Query-Based Entailment and Inseparability for $\mathcal{ALC}$ Ontologies

Elena Botoeva,[1] Carsten Lutz,[2] Vladislav Ryzhikov,[1] Frank Wolter,[3] Michael Zakharyaschev[4]
[1]Faculty of Computer Science, Free University of Bozen-Bolzano
[2]Fachbereich Informatik, University of Bremen
[3]Department of Computer Science, University of Liverpool
[4]Department of Computer Science, Birkbeck, University of London

Abstract

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

1 Introduction

In recent years, data access using description logic (DL) TBoxes has become one of the most important applications of DLs [Poggi *et al.*, 2008; Bienvenu and Ortiz, 2015], where the underlying idea is to use a TBox to specify semantics and background knowledge for the data (stored in an ABox), and thereby derive more complete query answers. A major research effort has led to the development of efficient algorithms and tools for a number of DLs ranging from DL-Lite [Calvanese *et al.*, 2007; Rodriguez-Muro *et al.*, 2013] via more expressive Horn DLs such as Horn-$\mathcal{ALC}$ [Eiter *et al.*, 2012; Trivela *et al.*, 2015] to DLs with all Boolean constructors such as $\mathcal{ALC}$ [Kollia and Glimm, 2013; Zhou *et al.*, 2015].

While query answering with DLs is now well-developed, this is much less the case for reasoning services that support ontology engineering and target query answering as an application. In ontology versioning, for example, one would like to know whether two versions of an ontology give the same answers to all queries formulated over a given vocabulary of interest, which means that the newer version can safely replace the older one [Konev *et al.*, 2012]. Similarly, if one wants to know whether an ontology can be safely replaced by a smaller subset (module), it is the answers to all queries that should be preserved [Kontchakov *et al.*, 2010]. In this context, the fundamental relationship between ontologies is thus not whether they are logically equivalent (have the same models), but whether they give the same answers to any relevant query. The resulting entailment problem can be formalized in two ways, with different applications. First, given a class $\mathcal{Q}$ of queries, knowledge bases (KBs) $\mathcal{K}_1$ and $\mathcal{K}_2$, and a signature Σ of relevant concept and role names, we say that $\mathcal{K}_1$ Σ-$\mathcal{Q}$-*entails* $\mathcal{K}_2$ if the answers to any Σ-query in $\mathcal{Q}$ over $\mathcal{K}_2$ are contained in the answers to the same query over $\mathcal{K}_1$. Further, $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ-$\mathcal{Q}$-*inseparable* if they Σ-$\mathcal{Q}$-entail each other. Note that a KB includes an ABox, and thus this notion of entailment is appropriate if the data is known and does not change frequently. Applications include data-oriented KB versioning and KB module extraction, KB forgetting [Wang *et al.*, 2014], and knowledge exchange [Arenas *et al.*, 2013].

If the data is not known or changes frequently, it is not KBs that should be compared, but TBoxes. Given a pair $\Theta = (\Sigma_1, \Sigma_2)$ that specifies a relevant signature Σ_1 for ABoxes and Σ_2 for queries, we say that a TBox $\mathcal{T}_1$ Θ-$\mathcal{Q}$-*entails* a TBox $\mathcal{T}_2$ if, for every Σ_1-ABox $\mathcal{A}$, the KB $(\mathcal{T}_1, \mathcal{A})$ Σ_2-$\mathcal{Q}$-entails $(\mathcal{T}_2, \mathcal{A})$. $\mathcal{T}_1$ and $\mathcal{T}_2$ are Θ-$\mathcal{Q}$-*inseparable* if they Θ-$\mathcal{Q}$-entail each other. Applications include data-oriented TBox versioning, TBox modularization and TBox forgetting [Kontchakov *et al.*, 2010].

In this paper, we concentrate on the most important choices for $\mathcal{Q}$, conjunctive queries (CQs) and unions thereof (UCQs); we also consider the practically relevant classes of rooted CQs (rCQs) and UCQs (rUCQs), in which every variable is connected to an answer variable. So far, CQ-entailment has been studied for Horn DL KBs [Botoeva *et al.*, 2014], $\mathcal{EL}$ TBoxes [Lutz and Wolter, 2010; Konev *et al.*, 2012], DL-Lite TBoxes [Kontchakov *et al.*, 2009], and also for OBDA specifications, that is, DL-Lite TBoxes with mappings [Bienvenu and Rosati, 2015]. No results are available for non-Horn DLs (neither in the KB nor in the TBox case) and for expressive Horn DLs in the TBox case. In particular, query entailment in non-Horn DLs has had the reputation of being a technically challenging problem.

This paper makes a first breakthrough into understanding query entailment and inseparability in these cases, with the main results summarized in Figures 1 and 2 (those marked with $(\star)$ are from [Botoeva *et al.*, 2014]). Three of them came as a real surprise to us. First, it turned out that CQ- and rCQ-entailment between $\mathcal{ALC}$ KBs is undecidable, even when the first KB is formulated in *Horn*-$\mathcal{ALC}$ (in fact, $\mathcal{EL}$) and without any signature restriction. This should be contrasted with the decidability of subsumption-based entailment between $\mathcal{ALC}$

Queries	$\mathcal{ALC}$	Horn-$\mathcal{ALC}$ to $\mathcal{ALC}$	$\mathcal{ALC}$ to Horn-$\mathcal{ALC}$	Horn-$\mathcal{ALC}$
CQ	undecidable		?	=ExpTime$^{(\star)}$
UCQ	?			
rCQ	undecidable		$\leq$2ExpTime	=ExpTime$^{(\star)}$
rUCQ	$\leq$2ExpTime			

Figure 1: KB query entailment.

Queries	$\mathcal{ALC}$	Horn-$\mathcal{ALC}$ to $\mathcal{ALC}$	$\mathcal{ALC}$ to Horn-$\mathcal{ALC}$	Horn-$\mathcal{ALC}$
CQ	undecidable		?	=2ExpTime
UCQ	?			
rCQ	undecidable		=ExpTime	=ExpTime
rUCQ	?			

Figure 2: TBox query entailment.

TBoxes [Ghilardi *et al.*, 2006] and of CQ-entailment between *Horn-$\mathcal{ALC}$* KBs [Botoeva *et al.*, 2014]. The second surprising result is that entailment between $\mathcal{ALC}$ KBs becomes decidable when CQs are replaced with rUCQs. For $\mathcal{ALC}$ TBoxes, CQ- and rCQ-entailment are undecidable as well. We obtain decidability for *Horn-$\mathcal{ALC}$* TBoxes (where CQ- und UCQ-entailments coincide) using the fact that non-entailment is always witnessed by tree-shaped ABoxes. As another surprise, CQ-entailment of *Horn-$\mathcal{ALC}$* TBoxes is 2ExpTime-complete while rCQ-entailment is only ExpTime-complete. This should be contrasted with the $\mathcal{EL}$ case, where both problems are ExpTime-complete [Lutz and Wolter, 2010]. All upper bounds and most lower bounds hold also for inseparability in place of entailment. A model-theoretic foundation for these results is a characterization of query entailment between KBs and TBoxes in terms of (partial) homomorphisms, which, in particular, enables the use of tree automata techniques to establish the upper bounds in Figs. 1 and 2. Omitted proofs are available in the full version [Botoeva *et al.*, 2016].

2 Preliminaries

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

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

We use $\bot$, $C_1 \sqcup C_2$ and $\forall R.C$ as abbreviations for $\neg\top$, $\neg(\neg C_1 \sqcap \neg C_2)$ and $\neg\exists R.\neg C$, respectively. A *concept inclusion* (CI) takes the form $C \sqsubseteq D$, where C and D are concepts. An *$\mathcal{ALC}$ TBox* is a finite set of CIs. In a *Horn-$\mathcal{ALC}$ TBox*, no concept of the form $\neg C$ occurs negatively and no $\exists R.\neg C$ occurs positively [Hustadt *et al.*, 2005; Kazakov, 2009]. An *$\mathcal{EL}$ TBox* does not contain $\neg$ at all. An *ABox*, $\mathcal{A}$, is a finite set of *assertions* of the form $A_k(a_i)$ or $R_k(a_i, a_j)$; $\mathsf{ind}(\mathcal{A})$ is the set of individual names in $\mathcal{A}$. Taken together, $\mathcal{T}$ and $\mathcal{A}$ form a *knowledge base* (KB) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$; we set $\mathsf{ind}(\mathcal{K}) = \mathsf{ind}(\mathcal{A})$.

The semantics is defined as usual based on interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ that comply with the *standard name assumption* in the sense that $a_i^{\mathcal{I}} = a_i$ [Baader *et al.*, 2003]. We write $\mathcal{I} \models \alpha$ if an inclusion or assertion α is true in $\mathcal{I}$. If $\mathcal{I} \models \alpha$, for all $\alpha \in \mathcal{T} \cup \mathcal{A}$, then we call $\mathcal{I}$ a *model* of $\mathcal{K}$ and write $\mathcal{I} \models \mathcal{K}$. $\mathcal{K}$ is *consistent* if it has a model; we then also say that $\mathcal{A}$ *is consistent with* $\mathcal{T}$. $\mathcal{K} \models \alpha$ means that $\mathcal{I} \models \alpha$ for all $\mathcal{I} \models \mathcal{K}$.

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

A tuple $\boldsymbol{a}$ in $\mathsf{ind}(\mathcal{K})$ is a *certain answer to a UCQ* $\boldsymbol{q}(\boldsymbol{x})$ *over* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for all $\mathcal{I} \models \mathcal{K}$; in this case we write $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$. If $\boldsymbol{x} = \emptyset$, the answer to $\boldsymbol{q}$ is 'yes' if $\mathcal{K} \models \boldsymbol{q}$ and 'no' otherwise. The problem of checking whether a tuple is a certain answer to a given (U)CQ over a given $\mathcal{ALC}$ KB is known to be ExpTime-complete for combined complexity [Lutz, 2008]. The ExpTime lower bound actually holds for *Horn-$\mathcal{ALC}$* [Krötzsch *et al.*, 2013].

A set $\boldsymbol{M}$ of models of a KB $\mathcal{K}$ is called *complete for* $\mathcal{K}$ if, for every UCQ $\boldsymbol{q}(\boldsymbol{x})$, we have $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$ iff $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for all $\mathcal{I} \in \boldsymbol{M}$. We call an interpretation $\mathcal{I}$ a *ditree interpretation* if the directed graph $G_{\mathcal{I}}$ with nodes $d \in \Delta^{\mathcal{I}}$ and edges $(d, e) \in R^{\mathcal{I}}$, for some R, is a tree and $R^{\mathcal{I}} \cap S^{\mathcal{I}} = \emptyset$, for any distinct roles R and S. $\mathcal{I}$ has *outdegree* n if $G_{\mathcal{I}}$ has outdegree n. A model $\mathcal{I}$ of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is *forest-shaped* if $\mathcal{I}$ is the disjoint union of ditree interpretations $\mathcal{I}_a$ with root a, for $a \in \mathsf{ind}(\mathcal{A})$, extended with all $R(a, b) \in \mathcal{A}$. The *outdegree* of $\mathcal{I}$ is the maximum outdegree of the $\mathcal{I}_a$. It is well known that the class $\boldsymbol{M}_{\mathcal{K}}^{fo}$ of all forest-shaped models of an $\mathcal{ALC}$ KB $\mathcal{K}$ of outdegree bounded by $|\mathcal{T}|$ is complete for $\mathcal{K}$ [Lutz, 2008]. If $\mathcal{K}$ is a *Horn-$\mathcal{ALC}$* KB, then a single member $\mathcal{I}_{\mathcal{K}}$ of $\boldsymbol{M}_{\mathcal{K}}^{fo}$ is complete for $\mathcal{K}$. $\mathcal{I}_{\mathcal{K}}$ is constructed using the standard chase procedure and called the *canonical model of* $\mathcal{K}$.

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

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

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

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

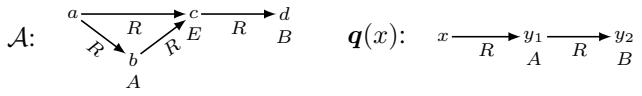
Now, let $\Sigma_1 = \{E, B\}$, $\mathcal{T}_1 = \emptyset$, and $\mathcal{T}_1' = \{E \sqsubseteq \exists R.B\}$. Let $\mathcal{A}_1 = \{E(a)\}$, $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A}_1)$, and $\mathcal{K}_1' = (\mathcal{T}_1', \mathcal{A}_1)$.

Then $\mathcal{K}_1$ and $\mathcal{K}_1'$ are Σ_1-rUCQ inseparable but not Σ_1-CQ inseparable. In fact, $\mathcal{K}_1' \models \exists x B(x)$ but $\mathcal{K}_1 \not\models \exists x B(x)$.

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

We only consider ABoxes that are consistent with both TBoxes because the problem whether a Σ_1-ABox consistent with $\mathcal{T}_2$ is also consistent with $\mathcal{T}_1$ is well understood: it is mutually polynomially reducible with the containment problem for ontology-mediated queries with CQs of the form $\exists x A(x)$, which is NEXPTIME-complete for $\mathcal{ALC}$ and EXPTIME-complete for *Horn-*$\mathcal{ALC}$ [Bienvenu *et al.*, 2012; 2014].

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



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

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

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

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

We now give model-theoretic criteria for Σ-entailment between KBs. The *product* $\prod \boldsymbol{\mathcal{I}}$ of a set $\boldsymbol{\mathcal{I}}$ of interpretations is defined as usual in model theory [Chang and Keisler, 1990, page 405]. Note that, for any CQ $\boldsymbol{q}(\boldsymbol{x})$ and any tuple $\boldsymbol{a}$ of individual names, $\prod \boldsymbol{\mathcal{I}} \models \boldsymbol{q}(\boldsymbol{a})$ iff $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for each $\mathcal{I} \in \boldsymbol{\mathcal{I}}$.

Suppose $\mathcal{I}_i$ is an interpretation for a KB $\mathcal{K}_i$, $i = 1, 2$. A function $h\colon \Delta^{\mathcal{I}_2} \to \Delta^{\mathcal{I}_1}$ is called a Σ-*homomorphism* if $u \in A^{\mathcal{I}_2}$ implies $h(u) \in A^{\mathcal{I}_1}$ and $(u, v) \in R^{\mathcal{I}_2}$ implies $(h(u), h(v)) \in R^{\mathcal{I}_1}$ for all $u, v \in \Delta^{\mathcal{I}_2}$, Σ-concept names A, and Σ-role names R, and $h(a) = a$ for all $a \in \mathsf{ind}(\mathcal{K}_2)$. It is known from database theory that homomorphisms characterize CQ-containment [Chandra and Merlin, 1977]. For KB Σ-query entailment, finite partial homomorphisms are required. We say that $\mathcal{I}_2$ is $n\Sigma$-*homomorphically embeddable into* $\mathcal{I}_1$ if, for any subinterpretation $\mathcal{I}_2'$ of $\mathcal{I}_2$ with $|\Delta^{\mathcal{I}_2'}| \leq n$, there is a Σ-homomorphism from $\mathcal{I}_2'$ to $\mathcal{I}_1$. If, additionally, we require $\mathcal{I}_2'$ to be Σ-connected then $\mathcal{I}_2$ is said to be *con-*$n\Sigma$-*homomorphically embeddable into* $\mathcal{I}_1$.

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

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

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

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

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

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

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

If both $\boldsymbol{M}_i$ are finite and contain only finite interpretations, then Theorem 6 provides a decision procedure for KB entailment. This applies, for example, to KBs with acyclic classical TBoxes [Baader *et al.*, 2003], and to KBs for which the chase terminates [Grau *et al.*, 2013].

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

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

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

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

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

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

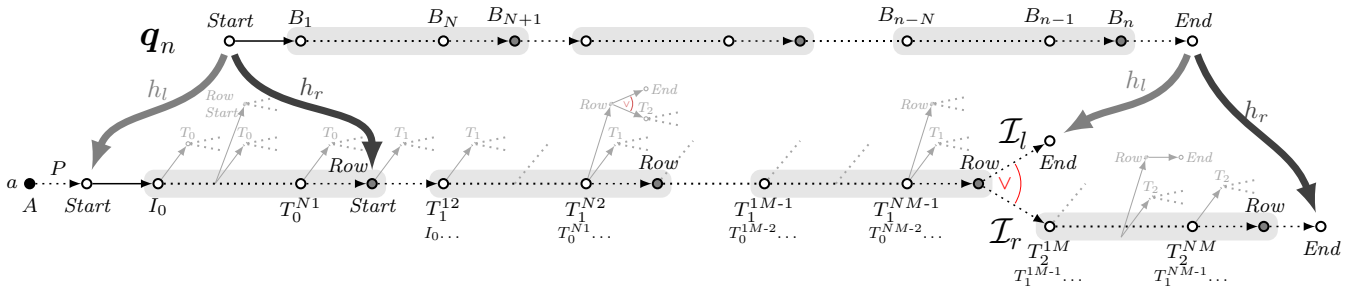


Figure 3: The structure of models $\mathcal{I}_l$ and $\mathcal{I}_r$ of $\mathcal{K}_2$, and homomorphisms $h_l\colon \boldsymbol{q}_n \to \mathcal{I}_l$ and $h_r\colon \boldsymbol{q}_n \to \mathcal{I}_r$.

of type T with *right*$(T) = W$; and every (i, M), for $i \leq N$, is covered by a tile of type T with *up*$(T) = C$.

Given an instance of this problem, we first describe a KB $\mathcal{K}_2 = (\mathcal{T}_2, \{A(a)\})$ that uses (among others) 3 concept names T_k, $k = 0, 1, 2$, for each tile type $T \in \mathfrak{T}$. If a point x in a model $\mathcal{I}$ of $\mathcal{K}_2$ is in T_k and *right*$(T) =$ *left*(T'), then x has an R-successor in T'_k. Thus, branches of $\mathcal{I}$ define (possibly infinite) horizontal rows of tilings with $\mathfrak{T}$. If a branch contains a point $y \in T_k$ with *right*$(T) = W$, then this y can be the last point in the row, which is indicated by an R-successor $z \in$ *Row* of y. In turn, z has R-successors in all $T_{(k+1) \bmod 3}$ that can be possible beginnings of the next row of tiles. To coordinate the *up* and *down* colours between the rows—which will be done by the CQs separating $\mathcal{K}_1$ and $\mathcal{K}_2$— we make every $x \in T_k$, starting from the second row, an instance of all $T'_{(k-1) \bmod 3}$ with *down*$(T) =$ *up*(T'). The row started by $z \in$ *Row* can be the last one in the tiling, in which case we require that each of its tiles T has *up*$(T) = C$. After the point in *Row* indicating the end of the final row, we add an R-successor in *End* for the end of tiling. The beginning of the first row is indicated by a P-successor in *Start* of the ABox element a, after which we add an R-successor in I_0 for the given initial tile type I; see the lowest branch in Fig. 3. To generate a tree with all possible branches described above, we only require $\mathcal{EL}$ axioms of the form $E \sqsubseteq D$ and $E \sqsubseteq \exists S.D$.

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

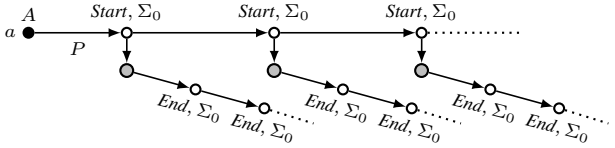
$$\exists \boldsymbol{x} \big(\mathit{Start}(x_0) \land \bigwedge_{i=0}^{n} R(x_i, x_{i+1}) \land \bigwedge_{i=1}^{n} B_i(x_i) \land \mathit{End}(x_{n+1}) \big)$$

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

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

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

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



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

This proves (*i*) for Σ-CQ entailment. For Σ-rCQ entailment, we slightly modify the construction, in particular, by adding $R(a, a)$ and *Row*(a) to the ABox $\{A(a)\}$, and a conjunct $R(y, x_0)$ with a free y to $\boldsymbol{q}_n$. (The loop $R(a, a)$ plays roughly the same role as the path between two *Start*-points in Fig. 3.) To prove (*ii*), we take $\mathcal{K}'_2 = \mathcal{K}_2 \cup \mathcal{K}_1$ and show that $\mathcal{K}_1$ Σ-CQ entails $\mathcal{K}_2$ iff $\mathcal{K}_1$ and $\mathcal{K}'_2$ are Σ-CQ inseparable. Finally, we prove (*iii*) by replacing non-Σ symbols in $\mathcal{K}_2$ with complex $\mathcal{ALC}$-concepts that cannot be used in CQs and extending the TBoxes appropriately; cf. [Lutz and Wolter, 2012, Lemma 21]. □

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

Theorem 9. (*i*) *The problem whether a Horn-$\mathcal{ALC}$ TBox Θ-$\mathcal{Q}$ entails an $\mathcal{ALC}$ TBox is undecidable for $\mathcal{Q} \in \{\mathit{CQ}, \mathit{rCQ}\}$.*

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

(*iii*) *For CQs,* (*i*) *and* (*ii*) *hold for full ABox signatures and for $\Theta = (\Sigma_1, \Sigma_2)$ with $\Sigma_1 = \Sigma_2$.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof. It is known that Horn-$\mathcal{ALC}$ is unravelling tolerant, that is, $(\mathcal{T}, \mathcal{A}) \models C(a)$ for a Horn-$\mathcal{ALC}$ TBox $\mathcal{T}$ and $\mathcal{EL}$-concept C iff $(\mathcal{T}, \mathcal{A}') \models C(a)$ for a finite sub-ABox $\mathcal{A}'$ of the tree-unravelling of $\mathcal{A}$ at a [Lutz and Wolter, 2012]. Thus, any witness ABox for non-entailment w.r.t. $\mathcal{EL}$-instance queries can be transformed into a tree-shaped witness ABox. The result follows by observing that if $\mathcal{T}_1$ does not Θ-rCQ-entail $\mathcal{T}_2$, then this is witnessed by an $\mathcal{EL}$-instance query and by applying Theorem 10 to the KBs. The bound on the outdegree is obtained by a careful analysis of derivations. □

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

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

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

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

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

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

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

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

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

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

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

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

Note that the ExpTime lower bound holds already for entailment of $\mathcal{EL}$ TBoxes and $\Sigma_1 = \Sigma_2$ [Lutz and Wolter, 2010]. We now study the non-rooted case, starting with an analogue of Theorem 12. As expected, moving to unrestricted queries corresponds to moving to unrestricted homomorphisms.

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

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

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

7 Future Work

We have made first steps towards understanding query entailment and inseparability for KBs and TBoxes in expressive DLs. Many problems remain to be addressed. From a theoretical viewpoint, it would be of interest to solve the open problems in Figures 1 and 2, and also consider other expressive DLs such as *DL-Lite*$^{\mathcal{H}}_{bool}$ [Artale *et al.*, 2009] or $\mathcal{ALCI}$. For example, if Theorem 10 could be generalized to UCQs (and Σ-homomorphisms), we would obtain a 2ExpTime upper bound for UCQ-entailment between $\mathcal{ALC}$ KBs using the same technique as for rUCQs. Also, our undecidability proof goes through for *DL-Lite*$^{\mathcal{H}}_{bool}$, but the other cases remain open. From a practical viewpoint, our model-theoretic criteria for query entailment are a good starting point for developing algorithms for approximations of query entailment based on simulations. Our undecidability and complexity results also indicate that rUCQ-entailment is more amenable to practical algorithms than, say, CQ-entailment and can be used as an approximation of the latter.

Acknowledgments. This work has been supported by the EU IP project Optique, grant n. FP7-318338, DFG grant LU 1417/2-1, and EPSRC UK grants EP/M012646/1 and EP/M012670/1 (iTract).

References

[Arenas *et al.*, 2013] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Exchanging OWL 2 QL knowledge bases. In *Proc. of IJCAI*. AAAI Press, 2013.

[Artale *et al.*, 2009] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The DL-Lite family and relations. *J. of Art. Intel. Research*, 2009.

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

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *RW*, 2015.

[Bienvenu and Rosati, 2015] Meghyn Bienvenu and Riccardo Rosati. Query-based comparison of OBDA specifications. In *Proc. of DL*, volume 1350. CEUR-WS, 2015.

[Bienvenu *et al.*, 2012] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Query containment in description logics reconsidered. In *Proc. of KR*, pages 221–231, 2012.

[Bienvenu *et al.*, 2014] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through Disjunctive Datalog, CSP, and MMSNP. *ACM Trans. on Database Systems*, 39(4), 2014.

[Botoeva *et al.*, 2014] Elena Botoeva, Roman Kontchakov, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Query inseparability for description logic knowledge bases. In *Proc. of KR*, pages 238–247, 2014.

[Botoeva *et al.*, 2016] Elena Botoeva, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Query-based entailment and inseparability for ALC ontologies (full version). Corr technical report, 2016. Available at arxiv.org/abs/1604.04164.

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

[Chandra and Merlin, 1977] Ashok K. Chandra and Philip M. Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Proc. of STOC*, pages 77–90, 1977.

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

[Eiter *et al.*, 2012] Thomas Eiter, Magdalena Ortiz, Mantas Simkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for Horn-SHIQ plus rules. In *Proc. of AAAI*, pages 726–733. AAAI Press, 2012.

[Ghilardi *et al.*, 2006] S. Ghilardi, C. Lutz, and F. Wolter. Did I damage my ontology? A case for conservative extensions in description logics. In *Proc. of KR*, 2006.

[Grau *et al.*, 2013] Bernardo Cuenca Grau, Ian Horrocks, Markus Krötzsch, Clemens Kupke, Despoina Magka, Boris Motik, and Zhe Wang. Acyclicity notions for existential rules and their application to query answering in ontologies. *J. of Art. Intel. Research*, 47:741–808, 2013.

[Hustadt *et al.*, 2005] Ulrich Hustadt, Boris Motik, and Ulrike Sattler. Data complexity of reasoning in very expressive description logics. In *Proc. IJCAI*, 2005.

[Kazakov, 2009] Yevgeny Kazakov. Consequence-driven reasoning for Horn SHIQ ontologies. In *Proc. of IJCAI*, 2009.

[Kollia and Glimm, 2013] Ilianna Kollia and Birte Glimm. Optimizing SPARQL query answering over OWL ontologies. *J. of Art. Intel. Research*, 48:253–303, 2013.

[Konev *et al.*, 2012] Boris Konev, Michel Ludwig, Dirk Walther, and Frank Wolter. The logical difference for the lightweight description logic EL. *J. of Art. Intel. Research*, 44:633–708, 2012.

[Kontchakov *et al.*, 2009] R. Kontchakov, L. Pulina, U. Sattler, T. Schneider, P. Seimer, F. Wolter, and M. Zakharyaschev. Minimal module extraction from *DL-Lite* ontologies using QBF solvers. In *Proc. of IJCAI*, 2009.

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

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

[Lutz and Wolter, 2010] Carsten Lutz and Frank Wolter. Deciding inseparability and conservative extensions in the description logic EL. *J. of Symb. Comp.*, 45(2), 2010.

[Lutz and Wolter, 2012] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of KR*, pages 297–307, 2012.

[Lutz, 2008] Carsten Lutz. The complexity of conjunctive query answering in expressive description logics. In *Proc. of IJCAR*, pages 179–193. Springer, 2008.

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

[Rodriguez-Muro *et al.*, 2013] Mariano Rodriguez-Muro, Roman Kontchakov, and Michael Zakharyaschev. Ontology-based data access: Ontop of databases. In *Proc. of ISWC*, pages 558–573. Springer, 2013.

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

[Wang *et al.*, 2014] Kewen Wang, Zhe Wang, Rodney W. Topor, Jeff Z. Pan, and Grigoris Antoniou. Eliminating concepts and roles from ontologies in expressive descriptive logics. *Computational Intelligence*, 30(2), 2014.

[Zhou *et al.*, 2015] Yujiao Zhou, Bernardo Cuenca Grau, Yavor Nenov, Mark Kaminski, and Ian Horrocks. Pagoda: Pay-as-you-go ontology query answering using a datalog reasoner. *J. of Art. Intel. Research*, 54:309–367, 2015.