditions 3 and 4 again by the construction of L_4. $\square$

By Theorem 46 and Lemma 52, we can decide whether $\mathcal{T}_1$ does (Σ_1, Σ_2)-rCQ entail $\mathcal{T}_2$ by checking whether there is no Γ-labelled tree that is i-proper for each $i \in \{0, \dots, 4\}$. We do this by constructing automata $\mathcal{A}_0, \dots, \mathcal{A}_4$ such that each $\mathcal{A}_i$ accepts exactly the Γ-labelled 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). 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

- an NTA $\mathfrak{A}_0$ that checks 0-properness and has constantly many states;
- 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;
- an NTA $\mathfrak{A}_3$ that checks 3-properness and has constantly many states.

It thus remains to construct

- a 2ABTA $\mathfrak{A}_2$ that checks 2-properness and whose number of states is polynomial in $|\mathcal{T}_2|$;
- an NTA $\mathfrak{A}_4$ that checks 4-properness and whose number of states is (single) exponential in $|\mathcal{T}_2|$.

In fact, the reason for mixing 2ABTAs and NTAs is that while $\mathfrak{A}_2$ is easier to construct as a 2ABTA, there is no obvious way to construct $\mathfrak{A}_4$ as a 2ABTA with only polynomially many states: it seems 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) $(\mathcal{T}_2, \mathcal{A}_{(T,L)}) \models A(x)$ implies $A \in L_2(x)$ for all $x \in T$ and $A \in \mathsf{CN}(\mathcal{T}_2)$;
(ii) $\mathcal{T}_2 \not\models \bigsqcap L_2(x) \sqsubseteq \bot$.

Note that, by Lemma 51, (i) and (ii) imply that $\mathcal{A}_{(T,L)}$ is consistent with $\mathcal{T}_2$. It is easy for a 2ABTA to verify (ii), alternatively one can simply refine Γ. To achieve (i), it suffices to guarantee the following conditions, for $x_1, x_2 \in T$:

- $A \in L_0(x_1)$ implies $A \in L_2(x_1)$;
- if $A_1, \dots, A_n \in L_2(x_1)$ and $\mathcal{T}_2 \models A_1 \sqcap \dots \sqcap A_n \sqsubseteq A$, then $A \in L_2(x_1)$;
- 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 $B \in L_2(x_2)$;
- 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 $B \in L_2(x_1)$;
- 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 $B \in L_2(x_1)$;
- 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 which are easily verified with a 2ABTA. 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 characterise 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 $\mathsf{ind}(\mathcal{A}) \times \mathsf{CN}(\mathcal{T}_2)$-labelled 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 $y_1, \dots, y_n$ with $V(y_i) = (a, A_i)$, for $1 \le i \le n$, and $\mathcal{T}_2 \models A_1 \sqcap \dots \sqcap A_n \sqsubseteq A$;
  - 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 53. *If $(\mathcal{T}_2, \mathcal{A}) \models A(a)$ and $\mathcal{A}$ is consistent with $\mathcal{T}_2$, 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 53 is based on the chase procedure, details can be found in [67].) We are now ready to construct the 2ABTA $\mathfrak{A}_{2,2}$. Since $\mathfrak{A}_{2,1}$ ensures that $\mathcal{A}_{(T,L)}$ is consistent with $\mathcal{T}_2$, by Lemma 53 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 $\mathfrak{A}_2 = (Q, \Gamma^-, \delta, q_0, F)$, where $Q = \{q_0\} \uplus \{q_A \mid A \in \mathsf{CN}(\mathcal{T}_2)\} \uplus \{q_{A,R}, q_R \mid A \in \mathsf{CN}(\mathcal{T}_2), R \in \mathsf{rol}(\mathcal{T}_2)\}$ and $F = \emptyset$ (i.e., exactly the finite runs are accepting). For all $(\sigma_0, \sigma_2) \in \Gamma^-$, set

$$
\begin{aligned}
\delta(q_0,(\sigma_0,\sigma_2)) &= \bigwedge_{A\in\sigma_2}(0,q_A)\wedge(\mathsf{leaf}\vee\bigwedge_{i\in 1..m}(i,q_0)),\\
\delta(q_A,(\sigma_0,\sigma_2)) &= \mathsf{true}, && \text{whenever } A\in\sigma_0 \text{ or } \top\sqsubseteq A\in\mathcal{T}_2,\\
\delta(q_A,(\sigma_0,\sigma_2)) &= \bigvee_{\mathcal{T}_2\models A_1\sqcap\dots\sqcap A_n\sqsubseteq A}((0,q_{A_1})\wedge\dots\wedge(0,q_{A_n}))\vee && \text{whenever } A\notin\sigma_0 \text{ and } \top\sqsubseteq A\notin\mathcal{T}_2,\\
&\quad \bigvee_{\exists R.B\sqsubseteq A\in\mathcal{T},\ R\in\Sigma_1}(((0,q_{R^-})\wedge(-1,q_B))\vee\bigvee_{i\in 1..m}(i,q_{B,R}))\vee\\
&\quad \bigvee_{B\sqsubseteq\forall R.A\in\mathcal{T},\ R\in\Sigma_1}((0,q_R)\wedge(-1,q_B))\vee\bigvee_{i\in 1..m}(i,q_{B,R^-})),\\
\delta(q_{A,R},(\sigma_0,\sigma_2)) &= (0,q_A), && \text{whenever } R\in\sigma_0,\\
\delta(q_{A,R},(\sigma_0,\sigma_2)) &= \mathsf{false}, && \text{whenever } R\notin\sigma_0,\\
\delta(q_R,(\sigma_0,\sigma_2)) &= \mathsf{true}, && \text{whenever } R\in\sigma_0,\\
\delta(q_R,(\sigma_0,\sigma_2)) &= \mathsf{false}, && \text{whenever } R\notin\sigma_0.
\end{aligned}
$$

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

We next discuss the construction of the NTA $\mathfrak{A}_4$, omitting most of the details because the construction is not difficult. Conditions 1 and 2 of 4-properness can be enforced by making sure that certain symbols from Γ do not occur. However, in the case of Condition 2, we have to decide during the automaton construction whether, for given sets $S_1\subseteq \mathsf{cl}(\mathcal{T}_1)$ and $S_2=\{\exists R_0.A,\forall R_0.B_1,\dots,\forall R_0.B_n\}\subseteq \mathsf{sub}(\mathcal{T}_2)$, there is a model $\mathcal{I}$ of $\mathcal{T}_1$ and a $d\in\Delta^{\mathcal{I}}$ such that

(a) $d\in C^{\mathcal{I}}$ iff $C\in S_1$ for all $C\in\mathsf{cl}(\mathcal{T}_1)$ and
(b) $(\mathcal{I}_{\mathcal{T}_2,S_2^{\downarrow}},a_\varepsilon)\not\leq_{\Sigma_2}(\mathcal{I},e)$ for all $(d,e)\in R_0^{\mathcal{I}}$.

We have to show that this check can be done in ExpTime. We give a sketch of a decision procedure based on nondeterministic Büchi automata on infinite trees that borrows ideas from the above constructions, but is much simpler.

Definition 54. A *nondeterministic Büchi tree automaton* (NBA) on infinite m-ary trees is a tuple $\mathfrak{A}=(Q,\Gamma,Q_0,\delta,F)$ where Q is a finite set of *states*, Γ a finite alphabet, $Q_0\subseteq Q$ a set of *initial states*, $\delta\colon Q\times\Gamma\to 2^{Q^m}$ a *transition function*, and $F\subseteq Q$ is an *acceptance condition*. Let (T,L) be a Γ-labelled m-ary tree. A *run* of $\mathfrak{A}$ on (T,L) is a Q-labelled m-ary tree (T,r) such that $r(\varepsilon)\in Q_0$ and $\langle r(x\cdot 1),\dots,r(x\cdot m)\rangle\in\delta(r(x),L(x))$, for each $x\in T$. We say that (T,r) is *accepting* if in all infinite paths $y_1y_2\cdots$ of T, the set $\{i\mid r(y_i)\in F\}$ is infinite. An infinite Γ-labelled tree (T,L) is *accepted* by $\mathfrak{A}$ if there is an accepting run of $\mathfrak{A}$ on (T,L). We use $\mathcal{L}(\mathfrak{A})$ to denote the set of all infinite Γ-labelled trees accepted by $\mathfrak{A}$.

The emptiness problem for NBAs can be solved in polynomial time. Our aim is to build an NBA $\mathfrak{B}$ such that the labelled trees accepted by $\mathfrak{B}$ represent tree interpretations $\mathcal{I}$ that satisfy Conditions (a) and (b). We make precise which trees should be accepted by $\mathfrak{B}$. Let Γ_0' be the set of all subsets of $\mathsf{cl}(\mathcal{T}_1)\cup\{R\in\mathsf{N_R}\mid R \text{ occurs in } \mathcal{T}_1\}$ that contain at most one role name and let $\Gamma'=(\Gamma_0'\times 2^{\mathsf{sub}(\mathcal{T}_2)})\cup\{\mathit{empty}\}$. For a Γ'-labelled tree (T,L) and a node x in T with $L(x)\neq\mathit{empty}$, we write $L_i(x)$ to denote the $i+1$st component of $L(x)$, for $i\in\{0,1\}$. Informally, the projection of a Γ'-labelled tree to the L_0-components represents $\mathcal{I}$ and the projection to the L_1-components contains bookkeeping information that helps to ensure Condition (b). A Γ'-labelled tree is *proper* if the following conditions hold, for $x_1,x_2\in T$:

- $L(\varepsilon)=(S_1,S_2)$;
- if $L(x_1)\neq\mathit{empty}$, then $L_0(x_1)$ is satisfiable with $\mathcal{T}_1$;
- if x_2 is a child of x_1 and $R\in L_0(x_2)$, then $\forall R.C\in L_0(x_1)$ implies $C\in L_0(x_2)$ for all $\forall R.C\in\mathsf{cl}(\mathcal{T}_1)$;
- if $\exists R.C\in L_0(x_1)$, then there is a child x_2 of x_1 such that $\{R,C\}\subseteq L_0(x_2)$;
- if x_2 is a child of x_1 and $L(x_1)=\mathit{empty}$, then $L(x_2)=\mathit{empty}$;
- if x_2 is a child of x_1, $L_0(x_2)$ contains the role name R, and $L_1(x_1)=\{\exists R.A,\forall R.B_1,\dots,\forall R.B_n\}$, then there is a Σ_2-concept name in $\mathsf{cl}_{\mathcal{T}_2}(\{A,B_1,\dots,B_n\})\setminus L_0(x_2)$ or $L_1(x_2)$ is a Σ_2-successor set for $\mathsf{cl}_{\mathcal{T}_2}(\{A,B_1,\dots,B_n\})$;
- there are only finitely many nodes x with $L_1(x)\neq\emptyset$.

In the conditions above, we assume that whenever a condition is posed on a component of the label of a node x, then $L(x)\neq\mathit{empty}$. Note that the L_1-component of a node label plays the same role as the L_4-component in the previous construction. Every proper Γ'-labelled tree (T,L) represents the following tree interpretation $\mathcal{I}_{(T,L)}$:

$$
\begin{aligned}
\Delta^{\mathcal{I}_{(T,L)}} &= \{x\in T\mid L(x)\neq\mathit{empty}\},\\
A^{\mathcal{I}_{(T,L)}} &= \{x\mid A\in L_0(x)\},\\
R^{\mathcal{I}_{(T,L)}} &= \{(x_1,x_2)\mid x_2 \text{ child of } x_1 \text{ and } R\in L_0(x_2)\}.
\end{aligned}
$$

Set $m'=|\mathcal{T}_1|$. The proof of the following lemma is similar to that of Lemma 52, but simpler.

Lemma 55. *There is an m'-ary proper Γ'-labelled tree (T, L) iff there is a model $\mathcal{I}$ of $\mathcal{T}_1$ and a $d \in \Delta^{\mathcal{I}}$ that satisfy Conditions (a) and (b) from before Definition 54; in fact, $\mathcal{I}_{(T,L)}$ is such a model.*

It is now straightforward to construct an NBA $\mathfrak{B}$ whose number of states is polynomial in $|\mathcal{T}_1|$ and exponential in $|\mathcal{T}_2|$ and which accepts exactly the m'-ary proper Γ'-labelled trees. Details are left to the reader.

8.2. 2ExpTime upper bound for Θ-CQ-entailment of Horn$\mathcal{ALC}$ TBoxes by $\mathcal{ALC}$ TBoxes

We now consider the case of non-rooted CQs. Our aim is to prove the following 2ExpTime upper bound:

Theorem 56. *Θ-CQ entailment of Horn$\mathcal{ALC}$ TBoxes by $\mathcal{ALC}$ TBoxes is in 2ExpTime.*

The proof again builds on the characterisations provided by Theorem 46. Since we are now working with CQs rather than rCQs, we have to consider Σ_2-homomorphic embeddability instead of con-Σ_2-homomorphic embeddability. Note that Lemma 47 also provides a characterisation in terms of simulations in that case, adding a third condition. We modify the previous construction to accommodate this additional condition.

Condition (2) of Lemma 47 tells us to avoid certain simulations. In the previous construction, we were able to do that by storing a single successor set in the L_4-component of each Γ-label, that is, it was sufficient to avoid at most one simulation into each individual of the ABox $\mathcal{A}_{(T,L)}$. In the current construction, this is no longer the case. We thus let the L_4-component of Γ-labels range over $2^{2^{\mathsf{sub}(\mathcal{T}_2)}}$ rather than $2^{\mathsf{sub}(\mathcal{T}_2)}$ and use it to store *sets of* successor sets. To address (3) in Lemma 47, we add an L_5-component to Γ-labels, which also ranges over $2^{2^{\mathsf{sub}(\mathcal{T}_2)}}$. The purpose of this component is to represent elements of the canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ from which we have to avoid a simulation into *any* individual in $\mathcal{A}_{(T,L)}$ and, in fact, into any element of the interpretation (partially) represented by the L_2-components of node labels. 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 define a notion of Σ_2-descendant set. While a Σ_2-successor set for $t \subseteq \mathsf{CN}(\mathcal{T}_2)$ represents a Σ_2-successor of an element d in a canonical model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ that satisfies $d \in A^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}}$ for all $A \in t$, a Σ_2-descendent set represents a *descendent* of such a d that is attached to its predecessor via a role name that is *not* in Σ_2, as in (3) of Lemma 47. Formally, for $t \subseteq \mathsf{CN}(\mathcal{T}_2)$, we define Γ_t to be the smallest set such that $t \in \Gamma_t$ and if $t' \in \Gamma_t$ and S is a successor set for $\mathsf{cl}_{\mathcal{T}_2}(t')$, then $S^{\downarrow} \in \Gamma_t$. A set $s \subseteq \mathsf{CN}(\mathcal{T}_2)$ is a Σ_2-*descendant set* for t if there is a $t' \in \Gamma_t$ and successor set $S = \{\exists R.A, \forall R.B_1, \dots, \forall R.B_n\}$ for $\mathsf{cl}_{\mathcal{T}_2}(t')$ with $R \notin \Sigma_2$ such that $s = S^{\downarrow}$.

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

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

A Γ-labelled tree (T, L) is *5-proper* if the following conditions are satisfied for all $x_1 \in T$:

- all $x \in T$ agree regarding their L_5-label;
- if $s \in L_5(x_1)$, then one of the following holds:
  - there is a Σ_2-concept name in $s \setminus L_1(x_1)$;
  - $L_4(x_1)$ contains a Σ_2-successor set for s.

Note that 4-properness and 5-properness together implement (2) and (3) of Lemma 47; in particular, Point (3) from Lemma 47 requires that $(\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}, d_0) \not\leq_{\Sigma_2} (\mathcal{I}_1, e)$ for any element e of $\mathcal{I}_1$ which can be broken down into the two cases above.

The proof of the following lemma is similar to that of Lemma 53:

Lemma 57. *There is an m-ary Γ-labelled tree that is i-proper for all $i \in \{0, \dots, 5\}$ iff there is a tree Σ_1-ABox $\mathcal{A}$ of outdegree at most m that is consistent with $\mathcal{T}_1$ 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 $\mathcal{I}_1$.*

We can now adapt the automata construction presented in the previous section. It is straightforward to construct an NTA $\mathfrak{A}_5$ with double exponentially many states that verifies 5-properness. Also, the NTA $\mathfrak{A}_4$ for 4-properness will now have double exponentially many states because L_4- and L_5-components are sets of sets of concepts rather than sets of concepts. In fact, we could dispense with NTAs altogether and use a 2ABTA that has exponentially many states, both for $\mathfrak{A}_4$ and $\mathfrak{A}_5$. The construction of $\mathfrak{A}_4$ needs to decide whether, for given sets $S_1 \subseteq \mathsf{cl}(\mathcal{T}_1)$ and $S_2, S_3 \subseteq 2^{\mathsf{CN}(\mathcal{T}_2)}$, there is a model $\mathcal{I}$ of $\mathcal{T}_1$ and a $d \in \Delta^{\mathcal{I}}$ such that

(a) $d \in C^{\mathcal{I}}$ iff $C \in S_1$, for all $C \in \mathsf{cl}(\mathcal{T}_1)$;
(b) $(\mathcal{I}_{\mathcal{T}_2,S}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}, d)$ for all $S \in S_2$;
(c) $(\mathcal{I}_{\mathcal{T}_2,S}, a_\varepsilon) \not\lesssim_{\Sigma_2} (\mathcal{I}, e)$ for all $S \in S_3$ and $e \in \Delta^{\mathcal{I}}$;

This check can be implemented in 2ExpTime using a decision procedure based on NBAs, mixing ideas from the corresponding construction in the previous section and the construction above. Overall, we obtain the 2ExpTime upper bound stated in Theorem 56.

8.3. 2ExpTime *lower bound for* Θ*-CQ-inseparability between Horn*$\mathcal{ALC}$ *TBoxes*

We prove a matching lower bound for the 2ExpTime upper bound established in Theorem 56 using a reduction of the word problem of exponentially space bounded ATMs (see Section 5.3). More precisely, we show the following:

Theorem 58. *(Σ, Σ)-CQ inseparability between the empty TBox and Horn$\mathcal{ALC}$ TBoxes is 2ExpTime-hard.*

Note that we obtain a 2ExpTime lower bound for Θ-CQ entailment as well since, clearly, the empty TBox (Σ, Σ)-CQ-entails a TBox $\mathcal{T}$ iff the empty TBox and $\mathcal{T}$ are (Σ, Σ)-CQ-inseparable. Let $M = (Q, \Gamma_I, \Gamma, q_0, \Delta)$ be an exponentially space bounded ATM whose word problem is 2ExpTime-hard, where Q is the finite set of *states*, Γ_I the *input alphabet*, $\Gamma \supseteq \Gamma_I$ the *tape alphabet* with *blank symbol* $\square \in \Gamma \setminus \Gamma_I$, $q_0 \in Q$ the *initial state*, and $\Delta \subseteq Q \times \Gamma \times Q \times \Gamma \times \{L, R\}$ the *transition relation*. We use $\Delta(q, \sigma)$ to denote the set of *transitions* $(q', \sigma', D) \in Q \times \Gamma \times \{L, R\}$ possible when M is in state q and reads σ, that is, $(q, \sigma, q', \sigma', D) \in \Delta$. 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 [60]). To simplify the reduction, we may also assume without loss of generality that M makes at least one step on every input, that it never reaches the last tape cell, and that every universal configuration has exactly two successor configurations.

Note that when M accepts an input w, this is witnessed by an *accepting computation tree* whose nodes are labelled with configurations such that the root is labelled with the initial configuration of M on w, the descendants of any non-leaf labelled with a universal (respectively, existential) configuration include all (respectively, one) of the successors of that configuration, and all leafs are labelled with accepting configurations.

Let w be an input to M. We aim to construct a *Horn$\mathcal{ALC}$* TBox $\mathcal{T}$ and a signature Σ such that M accepts w iff there is a tree Σ-ABox $\mathcal{A}$ such that

(a) $\mathcal{A}$ is consistent with $\mathcal{T}$ and
(b) $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ is not Σ-homomorphically embeddable into $\mathcal{I}_{\mathcal{T}_\emptyset,\mathcal{A}}$,

where $\mathcal{T}_\emptyset = \emptyset$. Note that this is equivalent to (Σ, Σ)-CQ-entailment of $\mathcal{T}$ by $\mathcal{T}_\emptyset$ due to Theorem 46 (2); that theorem additionally imposes a restriction on the outdegree of $\mathcal{A}$, but it is easy to go through the proofs and verify that the characterisation holds also without that restriction. We are going to construct $\mathcal{T}$ and Σ such that $\mathcal{A}$ represents an accepting computation tree of M on w.

When dealing with an input w of length n, in $\mathcal{A}$ 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 Fig. 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 shall see at the very end of the reduction why successor configurations are separated by two consecutive edges instead of a single one.

The above description is 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 (b) 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 enforced locally inside configuration sequences.

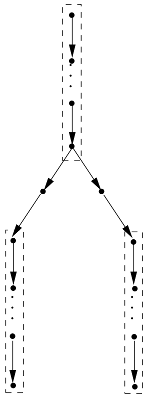


Fig. 7. Configuration tree (partial).

The signature Σ consists of the following symbols:

- the concept names $A_0, \ldots, A_{n-1}, \overline{A}_0, \ldots, \overline{A}_{n-1}$ 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, $\overline{A}_0$ are the lowest bit);
- the concept names $A'_0, \ldots, A'_{m-1}$ and $\overline{A}'_0, \ldots, \overline{A}'_{m-1}$, where $m = \lceil \log(2^n + 2) \rceil$, that serve as bits of another counter which is able to count from 0 to $2^n + 2$ and whose purpose will be explained later;
- the concept names A_σ, A'_σ, $\overline{A}_\sigma$, for each $\sigma \in \Gamma$;
- the concept names $A_{q,\sigma}$, $A'_{q,\sigma}$, $\overline{A}_{q,\sigma}$, for each $\sigma \in \Gamma$ and $q \in Q$;
- the concept names X_1, X_2 that mark the first and second successor configuration;
- the role name R.

From the above list, the 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 role of the concept names $\overline{A}_\sigma$ and $\overline{A}_{q,\sigma}$ will be explained later.

It thus remains to construct the TBox $\mathcal{T}$, which is the most laborious part of the reduction. We use $\mathcal{T}$ to verify the existence of a computation tree of M on input w in the ABox. For the time being, we are going to assume that $(*)$ holds and, in a second step, we will demonstrate how to actually achieve that. We start with verifying halting configurations, which must all be accepting in an accepting computation tree, in a bottom-up manner:

$$A_0 \sqcap \cdots \sqcap A_{n-1} \sqcap A_\sigma \sqcap A'_\sigma \sqsubseteq V, \tag{1}$$

$$A_i \sqcap \exists R.A_i \sqcap \bigsqcup_{j<i} \exists R.A_j \sqsubseteq \mathsf{ok}_i, \tag{2}$$

$$\overline{A}_i \sqcap \exists R.\overline{A}_i \sqcap \bigsqcup_{j<i} \exists R.A_j \sqsubseteq \mathsf{ok}_i, \tag{3}$$

$$A_i \sqcap \exists R.\overline{A}_i \sqcap \bigsqcap_{j<i} \exists R.\overline{A}_j \sqsubseteq \mathsf{ok}_i, \tag{4}$$

$$\overline{A}_i \sqcap \exists R.A_i \sqcap \bigsqcap_{j<i} \exists R.\overline{A}_j \sqsubseteq \mathsf{ok}_i, \tag{5}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V \sqcap A_\sigma \sqcap A'_\sigma \sqsubseteq V, \tag{6}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V \sqcap A_\sigma \sqcap A'_{q,\sigma'} \sqsubseteq V_{L,\sigma}, \tag{7}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V \sqcap A_{q_a,\sigma} \sqcap A'_\sigma \sqsubseteq V_{R,q_a}, \tag{8}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V_{L,\sigma} \sqcap A_{q_a,\sigma'} \sqcap A'_{\sigma'} \sqsubseteq V_{L,q_a,\sigma}, \tag{9}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V_{R,q_a} \sqcap A_\sigma \sqcap A'_{q,\sigma'} \sqsubseteq V_{R,q_a,\sigma}, \tag{10}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V_{D,q_a,\sigma} \sqcap A_{\sigma'} \sqcap A'_{\sigma'} \sqsubseteq V_{D,q_a,\sigma}, \tag{11}$$

$$\exists R.A_i \sqcap \exists R.\overline{A}_i \sqsubseteq \bot, \tag{12}$$

where σ, σ' range over Γ, q over Q, i over $0, \ldots, n-1$, and D over $\{L, R\}$. The first line starts the verification at the last tape cell, ensuring that at least one concept name A_σ and one concept name A'_σ is true (it also verifies that the symbol is identical in the current and previous configuration, assuming $(*)$; it is here that the assumption that M never reaches the last tape cell makes the construction easier). The following lines implement the verification of the remaining tape cells of the configuration. Lines (2)–(5) implement decrementation of a binary counter and the conjunct $\overline{A}_i$ in lines (6)–(11) prevents the counter from wrapping around once it has reached 0. We use several kinds of verification markers:

– with V, we indicate that we have not yet seen the head of the ATM;
– $V_{L,\sigma}$ indicates that the ATM made a step to the left to reach the current configuration, writing σ;
– $V_{R,q}$ indicates that the ATM made a step to the right to reach the current configuration, switching to state q;
– $V_{D,q,\sigma}$ indicates that the ATM moved in direction D to reach the current configuration, switching to state q and writing σ.

In the remaining reduction, we expect that a marker $V_{D,q,\sigma}$ has been derived at the first (thus top-most) cell of the configuration. This makes sure that there is exactly one head in the current and previous configuration, and that the head moved exactly one step between the previous and current position. Also note that the above CIs ensure that the tape content does not change for cells that were not under the head in the previous configuration, assuming (∗). Note that it is not immediately clear that lines (2)–(11) 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:

$$A_i \sqcap \overline{A}_i \sqsubseteq \bot, \tag{13}$$

$$A_{\sigma_1} \sqcap A_{\sigma_2} \sqsubseteq \bot, \qquad \text{if } \sigma_1 \neq \sigma_2, \tag{14}$$

$$A_{\sigma_1} \sqcap A_{q_2,\sigma_2} \sqsubseteq \bot, \tag{15}$$

$$A_{q_1,\sigma_1} \sqcap A_{q_2,\sigma_2} \sqsubseteq \bot, \qquad \text{if } (q_1,\sigma_1) \neq (q_2,\sigma_2), \tag{16}$$

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

$$\exists R.\exists R.(X_1 \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap (V_{D,q,\sigma} \sqcup V'_{D,q,\sigma})) \sqsubseteq L\mathsf{ok}, \tag{17}$$

$$\exists R.\exists R.(X_2 \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap (V_{D,q,\sigma} \sqcup V'_{D,q,\sigma})) \sqsubseteq R\mathsf{ok}, \tag{18}$$

$$A_0 \sqcap \cdots \sqcap A_{n-1} \sqcap A_\sigma \sqcap A'_\sigma \sqcap L\mathsf{ok} \sqcap R\mathsf{ok} \sqsubseteq V', \tag{19}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_\sigma \sqcap A'_\sigma \sqsubseteq V', \tag{20}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_\sigma \sqcap A'_{q,\sigma'} \sqsubseteq V'_{L,\sigma}, \tag{21}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V'_{R,q} \sqcap A_\sigma \sqcap A'_{q',\sigma'} \sqsubseteq V'_{R,q,\sigma}, \tag{22}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V'_{D,q,\sigma} \sqcap A_{\sigma'} \sqcap A'_{\sigma'} \sqsubseteq V'_{D,q,\sigma}, \tag{23}$$

where $\sigma, \sigma', \sigma''$ range over Γ, q and q' over Q, i over $0,\dots,n-1$, and D over $\{L,R\}$. We switch to different verification markers V', $V'_{L,\sigma}$, $V'_{R,q}$, $V'_{D,q,\sigma}$ to distinguish between halting and non-halting configurations. Note that the first verification step is different for non-halting configurations: we expect to see one successor marked with X_1 and one with X_2, 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 CIs 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:

$$\exists R.\exists R.(X_t \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap V_{q,\sigma,D'}) \sqsubseteq S^t_{q,\sigma,D'}, \tag{24}$$

$$\exists R.\exists R.(X_t \sqcap \overline{A}_0 \sqcap \cdots \sqcap \overline{A}_{n-1} \sqcap V'_{q,\sigma,D'}) \sqsubseteq S^t_{q,\sigma,D'}, \tag{25}$$

$$\exists R.(A_\sigma \sqcap S^t_{q,\sigma',D}) \sqsubseteq S^t_{q,\sigma',D}, \tag{26}$$

where q ranges over Q, σ and σ' over Γ, t over $\{1,2\}$, and i over $0,\dots,n-1$. Note that markers are propagated up exactly to the head position. One issue with the above is that additional $S^t_{q,\sigma,D}$-markers could be propagated up not from the successors that we have verified, but from surplus (unverified) successors. To prevent such undesired markers, we add the CIs

$$S^t_{q_1,\sigma_1,D_1} \sqcap S^t_{q_2,\sigma_2,D_2} \sqsubseteq \bot \tag{27}$$

for all $t \in \{1,2\}$ and all distinct $(q_1,\sigma_1,D_1),(q_2,\sigma_2,D_2) \in Q \times \Gamma \times \{L,R\}$. We can now implement the verification of the cells under the head in non-halting configurations. We take

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^1_{q_2,\sigma_2,D_2} \sqcap S^2_{q_3,\sigma_3,D_3} \sqsubseteq V'_{R,q_1}, \tag{28}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V'_{L,\sigma} \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^1_{q_2,\sigma_2,D_2} \sqcap S^2_{q_3,\sigma_3,D_3} \sqsubseteq V'_{L,q_1,\sigma}, \tag{29}$$

for all $(q_1, \sigma_1) \in Q \times \Gamma$ with q_1 a universal state and $\Delta(q_1, \sigma_1) = \{(q_2, \sigma_2, D_2), (q_3, \sigma_3, D_3)\}$, i from $0, \ldots, n-1$, and σ from Γ; moreover, we take

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V' \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^1_{q_2,\sigma_2,D_2} \sqcap S^2_{q_2,\sigma_2,D_2} \sqsubseteq V'_{R,q_1}, \tag{30}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap \overline{A}_i \sqcap \exists R.V'_{L,\sigma} \sqcap A_{q_1,\sigma_1} \sqcap A'_{\sigma_1} \sqcap S^1_{q_2,\sigma_2,D_2} \sqcap S^2_{q_2,\sigma_2,D_2} \sqsubseteq V'_{L,q_1,\sigma}, \tag{31}$$

for all $(q_1, \sigma_1) \in Q \times \Gamma$ with q_1 an existential state, for all $(q_2, \sigma_2, D_2) \in \Delta(q_1, \sigma_1)$, all i from $0, \ldots, n-1$, and all σ from Γ. It remains to verify the initial configuration. Let $w = \sigma_0 \cdots \sigma_{n-1}$, let $(C = j)$ be the conjunction over the concept names A_i, $\overline{A}_i$ that expresses j in binary, for $0 \leq j < n$, and let $(C \geq n)$ be the Boolean concept over the concept names A_i, $\overline{A}_i$ expressing that the counter value is at least n. Then we take

$$A_0 \sqcap \cdots \sqcap A_{n-1} \sqcap A_{\Box} \sqcap L\mathsf{ok} \sqcap R\mathsf{ok} \sqsubseteq V^I, \tag{32}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap (C \geq n) \sqcap \exists R.V^I \sqcap A_{\Box} \sqsubseteq V^I, \tag{33}$$

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap (C = i) \sqcap \exists R.V^I \sqcap A_{\sigma_i} \sqsubseteq V^I, \tag{34}$$

where i ranges over $1, \ldots, n-1$ 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

$$\mathsf{ok}_0 \sqcap \cdots \sqcap \mathsf{ok}_{n-1} \sqcap (C = 0) \sqcap \exists R.V^I \sqcap A_{q_0,\sigma_0} \sqcap S^1_{q,\sigma,D} \sqcap S^2_{q,\sigma,D} \sqsubseteq I \tag{35}$$

for all $(q, \sigma, D) \in \Delta(q_0, \sigma_0)$.

At this point, we have finished the verification of the computation tree, except that we have assumed but not yet established $(*)$. Achieving $(*)$ consists of two parts. In the first part, we use the concept names B_i, $\overline{B}_i$, $i < m$ (recall that $m = \lceil \log(2^n + 2) \rceil$) to implement an additional counter that serves the purpose of generating a path whose length is $2^n + 2$, the distance between two corresponding tape cells in consecutive configurations. Let $\alpha_0, \ldots, \alpha_{k-1}$ be the elements of $Q \cup (Q \times \Gamma)$. We add the following to $\mathcal{T}$:

$$\exists R.I \sqsubseteq \exists S. \bigsqcap_{\ell < k} \exists R.(A_{\alpha_\ell} \sqcap B_{\alpha_\ell} \sqcap (C_B = 0)) \tag{36}$$

$$B_{\alpha_\ell} \sqsubseteq \exists R.\top, \tag{37}$$

$$B_i \sqcap \bigsqcap_{j<i} B_j \sqsubseteq \forall R.\overline{B}_i, \tag{38}$$

$$\overline{B}_i \sqcap \bigsqcap_{j<i} B_j \sqsubseteq \forall R.B_i, \tag{39}$$

$$B_i \sqcap \bigsqcup_{j<i} \overline{B}_j \sqsubseteq \forall R.B_i, \tag{40}$$

$$\overline{B}_i \sqcap \bigsqcup_{j<i} \overline{B}_j \sqsubseteq \forall R.\overline{B}_i, \tag{41}$$

$$(C_B < 2^n + 1) \sqcap B_{\alpha_\ell} \sqsubseteq \forall R.B_{\alpha_\ell}, \tag{42}$$

$$(C_B = 2^n + 1) \sqcap B_{\alpha_\ell} \sqsubseteq \forall R.\overline{A}_{\alpha_\ell}, \tag{43}$$

where ℓ ranges over $0, \ldots, k-1$, i ranges over $0, \ldots, m$, and $(C_B = j)$ (respectively, $(C_B < j)$) denotes a Boolean concept expressing that the value of the $B_i/\overline{B}_i$-counter is j (respectively, smaller than j). We will explain shortly why we need to travel one more R-step (in the first line) after seeing I.

The above CIs generate, after the verification of the computation tree has ended successfully, a tree in the canonical model of the input ABox and of $\mathcal{T}$ as shown in Fig. 8. Note that the topmost edge is labelled with the role name S, which is *not* in Σ. To satisfy Condition (b) above, we must thus not (homomorphically) find the subtree rooted at the node with the incoming S-edge *anywhere* in the canonical model of the ABox and $\mathcal{T}_\emptyset$ (which is just a different presentation of $\mathcal{A}$). We use this effect to ensure that $(*)$ is satisfied *everywhere*. Note that the R-paths in Fig. 8 have length $2^n + 2$ and that we do not display the labelling with the concept names B_i, $\overline{B}_i$, B_α. These concept names are not in Σ and only serve the purpose of achieving the intended path length and of memorising α. Informally, every R-path in the tree represents one possible *copying defect*. The concept names of the form $\overline{A}_\alpha$ stand for the disjunction over all A'_β with $\beta \neq \alpha$. Although we have not done it so far, we can easily modify $\mathcal{T}$ to achieve that they are indeed used this way in the input ABox. For example, we can add the conjunct $\bigsqcap_{\sigma' \in \Gamma \setminus \{\sigma\}} \overline{A}_{\sigma'}$ to the left-hand side of the concept inclusion in (1), and likewise for (6), (7), and so on.

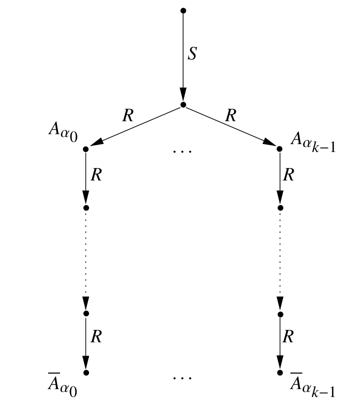


Fig. 8. Tree gadget.

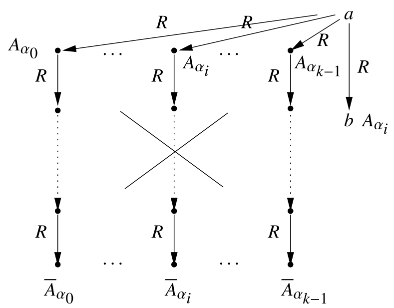


Fig. 9. Additional paths attached to computation tree. In the sequence of paths on the left, the path for A_{α_i} is missing.

If there is a copying defect somewhere in the ABox, then one of the R-paths in Fig. 8 can be homomorphically embedded. We have to ensure that the other paths can be embedded, too. The first step is to add the following CIs:

$$(C' = 2^n + 2) \sqcap \overline{A}_{\alpha_\ell} \sqsubseteq V'_\ell, \tag{44}$$

$$A'_i \sqcap \exists R.A'_i \sqcap \bigsqcup_{j<i} \exists R.A'_j \sqsubseteq \mathsf{ok}'_i, \tag{45}$$

$$\overline{A}'_i \sqcap \exists R.\overline{A}'_i \sqcap \bigsqcup_{j<i} \exists R.A'_j \sqsubseteq \mathsf{ok}'_i, \tag{46}$$

$$A'_i \sqcap \exists R.\overline{A}'_i \sqcap \bigsqcap_{j<i} \exists R.\overline{A}'_j \sqsubseteq \mathsf{ok}'_i, \tag{47}$$

$$\overline{A}'_i \sqcap \exists R.A'_i \sqcap \bigsqcap_{j<i} \exists R.\overline{A}'_j \sqsubseteq \mathsf{ok}'_i, \tag{48}$$

$$\mathsf{ok}'_0 \sqcap \cdots \sqcap \mathsf{ok}'_{n-1} \sqcap \overline{A}'_i \sqcap \exists R.V'_\ell \sqcap A_\sigma \sqcap A'_\sigma \sqsubseteq V'_\ell, \tag{49}$$

$$\exists R.((C' = 0) \sqcap V'_\ell \sqcap A_{\alpha_\ell}) \sqsubseteq V_\ell, \tag{50}$$

where ℓ ranges over $0, \dots, k-1$, i ranges over $0, \dots, m$, and $(C' = j)$ denotes a Boolean concept which expresses that the value of the $A'_i/\overline{A}'_i$-counter is j; recall that the concept names implementing this counter are in Σ. The purpose of the above CIs is to set the verification marker V_ℓ at an individual a whenever we find in the ABox an R-path with root a that is isomorphic to the R-path labelled with $A_{\alpha_\ell}/\overline{A}_{\alpha_\ell}$ in Fig. 8 (and additionally is decorated in an appropriate way with the concept names used by the $A'_i/\overline{A}'_i$-counter).

As the second step, it remains to add the verification markers V_ℓ to the left-hand side of the CIs in $\mathcal{T}$ in such a way that

(∗∗) whenever an ABox individual a that is part of the computation tree has an R-successor in that tree which is labelled with A_{α_ℓ}, then all verification markers V_j with $j \in \{0, \dots, \ell-1, \ell+1, \dots, k-1\}$ must be present at a.

Informally, (∗∗) achieves the presence of additional paths attached to nodes of the computation tree, as displayed in Fig. 9. There, a and b are nodes in the computation tree proper and since A_{α_i} holds at b, we attach to a all paths from Fig. 8 except the one for A_{α_i}. By what was achieved in the first step, we can thus homomorphically embed the R-tree in Fig. 8 at a iff there is a copying defect at the successor of a.

We next describe the modifications required to achieve (∗∗). Line (20) needs to be extended by adding to the left-hand side the conjunct $\bigsqcap_{j\in\{0,\dots,\ell-1,\ell+1,\dots,k-1\}} V_j \sqcap \exists R.\alpha_\ell$ where ℓ ranges over $0,\dots,k-1$. Here, we want $\exists R.\alpha_\ell$ to refer to the same R-successor whose existence is verified by the existing concept $\exists R.V'$ on the left-hand side of (20), or at least to a successor that has the same α_ℓ-label. This can be achieved by adding the CIs

$$\exists R.\alpha_\ell \sqcap \exists R.\alpha_{\ell'} \sqsubseteq \bot \tag{51}$$

where ℓ and ℓ' are distinct, ranging over $0,\dots,k-1$.

The same conjunct needs to be added to the left-hand sides of Lines (21)–(23), (28)–(31), and (33)–(35). We also need to add the conjunct into the scope of the outermost (but not innermost!) existential quantifier in (17) and (18) and to (36), outside the scope of the existential quantifier. Note that we indeed need to travel one more R-step after seeing I (the explanation of this was deferred until now): we always consider copying defects at R-successor of some individual name and thus also the root of our configuration tree should be the R-successor of some individual. Also note that we indeed need to separate successor configurations by two R-steps (the remaining deferred explanation). If we used only one R-step, then the branching ABox individual would *always* allow the R-tree from Fig. 8 to be homomorphically embedded, no matter whether there is a copying defect or not.

Lemma 59. *The following conditions are equivalent:*

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

Proof. (sketch) For (2) ⇒ (1), suppose M accepts w. The accepting computation tree of M on w can be represented as a Σ-ABox as detailed above alongside the construction of the TBox $\mathcal{T}$. The representation only uses the role name R and the concept names A_i, $\overline{A}_i$,A'_i, $\overline{A}'_i$, A_σ, $A_{q,\sigma}$, A'_σ, $A'_{q,\sigma}$, $\overline{A}_\sigma$, $\overline{A}_{\boldsymbol{q},\sigma}$, X_1, and X_2. 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. The resulting ABox $\mathcal{A}$ is consistent with $\mathcal{T}$. Moreover, since there are no copying defects, there is no homomorphism from $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_\emptyset,\mathcal{A}}$.

For (1) ⇒ (2), suppose there is a tree Σ-ABox $\mathcal{A}$ that satisfies (a) and (b). Because of (b), I must be true somewhere in $\mathcal{I}_{\mathcal{T},\mathcal{A}}$: otherwise, $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ does not contain anonymous elements and the identity is a homomorphism from $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_\emptyset,\mathcal{A}}$, contradicting (b). Since I is true somewhere in $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ and by the construction of $\mathcal{T}$, the ABox must contain the representation of an accepting computation tree of M on w, except satisfaction of (∗). For the same reason, $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ must contain a tree as shown in Fig. 8. As already been argued during the construction of $\mathcal{T}$, however, condition (∗) follows from the existence of such a tree in $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ together with (b). □

We remark that the above reduction also yields 2ExpTime hardness for (Σ,Σ)-CQ entailment in the DL $\mathcal{ELI}$ extending $\mathcal{EL}$ with inverse roles. In fact, CIs $D \sqsubseteq \forall r.C$ can be replaced by $\exists r^-.D \sqsubseteq C$ and disjunctions on the left-hand side can be removed with only a polynomial blowup. It thus remains to eliminate ⊥, which only occurs non-nested on the right-hand side of CIs. With the exception of the CIs in (27), this can be done as follows: replace $\mathcal{T}_\emptyset$ with a non-empty TBox $\mathcal{T}_1$ and rename $\mathcal{T}$ to $\mathcal{T}_2$ for uniformity; include all CIs with ⊥ on the right-hand side in $\mathcal{T}_1$ instead of in $\mathcal{T}_2$; then replace ⊥ with a fresh concept name D and further extend $\mathcal{T}_1$ with CIs which make sure that $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ contains an R-tree as in Fig. 8 whenever D is non-empty, which is straightforward. As a consequence, any ABox that satisfies the left-hand side of a ⊥-CI in the original TBox $\mathcal{T}$ cannot satisfy (b) from Lemma 59 and does not have to be considered.

For the excluded CIs, a different approach needs to be taken since these CIs rely on many CIs in $\mathcal{T}_2$ that are not included in $\mathcal{T}_1$. We only sketch the required modification: instead of introducing the concept names $S^t_{q_1,\sigma_1,D_1}$, one would propagate transitions inside the V'-markers. Thus, $S^1_{q_1,\sigma_1,D_1}$, $S^2_{q_2,\sigma_2,D_2}$, and V' would be integrated into a single marker $V'_{q_1,\sigma_1,D_1,q_2,\sigma_2,D_2}$, and likewise for $V_{L,q}$. The excluded CIs can then simply be dropped.

Theorem 60. *It is 2ExpTime-hard to decide whether an $\mathcal{ELI}$ TBox (Σ,Σ)-CQ entails an $\mathcal{ELI}$ TBox.*

A corresponding upper bound has recently been established in [68].

9. Related work

The comparison of logical theories has been an active research area almost since the invention of formal logic. Important concepts include Tarski's notion of *interpretability* [69] of one theory into another and the notion of *conservative extension*, which has been employed extensively in mathematical logic, in particular to compare theories of sets and numbers [70]. Conservative extensions have also been used to formalise modular software specification [71–73] and to enable modular ontology development [42,16,17]. Query entailment can be regarded as a generalisation of conservative extension where we do not require that one of the theories under consideration is included in the other and where conservativity depends on

Table 4
KB query inseparability [34].

DL	Complexity	DL	Complexity
$\mathcal{EL}(\mathcal{H}^{dr}_{\bot})$	P	–	–
DL-Lite$_{core}$	P	*DL-Lite*$^{\mathcal{H}}_{core}$	ExpTime
Horn$\mathcal{ALC}(\mathcal{H})$	ExpTime	*Horn*$\mathcal{ALC}(\mathcal{H})\mathcal{I}$	2ExpTime

Table 5
TBox query inseparability.

DL	Complexity	DL	Complexity
$\mathcal{EL}$	ExpTime [38]	*Horn*$\mathcal{ALC}(\mathcal{H})\mathcal{I}$	2ExpTime [68]
DL-Lite$_{core}$	in P [41]	*DL-Lite*$^{\mathcal{H}}_{core}$	ExpTime [34]

database queries in a signature of interest instead of formulas in the signature of the smaller theory. In an independent but closely related research field, various notions of equivalence between (extended) datalog programs have been proposed and investigated [74], often focusing on answer set programming [74–77].

The state of the art in the research of inseparability between description logic ontologies has recently been presented in great detail in [41]. This survey contains, in particular, a discussion of the relationships between concept-based, model-based, and query-based inseparability. In the first approach, one compares the concept inclusions entailed by the two versions of an ontology. In the second approach, one compares the models of the two versions. In contrast, in the query-based approach underpinning the present investigation, one compares the certain answers to database queries. It turns out that the three approaches exhibit rather different properties and require different model-theoretic and algorithmic techniques. While various forms of bisimulations and corresponding bisimulation-invariant tree automata are required to investigate concept-based inseparability, query-based inseparability relies on understanding homomorphisms between interpretations and products, which are then reflected in the games or automata required to design algorithms; we refer the reader to [41] for an in-depth discussion. Important notions that are closely related to query inseparability, such as knowledge exchange and entailment between OBDA specifications, are discussed in [34].

In what follows, we focus on summarising what is known about query inseparability between description logic ontologies, discussing both the KB and the TBox cases. All existing results are about Horn-DLs as the present paper is the first one to study query-based inseparability for expressive non Horn-DLs. As discussed in this paper, for Horn-DLs, there is no difference between CQ- and UCQ-inseparability, so we do not explicitly distinguish between them below.

We start with the KB case. In [34], CQ-inseparability between KBs is investigated for Horn-DLs ranging from the lightweight $\mathcal{EL}$ and *DL-Lite*$_{core}$ to *Horn*$\mathcal{ALCHI}$. The authors develop model-theoretic and game-theoretic characterisations of query inseparability. In contrast to the present investigation, the main complexity results, summarised in Table 4, are then obtained using the game-theoretic characterisations instead of reductions to the emptiness problem of tree-automata. It is also proved that rootedness does not affect the worst-case complexity of query entailment. Observe that the addition of the inverse role constructor leads to an exponential increase of the complexity of checking query inseparability.

CQ-inseparability between TBoxes has been investigated for $\mathcal{EL}$ terminologies (a restricted form of TBox) extended with role inclusions and domain and range restrictions [15,78], for (unrestricted TBoxes in) the description logic $\mathcal{EL}$ [38], and for variants of DL-Lite [41,34]. The algorithms presented in [15] are based on both model-theoretic and proof-theoretic methods. The authors focus not only on deciding inseparability but also on presenting the logical difference between TBoxes to the user. A versioning and modularisation system for acyclic $\mathcal{EL}$ TBoxes based on CQ-inseparability is presented and evaluated in [78]. The system makes intense use of the fact that, in this case, query inseparability can be decided in polynomial time. This is in contrast to general $\mathcal{EL}$ TBoxes for which ExpTime completeness of deciding CQ-inseparability is shown in [38]. The method is purely model-theoretic and based on the close relationship between concept and query inseparability for $\mathcal{EL}$. More recently, CQ inseparability has been investigated for Horn$\mathcal{ALCHI}$ and shown to be 2ExpTime-complete, using a subtle approach that combines a mosaic technique with automata [68]. The mentioned results are summarised in Table 5.

10. Conclusion and future work

We have made significant steps towards understanding query entailment and inseparability for KBs and TBoxes in expressive DLs. Our main—and rather unexpected—results are as follows:

– for $\mathcal{ALC}$-KBs, Σ-(r)UCQ inseparability is decidable and (r)CQ-inseparability is undecidable (even without restrictions on the signature);
– for *Horn*$\mathcal{ALC}$-TBoxes, Θ-rCQ inseparability is ExpTime complete and Θ-CQ inseparability is 2ExpTime complete.

The first result reflects a fundamental difference between the model-theoretic characterisations of inseparability for CQs and UCQs: while UCQ-inseparability can be characterised using (partial) homomorphisms between models of the respective KBs, CQ-inseparability requires the construction of products of the models of the respective KBs, a result which is at the core of our undecidability proof. The second result reflects a fundamental difference between homomorphisms whose domain is connected to ABox individuals (as required for rooted CQs) and those whose domain is not necessarily reachable from the ABox. Searching for the latter turns out to be much harder. Both results have important practical implications. The first one indicates that one should approximate CQ-inseparability using UCQ-inseparability when designing practical algorithms. Observe that this is a sound approximation as no two ontologies that are UCQ-inseparable can be separated by CQs. The second one indicates that it is worth focusing on rooted (U)CQs rather than all (U)CQs when designing practical algorithms for inseparability. The latter are likely to cover the vast majority of queries used in practice. We believe that our model-theoretic characterisations provide a good foundation for developing practical (approximation) algorithms.

Many problems remain open. The main one, which can be directly inferred from the tables presenting our results, is the decidability of UCQ-inseparability for $\mathcal{ALC}$ TBoxes. We conjecture that this problem is undecidable but have found no way of proving this. Another family of interesting open problems concerns the role of the signatures Σ and Θ in our investigation of the decidability/complexity of inseparability between KBs and TBoxes, respectively. Observe that admitting more symbols in Σ or Θ leads to sound approximations of the original inseparability problem: for example, if TBoxes are Θ'-CQ inseparable for a pair of signatures $\Theta' \supseteq \Theta$, then they are Θ-CQ inseparable as well. It would, therefore, be of great interest to understand the complexity of inseparability if Σ and Θ consist of *all* concept and role names (the 'full signature' case). We have been able to prove undecidability of full signature (r)CQ-inseparability for $\mathcal{ALC}$ KBs, but the complexity of full signature (r)UCQ-inseparability between $\mathcal{ALC}$ KBs remains open. Similarly, the decidability of full signature (r)CQ-inseparability and (r)UCQ-inseparability between $\mathcal{ALC}$ TBoxes remains open. The 'hiding technique' discussed in this paper might be a good starting point to attack those problems. Finally, it would be of interest to consider extensions of $\mathcal{ALC}$ with inverse roles, qualified number restrictions, nominals, and role inclusions. We conjecture that extensions of our results to DLs with qualified number restrictions and role inclusions are rather straightforward (though proofs might become significantly less transparent). The addition of inverse roles, however, might lead to non-trivial modifications of the model-theoretic criteria, see also [68].

Acknowledgements

We thank the anonymous reviewers for their very thorough and useful comments. This research was supported by the DFG grant LU 1417/2-1 (C. Lutz), the ERC consolidator grant CODA 647289, and the EPSRC joint grants EP/M012646/1 and EP/M012670/1 'iTract: Islands of Tractability in Ontology-Based Data Access' (F. Wolter and M. Zakharyaschev).

Appendix A. Proof of Theorem 22

For the proof of Theorem 22 (i), suppose that an instance $\mathfrak{T}$ of the rectangle tiling problem is given. Consider the KBs $\mathcal{K}^1_{\mathrm{rCQ}} = (\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A}_{\mathrm{rCQ}})$ and $\mathcal{K}^2_{\mathrm{rCQ}} = (\mathcal{T}^2_{\mathrm{rCQ}}, \mathcal{A}_{\mathrm{rCQ}})$ given in the proof sketch for Theorem 22 (i). It suffices to prove Lemmas 18 and 19 for the new KBs, the rCQs $q^r_n(y)$, and the signature Σ_{rCQ}.

Lemma 61. *The instance $\mathfrak{T}$ admits a rectangle tiling iff there exists $\boldsymbol{q}^r_n(a)$ such that $\mathcal{K}^2_{rCQ} \models \boldsymbol{q}^r_n(a)$.*

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

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

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

$$Row,\ block_1,\ block_1,\ block_2, \dots,\ block_{M-1}$$

(thus, $n = (N+1) \times M + 1$). In view of Lemma 11, we only need to prove $\mathcal{I} \models \boldsymbol{q}^r_n(a)$ for each minimal model $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}}$. Take such an $\mathcal{I}$. We have to show that there is an R-path $a, x_0, \dots, x_{n+1}$ in $\mathcal{I}$ such that $x_i \in B_i^{\mathcal{I}}$ and $x_{n+1} \in End^{\mathcal{I}}$.

First, we construct an auxiliary R-path $y_0, \dots, y_{n-N-1}$. We take $y_0 \in Row^{\mathcal{I}}$, the successor of a in $\mathcal{I}$, and $y_1 \in {I_0}^{\mathcal{I}}$, the successor of y_0 in $\mathcal{I}$, by (21) ($I_0 = T^{1,1}$). Then we take $y_2 \in (T^{2,1})^{\mathcal{I}}, \dots, y_N \in (T^{N,1})^{\mathcal{I}}$ by (6). We now have $right(T^{N,1}) = W$. By (7), we obtain $y_{N+1} \in {Row_1}^{\mathcal{I}}$. By (9), $y_{N+1} \in {Row_1}^{\mathcal{I}} \subseteq Row^{\mathcal{I}}$. We proceed in this way, starting with (5), till the moment we construct $y_{n-1} \in (T^{N,M-1})^{\mathcal{I}}$, for which we use (8) and (15) to obtain $y_n \in (Row^{halt}_k)^{\mathcal{I}} \subseteq Row^{\mathcal{I}}$, for some k. Note that $T^{\mathcal{I}} \subseteq \widehat{T}^{\mathcal{I}}$ by (10).

By (12), two cases are possible now.

Case 1: there is y such that $(y_n, y) \in R^{\mathcal{I}}$ and $y \in End^{\mathcal{I}}$. 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 z_1 such that $(y_n, z_1) \in R^{\mathcal{I}}$ and $z_1 \in (T^{halt}_k)^{\mathcal{I}}$, where $T = T^{1,M}$ and $up(T) = C$. We then use (13) and find $z_2, \dots, z_N, u, v$ such that $z_i \in (T^{halt}_k)^{\mathcal{I}}$, where $T = T^{i,M}$, $u \in Row^{\mathcal{I}}$ and $v \in End^{\mathcal{I}}$. 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 (11) and (16), we have $(T^{i,j})^{\mathcal{I}} \subseteq (\widehat{T}^{i,j-1})^{\mathcal{I}}$.

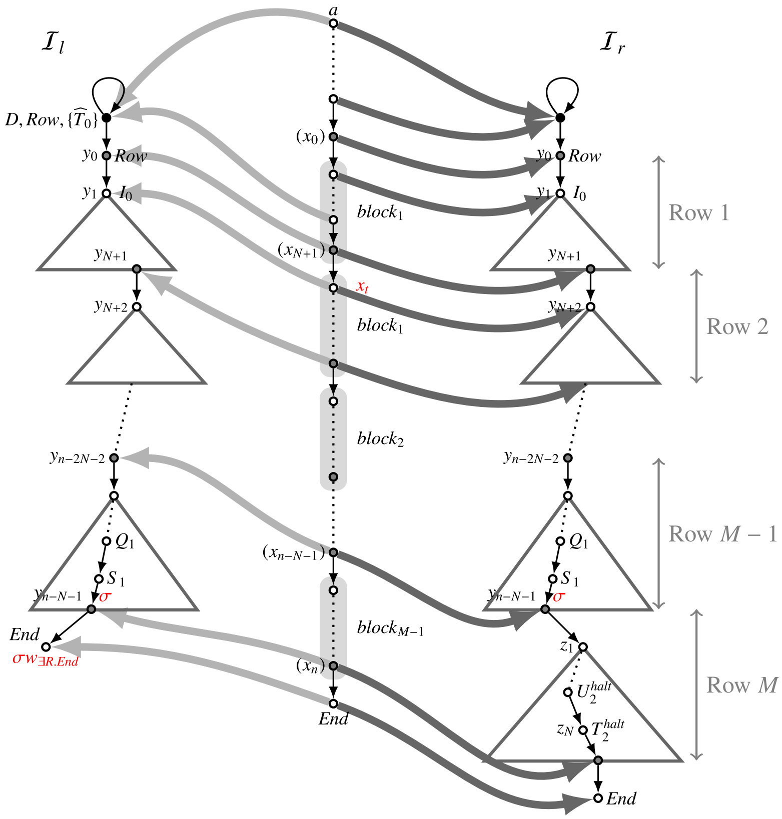


Fig. A.1. Two homomorphisms to minimal models.

($\Leftarrow$) Suppose $\mathcal{K}^2_{\text{rCQ}} \models \boldsymbol{q}^r_n(a)$ for some $n > 0$. Consider all the pairwise distinct pairs $(\mathcal{I}, h)$ such that $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\text{rCQ}}}$ and h is a homomorphism from $\boldsymbol{q}^r_n(a)$ to $\mathcal{I}$. Note that $h(\boldsymbol{q}^r_n)$ contains an or-node σ_h (which is an instance of Row^{halt}_k, for some k). We call $(\mathcal{I}, h)$ and h *left* if $h(x_{n+1}) = \sigma_h \cdot w_{\exists R.End}$, and *right* otherwise. It is not hard to see that there exist a left $(\mathcal{I}_l, h_l)$ and a right $(\mathcal{I}_r, h_r)$ with $\sigma_{h_l} = \sigma_{h_r}$ (if this is not the case, we can construct $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\text{rCQ}}}$ such that $\mathcal{I} \not\models \boldsymbol{q}^r_n(a)$).

Take $(\mathcal{I}_l, h_l)$ and $(\mathcal{I}_r, h_r)$ such that $\sigma_{h_l} = \sigma_{h_r} = \sigma$ and use them to construct the required tiling. Let $\sigma = aw_0 \cdots w_{n'}$. We have $h_l(x_n) = \sigma$, $h_l(x_{n+1}) = \sigma \cdot w_{\exists R.End}$. 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 $v_m = w_{\exists R.T_2^{halt}}$ (any k other than 2 is treated analogously). By (14), $right(T) = W$; by (13), $up(T) = C$. Suppose $w_{n'-1} = w_{\exists R.S_k}$. Now, we know that $k = 1$. By (8), $right(S) = W$. Consider the atom $B_{n-1}(x_{n-1})$ from $\boldsymbol{q}^r_n$. Both $aw_0 \cdots w_{n'-1}$ and $\sigma v_1 \cdots v_m$ are instances of B_{n-1}. By (10) and (16), $B_{n-1} = \widehat{S}_1$ and $down(T) = up(S)$. Suppose $v_{m-1} = w_{\exists R.U_2^{halt}}$. By (13), $right(U) = left(T)$ and $up(U) = C$. Suppose $w_{n'-2} = w_{\exists R.Q_1}$. By (6), $right(Q) = left(S)$. Consider the atom $B_{n-2}(x_{n-2})$ from $\boldsymbol{q}^r_n$. Both $aw_0 \cdots w_{n'-2}$ and $\sigma \cdots v_{m-1}$ are instances of B_{n-2}. By (10) and (16), $B_{n-2} = \widehat{Q}_1$ and $down(U) = up(Q)$. 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 $B_{n-N-1} = Row$. Thus, we have tiled the last two rows of the grid.

We proceed in this way until we have reached some variable x_t, for $t \geq 0$, of $\boldsymbol{q}^r_n$ that is mapped by h_l to $aw_0 w_1$ (see Fig. A.1). Note that this situation is guaranteed to occur. Indeed, $h_l(a) = a$, $h_l(x_0) \in \{a, aw_0\}$, $h_l(x_1) \in \{a, aw_0, aw_0 w_1\}$, etc. Clearly, the assumption that $h_l(x_i) \in \{a, aw_0\}$ for all i $(0 \leq i \leq n+1)$ leads to a contradiction. Let $h_r(x_t) = aw_0 \cdots w_s$, for some $s > 1$. Note that $s = N + 2$. By (21), it follows that $aw_0 w_1$ is an instance of I_0. Therefore, $B_t = \widehat{I}_0$ and, by (11), $aw_0 \cdots w_s$ is an instance of V_1, for some tile V such that $down(V) = up(I)$.

Thus, we have a 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. □

Lemma 62. *$\prod \boldsymbol{M}_{\mathcal{K}^2_{\text{rCQ}}}$ is con-$n\Sigma_{\text{rCQ}}$-homomorphically embeddable into $\mathcal{I}_{\mathcal{K}^1_{\text{rCQ}}}$ preserving $\{a\}$ for all $n \geq 1$ iff there does not exist an rCQ $\boldsymbol{q}^r_m(y)$ such that $\prod \boldsymbol{M}_{\mathcal{K}^2_{\text{rCQ}}} \models \boldsymbol{q}^r_m(a)$.*

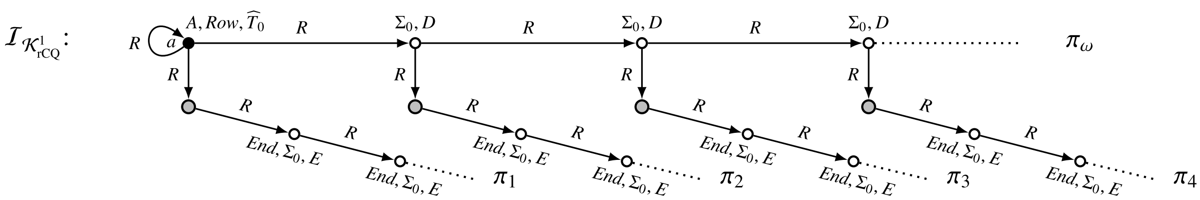
Proof. (⇒) Suppose otherwise, that is, $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}} \models \boldsymbol{q}^r_m(a)$ for some m. By the assumption, $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}}$ is con-$n\Sigma_{\mathrm{rCQ}}$-homomorphically embeddable into $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$ for $n = m+3$ (the length of $\boldsymbol{q}^r_m$). So we have $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}} \models \boldsymbol{q}^r_m(a)$, which is clearly impossible because none of the paths of $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$ contains the full sequence of symbols mentioned in $\boldsymbol{q}^r_m(y)$.

(⇐) Suppose $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}} \not\models \boldsymbol{q}^r_m(a)$ for all m. Take any subinterpretation of $\prod \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}}$ whose domain contains n elements connected to a. Recall from the proof of Theorem 6 that we can regard the Σ_{rCQ}-reduct of this subinterpretation as a Σ_{rCQ}-rCQ, and so denote it by $\boldsymbol{q}(y)$. Clearly, $\boldsymbol{q}$ is tree shaped plus the atom $R(y, y)$. We know that there is no Σ_{rCQ}-homomorphism from $\boldsymbol{q}^r_m(y)$ into $\boldsymbol{q}(y)$ for any m; in particular, $\boldsymbol{q}(y)$ does not have a subquery of the form $\boldsymbol{q}^r_m(y)$. We have to show that $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}} \models \boldsymbol{q}(a)$. We show how to map $\boldsymbol{q}(y)$ starting from a.

We call a variable x in $\boldsymbol{q}(y)$ a *gap* if there exists no $B \in \Sigma_{\mathrm{rCQ}}$ such that $B(x)$ is in $\boldsymbol{q}(y)$. Since $\boldsymbol{q}(y)$ does not contain a subquery of the form $\boldsymbol{q}^r_m(y)$, we know that every path ρ starting from y in $\boldsymbol{q}(y)$ either:

(a) does not contain $\mathit{End}(x)$, or
(b) contains $\mathit{End}(x)$ and contains a gap x' that occurs between the y and x.

If all paths ρ starting from y in $\boldsymbol{q}(y)$ are of type **(a)** we map $\boldsymbol{q}(y)$ on the path π_ω:



Otherwise, let y be the current variable and a the current image. Let $x_1, \dots, x_k$ be all successor gaps and $z_1, \dots, z_l$ all successor non-gaps of the current variable in $\boldsymbol{q}(y)$. We map all x_i to the vertical successor and all z_i to the horizontal successor of the current image. All the rest of the paths starting from x_i can then be mapped to an appropriate π_i. We then consider each z_i as the current variable, and the point where it has been mapped as the current image, and continue analogously. Thus, the paths ρ not containing gaps and $\mathit{End}(x)$ atoms would result in being mapped to π_ω, while the paths with gaps would each result in being mapped to an appropriate π_i. □

We now prove Theorem 22 (*ii*). We set $\mathcal{K}_2 = \mathcal{K}^2_{\mathrm{rCQ}} \cup \mathcal{K}^1_{\mathrm{rCQ}}$ and show that the following are equivalent:

(1) $\mathcal{K}^1_{\mathrm{rCQ}}$ Σ_{rCQ}-rCQ entails $\mathcal{K}^2_{\mathrm{rCQ}}$;
(2) $\mathcal{K}^1_{\mathrm{rCQ}}$ and $\mathcal{K}_2$ are Σ_{rCQ}-rCQ inseparable.

Let $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$ be the canonical model of $\mathcal{K}^1_{\mathrm{rCQ}}$ and $\boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}}$ the set of minimal models of $\mathcal{K}^2_{\mathrm{rCQ}}$. Again, one can easily show that the following set $\boldsymbol{M}_{\mathcal{K}_2}$ is complete for $\mathcal{K}_2$:

$$\boldsymbol{M}_{\mathcal{K}_2} = \{\mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}} \mid \mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}}\},$$

where $\mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$ is the interpretation that results from merging the roots a of $\mathcal{I}$ and $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$. Now (2) ⇒ (1) is trivial. For the converse, suppose $\mathcal{K}^1_{\mathrm{rCQ}}$ Σ_{rCQ}-rCQ entails $\mathcal{K}^2_{\mathrm{rCQ}}$. It directly follows that $\mathcal{K}_2$ Σ_{rCQ}-rCQ entails $\mathcal{K}^1_{\mathrm{rCQ}}$. So it remains to show that $\mathcal{K}^1_{\mathrm{rCQ}}$ Σ_{rCQ}-rCQ entails $\mathcal{K}_2$. Suppose this is not the case. Without loss of generality, we may assume that there is a Σ_{rCQ}-rCQ $\boldsymbol{q}(y)$, a ditree with one answer variable y not mentioning D and E, such that $\mathcal{K}_2 \models \boldsymbol{q}(a)$ and $\mathcal{K}^1_{\mathrm{rCQ}} \not\models \boldsymbol{q}(a)$. We can assume $\boldsymbol{q}$ to be a *smallest* rCQ with this property. Consider the various cases of $\boldsymbol{q}(y)$:

– $\boldsymbol{q}(y)$ does not contain *End* atoms: but then $\mathcal{K}^1_{\mathrm{rCQ}} \models \boldsymbol{q}(a)$ (see the proof of Lemma 62), contrary to our assumption.
– $\boldsymbol{q}(y)$ contains *End* atoms and, on each path from y to an *End* atom, there is a variable x that does not appear in $\boldsymbol{q}(y)$ in any atom of the form $B(x)$, for a concept name $B \in \Sigma$. But then $\mathcal{K}^1_{\mathrm{rCQ}} \models \boldsymbol{q}(a)$ (see the proof of Lemma 62), contrary to our assumption.
– $\boldsymbol{q}(y)$ contains *End* atoms and a path from y to an *End* atom such that each variable x on this path appears in an atom of the form $B(x)$, for a concept name $B \in \Sigma$. Denote this path by $\boldsymbol{q}'(y)$, and observe that $\boldsymbol{q}'(y)$ is a query of the form $\boldsymbol{q}^r_n(y)$. Then $\mathcal{K}^1_{\mathrm{rCQ}} \not\models \boldsymbol{q}'(a)$ by the construction of $\mathcal{K}^1_{\mathrm{rCQ}}$, moreover there is no subquery $\boldsymbol{q}''$ of $\boldsymbol{q}'(y)$ such that there is a model $\mathcal{I} \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}}$ and $\mathcal{I} \uplus \mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}} \models \boldsymbol{q}'(a)$ by mapping $\boldsymbol{q}''$ entirely into $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$. So it must be that $\mathcal{K}^2_{\mathrm{rCQ}} \models \boldsymbol{q}'(a)$. But now, as $\mathcal{K}^1_{\mathrm{rCQ}} \models \mathcal{K}^2_{\mathrm{rCQ}}$, we know that $\mathcal{K}^2_{\mathrm{rCQ}} \not\models \boldsymbol{q}^r_n(a)$ for each n, which is again a contradiction.

The contradictions arise from the assumption that $\mathcal{K}^1_{\mathrm{rCQ}}$ does not Σ_{rCQ}-rCQ entail $\mathcal{K}_2$.

Appendix B. Proof of Theorem 43 for rooted CQs

We show that it is undecidable whether an $\mathcal{EL}$ TBox is Θ-rCQ inseparable from an $\mathcal{ALC}$ TBox. For the proof we require homomorphisms between ABoxes and the observation that they preserve certain answers. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be ABoxes. A map h from $\mathsf{ind}(\mathcal{A}_1)$ to $\mathsf{ind}(\mathcal{A}_2)$ is called an *ABox-homomorphism* if $A(a) \in \mathcal{A}_1$ implies $A(h(a)) \in \mathcal{A}_2$ for all concept names A, and $R(a,b) \in \mathcal{A}_1$ implies $R(h(a),h(b)) \in \mathcal{A}_2$ for all role names R. The following is shown in [64].

Proposition 63. *Let $\mathcal{T}$ be an $\mathcal{ALC}$ TBox, $\mathcal{A}$, $\mathcal{A}'$ be ABoxes, and $h\colon \mathcal{A} \to \mathcal{A}'$ an ABox homomorphism. Then*

- *$\mathcal{A}$ is consistent with $\mathcal{T}$ if $\mathcal{A}'$ is consistent with $\mathcal{T}$, and*
- *$(\mathcal{T},\mathcal{A}) \models \boldsymbol{q}(\boldsymbol{a})$ implies $(\mathcal{T},\mathcal{A}') \models \boldsymbol{q}(h(\boldsymbol{a}))$ for all CQs $\boldsymbol{q}(\boldsymbol{x})$.*

To prove the undecidability of the problem whether an $\mathcal{EL}$ TBox is Θ-rCQ inseparable from an $\mathcal{ALC}$ TBox, we use the TBoxes constructed in the proof of Theorem 22. Recall the KBs $\mathcal{K}^1_{\mathrm{rCQ}} = (\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A}_{\mathrm{rCQ}})$, $\mathcal{K}^2_{\mathrm{rCQ}} = (\mathcal{T}^2_{\mathrm{rCQ}}, \mathcal{A}_{\mathrm{rCQ}})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A}_{\mathrm{rCQ}})$, where $\mathcal{T}_2 = \mathcal{T}^1_{\mathrm{rCQ}} \cup \mathcal{T}^2_{\mathrm{rCQ}}$. Set $\Theta = (\Sigma_1, \Sigma_2)$, where $\Sigma_1 = \mathsf{sig}(\mathcal{A}_{\mathrm{rCQ}})$ and $\Sigma_2 = \Sigma_{\mathrm{rCQ}}$. We aim to show that the following conditions are equivalent:

(1) $\mathcal{K}^1_{\mathrm{rCQ}}$ and $\mathcal{K}_2$ are Σ_{rCQ}-rCQ inseparable;
(2) $\mathcal{T}^1_{\mathrm{rCQ}}$ and $\mathcal{T}_2$ are Θ-rCQ inseparable.

The implication (2) $\Rightarrow$ (1) is straightforward: if $\mathcal{K}^1_{\mathrm{rCQ}}$ and $\mathcal{K}_2$ are not Σ_{rCQ}-CQ inseparable then the ABox $\mathcal{A}_{\mathrm{rCQ}}$ witnesses that $\mathcal{T}^1_{\mathrm{rCQ}}$ and $\mathcal{T}_2$ are not Θ-rCQ inseparable. Conversely, suppose $\mathcal{T}^1_{\mathrm{rCQ}}$ and $\mathcal{T}_2$ are not Θ-rCQ inseparable. Take a Σ_1-ABox $\mathcal{A}$ such that $(\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A})$ and $(\mathcal{T}_2, \mathcal{A})$ are not Σ_2-rCQ inseparable. Clearly, $(\mathcal{T}_2, \mathcal{A})$ Σ_2-rCQ entails $(\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A})$. Thus, $(\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A})$ does not Σ_2-rCQ entail $(\mathcal{T}_2, \mathcal{A})$. The canonical model $\mathcal{I}_1$ of the $\mathcal{EL}$ KB $(\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A})$ can be constructed by taking, for every $A(b) \in \mathcal{A}$, a copy of the canonical model $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$ and hooking the two R-successors of a in $\mathcal{I}_{\mathcal{K}^1_{\mathrm{rCQ}}}$ (together with the subinterpretations they root) as fresh R-successors to b. On the other hand, the class $\boldsymbol{M}$ of minimal models of $(\mathcal{T}_2, \mathcal{A})$ is obtained from $\mathcal{I}_1$ by hooking to every b with $A(b) \in \mathcal{A}$ a copy of a minimal model $\mathcal{I}_b \in \boldsymbol{M}_{\mathcal{K}^2_{\mathrm{rCQ}}}$ by identifying the root a of $\mathcal{I}_b$ with b.

Now consider a Σ_2-rCQ $\boldsymbol{q}(\boldsymbol{a})$ with $(\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A}) \not\models \boldsymbol{q}(\boldsymbol{a})$ and $(\mathcal{T}_2, \mathcal{A}) \models \boldsymbol{q}(\boldsymbol{a})$. Suppose $\boldsymbol{q}(\boldsymbol{a})$ is the smallest rCQ with this property. Using the description of the canonical model $\mathcal{I}_1$ of $(\mathcal{T}^1_{\mathrm{rCQ}}, \mathcal{A})$ and the class $\boldsymbol{M}$ of minimal models of $(\mathcal{T}_2, \mathcal{A})$, one can show in the same way as in the proof of Theorem 22 (*ii*) given in the appendix above that there must be a path in $\boldsymbol{q}$ from an answer variable to an *End* atom such that each variable x on this path appears in an atom of the form $B(x)$ with $B \in \Sigma_{\mathrm{rCQ}}$. But then $\boldsymbol{q}$ contains a query of the form $\boldsymbol{q}^r_n(x)$ (see again the proof of Theorem 22 (*ii*)) such that $(\mathcal{T}_2, \mathcal{A}) \models \boldsymbol{q}^r_n(a)$ for some individual a and $n > 0$. Observe that the map $h\colon \mathsf{ind}(\mathcal{A}) \to \{a\}$ is an ABox-homomorphism from the ABox $\mathcal{A}$ onto the ABox $\mathcal{A}_{\mathrm{rCQ}}$. It follows from Proposition 63 that $(\mathcal{T}_2, \mathcal{A}_{\mathrm{rCQ}}) \models \boldsymbol{q}^r_n(h(a))$, for some n. We know from the proof of Theorem 22 that $\mathcal{K}^1_{\mathrm{rCQ}} \not\models q^r_n(a)$. Thus, $\mathcal{K}^1_{\mathrm{rCQ}}$ and $\mathcal{K}_2$ are not Σ_{rCQ}-rCQ inseparable, as required.

Appendix C. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.artint.2018.09.003.

References

[1] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati, Linking data to ontologies, J. Data Semant. 10 (2008) 133–173.
[2] M. Bienvenu, M. Ortiz, Ontology-mediated query answering with data-tractable description logics, in: 11th Reasoning Web International Summer School Tutorial Lectures (RW 2015), 2015, pp. 218–307.
[3] R. Kontchakov, M. Rodriguez-Muro, M. Zakharyaschev, Ontology-based data access with databases: a short course, in: 9th Reasoning Web International Summer School Tutorial Lectures (RW 2013), 2013, pp. 194–229.
[4] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati, Tractable reasoning and efficient query answering in description logics: the DL-Lite family, J. Autom. Reason. 39 (2007) 385–429.
[5] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodriguez-Muro, R. Rosati, M. Ruzzi, D.F. Savo, The MASTRO system for ontology-based data access, Semant. Web 2 (2011) 43–53.
[6] M. Rodriguez-Muro, R. Kontchakov, M. Zakharyaschev, Ontology-based data access: Ontop of databases, in: Proceedings of the 12th International Semantic Web Conference (ISWC 2013), Springer, 2013, pp. 558–573.
[7] T. Eiter, M. Ortiz, M. Simkus, T. Tran, G. Xiao, Query rewriting for Horn-SHIQ plus rules, in: Proceedings of the 26th National Conference on Artificial Intelligence (AAAI 2012), AAAI Press, 2012, pp. 726–733.
[8] D. Trivela, G. Stoilos, A. Chortaras, G.B. Stamou, Optimising resolution-based rewriting algorithms for OWL ontologies, J. Web Semant. 33 (2015) 30–49.
[9] I. Kollia, B. Glimm, Optimizing SPARQL query answering over OWL ontologies, J. Artif. Intell. Res. 48 (2013) 253–303.
[10] Y. Zhou, B.C. Grau, Y. Nenov, M. Kaminski, I. Horrocks, Pagoda: pay-as-you-go ontology query answering using a datalog reasoner, J. Artif. Intell. Res. 54 (2015) 309–367.

[11] N.F. Noy, M.A. Musen, PromptDiff: a fixed-point algorithm for comparing ontology versions, in: Proceedings of the 18th National Conference on Artificial Intelligence (AAAI 2002), AAAI Press, Menlo Park, CA, USA, 2002, pp. 744–750.
[12] M.C.A. Klein, D. Fensel, A. Kiryakov, D. Ognyanov, Ontology versioning and change detection on the Web, in: Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, in: Lecture Notes in Computer Science, vol. 2473, Springer Verlag, Berlin/Heidelberg, Germany, 2002, pp. 247–259.
[13] T. Redmond, M. Smith, N. Drummond, T. Tudorache, Managing change: an ontology version control system, in: Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED 2008), in: CEUR Workshop Proceedings, vol. 432, 2008.
[14] E. Jimenez-Ruiz, B. Cuenca Grau, I. Horrocks, R.B. Llavori, Supporting concurrent ontology development: framework, algorithms and tool, Data Knowl. Eng. 70 (2011) 146–164.
[15] B. Konev, M. Ludwig, D. Walther, F. Wolter, The logical difference for the lightweight description logic EL, J. Artif. Intell. Res. 44 (2012) 633–708.
[16] H. Stuckenschmidt, C. Parent, S. Spaccapietra (Eds.), Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization, Lecture Notes in Computer Science, vol. 5445, Springer, 2009.
[17] O. Kutz, T. Mossakowski, D. Lücke, Carnap, Goguen, and the hyperontologies: logical pluralism and heterogeneous structuring in ontology design, Log. Univers. 4 (2010) 255–333.
[18] B. Cuenca Grau, I. Horrocks, Y. Kazakov, U. Sattler, Modular reuse of ontologies: theory and practice, J. Artif. Intell. Res. 31 (2008) 273–318.
[19] R. Kontchakov, F. Wolter, M. Zakharyaschev, Logic-based ontology comparison and module extraction, with an application to DL-Lite, Artif. Intell. 174 (2010) 1093–1141.
[20] A.A. Romero, M. Kaminski, B.C. Grau, I. Horrocks, Module extraction in expressive ontology languages via datalog reasoning, J. Artif. Intell. Res. 55 (2016) 499–564.
[21] G. De Giacomo, M. Lenzerini, A. Poggi, R. Rosati, On instance-level update and erasure in description logic ontologies, J. Log. Comput. 19 (2009) 745–770.
[22] H. Liu, C. Lutz, M. Milicic, F. Wolter, Foundations of instance level updates in expressive description logics, Artif. Intell. 175 (2011) 2170–2197.
[23] Z. Wang, K. Wang, R.W. Topor, Revising general knowledge bases in description logics, in: Proceedings of the 12th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010), AAAI Press, 2010.
[24] Z. Wang, K. Wang, R.W. Topor, DL-Lite ontology revision based on an alternative semantic characterization, ACM Trans. Comput. Log. 16 (2015) 31:1–31:37.
[25] B. Konev, D. Walther, F. Wolter, Forgetting and uniform interpolation in large-scale description logic terminologies, in: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009), 2009, pp. 830–835.
[26] Z. Wang, K. Wang, R.W. Topor, J.Z. Pan, Forgetting for knowledge bases in DL-Lite, Ann. Math. Artif. Intell. 58 (2010) 117–151.
[27] C. Lutz, F. Wolter, Foundations for uniform interpolation and forgetting in expressive description logics, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011), IJCAI/AAAI, 2011, pp. 989–995.
[28] K. Wang, Z. Wang, R.W. Topor, J.Z. Pan, G. Antoniou, Eliminating concepts and roles from ontologies in expressive descriptive logics, Comput. Intell. 30 (2014) 205–232.
[29] P. Koopmann, R.A. Schmidt, Forgetting and uniform interpolation for ALC-ontologies with ABoxes, in: DL-14, vol. 1193, 2014, pp. 245–257.
[30] N. Nikitina, S. Rudolph, (Non-)succinctness of uniform interpolants of general terminologies in the description logic EL, Artif. Intell. 215 (2014) 120–140.
[31] P. Koopmann, R.A. Schmidt, Uniform interpolation and forgetting for ALC ontologies with ABoxes, in: Proceedings of the 29th National Conference on Artificial Intelligence (AAAI 2015), AAAI Press, 2015, pp. 175–181.
[32] M. Arenas, E. Botoeva, D. Calvanese, V. Ryzhikov, Exchanging OWL 2 QL knowledge bases, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013), AAAI Press, 2013, pp. 703–710.
[33] M. Arenas, E. Botoeva, D. Calvanese, V. Ryzhikov, Knowledge base exchange: the case of OWL 2 QL, Artif. Intell. 238 (2016) 11–62.
[34] E. Botoeva, R. Kontchakov, V. Ryzhikov, F. Wolter, M. Zakharyaschev, Games for query inseparability of description logic knowledge bases, Artif. Intell. 234 (2016) 78–119.
[35] P. Shvaiko, J. Euzenat, Ontology matching: state of the art and future challenges, IEEE Trans. Knowl. Data Eng. 25 (2013) 158–176.
[36] A. Schaerf, Query Answering in Concept-Based Knowledge Representation Systems: Algorithms, Complexity, and Semantic Issues, Ph.D. thesis, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", 1994.
[37] E. Botoeva, R. Kontchakov, V. Ryzhikov, F. Wolter, M. Zakharyaschev, Query inseparability for description logic knowledge bases, in: Proceedings of the 14th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2014), 2014, pp. 238–247.
[38] C. Lutz, F. Wolter, Deciding inseparability and conservative extensions in the description logic EL, J. Symb. Comput. 45 (2010) 194–228.
[39] R. Kontchakov, L. Pulina, U. Sattler, T. Schneider, P. Seimer, F. Wolter, M. Zakharyaschev, Minimal module extraction from DL-Lite ontologies using QBF solvers, in: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009), 2009, pp. 836–840.
[40] M. Bienvenu, R. Rosati, Query-based comparison of mappings in ontology-based data access, in: Proceedings of the 15th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2016), 2016, pp. 197–206.
[41] E. Botoeva, B. Konev, C. Lutz, V. Ryzhikov, F. Wolter, M. Zakharyaschev, Inseparability and conservative extensions of description logic ontologies: a survey, in: 12th Reasoning Web International Summer School Tutorial Lectures (RW 2016), 2016, pp. 27–89.
[42] S. Ghilardi, C. Lutz, F. Wolter, Did I damage my ontology? A case for conservative extensions in description logics, in: P. Doherty, J. Mylopoulos, C. Welty (Eds.), Proceedings of the 10th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2006), 2006, pp. 187–197.
[43] J.C. Jung, C. Lutz, M. Martel, T. Schneider, F. Wolter, Conservative extensions in guarded and two-variable fragments, in: Proceedings of the 39th International Coll. on Automata, Languages and Programming (ICALP), 2017, pp. 108:1–108:14.
[44] E. Botoeva, C. Lutz, V. Ryzhikov, F. Wolter, M. Zakharyaschev, Query-based entailment and inseparability for ALC ontologies, in: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), 2016, pp. 1001–1007.
[45] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P.F. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation and Applications, Cambridge University Press, 2003.
[46] F. Baader, S. Brandt, C. Lutz, Pushing the EL envelope, in: Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005), 2005, pp. 364–369.
[47] U. Hustadt, B. Motik, U. Sattler, Reasoning in description logics by a reduction to disjunctive Datalog, J. Autom. Reason. 39 (2007) 351–384.
[48] Y. Kazakov, Consequence-driven reasoning for Horn SHIQ ontologies, in: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009), 2009, pp. 2040–2045.
[49] B. Glimm, C. Lutz, I. Horrocks, U. Sattler, Answering conjunctive queries in the $\mathcal{SHIQ}$ description logic, J. Artif. Intell. Res. 31 (2008) 150–197.
[50] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati, Data complexity of query answering in description logics, in: Proceedings of the 10th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2006), 2006, pp. 260–270.
[51] D. Calvanese, T. Eiter, M. Ortiz, Answering regular path queries in expressive description logics: an automata-theoretic approach, in: Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI), 2007, pp. 391–396.
[52] A.K. Chandra, P.M. Merlin, Optimal implementation of conjunctive queries in relational data bases, in: Proceedings of the 9th ACM Symposium on Theory of Computing (STOC'77), 1977, pp. 77–90.

[53] C. Lutz, The complexity of conjunctive query answering in expressive description logics, in: A. Armando, P. Baumgartner, G. Dowek (Eds.), Proceedings of the 4th International Joint Conference on Automated Reasoning (IJCAR 2008), in: LNAI, vol. 5195, Springer, 2008, pp. 179–193.
[54] C.C. Chang, H.J. Keisler, Model Theory, Studies in Logic and the Foundations of Mathematics, vol. 73, Elsevier, 1990.
[55] B.C. Grau, I. Horrocks, M. Krötzsch, C. Kupke, D. Magka, B. Motik, Z. Wang, Acyclicity notions for existential rules and their application to query answering in ontologies, J. Artif. Intell. Res. 47 (2013) 741–808.
[56] P. van Emde Boas, The convenience of tiling, in: A. Sorbi (Ed.), Complexity, Logic and Recursion Theory, in: Lecture Notes in Pure and Applied Mathematics, vol. 187, Marcel Dekker Inc., 1997, pp. 331–363.
[57] C. Lutz, F. Wolter, Non-uniform data complexity of query answering in description logics, in: Proceedings of the 13th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2012), 2012, pp. 297–307.
[58] M.O. Rabin, Automata on Infinite Objects and Church's Problem, American Mathematical Society, Boston, MA, USA, 1972.
[59] M.Y. Vardi, Reasoning about the past with two-way automata, in: Proceedings of the 25th International Coll. on Automata, Languages and Programming (ICALP), in: Lecture Notes in Computer Science, vol. 1443, Springer, 1998, pp. 628–641.
[60] A.K. Chandra, D. Kozen, L.J. Stockmeyer, Alternation, J. ACM 28 (1981) 114–133.
[61] M. Bienvenu, C. Lutz, F. Wolter, Query containment in description logics reconsidered, in: Proceedings of the 13th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2012), 2012, pp. 221–231.
[62] M. Bienvenu, B. ten Cate, C. Lutz, F. Wolter, Ontology-based data access: a study through Disjunctive Datalog, CSP, and MMSNP, ACM Trans. Database Syst. 39 (2014) 33:1–33:44.
[63] M. Bienvenu, P. Hansen, C. Lutz, F. Wolter, First-order rewritability and containment of conjunctive queries in Horn description logics, in: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), 2016.
[64] F. Baader, M. Bienvenu, C. Lutz, F. Wolter, Query and predicate emptiness in ontology-based data access, J. Artif. Intell. Res. 56 (2016) 1–59.
[65] D.E. Muller, P.E. Schupp, Alternating automata on infinite trees, Theor. Comput. Sci. 54 (1987) 267–276.
[66] W. Thomas, Languages, automata, and logic, in: Handbook of Formal Language Theory, III, 1997, pp. 389–455.
[67] M. Bienvenu, C. Lutz, F. Wolter, First-order rewritability of atomic queries in Horn description logics, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013), 2013, pp. 754–760.
[68] J.C. Jung, C. Lutz, M. Martel, T. Schneider, Query conservative extensions in horn description logics with inverse roles, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017), 2017, pp. 1116–1122.
[69] A. Tarski, A. Mostowski, R. Robinson, Undecidable Theories, North-Holland, 1953.
[70] W. Rautenberg, A Concise Introduction to Mathematical Logic, Springer, 2010.
[71] J.A. Goguen, R.M. Burstall, Institutions: abstract model theory for specification and programming, J. ACM 39 (1992) 95–146.
[72] T. Maibaum, Conservative extensions, interpretations between theories and all that!, in: Proceedings of the 7th International Conference on Theory and Practice of Software Development (TAPSOFT), in: LNCS, Springer Verlag, 1997.
[73] J.G.R. Diaconescu, P. Stefaneas, Logical support for modularisation, in: G. Huet, G. Plotkin (Eds.), Logical Environments, 1993.
[74] S. Woltran, Equivalence between extended datalog programs – a brief survey, in: Datalog Reloaded, 2010, pp. 106–119.
[75] V. Lifschitz, D. Pearce, A. Valverde, Strongly equivalent logic programs, ACM Trans. Comput. Log. 2 (2001) 526–541.
[76] T. Eiter, M. Fink, S. Woltran, Semantical characterizations and complexity of equivalences in answer set programming, ACM Trans. Comput. Log. 8 (2007) 17.
[77] A. Harrison, V. Lifschitz, D. Pearce, A. Valverde, Infinitary equilibrium logic and strongly equivalent logic programs, Artif. Intell. 246 (2017) 22–33.
[78] B. Konev, M. Ludwig, F. Wolter, Logical difference computation with CEX2.5, in: Proceedings of the 6th International Joint Conference on Automated Reasoning (IJCAR 2012), in: Lecture Notes in Computer Science, Springer, Berlin/Heidelberg, Germany, 2012, pp. 371–377.