Query Conservative Extensions in Horn Description Logics with Inverse Roles

Jean Christoph Jung, Carsten Lutz, Mauricio Martel, and **Thomas Schneider**
Fachbereich Informatik, Universität Bremen, Germany

Abstract

We investigate the decidability and computational complexity of query conservative extensions in Horn description logics (DLs) with inverse roles. This is more challenging than without inverse roles because characterizations in terms of unbounded homomorphisms between universal models fail, blocking the standard approach to establishing decidability. We resort to a combination of automata and mosaic techniques, proving that the problem is 2ExpTime-complete in Horn-$\mathcal{ALCHIF}$ (and also in Horn-$\mathcal{ALC}$ and in $\mathcal{ELI}$). We obtain the same upper bound for deductive conservative extensions, for which we also prove a coNExpTime lower bound.

1 Introduction

In the past years, access of incomplete data mediated by description logic (DL) ontologies has gained increasing importance [Poggi *et al.*, 2008; Bienvenu and Ortiz, 2015]. The main idea is to specify domain knowledge and semantics of the data in the ontology, resulting in more complete answers to queries. Significant research activity has led to efficient algorithms and tools for a wide range of DLs such as DL-Lite [Calvanese *et al.*, 2007], more expressive Horn-DLs [Eiter *et al.*, 2012; Trivela *et al.*, 2015; Bienvenu *et al.*, 2016], and "full Boolean" DLs such as $\mathcal{ALC}$ [Kollia and Glimm, 2013; Zhou *et al.*, 2015].

In contrast to query answering, which is by now well-understood, there is a need to develop reasoning services for ontology engineering that are tailored towards query-centric applications and support tasks such as ontology versioning and module extraction from ontologies. For example, if one wants to safely replace an ontology with a new version or with a smaller subset of itself (a module), then the new ontology should preserve the answers to all queries over all ABoxes (which store the data) [Kontchakov *et al.*, 2010]. The same guarantee ensures that one can safely replace an ontology with another version in an application [Konev *et al.*, 2012]. In both cases, ontologies need to be tested not for their logical equivalence, but for giving the same answers to relevant queries over relevant datasets.

This requirement can be formalized using conservative extensions. In the following, we use the DL term *TBox* instead of *ontology*. A TBox $\mathcal{T}_2 \supseteq \mathcal{T}_1$ is a (Γ, Σ)-*query conservative extension* of a TBox $\mathcal{T}_1$, where Γ and Σ are signatures of concept/role names relevant for data and queries, respectively, if all Σ-queries give the same answers w.r.t. $\mathcal{T}_1$ and $\mathcal{T}_2$, for every Γ-ABox. Note that the subset relationship $\mathcal{T}_2 \supseteq \mathcal{T}_1$ is natural when replacing a TBox with a module, but not in versioning, so we might not want to insist on it. In this more general case, $\mathcal{T}_1$ and $\mathcal{T}_2$ are called (Γ, Σ)-*query inseparable*. Conservativity and inseparability of *TBoxes*, as defined above, are useful when knowledge is considered static and data changes frequently. Variants of these notions for *knowledge bases (KBs)*, which consist of a TBox and an ABox, can be used for applications with static data [Wang *et al.*, 2014; Arenas *et al.*, 2016].

We also consider the basic notion of query entailment: $\mathcal{T}_1$ (Γ, Σ)-*query entails* $\mathcal{T}_2$ if all Σ-queries give *at least* the answers w.r.t. $\mathcal{T}_1$ that they give w.r.t. $\mathcal{T}_2$, on any Γ-ABox. Query inseparability and conservativity are special cases of entailment: inseparability is bidirectional entailment and conservativity is entailment with the assumption that $\mathcal{T}_1 \subseteq \mathcal{T}_2$. It thus suffices to prove upper bounds for query entailment and lower bounds for conservative extensions. As a query language, we concentrate on conjunctive queries (CQs); since we work with Horn-DLs and quantify over the queries, this is equivalent to using unions of CQs (UCQs) or positive existential queries (PEQs). CQ entailment has been studied for various DLs [Kontchakov *et al.*, 2009; Lutz and Wolter, 2010; Konev *et al.*, 2012; Botoeva *et al.*, 2016c], also in the KB version [Botoeva *et al.*, 2016b; Botoeva *et al.*, 2016c] and for OBDA specifications [Bienvenu and Rosati, 2015], see also the survey [Botoeva *et al.*, 2016a]. Nevertheless, there is still a notable gap in our understanding of this notion: query entailment between TBoxes is poorly understood in Horn DLs with inverse roles, often considered a crucial feature, for which there do not seem to be any available results. This is for a reason: it has been observed in [Botoeva *et al.*, 2016a; Botoeva *et al.*, 2016b] that standard techniques for Horn DLs without inverse roles fail when inverse roles are added.

In fact, for Horn-DLs without inverse roles query entailment can be characterized by the existence of homomorphisms between universal models [Lutz and Wolter, 2010; Botoeva *et al.*, 2016a]. The resulting characterizations provide an important foundation for decision procedures, often based on tree automata [Botoeva *et al.*, 2016a]. In the presence of inverse roles, however, such characterizations are only correct if we require the existence of n-*bounded* homomorphisms, *for any* n [Botoeva *et al.*, 2016a; Botoeva *et al.*, 2016b]. It is not obvious how the existence of such infinite families of bounded homomorphisms can be verified using tree automata (or related techniques) and, consequently, decidability results for query conservative extensions in Horn-DLs with inverse roles are difficult to obtain. The only result we are aware of concerns inseparability *of KBs*, and it is proved using intricate game-theoretic techniques.

In this paper, we develop decision procedures for query entailment and related problems in Horn DLs with inverse roles. The main idea is to provide a more refined characterization, mixing unbounded and bounded homomorphisms and using bounded homomorphisms only in places where this is strictly necessary. We can then deal with the "unbounded part" using tree automata while the "bounded part" is addressed by precomputing relevant information using a mosaic technique. In this way, we establish decidability and a 2EXPTIME upper bound for query entailment (and thus inseparability and conservativity) in Horn-$\mathcal{ALCHIF}$. Together with lower bounds from [Botoeva *et al.*, 2016c], we get 2EXPTIME-completeness for all fragments of Horn-$\mathcal{ALCHIF}$ that contain $\mathcal{ELI}$ or Horn-$\mathcal{ALC}$.

We additionally study the case of deductive entailment between TBoxes, i.e., the question whether $\mathcal{T}_1$ entails at least the same concept and role inclusions as well as functionality assertions over Σ as $\mathcal{T}_2$. This problem too has not previously been studied for Horn DLs with inverse roles. We consider $\mathcal{ELHIF}_\bot$-TBoxes and show that deductive entailment is equivalent to a restricted version of query entailment. We obtain a model theoretic characterization, a decision procedure, and a 2EXPTIME upper complexity bound. We also give a CONEXPTIME lower bound.

Omitted proofs can be found in the long version here: www.informatik.uni-bremen.de/tdki/research/papers.html

2 Preliminaries

2.1 Horn-$\mathcal{ALCHIF}$

We introduce Horn-$\mathcal{ALCHIF}$, a member of the Horn-$\mathcal{SHIQ}$ family of DLs whose reasoning problems have been widely studied [Hustadt *et al.*, 2007; Krötzsch *et al.*, 2007; Eiter *et al.*, 2008; Kazakov, 2009; Lutz and Wolter, 2012; Ibáñez-García *et al.*, 2014]. Let $\mathsf{N_C}, \mathsf{N_R}, \mathsf{N_I}$ be sets of concept, role, and individual names. A *role* is either a role name r or an *inverse role* r^-. As usual, we identify $(r^-)^-$ and r, allowing to switch between roles names and their inverses easily. A *concept inclusion (CI)* is of the form $L \sqsubseteq R$, where L and R are concepts defined by the syntax rules

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

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

with A ranging over concept names and r over roles. A *role inclusion (RI)* is of the form $r \sqsubseteq s$ with r, s roles and a *functionality assertion (FA)* is of the form $\mathsf{func}(r)$ with r a role. $\mathcal{ELI}_\bot$*-concepts* are expressions that are built according to the syntax rule for L above, but do not use "$\sqcup$".

A *Horn-$\mathcal{ALCHIF}$ TBox* $\mathcal{T}$ is a set of CIs, RIs, and FAs. An $\mathcal{ELHIF}_\bot$ *TBox* is a set of $\mathcal{ELI}_\bot$-CIs, RIs, and FAs. To avoid dealing with rather messy technicalities that do neither seem to be very illuminating from a theoretical viewpoint nor too useful from a practical one,[1] we generally assume that functional roles cannot have any subroles, that is, $r \sqsubseteq s \in \mathcal{T}$ implies $\mathsf{func}(s) \notin \mathcal{T}$. We conjecture that our main results also hold without that restriction. An *ABox* $\mathcal{A}$ is a non-empty set of *concept and role assertions* of the form $A(a)$ and $r(a, b)$, where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $a, b \in \mathsf{N_I}$. We write $\mathsf{ind}(\mathcal{A})$ for the set of individuals in $\mathcal{A}$.

The semantics is defined as usual in terms of interpretations $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ complying with the standard name assumption, i.e., $a^\mathcal{I} = a$ for all $a \in \mathsf{N_I}$ [Baader *et al.*, 2017]. An interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ if it satisfies all inclusions and assertions in it, and likewise for ABoxes. $\mathcal{A}$ is *consistent* with $\mathcal{T}$ if $\mathcal{T}$ and $\mathcal{A}$ have a common model.

A *signature* Σ is a set of concept and role names. A Σ-ABox is an ABox that uses only concept and role names from Σ, and likewise for Σ-$\mathcal{ELI}_\bot$-concepts and other syntactic objects.

Generally and without further notice, we work with Horn-$\mathcal{ALCHIF}$ TBoxes that are in a certain nesting-free normal form, that is, they contain only CIs of the form

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

where A, B, A_1, A_2 are concept names and r, s are roles. It is well-known that every Horn-$\mathcal{ALCHIF}$ TBox $\mathcal{T}$ can be converted into a TBox $\mathcal{T}'$ in normal form (introducing additional concept names) such that $\mathcal{T}$ is a logical consequence of $\mathcal{T}'$ and every model of $\mathcal{T}$ can be extended to one of $\mathcal{T}'$ by interpreting the additional concept names, see e.g. [Bienvenu *et al.*, 2016]. As a consequence, all results obtained in this paper for TBoxes in normal form lift to the general case.

2.2 Query Conservative Extensions and Entailment

A *conjunctive query (CQ)* is of the form $q(\mathbf{x}) = \exists \mathbf{y}\, \varphi(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are tuples of variables and $\varphi(\mathbf{x}, \mathbf{y})$ is a conjunction of *atoms* of the form $A(v)$ or $r(v, w)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $v, w \in \mathbf{x} \cup \mathbf{y}$. We call $\mathbf{x}$ *answer variables* and $\mathbf{y}$ *quantified variables* of q. A CQ q is *tree-shaped* if it does not contain atoms of the form $r(x, x)$ and the undirected graph $(\mathbf{x} \cup \mathbf{y},\ \{\{v, w\} \mid r(v, w) \text{ is an atom in } q\})$ is a tree; tree-shaped CQs are thus connected and may contain multi-edges. A tree-shaped CQ q is *strongly tree-shaped* or an *stCQ* if the root is the one and only answer variable and q has no multi-edges, that is, for any distinct variables z, z' in q, there is at most one role atom that contains both z and z'.

A *match* of q in an interpretation $\mathcal{I}$ is a function $\pi : \mathbf{x} \cup \mathbf{y} \to \Delta^\mathcal{I}$ such that $\pi(v) \in A^\mathcal{I}$ for every atom $A(v)$ of q and $(\pi(v), \pi(w)) \in r^\mathcal{I}$ for every atom $r(v, w)$ of q. We write $\mathcal{I} \models q(a_1, \ldots, a_n)$ if there is a match of q in $\mathcal{I}$ with $\pi(x_i) = a_i$ for all $i < n$. A tuple $\mathbf{a}$ of elements from $\mathsf{N_I}$ is a *certain answer* to q over an ABox $\mathcal{A}$ given a TBox $\mathcal{T}$, written $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$, if $\mathcal{I} \models q(\mathbf{a})$ for all models of $\mathcal{T}$ and $\mathcal{A}$.

[1]E.g., out of 439 available ontologies in BioPortal [Matentzoglu and Parsia, 2017], only 21 ($\leq$ 4.8%) contain the described pattern. A significant fraction of the occurrences of the pattern appear to be due to modeling mistakes.

Definition 1 *Let Γ, Σ be signatures and $\mathcal{T}_1, \mathcal{T}_2$ Horn-$\mathcal{ALCHIF}$ TBoxes. We say that $\mathcal{T}_1$ (Γ, Σ)*-CQ entails *$\mathcal{T}_2$, written $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, if for all Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, all Σ-CQs $q(\mathbf{x})$ and all tuples $\mathbf{a} \subseteq \mathsf{ind}(\mathcal{A})$, $\mathcal{T}_2, \mathcal{A} \models q(\mathbf{a})$ implies $\mathcal{T}_1, \mathcal{A} \models q(\mathbf{a})$. If in addition $\mathcal{T}_1 \subseteq \mathcal{T}_2$, we say that $\mathcal{T}_2$* is a (Γ, Σ)-CQ conservative extension *of $\mathcal{T}_1$. If $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ and vice versa, then $\mathcal{T}_1$ and $\mathcal{T}_2$ are* (Γ, Σ)-CQ inseparable.

We also consider (Γ, Σ)-*stCQ entailment*, denoted $\models^{\mathrm{stCQ}}_{\Gamma,\Sigma}$ and defined in the obvious way by replacing CQs with stCQs.

If $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ because $\mathcal{T}_2, \mathcal{A} \models q(\mathbf{a})$ but $\mathcal{T}_1, \mathcal{A} \not\models q(\mathbf{a})$ for some Γ-ABox $\mathcal{A}$ consistent with both $\mathcal{T}_i$, Σ-CQ $q(\mathbf{x})$ and $\mathbf{a}$, we call the triple $(\mathcal{A}, q, \mathbf{a})$ a *witness* to non-entailment.

Example 2 *Let $\mathcal{T}_1 = \{\mathit{PhDStud} \sqsubseteq \exists \mathit{advBy}.\mathit{Prof}, \mathit{adv} \sqsubseteq \mathit{advBy}^-\}$ and $\mathcal{T}_2 = \mathcal{T}_1 \cup \{\mathsf{func}(\mathit{advBy})\}$, $\Sigma = \{\mathit{Prof}\}$ and $\Gamma = \{\mathit{PhDStud}, \mathit{adv}\}$. Then $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ because of the witness $(\{\mathit{PhDStud}(\mathit{john}), \mathit{adv}(\mathit{mary}, \mathit{john})\}, \mathit{Prof}(x), \mathit{mary})$.*

If we drop from Definition 1 the condition that $\mathcal{A}$ must be consistent with both $\mathcal{T}_1$ and $\mathcal{T}_2$, then we obtain an alternative notion of CQ entailment that we call *CQ entailment with inconsistent ABoxes*. While this new notion trivially implies CQ entailment in the original sense, the converse fails.

Example 3 *Let $\mathcal{T}_1 = \emptyset$, $\mathcal{T}_2 = \{A_1 \sqcap A_2 \sqsubseteq \bot\}$ and $\Gamma = \{A_1, A_2\}$, $\Sigma = \{B\}$. Then $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ but $\mathcal{T}_1$ does not (Γ, Σ)-CQ entail $\mathcal{T}_2$ with inconsistent ABoxes.*

The following lemma relates the two notions of CQ entailment. *CQ evaluation* is the problem to decide, given a TBox $\mathcal{T}$, an ABox $\mathcal{A}$, a CQ q, and a tuple $\mathbf{a} \in \mathsf{ind}(\mathcal{A})$, whether $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$.

Lemma 4 *CQ entailment with inconsistent ABoxes can be decided in polynomial time given access to oracles deciding CQ entailment and CQ evaluation.*

Consequently and since CQ evaluation is in ExpTime in Horn-$\mathcal{ALCHIF}$ [Eiter *et al.*, 2008], all complexity results obtained in this paper also apply to CQ entailment with inconsistent ABoxes.

It is easy to see that $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ if there is a Γ-role r and a Σ-role s with $\mathcal{T}_2 \models r \sqsubseteq s$ but $\mathcal{T}_1 \not\models r \sqsubseteq s$. We write $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$ if there are no such r and s. Clearly, $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$ can be decided via $|\Gamma| \cdot |\Sigma|$ many Horn-$\mathcal{ALCHIF}$ subsumption tests, thus in ExpTime [Tobies, 2001]. It is thus safe to assume $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$ when deciding CQ entailment, which we will generally do from now on to avoid dealing with special cases.

2.3 Deductive Conservative Extensions

Another natural notion of entailment is deductive entailment, which generalizes the notion of deductive conservative extensions [Ghilardi *et al.*, 2006; Lutz *et al.*, 2007; Konev *et al.*, 2009; Lutz and Wolter, 2010], and which separates two TBoxes in terms of concept and role inclusions and functionality assertions, instead of ABoxes and queries.

Definition 5 *Let Σ be a signature and let $\mathcal{T}_1$ and $\mathcal{T}_2$ be $\mathcal{ELHIF}_\bot$ TBoxes. We say that $\mathcal{T}_1$* Σ-deductively entails *$\mathcal{T}_2$, written $\mathcal{T}_1 \models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$, if for all Σ-$\mathcal{ELI}_\bot$-concept inclusions α and all Σ-RIs and Σ-FAs α: $\mathcal{T}_2 \models \alpha$ implies $\mathcal{T}_1 \models \alpha$. If additionally $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then we say that $\mathcal{T}_2$* is a Σ-deductive conservative extension *of $\mathcal{T}_1$. If $\mathcal{T}_1 \models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$ and vice versa, then $\mathcal{T}_1$ and $\mathcal{T}_2$ are* Σ-deductively inseparable.

Although closely related, it is not difficult to see that deductive and query entailment are orthogonal.

Example 6 *(1) Let $\mathcal{T}_1$, $\mathcal{T}_2$ be as in Example 3 and $\Sigma = \{A_1, A_2, B\}$. Then $\mathcal{T}_1 \models^{\mathrm{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$, but $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$.*

(2) Let $\mathcal{T}_1 = \emptyset$ and $\mathcal{T}_2 = \{A \sqsubseteq \exists r.B\}$, and $\Sigma = \{A, B\}$. Then $\mathcal{T}_1 \models^{\mathrm{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$, but $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Sigma,\Sigma} \mathcal{T}_2$ as witnessed by $(\{A(a)\}, \exists x B(x), a)$. However, $\mathcal{T}_1 \models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$.

Nevertheless, the two notions are sufficiently closely related so that we have the following.

Lemma 7 *In $\mathcal{ELHIF}_\bot$, deductive entailment can be decided in polynomial time given access to oracles for stCQ entailment and stCQ evaluation.*

2.4 Homomorphisms and the Universal Model

For interpretations $\mathcal{I}_1, \mathcal{I}_2$ and a signature Σ, a Σ-*homomorphism* from $\mathcal{I}_1$ to $\mathcal{I}_2$ is a total function $h : \Delta^{\mathcal{I}_1} \to \Delta^{\mathcal{I}_2}$ such that (1) $h(a) = a$ for all $a \in \mathsf{N_I}$, (2) $h(d) \in A^{\mathcal{I}_2}$ for all $d \in A^{\mathcal{I}_1}$, $A \in \mathsf{N_C} \cap \Sigma$, and (3) $(h(d), h(d')) \in r^{\mathcal{I}_2}$ for all $(d, d') \in r^{\mathcal{I}_1}$, $r \in \mathsf{N_R} \cap \Sigma$. If there is a Σ-homomorphism from $\mathcal{I}_1$ to $\mathcal{I}_2$, we write $\mathcal{I}_1 \to_\Sigma \mathcal{I}_2$.

Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox in normal form and $\mathcal{A}$ an ABox consistent with $\mathcal{T}$. A *type for* $\mathcal{T}$ is a set $t \subseteq \mathsf{sub}(\mathcal{T}) \cap \mathsf{N_C}$ such that $\mathcal{T} \models \bigsqcap t \sqsubseteq A$ implies $A \in t$ for all concept names A. For $a \in \mathsf{ind}(\mathcal{A})$, let $\mathsf{tp}_\mathcal{T}(a) = \{A \mid \mathcal{T}, \mathcal{A} \models A(a)\}$ be the *type of a relative to* $\mathcal{T}$. When $a \in \mathsf{ind}(\mathcal{A})$, t, t' are types for $\mathcal{T}$, and r is a role, we write

- $a \leadsto^{\mathcal{T},\mathcal{A}}_r t$ if $\mathcal{T}, \mathcal{A} \models \exists r.\bigsqcap t(a)$ and t is maximal with this condition, and
- $t \leadsto^{\mathcal{T}}_r t'$ if $\mathcal{T} \models \bigsqcap t \sqsubseteq \exists r.\bigsqcap t'$ and t' is maximal with this condition.

A *path for $\mathcal{A}$ and $\mathcal{T}$* is a finite sequence $\pi = a r_0 t_1 \cdots t_{n-1} r_{n-1} t_n$, $n \geq 0$, with $a \in \mathsf{ind}(\mathcal{A})$, $r_0, \ldots, r_{n-1}$ roles, and $t_1, \ldots, t_n$ types for $\mathcal{T}$ such that

(i) $a \leadsto^{\mathcal{T},\mathcal{A}}_{r_0} t_1$ and, if $\mathsf{func}(r_0) \in \mathcal{T}$, then there is no $b \in \mathsf{ind}(\mathcal{A})$ such that $\mathcal{T}, \mathcal{A} \models r_0(a, b)$;

(ii) for every $1 \leq i < n$, we have $t_i \leadsto^{\mathcal{T}}_{r_i} t_{i+1}$ and, if $\mathsf{func}(r) \in \mathcal{T}$, then $r_{i-1} \neq r_i^-$.

When $n > 0$, we use $\mathsf{tail}(\pi)$ to denote t_n. Let Paths be the set of all paths for $\mathcal{A}$ and $\mathcal{T}$. The *universal model* $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ of $\mathcal{T}$ and $\mathcal{A}$ is defined as follows:

$$\Delta^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} = \mathsf{Paths}$$

$$A^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} = \{a \in \mathsf{ind}(\mathcal{A}) \mid \mathcal{T}, \mathcal{A} \models A(a)\} \cup \{\pi \in \Delta^{\mathcal{I}} \setminus \mathsf{ind}(\mathcal{A}) \mid \mathcal{T} \models \bigsqcap \mathsf{tail}(\pi) \sqsubseteq A\}$$

$$r^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} = \{(a, b) \in \mathsf{ind}(\mathcal{A})^2 \mid s(a, b) \in \mathcal{A}, \mathcal{T} \models s \sqsubseteq r\} \cup \{(\pi, \pi s t) \mid \pi s t \in \mathsf{Paths} \text{ and } \mathcal{T} \models s \sqsubseteq r\} \cup \{(\pi s t, \pi) \mid \pi s t \in \mathsf{Paths} \text{ and } \mathcal{T} \models s^- \sqsubseteq r\}$$

It is standard to prove that $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ is indeed a model of $\mathcal{T}$ and $\mathcal{A}$ and that it is universal in the sense that for every model $\mathcal{I}$

of $\mathcal{T}$ and $\mathcal{A}$, we have $\mathcal{I} \to \mathcal{I}_{\mathcal{T},\mathcal{A}}$. Consequently, $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$ iff $\mathcal{I}_{\mathcal{T},\mathcal{A}} \models q(\mathbf{a})$, for all CQs $q(\mathbf{x})$ and tuples $\mathbf{a}$ of individuals.

We also need universal models of a TBox $\mathcal{T}$ and a type t, instead of an ABox. More precisely, we define $\mathcal{I}_{\mathcal{T},t} = \mathcal{I}_{\mathcal{T},\mathcal{A}_t}$ where $\mathcal{A}_t = \{A(a) \mid A \in t\}$ for a fixed $a \in \mathsf{N_I}$.

3 Model-theoretic Characterization

We aim to provide a model-theoretic characterization of query entailment that will be the basis for our decision procedure later on. The first step towards this characterization consists in showing that non-entailment is always witnessed by tree-shaped ABoxes and tree-shaped CQs with at most one answer variable. Here, an ABox $\mathcal{A}$ is *tree-shaped* if it does not contain an assertion of the form $r(a, a)$, the undirected graph $G_\mathcal{A} = (\mathsf{ind}(\mathcal{A}), \{\{a, b\} \mid r(a, b) \in \mathcal{A}\})$ is a tree, and for any $a, b \in \mathsf{ind}(\mathcal{A})$, $\mathcal{A}$ contains at most one role assertion that involves both a and b.

Lemma 8 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\mathsf{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. If $\mathcal{T}_1 \not\models^{\mathsf{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, then there is a witness $(\mathcal{A}, q, \mathbf{a})$ where $\mathcal{A}$ and q are tree-shaped and $|\mathbf{a}| \leq 1$, i.e., q has at most one answer variable. If $\mathcal{T}_1 \not\models^{\mathsf{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, then there is such a witness where additionally q is an stCQ.*

Our goal is to characterize query entailment in terms of homomorphisms between (universal) models. Homomorphisms are natural because answers to CQs are preserved under homomorphisms (both on interpretations and on ABoxes). In fact, they are preserved even under bounded homomorphisms if the bound is not smaller than the number of variables in the CQ.

Let $\mathcal{I}_1, \mathcal{I}_2$ be interpretations, $d \in \Delta^{\mathcal{I}_1}$, and $n \geq 0$. We say that there is an *n-bounded Σ-homomorphism* from $\mathcal{I}_1$ to $\mathcal{I}_2$, written $\mathcal{I}_1 \to^n_\Sigma \mathcal{I}_2$, if for any subinterpretation $\mathcal{I}'_1$ of $\mathcal{I}_1$ with $|\Delta^{\mathcal{I}'_1}| \leq n$, we have $\mathcal{I}'_1 \to_\Sigma \mathcal{I}_2$. Moreover, we write $\mathcal{I}_1 \to^{\mathsf{fin}}_\Sigma \mathcal{I}_2$ if $\mathcal{I}_1 \to^n_\Sigma \mathcal{I}_2$ for any n. The following characterization follows from the definition of CQ entailment, Lemma 8, and the connection between CQs and suitably bounded homomorphisms.

Lemma 9 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\mathsf{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. Then $\mathcal{T}_1 \models^{\mathsf{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff for all tree-shaped Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.*

Ideally, we would like to use Lemma 9 as a basis for a decision procedure based on tree automata. To this end, it is useful that the ABox $\mathcal{A}$ and models $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ and $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ in the lemma are tree-shaped. What is problematic is that Lemma 9 speaks about bounded homomorphisms, for *any* bound (corresponding to the bounded size of CQs), since it does not seem possible to verify such a condition using automata. We would thus like to replace bounded homomorphisms with unbounded ones, which does not compromise the characterization in the case of Horn-DLs without inverse roles [Lutz and Wolter, 2010; Botoeva *et al.*, 2016c]. However, this is not true already for $\mathcal{ELI}$ TBoxes [Botoeva *et al.*, 2016a]:

Example 10 *Let $\mathcal{T}_1 = \{A \sqsubseteq \exists s.B,\ B \sqsubseteq \exists r^-.B\}$, $\mathcal{T}_2 = \{A \sqsubseteq \exists s.B,\ B \sqsubseteq \exists r.B\}$, $\Gamma = \{A\}$, and $\Sigma = \{r\}$. Then both $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ and $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ contain an infinite r-path; the r-path in $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ has a final element while the one in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ does not. Hence $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \not\to_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$, but $\mathcal{T}_1 \models^{\mathsf{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ (see Thm. 11 below).*

We now show that it is possible to refine Lemma 9 so that it makes a much more careful statement in which bounded homomorphisms are *partly* replaced by unbounded ones. It is then possible to check the unbounded homomorphism part of the characterization using tree automata as desired, and to deal with bounded homomorphisms using a mosaic technique that "precompiles" relevant information about unbounded homomorphisms to be used in the automaton construction.

We start with introducing relevant notation. For a signature Σ, we use $\mathcal{I}|^{\mathsf{con}}_\Sigma$ to denote the restriction of the interpretation $\mathcal{I}$ to those elements that can be reached from an ABox individual by traveling along Σ-roles (forwards or backwards). *Tree-shaped interpretations* are defined analogously to tree-shaped CQs (thus multi-edges are allowed). For a TBox $\mathcal{T}$, an ABox $\mathcal{A}$, and $a \in \mathsf{ind}(\mathcal{A})$, we use $\mathcal{I}_{\mathcal{T},\mathcal{A}}|_a$ to denote the subtree interpretation in the universal model $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ rooted at a. A *Σ-subtree in $\mathcal{I}_{\mathcal{T},\mathcal{A}}$* is a maximal tree-shaped, Σ-connected sub-interpretation $\mathcal{I}$ of $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ that does not comprise any ABox individuals. The *root* of $\mathcal{I}$ is the (unique) element of $\Delta^\mathcal{I}$ that can be reached from an ABox individual on a shortest path among all element of $\Delta^\mathcal{I}$. The refined characterization uses simulations instead of homomorphisms for the stCQ case because they are insensitive to multi-edges. Given a signature Σ and two interpretations $\mathcal{I}, \mathcal{J}$, a *Σ-simulation of $\mathcal{I}$ in $\mathcal{J}$* is a relation $\sigma \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{J}$ such that: (1) $(a, a) \in \sigma$ for all $a \in \mathsf{N_I}$, (2) if $d \in A^\mathcal{I}$ with $A \in \Sigma$ and $(d, e) \in \sigma$, then $e \in A^\mathcal{J}$, and (3) if $(d, d') \in r^\mathcal{I}$ with r a Σ-role and $(d, e) \in \sigma$, then there is some e' with $(e, e') \in r^\mathcal{J}$ and $(d', e') \in \sigma$. We write $\mathcal{I} \preceq_\Sigma \mathcal{J}$ if there is a Σ-simulation of $\mathcal{I}$ in $\mathcal{J}$.

Theorem 11 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\mathsf{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. Then $\mathcal{T}_1 \models^{\mathsf{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff for all tree-shaped Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, and for all tree-shaped, finitely branching models $\mathcal{I}_1$ of $\mathcal{A}$ and $\mathcal{T}_1$, the following hold:*

(1) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \to_\Sigma \mathcal{I}_1$;

(2) for all Σ-subtrees $\mathcal{I}$ in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, one of the following holds:

- *(a) $\mathcal{I} \to_\Sigma \mathcal{I}_1$;*
- *(b) $\mathcal{I} \to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathsf{tp}_{\mathcal{I}_1}(a)}$ for some $a \in \mathsf{ind}(\mathcal{A})$.*

Furthermore, $\mathcal{T}_1 \models^{\mathsf{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq_\Sigma \mathcal{I}_1$ for all $\mathcal{A}$ and $\mathcal{I}_1$ as above iff $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

4 Decidability and Complexity

We prove that, in Horn-$\mathcal{ALCHIF}$, CQ entailment can be decided in 2EXPTIME. By existing lower bounds, the former is thus 2EXPTIME-complete in all fragments of Horn-$\mathcal{ALCHIF}$ that contain $\mathcal{ELI}$ or Horn-$\mathcal{ALC}$. Moreover, stCQ entailment in Horn-$\mathcal{ALCHIF}$ and deductive entailment in $\mathcal{ELHIF}_\bot$ can also be decided in 2EXPTIME. We establish a CONEXPTIME lower bound and leave the precise complexity open.

To obtain the upper bounds, we use a combination of tree automata and mosaics to implement the characterization in Theorem 11. We start with a mosaic-based decision procedure for Condition (2b). Note that a Σ-subtree $\mathcal{I}$ in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ can be uniquely identified by the type t_2 of its root. It therefore suffices to show the following.

Theorem 12 *Given two Horn-$\mathcal{ALCHIF}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and types t_i for $\mathcal{T}_i$, $i \in \{1,2\}$, it can be decided in time $2^{2^{p(|\mathcal{T}_2|\log|\mathcal{T}_1|)}}$ whether $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_{\Sigma} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T}_1,t_1}$, p a polynomial.*

Although we cannot get rid of bounded homomorphisms in Theorem 11, a central idea for applying a mosaic approach to prove Theorem 12 is to first replace bounded homomorphisms with unbounded ones. To make this possible, we also replace $\mathcal{I}_{\mathcal{T}_1,t_1}$ with a suitable class of interpretations used as targets for the unbounded homomorphisms.

To illustrate, consider Example 10 and let $t_1 = t_2 = \{B\}$. The difference between $\mathcal{I}_{\mathcal{T}_2,t_2} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T}_1,t_1}$ and $\mathcal{I}_{\mathcal{T}_2,t_2} \rightarrow_{\Sigma} \mathcal{I}_{\mathcal{T}_1,t_1}$ is that unbounded homomorphisms fail once they "reach the root" of $\mathcal{I}_{\mathcal{T}_1,t_1}$ while bounded homomorphisms can, depending on the bound, map the root of $\mathcal{I}_{\mathcal{T}_2,t_2}$ deeper and deeper into $\mathcal{I}_{\mathcal{T}_1,t_1}$, thus never reaching its root. The latter is possible because $\mathcal{I}_{\mathcal{T}_1,t_1}$ is regular in the sense that any two elements which have the same type root isomorphic subtrees. This is of course not only true in this example, but by construction in any universal model. To transition back from bounded to unbounded homomorphisms, we replace $\mathcal{I}_{\mathcal{T}_1,t_1}$ with a class of (finite and infinite) interpretations that can be seen as a "backwards regularization" of $\mathcal{I}_{\mathcal{T}_1,t_1}$. In our concrete example, we would include an interpretation where a predecessor is added to the root of $\mathcal{I}_{\mathcal{T}_1,t_1}$ because $\mathcal{I}_{\mathcal{T}_1,t_1}$ contains an element of the same type as the root that has such a predecessor, an interpretation where that predecessor has a predecessor, and so on, even ad infinitum. We will now make this precise.

An interpretation $\mathcal{I}$ is *quasi tree-shaped* if:

1. $\Delta^{\mathcal{I}} \subseteq (\{-1\} \cup \mathbb{N})^*$;
2. $(d,e) \in r^{\mathcal{I}}$ implies that $e = d \cdot c$ or $d = e \cdot c$ for some $c \in \{-1\} \cup \mathbb{N}$.

For $d, e \in \Delta^{\mathcal{I}}$, we say that e is a *successor* of d if $e = d \cdot c$ for some $c \in \mathbb{N}$ or $d = e \cdot -1$. By this convention, quasi tree-shaped interpretations can be viewed as directed graphs. The directedness does not correspond to the distinction between roles and inverse roles; in particular, there can be several role edges in both directions between the same d and e. Quasi tree-shaped interpretations can be viewed as a finite or infinite trees that need not have a root as they can extend indefinitely not only downwards but also upwards.

Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox and let $\mathsf{tp}(\mathcal{T})$ be the set of all types for $\mathcal{T}$ consistent with $\mathcal{T}$. For every $t_0 \in \mathsf{tp}(\mathcal{T})$, we use $\mathsf{tp}(\mathcal{T}, t_0)$ to denote the set of all $t \in \mathsf{tp}(\mathcal{T})$ that occur in the universal model $\mathcal{I}_{\mathcal{T},t_0}$ of t_0 and $\mathcal{T}$. Furthermore, given a quasi tree-shaped interpretation $\mathcal{I}$ and an element $d \in \Delta^{\mathcal{I}}$, the *1-neighborhood of d in $\mathcal{I}$* is a tuple $n_1^{\mathcal{I}}(d) = (t^-, \rho, t, S)$ such that (a) $t = \mathsf{tp}_{\mathcal{I}}(d)$; (b) if there is a predecessor $d_0 \in \Delta^{\mathcal{I}}$ of d, then $t^- = \mathsf{tp}_{\mathcal{I}}(d_0)$ and $\rho = \{r \mid (d_0, d) \in r^{\mathcal{I}}\}$, otherwise $\rho = t^- = \bot$; (c) S is the set of all pairs (ρ', t') such that there is a successor d' of d such that $t' = \mathsf{tp}_{\mathcal{I}}(d')$ and $\rho' = \{r \mid (d, d') \in r^{\mathcal{I}}\}$. We write $(t_1^-, \rho_1, t_1, S_1) \sqsubseteq (t_2^-, \rho_2, t_2, S_2)$ if $t_1 = t_2$, $S_1 \subseteq S_2$ and, if $\rho_1 \neq \bot$, then $\rho_1 = \rho_2$ and $t_1^- = t_2^-$.

In the following, we define a class $\mathsf{can}_\omega(\mathcal{T}, t_0)$ of quasi tree-shaped models of $\mathcal{T}$. To construct a model from this class, choose a type $t \in \mathsf{tp}(\mathcal{T}, t_0)$ and define $\mathcal{I} = (\{d_0\}, \cdot^{\mathcal{I}})$ such that $\mathsf{tp}_{\mathcal{I}}(d_0) = t$. Then extend $\mathcal{I}$ by applying the following rule exhaustively in a fair way:

(R) Let $d \in \Delta^{\mathcal{I}}$. Choose some $e \in \Delta^{\mathcal{I}_{\mathcal{T},t_0}}$ such that $n_1^{\mathcal{I}}(d) \sqsubseteq n_1^{\mathcal{I}_{\mathcal{T},t_0}}(e)$, and add to d the predecessor and/or successors required to achieve $n_1^{\mathcal{I}}(d) = n_1^{\mathcal{I}_{\mathcal{T},t_0}}(e)$.

The potentially infinite class $\mathsf{can}_\omega(\mathcal{T}, t_0)$ is the set of all interpretations $\mathcal{I}$ obtained as a limit of this construction.

Lemma 13 *Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox, $t_0 \in \mathsf{tp}(\mathcal{T})$, and $\mathcal{I}$ a tree-shaped interpretation. Then $\mathcal{I} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T},t_0}$ iff there is a $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}, t_0)$ with $\mathcal{I} \rightarrow_{\Sigma} \mathcal{J}$.*

We can now use Lemma 13 to devise the mosaic-based procedure for deciding the existence of a bounded homomorphism. Let $\mathcal{T}_1, \mathcal{T}_2$ be as in Theorem 11. We denote with $\mathsf{rol}(\mathcal{T}_i)$ the set of all roles r, r^- such that the (possibly inverse) role r occurs in $\mathcal{T}_i$. Moreover, for a set of roles ρ, denote with $\rho|_\Sigma$ the restriction of ρ to Σ-roles.

Fix now some $t_1 \in \mathsf{tp}(\mathcal{T}_1)$. Intuitively, a mosaic for t_1 represents a possible 1-neighborhood of some element in $\mathcal{I}_{\mathcal{T}_1,t_1}$ together with a decoration with sets of types for $\mathcal{T}_2$ that can be homomorphically embedded into the neighborhood. Formally, a *mosaic for* t_1 is a tuple $M = (t^-, \rho, t, S, \ell)$ such that $(t^-, \rho, t, S) = n_1^{\mathcal{I}_{\mathcal{T}_1,t_1}}(d)$ for some $d \in \Delta^{\mathcal{I}_{\mathcal{T}_1,t_1}}$ and $\ell : \{t^-, t\} \cup S \rightarrow 2^{\mathsf{tp}(\mathcal{T}_2)}$ satisfies the following condition:

(M) For all $\widehat{t} \in \ell(t)$ we have $\widehat{t} \cap \Sigma \subseteq t$ and, for all $\widehat{t'} \in \mathsf{tp}(\mathcal{T}_2)$, $r \in \mathsf{rol}(\mathcal{T}_2)$ with $\widehat{t} \rightsquigarrow_r^{\mathcal{T}_2} \widehat{t'}$, one of the following holds for $\sigma = \{s \in \mathsf{rol}(\mathcal{T}_2) \mid \mathcal{T}_2 \models r \sqsubseteq s\}$:

(a) $\sigma|_\Sigma = \emptyset$;

(b) $t^- \neq \bot$, $\sigma|_\Sigma \subseteq \rho^-$, and $\widehat{t'} \in \ell(t^-)$;

(c) there is $(\rho', t') \in S$ with $\widehat{t'} \in \ell(\rho', t')$ and $\sigma|_\Sigma \subseteq \rho'$.

To ease notation, we use t_M^- to denote t^-, ρ_M to denote ρ, and likewise for the other components of a mosaic M. Let $\mathcal{M}$ be the set of all mosaics for t_1 and $\mathcal{M}' \subseteq \mathcal{M}$. An $M \in \mathcal{M}'$ is *good in* $\mathcal{M}'$ if the following conditions are satisfied:

1. for each $(\rho, t) \in S_M$, there is an $N \in \mathcal{M}'$ such that $(t_M, \rho, t) = (t_N^-, \rho_N, t_N)$, $\ell_M(\rho, t) = \ell_N(t_N)$, and $\ell_M(t_M) = \ell_N(t_N^-)$.
2. if $t_M^- \neq \bot$, there is $N \in \mathcal{M}'$ with $(\rho_M, t_M) \in S_N$, $t_M^- = t_N$, $\ell_M(t_M^-) = \ell_N(t_N)$, and $\ell_M(t_M) = \ell_N(\rho_M, t_M)$.

Let $\mathcal{M}_0, \mathcal{M}_1, \dots$ be the sequence obtained by starting with $\mathcal{M}_0 = \mathcal{M}$ and defining $\mathcal{M}_{i+1}$ to be $\mathcal{M}_i$ when all mosaics that are not good in $\mathcal{M}_i$ have been removed. Assume that $\mathcal{M}_p$ is where the sequence stabilizes.

Lemma 14 *Let $t_i \in \mathsf{tp}(\mathcal{T}_i)$ for $i \in \{1,2\}$. Then there is a $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}_1, t_1)$ such that $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_{\Sigma} \rightarrow_{\Sigma} \mathcal{J}$ iff $\mathcal{M}_p$ contains a mosaic M with $t_2 \in \ell_M(t_M)$.*

Since there are at most $2^{|\mathcal{T}_1| 2^{|\mathcal{T}_2|}}$ mosaics for t_1, we obtain the desired Theorem 12.

We now develop the decision procedure for CQ and stCQ entailment in Horn-$\mathcal{ALCHIF}$, based on Theorems 11 and 12. Our main tool are *alternating two-way tree automata with counting (2ATA$_c$)*, an extension of alternating tree automata over *unranked* trees [Grädel and Walukiewicz, 1999] with the ability to count. A *tree* is a non-empty (potentially infinite) set of words $T \subseteq (\mathbb{N} \setminus 0)^*$ closed under prefixes. We assume that trees are finitely branching, i.e., for every $w \in T$, the set $\{i \mid w \cdot i \in T\}$ is finite. For any $w \in (\mathbb{N} \setminus 0)^*$, we set $w \cdot 0 := w$. If $w = n_0 n_1 \cdots n_k$, $k \geq 0$, we set $w \cdot -1 := n_0 \cdots n_{k-1}$. For an alphabet Θ, a Θ-*labeled tree* is a pair (T, L) with T a tree and $L : T \to \Theta$ a node labeling function.

A *2ATA$_c$* is a tuple $\mathfrak{A} = (Q, \Theta, q_0, \delta, \Omega)$ where Q is a finite set of *states*, Θ is the *input alphabet*, $q_0 \in Q$ is the *initial state*, δ is a *transition function*, and $\Omega : Q \to \mathbb{N}$ is a *priority function*. The transition function δ maps every state q and input letter $a \in \Theta$ to a positive Boolean formula $\delta(q, a)$ over the truth constants true and false and *transition atoms* of the form q, $\langle - \rangle q$, $[-]q$, $\Diamond_n q$ and $\Box_n q$. Informally, a transition q expresses that a copy of $\mathfrak{A}$ is sent to the current node in state q; $\langle - \rangle q$ means that a copy is sent in state q to the predecessor node, which is required to exist; $[-]q$ means the same except that the predecessor node is not required to exist; $\Diamond_n q$ (resp., $\Box_n q$) means that a copy of q is sent to n (resp., to all but n) successors. The semantics of 2ATA$_c$ is given in terms of runs as usual, please see the appendix. We use $L(\mathfrak{A})$ to denote the set of trees accepted by $\mathfrak{A}$. It is standard to verify closure of 2ATA$_c$ under intersection. The following is obtained via reduction to standard alternating parity tree automata [Vardi, 1998].

Theorem 15 *The emptiness problem for 2ATA$_c$ can be solved in time exponential in the number of states.*

Let $\mathcal{T}_1, \mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes and Γ, Σ signatures. We aim to show that one can construct a 2ATA$_c$ $\mathfrak{A}$ such that $L(\mathfrak{A}) \neq \emptyset$ iff $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$. In fact, $\mathfrak{A}$ is the intersection of four 2ATA$_c$ $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3, \mathfrak{A}_4$. They run over Θ-labeled trees with $\Theta = 2^{\Theta_0} \times 2^{\Theta_1} \times 2^{\Theta_2}$, where $\Theta_0 = \Gamma \cup \{r^- \mid r \in \Gamma\}$ and $\Theta_i = \mathsf{sig}(\mathcal{T}_i) \cup \{r^- \mid r \in \mathsf{sig}(\mathcal{T}_i)\}$ for $i = 1, 2$. For a Θ-labeled tree (T, L), we use L_i, $i \in \{0, 1, 2\}$ to refer to the i-th component of L, that is, $L(n) = (L_0(n), L_1(n), L_2(n))$, for all $n \in T$. The component L_0 represents a (possibly infinite) ABox $\mathcal{A} = \{A(n) \mid A \in L_0(n)\} \cup \{r(n \cdot -1, n) \mid n \neq \varepsilon, r \in L_0(n)\}$, where $r^-(a, b)$ is identified with $r(b, a)$. The 2ATA$_c$ $\mathfrak{A}_1$ accepts a Θ-labeled tree (T, L) iff $\mathcal{A}$ is finite, tree-shaped (and thus connected) and includes the root of T, and it is straightforward to construct.

Components L_1, L_2 give rise to interpretations $\mathcal{I}_1 = (T, \cdot^{\mathcal{I}_1})$ and $\mathcal{I}_2 = (\mathsf{ind}(\mathcal{A}), \cdot^{\mathcal{I}_2})$, where for $i \in \{1, 2\}$:

$$A^{\mathcal{I}_i} = \{n \mid A \in L_i(n)\}$$

$$r^{\mathcal{I}_i} = \{(n, n \cdot -1) \mid r^- \in L_i(n)\} \cup \{(n \cdot -1, n) \mid r \in L_i(n)\}$$

$\mathfrak{A}_2$ verifies that $\mathcal{I}_1$ is a model of $\mathcal{A}$ and $\mathcal{T}_1$, which is standard, too. $\mathfrak{A}_3$ verifies that $\mathcal{A}$ is consistent with $\mathcal{T}_2$, and $\mathcal{I}_2$ is $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ restricted to $\mathsf{ind}(\mathcal{A})$. This involves computing the type of an ABox element without having access to the anonymous (that is: non-ABox) part of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, using a characterization of ABox entailments [Bienvenu *et al.*, 2013] in terms of derivation trees. Finally, $\mathfrak{A}_4$ verifies that either (1) or (2) from Theorem 11 is *not* satisfied. For (1), $\mathfrak{A}_4$ sends a copy of itself to every tree $\mathcal{I}$ starting at an ABox element in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, and attempts to show that $\mathcal{I}$ cannot be homomorphically embedded into a corresponding tree in $\mathcal{I}_1$. This attempt is successful if either incompatible types are found in the root or, recursively, there is some successor of the current type in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ that cannot be mapped to any neighbor in $\mathcal{I}_1$. Since the anonymous part of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is not explicit in the input, the current type is stored in the states, and the generating relation $t \rightsquigarrow^{\mathcal{T}_2}_r t'$ is "hard-coded" into the transition function. For Condition (2a), $\mathfrak{A}_4$ non-deterministically guesses a Σ-subtree $\mathcal{I}$ and proceeds as in (1); Condition (2b) is verified based on Theorem 12 by pre-computing $\to^{\mathsf{fin}}_{\Sigma}$. Thus the number of states of $\mathfrak{A}_4$ is exponential in $\mathcal{T}_2$ (because of the types) but only polynomial in $|\mathcal{T}_1|$. Automata $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$ have polynomially many states.

In the special case of stCQ entailment, we simply replace $\mathfrak{A}_4$ with a 2ATA$_c$ $\mathfrak{A}'_4$ that refutes the simulation condition of Theorem 11 analogously to how $\mathfrak{A}_4$ refutes Condition (1).

To obtain the desired upper complexity bounds for query and deductive entailment, we observe that, in both cases, $\mathfrak{A}$ can be constructed in time polynomial in $|\mathcal{T}_1|$ and exponential in $|\mathcal{T}_2|$, and the emptiness check adds an exponential blowup (Theorem 15). For deductive entailment, we use the reduction to stCQ entailment (Lemma 7).

Theorem 16 *In Horn-$\mathcal{ALCHIF}$, the following problems can be decided in time $2^{2^{p(|\mathcal{T}_2| \log |\mathcal{T}_1|)}}$, p a polynomial: (Γ, Σ)-CQ entailment, (Γ, Σ)-CQ inseparability, and (Γ, Σ)-CQ conservative extensions. The same holds for Σ-deductive entailment, Σ-deductive inseparability, and Σ-deductive conservative extensions in $\mathcal{ELHIF}_\bot$.*

Matching lower bounds for all problems except deductive entailment are provided by [Botoeva *et al.*, 2016c]. They hold even in the case where $\Gamma = \Sigma$.

Corollary 17 *In any fragment of Horn-$\mathcal{ALCHIF}$ that contains $\mathcal{ELI}$ or Horn-$\mathcal{ALC}$, the following problems are* 2EXPTIME*-complete: (Γ, Σ)-CQ entailment, (Γ, Σ)-CQ inseparability, and (Γ, Σ)-CQ conservative extensions.*

In the description logic $\mathcal{EL}$, which is $\mathcal{ELI}$ without inverse roles, deductive conservative extensions and deductive Σ-entailment are EXPTIME-complete [Lutz and Wolter, 2010]. This raises the question whether the upper bound for deductive entailment reported in Theorem 16 is tight. While we leave this question open, we observe that the transition from $\mathcal{EL}$ to $\mathcal{ELI}$ does increase the complexity of deductive conservative extensions and related problems to at least CONEXPTIME. We consider this a surprising result since in reasoning problems that are not defined in terms of conjunctive queries, adding inverse roles does typically not result in an increase of complexity. The following is established by a non-trivial reduction of a tiling problem.

Theorem 18 *In any DL between $\mathcal{ELI}$ and $\mathcal{ELHIF}_\bot$, deductive conservative extensions, deductive Σ-entailment, and deductive Σ-inseparability are* CONEXPTIME*-hard.*

5 Conclusion

As future work, it would be interesting to close the gap in complexity between CONEXPTIME and 2EXPTIME for deductive entailment in $\mathcal{ELI}$ and $\mathcal{ELHIF}_\bot$. Furthermore, it would be interesting to extend the results to ontology languages from the family of Datalog+/- (aka existential rules), in particular to frontier-guarded TGDs.

Acknowledgments

This work was partially supported by DFG project LU 1417/2 and ERC Consolidator Grant 647289 CODA.

References

[Arenas *et al.*, 2016] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Knowledge base exchange: The case of OWL 2 QL. *Artif. Intell.*, 238:11–62, 2016.

[Baader *et al.*, 2017] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logics*. Cambride University Press, 2017.

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *Proc. RW*, pages 218–307, 2015.

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

[Bienvenu *et al.*, 2013] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. First order-rewritability of atomic queries in Horn description logics. In *Proc. IJCAI*, pages 754–760, 2013.

[Bienvenu *et al.*, 2016] Meghyn Bienvenu, Peter Hansen, Carsten Lutz, and Frank Wolter. First order-rewritability and containment of conjunctive queries in Horn description logics. In *Proc. IJCAI*, pages 965–971, 2016.

[Botoeva *et al.*, 2016a] Elena Botoeva, Boris Konev, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Inseparability and conservative extensions of description logic ontologies: A survey. In *Proc. RW*, 2016.

[Botoeva *et al.*, 2016b] Elena Botoeva, Roman Kontchakov, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Games for query inseparability of description logic knowledge bases. *Artif. Intell.*, 234:78–119, 2016.

[Botoeva *et al.*, 2016c] Elena Botoeva, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Query-based entailment and inseparability for $\mathcal{ALC}$ ontologies. In *Proc. IJCAI*, pages 1001–1007, 2016.

[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. Autom. Reas.*, 39(3):385–429, 2007.

[Eiter *et al.*, 2008] Thomas Eiter, Georg Gottlob, Magdalena Ortiz, and Mantas Šimkus. Query answering in the description logic Horn-$\mathcal{SHIQ}$. In *Proc. JELIA*, volume 5293 of *LNCS*, pages 166–179. Springer, 2008.

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

[Emerson and Jutla, 1991] E. Allen Emerson and Charanjit S. Jutla. Tree automata, mu-calculus and determinacy (extended abstract). In *Proc. FOCS*, pages 368–377, 1991.

[Ghilardi *et al.*, 2006] Silvio Ghilardi, Carsten Lutz, and Frank Wolter. Did I damage my ontology? A case for conservative extensions in description logics. In *Proc. KR*, pages 187–197, 2006.

[Grädel and Walukiewicz, 1999] Erich Grädel and Igor Walukiewicz. Guarded fixed point logic. In *Proc. LICS*, pages 45–54, 1999.

[Grädel, 1989] Erich Grädel. Dominoes and the complexity of subclasses of logical theories. *Ann. Pure Appl. Logic*, 43(1):1–30, 1989.

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

[Ibáñez-García *et al.*, 2014] Yazmín Ibáñez-García, Carsten Lutz, and Thomas Schneider. Finite model reasoning in Horn description logics. In *Proc. KR*, 2014.

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

[Kollia and Glimm, 2013] Ilianna Kollia and Birte Glimm. Optimizing SPARQL query answering over OWL ontologies. *J. Artif. Intell. Res.*, 48:253–303, 2013.

[Konev *et al.*, 2009] Boris Konev, Carsten Lutz, Dirk Walther, and Frank Wolter. Formal properties of modularisation. In H. Stuckenschmidt, S. Spacciapietra, and C. Parent, editors, *Modular Ontologies*, volume 5445 of *LNCS*, pages 25–66. Springer, 2009.

[Konev *et al.*, 2012] Boris Konev, Michel Ludwig, Dirk Walther, and Frank Wolter. The logical difference for the lightweight description logic $\mathcal{EL}$. *J. Artif. Intell. Res.*, 44:633–708, 2012.

[Kontchakov *et al.*, 2009] Roman Kontchakov, Luca Pulina, Ulrike Sattler, Thomas Schneider, P. Selmer, Frank Wolter, and Michael Zakharyaschev. Minimal module extraction from DL-Lite ontologies using QBF solvers. In *Proc. IJCAI*, pages 836–840, 2009.

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

[Krötzsch *et al.*, 2007] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Complexity boundaries for Horn description logics. In *Proc. AAAI*, pages 452–457, 2007.

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

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

[Lutz *et al.*, 2007] Carsten Lutz, Dirk Walther, and Frank Wolter. Conservative extensions in expressive description logics. In *Proc. IJCAI*, pages 453–458, 2007.

[Matentzoglu and Parsia, 2017] Nico Matentzoglu and Bijan Parsia. BioPortal Snapshot 30 March 2017 (data set), 2017. http://doi.org/10.5281/zenodo.439510.

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

[Tobies, 2001] Stephan Tobies. *Complexity Results and Practical Algorithms for Logics in Knowledge Representation*. PhD thesis, RWTH Aachen, 2001.

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

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

[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. *Comput. Intell.*, 30(2):205–232, 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. Artif. Intell. Res.*, 54:309–367, 2015.

A Proofs for Section 2

A.1 Additional Definitions and Properties

If an interpretation $\mathcal{I}$ is a common model of a TBox $\mathcal{T}$ and ABox $\mathcal{A}$, then we also write $\mathcal{I} \models (\mathcal{T}, \mathcal{A})$ and call $\mathcal{I}$ a *model of* $(\mathcal{T}, \mathcal{A})$.

The following is standard to prove.

Lemma 19 *For every Horn-$\mathcal{ALCHIF}$ TBox $\mathcal{T}$ in normal form and ABox $\mathcal{A}$ consistent with $\mathcal{T}$, the following hold:*

(1) $\mathcal{I}_{\mathcal{T},\mathcal{A}} \models (\mathcal{T}, \mathcal{A})$.

(2) For all models $\mathcal{I}$ of $(\mathcal{T}, \mathcal{A})$, we have $\mathcal{I}_{\mathcal{T},\mathcal{A}} \to \mathcal{I}$.

(3) For all types t, t' for $\mathcal{T}$ with $t \subseteq t'$, we have $\mathcal{I}_{\mathcal{T},t} \to \mathcal{I}_{\mathcal{T},t'}$.

(4) $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$ iff $\mathcal{I}_{\mathcal{T},\mathcal{A}} \models q(\mathbf{a})$, for all CQs $q(\mathbf{x})$ and tuples $\mathbf{a}$ of individuals.

In the following subsections, we need to deal with the question whether any Γ-ABox that is inconsistent with some TBox $\mathcal{T}_2$ is inconsistent with another TBox $\mathcal{T}_1$. We say that $\mathcal{T}_1$ Γ-*inconsistency entails* $\mathcal{T}_2$, written $\mathcal{T}_1 \models^{\perp}_{\Gamma} \mathcal{T}_2$, if for all Γ-ABoxes $\mathcal{A}$: if $\mathcal{A}$ is inconsistent with $\mathcal{T}_2$, then $\mathcal{A}$ is inconsistent with $\mathcal{T}_1$.

A.2 Proof of Lemma 4

Lemma 4 *CQ entailment with inconsistent ABoxes can be decided in polynomial time given access to oracles deciding CQ entailment and CQ evaluation.*

To prove Lemma 4, we proceed in two steps: first, we show how to deal with inconsistency entailment as defined in Section A.1; second we show how to use this type of entailment to deal with CQ entailment with inconsistent ABoxes.

For the first step, we can reduce inconsistency entailment to CQ entailment because, if $\mathcal{A}$ is inconsistent with $\mathcal{T}$, then either (a) $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ contains a B-instance for some B with $B \sqsubseteq \perp \in \mathcal{T}$, or (b) $\mathcal{A}$ contains a "fork" $\{r(a, b), r(a, c)\}$ that is prohibited by $\mathsf{func}(r) \in \mathcal{T}$. We write $\mathcal{T}_1 \models^{\text{fork}}_{\Gamma} \mathcal{T}_2$ if for all Γ-ABoxes $\mathcal{A} = \{r(a, b), r(a, c)\}$: if $\mathcal{A}$ is inconsistent with $\mathcal{T}_2$, then also with $\mathcal{T}_1$. Thus, if we have a witness ABox $\mathcal{A}$ for $\mathcal{T}_1 \not\models^{\perp}_{\Gamma} \mathcal{T}_2$, then $\mathcal{A}$ is inconsistent with $\mathcal{T}_2$ either by Case (a) – which we can detect via CQ entailment if we allow a fresh concept name in the CQ and slightly modify the TBoxes – or by Case (b), which implies $\mathcal{T}_1 \not\models^{\text{fork}}_{\Gamma} \mathcal{T}_2$.

Lemma 20 *Let Γ be a signature and let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes. Furthermore, let A be a fresh concept name and $\mathcal{T}_i^A$ be obtained from $\mathcal{T}_i$ by replacing each occurrence of $\perp$ with A and adding the axioms $\exists s.A \sqsubseteq A$ and $\exists s^-.A \sqsubseteq A$ for every role s occurring in $\mathcal{T}_i$, for $i = 1, 2$. Then the following are equivalent.*

(1) $\mathcal{T}_1 \models^{\perp}_{\Gamma} \mathcal{T}_2$

(2) $\mathcal{T}_1^A \models^{\text{CQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$ and $\mathcal{T}_1 \models^{\text{fork}}_{\Gamma} \mathcal{T}_2$

(3) $\mathcal{T}_1^A \models^{\text{stCQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$ and $\mathcal{T}_1 \models^{\text{fork}}_{\Gamma} \mathcal{T}_2$

Note that we only need the equivalence between (1) and (2) to prove Lemma 4. However, we will need (3) later to prove Lemma 7. Indeed, (2) and (3) are obviously equivalent given the primitive query signature $\{A\}$ and the propagation of A throughout the $\mathcal{T}_i^A$.

Proof.

"1 ⇒ 2". We prove the contrapositive. Assume $\mathcal{T}_1^A \not\models^{\text{CQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$ or $\mathcal{T}_1 \not\models^{\text{fork}}_{\Gamma} \mathcal{T}_2$. In case $\mathcal{T}_1 \not\models^{\text{fork}}_{\Gamma} \mathcal{T}_2$, every witness ABox is a witness for $\mathcal{T}_1 \not\models^{\perp}_{\Gamma} \mathcal{T}_2$ too.

In case $\mathcal{T}_1^A \not\models^{\text{CQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$ is violated, consider a witness $(\mathcal{A}, q, \mathbf{a})$. Since A is the only symbol allowed in q, all atoms of q have the form $A(z)$ for arbitrary variables z. If q consists of several atoms, then it is disconnected and we can omit all but one atom from q and still have a witness (see also proof of Lemma 8, Property d). Hence we can assume w.l.o.g. that q is of the form (i) $q(x) = A(x)$ or (ii) $q() = \exists y\, A(y)$ and, furthermore, that $\mathcal{A}$ and thus the universal models $\mathcal{I}_{\mathcal{T}_i,\mathcal{A}}$ are connected. (Due to the "propagation" of A in the $\mathcal{T}_i$, we can even assume that q is of the form (i) only, but that does not matter in the following argumentation.) We now have:

- $\mathcal{A}$ is inconsistent with $\mathcal{T}_2$:

  Assume to the contrary that $\mathcal{A}$ is consistent with $\mathcal{T}_2$ and consider the universal model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ for $\mathcal{T}_2$ and $\mathcal{A}$ (Section 2.4). Clearly, for all domain elements d of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, we have $\mathcal{T}_2 \not\models \bigsqcap \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}}(d) \sqsubseteq \perp$. Since A is fresh and by the definition of $\mathcal{T}_2^A$ we get $\mathcal{T}_2^A \not\models \bigsqcap \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}}(d) \sqsubseteq A$. Now Lemma 19 (1) for $\mathcal{T}_2^A$ implies that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models (\mathcal{T}_2, \mathcal{A})$; hence $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ satisfies all axioms in $\mathcal{T}_2^A$ that have been taken over from $\mathcal{T}_2$ without modification, i.e., all axioms that are not of the form $B \sqsubseteq A$. But axioms of the latter form are also satisfied because $\mathcal{T}_2^A \not\models \bigsqcap \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}}(d) \sqsubseteq A$ for every domain element d. Hence $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models (\mathcal{T}_2^A, \mathcal{A})$. Now, since $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ has no A-instance, we cannot have $\mathcal{T}_2^A, \mathcal{A} \models q(\mathbf{a})$ for any $\{A\}$-query q; contradicting the assumption that $(\mathcal{A}, q, \mathbf{a})$ is a witness.

- $\mathcal{A}$ is consistent with $\mathcal{T}_1$:

  Since $(\mathcal{A}, q, \mathbf{a})$ is a witness, we have $\mathcal{I}_{\mathcal{T}_1^A,\mathcal{A}} \not\models q(\mathbf{a})$ by Lemma 19 (4). Due to the additional axioms in the definition of $\mathcal{T}_1^A$, which "propagate" A into every domain element of the connected (see above) universal model $\mathcal{I}_{\mathcal{T}_1^A,\mathcal{A}}$, we have $\mathcal{T}_1^A \not\models \bigsqcap \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_1^A,\mathcal{A}}}(d) \sqsubseteq A$ for all domain elements d. Since A is fresh, we have $\mathcal{T}_1 \not\models \bigsqcap \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_1^A,\mathcal{A}}}(d) \sqsubseteq \perp$. With the same reasoning as above, we get $\mathcal{I}_{\mathcal{T}_1^A,\mathcal{A}} \models (\mathcal{T}_1, \mathcal{A})$; hence $\mathcal{A}$ is consistent with $\mathcal{T}_1$.

Consequently $\mathcal{T}_1 \not\models^{\perp}_{\Gamma} \mathcal{T}_2$, as desired.

"2 ⇒ 3". This is immediate because $\mathcal{T}_1^A \models^{\text{CQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$ implies $\mathcal{T}_1^A \models^{\text{stCQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$.

"3 ⇒ 1". We prove the contrapositive. Assume $\mathcal{T}_1 \not\models^{\perp}_{\Gamma} \mathcal{T}_2$, i.e., there is a Γ-ABox $\mathcal{A}$ that is is inconsistent with $\mathcal{T}_2$ but consistent with $\mathcal{T}_1$. We need to show that $\mathcal{T}_1^A \not\models^{\text{stCQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$ or $\mathcal{T}_1 \not\models^{\text{fork}}_{\Gamma} \mathcal{T}_2$.

From $\mathcal{A}$ being inconsistent with $\mathcal{T}_2$, we first conclude that one of the following two properties must hold.

(i) There is some $d \in B^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}}$ with $B \sqsubseteq \bot \in \mathcal{T}_2$.

(ii) $\mathcal{A}$ contains a "fork" $\mathcal{A}^- = \{r(a,b), r(a,c)\}$ such that $\mathcal{A}^-$ is inconsistent with $\mathcal{T}_2$.

Indeed, if neither (i) nor (ii) holds, then we have $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models (\mathcal{T}_2, \mathcal{A})$, contradicting the inconsistency of $\mathcal{A}$ with $\mathcal{T}_2$: First, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models \mathcal{A}$ follows directly from the construction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$. Second, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models \mathcal{T}_2$ can be shown analogously to the (omitted) standard proof of Lemma 19 (1), via a case distinction over the axioms in $\mathcal{T}_2$, using "not (i)" and "not (ii)" instead of the assumption that $\mathcal{A}$ is consistent with $\mathcal{T}_2$.

Now first assume that (ii) holds. Since $\mathcal{A}$ is consistent with $\mathcal{T}_1$, so is $\mathcal{A}^-$. Hence $\mathcal{T}_1 \not\models^{\text{fork}}_{\Gamma} \mathcal{T}_2$.

In case (ii) does not hold, (i) must hold. To show that $\mathcal{T}_1^A \not\models^{\text{stCQ}}_{\Gamma,\{A\}} \mathcal{T}_2^A$, consider the stCQ $q = A(x)$ and some $a \in \mathsf{ind}(\mathcal{A})$ to which the element d from (i) is connected in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, i.e., if $d \in \mathsf{ind}(\mathcal{A})$, then choose $a = d$; otherwise choose a such that d is in the subtree $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|_a$. We then have:

- $\mathcal{A}$ is consistent with $\mathcal{T}_2^A$:

  Since $\mathcal{T}_2^A$ does not contain $\bot$ and $\mathcal{A}$ does not contain forks as in (ii), $\mathcal{A}$ is consistent with $\mathcal{T}_2^A$ is consistent, as witnessed by the universal model $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}}$ (we again refer to the standard proof of Lemma 19 (1); except that the FA case in the ABox part of $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}}$ is now due to "not (ii)").

- $\mathcal{A}$ is consistent with $\mathcal{T}_1^A$:

  It is not difficult to see that $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \models (\mathcal{T}_1^A, \mathcal{A})$: From $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \models (\mathcal{T}_1, \mathcal{A})$, it follows that $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ is a model of $\mathcal{A}$ and satisfies all axioms in $\mathcal{T}_1^A$ that $\mathcal{T}_1^A$ shares with $\mathcal{T}_1$. The modified axioms $B \sqsubseteq A$ with $B \sqsubseteq \bot \in \mathcal{T}_1$ are satisfied, too, because $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ cannot have any B-instances. Finally, the additional propagation axioms are satisfied because $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ has no A-instance as A is fresh.

- $\mathcal{T}_2^A, \mathcal{A} \models q(a)$:

  Due to (i), we have $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models \exists y\, B(y)$ for some $B \sqsubseteq \bot \in \mathcal{T}_2$. Hence $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}} \models \exists y\, B(y)$, which follows from the construction of both universal models (in fact the only difference between $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}}$ and $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is that some domain elements of $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}}$ may be A-instances). Hence $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}}$ has a B-instance in the subtree $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|_a$ and thus, by construction, an A-instance. By the "propagation" of A in $\mathcal{T}_2^A$, we have that a is an instance of A in $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}}$; hence $\mathcal{I}_{\mathcal{T}_2^A,\mathcal{A}} \models A(a) = q$.

- $\mathcal{T}_1^A, \mathcal{A} \not\models q(a)$:

  Follows from $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \models (\mathcal{T}_1^A, \mathcal{A})$ (as shown above) and $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \not\models q(a)$ (given the lack of A-instances).

❑

Proposition 21 *Fork entailment $\mathcal{T}_1 \models^{\text{fork}}_{\Gamma} \mathcal{T}_2$ can be (Turing) reduced in polynomial time to stCQ evaluation.*

Proof. Perform $2|\Gamma|$ many ABox consistency checks by evaluating the stCQ $A(a)$ on both $\mathcal{T}_i$, where A is a concept name that does not occur in any of the $\mathcal{T}_i$. ❑

For the second step, we can now reduce CQ entailment with inconsistent ABoxes to the disjunction of our original notion of CQ entailment and inconsistency entailment. We need an additional notion: Given a TBox $\mathcal{T}$ and signatures Γ, Σ, we say that $\mathcal{T}$ is *(Γ, Σ)-universal* if

($*$) for all Γ-ABoxes $\mathcal{A}$ and Σ-CQs $q(\mathbf{x})$ and all tuples $\mathbf{a} \subseteq \mathsf{ind}(\mathcal{A})$ with $|\mathbf{a}| = |\mathbf{x}|$, we have $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$.

Lemma 22 *Let Γ, Σ be signatures and let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes. Then $\mathcal{T}_1$ (Γ, Σ)-CQ entails $\mathcal{T}_2$ with inconsistent ABoxes iff one of the two following conditions holds.*

(1) $\mathcal{T}_1 \models^{\text{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ and $\mathcal{T}_1 \models^{\bot}_{\Gamma} \mathcal{T}_2$

(2) $\mathcal{T}_1$ is (Γ, Σ)-universal.

Proof. We prove both implications via contraposition.

"⇒". Assume (1) and (2) are both false, i.e., $\mathcal{T}_1$ is not (Γ, Σ)-universal and either (a) $\mathcal{T}_1 \not\models^{\text{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ or (b) $\mathcal{T}_1 \not\models^{\bot}_{\Gamma} \mathcal{T}_2$. In case (a), $\mathcal{T}_1$ trivially does not (Γ, Σ)-CQ entail $\mathcal{T}_2$ with inconsistent ABoxes. In case (b), consider a witness Γ-ABox $\mathcal{A}$. Since $\mathcal{T}_1$ is not (Γ, Σ)-universal, there is a Γ-ABox $\mathcal{A}'$, a Σ-CQ $q(\mathbf{x})$ and a tuple $\mathbf{a} \subseteq \mathsf{ind}(\mathcal{A}')$ with $|\mathbf{a}| = |\mathbf{x}|$ such that $\mathcal{T}_1, \mathcal{A}' \not\models q(\mathbf{a})$. We assume w.l.o.g. that $\mathcal{A}$ and $\mathcal{A}'$ use distinct sets of individuals. We set $\mathcal{A}'' = \mathcal{A} \cup \mathcal{A}'$ and have:

- $\mathcal{T}_2, \mathcal{A}'' \models q(\mathbf{a})$ because $\mathcal{A}$ is inconsistent with $\mathcal{T}_2$ and so is $\mathcal{A}''$.
- $\mathcal{T}_1, \mathcal{A}'' \not\models q(\mathbf{a})$: let $\mathcal{J}$ be the disjoint union of the universal model $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ and the model $\mathcal{I}$ witnessing $\mathcal{T}_1, \mathcal{A}' \not\models q(\mathbf{a})$. Clearly $\mathcal{J} \models (\mathcal{T}_1, \mathcal{A}'')$ but $\mathcal{J} \not\models q(\mathbf{a})$.

Hence $\mathcal{T}_1$ does not (Γ, Σ)-CQ entail $\mathcal{T}_2$ with inconsistent ABoxes, as desired.

"⇐". Assume $\mathcal{T}_1$ does not (Γ, Σ)-CQ entail $\mathcal{T}_2$ with inconsistent ABoxes and consider a witness $(\mathcal{A}, q, \mathbf{a})$. Then it is immediate that (2) does not hold. Furthermore, if $\mathcal{A}$ is consistent with both $\mathcal{T}_1$ and $\mathcal{T}_2$, then $\mathcal{T}_1 \not\models^{\text{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$. Otherwise $\mathcal{A}$ must be inconsistent with $\mathcal{T}_2$ but consistent with $\mathcal{T}_1$; hence $\mathcal{T}_1 \not\models^{\bot}_{\Gamma} \mathcal{T}_2$. Therefore (1) does not hold either. ❑

Proposition 23 *(Γ, Σ)-universality can be (Turing) reduced in polynomial time to stCQ evaluation.*

Proof. It suffices to check Condition $(*)$ above (i) for all singleton Γ-ABoxes $\{A(a)\}$ and all single-atom Σ-CQs $B(x)$ or $r(x, x)$, and (ii) for all two-element Γ-ABoxes $\{r(a, b)\}$ and all Σ-CQs as in (i) but with possibly two distinct answer variables. ❑

Lemma 4 is now a direct consequence of Lemmas 22 and 20, and Propositions 21 and 23.

A.3 Proof of Lemma 7

Lemma 7 *In $\mathcal{ELHIF}_\bot$, deductive entailment can be decided in polynomial time given access to oracles for stCQ entailment and stCQ evaluation.*

Lemma 7 is an immediate consequence of the following lemma because the additional $\mathcal{T}_1 \models^{\bot}_{\Sigma} \mathcal{T}_2$ can be reduced to *rst*CQ entailment and stCQ evaluation via Lemma 20.

Lemma 24 *Let Σ be a signature and $\mathcal{T}_1, \mathcal{T}_2$ $\mathcal{ELHIF}_\bot$ TBoxes such that $\mathcal{T}_1 \models^{\text{RI}}_{\Sigma,\Sigma} \mathcal{T}_2$. Then*

$$\mathcal{T}_1 \models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2 \quad \textit{iff} \quad \mathcal{T}_1 \models^{\text{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2 \quad \textit{and} \quad \mathcal{T}_1 \models^{\bot}_{\Sigma} \mathcal{T}_2.$$

Proof. We prove both implications via contraposition.

"$\Leftarrow$". We assume that $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$. In case this is witnessed by a Σ-FA $\mathsf{func}(r)$, we immediately get a witness Σ-ABox = $\{r(a,b), r(a,c)\}$ for $\mathcal{T}_1 \not\models^{\bot}_{\Sigma} \mathcal{T}_2$ and are done.

Otherwise, $\mathcal{T}_1$ contains all Σ-FAs from $\mathcal{T}_2$, and there is a witness Σ-CI $C \sqsubseteq D$ (witness RIs are excluded by the assumption $\mathcal{T}_1 \models^{\text{RI}}_{\Sigma,\Sigma} \mathcal{T}_2$). Since $\mathcal{ELI}_\bot$-concepts that contain $\bot$ are equivalent to $\bot$, the left-hand side C cannot contain $\bot$ (i.e., is an $\mathcal{ELI}$ concept) and, if D does, then $C \sqsubseteq \bot$ is a witness. We show that such witnesses give rise to either a witness $\mathcal{A}_C$ for $\mathcal{T}_1 \not\models^{\bot}_{\Sigma} \mathcal{T}_2$ or a witness $(\mathcal{A}_C, q_D, a)$ for $\mathcal{T}_1 \not\models^{\text{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$ with $q_D(x)$ an *stCQ*.

We first consider the case that there is a witness $C \sqsubseteq \bot$ with C an $\mathcal{ELI}$ concept. We can construct from C in the obvious way a tree-shaped Σ-ABox $\mathcal{A}_C$ and root a: $\mathcal{A}$ reflects the tree structure of C; however, to respect the Σ-FAs in $\mathcal{T}_1$ (and thus those in $\mathcal{T}_2$), we need to merge the subtrees of all nodes that are r-neighbors of the same node, whenever $\mathsf{func}(r) \in \mathcal{T}_1$. Consider the universal model $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}_C}$[2] and observe that $a \in C^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}_C}}$ from the construction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}_C}$. Since $\mathcal{T}_2 \models C \sqsubseteq \bot$, we have that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}_C}$ is not a model of $\mathcal{T}_2$. Hence, by the contrapositive of Lemma 19 (1), $\mathcal{A}_C$ is inconsistent with $\mathcal{T}_2$. On the other hand, since $\mathcal{T}_1 \not\models C \sqsubseteq \bot$, there is a model $\mathcal{I} \models \mathcal{T}_1$ and an instance $d \in C^{\mathcal{I}}$. We can turn $\mathcal{I}$ into a model of $\mathcal{A}_C$ by interpreting the ABox individuals accordingly ("partial" unraveling might be necessary to ensure that the standard name assumption is respected), witnessing the consistency of $\mathcal{A}$ with $\mathcal{T}_1$. We thus have $\mathcal{T}_1 \not\models^{\bot}_{\Sigma} \mathcal{T}_2$ and are done.

In the second case, *all* witnesses $C \sqsubseteq D$ consist solely of $\mathcal{ELI}$ concepts C, D. We construct the same ABox $\mathcal{A}_C$ with root a from C and transform D into a Σ-stCQ $q_D(x)$ with a single answer variable that represents the tree shape of D. Now $(\mathcal{A}_C, q_D, a)$ is a witness to $\mathcal{T}_1 \not\models^{\text{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$ for the following reasons.

- $\mathcal{A}_C$ is consistent with $\mathcal{T}_1$: a model can be obtained in the obvious way from the model witnessing $\mathcal{T}_1 \not\models C \sqsubseteq D$ (possibly involving "partial" unraveling as above).
- $\mathcal{A}_C$ is consistent with $\mathcal{T}_2$: since $C \sqsubseteq \bot$ is not a witness to $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$, there must be a model $\mathcal{I} \models \mathcal{T}_2$ with $d \in C^{\mathcal{I}}$. We claim that we can turn $\mathcal{I}$ into a model of $\mathcal{A}_C$ by interpreting the ABox individuals without violating the standard name assumption. If we assume to the contrary that this is not possible, then there are subconcepts $C_1, \dots, C_n$ of C corresponding to subtrees that have been merged in the construction of $\mathcal{A}_C$, such that $\mathcal{T}_2 \models C_1 \sqcap \dots \sqcap C_n \sqsubseteq \bot$. However, $\mathcal{T}_1 \not\models C_1 \sqcap \dots \sqcap C_n \sqsubseteq \bot$ because $\mathcal{A}_C$ is consistent with $\mathcal{T}_1$, as shown previously. Hence $C_1 \sqcap \dots \sqcap C_n \sqsubseteq \bot$ would be a witness to $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$, which we have ruled out – a contradiction.
- $\mathcal{T}_2, \mathcal{A}_C \models q_D(a)$, witnessed by $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}_C}$, together with $a \in C^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}_C}}$ and $\mathcal{T}_2 \models C \sqsubseteq D$.
- $\mathcal{T}_1, \mathcal{A}_C \not\models q_D(a)$: take a model $\mathcal{I}$ witnessing $\mathcal{T}_1 \not\models C \sqsubseteq D$ and an element $d \in C^{\mathcal{I}} \setminus D^{\mathcal{I}}$. As in the previous case, we can turn $\mathcal{I}$ into a model $\mathcal{J}$ of $\mathcal{A}_C$ by interpreting the ABox individuals (again involving unraveling if necessary), obtaining $\mathcal{J} \not\models q_D(a)$.

"$\Rightarrow$". Assume $\mathcal{T}_1 \not\models^{\text{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$ or $\mathcal{T}_1 \not\models^{\bot}_{\Sigma} \mathcal{T}_2$.

In case $\mathcal{T}_1 \not\models^{\bot}_{\Sigma} \mathcal{T}_2$, consider a witness Σ-Box $\mathcal{A}$ and assume w.l.o.g. that $\mathcal{A}$ is tree-shaped. Let $a \in \mathsf{ind}(\mathcal{A})$ be its root. We can assume that $\mathcal{T}_1$ contains all Σ-FAs from $\mathcal{T}_2$ (otherwise $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$ and we are done). We turn $\mathcal{A}$ into a Σ-$\mathcal{ELI}$ concept $C_{\mathcal{A}}$ in the obvious way. Then $C_{\mathcal{A}} \sqsubseteq \bot$ is a witness to $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$:

- $\mathcal{T}_2 \models C_{\mathcal{A}} \sqsubseteq \bot$ because, if there were a model $\mathcal{I}$ of $\mathcal{T}_2$ with $d \in C_{\mathcal{A}}^{\mathcal{I}}$, we could turn it into a model of $(\mathcal{T}_2, \mathcal{A})$ by interpreting the ABox individuals accordingly (possibly involving partial unraveling as above), which would contradict the assumption that $\mathcal{A}$ is a witness to $\mathcal{T}_1 \not\models^{\bot}_{\Sigma} \mathcal{T}_2$.
- $\mathcal{T}_1 \not\models C_{\mathcal{A}} \sqsubseteq \bot$, witnessed by $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

In case $\mathcal{T}_1 \not\models^{\text{stCQ}}_{\Sigma,\Sigma} \mathcal{T}_2$, by Lemma 8 there is a witness $(\mathcal{A}, q, a)$ with $\mathcal{A}$ tree-shaped and q a Σ-stCQ with exactly one answer variable. We construct $C_{\mathcal{A}}$ as above and another Σ-$\mathcal{ELI}$ concept D_q from q in the obvious way. It can be shown analogously to the previous case that $C_{\mathcal{A}} \sqsubseteq D_q$ is a witness to $\mathcal{T}_1 \not\models^{\mathcal{ELHIF}_\bot}_{\Sigma} \mathcal{T}_2$. ❏

B Proofs for Section 3

B.1 Unraveling ABoxes

To obtain tree-shaped ABoxes or CQs, we use unraveling, which needs to be more cautious in the presence of inverse roles and functionality. In particular, we need to ensure that, whenever a role is functional in an ABox, then so it is in its unraveling. We define an unraveling for Horn-$\mathcal{ALCHIF}$ similar to the one for Horn-$\mathcal{ALCIF}$ in [Lutz and Wolter, 2012].

Let $\mathcal{A}$ be an ABox. The unraveling $U^a_{\mathcal{A}}$ of $\mathcal{A}$ at an individual $a \in \mathsf{ind}(\mathcal{A})$ is the following ABox:

- $\mathsf{ind}(U^a_{\mathcal{A}})$ is the set of sequences $b_0 r_0 b_1 \cdots r_{n-1} b_n$ with $n \geq 0$, where $b_0 = a$, $b_i \in \mathsf{ind}(\mathcal{A})$ for all $0 \leq i \leq n$, $r_i(b_i, b_{i+1}) \in \mathcal{A}$ for all $0 \leq i < n$, and $(b_{i-1}, r^-_{i-1}) \neq$

[2]The assumption that $\mathcal{A}$ is consistent with $\mathcal{T}$ is not needed for the construction of $\mathcal{I}_{\mathcal{T},\mathcal{A}}$, only for the proof of Lemma 19 (1).

(b_{i+1}, r_i) (the latter inequality is needed to ensure preservation of functionality).

- The concept assertions in $U^a_\mathcal{A}$ are all assertions of the shape $C(\alpha)$ such that $\alpha = b_0 \cdots b_{n-1} r_{n-1} b_n \in \mathsf{ind}(\mathcal{A})$ and $C(b_n) \in \mathcal{A}$. The role assertions in $U^a_\mathcal{A}$ are all assertions of the shape $r(b_0 \cdots b_{n-1}, \alpha)$ such that $\alpha = b_0 \cdots b_{n-1} r_{n-1} b_n \in \mathsf{ind}(\mathcal{A})$.

The following is standard to prove:

Proposition 25 *Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox, $\mathcal{A}$ an ABox, and $a \in \mathsf{ind}(\mathcal{A})$. If $\mathcal{A}$ is consistent with $\mathcal{T}$, then so is $U^a_\mathcal{A}$.*

B.2 Proof of Lemma 8

We reformulate the lemma to make its statement more explicit.

Lemma 8, reformulated equivalently. *Let $\mathcal{T}_1, \mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. If $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, then there is a tree-shaped Γ-ABox $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, and a tree-shaped Σ-CQ q such that one of the following holds:*

(1) q has a single answer variable and there is an $a \in \mathsf{ind}(\mathcal{A})$ such that $\mathcal{T}_2, \mathcal{A} \models q(a)$ but $\mathcal{T}_1, \mathcal{A} \not\models q(a)$;

(2) q is Boolean and $\mathcal{T}_2, \mathcal{A} \models q$ but $\mathcal{T}_1, \mathcal{A} \not\models q$.

If $\mathcal{T}_1 \not\models^{\mathrm{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, then there is a tree-shaped Γ-ABox $\mathcal{A}$ and a tree-shaped Σ-stCQ q with (1).

Proof. Unrestricted CQs. Assume $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$, i.e., $\mathcal{T}_2, \mathcal{A} \models q(\mathbf{a})$ and $\mathcal{T}_1, \mathcal{A} \not\models q(\mathbf{a})$, for some Γ-ABox $\mathcal{A}$ consistent with both $\mathcal{T}_i$, some Σ-CQ q and some tuple $\mathbf{a}$. Lemma 19 (4) yields $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q(\mathbf{a})$ and $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \not\models q(\mathbf{a})$. We first show that the following properties of q and $\mathbf{a}$ are without loss of generality:

(a) Every match of $q(\mathbf{x})$ into $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ maps every quantified variable into the anonymous part.

(b) $q(\mathbf{x})$ does not contain atoms of the form $r(x_1, x_2)$ with x_1, x_2 answer variables.

(c) If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{a} = (a_1, \ldots, a_n)$, then $a_i \neq a_j$ for all i, j with $1 \leq i < j \leq n$.

(d) $q(\mathbf{x})$ is connected.

For (a), take a match π of q in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ and a quantified variable y such that $\pi(y) = b \in \mathsf{ind}(\mathcal{A})$. Obtain $q'(\mathbf{x}, y)$ from $q(\mathbf{x})$ by removing the quantification over y, thus making y an answer variable. Clearly, we have $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q'(\mathbf{a}, b)$ and $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \not\models q'(\mathbf{a}, b)$, and thus $\mathcal{T}_2, \mathcal{A} \models q'(\mathbf{a}, b)$ and $\mathcal{T}_1, \mathcal{A} \not\models q'(\mathbf{a}, b)$.

For (b), observe that such atoms can always be dropped, since they cannot be inferred via $\mathcal{T}_1$ or $\mathcal{T}_2$: Let $q(\mathbf{x}) = \exists \mathbf{y}\, (r(x_1, x_2) \wedge \varphi(\mathbf{x}', \mathbf{y}))$ with $x_1, x_2 \in \mathbf{x}$, and let $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q(\mathbf{a})$ be witnessed by the match π with $\pi(x_i) = a_i$, $i = 1, 2$. Construct the CQ $q(\mathbf{x}') = \exists \mathbf{y}\, \varphi(\mathbf{x}', \mathbf{y})$ by dropping the atom $r(x_1, x_2)$ (and thus possibly removing x_1 and/or x_2 from the free variables). It is clear that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q'(\mathbf{a}')$ for the corresponding restriction $\mathbf{a}'$ of the tuple $\mathbf{a}$; thus it suffices to show that $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \not\models q'(\mathbf{a}')$.

From $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q(\mathbf{a})$ we can conclude that $(a_1, a_2) \in r^{\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}}$. By construction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ there is some Γ-role r' with $r'(a_1, a_2) \in \mathcal{A}$ and $\mathcal{T}_2 \models r' \sqsubseteq r$ (which includes the possibility $r' = r$, i.e., $r(a_1, a_2) \in \mathcal{A}$). Due to $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$, we also have $\mathcal{T}_1 \models r' \sqsubseteq r$ and hence $(a_1, a_2) \in r^{\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}}$. This implies the desired $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \not\models q'(\mathbf{a}')$ because, otherwise, any match π of q' in $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ with $\pi(x_i) = a_i$, $i = 1, 2$, could be extended to a match of q.

This construction does not introduce any violations of (a).

For (c), observe that, whenever $a_i = a_j$ for some i, j with $1 \leq i < j \leq n$, we can always drop x_j and a_j: Let $\mathbf{x}'$ and $\mathbf{a}'$ be $\mathbf{x}$ and $\mathbf{a}$ with x_j and a_j removed, and transform $q(\mathbf{x})$ into $q'(\mathbf{x}')$ by replacing every occurrence of x_j with x_i. Now $\mathcal{T}_2, \mathcal{A} \models q'(\mathbf{a}')$ and $\mathcal{T}_1, \mathcal{A} \not\models q'(\mathbf{a}')$. This construction does not introduce any violations of (a) or (b).

For (d), observe that $\mathcal{T}_2, \mathcal{A} \models q(\mathbf{a})$ and $\mathcal{T}_1, \mathcal{A} \not\models q(\mathbf{a})$ implies $\mathcal{T}_2, \mathcal{A} \models q'(\mathbf{a})$ and $\mathcal{T}_1, \mathcal{A} \not\models q'(\mathbf{a})$ for some connected component q' of q. This construction does not introduce any violations of (a), (b), or (c). While this is easy to see for (b) and (c), Property (a) requires a closer look: If the possibly disconnected CQ q satisfies (a) and has at least one match π in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, then every match of any connected component q' in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ can be extended to a match of q in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ via π restricted to the remaining connected components. Since the match of q satisfies (a), so does the match of q'.

Thus, as long as q violates any of the above properties, we apply the corresponding modification as described and in the order given. From now on, we assume that q satisfies properties (a) to (d). Furthermore, they imply:

(e) $q(\mathbf{x})$ does *not* contain a *proper path* between any two answer variables, which is a non-empty sequence of atoms $r_1(z_1, z_2), r_2(z_2, z_3), \ldots, r_n(z_n, z_{n+1})$ with variables $z_1, z_{n+1} \in \mathbf{x}$ and $z_i \in \mathbf{y}$ for $1 < i \leq n$, and with roles r_i such that $z_{i+1} \neq z_{i-1}$ for every $1 < i \leq n$.

To show this, assume the opposite, i.e., $q(\mathbf{x})$ contains a proper path as above between two answer variables x, x'. By (b) we have $n > 1$. By (a) and (c), π maps all z_i with $1 < i \leq n$ to the anonymous part of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$. However, there is no corresponding proper path between any two ABox individuals in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$; a contradiction.

Assume now that $q(\mathbf{x}) = \exists \mathbf{y}\, \varphi(\mathbf{x}, \mathbf{y})$ is not tree-shaped, i,e., there is a cycle $r_1(z_1, z_2), r_2(z_2, z_3), \ldots, r_n(z_n, z_{n+1})$ with variables $z_i \in \mathbf{x} \cup \mathbf{y}$, $z_1 = z_{n+1}$, and roles r_i such that $z_{i+1} \neq z_{i-1}$ for every $1 < i \leq n$ and $z_2 \neq z_n$. By (e), we have $z_i \in \mathbf{y}$ for all $1 \leq i \leq n + 1$. Let π be a match of $q(\mathbf{x})$ in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$. By (a), π maps all variables to the anonymous part of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ which, by construction, is acyclic. Hence π cannot satisfy the properties of a match; contradiction.

Assume now that $\mathbf{x}$ in $q(\mathbf{x})$ contains more than one answer variable, say $x \neq x'$, matched by a and a' in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, with $a \neq a'$ due to (c). By (d), q is connected, and thus, there is a path from x to x' in q. Since $x \neq x'$, there is even a proper path; contradicting (e).

Thus, we now have that q is tree-shaped and behaves as required by (1) or (2). It remains to transform $\mathcal{A}$ into a tree-shaped ABox: In case q is Boolean, we get from (a) and (c) that every match of q in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is into the anonymous subtree rooted at some ABox individual a; in case q has one answer variable, let $\mathbf{a} = a$. Consider the unraveling $U^a_\mathcal{A}$ of $\mathcal{A}$ at a. Clearly, $\mathcal{T}_2, U^a_\mathcal{A} \models q(a)$ and $\mathcal{T}_1, U^a_\mathcal{A} \not\models q(a)$, which is still consistent with both $\mathcal{T}_i$, due to Proposition 25. By compactness, there is a finite subset $\mathcal{B} \subseteq U^a_\mathcal{A}$ with $\mathcal{T}_1, \mathcal{B} \models q(a)$ and $\mathcal{T}_2, \mathcal{B} \not\models q(a)$. Clearly, we can also assume that $\mathcal{B}$ is connected.

stCQs. Since stCQs are already tree-shaped and have exactly one answer variable, the previous argument for unrestricted CQs reduces to observing Properties (a) and (c) and unraveling the witness ABox as described. ❑

B.3 Proof of Lemma 9

Lemma 9 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. Then $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff for all tree-shaped Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \rightarrow^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$.*

Proof. We prove both implications via contraposition.

"⇐". Assume $\mathcal{T}_1 \not\models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ and consider a witness $(\mathcal{A}, q, \mathbf{a})$, By Lemma 8, we can assume that $\mathcal{A}$ is tree-shaped. From Lemma 19 (4) we get $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q(\mathbf{a})$ and $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \not\models q(\mathbf{a})$. If we take the finite subinterpretation $\mathcal{I}$ of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ given by a match of q in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, then we must have $\mathcal{I} \not\rightarrow_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ because of $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \not\models q(\mathbf{a})$. Hence $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \not\rightarrow^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

"⇒". Assume $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \not\rightarrow^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$, i.e., there is a finite subinterpretation $\mathcal{I}$ of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with $\mathcal{I} \not\rightarrow_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$. Let $\mathbf{a}$ be the ABox individuals in $\mathcal{I}$ and let $q_\mathcal{I}$ be $\mathcal{I}$ viewed as a CQ whose variables correspond to the domain elements of $\mathcal{I}$ and the ABox individuals are represented by answer variables. Then it can be verified that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q_\mathcal{I}(\mathbf{a})$ and $\mathcal{I}_1 \not\models q_\mathcal{I}(\mathbf{a})$. ❑

B.4 Proof of Theorem 11

To prove the second part of Theorem 11 (the stCQ case), we need a bounded variant of simulations, analogously to bounded homomorphisms. We write $\mathcal{I}_1 \preceq^n_\Sigma \mathcal{I}_2$ if for any subinterpretation $\mathcal{I}'_1$ of $\mathcal{I}_1$ with $|\Delta^{\mathcal{I}'_1}| \leq n$, we have $\mathcal{I}'_1 \preceq_\Sigma \mathcal{I}_2$. Moreover, we write $\mathcal{I}_1 \preceq^{\mathsf{fin}}_\Sigma \mathcal{I}_2$ if $\mathcal{I}_1 \preceq^n_\Sigma \mathcal{I}_2$ for any n.

We begin with two useful insights about bounded homomorphisms (and simulations) and their connection to unbounded ones. We use $\mathcal{I}_1|^d_n$ to denote the restriction of $\mathcal{I}_1$ to elements that can be reached by starting at d and traveling along at most n role edges (forwards or backwards).

The first insight is straightforward.

Fact 26 *Let Σ be a signature and $\mathcal{I}_1, \mathcal{I}_2$ be interpretations such that $\mathcal{I}_1$ is finitely branching.*

(1) The following are equivalent.

(a) $\mathcal{I}_1 \rightarrow^{\mathsf{fin}}_\Sigma \mathcal{I}_2$

(b) For every $d \in \Delta^{\mathcal{I}_1}$ and every $i \geq 0 : \mathcal{I}_1|^d_i \rightarrow_\Sigma \mathcal{I}_2$

(2) The following are equivalent.

(a) $\mathcal{I}_1 \preceq^{\mathsf{fin}}_\Sigma \mathcal{I}_2$

(b) For every $d \in \Delta^{\mathcal{I}_1}$ and every $i \geq 0 : \mathcal{I}_1|^d_i \preceq_\Sigma \mathcal{I}_2$

We will thus use Conditions (1b) and (2b) as alternative characterizations of bounded homomorphisms and simulations.

The second insight shows that, under additional conditions, we can extract an unbounded homomorphism from a suitable family of bounded ones.

Lemma 27 *Let $\mathcal{I}_1, \mathcal{I}_2$ be finitely branching interpretations and let $\mathcal{I}_1$ be Σ-connected.*

(1) If there are $d_0 \in \Delta^{\mathcal{I}_1}$ and $e_0 \in \Delta^{\mathcal{I}_2}$ such that for each $i \geq 0$ there is a Σ-homomorphism h_i from $\mathcal{I}_1|^{d_0}_i$ to $\mathcal{I}_2$ with $h_i(d_0) = e_0$, then $\mathcal{I}_1 \rightarrow_\Sigma \mathcal{I}_2$.

(2) If there are $d_0 \in \Delta^{\mathcal{I}_1}$ and $e_0 \in \Delta^{\mathcal{I}_2}$ such that for each $i \geq 0$ there is a Σ-simulation ρ_i of $\mathcal{I}_1|^{d_0}_i$ in $\mathcal{I}_2$ with $(d_0, e_0) \in \rho_i$, then $\mathcal{I}_1 \preceq_\Sigma \mathcal{I}_2$.

Proof. We only show (1); Part (2) is analogous. We are going to construct a Σ-homomorphism h from $\mathcal{I}_1$ to $\mathcal{I}_2$ step by step, obtaining the desired homomorphism in the limit. We will take care that, at all times, the domain of h is finite and

(∗) there is a sequence $h_0, h_1, \ldots$ with h_i a Σ-homomorphism from $\mathcal{I}_1|^{d_0}_i$ to $\mathcal{I}_2$ such that whenever $h(d)$ is already defined, then $h_i(d) = h(d)$ for all $i \geq 0$.

Start with setting $h(d_0) = e_0$. The original sequence h_0, h_1 from the lemma witnesses (∗). Now consider the set Λ that consists of all elements $d \in \Delta^{\mathcal{I}_1}$ such that $h(d)$ is undefined and there is an $e \in \Delta^{\mathcal{I}_1}$ with $h(e)$ defined and such that d is reachable from e along a Σ-role edge. Since the domain of h is finite and $\mathcal{I}_1$ is finitely branching, Λ is finite. By (∗), since every $d \in \Lambda$ is reachable in one step from an element e such that $h(e)$ is defined, and since $\mathcal{I}_2$ is finitely branching, for each $d \in \Lambda$ there are only finitely many e' such that $h_i(d) = e'$ for some i. Thus there must be a function $\delta : \Lambda \rightarrow \Delta^{\mathcal{I}_2}$ such that, for infinitely many i, we have $h_i(d) = \delta(d)$ for all $d \in \Lambda$. Extend h accordingly, that is, set $h(d) = \delta(d)$ for all $d \in \Lambda$. Clearly, the sequence $h_0, h_1, \ldots$ from (∗) before the extension is no longer sufficient to witness (∗) after the extension. We fix this by skipping homomorphisms that do not respect δ, that is, define a new sequence $h'_0, h'_1, \ldots$ by using as h'_i the restriction of h_j to the domain of $\mathcal{I}_1|^{d_0}_i$ where $j \geq i$ is smallest such that $h_j(d) = \delta(d)$ for all $d \in \Lambda$. This finishes the construction. Note that we will automatically have $h(a) = a$ for all individual names a (as required), no matter whether d_0 is an individual name or not. ❑

We are now ready to prove Theorem 11.

Theorem 11 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Horn-$\mathcal{ALCHIF}$ TBoxes with $\mathcal{T}_1 \models^{\mathrm{RI}}_{\Gamma,\Sigma} \mathcal{T}_2$. Then $\mathcal{T}_1 \models^{\mathrm{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff for all tree-shaped Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, and for all tree-shaped, finitely branching models $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$, the following conditions are satisfied:*

(1) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \rightarrow_\Sigma \mathcal{I}_1$;

(2) for all Σ-subtrees $\mathcal{I}$ in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$, one of the following holds:

(a) $\mathcal{I} \rightarrow_\Sigma \mathcal{I}_1$;

(b) $\mathcal{I} \rightarrow^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathsf{tp}_{\mathcal{I}_1}(a)}$ for some $a \in \mathsf{ind}(\mathcal{A})$.

Furthermore, $\mathcal{T}_1 \models^{\mathrm{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ iff $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq_\Sigma \mathcal{I}_1$ for all $\mathcal{A}$ and $\mathcal{I}_1$ as above iff $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

Proof. Unrestricted CQs, "if". We show the contrapositive. Thus first assume that $\mathcal{T}_1 \not\models^{\mathsf{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$. By Lemma 8, there is a tree-shaped Γ-ABox $\mathcal{A}$ consistent with both $\mathcal{T}_i$, and a tree-shaped Σ-CQ q such that either

(1′) q has a single answer variable and there is an element $a \in \mathsf{ind}(\mathcal{A})$ such that $\mathcal{T}_2, \mathcal{A} \models q(a)$ but $\mathcal{T}_1, \mathcal{A} \not\models q(a)$ or

(2′) q is Boolean and $\mathcal{T}_2, \mathcal{A} \models q$ but $\mathcal{T}_1, \mathcal{A} \not\models q$.

In case (1′) holds, q is connected. Let h be a match of q in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$; in particular $h(x) = a$. Since q contains an answer variable, we must have $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \not\to_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ as otherwise the composition of h and the witnessing homomorphism shows $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \models q(a)$, which is not the case. Thus Condition (1) is violated for $\mathcal{I}_1 = \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

In case (2′) holds, consider again a match h of q in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$. Let $\mathcal{I}'_{\mathcal{T}_2,\mathcal{A}}$ be the restriction of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ to the elements in the range of h. Clearly, we have $\mathcal{I}'_{\mathcal{T}_2,\mathcal{A}} \not\to_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$. Consequently, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \not\to^n_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ where n is the number of variables in q, implying that Conditions (2a) and (2b) are both false.

Unrestricted CQs, "only if". Assume that $\mathcal{T}_1 \models^{\mathsf{CQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ and let $\mathcal{A}$ be a tree-shaped Γ-ABox consistent with both $\mathcal{T}_i$. We first show the following:

Claim. For all models $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$, we have $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \to^{\mathsf{fin}}_\Sigma \mathcal{I}_1$.

Proof of claim: Assume to the contrary that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \not\to^{\mathsf{fin}}_\Sigma \mathcal{I}_1$. Then $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \not\to^n_\Sigma \mathcal{I}_1$ for some n, that is, there is a subinterpretation $\mathcal{I}$ of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with $|\Delta^\mathcal{I}| \leq n$ such that $\mathcal{I} \not\to_\Sigma \mathcal{I}_1$. Let $\mathbf{a}$ be the ABox individuals in $\mathcal{I}$ and let $q_\mathcal{I}$ be $\mathcal{I}$ viewed as a CQ whose variables correspond to the domain elements of $\mathcal{I}$ and the ABox individuals are represented by answer variables. Then it can be verified that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q_\mathcal{I}(\mathbf{a})$ and $\mathcal{I}_1 \not\models q_\mathcal{I}(\mathbf{a})$.

Condition 1 is a consequence of Lemma 27: Fix a tree-shaped, finitely branching model $\mathcal{I}_1 \models (\mathcal{T}_1, \mathcal{A})$ and let $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma$ be the disjoint union of the connected interpretations $\mathcal{I}_1, \ldots, \mathcal{I}_k$. In each $\mathcal{I}_i$, we find at least one individual a_i from $\mathsf{ind}(\mathcal{A})$. Let $\ell \in \{1, \ldots, k\}$. By the claim above and Fact 26, we find a sequence $h_0, h_1, \ldots$ such that h_i is a Σ-homomorphism from $\mathcal{I}_\ell|^{a_\ell}_i$ to $\mathcal{I}_1$. Note that we must have $h_i(a_\ell) = a_\ell$ for all i. Thus, Lemma 27 yields $\mathcal{I}_\ell \to_\Sigma \mathcal{I}_1$ and, in summary, $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \to_\Sigma \mathcal{I}_1$.

Now for Condition 2. Let $\mathcal{I}$ be a Σ-subtree in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with root d_0. By the claim above and Fact 26, there is a sequence $h_0, h_1, \ldots$ such that h_i is a Σ-homomorphism from $\mathcal{I}|^{d_0}_i$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

First assume that there is an $e_0 \in \Delta^{\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}}$ such that $h_i(d_0) = e_0$ for infinitely many i. Construct a new sequence $h'_0, h'_1, \ldots$ with h'_i a Σ-homomorphism from $\mathcal{I}|^{d_0}_i$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ by skipping homomorphisms that do not map d_0 to e_0, that is, h'_i is the restriction of h_j to the domain of $\mathcal{I}|^{d_0}_i$ where $j \geq i$ is smallest such that $h_j(d_0) = e_0$. Clearly, $h'_i(d_0) = e_0$ for all i. Thus, Lemma 27 yields $\mathcal{I} \to_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ and thus, by Lemma 19 (2) $\mathcal{I} \to_\Sigma \mathcal{I}_1$ for every tree-shaped, finitely branching model $\mathcal{I}_1 \models (\mathcal{T}_1, \mathcal{A})$.

It remains to deal with the case that there is no $e_0 \in \Delta^{\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}}$ such that $h_i(d_0) = e_0$ for infinitely many i. We can assume that there is an $a_0 \in \mathsf{ind}(\mathcal{A})$ such that $h_i(d_0) \in \Delta^{\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}|_{a_0}}$ for all i; in fact, there must be an a_0 such that $h_i(d_0) \in \Delta^{\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}|_{a_0}}$ for infinitely many i and we can again skip homomorphisms to achieve this for all i. It is important to note that the remaining homomorphisms do not necessarily map all ancestors of d_0 in $\mathcal{I}$ to elements in $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}|_{a_0}$ due to the presence of inverse roles. Now, since $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ is are finitely branching, for all $i, n \geq 0$ we must find a $j \geq i$ such that $h_j(d_0)$ is a domain element whose distance from a_0 exceeds n (otherwise the previous case would apply). We can use this fact to construct a sequence $h'_0, h'_1, \ldots$ with h'_i a Σ-homomorphism from $\mathcal{I}|^{d_0}_i$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}|_{a_0}$. It is easy to verify that this implies $\mathcal{I} \to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}|_{a_0}$; in fact, $h'_0, h'_1, \ldots$ can again be found by skipping homomorphisms.

If we now fix an arbitrary (tree-shaped, finitely branching) model $\mathcal{I}_1 \models (\mathcal{T}_1, \mathcal{A})$, by Lemma 19 (2) and (3) we have $\mathsf{tp}_{\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}}(a_0) \subseteq \mathsf{tp}_{\mathcal{I}_1}(a_0)$ and thus $\mathcal{I}_{\mathcal{T}_1,\mathsf{tp}_{\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}}(a_0)} \to_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathsf{tp}_{\mathcal{I}_1}(a_0)}$. Hence $\mathcal{I} \to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathsf{tp}_{\mathcal{I}_1}(a_0)}$ as required.

stCQs. We need to show that the following three conditions are equivalent.

(i) $\mathcal{T}_1 \models^{\mathsf{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$

(ii) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq_\Sigma \mathcal{I}_1$ for all tree-shaped Γ-ABoxes $\mathcal{A}$ consistent with $\mathcal{T}_1$ and $\mathcal{T}_2$, and for all tree-shaped, finitely branching models $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A})$.

(iii) $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq_\Sigma \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ for all $\mathcal{A}$ as above.

(ii) ⇔ (iii). The "only if" direction follows from Lemma 19 (1); the "if" direction follows from $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}} \preceq_\Sigma \mathcal{I}_1$, which is a direct consequence of Lemma 19 (2).

(ii) ⇒ (i). This implication is analogous to the "if" direction of the case for unrestricted CQs above, except that the witness stCQ is rooted and connected, which rules out Case (2′) and thus Condition (2).

(i) ⇒ (ii). Assume that $\mathcal{T}_1 \models^{\mathsf{stCQ}}_{\Gamma,\Sigma} \mathcal{T}_2$ and let $\mathcal{A}$ be a tree-shaped Σ-ABox consistent with both $\mathcal{T}_i$. We first show the following:

Claim. For all models $\mathcal{I}_1$ of $(\mathcal{T}_1, \mathcal{A}) : \mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \preceq^{\mathsf{fin}}_\Sigma \mathcal{I}_1$.

Proof of claim: Assume to the contrary that $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \not\preceq^{\mathsf{fin}}_\Sigma \mathcal{I}_1$. Then $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_\Sigma \not\preceq^n_\Sigma \mathcal{I}_1$ for some n, that is, there is a subinterpretation $\mathcal{I}$ of $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ with $|\Delta^\mathcal{I}| \leq n$ such that $\mathcal{I} \not\preceq_\Sigma \mathcal{I}_1$. We can assume w.l.o.g. that $\mathcal{I}$ is connected and contains at least one ABox individual (otherwise we just extend $\mathcal{I}$ and increase n accordingly). Let $\mathbf{a}$ be the ABox individuals in $\mathcal{I}$ and let $q_\mathcal{I}$ be $\mathcal{I}$ viewed as a tree-shaped CQ whose variables correspond to the domain elements of $\mathcal{I}$ and the ABox individuals are represented by answer variables. Clearly $\mathcal{I} \models q(\mathbf{a})$ and thus $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q(\mathbf{a})$; let π be a match of q in $\mathcal{I}$. To transform q into an stCQ, perform the following operations.

- Remove all binary atoms involving only answer variables (see Condition (b) in the proof of Lemma 8).
- Restrict the resulting CQ to one connected component, with exactly one answer variable x (see Condition (d) in the proof of Lemma 8); then x is the root of the tree q. Let $a = \pi(x)$.
- "Split" multi-edges along the tree structure of q: if there are n binary atoms involving variables z_1, z_2 of q with z_2 being a child of z_1 in the tree q, introduce n copies of z_2

and its subtree, and redirect each of the n original atoms to its corresponding copy. Apply this step exhaustively.

The result of this transformation is an stCQ q', which still satisfies $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}} \models q'(a)$. On the other hand, $\mathcal{I}_1 \not\models q'(a)$ because, otherwise, a match π' of $q'(x)$ in $\mathcal{I}_1$ would give rise to a simulation of $\mathcal{I}$ in $\mathcal{I}_1$.

Having established the claim, we proceed as follows: Let a be an ABox individual in $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_{\Sigma}$. By the claim and Fact 26, there is a sequence $h_0, h_1, \dots$ such that h_i is a Σ-homomorphism from $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^a_i$ to $\mathcal{I}_1$. Obviously $h_i(a) = a$ for all i. From Lemma 27 we obtain $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}|^{\mathsf{con}}_{\Sigma} \preceq_\Sigma \mathcal{I}_1$ as desired. ❑

C Proofs for Section 4

C.1 Proof of Theorem 12

Theorem 12 *Given two Horn-$\mathcal{ALCHIF}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and types t_i for $\mathcal{T}_i$, $i \in \{1,2\}$, it can be decided in time $2^{2^{p(|\mathcal{T}_2|\log|\mathcal{T}_1|)}}$ whether $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_{\Sigma} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T}_1,t_1}$, p a polynomial.*

Proof. By Lemma 13, we can decide $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_{\Sigma} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T}_1,t_1}$ by checking whether there is a $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}_1, t_1)$ with $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_{\Sigma} \rightarrow_\Sigma \mathcal{J}$. By Lemma 14, this can be done by constructing the corresponding set $\mathcal{M}$ of mosaics for t_1, removing all mosaics that are not good, and checking whether the remaining set $\mathcal{M}_p$ contains a mosaic M with $t_2 \in T_M$.

The desired upper time bound is now a consequence of the following observations:

- The size of each 1-neighborhood in $\mathcal{I}_{\mathcal{T}_1,t_1}$ is bounded by $q(|\mathcal{T}_1|)$, for a polynomial q.
- The number of mosaics for t_1 is bounded by $2^{q'(|\mathcal{T}_1|)2^{|\mathcal{T}_2|}}$ for a polynomial q': there are at most $2^{|\mathcal{T}_1|^2}$ many 1-neighborhoods in $\mathcal{I}_{\mathcal{T}_1,t_1}$, and each such neighborhood admits at most $2^{|\mathcal{T}_1|q(|\mathcal{T}_1|)2^{|\mathcal{T}_2|}}$ many decorations with sets of types.
- Given a tuple (t^-, ρ, t, S, ℓ), one can decide in time $2^{\hat{q}(|\mathcal{T}_1|)}$, $\hat{q}$ a polynomial, whether (t^-, ρ, t, S) is a 1-neighborhood. Moreover, we can decide in time $2^{\hat{q}'(|\mathcal{T}_1|\cdot|\mathcal{T}_2|)}$, $\hat{q}'$ a polynomial, whether **(M)** is satisfied.
- Conditions 1 and 2 of a mosaic being good can be checked in the desired time.

❑

C.2 Proof of Lemma 13

Lemma 13 *Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox, $t_0 \in \mathsf{tp}(\mathcal{T})$, and $\mathcal{I}$ a tree-shaped interpretation. Then $\mathcal{I} \rightarrow^{\mathsf{fin}}_{\Sigma} \mathcal{I}_{\mathcal{T},t_0}$ iff there is a $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}, t_0)$ with $\mathcal{I} \rightarrow_\Sigma \mathcal{J}$.*

Proof. **“⇒”.** Let d_0 be the root of $\mathcal{I}$. By Fact 26, there is a sequence $h_0, h_1, \dots$ such that h_i is a Σ-homomorphism from $\mathcal{I}|^{d_0}_i$ to $\mathcal{I}_{\mathcal{T},t_0}$. Note that the set $\mathsf{tp}(\mathcal{T})$ is finite, and that $\mathcal{I}_{\mathcal{T},t_0}$ is finitely branching. By skipping homomorphisms, we can thus construct a new sequence $h'_0, h'_1, \dots$ such that h'_i is a Σ-homomorphism from $\mathcal{I}|^{d_0}_i$ to $\mathcal{I}_{\mathcal{T},t_0}$ and, additionally, for every $0 \le i \le j$ and $d \in \Delta^{\mathcal{I}|^{d_0}_i}$ the following properties hold:

(i) $n_1^{\mathcal{I}_{\mathcal{T},t_0}}(h'_i(d)) = n_1^{\mathcal{I}_{\mathcal{T},t_0}}(h'_j(d))$, and

(ii) If e is a successor of d in $\mathcal{I}$, then $h'_i(e)$ is a successor of $h'_i(d)$ in $\mathcal{I}_{\mathcal{T},t_0}$ iff $h'_j(e)$ is a successor of $h'_j(d)$.

Guided by h'_i, we construct a sequence of interpretations $\mathcal{J}_0, \mathcal{J}_1, \dots$ and a sequence $g_0, g_1, \dots$ with g_i a Σ-homomorphism from $\mathcal{I}|^{d_0}_i$ to $\mathcal{J}_i$ such that for all $0 \le i \le j$ and d in the domain of $\mathcal{I}|^{d_0}_i$, we have $g_i(d) = g_j(d)$. Throughout the construction, we maintain the invariant

$$n_1^{\mathcal{J}_i}(g_i(d)) \sqsubseteq n_1^{\mathcal{I}_{\mathcal{T},t_0}}(h'_i(d)) \qquad (*)$$

for all i, d such that $g_i(d)$ is defined.

We start with $\mathcal{J}_0 = (\{e_0\}, \cdot^{\mathcal{J}_0})$ such that $\mathsf{tp}_{\mathcal{J}_0}(e_0) = \mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(h'_0(d_0))$ and $g_0(d_0) = e_0$. Clearly $(*)$ is satisfied. Assuming that $\mathcal{J}_i$ and g_i are already defined, we extend them to $\mathcal{J}_{i+1}$ and g_{i+1} by doing the following for every $(d, d') \in \rho^{\mathcal{I}}$ with $d \in \Delta^{\mathcal{I}|^{d_0}_i}$ and $d' \notin \Delta^{\mathcal{I}|^{d_0}_i}$. By invariant $(*)$ and Item (i), we have $n_1^{\mathcal{J}_i}(g_i(d)) \sqsubseteq n_1^{\mathcal{I}_{\mathcal{T},t_0}}(h'_j(d))$ for all $j \ge i$; thus, we can apply **(R)** to $g_i(d)$ and $h'_i(d)$. More precisely, we obtain $\mathcal{J}_{i+1}$ by adding a predecessor and/or successors to achieve

$$n_1^{\mathcal{J}_{i+1}}(g_i(d)) = n_1^{\mathcal{I}_{\mathcal{T},t_0}}(h'_i(d)). \qquad (**)$$

To define $g_{i+1}(d')$, we distinguish two cases according to Item (ii):

- $h'_j(d')$ is a successor of $h'_j(d)$ for all $j \ge i$. Then there is some (ρ', t') in component S of $n_1^{\mathcal{I}_{\mathcal{T},t_0}}(h'_i(d))$ such that $(h'_i(d), h'_i(d')) \in \rho^{\mathcal{I}_{\mathcal{T},t_0}}$ (ρ maximal) and $\mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(h_i(d')) = t'$. By $(**)$ that pair is also in component S of $n_1^{\mathcal{J}_{i+1}}(g_i(d))$. Take a corresponding ρ'-successor e' of e in $\mathcal{J}_{i+1}$ and set $g_{i+1}(d') = e'$. Clearly $(*)$ is satisfied.
- $h'_j(d)$ is a successor of $h'_j(d')$ for all $j \ge i$. Then $t^- = \mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(h_i(d'))$ and ρ is maximal with $(h'_i(d'), h'_i(d)) \in \rho^{\mathcal{I}_{\mathcal{T},t_0}}$. By $(**)$, the t^-- and ρ-component in $n_1^{\mathcal{J}_{i+1}}(g_i(d))$ are identical. Take a corresponding ρ-predecessor e' of e in $\mathcal{J}_{i+1}$ and set $g_{i+1}(d') = e'$. Clearly $(*)$ is satisfied.

The construction of $\mathcal{J}$ and h is finished by setting $h = \bigcup_{i \geqslant 0} g_i$ and $\mathcal{J}' = \bigcup_{i \geqslant 0} \mathcal{J}_i$, and defining $\mathcal{J}$ as the result of exhaustive application of rule **(R)** to $\mathcal{J}'$.

“⇐”. It suffices to show $\mathcal{J} \rightarrow^{\mathsf{fin}} \mathcal{I}_{\mathcal{T},t_0}$.[3] To this end, denote with $\mathcal{J}_i$, $i \ge 0$, the finite submodel of $\mathcal{J}$ obtained after i rule applications, and with d_i the root of $\mathcal{J}_i$. We verify the following claim, which implies $\mathcal{J} \rightarrow^{\mathsf{fin}} \mathcal{I}_{\mathcal{T},t_0}$.

Claim. For all $i \ge 0$, we have:

(i) there is an $e_0 \in \Delta^{\mathcal{I}_{\mathcal{T},t_0}}$ with $\mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(e_0) = \mathsf{tp}_{\mathcal{J}_i}(d_i)$;

(ii) for all $e_0 \in \Delta^{\mathcal{I}_{\mathcal{T},t_0}}$ with $\mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(e_0) \supseteq \mathsf{tp}_{\mathcal{J}_i}(d_i)$, we have $(\mathcal{J}_i, d_i) \rightarrow (\mathcal{I}_{\mathcal{T},t_0}, e_0)$.

[3] We write $\mathcal{I} \rightarrow^{\mathsf{fin}} \mathcal{J}$ to denote that, for every $n \ge 0$, there are n-bounded homomorphisms from $\mathcal{I}$ to $\mathcal{J}$, without restricting the signature.

We prove the claim by induction on i. For $i = 0$, Points (i) and (ii) are clear by definition of $\mathcal{J}_0$. For the inductive step, consider $\mathcal{J}_{i+1}$ and suppose **(R)** has been applied to some $d \in \Delta^{\mathcal{J}_i}$ and $e \in \Delta^{\mathcal{I}_{\mathcal{T},t_0}}$.

Observe that Point (i) is trivially preserved when d is not the root of $\mathcal{J}_i$. In case $d = d_i$, it is preserved by the condition on the choice of e in **(R)**: e has the same type as d_i and, by construction, the predecessor e' of e (if it exists) has the same type as d_{i+1}.

For Point (ii), we distinguish two cases:

- The extension of $\mathcal{J}_i$ to $\mathcal{J}_{i+1}$ has not added any predecessors to d. In particular, we then have $d_{i+1} = d_i$. Let e_0 be as in (ii), i.e., $\mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(e_0) \supseteq \mathsf{tp}_{\mathcal{J}_i}(d_{i+1}) = \mathsf{tp}_{\mathcal{J}_i}(d_i)$. By induction hypothesis, there is a homomorphism $h : (\mathcal{J}_i, d_{i+1}) \to (\mathcal{I}_{\mathcal{T},t_0}, e_0)$. We extend h to the domain of $\mathcal{J}_{i+1}$ by doing the following for each newly added successor d' of d.
  Let $\mathsf{tp}_{\mathcal{J}_{i+1}}(d) = t$ and $\mathsf{tp}_{\mathcal{J}_{i+1}}(d') = t'$ and ρ maximal with $(d, d') \in \rho^{\mathcal{J}_{i+1}}$. By the choice of e in **(R)**, e is of type t and has a ρ-successor of type t'. By construction of the universal model, there is some $r \in \rho$ with $t \rightsquigarrow^{\mathcal{T}}_r t'$ and $\rho = \{s \mid \mathcal{T} \models r \sqsubseteq s\}$. Denote with $\hat{t} = \mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(h(d))$. The definition of a homomorphism yields $t \subseteq \hat{t}$. Thus, there is $\hat{t}' \supseteq t'$ such that $\hat{t} \rightsquigarrow^{\mathcal{T}}_r \hat{t}'$. By definition of the universal model, $h(d)$ has a ρ-successor of type $\hat{t}'$ or a ρ-predecessor of type $\hat{t}''$, for $\hat{t}'' \supseteq \hat{t}'$. We extend h by setting $h(d')$ to that predecessor or successor, respectively.
- The extension of $\mathcal{J}_i$ to $\mathcal{J}_{i+1}$ has added a ρ-predecessor d' to d. Then $d = d_i$ and $d' = d_{i+1}$. Let $\mathsf{tp}_{\mathcal{J}_{i+1}}(d) = t$ and $\mathsf{tp}_{\mathcal{J}_{i+1}}(d') = t'$. By construction of the universal model, there is $r \in \rho$ with $t' \rightsquigarrow^{\mathcal{T}}_r t$ and $\rho = \{s \mid \mathcal{T} \models r \sqsubseteq s\}$. Let e_0 be as in (ii), that is, $\hat{t}' := \mathsf{tp}_{\mathcal{I}_{\mathcal{T},t_0}}(e_0) \supseteq t'$. We then have that $\hat{t}' \rightsquigarrow_r \hat{t}$ for some $\hat{t} \supseteq t$. By definition of the universal model, e_0 has a ρ-successor of type $\hat{t}$ or a ρ-predecessor of type $\hat{t}'' \supseteq t$. Let this element be $\overline{e}_0$. By induction hypothesis, there is a homomorphism $h : (\mathcal{J}_i, d) \to (\mathcal{I}_{\mathcal{T},t_0}, \overline{e}_0)$. We extend h by first setting $h(d') = e_0$ and then extending h to all successors of d as in the previous case.

It should be clear that h, updated as above, witnesses $(\mathcal{J}_{i+1}, d_{i+1}) \to (\mathcal{I}_{\mathcal{T},t_0}, e_0)$. ❑

C.3 Proof of Lemma 14

Lemma 14 *Let $t_i \in \mathsf{tp}(\mathcal{T}_i)$ for $i \in \{1, 2\}$. Then there is a $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}_1, t_1)$ such that $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_\Sigma \to_\Sigma \mathcal{J}$ iff $\mathcal{M}_p$ contains a mosaic M with $t_2 \in \ell_M(t_M)$.*

Proof. **"⇒".** Let h be a Σ-homomorphism from $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_\Sigma$ to some $\mathcal{J} \in \mathsf{can}_\omega(\mathcal{T}_1, t_1)$. For every $d \in \Delta^{\mathcal{J}}$, denote with $T_h(d)$ the set of all types mapped to d by h, that is,

$$T_h(d) = \left\{\mathsf{tp}_{\mathcal{I}_{\mathcal{T}_2,t_2}}(e) \mid h(e) = d,\ e \in \Delta^{\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_\Sigma}\right\}.$$

For every element $d \in \Delta^{\mathcal{J}}$, we define a tuple $M(d) = (t^-, \rho, t, S, \ell)$ by taking:

- $(t^-, \rho, t, S) = n_1^{\mathcal{J}}(d)$;
- $\ell(t) = T_h(d)$;
- If there is a predecessor d' of d, then $\ell(t^-) = T_h(d')$; otherwise, set $\ell(t^-) = \emptyset$ (not important);
- For every successor d' of d with $\mathsf{tp}_{\mathcal{J}}(d') = t'$ and $\rho' = \{r \mid (d, d') \in r^{\mathcal{J}}\}$ add $(\rho', t') \in S$ and set $\ell(\rho', t') = T_h(d')$;

It is easy to verify that every $M(d) = (t^-, \rho, t, S, \ell)$ obtained in this way is actually a mosaic: By definition of $\mathcal{J}$, we know that $(t^-, \rho, t, S) = n_1^{\mathcal{I}_{\mathcal{T}_1,t_1}}(d')$ for some $d' \in \Delta^{\mathcal{I}_{\mathcal{T}_1,t_1}}$. Moreover, by definition of the universal model $\mathcal{I}_{\mathcal{T}_2,t_2}$ and the fact that h is a homomorphism, Condition **(M)** is satisfied.

Let $\mathcal{M}(\mathcal{J}) = \{M(d) \mid d \in \Delta^{\mathcal{J}}\}$. It follows from the construction that all mosaics in $\mathcal{M}(\mathcal{J})$ are good in $\mathcal{M}(\mathcal{J})$; hence $\mathcal{M}(\mathcal{J}) \subseteq \mathcal{M}_p$. Finally, let d_0 be the root of $\mathcal{I}_{\mathcal{T}_2,t_2}$. By definition of $M := M(h(d_0))$, we have $t_2 \in \ell_M(t_M)$.

"⇐". Assume $\mathcal{M}_p$ contains a mosaic M with $t_2 \in \ell_M(t_M)$. We define the interpretation $\mathcal{J}$ as the limit of the following process. We maintain a partial function $q : \Delta^{\mathcal{J}} \to \mathcal{M}_p$, intuitively mapping each domain element of $\mathcal{J}$ to the mosaic that gave rise to it. Throughout the construction, the following invariant is preserved:

$$\text{If } q(d) = (t^-, \rho, t, S, \ell), \text{ then } n_1^{\mathcal{J}}(d) = (t^-, \rho, t, S). \quad (*)$$

We start with defining $\mathcal{J}$ as the interpretation corresponding to the 1-neighborhood represented by M, and define $q(e_0) = M$, where e_0 is the "center" of that 1-neighborhood. By definition, the invariant $(*)$ is satisfied. Then extend $\mathcal{J}$ by applying the following step exhaustively in a fair way: Choose some $d \in \mathcal{J}$ such that $q(d)$ is undefined, and:

- If d has a predecessor d' such that $q(d') = M'$ then, due to $(*)$, there is $(\rho, t) \in S_{M'}$ such that $(d', d) \in \rho^{\mathcal{J}}$ and $\mathsf{tp}_{\mathcal{J}}(d) = t$. Let $N \in \mathcal{M}_p$ be the mosaic that exists according to Condition 1 of being good for $(\rho, t) \in S_{M'}$. Then extend $\mathcal{J}$ such that $n_1^{\mathcal{J}}(d) = (t_N^-, \rho_N, t_N, S_N)$ and set $q(d) = N$.
- If d has a successor d' such that $q(d') = M'$ then, due to $(*)$, we know that $t^-_{M'} = \mathsf{tp}_{\mathcal{J}}(d) \neq \bot$. Let $N \in \mathcal{M}_p$ be the mosaic that exists according to Condition 2 of being good. Then extend $\mathcal{J}$ such that $n_1^{\mathcal{J}}(d) = (t_N^-, \rho_N, t_N, S_N)$ and set $q(d) = N$.

It is immediate from the construction that these steps preserve $(*)$, and that always one of the cases applies. Moreover, by construction, any interpretation $\mathcal{J}$ obtained in the limit of such a process is an element of $\mathsf{can}_\omega(\mathcal{T}_1, t_1)$. It thus remains to construct a Σ-homomorphism h witnessing $\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_\Sigma \to_\Sigma \mathcal{J}$. We proceed again inductively, maintaining the invariant:

$$\text{If } h(d) \text{ is defined, then } \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_2,t_2}}(d) \in \ell_{q(h(d))}(t_{q(h(d))}). \quad (\dagger)$$

Let d_0 be the root of $\mathcal{I}_{\mathcal{T}_2,t_2}$. We start with setting $h(d_0) = e_0$, where e_0 is as above. By the assumption that $t_2 \in \ell_M(t_M)$, invariant $(\dagger)$ is satisfied. Now, exhaustively apply the following step. Choose $d \in \Delta^{\mathcal{I}_{\mathcal{T}_2,t_2}|^{\mathsf{con}}_\Sigma}$ such that $h(d)$ is not defined but $h(d') = e$ is defined for the predecessor d' of d. Let

$t = \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_2,t_2}}(d)$, $t' = \mathsf{tp}_{\mathcal{I}_{\mathcal{T}_2,t_2}}(d')$, and $M' = q(d')$. By definition of $\mathcal{I}_{\mathcal{T}_2,t_2}$, we know that $t' \rightsquigarrow_r^{\mathcal{T}_2} t$ for some $r \in \mathsf{rol}(\mathcal{T}_2)$. Let $\sigma = \{s \mid \mathcal{T} \models r \sqsubseteq s\}$. By invariant (†), we know that $t' \in \ell_{M'}(t_{M'})$. Thus, one of (a)–(c) in Condition **(M)** applies. Since $d, d' \in \Delta^{\mathcal{I}_{\mathcal{T}_2,t_2}|_\Sigma^{\mathsf{con}}}$, we know that $\sigma|_\Sigma \neq \emptyset$, thus only (b) or (c) are possible. In case of (b), we extend h by setting $h(d)$ to the predecessor of $h(d')$. In case of (c), we extend h by setting $h(d)$ to the according successor of $h(d')$. Note that h extended like this satisfies the homomorphism conditions and preserves (†) due to the conditions in (b) and (c). ❑

C.4 Proof of Theorem 15

We first make precise the semantics of 2ATA$_c$. Let (T, L) be a Θ-labeled tree and $\mathfrak{A} = (Q, \Theta, q_0, \delta, \Omega)$ a 2ATA$_c$. A *run of* $\mathfrak{A}$ *over* (T, L) is a $T \times Q$-labeled tree (T_r, r) such that $\varepsilon \in T_r$, $r(\varepsilon) = (\varepsilon, q_0)$, and for all $y \in T_r$ with $r(y) = (x, q)$ and $\delta(q, V(x)) = \theta$, there is an assignment v of truth values to the transition atoms in θ such that v satisfies θ and:

- if $v(q') = 1$, then $r(y') = (x, q')$ for some successor y' of y in T_r;
- if $v(\langle - \rangle q') = 1$, then $x \neq \varepsilon$ and $r(y') = (x \cdot -1, q')$ for some successor y' of y in T_r;
- if $v([-]q') = 1$, then $x = \varepsilon$ or $r(y') = (x \cdot -1, q')$ for some successor y' of y in T_r;
- if $v(\Diamond_n q') = 1$, then there are pairwise different $i_1, \dots, i_n$ such that, for each j, there is some successor y' of y in T_r with $r(y') = (x \cdot i_j, q')$;
- if $v(\Box_n q') = 1$, then for all but n successors x' of x, there is a successor y' of y in T_r with $r(y') = (x', q')$.

Let $\gamma = i_0 i_1 \cdots$ be an infinite path in T_r and denote, for all $j \geq 0$, with q_j the state such that $r(i_j) = (x, q_j)$. The path γ is *accepting* if the largest number m such that $\Omega(q_j) = m$ for infinitely many j is even. A run (T_r, r) is accepting, if all infinite paths in T_r are accepting. $\mathfrak{A}$ accepts a tree if $\mathfrak{A}$ has an accepting run over it.

Theorem 15 *The emptiness problem for 2ATA$_c$ can be solved in time exponential in the number of states.*

The proof is by reduction to the emptiness problem of standard two-way alternating tree automata on trees of some fixed outdegree [Vardi, 1998]. We need to introduce strategy trees similar to [Vardi, 1998, Section 4]. A *strategy tree for* $\mathfrak{A}$ is a tree (T, τ) where τ labels every node in T with a subset $\tau(x) \subseteq 2^{Q \times \mathbb{N} \cup \{-1\} \times Q}$, that is, with a graph with nodes from Q and edges labeled with natural numbers or -1. Intuitively, $(q, i, p) \in \tau(x)$ expresses that, if we reached node x in state q, then we should send a copy of the automaton in state p to $x \cdot i$. For each label ζ, we define $\mathsf{state}(\zeta) = \{q \mid (q, i, q') \in \zeta\}$, that is, the set of sources in the graph ζ. A strategy tree is *on an input tree* (T', L) if $T = T'$, $q_0 \in \mathsf{state}(\tau(\varepsilon))$, and for every $x \in T$, the following conditions are satisfied:

(i) if $(q, i, p) \in \tau(x)$, then $x \cdot i \in T$;

(ii) if $(q, i, p) \in \tau(x)$, then $p \in \mathsf{state}(\tau(x \cdot i))$;

(iii) if $q \in \mathsf{state}(\tau(x))$, then the truth assignment $v_{q,x}$ defined below satisfies $\delta(q, L(x))$:

- $v_{q,x}(p) = 1$ iff $(q, 0, p) \in \tau(x)$;
- $v_{q,x}(\langle - \rangle p) = 1$ iff $(q, -1, p) \in \tau(x)$;
- $v_{q,x}([-]p) = 1$ iff $x = \varepsilon$ or $(q, -1, p) \in \tau(x)$;
- $v_{q,x}(\Diamond_n p) = 1$ iff $(q, i, p) \in \tau(x)$ for n pairwise distinct $i \geq 1$;
- $v(\Box_n p) = 1$ iff for all but at most n values $i \geq 1$ with $x \cdot i \in T$, we have $(q, i, p) \in \tau(x)$.

A *path* β in a strategy tree (T, τ) is a sequence $\beta = (u_1, q_1)(u_2, q_2) \cdots$ of pairs from $T \times Q$ such that for all $i > 0$, there is some c_i such that $(q_i, c_i, q_{i+1}) \in \tau(u_i)$ and $u_{i+1} = u_i \cdot c_i$. Thus, β is obtained by moves prescribed in the strategy tree. We say that β is accepting if the largest number m such that $\Omega(q_i) = m$, for infinitely many i, is even. A strategy tree (T, τ) is accepting if all infinite paths in (T, τ) are accepting.

Lemma 28 *A 2ATA$_c$ accepts an input tree iff there is an accepting strategy tree on the input tree.*

Proof. The "if"-direction is immediate: just read off an accepting run from the accepting strategy tree.

For the "only if"-direction, we observe that acceptance of an input tree can be defined in terms of a parity game between Player 1 (trying to show that the input is accepted) and Player 2 (trying to challenge that). The initial configuration is (ε, q_0) and Player 1 begins. Consider a configuration (x, q). Player 1 chooses a satisfying truth assignment v of $\delta(q, L(x))$. Player 2 chooses an atom α with $v_{q,x}(\alpha) = 1$ and determines the next configuration as follows:

- if $\alpha = p$, then the next configuration is (x, p),
- if $\alpha = \langle - \rangle p$, then the next configuration is $(x \cdot -1, p)$ unless $x = \varepsilon$; in this case, Player 1 loses immediately;
- if $\alpha = [-]p$, then the next configuration is $(x \cdot -1, p)$ unless $x = \varepsilon$; in this case, Player 2 loses immediately;
- if $\alpha = \Diamond_n p$, then Player 1 selects pairwise distinct $i_1, \dots, i_n$ with $x \cdot i_j \in T$, for all j (and loses if she cannot); Player 2 then chooses some i_j and the next configuration is $(x \cdot i_j, p)$;
- if $\alpha = \Box_n p$, then Player 1 selects n values $i_1, \dots, i_n$; Player 2 then chooses some $\ell \notin \{i_1, \dots, i_n\}$ such that $x \cdot \ell \in T$ (and loses if he cannot) and the next configuration is $(x \cdot \ell, p)$.

Player 1 wins an infinite play $(x_0, q_0)(x_1, q_1) \cdots$ if the largest number m such that $\Omega(q_i) = m$, for infinitely many i, is even. It is not difficult to see that Player 1 has a winning strategy on an input tree iff $\mathfrak{A}$ accepts the input tree.

Observe now that the defined game is a parity game and thus Player 1 has a winning strategy iff she is has a *memoryless* winning strategy [Emerson and Jutla, 1991]. It remains to observe that a memoryless winning strategy is nothing else than an accepting strategy tree. ❑

Lemma 29 *If $L(\mathfrak{A}) \neq \emptyset$, then there is some $(T, L) \in L(\mathfrak{A})$ such that T has outdegree at most $n \cdot C$, where n is the number of states in $\mathfrak{A}$ and C is the largest number in (some transition $\Diamond_m p$ or $\Box_m p$ in) δ.*

Proof. Let (T, L) be an input tree and τ an accepting strategy tree on T, and let C be the largest number appearing in δ. We inductively construct a tree (T', L') with $T' \subseteq T$ and L' the restriction of L to T' and an accepting strategy tree τ' on (T', L'). For the induction base, we start with $T' = \{\varepsilon\}$ and τ' the empty mapping. For the inductive step, assume that $\tau'(x)$ is still undefined for some $x \in T'$, and proceed as follows:

1. For every $(q, i, p) \in \tau(x)$ with $i \in \{-1, 0\}$, add $(q, i, p) \in \tau'(x)$.
2. For every $p \in Q$, define $N_p = \{i \geq 1 \mid (q, i, p) \in \tau(x), x \cdot i \in T\}$ and let $N'_p \subseteq N_p$ be a subset of N_p with precisely $\min(C, |N_p|)$ elements. Then:
   (a) for all $i \in N'_p$, add $x \cdot i \in T'$;
   (b) for all $(q, i, p) \in \tau(x)$ with $i \in N'_p$, add $(q, i, p) \in \tau'(x)$;
   (c) for all $q \in \mathsf{state}(x)$ and $i \in N'_p$, add $(q, i, p) \in \tau'(x)$.

By Step 2 above, T' has outdegree bounded by $|Q| \cdot C$. It remains to show that τ' is an accepting strategy tree on T'. Observe first that, by construction, $q_0 \in \mathsf{state}(\tau'(\varepsilon))$.

We verify Conditions (i)–(iii) of a strategy tree being on an input tree. Conditions 1 follows directly from the construction. For (ii), assume that $(q, i, p) \in \tau'(x)$. By construction, there is some q' with $(q', i, p) \in \tau(x)$, and, by Condition (ii) $p \in \mathsf{state}(\tau(x \cdot i))$. Hence, there is some $(p, j, p') \in \mathsf{state}(\tau(x \cdot i))$. By construction, there is also some $(p, j', p') \in \mathsf{state}(\tau'(x \cdot i)$, thus $p \in \mathsf{state}(x \cdot i)$. For Condition (iii), take any $x \in T'$ and $q \in \mathsf{state}(\tau'(x))$. As $q \in \mathsf{state}(\tau(x))$, we know that the truth assignment $v_{q,x}$ defined for τ in Condition (iii) satisfies $\delta(q, L(x))$. We show that for all transitions α with $v_{q,x}(\alpha) = 1$, we also have $v'_{q,x}(\alpha) = 1$, where $v'_{q,x}$ is the truth assignment defined for τ'. By Step 1 of the construction, this is true for all α of the shape p, $\langle - \rangle p$, and $[-]p$. Let now be $\alpha = \Diamond_k p$, that is, there are k pairwise distinct $i \geq 1$ such that $(q, i, p) \in \tau(x)$. By the choice of C, we have $|N'_P| \geq k$. By Step 2(c), we know that there are k pairwise distinct i such that $(q, i, p) \in \tau'(x)$, hence $v'_{q,x}(\alpha) = 1$. Consider now $\alpha = \Box_k p$, that is, for all but at most k values $i \geq 1$ with $x \cdot i \in T$, we have $(q, i, p) \in \tau(x)$. By Step 2(b), this remains true for τ', hence $v'_{q,x}(\alpha) = 1$.

We finally argue that τ' is also accepting. Let $\beta = (u_1, q_1)(u_2, q_2) \cdots$ be an infinite path in (T', τ'). We construct an infinite path $\beta' = (u'_1, q_1)(u'_2, q_2)(u'_3, q_3) \cdots$ in (T, τ) as follows:

– $u'_1 = u_1$;
– Let $u_{i+1} = u_i \cdot \ell$ for some ℓ with $(q_i, \ell, q_{i+1}) \in \tau'(x)$. If $\ell \in \{0, 1\}$, we have $(q_i, \ell, q_{i+1}) \in \tau(x)$, by Step 1. We set $u'_{i+1} = u'_i \cdot \ell$. If $\ell \geq 0$ then, by Step 2(c), there is some ℓ' with $(q_i, \ell', q_{i+1}) \in \tau(x)$ and $x \cdot \ell' \in T'$. Set $u'_{i+1} = u'_i \cdot \ell'$.

Since every infinite path in (T, τ) is accepting, so is β', and thus β. ❏

We now reduce to reduce the emptiness problem of 2ATA$_c$ to the emptiness of alternating automata running on trees of fixed outdegree [Vardi, 1998], which we recall here. A tree T is *k-ary* if every node has exactly k. A *two-way alternating tree automaton over k-ary trees (2ATA^k)* that are Θ-labeled is a tuple $\mathcal{A} = (Q, \Theta, q_0, \delta, \Omega)$ where Q is a finite set of *states*, Θ is the *input alphabet*, $q_0 \in Q$ is an *initial state*, δ is the *transition function*, and $\Omega : Q \to \mathbb{N}$ is a *priority function*. The transition function maps a state q and some input letter θ to a *transition condition* $\delta(q, \theta)$, which is a positive Boolean formula over the truth constants true, false, and transitions of the form $(i, q) \in [k] \times Q$ where $[k] = \{-1, 0, \ldots, k\}$. A *run* of $\mathcal{A}$ on a Θ-labeled tree (T, L) is a $T \times Q$-labeled tree (T_r, r) such that

1. $r(\varepsilon) = (\varepsilon, q_0)$;
2. for all $x \in T_r$ with $r(x) = (w, q)$ and $\delta(q, \tau(w)) = \varphi$, there is a (possibly empty) set $\mathcal{S} = \{(m_1, q_1), \ldots, (m_n, q_n)\} \subseteq [k] \times Q$ such that $\mathcal{S}$ satisfies φ and for $1 \leq i \leq n$, we have $x \cdot i \in T_r$, $w \cdot m_i$ is defined, and $\tau_r(x \cdot i) = (w \cdot m_i, q_i)$.

Accepting runs and accepted trees are defined as for 2ATA$_c$s. The emptiness problem for 2ATA^{k}s can be solved in time exponential in the number of states [Vardi, 1998].

Theorem 15 *The emptiness problem for 2ATA$_c$ can be solved in time exponential in the number of states.*

Proof. Let $\mathfrak{A} = (Q, \Theta, q_0, \delta, \Omega)$ be an 2ATA$_c$ with n states and C the largest number in δ. We translate $\mathfrak{A}$ to a to a 2ATA^k $\mathfrak{A}' = (Q', \Theta', q'_0, \delta', \Omega)$ with $k = n \cdot C$, the bound from Lemma 29. Set $Q' = Q \cup \{q'_0, q_1, q_r, q_\perp\}$ and $\Theta' = (\Theta \cup \{d_\perp\}) \times \{0, 1\}$. The extended alphabet and the extra states are used to simulate transitions of the form $[-]p$ and to allow for input trees of outdegree less than k.

We obtain δ' from δ by replacing q with $(0, 1)$, $\langle - \rangle q$ with $(-1, q)$ and $[-]q$ with $(0, q_r) \vee (-1, q)$. Moreover, we replace

– $\Diamond_n q$ with $\bigvee_{X \in \binom{\{1, \ldots, N\}}{n}} \bigwedge_{i \in X} (i, q)$, and
– $\Box_n q$ with $\bigvee_{X \in \binom{\{1, \ldots, N\}}{n}} \bigwedge_{i \in \{1, \ldots, N\} \setminus X} (i, q)$,

where, as usual, $\binom{M}{m}$ denotes the set of all m-elementary subsets of a set M. To deal with the case of smaller outdegree, we use the fresh symbol $d_\perp$ as follows:

– For all $q \in Q'$: $\delta(q, (d_\perp, b)) = \begin{cases} \mathsf{true} & \text{if } b = 0 \\ \mathsf{false} & \text{if } b = 1 \end{cases}$

To enforce the intended labeling in the second component and the correct behaviour for q_r, we set:

$$\begin{aligned} \delta'(q'_0, (\theta, b)) &= \begin{cases} \mathsf{false} & \text{if } b = 0 \\ q_0 \wedge \bigwedge_{i=1}^{k} (i, q_1) & \text{otherwise} \end{cases} \\ \delta'(q_1, (\theta, b)) &= \begin{cases} \bigwedge_{i=1}^{k} (i, q_1) & \text{if } b = 0 \\ \mathsf{false} & \text{otherwise} \end{cases} \\ \delta'(q_r, (\theta, b)) &= \begin{cases} \mathsf{true} & \text{if } b = 1 \\ \mathsf{false} & \text{otherwise} \end{cases} \end{aligned}$$

Using Lemma 29, it is easy to verify that $L(\mathfrak{A})$ is empty iff $L(\mathfrak{A}')$ is empty. Moreover, since emptiness of 2ATA^{k}s can be checked in exponential time in the number of states, this finishes the proof of Theorem 15. ❏

C.5 Proof of Theorem 16

Theorem 16 *In Horn-$\mathcal{ALCHIF}$, the following problems can be decided in time $2^{2^{p(|\mathcal{T}_2|^{\log|\mathcal{T}_1|})}}$, p a polynomial: (Γ,Σ)-CQ entailment, inseparability, and conservative extensions. The same holds for Σ-deductive entailment in $\mathcal{ELHIF}_\bot$.*

We show the following lemma which, together with Theorem 15 and Lemma 7, implies Theorem 16.

Lemma 30 *There are 2ATA$_c$ $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3, \mathfrak{A}_4, \mathfrak{A}'_4$ such that:*

- *$\mathfrak{A}_1$ accepts (T,L) iff $\mathcal{A}$ is finite, tree-shaped, and contains ε;*
- *$\mathfrak{A}_2$ accepts (T,L) iff $\mathcal{I}_1$ is a model of $\mathcal{A}$ and $\mathcal{T}_1$;*
- *$\mathfrak{A}_3$ accepts (T,L) iff $\mathcal{A}$ is consistent with $\mathcal{T}_2$, and $\mathcal{I}_2$ is $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ restricted to* $\mathsf{ind}(\mathcal{A})$*;*
- *$\mathfrak{A}_4$ accepts (T,L) iff either (1) or (2) from Theorem 11 is* not *satisfied, when $\mathcal{I}_{\mathcal{T}_2,\mathcal{A}}$ is replaced with $\mathcal{I}_2$.*
- *$\mathfrak{A}'_4$ accepts (T,L) iff $\mathcal{I}_2|^{\mathsf{con}}_\Sigma \not\preceq_\Sigma \mathcal{I}_1$.*

The number of states of $\mathfrak{A}_1$ and $\mathfrak{A}_2$ is polynomial in $|\mathcal{T}_1|$ (and independent of $\mathcal{T}_2$); the number of states of $\mathfrak{A}_3$ is polynomial in $|\mathcal{T}_2|$ (and independent of $\mathcal{T}_1$), and the number of states of $\mathfrak{A}_4, \mathfrak{A}'_4$ is exponential in $|\mathcal{T}_2|$ (and independent of $\mathcal{T}_1$). All automata can be constructed in time polynomial in $|\mathcal{T}_1|$ and double-exponential in $|\mathcal{T}_2|$.

The construction of the automaton $\mathfrak{A}_1$ is straightforward, so we concentrate on $\mathfrak{A}_2$, $\mathfrak{A}_3$, and $\mathfrak{A}_4$.

In what follows, we use $\Diamond q$ and $\Box q$ to abbreviate $\Diamond_1 q$ and $\Box_0 q$, respectively. We define $\mathfrak{A}_2 = (Q_2, \Theta, q_0, \delta_2, \Omega_2)$ where

$$\begin{aligned} Q_2 = {} & \{q_0, q_\mathcal{A}\} \cup \{q_\alpha \mid \alpha \in \mathcal{T}_1\} \cup \{q_\rho, \overline{q}_\rho \mid \rho \in \Theta_1\} \cup \\ & \{q_{r,B}, q^\downarrow_{r,B}, \overline{q}_{r,B}, \overline{q}^\downarrow_{r,B} \mid \exists r.B \in \mathsf{cl}(\mathcal{T}_1)\}, \end{aligned}$$

and Ω_2 assigns 0 to all states. The idea of $\mathfrak{A}_2$ is to check that the ABox is satisfied, realized in state $q_\mathcal{A}$, and that every axiom TBox axiom in $\mathcal{T}_1$ is satisfied everywhere, realized using states q_α below. Formally, the transition function δ_2 is given as follows, for $\sigma = (L_0, L_1, L_2)$:

$$\begin{aligned}
\delta_2(q_0,\sigma) &= \Box q_0 \wedge q_\mathcal{A} \wedge \textstyle\bigwedge_{\alpha\in\mathcal{T}_1} q_\alpha \\
\delta_2(q_\mathcal{A},\sigma) &= \bigwedge_{\rho\in L_0} q_\rho \\
\delta_2(q_{\mathsf{func}}(r),\sigma) &= (q_{r^-} \wedge \Box\overline{q}_r) \vee (\overline{q}_{r^-} \wedge \Box_1\overline{q}_r) \\
\delta_2(q_{r\sqsubseteq s},\sigma) &= \overline{q}_r \vee q_s \\
\delta_2(q_{A_1\sqcap A_2\sqsubseteq B},\sigma) &= \overline{q}_{A_1} \vee \overline{q}_{A_2} \vee q_B \\
\delta_2(q_{A\sqsubseteq\bot},\sigma) &= \overline{q}_A \\
\delta_2(q_{\top\sqsubseteq A},\sigma) &= q_A \\
\delta_2(q_{A\sqsubseteq\exists r.B},\sigma) &= \overline{q}_A \vee q_{r,B} \\
\delta_2(q_{\exists r.A\sqsubseteq B},\sigma) &= \overline{q}_{r,A} \vee q_B \\
\delta_2(q_{r,B},\sigma) &= \Diamond q^\downarrow_{r,B} \vee (q_{r^-} \wedge \langle -\rangle q_B) \\
\delta_2(\overline{q}_{r,B},\sigma) &= \Box\overline{q}^\downarrow_{r,B} \wedge (\overline{q}_{r^-} \vee [-]\overline{q}_B) \\
\delta_2(q^\downarrow_{r,B},\sigma) &= q_r \wedge q_B \\
\delta_2(\overline{q}^\downarrow_{r,B},\sigma) &= \overline{q}_r \vee \overline{q}_B
\end{aligned}$$

Finally, we set for all $\rho \in \Theta_1$:

$$\delta_2(q_\rho,\sigma) = \begin{cases} \mathsf{true} & \text{if } \rho \in L_1 \\ \mathsf{false} & \text{if } \rho \notin L_1 \end{cases}$$

$$\delta_2(\overline{q}_\rho,\sigma) = \begin{cases} \mathsf{true} & \text{if } \rho \notin L_1 \\ \mathsf{false} & \text{if } \rho \in L_1 \end{cases}$$

Automaton $\mathfrak{A}_3$ relies on a syntactic characterization of ABox entailment [Bienvenu *et al.*, 2013], which we introduce first.

Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox and $\mathcal{A}$ a tree-shaped ABox. A *derivation tree* for an assertion $A_0(a_0)$ in $\mathcal{A}$ w.r.t. $\mathcal{T}$ with $A_0 \in \mathsf{N_C}$ is a finite $\mathsf{ind}(\mathcal{A}) \times \mathsf{N_C}$-labeled tree (T,V) that satisfies the following conditions:

1. $V(\varepsilon) = (a_0, A_0)$;
2. if $V(x) = (a, A)$ and neither $A(a) \notin \mathcal{A}$ nor $\top \sqsubseteq A \in \mathcal{T}$, then one of the following holds:
   - (i) x has successors $y_1, \dots, y_k$, $k \geq 1$ with $V(y_i) = (a, A_i)$ for $1 \leq i \leq k$ and $\mathcal{T} \models A_1 \sqcap \dots \sqcap A_k \sqsubseteq A$;
   - (ii) x has a single successor y with $V(y) = (b, B)$ and there is an $\exists r.B \sqsubseteq A \in \mathcal{T}$ and an $s(a,b) \in \mathcal{A}$ such that $\mathcal{T} \models s \sqsubseteq r$;
   - (iii) x has a single successor y with $V(y) = (b, B)$ and there is a $B \sqsubseteq \exists r.A \in \mathcal{T}$ such that $r(b,a) \in \mathcal{A}$ and $\mathsf{func}(r) \in \mathcal{T}$.

Note that the first item of Point 2 above requires $\mathcal{T} \models A_1 \sqcap \dots \sqcap A_n \sqsubseteq A$ instead of $A_1 \sqcap A_2 \sqsubseteq A \in \mathcal{T}$ to 'shortcut' anonymous parts of the universal model. In fact, the derivation of A from $A_1 \sqcap \dots \sqcap A_n$ by $\mathcal{T}$ can involve the introduction of anonymous elements.

The main property of derivation trees is the following.

Lemma 31 *Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox and $\mathcal{A}$ an ABox consistent with $\mathcal{T}$. Then for all assertions $A(a)$ with $A \in \mathsf{N_C}$, and $a \in \mathsf{ind}(\mathcal{A})$ we have $\mathcal{T}, \mathcal{A} \models A(a)$ iff there is a derivation tree for $A(a)$ in $\mathcal{A}$ w.r.t. $\mathcal{T}$.*

Proof. The "if"-direction is immediate, so we concentrate on the "only if"-direction. We construct a sequence of interpretations $\mathcal{I}_0, \mathcal{I}_1, \dots$ by the following procedure. We start with setting:

$$\begin{aligned} \Delta^{\mathcal{I}_0} &= \mathsf{ind}(\mathcal{A}) \\ A^{\mathcal{I}_0} &= \{a \mid A(a) \in \mathcal{A}\} \\ r^{\mathcal{I}_0} &= \{(a,b) \mid r(a,b) \in \mathcal{A}\} \end{aligned}$$

For every $i \geq 0$, we obtain $\mathcal{I}_{i+1}$ from $\mathcal{I}_i$ by setting $\mathcal{I}_{i+1} = \mathcal{I}_i$ and applying the following rules to all $d, e \in \Delta^{\mathcal{I}_i}$:

1. If $d \in (A_1 \sqcap A_2)^{\mathcal{I}_i}$, but $d \notin A^{\mathcal{I}_i}$ for some $A_1 \sqcap A_2 \sqsubseteq A \in \mathcal{T}$, then add $d \in A^{\mathcal{I}_{i+1}}$;
2. If $d \in (\exists r.B)^{\mathcal{I}_i}$, but $d \notin A^{\mathcal{I}_i}$ for some $\exists r.B \sqsubseteq A \in \mathcal{T}$, then add $d \in A^{\mathcal{I}_{i+1}}$;
3. If $(d,e) \in r^{\mathcal{I}_i}$ but $(d,e) \notin s^{\mathcal{I}_i}$, for some s with $\mathcal{T} \models r \sqsubseteq s$, then add $(d,e) \in s^{\mathcal{I}_{i+1}}$;
4. If $d \in A^{\mathcal{I}_i}$, but $d \notin (\exists r.B)^{\mathcal{I}_i}$ for some $A \sqsubseteq \exists r.B \in \mathcal{T}$, then:

(a) if there is e with $(d,e) \in r^{\mathcal{I}_i}$ and $\mathsf{func}(r) \in \mathcal{T}$ then add $e \in B^{\mathcal{I}_{i+1}}$;

(b) otherwise add a fresh domain element e with $(d,e) \in r^{\mathcal{I}_{i+1}}$ and $e \in B^{\mathcal{I}_{i+1}}$.

Let $\mathcal{I}$ be defined as $\Delta^{\mathcal{I}} = \bigcup_{i\geq 0} \mathcal{I}_i$, $A^{\mathcal{I}_i} = \bigcup_{i\geq 0} A^{\mathcal{I}_i}$, and $r^{\mathcal{I}} = \bigcup_{i\geq 0} r^{\mathcal{I}_i}$. It is standard to verify the following:

Claim 1. $\mathcal{I} \to \mathcal{J}$ for all models $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$.

By definition of $\mathcal{I}_0$, we have $\mathcal{I} \models \mathcal{A}$. Moreover, we have $\mathcal{I} \models \mathcal{T}'$ where $\mathcal{T}' \subseteq \mathcal{T}$ is obtained from $\mathcal{T}$ by dropping all CIs of the form $A \sqsubseteq \bot$ and all FAs. Since $\mathcal{A}$ is consistent with $\mathcal{T}$, there is a model $\mathcal{J}$ of $\mathcal{A}$ and $\mathcal{T}$; in particular, $A^{\mathcal{J}} = \emptyset$ for all $A \sqsubseteq \bot \in \mathcal{T}$. By Claim 1, we have $\mathcal{I} \to \mathcal{J}$, and thus $A^{\mathcal{I}} = \emptyset$. For the FAs $\mathsf{func}(s)$, observe that they are obeyed by $\mathcal{A}$ (because of consistency with $\mathcal{T}$) and that they are preserved, by rule 4(a). Thus, $\mathcal{I}$ is a model of $\mathcal{T}$.

Claim 2. For all $i \geq 0$, we have:

(a) For all $a \in \mathsf{ind}(\mathcal{A})$: if $a \in A^{\mathcal{I}_i}$, then there is a derivation tree for $A(a)$ in $\mathcal{A}$ w.r.t. $\mathcal{T}$.

(b) If e was created because of d in Rule 4(b), then we have $\mathcal{T} \models \bigsqcap\{A \mid d \in A^{\mathcal{I}_i}\} \sqsubseteq \exists r.\bigsqcap\{A \mid e \in A^{\mathcal{I}_i}\}$ for all r with $(d,e) \in r^{\mathcal{I}_i}$.

Proof of Claim 2. It is standard to show Part (b) of the Claim. We show Part (a) by induction on i. By construction of $\mathcal{I}_0$, it is true for $i = 0$. Consider $\mathcal{I}_{i+1}$. If $a \in A^{\mathcal{I}_{i+1}}$ because of Rule 1, construct a derivation tree of type (i) from the derivation trees for $A_1(a)$ and $A_2(a)$ which exist due to the induction hypothesis. If $a \in A^{\mathcal{I}_{i+1}}$ because of Rule 2, there is some $d \in B^{\mathcal{I}_i}$ with $(a,d) \in r^{\mathcal{I}_i}$ and $\exists r.B \sqsubseteq A \in \mathcal{T}$. If $d \in \mathsf{ind}(\mathcal{A})$, then there is some $s(a,d) \in \mathcal{A}$ with $\mathcal{T} \models s \sqsubseteq r$, by Rule 3. We can thus construct a derivation of type (ii) from the derivation tree of $B(d)$, which exists due to induction hypothesis. If $d \notin \mathsf{ind}(\mathcal{A})$, then d was created because of a in Rule 4(b). By Part (b) of the Claim, we have $\mathcal{T} \models \bigsqcap\{A' \mid a \in A'^{\mathcal{I}}\} \sqsubseteq \exists r.B$. Hence, $\mathcal{T} \models \bigsqcap\{A' \mid a \in A'^{\mathcal{I}}\} \sqsubseteq A$, and we can construct a derivation tree of type (i) for $A(a)$. If $a \in A^{\mathcal{I}_{i+1}}$ because of Rule 4(a), there is $(d,a) \in r_i^{\mathcal{I}}$ and $d \in B^{\mathcal{I}_i}$, and $B \sqsubseteq \exists r.A, \mathsf{func}(r) \in \mathcal{T}$. If $d \in \mathsf{ind}(\mathcal{A})$, we can construct a derivation tree of type (iii) for $A(a)$ from the derivation tree of $B(d)$ which exists by induction. If $d \notin \mathsf{ind}(\mathcal{A})$, then d was created because of a in Rule 4(b). By Part (b) of the Claim, we have $\mathcal{T} \models \bigsqcap\{A' \mid a \in A'^{\mathcal{I}}\} \sqsubseteq \exists r^-.B$. Hence, $\mathcal{T} \models \bigsqcap\{A' \mid a \in A'^{\mathcal{I}}\} \sqsubseteq A$, and construct a derivation tree of type (i) for $A(a)$ based on this. This finishes the proof of Claim 2 and the Lemma. ❑

In the following Lemma, we characterize consistency of ABoxes with TBoxes.

Lemma 32 *Let $\mathcal{T}$ be a Horn-$\mathcal{ALCHIF}$ TBox and $\mathcal{A}$ an ABox. Then $\mathcal{A}$ is consistent with $\mathcal{T}$ iff the following points are satisfied for all $a \in \mathsf{ind}(\mathcal{A})$:*

1. *the following ABox $\mathcal{A}_a$ is consistent with $\mathcal{T}$:*

   $$\mathcal{A}_a = \{B(a) \mid B(a) \text{ has a derivation tree in } \mathcal{A} \text{ w.r.t. } \mathcal{T}\}$$

2. *for all $\mathsf{func}(s) \in \mathcal{T}$, there is at most one $b \in \mathsf{ind}(\mathcal{A})$ with $s(a,b) \in \mathcal{A}$.*

Proof. The "only if"-direction is immediate, so we concentrate on the "if"-direction. Assume that all $a \in \mathsf{ind}(\mathcal{A})$ satisfy both items above. By the first item, there is a model $\mathcal{I}_a$ of $\mathcal{A}_a$ and $\mathcal{T}$. Since we are considering Horn-$\mathcal{ALCHIF}$, there is also a tree-model $\mathcal{I}_a$ with root $d_a \in \Delta^{\mathcal{I}_a}$ satisfying, for all concept names $B \in \mathsf{N_C}$:

$(*)$ $d_a \in B^{\mathcal{I}_a}$ iff $\mathcal{T}, \mathcal{A}_a \models B(a)$.

We construct an interpretation $\mathcal{I}$ as follows. Start with $\mathcal{I}_0$ by taking

$$\Delta^{\mathcal{I}_0} = \mathsf{ind}(\mathcal{A})$$
$$A^{\mathcal{I}_0} = \{a \mid A(a) \text{ has a derivation tree in } \mathcal{A} \text{ w.r.t. } \mathcal{T}\}$$
$$r^{\mathcal{I}_0} = \{(a,b) \mid s(a,b) \in \mathcal{A}, \mathcal{T} \models s \sqsubseteq r\}$$

Now, obtain $\mathcal{I}$ from $\mathcal{I}_0$ by performing the following operation for every $a \in \mathsf{ind}(\mathcal{A})$ and $b \in \Delta^{\mathcal{I}_a}$ such that $(d_a, b) \in \rho^{\mathcal{I}}$ for some set of roles ρ which contains no role r such that there is a' with $r(a,a') \in \mathcal{A}$. Extend $\mathcal{I}$ by adding the sub-interpretation of $\mathcal{I}_a$ rooted at b as a ρ-successor of a.

Based on $(*)$ and the assumptions, it is straightforward to show that $\mathcal{I}$ is a model of $\mathcal{A}$ and $\mathcal{T}$. ❑

We are now ready to give the automaton $\mathfrak{A}_3$. We take $\mathfrak{A}_3 = (Q_3, \Theta, q_0, \delta_3, \Omega_3)$ where

$$\begin{aligned} Q_3 = \{q_0, q_{0r}\} &\cup \{q_A, \overline{q}_A \mid A \in \Theta_2 \cap \mathsf{N_C}\} \cup \\ &\{q_r, \overline{q}_r, q_r^{\mathcal{A}}, \overline{q}_r^{\mathcal{A}}, q_r^f, q_{\neg r} \mid r \in \Theta_2 \setminus \mathsf{N_C}\} \cup \\ &\{q_{r,B}, \overline{q}_{r,B} \mid r \in \Theta_2 \cap \mathsf{N_R}, B \in \Theta_2 \cap \mathsf{N_C}\} \end{aligned}$$

and Ω_3 assigns zero to all states, except for states of the form q_A, to which it assigns 1. The automaton $\mathfrak{A}_3$ ensures that, for all $n \in \mathsf{ind}(\mathcal{A})$ we have:

(i) $A \in L_2(n)$ iff there is a derivation tree for $A(n)$ in $\mathcal{A}$,

(ii) for all $n \neq \varepsilon$, $r \in L_2(n)$ iff there is some s such that $s(n \cdot -1, n) \in \mathcal{A}$ and $\mathcal{T}_2 \models s \sqsubseteq r$.

Intuitively, these points ensure that the represented interpretation $\mathcal{I}_2$ is the universal model of $\mathcal{T}_2$ and $\mathcal{A}$, in case $\mathcal{A}$ is consistent with $\mathcal{T}_2$. Having (i) and (ii), we can check inconsistency of $\mathcal{A}$ with $\mathcal{T}_2$ based on Lemma 32, that is, we verify the following conditions for all $n \in \mathsf{ind}(\mathcal{A})$:

(iii) the set $L_2(n) \cap \mathsf{N_C}$ is consistent with $\mathcal{T}_2$;

(iv) for each s with $\mathsf{func}(s) \in \mathcal{T}$, there are no $n_1 \neq n_2$ such that both $s(n, n_1) \in \mathcal{A}$ and $s(n, n_2) \in \mathcal{A}$.

For Point (i), we use states q_A for the "if" part, and states $\overline{q}_A$ for the "only if" part; for Point (ii), we use states q_r and $\overline{q}_r$, respectively. Intuitively, a state q_A assigned to some node n is an obligation to verify the existence of a derivation tree for $A(n)$. Conversely, $\overline{q}_A$ is the obligation that there is *no* such derivation tree. Similar obligations hold for q_r and $\overline{q}_r$. For Point (iii), we precompute the set of consistent types and check (iii) while visiting all $n \in \mathsf{ind}(\mathcal{A})$. Point (iv) can be checked directly on $\mathcal{A}$, that is, independent from $\mathcal{T}_2$. The automaton starts with the following transitions, where we assume $\sigma = (L_0, L_1, L_2)$:

- $\delta_3(q_0, \sigma) = \mathsf{true}$ if $L_0 = \emptyset$;
- $\delta_3(q_0, \sigma) = \mathsf{false}$ if $L_0 \neq \emptyset$ and $L_2 \cap \mathsf{N_c}$ inconsistent with $\mathcal{T}_2$, c.f. Point (iii);

– if $L_0 \neq \emptyset$ and $L_2 \cap \mathsf{N_C}$ consistent with $\mathcal{T}_2$, then

$$\delta_3(q_0, \sigma) = \Box q_0 \wedge \Box q_{0r} \wedge \bigwedge_{A \in L_2 \cap \mathsf{N_C}} q_A \wedge \bigwedge_{A \in (\Theta_2 \cap \mathsf{N_C}) \setminus L_2} \overline{q}_A.$$

– $\delta_3(q_{0r}, \sigma) = \mathsf{true}$ if $L_0 = \emptyset$;

– if $L_0 \neq \emptyset$, then

$$\delta_3(q_r, \sigma) = \bigwedge_{\mathsf{func}(r) \in \mathcal{T}_2} q_r^f \wedge \bigwedge_{r \in L_2 \cap \mathsf{N_R}} q_r \wedge \bigwedge_{r \in (\Theta_2 \cap \mathsf{N_R}) \setminus L_2} \overline{q}_r$$

– $\delta_3(q_r^f, \sigma) = \begin{cases} \Box q_{\neg r} & \text{if } r^- \in L_0 \\ \Box_1 q_{\neg r} & \text{if } r^- \notin L_0 \end{cases}$

– $\delta_3(q_{\neg r}, \sigma) = \begin{cases} \mathsf{true} & \text{if } r \notin L_0 \\ \mathsf{false} & \text{otherwise} \end{cases}$

Now, for states q_A, we directly implement the conditions of a derivation tree. Finiteness of the derivation is ensured by the priority of states of the form q_A. The relevant transitions are as follows:

– $\delta_3(q_A, \sigma) = \mathsf{false}$ if $L_0 = \emptyset$;

– $\delta_3(q_A, \sigma) = \mathsf{true}$ if $A \in L_0$;

– if $A \notin L_0$ and $L_0 \neq \emptyset$, then

$$\delta_3(q_A, \sigma) = \bigvee_{\mathcal{T}_2 \models A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A} (q_{A_1} \wedge \cdots \wedge q_{A_n}) \vee \bigvee_{\exists r.B \sqsubseteq A \in \mathcal{T}_2, \mathcal{T}_2 \models s \sqsubseteq r} (q_{s^-}^{\mathcal{A}} \wedge \langle - \rangle q_B) \vee \Diamond q_{s,B} \vee \bigvee_{B \sqsubseteq \exists r.A \in \mathcal{T}_2, \mathsf{func}(r) \in \mathcal{T}_2} (q_s^{\mathcal{A}} \wedge \langle - \rangle q_B) \vee \Diamond q_{s^-,B}$$

– $\delta_3(q_r^{\mathcal{A}}, \sigma) = \begin{cases} \mathsf{true} & \text{if } r \in L_0; \\ \mathsf{false} & \text{otherwise;} \end{cases}$

– $\delta_3(q_{s,B}, \sigma) = q_s^{\mathcal{A}} \wedge q_B$.

The transitions for $\overline{q}_A$ are obtained by taking the "complement" of the ones for q_A. More precisely, we define $\delta_3(\overline{q}, \sigma) = \overline{\delta_3(q, \sigma)}$, where $\overline{\varphi}$ is obtained from φ by exchanging $\wedge$ and $\vee$, $\Diamond$ and $\Box$, $\langle - \rangle$ and $[-]$, and true and false, and replacing every state p with $\overline{p}$; see the following set of transitions.

– $\delta_3(\overline{q}_A, \sigma) = \mathsf{true}$ if $L_1 = \emptyset$;

– $\delta_3(\overline{q}_A, \sigma) = \mathsf{false}$ if $A \in L_1$;

– if $A \notin L_1$ and $L_1 \neq \emptyset$, then

$$\delta_3(\overline{A}, \sigma) = \bigwedge_{\mathcal{T}_2 \models A_1 \sqcap \cdots \sqcap A_n \sqsubseteq A} (\overline{q}_{A_1} \vee \cdots \vee \overline{q}_{A_n}) \wedge \bigwedge_{\exists r.B \sqsubseteq A \in \mathcal{T}_2, \mathcal{T}_2 \models s \sqsubseteq r} (\overline{q}_{s^-}^{\mathcal{A}} \wedge [-]\overline{q}_B) \wedge \Box \overline{q}_{s,B} \wedge \bigwedge_{B \sqsubseteq \exists r.A \in \mathcal{T}_2, \mathsf{func}(r) \in \mathcal{T}_2} (\overline{q}_s^{\mathcal{A}} \vee [-]\overline{q}_B) \wedge \Box \overline{q}_{s^-,B}$$

– $\delta_3(\overline{q}_r^{\mathcal{A}}, \sigma) = \begin{cases} \mathsf{false} & \text{if } r \in L_0; \\ \mathsf{true} & \text{otherwise;} \end{cases}$

– $\delta_3(\overline{q}_{s,B}, \sigma) = \overline{q}_s^{\mathcal{A}} \vee \overline{q}_B$.

Finally, states q_r and $\overline{q}_r$ at some node n represent the obligation to verify that the role atom $r(n \cdot -1, n)$ follows, respectively does not follow, from $\mathcal{T}$ and $\mathcal{A}_2$. This is realized by the following transitions which implement Point (ii) above.

$$\delta_3(q_r, \sigma) = \bigvee_{\mathcal{T}_2 \models s \sqsubseteq r} q_s^{\mathcal{A}} \qquad \delta_3(\overline{q}_r, \sigma) = \bigwedge_{\mathcal{T}_2 \models s \sqsubseteq r} \overline{q}_s^{\mathcal{A}}$$

For the automaton $\mathfrak{A}_4$, we take $\mathfrak{A}_4 = (Q_4, \Theta, q_0, \delta_4, \Omega_4)$ where

$$Q_4 = \{q_0, q_1, q_r\} \cup \{q_t, q_t^3, q_t^{3b} \mid t \in \mathsf{tp}(\mathcal{T}_2)\} \cup \{q_{\rho,t}, q_{\rho,t}^{\downarrow} \mid t \in \mathsf{tp}(\mathcal{T}_2), \rho \text{ set of } \mathsf{sig}(\mathcal{T}_2)\text{-roles}\},$$

and Ω_4 assigns zero to all states, except for states of the form q_t, $t \in \mathsf{tp}(\mathcal{T}_2)$, to which it assigns one. For some $n \in \mathsf{ind}(\mathcal{A})$, denote with $\mathcal{J}_n$ the universal model of the type $\{A(n) \mid A \in L_2(n)\}$ and $\mathcal{T}_2$. The automaton ensures that indeed (1) or (2) from Theorem 11 is not satisfied, by verifying that there is some $n \in \mathsf{ind}(\mathcal{A})$ such that one of the following conditions holds:

1. there is $r \in \Sigma$ and $n' \in \mathsf{ind}(\mathcal{A})$ such that $(n, n') \in r^{\mathcal{I}_2}$, but $(n, n') \notin r^{\mathcal{I}_1}$;
2. $\mathcal{J}_n \not\to_\Sigma \mathcal{I}_1$;
3. there is a Σ-subtree $\mathcal{J}$ of $\mathcal{J}_n$ such that
   (a) $\mathcal{J} \not\to_\Sigma \mathcal{I}_1$, and
   (b) $\mathcal{J} \not\to_\Sigma^{\mathsf{fin}} \mathcal{I}_{\mathcal{T}_1, \mathsf{tp}_{\mathcal{I}_1}(m)}$, for all m with $L_0(m) \neq \emptyset$.

Condition 1 is straightforward (realized in state q_r). For Condition 2, we use states q_t with $t \in \mathsf{tp}(\mathcal{T}_2)$. A state q_t assigned to a node n represents the obligation to verify that there is *no* Σ-homomorphism from the universal model of t and $\mathcal{T}_2$ to $\mathcal{I}_1$ that maps the root to n. This is the case if either the root cannot be mapped to n, or, recursively, there is some ρ-successor t' of t in the universal model such that the universal model of t' and $\mathcal{T}_2$ cannot be mapped to any ρ-neighbor of n. This process is finite because of priority 1 for all q_t with $t \in \mathsf{tp}(\mathcal{T}_2)$. For Condition 3, we precompute the set $R_\Sigma(t)$ of all types of roots of Σ-subtrees which appear in the universal model of t and $\mathcal{T}_2$, and the relation $\to_\Sigma^{\mathsf{fin}}$ according to Theorem 12. Thus, the sets $R_\Sigma(t)$ and the test for finite homomorphisms can be used directly in the transition condition, see states q_1 and q_t^{3b}, respectively. Using states q_t^3, the automaton ensures that a given root t of a Σ-subtree satisfies 3(a) and 3(b).

Let $t|_\Sigma$ and $\rho|_\Sigma$ denote the restriction of t and ρ, respectively, to symbols from Σ. For $\sigma = (L_1, L_2, L_3)$, we take the following transitions:

$$\delta_3(q_0,\sigma) = \begin{cases} \Diamond q_0 \vee q_1 & \text{if } L_0 \neq \emptyset \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta_3(q_1,\sigma) = q_r \vee q_t \vee \textstyle\bigvee_{t' \in \mathsf{R}_\Sigma(t)} q^3_{t'} \quad \text{for } t = L_2 \cap \mathsf{N_C}$$

$$\delta_3(q_t,\sigma) = \begin{cases} \mathsf{true} & \text{if } t|_\Sigma \not\subseteq L_1 \\ \bigvee_{t' \mid t \leadsto^{\mathcal{T}_2}_\rho t'} q_{\rho,t'} & \text{otherwise} \end{cases}$$

$$\delta_3(q_{\rho,t},\sigma) = \begin{cases} \Box q^{\downarrow}_{\rho,t} & \text{if } \rho^-|_\Sigma \not\subseteq L_1 \\ \Box q^{\downarrow}_{\rho,t} \wedge \langle - \rangle q_t & \text{if } \rho^-|_\Sigma \subseteq L_1 \end{cases}$$

$$\delta_3(q^{\downarrow}_{\rho,t},\sigma) = \begin{cases} \mathsf{true} & \text{if } \rho|_\Sigma \not\subseteq L_1 \\ q_t & \text{if } \rho|_\Sigma \subseteq L_1 \end{cases}$$

$$\delta_3(q^3_t,\sigma) = \Box q^3_t \wedge [-] q^3_t \wedge q_t \wedge q^{3b}_t$$

$$\delta_3(q^{3b}_t,\sigma) = \begin{cases} \mathsf{true} & \text{if } L_0 = \emptyset \text{ or } \mathcal{I}_{\mathcal{T}_2,t}|^{\mathsf{con}}_\Sigma \not\to^{\mathsf{fin}}_\Sigma \mathcal{I}_{\mathcal{T}_1, L_1 \cap \mathsf{N_C}} \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta_3(q_r,\sigma) = \begin{cases} \mathsf{true} & \text{if there is } \Sigma\text{-role } s \in L_2 \setminus L_1 \\ \mathsf{false} & \text{otherwise} \end{cases}$$

The automaton $\mathfrak{A}'_4$ is a variant of $\mathfrak{A}_4$ which drops states q^3_t, q^{3b}_t, and all $q_{\rho,t}, q^{\downarrow}_{\rho,t}$ with $|\rho| > 1$ (and all according transitions), and replaces the transitions for q_1 and q_t as follows:

$$\delta_3(q_1,\sigma) = q_r \vee q_t \quad \text{for } t = L_2 \cap \mathsf{N_C}$$

$$\delta_3(q_t,\sigma) = \begin{cases} \mathsf{true} & \text{if } t|_\Sigma \not\subseteq L_1 \\ \displaystyle\bigvee_{t' \mid t \leadsto^{\mathcal{T}_2}_\rho t'} \bigvee_{r \in \rho} q_{\{r\},t'} & \text{otherwise} \end{cases}$$

In this way it verifies that either Condition 1 above is satisfied or the variant 2' of Condition 2 is satisfied, for some $n \in \mathsf{ind}(\mathcal{A})$:

1. there is $r \in \Sigma$ and $n' \in \mathsf{ind}(\mathcal{A})$ such that $(n,n') \in r^{\mathcal{I}_2}$, but $(n,n') \notin r^{\mathcal{I}_1}$;

2.' $\mathcal{J}_n \not\preceq_\Sigma \mathcal{I}_1$.

This finishes the proof of Lemma 30.

C.6 Proof of Theorem 18

Theorem 18 *In any DL between $\mathcal{ELI}$ and $\mathcal{ELHIF}_\bot$, deductive conservative extensions, deductive Σ-entailment, and deductive Σ-inseparability are* CONEXPTIME*-hard.*

The proof is by reduction of a NEXPTIME-complete tiling problem, where the aim is to tile a $2^n \times 2^n$-grid, to the complement of stCQ-conservative extensions. This tiling problem was introduced as a special case of the *origin constrained domino problem* by Grädel [1989], and its NEXPTIME-hardness follows from Grädel's Theorem 3.3. An instance is given by a tuple $P = (\mathfrak{T}, \mathfrak{T}_0, H, V)$, where $\mathfrak{T}$ is a finite set of *tile types*, $\mathfrak{T}_0 \subseteq \mathfrak{T}$ is a set of *distinguished tiles* to be placed on position $(0,0)$ of the grid, and H and V are horizontal and vertical matching conditions. Let $|\mathfrak{T}| = n$. A *solution* to P is a function $\tau : 2^n \times 2^n \to \mathfrak{T}$ such that

- if $\tau(i,j) = t$ and $\tau(i+1,j) = t'$ then $(t,t') \in H$, for all $i < 2^n - 1$, $j < 2^n$,
- if $\tau(i,j) = t$ and $\tau(i,j+1) = t'$ then $(t,t') \in V$, for all $i < 2^n$, $j < 2^n - 1$,
- $\tau(0,0) \in \mathfrak{T}_0$.

We can assume w.l.o.g. that for every tile $t \in \mathfrak{T}$, there is a t' with $(t',t) \in V$.

Let $P = (\mathfrak{T}, \mathfrak{T}_0, H, V)$. We show how to construct $\mathcal{ELI}$ TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ such that $\mathcal{T}_1 \cup \mathcal{T}_2$ is a $(\mathsf{sig}(\mathcal{T}_1), \mathsf{sig}(\mathcal{T}_1))$-stCQ-conservative extension of $\mathcal{T}_1$ iff there is no solution for P. Hence, stCQ-conservative extensions, (Γ,Σ)-stCQ entailment, and (Γ,Σ)-stCQ inseparability are CONEXPTIME-hard in $\mathcal{ELI}$ (and any DL that contains it as a fragment). Since $\mathcal{T}_1$ and $\mathcal{T}_2$ are formulated in $\mathcal{ELI}$, we trivially have $\mathcal{T}_1 \models^{\bot}_{\mathsf{sig}(\mathcal{T}_1)} \mathcal{T}_1 \cup \mathcal{T}_2$. Thus, hardness of deductive conservative extensions follows from Lemma 24, in all DLs between $\mathcal{ELI}$ and $\mathcal{ELHIF}_\bot$ since $\mathcal{T}_1$ and $\mathcal{T}_2$ are formulated in $\mathcal{ELI}$ and Lemma 24 covers deductive conservative extensions in $\mathcal{ELHIF}_\bot$. This also implies hardness of deductive Σ-entailment and of deductive Σ-inseparability in the mentioned DLs.

The intuitions and correctness proofs are based on the following characterization of stCQ-conservative extensions.

Lemma 33 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be $\mathcal{ELI}$ TBoxes such that all role names in $\mathcal{T}_2$ are in $\mathsf{sig}(\mathcal{T}_1)$. Then $\mathcal{T}_1 \cup \mathcal{T}_2$ is a $(\mathsf{sig}(\mathcal{T}_1), \mathsf{sig}(\mathcal{T}_1))$-stCQ-conservative extension of $\mathcal{T}_1$ iff $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}} \to_{\mathsf{sig}(\mathcal{T}_1)} \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ for all tree-shaped $\mathsf{sig}(\mathcal{T}_1)$-ABoxes $\mathcal{A}$.*

Proof. An interpretation is *strongly tree-shaped* if it is tree-shaped and does not contain multi-edges, that is, any $d, d' \in \Delta^{\mathcal{I}}$ are involved in at most one role edge. Since $\mathcal{T}_1$ and $\mathcal{T}_2$ are formulated in $\mathcal{ELI}$ (and thus do not contain role inclusions), for any tree-shaped ABox $\mathcal{A}$ the universal models $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ and $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ are strongly tree-shaped. The assumption on role names in $\mathcal{T}_2$ made in the lemma implies that every element in $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ can be reached from an ABox individual by traveling only along $\mathsf{sig}(\mathcal{T}_1)$-roles. Together, this implies the following:

($*$) there is a $\mathsf{sig}(\mathcal{T}_1)$-simulation from $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$ iff there is a $\mathsf{sig}(\mathcal{T}_1)$-homomorphism from $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

From Theorem 11, we get that $\mathcal{T}_1 \cup \mathcal{T}_2$ is a $(\mathsf{sig}(\mathcal{T}_1), \mathsf{sig}(\mathcal{T}_1))$-stCQ-conservative extension of $\mathcal{T}_1$ iff $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}|^{\mathsf{con}}_{\mathsf{sig}(\mathcal{T}_1)} \preceq_{\mathsf{sig}(\mathcal{T}_1)} \mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$. But $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}|^{\mathsf{con}}_{\mathsf{sig}(\mathcal{T}_1)} = \mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ by the assumption on role names in $\mathcal{T}_2$ made in the lemma and simulations can be replaced with homorphisms by ($*$). ❏

We will build $\mathcal{T}_1$ and $\mathcal{T}_2$ such that the same single role name r is used in $\mathcal{T}_1$ and $\mathcal{T}_2$, thus the assumption in Lemma 33 will be satisfied.

For a clearer presentation, we proceed in two steps. We first define $\mathcal{T}_1$ and $\mathcal{T}_2$ to be an $\mathcal{ELIU}$-TBox, i.e., on both sides of CIs we allow concepts of the following form:

$$L, L' ::= \top \mid A \mid L \sqcap L' \mid L \sqcup L' \mid \exists r.L.$$

The only non-trivial use of disjunction will be on the right-hand side of a CI in $\mathcal{T}_2$. In a second step, we show how to remove disjunction.

We use S to abbreviate the role composition $r; r^-$, writing for example $\exists S.C$ for $\exists r.\exists r^-.C$. Note that S behaves like a reflexive-symmetric role.[4] Ideally, we would like $\mathcal{T}_1$ to be empty (except introducing the required symbols) and $\mathcal{T}_2$ to verify the existence of an S-path in the input ABox whose individuals represent the grid positions along with a tiling, row by row from left to right, starting at the lower left corner and ending at the upper right corner. The positions in the grid are represented in binary by the concept names $X_1, \dots, X_{2n}$ in the ABox where $X_1, \dots, X_n$ indicate the horizontal position and $X_{n+1}, \dots, X_{2n}$ the vertical position. The tiling is represented by concept names T_t, $t \in \mathfrak{T}$. The verification is done by propagating a concept name as a marker bottom up and while doing this, verifying the horizontal matching condition. Under the assumption that an additional labeling with concept names T'_t, $t \in \mathfrak{T}$, is such that

($*$) every point in the path is labeled with T'_t if its descendant at distance exactly 2^n (that is, the grid position immediately below it) is labeled with T_t,

the vertical matching condition is also verified.

For several reasons, this program cannot quite be implemented in the desired way. First, we still have to make sure that ($*$) actually holds. This is done as follows. We install yet another labeling with concept names $\overline{T}_t$, $t \in \mathfrak{T}$, such that a node is labeled with $\overline{T}_t$ if it is not labeled with T_t. Then $\mathcal{T}_2$ checks for a violation of ($*$) in the following way: when the propagation reaches the final individual the verified path, the generation of a finite anonymous S-path is triggered. That path homomorphically embeds into the S-path in the ABox in many different ways since S is reflexiv-symmetric. In fact, for any individual on the ABox path we can find a homomorphism such that the endpoint of the anonymous path maps to that individual because (a) the anonymous path is long enough to reach the first individual on the ABox path and (b) the homomorphism can always 'fold' the reflexive-symmetric role S in a suitable way. At the end of the anonymous path, we then guess (using disjunction) a tile t, make T'_t true, continue building the anonymous path for another 2^n steps (in a way such that it cannot fold), and finally make $\overline{T}_t$ true. Let us pretend for a second that our TBoxes are formulated in $\mathcal{ELI}$. If ($*$) is violated, then the guess can be made such that the anonymous path homomorphically maps into the ABox path. Otherwise, this is not the case. Clearly, the latter can occur only if P has a solution.

The fact that S is reflexive-symmetric allows the mentioned folding of the existential path. However, it poses some complications in the verification of the S-path in the ABox because we must be careful not to confuse successors with predecessors. To this end, every grid position is actually represented by three consecutive individuals labeled with the concept names B_0, B_1, B_2, respectively. All these individuals are labeled identically regarding the X-counter and the concept names T_t. We are going to enforce ($*$) for the B_2-individuals and only these individuals also receive T'_t and $\overline{T}_t$ labels (any other B_i would work as well). Another problem is that $\mathcal{T}_2$ cannot check all possible kinds of defects. In particular, it cannot detect the defect that an element is labeled with more than one tile or that there are multiple successors in the ABox that have an incompatible labeling with the counter concept names. We thus use $\mathcal{T}_1$ to check for such defects. If found, it will generate a defect of the kind that $\mathcal{T}_2$ *can* verify, that is, a violation of ($*$).

[4]We will make sure that all 'relevant domain elements' have an r-successor, which guarantees reflexivity.

We start with assembling $\mathcal{T}_2$, which uses a single role name r via the abbreviation S introduced above and the following concept names.

- jointly with $\mathcal{T}_1$:
  - $X_1, \dots, X_{2n}, \overline{X}_1, \dots, \overline{X}_{2n}$ for the binary representation of the horizontal and vertical grid positions on the ABox path
  - B_0, B_1, B_2 for distinguishing successors and predecessors on the ABox path (these concept names implement a unary counter that counts modulo three)
  - $T_t, \overline{T}_t$, $t \in \mathfrak{T}$, representing tile types present/not present at individuals on the ABox path
  - T'_t, $t \in \mathfrak{T}$, representing tile types present at the descendant at distance exactly $3 \cdot 2^n$ from the given individual on the ABox path
- additionally:
  - L as a verification marker to be propagated along the ABox path
  - ok_i, $1 \leq i \leq 2n$, to indicate that the incrementation of the counter values at an ABox individual is correct regarding the i-th bit (the 1st bit being that of least value)
  - $Y_1, \dots, Y_{2n}, \overline{Y}_1, \dots, \overline{Y}_{2n}$ for counting the length of the anonymous path
  - $Y'_1, \dots, Y'_n, \overline{Y}'_1, \dots, \overline{Y}'_n$ implement another counter on the anonymous path, used to continue extending the path by exactly $3 \cdot 2^n$ positions to reach the grid position immediately below
  - B'_0, B'_1, B'_2 for distinguishing successors and predecessors on the anonymous path
  - M_t, $t \in \mathfrak{T}$, for memorizing a tile type on the anonymous path.

$\mathcal{T}_2$ consists of the following CIs.

1. The initial grid position starts the propagation:
$$\overline{X}_1 \sqcap \dots \sqcap \overline{X}_{2n} \sqcap B_0 \sqcap \bigsqcup_{t \in \mathfrak{T}_0} T_t \sqsubseteq L$$
2. The verification proceeds upwards. We first verify that the counter is incremented properly when moving upwards along the S-path in the ABox:
$$B_0 \sqcap X_i \sqcap \exists S.(B_2 \sqcap X_i) \sqcap \bigsqcup_{1 \leq j < i} \exists S.(B_2 \sqcap X_j) \sqsubseteq \mathsf{ok}_i$$
$$B_0 \sqcap \overline{X}_i \sqcap \exists S.(B_2 \sqcap \overline{X}_i) \sqcap \bigsqcup_{1 \leq j < i} \exists S.(B_2 \sqcap X_j) \sqsubseteq \mathsf{ok}_i$$
$$B_0 \sqcap X_i \sqcap \exists S.(B_2 \sqcap \overline{X}_i) \sqcap \textstyle\bigsqcap_{1 \leq j < i} \exists S.(B_2 \sqcap \overline{X}_j) \sqsubseteq \mathsf{ok}_i$$
$$B_0 \sqcap \overline{X}_i \sqcap \exists S.(B_2 \sqcap X_i) \sqcap \textstyle\bigsqcap_{1 \leq j < i} \exists S.(B_2 \sqcap \overline{X}_j) \sqsubseteq \mathsf{ok}_i$$
$$B_{j+1} \sqcap X_i \sqcap \exists S.(B_j \sqcap X_i) \sqsubseteq \mathsf{ok}_i$$
$$B_{j+1} \sqcap \overline{X}_i \sqcap \exists S.(B_j \sqcap \overline{X}_i) \sqsubseteq \mathsf{ok}_i$$
where i ranges over $1..2n$ and j over $\{0, 1\}$. These inclusions only work under the assumption that no individual

has two S-neighbors that are labeled with the same B_i but are labeled differently regarding X_j and $\overline{X}_j$ for some j. We shall prevent this situation later using $\mathcal{T}_1$.

3. We next make a verification step inside a row of the grid:

$$B_0 \sqcap \mathsf{ok}_1 \sqcap \cdots \sqcap \mathsf{ok}_n \sqcap T_{t_2} \sqcap \overline{X}_i \sqcap \exists S.(B_2 \sqcap L \sqcap T_{t_1}) \sqsubseteq L$$
$$B_1 \sqcap \mathsf{ok}_1 \sqcap \cdots \sqcap \mathsf{ok}_n \sqcap T_{t_2} \sqcap \overline{X}_i \sqcap \exists S.(B_0 \sqcap L \sqcap T_{t_1}) \sqsubseteq L$$
$$B_2 \sqcap \mathsf{ok}_1 \sqcap \cdots \sqcap \mathsf{ok}_n \sqcap T_{t_2} \sqcap \textstyle\bigsqcap_{t \in \mathfrak{T} \setminus \{t_2\}} \overline{T}_t \sqcap T'_{t_3} \sqcap \overline{X}_i \sqcap \exists S.(B_1 \sqcap L \sqcap T_{t_2}) \sqsubseteq L$$

where t_1, t_2, t_3 range over $\mathfrak{T}$ such that $(t_1, t_2) \in H$ and $(t_3, t_2) \in V$ and i ranges over $1..n$. The use of $\overline{X}_i$ on the left-hand sides ensures that we move inside a row. In the first line, we make a move between horizontally neighboring grid positions, verifying the horizontal matching condition. In the other lines, we move along the three points representing the same grid position, ensuring that they are all labeled by the same tile.[5]

4. We also have to consider the case where we jump from one grid row to the next, ignoring the tiling condition:

$$B_0 \sqcap \mathsf{ok}_1 \sqcap \cdots \sqcap \mathsf{ok}_n \sqcap T_t \sqcap \overline{X}_i \sqcap \exists S.(B_2 \sqcap L) \sqsubseteq L$$

where t ranges over $\mathfrak{T}$ and i ranges over $n+1..2n$. The use of $\overline{X}_i$ on the left-hand side ensures that we are not yet in the topmost row.

5. When the final individual of the ABox path is reached (maximum counter value and B_2-label), we make an extra step in the ABox to a B_0-labeled individual and then generate the first object of an existential path:

$$B_0 \sqcap \exists S.(B_2 \sqcap X_1 \sqcap \cdots \sqcap X_{2n} \sqcap L) \sqsubseteq \exists S.C$$

where

$$C = Y_1 \sqcap \cdots \sqcap Y_{2n} \sqcap B'_2$$

The purpose of the extra step will be explained later on.

6. Here and in the following, we use the abbreviation

$$\exists S^{(3)}.(X_1, X_2, X_3) := \exists S.(X_1 \sqcap \exists S.(X_2 \sqcap \exists S.X_3))$$

for concept names X_1, X_2, X_3.

We continue building the path, decrementing the Y-counter:

$$Y_i \sqcap B'_2 \sqsubseteq \exists S^{(3)}.(B'_1, B'_0, B'_2)$$
$$B'_j \sqcap \exists S.(B'_{j+1} \sqcap Y_i) \sqsubseteq Y_i$$
$$B'_j \sqcap \exists S.(B'_{j+1} \sqcap \overline{Y}_i) \sqsubseteq \overline{Y}_i$$
$$B'_2 \sqcap \exists S.(B'_0 \sqcap \textstyle\bigsqcap_{1 \leq j \leq i} \overline{Y}_i) \sqsubseteq Y_i$$
$$B'_2 \sqcap \exists S.(B'_0 \sqcap Y_i \sqcap \textstyle\bigsqcap_{1 \leq j < i} \overline{Y}_i) \sqsubseteq \overline{Y}_i$$
$$B'_2 \sqcap \exists S.(B'_0 \sqcap Y_i \sqcap \bigsqcup_{1 \leq j < i} Y_i) \sqsubseteq Y_i$$
$$B'_2 \sqcap \exists S.(B'_0 \sqcap \overline{Y}_i \sqcap \bigsqcup_{1 \leq j < i} Y_i) \sqsubseteq \overline{Y}_i$$

where i ranges over $1..2n$ and j over $\{0, 1\}$. It is essential to use different B'_i and counter concepts $Y_i, \overline{Y}_i$ than in the ABox; otherwise the anonymous path could not homomorphically embed into the ABox path in a folded way. It would actually suffice to build a path of length $3 \cdot (2^{2n-1})$ because no violation of $(*)$ can start in the bottommost row. However, overcounting does not compromise correctness.

7. At the end of the anonymous path, we implement a violation of $(*)$ as described above: we guess a tile t involved in the violation, make sure that T'_t holds at the current point, start a new counter, travel exactly $3 \cdot 2^n$ steps (without any folding), and verify that $\overline{T}_t$ holds where we arrive:

$$\overline{Y}_1 \sqcap \cdots \sqcap \overline{Y}_{2n} \sqcap B'_2 \sqsubseteq Y'_1 \sqcap \cdots \sqcap Y'_n \sqcap \bigsqcup_{t \in \mathfrak{T}} (T'_t \sqcap M_t)$$
$$Y'_i \sqcap B_2 \sqsubseteq \exists S^{(3)}.(B_0, B_1, B_2)$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap Y'_i) \sqsubseteq Y'_i$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap \overline{Y}'_i) \sqsubseteq \overline{Y}'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap \textstyle\bigsqcap_{1 \leq j \leq i} \overline{Y}'_i) \sqsubseteq Y'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Y'_i \sqcap \textstyle\bigsqcap_{1 \leq j < i} \overline{Y}'_i) \sqsubseteq \overline{Y}'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Y'_i \sqcap \bigsqcup_{1 \leq j < i} Y'_i) \sqsubseteq Y'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap \overline{Y}'_i \sqcap \bigsqcup_{1 \leq j < i} Y'_i) \sqsubseteq \overline{Y}'_i$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap M_t) \sqsubseteq M_t$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Y'_i \sqcap M_t) \sqsubseteq M_t$$
$$\overline{Y}'_1 \sqcap \cdots \sqcap \overline{Y}'_n \sqcap M_t \sqsubseteq \overline{T}_t$$

where i ranges over $1..n$ and t over $\mathfrak{T}$. Here we use the same B_i as in the ABox to avoid folding.

This finishes the definition of $\mathcal{T}_2$. We now define $\mathcal{T}_1$, which uses the following additional concept names.

- D for indicating the occurrence of a defect
- $Z_1, \ldots, Z_n$ for an additional counter.

$\mathcal{T}_1$ consists of the following CIs.

1. Tiles are mutually exclusive: for all distinct $t, t' \in \mathfrak{T}$:

$$T_t \sqcap T_{t'} \sqsubseteq D$$

where D starts a path that implements a violation of $(*)$, to be implemented below;

2. The problematic situation described at the end of Item 2 in the definition of $\mathcal{T}_2$ cannot occur:

$$\exists S.(B_k \sqcap X_i) \sqcap \exists S.(B_k \sqcap \overline{X}_i) \sqsubseteq D$$

where k ranges over $\{0, 1, 2\}$ and i over $1..2n$;

[5] Note that we expext to see T'_t labels also in row 0; this is why we assume that for every $t \in \mathfrak{T}$, there is a $(t', t) \in V$; we could avoid the assumption at the cost of dealing with row 0 as a special case.

3. We next implement the path triggered by D:

$$B_2 \sqcap (D \sqcup \exists S.D \sqcup \exists S.\exists S.D) \sqsubseteq Z_1 \sqcap \cdots \sqcap Z_n \sqcap T'_t$$
$$Z_i \sqcap B_2 \sqsubseteq \exists S^{(3)}.(B_0, B_1, B_2)$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap Z_i) \sqsubseteq Z_i$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap \overline{Z}_i) \sqsubseteq \overline{Z}_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap \textstyle\bigsqcap_{1 \leq j \leq i} \overline{Z}_i) \sqsubseteq Z_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Z_i \sqcap \textstyle\bigsqcap_{1 \leq j < i} \overline{Z}_i) \sqsubseteq \overline{Z}_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Z_i \sqcap \bigsqcup_{1 \leq j < i} Z_i) \sqsubseteq Z_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap \overline{Z}_i \sqcap \bigsqcup_{1 \leq j < i} Z_i) \sqsubseteq \overline{Z}_i$$
$$\overline{Z}_1 \sqcap \cdots \sqcap \overline{Z}_n \sqsubseteq \overline{T}_t$$

where i ranges over $1..n$ and $t \in \mathfrak{T}$ is fixed.

Before we eliminate the disjunction used in $\mathcal{T}_2$, let us mention the central property of $\mathcal{T}_1$ and $\mathcal{T}_2$ that can be used to show correctness of the reduction. Since $\mathcal{T}_2$ uses disjunction, a universal model for $\mathcal{T}_2$ and an ABox $\mathcal{A}$ is not guaranteed to exist. Instead, there is a set of models for $\mathcal{T}_2$ and $\mathcal{A}$ that is universal in the sense that for every model $\mathcal{I}$ of $\mathcal{T}_2$ and $\mathcal{A}$, there is a model in the set that admits a homomorphism into $\mathcal{I}$. We refrain from giving a formal definition. Now, the central property of $\mathcal{T}_1$ and $\mathcal{T}_2$ is as follows: P has a solution iff there is a tree-shaped $\mathsf{sig}(\mathcal{T}_1)$-ABox $\mathcal{A}$ consistent with $\mathcal{T}_1 \cup \mathcal{T}_2$ such that, for every $\mathcal{I}$ in the universal set of models for $\mathcal{T}_1 \cup \mathcal{T}_2$ and $\mathcal{A}$, there is no $\mathsf{sig}(\mathcal{T}_1)$-homomorphism from $\mathcal{I}$ to $\mathcal{I}_{\mathcal{T}_1,\mathcal{A}}$.

We now show how to get rid of disjunction. The central property of $\mathcal{T}_1$ and $\mathcal{T}_2$ will essentially be preserved, with a single universal model playing the role of the universal set of models. The disjunctions on the left-hand sides of CIs only serve as abbreviations and can easily be removed with only a polynomial blowup of TBox sizes. What remains is the disjunction in Item 7 of the definition of $\mathcal{T}_2$. To get rid of it, we need to modify both $\mathcal{T}_1$ and $\mathcal{T}_2$:

- In the first line of Point 7 of the definition of $\mathcal{T}_2$, the disjunction is replaced with a conjunction, generating $|\mathfrak{T}|$ many defective chains at once:

  $$\overline{Y}_1 \sqcap \cdots \sqcap \overline{Y}_{2n} \sqsubseteq \bigsqcap_{t \in \mathfrak{T}} \exists S.(Y'_1 \sqcap \cdots \sqcap Y'_n \sqcap B_2 \sqcap T'_t \sqcap M_t)$$

  The resulting universal model is illustrated in Figure 1 where we assume $\mathfrak{T} = \{t_1, t_2, t_3\}$, showing the final individual on the ABox path, the extra step from Item 5 of the definition of $\mathcal{T}_2$, the anonymous path, and the branching gadget attached to the end of it.

- We have now generated too many paths and thus the desired homomorphism may not exist even if $(*)$ is violated in the ABox. We compensate by enforcing in $\mathcal{T}_1$ that when a B_2-individual in the ABox path is labeled with T'_t, then it's B_0-predecessor on that path roots $|\mathfrak{T}| - 1$ many additional paths, realizing every possible violation of $(*)$ except the one induced by $t \in \mathfrak{T}$. This is illustrated in Figure 2 where we again assume $\mathfrak{T} = \{t_1, t_2, t_3\}$, showing a B_2-individual labeled with T'_{t_1} and the extra successors of it's B_0-predecessor generated by $\mathcal{T}_1$.

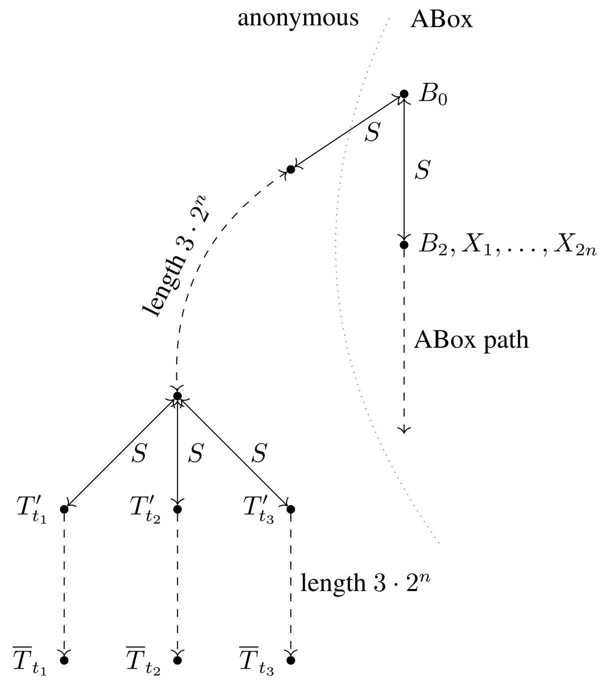

Figure 1: Part of the universal model of $\mathcal{T}_2$.

Note that this explains the extra step in Item 5 of the definition of $\mathcal{T}_2$ since also the final B_2-element of the ABox path must have a B_0-predecessor.

We add the following to $\mathcal{T}_1$, using fresh concept names M'_t, $t \in \mathfrak{T}$:

$$B_0 \sqcap \exists S.(B_2 \sqcap T'_t) \sqsubseteq \textstyle\bigsqcap_{t' \in \mathfrak{T} \setminus \{t\}} \exists S.(Y'_1 \sqcap \cdots \sqcap Y'_n \sqcap B_2 \sqcap M'_{t'})$$
$$Y'_i \sqcap B_2 \sqsubseteq \exists S^{(3)}.(B_0, B_1, B_2)$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap Y'_i) \sqsubseteq Y'_i$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap \overline{Y}'_i) \sqsubseteq \overline{Y}'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap \textstyle\bigsqcap_{1 \leq j \leq i} \overline{Y}'_i) \sqsubseteq Y'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Y'_i \sqcap \textstyle\bigsqcap_{1 \leq j < i} \overline{Y}'_i) \sqsubseteq \overline{Y}'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Y'_i \sqcap \bigsqcup_{1 \leq j < i} Y'_i) \sqsubseteq Y'_i$$
$$B_0 \sqcap \exists S.(B_2 \sqcap \overline{Y}'_i \sqcap \bigsqcup_{1 \leq j < i} Y'_i) \sqsubseteq \overline{Y}'_i$$
$$B_{j+1} \sqcap \exists S.(B_j \sqcap M'_t) \sqsubseteq Y'_i \sqcap M'_t$$
$$B_0 \sqcap \exists S.(B_2 \sqcap Y'_i \sqcap M'_t) \sqsubseteq M'_t$$
$$\overline{Y}'_1 \sqcap \cdots \sqcap \overline{Y}'_n \sqcap M'_t \sqsubseteq \overline{T}_t$$

where i ranges over $1..n$ and t over $\mathfrak{T}$.

Lemma 34 *$\mathcal{T}_1 \cup \mathcal{T}_2$ is a $(\mathsf{sig}(\mathcal{T}_1), \mathsf{sig}(\mathcal{T}_1))$-stCQ-conservative extension of $\mathcal{T}_1$ iff there is no solution for P.*

Proof. By Lemma 33, it suffices to show the following.

Claim. P has a solution iff there is a tree-shaped $\mathsf{sig}(\mathcal{T}_1)$-ABox $\mathcal{A}$ such that there is no $\mathsf{sig}(\mathcal{T}_1)$-homomorphism from $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1, \mathcal{A}}$.

We now sketch a proof of the claim.

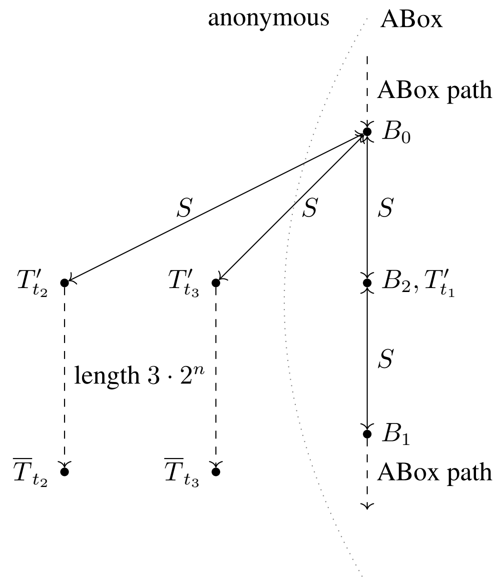

Figure 2: Extra successors to eliminate disjunction

"$\Rightarrow$". Assume that P has a solution. Let $\mathcal{A}$ be the ABox that contains a single S-path of length $3 \cdot 2^n$ which correctly encodes the solution to P via the concept names $X_i, \overline{X}_i, B_i, T_t, \overline{T}_t, T'_t$. Since $\mathcal{A}$ correctly encodes the tiling, there is no $\mathsf{sig}(\mathcal{T}_1)$-homomorphism from $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1, \mathcal{A}}$; in particular, the anonymous path described by $\mathcal{T}_2$ that ends in $|\mathfrak{T}|$ many violations of $(*)$, each represented via a path, cannot be mapped to $\mathcal{I}_{\mathcal{T}_1, \mathcal{A}}$.

"$\Leftarrow$". Assume P has no solution and let $\mathcal{A}$ be a tree-shaped $\mathsf{sig}(\mathcal{T}_1)$-ABox. If $\mathcal{A}$ contains no path of length $3 \cdot 2^n$ that is labeled in the desired way with the concept names $X_i, \overline{X}_i, B_i, T_t, \overline{T}_t, T'_t$ (and which does not necessarily satisfy $(*)$), then the generation of the anonymous path in $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ is not triggered and the identity is a $\mathsf{sig}(\mathcal{T}_1)$-homomorphism from $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1, \mathcal{A}}$. If there is such a path, then it violates $(*)$ since P has no solution. Consequently, the anonymous path in $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ homomorphically maps to $\mathcal{I}_{\mathcal{T}_1, \mathcal{A}}$, which is sufficient to show that there is a homomorphism from $\mathcal{I}_{\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}}$ to $\mathcal{I}_{\mathcal{T}_1, \mathcal{A}}$. ❏