Non-Uniform Data Complexity of Query Answering in Description Logics

Carsten Lutz
Fachbereich Informatik
Universität Bremen, Germany

Frank Wolter
Department of Computer Science
University of Liverpool, UK

Abstract

In ontology-based data access (OBDA), ontologies are used as an interface for querying instance data. Since in typical applications the size of the data is much larger than the size of the ontology and query, data complexity is the most important complexity measure. In this paper, we propose a new method for investigating data complexity in OBDA: instead of classifying whole logics according to their complexity, we aim at classifying *each individual ontology* within a given master language. Our results include a P/coNP-dichotomy theorem for ontologies of depth one in the description logic $\mathcal{ALCFI}$, the equivalence of a P/coNP-dichotomy theorem for $\mathcal{ALC}$/$\mathcal{ALCI}$-ontologies of unrestricted depth to the famous dichotomy conjecture for CSPs by Feder and Vardi, and a non-P/coNP-dichotomy theorem for $\mathcal{ALCF}$-ontologies.

1 Introduction

In recent years, the use of ontologies to access instance data has become increasingly popular (Poggi et al. 2008; Dolby et al. 2008). The general idea is that an ontology provides an enriched vocabulary or conceptual model for the application domain, thus serving as an interface for querying instance data and allowing to derive additional facts. In this emerging area, called ontology-based data access, it is a central research goal to identify ontology languages for which query answering scales to large amounts of instance data. Since the size of the data is typically very large compared to the size of the ontology and the size of the query, the central measure for such scalability is provided by *data complexity*—the complexity of query answering where only the data is considered to be an input, but both the query and the ontology are fixed.

In description logic (DL), ontologies take the form of a TBox, instance data is stored in an ABox, and the most important class of queries are conjunctive queries (CQs). A fundamental observation regarding this setup is that, for expressive DLs such as $\mathcal{ALC}$ and $\mathcal{SHIQ}$, the complexity of query answering is coNP-complete and thus intractable (when speaking of complexity, we *always* mean data complexity; references are given at the end of this section). The most popular strategy to avoid this problem is to replace $\mathcal{ALC}$ and $\mathcal{SHIQ}$ with less expressive DLs that *are Horn* in the sense that they can be embedded into the Horn fragment of first-order (FO) logic. Horn DLs in this sense include logics from the $\mathcal{EL}$ and DL-Lite families as well as Horn-$\mathcal{SHIQ}$, a large fragment of $\mathcal{SHIQ}$ for which CQ-answering is still in PTIME.

It thus seems that the data complexity of query answering in a DL context is well-understood. However, all results discussed above are on the *level of logics*, i.e., each result concerns a class of TBoxes that is defined in a syntactic way in terms of expressibility in a certain logic, but no attempt is made to identify more structure *inside* these classes. The aim of this paper is to advocate a fresh look on the subject, by taking a novel approach. Specifically, we initiate a *non-uniform* study of the complexity of query answering by considering data complexity on the *level of individual TBoxes*. We say that CQ-answering w.r.t. a TBox $\mathcal{T}$ is in PTIME if for every CQ q, there is a PTIME algorithm that computes, given an ABox $\mathcal{A}$, the answers to q in $\mathcal{A}$ w.r.t. $\mathcal{T}$; CQ-answering w.r.t. $\mathcal{T}$ is coNP-hard if there exists a Boolean CQ q such that it is coNP-hard to answer q in ABoxes $\mathcal{A}$ w.r.t. $\mathcal{T}$. Other complexities can be defined similarly. The ultimate goal of our approach is as follows:

For a fixed master DL $\mathcal{L}$, classify all TBoxes $\mathcal{T}$ in $\mathcal{L}$ according to the complexity of CQ-answering w.r.t. $\mathcal{T}$.

In this paper, we consider as master DLs the basic expressive DL $\mathcal{ALC}$, its extensions $\mathcal{ALCI}$ with inverse roles and $\mathcal{ALCF}$ with functional roles, and their union $\mathcal{ALCFI}$. It turns out that, even for $\mathcal{ALC}$, fully achieving the above goal is far beyond the scope of a single research paper. In fact, we show that a full classification of the complexity of $\mathcal{ALC}$-TBoxes is essentially equivalent to a full classification of the complexity of non-uniform constraint satisfaction problems with finite templates (CSPs). The latter is a major research programme ongoing for many years that combines complexity theory, graph theory, logic, and algebra; see below for references and additional details.

In the current paper, we mainly concentrate on understanding the boundary between PTIME and coNP-hardness of CQ-answering w.r.t. DL TBoxes, mostly neglecting other relevant classes such as AC^0, LOGSPACE, and NLOGSPACE. Our main results are as follows.

1. There is a PTIME/coNP-dichotomy for CQ-answering w.r.t. $\mathcal{ALCFI}$-TBoxes of depth one, i.e., TBoxes in which existential/universal restrictions are not nested.

Note that this is a relevant case since most TBoxes from practical applications have depth one. In particular, all TBoxes formulated in DL-Lite and its extensions proposed in (Calvanese et al. 2006; Artale et al. 2009) have depth one, and the same is true for more than 85 percent of all TBoxes in the TONES ontology repository (http://owl.cs.manchester.ac.uk/repository/).

2. There is a PTIME/coNP-dichotomy for CQ-answering w.r.t. $\mathcal{ALC}$-TBoxes if and only if Feder and Vardi's dichotomy conjecture for CSPs is true; the same holds for $\mathcal{ALCI}$-TBoxes.

The proof of this result establishes the close link between CQ-answering in $\mathcal{ALC}$ and CSP that was mentioned above. While dichotomy questions are mainly of theoretical interest, linking these two worlds is potentially very relevant also for applied DL research.

3. There is no PTIME/CONP-dichotomy for CQ-answering w.r.t. $\mathcal{ALCF}$-TBoxes (unless PTIME $=$ NP).

This is proved by showing that, for every problem in coNP, there is an $\mathcal{ALCF}$-TBox for which CQ-answering has the same complexity (up to polytime reductions); it then remains to apply Ladner's Theorem, which guarantees the existence of NP-intermediate problems. Consequently, we cannot expect an exhaustive classification of the complexity of CQ-answering w.r.t. $\mathcal{ALCF}$-TBoxes.

To prove these results, we introduce two new notions that are of independent interest and general utility. The first one is *materializability* of a TBox $\mathcal{T}$, which means that answering a CQ over an ABox $\mathcal{A}$ w.r.t. $\mathcal{T}$ can be reduced to query evaluation in a single model of $\mathcal{A}$ and $\mathcal{T}$. Note that such models play a crucial role in the context of Horn DLs, where they are often called least models or canonical models. In contrast to the Horn DL case, however, we only require the *existence* of such a model without making any assumptions about its form or construction.

4. If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is not materializable, then CQ-answering w.r.t. $\mathcal{T}$ is CONP-hard.

Perhaps in contrary to the intuitions that arise from the experience with Horn-DLs, materializability of a TBox $\mathcal{T}$ is *not* a sufficient condition for CQ-answering w.r.t. $\mathcal{T}$ to be in PTIME (unless PTIME = NP). This leads us to study the notion of *unraveling tolerance* of a TBox $\mathcal{T}$, meaning that answers to tree-shaped CQs over an ABox $\mathcal{A}$ w.r.t. $\mathcal{T}$ are preserved under unraveling the ABox $\mathcal{A}$. In CSP, unraveling tolerance corresponds to the existence of tree obstructions, a notion that characterizes the well-known arc consistency condition (Krokhin 2010; Dechter 2003). It can be shown that every TBox formulated in Horn-$\mathcal{ALCFI}$ (the intersection of $\mathcal{ALCFI}$ and Horn-$\mathcal{SHIQ}$) is unraveling tolerant and that there are unraveling tolerant TBoxes which are not equivalent to any Horn-$\mathcal{ALCFI}$-TBox. Thus, the following result yields a rather general (and uniform!) PTIME upper bound for CQ-answering.

5. If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is unraveling tolerant, then CQ-answering w.r.t. $\mathcal{T}$ is in PTIME.

Although the above result is rather general, unraveling tolerance of a TBox $\mathcal{T}$ is *not* a necessary condition for CQ-answering w.r.t. $\mathcal{T}$ to be in PTIME (unless PTIME = NP). However, for $\mathcal{ALCFI}$-TBoxes $\mathcal{T}$ *of depth one*, being materializable and being unraveling tolerant turns out to be equivalent. We thus obtain that CQ-answering w.r.t. $\mathcal{T}$ is in PTIME iff $\mathcal{T}$ is materializable iff $\mathcal{T}$ is unraveling tolerant while, otherwise, CQ-answering w.r.t. $\mathcal{T}$ is CONP-hard. This establishes the first main goal above.

Our framework also allows to formally capture some intuitions and beliefs commonly held in the context of CQ-answering in DLs. For example, we show that for every $\mathcal{ALCFI}$-TBox $\mathcal{T}$, CQ-answering is in PTIME iff answering positive existential queries is in PTIME iff answering $\mathcal{ELI}$-instance queries (tree-shaped CQs) is in PTIME. This implies that all results mentioned above apply not only to CQ answering, but also to answering queries in any of these other languages. In fact, the use of multiple query languages and in particular of $\mathcal{ELI}$-instance queries does not only yield additional results, but is also at the heart of our proof strategies, which would not work for CQs alone.

Another interesting observation in this spirit is that an $\mathcal{ALCFI}$-TBox is materializable iff it is convex, a condition that is also called the disjunction property and plays a central role in attaining PTIME complexity for standard reasoning in Horn DLs such as $\mathcal{EL}$, DL-Lite, and Horn-$\mathcal{SHIQ}$; see for example (Baader, Brandt, and Lutz 2005; Krisnadhi and Lutz 2007) for more details..

Most proofs are deferred to the long version, available at http://www.csc.liv.ac.uk/~frank/publ/publ.html.

Related Work

An early reference on data complexity in DLs is (Schaerf 1993), showing CONP-hardness of instance queries in the moderately expressive DL $\mathcal{ALE}$. A CONP upper bound for instance queries in the much more expressive $\mathcal{SHIQ}$ was obtained in (Hustadt, Motik, and Sattler 2007) and generalized to CQs in (Glimm et al. 2008). Horn-$\mathcal{SHIQ}$ was first defined in (Hustadt, Motik, and Sattler 2007), where also a PTIME upper bound for instance queries is established; the generalization to CQs can be found in (Eiter et al. 2008). See also (Krisnadhi and Lutz 2007; Calvanese et al. 2006) and references therein for data complexity in DLs and (Barany, Gottlob, and Otto 2010; Baget et al. 2011) for related work beyond standard DLs.

To the best of our knowledge, the current paper presents the first study of data complexity in OBDA at the level of individual TBoxes and the first formal link between OBDA and CSP. There is, however, a vague technical similarity to the link between view-based query processing for regular path queries (RPQs) and CSP found in (Calvanese et al. 2000; 2003b; 2003a). In this case, the recognition problem for perfect rewritings for RPQs can be polynomially reduced to non-uniform CSP and vice versa.

The work on CSP dichotomies started with Schaefer's PTIME/NP-dichotomy theorem, stating that every binary CSP is in PTIME or NP-hard (Schaefer 1978). Here, a binary CSP is defined by a relational structure $\mathcal{B}$ whose domain consists of two elements and the problem is to decide for a given relational structure $\mathcal{C}$ over the same relation symbols, whether there is a homomorphism from $\mathcal{C}$ to $\mathcal{B}$. To appreciate Schaefer's result, recall that Ladners theorem guarantees, in general, the existence of problems that are NP-intermediate and thus neither in PTIME nor NP-hard, unless PTIME = NP (Ladner 1975). Schaefer's theorem was followed by a dichotomy result for CSPs with graph templates (Hell and Nesetril 1990) and the seminal Feder-Vardi PTIME/NP-dichotomy conjecture for all CSPs (Feder and Vardi 1993), confirmed for ternary CSPs in (Bulatov 2002). Interesting results have also been obtained for other complexity classes such as AC^0 (Allender et al. 2005; Larose, Loten, and Tardif 2007). The state of the art is summarized, for example, in (Bulatov, Jeavons, and Krokhin 2005; Kun and Szegedy 2009; Bulatov 2011).

2 Preliminaries

We start with introducing the DL $\mathcal{ALC}$ and its extensions $\mathcal{ALCI}$ and $\mathcal{ALCFI}$. As usual, we use $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ to denote countably infinite sets of *concept names*, *role names*, and *individual names*, respectively. *$\mathcal{ALC}$-concepts* are constructed according to the rule

$$C, D \;:=\; \top \mid \bot \mid A \mid C \sqcap D \mid C \sqcup D \mid \neg C \mid \exists r.C \mid \forall r.C$$

where A ranges over $\mathsf{N_C}$ and r ranges over $\mathsf{N_R}$. *$\mathcal{ALCI}$-concepts* admit, in addition, *inverse roles* from the set $\mathsf{N_R^-} = \{r^- \mid r \in \mathsf{N_R}\}$. To avoid heavy notation, we set $r^- = s$ if $r = s^-$ for a role name s. An *$\mathcal{ALC}$-TBox* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, where C, D are $\mathcal{ALC}$-concepts, and likewise for $\mathcal{ALCI}$-TBoxes. An *$\mathcal{ALCFI}$-TBox* is an $\mathcal{ALCI}$-TBox that additionally admits functionality assertions $\mathsf{func}(r)$, where $r \in \mathsf{N_R} \cup \mathsf{N_R^-}$.

An *ABox* $\mathcal{A}$ is a finite set of assertions of the form $A(a)$ and $r(a, b)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. In some cases, we drop the finiteness condition on ABoxes and then explicitly speak about *infinite ABoxes*. We use $\mathsf{Ind}(\mathcal{A})$ to denote the set of individual names used in the ABox $\mathcal{A}$ and sometimes write $r^-(a, b) \in \mathcal{A}$ instead of $r(b, a) \in \mathcal{A}$.

The semantics of DLs is given by *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set and $\cdot^\mathcal{I}$ maps each concept name $A \in \mathsf{N_C}$ to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$, each role name $r \in \mathsf{N_R}$ to a binary relation $r^\mathcal{I}$ on $\Delta^\mathcal{I}$, and each individual name a to an element $a^\mathcal{I} \in \Delta^\mathcal{I}$. We make the *unique name assumption*, i.e., $a^\mathcal{I} \neq b^\mathcal{I}$ whenever $a \neq b$. The extension $C^\mathcal{I} \subseteq \Delta^\mathcal{I}$ of a concept C under the interpretation $\mathcal{I}$ is defined as usual, see (Baader et al. 2003). For the purposes of this paper, it is often convenient to work with interpretations that interpret only *some* individual names, but not all. In this case, we use $\mathsf{Ind}(\mathcal{I})$ to denote the set of individual names interpreted by $\mathcal{I}$.

We say that $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, an assertion $A(a)$ if $a \in \mathsf{Ind}(\mathcal{I})$ and $a^\mathcal{I} \in C^\mathcal{I}$, an assertion $r(a, b)$ if $a, b \in \mathsf{Ind}(\mathcal{I})$ and $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$, and a functionality assertion $\mathsf{func}(r)$ if $r^\mathcal{I}$ is a function. Finally, $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ (ABox $\mathcal{A}$) if it satisfies all inclusions in $\mathcal{T}$ (all assertions in $\mathcal{A}$). The class of all models of $\mathcal{T}$ and $\mathcal{A}$ is denoted by $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$. We call an ABox $\mathcal{A}$ *consistent w.r.t. a TBox* $\mathcal{T}$ if $\mathsf{Mod}(\mathcal{T}, \mathcal{A}) \neq \emptyset$.

Throughout this paper, we consider several query languages which can all be seen as fragments of *positive existential queries (PEQs)*. A PEQ $q(\vec{x})$ is a first-order formula with free variables $\vec{x}$ constructed from atoms $A(t)$, $r(t, t')$, and $t = t'$, (where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and t, t' range over individual names and variables) using conjunction, disjunction, and existential quantification. The variables in $\vec{x}$ are the *answer variables* of q. A PEQ without answer variables is *Boolean*. We say that a tuple $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$ of the same arity as $\vec{x}$ is an *answer to $q(\vec{x})$ in an interpretation $\mathcal{I}$* if $\mathcal{I} \models q[\vec{a}]$, where $q[\vec{a}]$ results from replacing the answer variables $\vec{x}$ in $q(\vec{x})$ with $\vec{a}$. Moreover, $\vec{a}$ is a *certain answer to $q(\vec{x})$ in $\mathcal{A}$ w.r.t. $\mathcal{T}$*, in symbols $(\mathcal{T}, \mathcal{A}) \models q(\vec{a})$, if $\mathcal{I} \models q[\vec{a}]$ for all $\mathcal{I} \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$. The set of all certain answers is denoted with $\mathsf{cert}_\mathcal{T}(q, \mathcal{A}) = \{\vec{a} \mid (\mathcal{T}, \mathcal{A}) \models q(\vec{a})\}$.

For Boolean queries q, we write $(\mathcal{T}, \mathcal{A}) \models q$ instead of $\mathsf{cert}_\mathcal{T}(q, \mathcal{A}) = \{()\}$ with $()$ the empty tuple; we then speak of deciding $(\mathcal{T}, \mathcal{A}) \models q$ rather than of computing $\mathsf{cert}_\mathcal{T}(q, \mathcal{A})$.

Example 1. (1) Let $\mathcal{T}_r = \{\exists r.A \sqsubseteq A\}$ and $q_0(x) = A(x)$. For any ABox $\mathcal{A}$, $\mathsf{cert}_{\mathcal{T}_r}(q_0, \mathcal{A})$ is the set of all $a \in \mathsf{Ind}(\mathcal{A})$ such that there is an r-path in $\mathcal{A}$ from a to some b with $A(b) \in \mathcal{A}$; i.e., there are $r(a_0, a_1), \dots, r(a_{n-1}, a_n) \in \mathcal{A}$, $n \geq 0$, with $a_0 = a$, $a_n = b$, and $A(b) \in \mathcal{A}$.

(2) Consider an undirected graph represented as an ABox $\mathcal{A}$ with assertions $r(a, b), r(b, a) \in \mathcal{A}$ iff there is an edge between a and b. Let $A_1, \dots, A_k, M$ be fresh concept names. Then $\mathcal{A}$ is k-colorable iff $(\mathcal{T}_k, \mathcal{A}) \not\models \exists x.M(x)$, where

$$\begin{array}{rcl} \mathcal{T}_k & = & \{A_i \sqcap A_j \sqsubseteq M \mid 1 \leq i < j \leq k\} \cup \\ & & \{A_i \sqcap \exists r.A_i \sqsubseteq M \mid 1 \leq i \leq k\} \cup \\ & & \{\top \sqsubseteq \bigsqcup_{1 \leq i \leq k} A_i\}. \end{array}$$

⊣

As additional query languages, we consider *conjunctive queries (CQs)*, which are PEQs without disjunction, as well as the following two weaker languages that are frequently used in an OBDA context.

Recall that *$\mathcal{EL}$-concepts* are constructed from $\mathsf{N_C}$ and $\mathsf{N_R}$ using conjunction, existential restriction, and the $\top$-concept (Baader, Brandt, and Lutz 2005). *$\mathcal{ELI}$-concepts* additionally admit inverse roles. If C is an $\mathcal{ELI}$-concept and $a \in \mathsf{N_I}$, then $C(a)$ is called an *$\mathcal{ELI}$-query (ELIQ)*; if C is an $\mathcal{EL}$-concept, then $C(a)$ is called an *$\mathcal{EL}$-query (ELQ)*. Note that every ELIQ (and, therefore, every ELQ) can be regarded as an acyclic Boolean CQ. For example, the ELIQ $\exists r.(A \sqcap \exists s^-.B)(a)$ is equivalent to the Boolean CQ

$$\exists x \exists y.(r(a, x) \wedge A(x) \wedge s(y, x) \wedge B(y)).$$

In what follows, we will not distinguish between an ELIQ and its translation into a Boolean CQ and freely apply notions introduced for PEQs also to ELIQs and ELQs.

For an ABox $\mathcal{A}$, we denote by $\mathcal{I}_\mathcal{A}$ the interpretation with $\Delta^{\mathcal{I}_\mathcal{A}} = \mathsf{Ind}(\mathcal{A})$, $a^{\mathcal{I}_\mathcal{A}} = a$ for all $a \in \mathsf{Ind}(\mathcal{A})$, and

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

for any $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. Note that $\mathsf{Ind}(\mathcal{I}) = \mathsf{Ind}(\mathcal{A})$.

In what follows, we sometimes slightly abuse notation and use PEQ to denote the set of all first-order queries, and likewise for CQ, ELIQ, and ELQ. We now introduce the main notions investigated in this paper.

Definition 2 (Complexity)**.** Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox and let $\mathcal{Q} \in \{\text{CQ}, \text{PEQ}, \text{ELIQ}, \text{ELQ}\}$. Then

- *$\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ is in* PTIME if for every $q(\vec{x}) \in \mathcal{Q}$, there is a polytime algorithm that computes, given an ABox $\mathcal{A}$, the set $\mathsf{cert}_\mathcal{T}(q, \mathcal{A})$;
- *$\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ is* CONP*-hard* if there is a Boolean $q \in \mathcal{Q}$ such that, given an ABox $\mathcal{A}$, it is coNP-hard to decide whether $(\mathcal{T}, \mathcal{A}) \models q$.

Note that $\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ is in PTIME iff for every *Boolean* query $q \in \mathcal{Q}$, there is a polytime algorithm deciding, given an ABox $\mathcal{A}$, whether $(\mathcal{T}, \mathcal{A}) \models q$. We give some examples that illustrate the above notions.

Example 3. (1) CQ-answering w.r.t. $\mathcal{T}_r$ from Example 1 is in PTIME since for any ABox $\mathcal{A}$, $\mathsf{cert}_{\mathcal{T}_r}(q, \mathcal{A})$ can be computed as follows. Let $\mathcal{A}'$ by the ABox obtained from $\mathcal{A}$ by adding $A(a)$ to $\mathcal{A}$ if there is an r-path from a to some b with $A(b) \in \mathcal{A}$. Then $\mathsf{cert}_{\mathcal{T}_r}(q, \mathcal{A}) = \{\vec{a} \mid \mathcal{I}_{\mathcal{A}'} \models q(\vec{a})\}$ can be computed in PTIME (actually in AC^0) by evaluating the PEQ q in the structure $\mathcal{I}_{\mathcal{A}'}$.

(2) Consider the TBoxes $\mathcal{T}_k$ from Example 3 that express k-colorability . For $k \geq 3$, CQ-answering w.r.t. $\mathcal{T}_k$ is CONP-hard since k-colorability is NP-hard. However, in contrast to the tractability of 2-colorability, CQ-answering w.r.t. $\mathcal{T}_2$ is CONP-hard as well. This follows from Theorem 11 below and, intuitively, is the case because $\mathcal{T}_2$ 'entails a disjunction': for $\mathcal{A} = \{B(a)\}$, we have $(\mathcal{T}_2, \mathcal{A}) \models A_1(a) \vee A_2(a)$, but neither $(\mathcal{T}_2, \mathcal{A}) \models A_1(a)$ nor $(\mathcal{T}_2, \mathcal{A}) \models A_2(a)$. ⊣

Interestingly, PTIME upper bounds (and the observations in Example 3) do not depend on whether we consider PEQs, CQs, or ELIQs.

Theorem 4. *For all $\mathcal{ALCFI}$-TBoxes $\mathcal{T}$,*

1. *CQ-answering w.r.t. $\mathcal{T}$ is in* PTIME *iff PEQ-answering w.r.t. $\mathcal{T}$ is in* PTIME *iff ELIQ-answering w.r.t. $\mathcal{T}$ is in* PTIME.
2. *ELIQ-answering w.r.t. $\mathcal{T}$ is in* PTIME *iff ELQ-answering w.r.t. $\mathcal{T}$ is in* PTIME*, provided that $\mathcal{T}$ is an $\mathcal{ALCF}$-ABox.*

The proof is based on Theorems 9 and 11 below. Theorem 4 gives a uniform explanation for the fact that, in the traditional logic-centered approach to data complexity in OBDA, the complexity of answering PEQs, CQs, and ELIQs has turned out to be identical for many DLs. It allows us to (sometimes) speak of the 'complexity of query answering' without reference to a concrete query language.

3 Materializability

We introduce materializability of a TBox $\mathcal{T}$ as a central tool for analyzing the complexity of query answering. Our main result is that *non*-materializability of a TBox is a sufficient condition for query answering being CONP-hard.

Definition 5 (Materializable)**.** Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox and $\mathcal{Q} \in \{\text{CQ}, \text{PEQ}, \text{ELIQ}, \text{ELQ}\}$. Then,

- a model $\mathcal{I}$ of $\mathcal{T}$ and an ABox $\mathcal{A}$ is a *$\mathcal{Q}$-materialization of $\mathcal{T}$ and $\mathcal{A}$* if for all queries $q(\vec{x}) \in \mathcal{Q}$ and potential answers $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{I} \models q[\vec{a}]$ iff $(\mathcal{T}, \mathcal{A}) \models q(\vec{a})$;
- $\mathcal{T}$ is *$\mathcal{Q}$-materializable* if for every ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$, there is a $\mathcal{Q}$-materialization of $\mathcal{T}$ and $\mathcal{A}$.

It can be proved that, in Example 3 (1), the interpretation $\mathcal{I}_{\mathcal{A}'}$ is a PEQ-materialization of $\mathcal{T}_r$ and $\mathcal{A}$. Note that a $\mathcal{Q}$-materialization can be viewed as a more abstract version of a *canonical model* as often used in the context of 'Horn DLs' such as $\mathcal{EL}$ and DL-Lite (Lutz, Toman, and Wolter 2009; Kontchakov et al. 2010). In fact, the ELQ-materialization in the next example is exactly the 'compact canonical model' from (Lutz, Toman, and Wolter 2009).

Example 6. Let $\mathcal{T} = \{A \sqsubseteq \exists r.A\}$ and $\mathcal{A}$ be an ABox with at least one assertion of the form $A(a)$. To obtain an ELQ-materialization $\mathcal{M}_\mathcal{A}$ of $\mathcal{T}$ and $\mathcal{A}$, start with the interpretation $\mathcal{I}_\mathcal{A}$, add a fresh domain element d_r, and set

$$\begin{aligned} A^{\mathcal{M}_\mathcal{A}} &= A^{\mathcal{I}_\mathcal{A}} \cup \{d_r\} \\ r^{\mathcal{M}_\mathcal{A}} &= r^{\mathcal{I}_\mathcal{A}} \cup \{(a, d_r) \mid A(a) \in \mathcal{A}\} \cup \{(d_r, d_r)\}. \end{aligned}$$ ⊣

Trivially, a PEQ-materialization is a CQ-materialization is an ELIQ-materialization is an ELQ-materialization. We show below as part of Lemma 8 that the converse holds for the CQ-PEQ case. However, the following example demonstrates that ELQ-materializations are different from ELIQ-materialization. A similar argument separates ELIQ-materializations from CQ-materializations.

Example 7. Let $\mathcal{T}$ be as in Example 6,

$$\begin{aligned} \mathcal{A} &= \{B_1(a), B_2(b), A(a), A(b)\} \text{ and} \\ q &= (B_1 \sqcap \exists r.\exists r^-.B_2)(a), \end{aligned}$$

Then the ELQ-materialization $\mathcal{M}_\mathcal{A}$ from Example 6 is not a $\mathcal{Q}$-materialization for any $\mathcal{Q} \in \{\text{ELIQ}, \text{CQ}, \text{PEQ}\}$. For example, we have $\mathcal{M}_\mathcal{A} \models q$, but $(\mathcal{T}, \mathcal{A}) \not\models q$. An ELIQ/CQ/PEQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ is obtained by unfolding $\mathcal{M}_\mathcal{A}$: instead of using only one additional individual d_r as a witness for $\exists r.A$, we attach to both a ad b an infinite r-path of elements that satisfy A. Note that every CQ/PEQ-materialization of $\mathcal{T}_r$ and $\mathcal{A}$ must be infinite. ⊣

Before linking materializations to the complexity of query answering, we characterize them semantically in terms of simulations and homomorphisms. This is interesting in its own right and establishes a close connection between materialization and initial models as studied in model theory, algebraic specification, and logic programming (Malcev 1971; Meseguer and Goguen 1985; Makowsky 1987). It also allows us to show that, despite the discrepancies between materiali*zations* for different query languages pointed out above, materializa*bility* coincides for PEQs, CQs, and ELIQs (and ELQs when the TBox is formulated in $\mathcal{ALCF}$).

Analyzing Materializability

A *simulation from an interpretation* $\mathcal{I}_1$ *to an interpretation* $\mathcal{I}_2$ is a relation $S \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ such that

1. for all $A \in \mathsf{N_C}$: if $d_1 \in A^{\mathcal{I}_1}$ and $(d_1, d_2) \in S$, then $d_2 \in A^{\mathcal{I}_2}$;
2. for all $r \in \mathsf{N_R}$: if $(d_1, d_2) \in S$ and $(d_1, d_1') \in r^{\mathcal{I}_1}$, then there exists $d_2' \in \Delta^{\mathcal{I}_2}$ such that $(d_1', d_2') \in S$ and $(d_2, d_2') \in r^{\mathcal{I}_2}$;
3. for all $a \in \mathsf{Ind}(\mathcal{I}_1)$: $a \in \mathsf{Ind}(\mathcal{I}_2)$ and $(a^{\mathcal{I}_1}, a^{\mathcal{I}_2}) \in S$.

We call S an *i-simulation* if Condition 2 is satisfied also for *inverse* roles and a *homomorphism* if S is a function. An interpretation $\mathcal{I}$ is called *hom-initial in a class* $\mathbb{K}$ *of interpretations* if for every $\mathcal{J} \in \mathbb{K}$, there exists a homomorphism from $\mathcal{I}$ to $\mathcal{J}$. $\mathcal{I}$ is called *sim-initial (i-sim-initial) in a class* $\mathbb{K}$ *of interpretations* if for every $\mathcal{J} \in \mathbb{K}$, there exists a simulation (i-simulation) from $\mathcal{I}$ to $\mathcal{J}$.

An interpretation $\mathcal{I}$ is *generated* if every $d \in \Delta^{\mathcal{I}}$ is reachable from some $a^{\mathcal{I}}$, $a \in \mathsf{Ind}(\mathcal{I})$, in the undirected graph $(\Delta^{\mathcal{I}}, \{\{d, d'\} \mid (d, d') \in \bigcup_{r \in \mathsf{N_R}} r^{\mathcal{I}}\})$. The next result relates simulations and homomorphisms to materializations.

Lemma 8. *Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox, $\mathcal{A}$ an ABox, and $\mathcal{I} \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$. Then $\mathcal{I}$ is*

1. *an ELIQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is i-sim-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$;*
2. *a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is a PEQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is hom-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$, provided that $\mathcal{I}$ is countable and generated;*
3. *an ELQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is sim-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$, provided that $\mathcal{T}$ is an $\mathcal{ALCF}$-TBox.*

Proof. (Sketch) The proofs of "$\Leftarrow$" are straightforward since matches of PEQs, CQs, and ELIQs are preserved under i-simulations and homomorphisms, and matches of ELQs are preserved under simulations. We thus concentrate on "$\Rightarrow$".

(1) Assume $\mathcal{I}$ is an ELIQ-materialization and let $\mathcal{J} \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$. If $\mathcal{J}$ has finite outdegree, an i-simulation from $\mathcal{I}$ to $\mathcal{J}$ can be constructed in the same way as in standard proofs showing that simulations characterize the expressive power of $\mathcal{EL}$-concepts (Lutz, Piro, and Wolter 2011). If $\mathcal{J}$ has infinite outdegree, then one can construct a selective unfolding $\mathcal{J}^* \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$ of $\mathcal{J}$ whose outdegree is finite and such that there is a homomorphism from $\mathcal{J}^*$ to $\mathcal{J}$. It remains to compose an i-simulation from $\mathcal{I}$ to $\mathcal{J}^*$ with the homomorphism from $\mathcal{J}^*$ to $\mathcal{J}$.

For (2), we show that any countable and generated CQ-materialization is hom-initial. If $\mathcal{I}$ is such a CQ-materialization and $\mathcal{J} \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$, then by the semantics of CQs we can find a homomorphism from any *finite subinterpretation* of $\mathcal{I}$ to $\mathcal{J}$. If $\mathcal{J}$ is of finite outdegree, we can assemble all those homomorphisms into a homomorphism from $\mathcal{I}$ to $\mathcal{J}$ in a direct way (using that $\mathcal{I}$ is countable and generated). For $\mathcal{J}$ of non-finite outdegree, we compose the homorphism from $\mathcal{I}$ to $\mathcal{J}^*$ with the homomorphism from $\mathcal{J}^*$ to $\mathcal{J}$, with $\mathcal{J}^*$ constructed as in (1). The claim for $\mathcal{ALCF}$-TBoxes is proved similarly to (1). ❑

In Point 2 of Lemma 8, we cannot drop the condition that $\mathcal{I}$ is generated without losing correctness, please see the long version for details. It is open whether the same is true for countability.

We now show that materializability coincides for the query languages studied in this paper.

Theorem 9. *Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox. Then*

1. *$\mathcal{T}$ is PEQ-materializable iff $\mathcal{T}$ is CQ-materializable iff $\mathcal{T}$ is ELIQ-materializable;*
2. *the above is the case iff $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$ contains a hom-initial $\mathcal{I}$ for every ABox $\mathcal{A}$ iff $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$ contains an i-sim-initial $\mathcal{I}$ for every ABox $\mathcal{A}$;*
3. *the above is the case iff $\mathcal{T}$ is ELQ-materializable iff $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$ contains a sim-initial $\mathcal{I}$ for every ABox $\mathcal{A}$, provided that $\mathcal{T}$ is an $\mathcal{ALCF}$-TBox.*

This theorem is essentially a consequence of Lemma 8. The proof of "$\Leftarrow$" in Point 2 employs a selective unfolding technique (similar to the one used in the the proof of Lemma 8) to transform an i-simulation into a homomorphism. Due to this technique, the conditions of generatedness and countability from Point 2 of Lemma 8 can be avoided in Theorem 9.

Because of Theorem 9, we sometimes speak of *materializability* without reference to a query language and of *materializations* instead of PEQ-materializations. Interestingly, materializability turns out to (also) be equivalent to the disjunction property, which is sometimes also called convexity and plays a central role in attaining PTIME complexity for standard reasoning in DLs (Baader, Brandt, and Lutz 2005). This observation will be useful for the proof of our main theorem below.

A TBox $\mathcal{T}$ has the *ABox disjunction property* if for all ABoxes $\mathcal{A}$ and ELIQs $C_1(a_1), \ldots, C_n(a_n)$, it follows from $(\mathcal{T}, \mathcal{A}) \models C_1(a_1) \vee \ldots \vee C_n(a_n)$ that $(\mathcal{T}, \mathcal{A}) \models C_i(a_i)$ for some $i \leq n$.

Theorem 10. *An $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is materializable iff it has the disjunction property.*

Proof. For the nontrivial "$\Leftarrow$" direction, let $\mathcal{A}$ be an ABox that is consistent w.r.t. $\mathcal{T}$ and such that there is no ELIQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. Then $\mathcal{T} \cup \mathcal{A} \cup \Gamma$ is not satisfiable, where

$$\Gamma = \{\neg C(a) \mid (\mathcal{T}, \mathcal{A}) \not\models C(a), a \in \mathsf{Ind}(\mathcal{A}), C(a) \text{ ELIQ}\}.$$

In fact, any satisfying interpretation would be an ELIQ-materialization. By compactness, there is a finite subset Γ' of Γ such that $\mathcal{T} \cup \mathcal{A} \cup \Gamma'$ is not satisfiable, i.e. $(\mathcal{T}, \mathcal{A}) \models \bigvee_{\neg C(a) \in \Gamma'} C(a)$. By definition of Γ', $(\mathcal{T}, \mathcal{A}) \not\models C(a)$, for all $\neg C(a) \in \Gamma'$. Thus, $\mathcal{T}$ lacks the ABox disjunction property. ❑

Materializability and coNP-hardness

Based on Theorems 9 and 10, we now establish the main result on materializability.

Theorem 11. *If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ ($\mathcal{ALCF}$-TBox $\mathcal{T}$) is not materializable, then ELIQ-answering (ELQ-answering) is* coNP*-hard w.r.t. $\mathcal{T}$.*

The proof exploits failure of the ABox disjunction property to generalize the reduction of 2+2-SAT used in (Schaerf 1993) to show that ELQ-answering in a variant of $\mathcal{EL}$ is coNP-hard.

The converse of Theorem 11 fails, i.e., there are TBoxes that are materializable, but for which ELIQ-answering is coNP-hard. In fact, materializations of such a TBox $\mathcal{T}$ and ABox $\mathcal{A}$ are guaranteed to exist, but cannot always be computed in PTime (unless PTime = coNP). Technically, this follows from Theorem 20 later on which states that for every non-uniform CSP, there is a materializable $\mathcal{ALC}$-TBox for which Boolean CQ-answering has the same complexity, up to complementation of the complexity class.

Theorem 11 also allows us to prove Theorem 4 (for this purpose, it is crucial for Theorem 11 to refer to ELIQs and ELQs rather than CQs or PEQs).

Proof of Theorem 4 (sketch). By Theorem 11, it is sufficient to consider materializable TBoxes when proving Theorem 4. To show, for example, that if CQ-answering w.r.t. $\mathcal{T}$ is in PTime then PEQ-answering w.r.t. $\mathcal{T}$ is in PTime, one can first transform a PEQ $q(\vec{x})$ into an equivalent union of CQs $\bigsqcup_{i \in I} q_i(\vec{x})$. CQ-materializability of $\mathcal{T}$ implies that, for any ABox $\mathcal{A}$, we have $\mathsf{cert}_\mathcal{T}(q,\mathcal{A}) = \bigcup_{i \in I} \mathsf{cert}_\mathcal{T}(q_i,\mathcal{A})$. It thus remains to note that each set $\mathsf{cert}_\mathcal{T}(q_i,\mathcal{A})$ can be computed in PTime. The remaining reductions are more involved, but based on similar ideas. □

4 Unraveling Tolerance

We develop a condition on TBoxes, called unraveling tolerance, that is sufficient for PTime query answering and strictly generalizes syntactic 'Horn conditions' such as the ones used to define the DL Horn-$\mathcal{SHIQ}$, which was designed as a maximal DL with PTime query answering (Hustadt, Motik, and Sattler 2007; Eiter et al. 2008). Unraveling tolerance is based on an unraveling operation on ABoxes, in the same spirit as the well-known unraveling of an interpretation into a tree interpretation. More precisely, the *unraveling* $\mathcal{A}^u$ of an ABox $\mathcal{A}$ is the following (possibly infinite) ABox:

- $\mathsf{Ind}(\mathcal{A}^u)$ is the set of sequences $b_0 r_0 b_1 \cdots r_{n-1} b_n$, $n \geq 0$, with $b_0, \ldots, b_n \in \mathsf{Ind}(\mathcal{A})$ and $r_0, \ldots, r_{n-1} \in \mathsf{N_R} \cup \mathsf{N_R^-}$ such that for all $i < n$, we have $r_i(b_i, b_{i+1}) \in \mathcal{A}$ and $(b_{i-1}, r_{i-1}^-) \neq (b_{i+1}, r_i)$ when $i > 0$;
- for each $C(b) \in \mathcal{A}$ and $\alpha = b_0 \cdots b_n \in \mathsf{Ind}(\mathcal{A}^u)$ with $b_n = b$, we have $C(\alpha) \in \mathcal{A}^u$;
- for each $\alpha = b_0 r_0 \cdots r_{n-1} b_n \in \mathsf{Ind}(\mathcal{A}^u)$ with $n > 0$, we have $r_{n-1}(b_0 \cdots b_{n-1}, \alpha) \in \mathcal{A}^u$.

For all $\alpha = b_0 \cdots b_n \in \mathsf{Ind}(\mathcal{A}^u)$, we write $\mathsf{tail}(\alpha)$ to denote b_n. Note that the condition $(b_{i-1}, r_{i-1}^-) \neq (b_{i+1}, r_i)$ is needed to ensure that functional roles can still be interpreted in a functional way after unraveling.

Definition 12 (Unraveling Tolerance)**.** A TBox $\mathcal{T}$ is *unraveling tolerant* if for all ABoxes $\mathcal{A}$ and ELIQs q, we have that $(\mathcal{T},\mathcal{A}) \models q$ implies $(\mathcal{T},\mathcal{A}^u) \models q$.

It is not hard to prove that the converse direction '$(\mathcal{T},\mathcal{A}^u) \models q$ implies $(\mathcal{T},\mathcal{A}) \models q$' is true for *all* $\mathcal{ALCFI}$-TBoxes. Note that it makes no sense to define unraveling tolerance for queries that are not necessarily tree shaped, such as CQs.

Example 13. (1) The $\mathcal{ALC}$-TBox $\mathcal{T}_1 = \{A \sqsubseteq \forall r.B\}$ is unraveling tolerant. This can be proved by showing that (i) for any (finite or infinite) ABox $\mathcal{A}$, the interpretation $\mathcal{I}_\mathcal{A}^+$ that is obtained from $\mathcal{I}_\mathcal{A}$ by setting $B^{\mathcal{I}_\mathcal{A}^+} = B^{\mathcal{I}_\mathcal{A}} \cup (\exists r^-.A)^{\mathcal{I}_\mathcal{A}}$ is an ELIQ-materialization of $\mathcal{T}_1$ and $\mathcal{A}$; and (ii) $\mathcal{I}_\mathcal{A}^+ \models C(a)$ iff $\mathcal{I}_{\mathcal{A}^u}^+ \models C(a)$ for all ELIQs $C(a)$. The proof of (ii) is based on a simple induction on the structure of the $\mathcal{ELI}$-concept C. As witnessed by the ABox $\mathcal{A} = \{r(a,b), A(a)\}$ and ELIQ $B(b)$, the use of inverse roles in the definition of $\mathcal{A}^u$ is crucial here despite the fact that $\mathcal{T}_1$ does not use inverse roles.

(2) A simple example for an $\mathcal{ALC}$-TBox that is not unraveling tolerant is $\mathcal{T}_2 = \{A \sqcap \exists r.A \sqsubseteq B, \neg A \sqcap \exists r.\neg A \sqsubseteq B\}$. For $\mathcal{A} = \{r(a,a)\}$, it is easy to see that we have $(\mathcal{T}_2,\mathcal{A}) \models B(a)$ (use a case distinction on the truth value of A at a!), but $(\mathcal{T}_2,\mathcal{A}^u) \not\models B(a)$. ⊣

Before we show that unraveling tolerance indeed implies PTime query answering, we first demonstrate the generality of this property by relating it to Horn-$\mathcal{ALCFI}$, the $\mathcal{ALCFI}$-fragment of Horn-$\mathcal{SHIQ}$. Different versions of Horn-$\mathcal{SHIQ}$ have been proposed in the literature, giving rise to different versions of Horn-$\mathcal{ALCFI}$ (Hustadt, Motik, and Sattler 2007; Krötzsch, Rudolph, and Hitzler 2007; Eiter et al. 2008; Kazakov 2009). As the original definition from (Hustadt, Motik, and Sattler 2007) based on polarity is rather technical, we prefer to work with the following, more direct syntax. A *Horn-$\mathcal{ALCFI}$-TBox* has the form $\mathcal{T} = \{\top \sqsubseteq C_\mathcal{T}\} \cup \mathcal{F}$, where $\mathcal{F}$ is a set of functionality assertions and $C_\mathcal{T}$ is built according to the topmost rule in

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

where r ranges over $\mathsf{N_R} \cup \mathsf{N_R^-}$ and $L \rightarrow R := \neg L \sqcup R$. By applying some simple transformations, it is not hard to show that every Horn-$\mathcal{ALCFI}$-TBox according to the original, polarity-based definition is equivalent to a Horn-$\mathcal{ALCFI}$-TBox of the form introduced here. Although not important in our context, we note that even a polytime transformation is possible.

Theorem 14.
Every Horn-$\mathcal{ALCFI}$-TBox is unraveling tolerant.

Proof. (hint) Based on a generalization of the argument in Example 13 (1), where the ad hoc materialization $\mathcal{I}_\mathcal{A}^+$ is replaced with a systematically constructed *canonical model* of $\mathcal{T}$ and $\mathcal{A}$. ❑

Theorem 14 shows that unraveling tolerance and Horn logic are closely related. Yet, the next example shows that there are unraveling tolerant $\mathcal{ALCFI}$-TBoxes that are not equivalent to any Horn sentence of FO. Since any Horn-$\mathcal{ALCFI}$-TBox is equivalent to such a sentence, it follows that unraveling tolerant $\mathcal{ALCFI}$-TBoxes strictly generalize Horn-$\mathcal{ALCFI}$-TBoxes. This increased generality will pay off in

Section 5 when we establish a dichotomy result for TBoxes of depth one.

Example 15. Take the $\mathcal{ALC}$-TBox

$$\mathcal{T} = \{\exists r.(A \sqcap \neg B_1 \sqcap \neg B_2) \sqsubseteq \exists r.(\neg A \sqcap \neg B_1 \sqcap \neg B_2)\}.$$

One can show as in Example 13 (1) that $\mathcal{T}$ is unraveling tolerant; here, the materialization is actually $\mathcal{I}_\mathcal{A}$ itself instead of some $\mathcal{I}_\mathcal{A}^+$, i.e., as far as ELIQ (and even PEQ) answering is concerned, $\mathcal{T}$ cannot be distinguished from the empty TBox.

It is well-known that FO Horn sentences are preserved under direct products of interpretations (Chang and Keisler 1990). To show that $\mathcal{T}$ is not equivalent to any such sentence, it thus suffices to show that $\mathcal{T}$ is not preserved under direct products. This is simple: let $\mathcal{I}_1$ and $\mathcal{I}_2$ consist of a single r-edge between elements d and e, and let $e \in (A \sqcap B_1 \sqcap \neg B_2)^{\mathcal{I}_1}$ and $e \in (A \sqcap \neg B_1 \sqcap B_2)^{\mathcal{I}_2}$; then the direct product $\mathcal{I}$ of $\mathcal{I}_1$ and $\mathcal{I}_2$ still has the r-edge between (d,d) and (e,e) and satisfies $(e,e) \in (A \sqcap \neg B_1 \sqcap \neg B_2)^\mathcal{I}$, thus is not a model of $\mathcal{T}$. ⊣

We now establish the PTIME upper bound for unraveling tolerant TBoxes.

Theorem 16. *If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is unraveling tolerant, then PEQ-answering w.r.t. $\mathcal{T}$ is in* PTIME.

Proof.(sketch) Let $\mathcal{T}$ be unraveling tolerant. By Theorem 4, it suffices to show that ELIQ-answering w.r.t. $\mathcal{T}$ is in PTIME. Let $\mathcal{A}$ be an ABox and $q = C_0(a_0)$ an ELIQ. Let $\mathsf{cl}(\mathcal{T}, C_0)$ denote the closure under single negation of the set of subconcepts of $\mathcal{T}$ and C_0. $\mathsf{tp}(\mathcal{T}, C_0)$ denotes the set of all *types* (aka Hintikka sets or maximal consistent sets) over $\mathsf{cl}(\mathcal{T}, C_0)$. A *type assignment* is a map $\mathsf{Ind}(\mathcal{A}) \rightarrow 2^{\mathsf{tp}(\mathcal{T},q)}$.

The PTIME algorithm for checking whether $(\mathcal{T}, \mathcal{A}) \models q$ is based on the computation of a sequence of type assignments $\pi_0, \pi_1, \ldots$ as follows. For every $a \in \mathsf{Ind}(\mathcal{A})$, $\pi_0(a)$ is the set of types $t \in \mathsf{tp}(\mathcal{T}, q)$ such that $A(a) \in \mathcal{A}$ implies $A \in t$. Then, $\pi_{i+1}(a)$ is defined as the set of types $t_a \in \pi_i(a)$ such that for all $r(a,b) \in \mathcal{A}$, r a role name or the inverse thereof, there is a type $t_b \in \pi_i(b)$ such that $t_a \rightsquigarrow_r t_b$, where we write $t_a \rightsquigarrow_r t_b$ if the following conditions are satisfied: if $C \in t_b$ then $\exists r.C \in t_a$, for all $\exists r.C \in \mathsf{cl}(\mathcal{T}, C_0)$; if $C \in t_a$ then $\exists r^-.C \in t_b$, for all $\exists r^-.C \in \mathsf{cl}(\mathcal{T}, C_0)$; $\exists r.C \in t_a$ iff $C \in t_b$, for all $\exists r.C \in \mathsf{cl}(\mathcal{T}, C_0)$ with $\mathsf{func}(r) \in \mathcal{T}$; $\exists r^-.C \in t_b$ iff $C \in t_a$, for all $\exists r^-.C \in \mathsf{cl}(\mathcal{T}, C_0)$ with $\mathsf{func}(r^-) \in \mathcal{T}$.

Clearly, the sequence $\pi_0, \pi_1, \ldots$ stabilizes after at most $\mathcal{O}(|\mathcal{A}|)$ steps and can be computed in time polynomial in $|\mathcal{A}|$ (since the cardinality of $\mathsf{tp}(\mathcal{T}, q)$ is bounded by a constant). Let π be the final type assignment in the sequence. In the long version, we show that $(\mathcal{T}, \mathcal{A}) \models q$ iff $C_0 \in t$ for all $t \in \pi(a_0)$. ❑

By Theorems 4 and 14 and since we actually exhibit a *uniform* algorithm for ELIQ-answering w.r.t. unraveling tolerant TBoxes, Theorem 16 also reproves the known PTIME upper bound for CQ-answering in Horn-$\mathcal{ALCFI}$ (Eiter et al. 2008).

By Theorems 11 and 16, unraveling tolerance implies materializability unless PTIME $=$ NP. Based on the disjunction property, this implication can also be proved without the side condition.

Lemma 17. *Every unraveling tolerant $\mathcal{ALCFI}$-TBox is materializable.*

The converse of Lemma 17 and, more generally, of Theorem 16 fails (unless PTIME $=$ NP). In fact, while unraveling tolerance is a sufficient condition for PTIME query answering, it is not a necessary one. An example is given in Section 6, where it is shown that the TBox $\mathcal{T}_2$ from Example 1 that represents 2-colorability has PTIME query answering, but is not unraveling tolerant.

The PTIME algorithm in Theorem 16 resembles the standard *arc consistency* algorithm for CSPs (Dechter 2003). This link to CSPs can be formalized for $\mathcal{ALCI}$-TBoxes using the templates $\mathcal{I}_{\mathcal{T},q}$ constructed in the proof of Theorems 22 and 24 below: it is known that a CSP can be solved using arc consistency iff it has tree obstructions (Krokhin 2010). Also, one can show that an $\mathcal{ALCI}$-TBox $\mathcal{T}$ is unraveling tolerant iff all templates $\mathcal{I}_{\mathcal{T},q}$ from Theorem 24 have tree obstructions. Consequently, for any $\mathcal{ALCI}$-TBox $\mathcal{T}$, ELIQs can be answered using an arc consistency algorithm iff $\mathcal{T}$ is unraveling tolerant.

5 Dichotomy for Depth One

We establish a dichotomy between PTIME and coNP for TBoxes of depth one, i.e., sets of CIs $C \sqsubseteq D$ such that the maximum nesting depth of the constructors $\exists r.E$ and $\forall r.E$ in C and D is one.[1] All examples given in the present paper up to this point use TBoxes of depth one.

Our main observation is that, when the depth of TBoxes is restricted to one, we can prove a converse of Theorem 17.

Theorem 18. *Every materializable $\mathcal{ALCFI}$-TBox of depth one is unraveling tolerant.*

Proof. (sketch) Let $\mathcal{T}$ be a materializable TBox of depth one, $\mathcal{A}$ an ABox, and q an ELIQ with $(\mathcal{T}, \mathcal{A}^u) \not\models q$. We have to show that $(\mathcal{T}, \mathcal{A}) \not\models q$. It follows from $(\mathcal{T}, \mathcal{A}^u) \not\models q$ that $\mathcal{A}^u$ is consistent w.r.t. $\mathcal{T}$ and thus there is a materialization $\mathcal{I}^u$ for $\mathcal{T}$ and $\mathcal{A}^u$ (even though $\mathcal{A}^u$ can be infinite, see long version). We have $\mathcal{I}^u \not\models q$ and our aim is to convert $\mathcal{I}^u$ into a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $\mathcal{I} \not\models q$. This is done in two steps.

As a preliminary to the first step, we note that $\mathcal{I}^u$ can be assumed w.l.o.g. to have forest-shape, i.e., $\mathcal{I}^u$ can be constructed by selecting a tree-shaped interpretation $\mathcal{I}_\alpha$ with root α for each $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$, then taking the disjoint union of all these interpretations, and finally adding role edges (α, β) to $r^{\mathcal{I}^u}$ whenever $r(\alpha, \beta) \in \mathcal{A}^u$. In fact, to achieve the desired shape we can simply unravel $\mathcal{I}^u$ starting from the elements $\mathsf{Ind}(\mathcal{A}^u) \subseteq \Delta^{\mathcal{I}^u}$ and then use Point 1 of Lemma 8 and the fact that there is an i-simulation from the unraveling of $\mathcal{I}^u$ to $\mathcal{I}^u$ to show that the obtained model is still a materialization of $\mathcal{T}$ and $\mathcal{A}$.

[1] Our results even apply to TBoxes that have depth one *after* replacing all $\mathcal{ELI}$-subconcepts with concept names, since $\mathcal{ELI}$-concept definitions do not affect the complexity of ELIQ-answering. This captures >90% of the TONES repository TBoxes.

Now, step one of the construction is to uniformize $\mathcal{I}^u$ such that for all $\alpha, \beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$, the tree component $\mathcal{I}_\alpha$ of $\mathcal{I}^u$ is isomorphic to the tree component $\mathcal{I}_\beta$ of $\mathcal{I}^u$. To achieve this while preserving the property that $\mathcal{I}^u \not\models q$, we rely on the self-similarity of the ABox $\mathcal{A}^u$: for all $\alpha, \beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$, we can find an automorphism on $\mathcal{A}^u$ that maps α to β.

Step two is to construct the desired model $\mathcal{I}$ of $\mathcal{T}$ and the original ABox $\mathcal{A}$, starting from the uniformized version of $\mathcal{I}^u$: take the disjoint union of all the tree components $\mathcal{I}_a$ of $\mathcal{I}_u$, with $a \in \mathsf{Ind}(\mathcal{A})$ (note that $\mathsf{Ind}(\mathcal{A}) \subseteq \mathsf{Ind}(\mathcal{A}^u)$), and add (a, b) to $r^\mathcal{I}$ whenever $r(a, b) \in \mathcal{A}$. Due to the uniformity of $\mathcal{I}^u$, we can find an i-simulation from $\mathcal{I}$ to $\mathcal{I}_u$. Since matches of ELIQs are preserved under i-simulations, $\mathcal{I}_u \not\models q$ thus implies $\mathcal{I} \not\models q$. ❑

The desired dichotomy follows: If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ of depth one is materializable, then PEQ-answering w.r.t. $\mathcal{T}$ is in PTIME by Theorems 18 and 16. Otherwise, ELIQ-answering w.r.t. $\mathcal{T}$ is coNP-complete by Theorem 11.

Theorem 19 (Dichotomy). *For every $\mathcal{ALCFI}$-TBox $\mathcal{T}$ of depth one, one of the following is true:*

- *$\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ is in* PTIME *for any $\mathcal{Q} \in$ {PEQ,CQ,ELIQ};*
- *$\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ is* coNP*-complete for any $\mathcal{Q} \in$ {PEQ,CQ,ELIQ}.*

We close this section by a brief discussion of why analyzing the complexity of query answering is easier for TBoxes of depth one than for TBoxes of unrestricted depth, when there is no such difference for other reasoning problems such as subsumption. Of course, every TBox can be converted to a TBox of depth one by introducing additional concept names A_C that replace compound concepts C which occur as an argument in $\exists r.C$ or $\forall r.C$. The trouble is that these concept names can then be used in a CQ, which results in an 'import' of C into the query language. This is obviously problematic, for example when C has the form $\forall r.D$, which is otherwise not expressible as a CQ. In the next section, we will use this effect to reduce CSPs to query answering with TBoxes of depth > 1.

6 Query Answering in $\mathcal{ALC}/\mathcal{ALCI}$ = CSP

We show that query answering w.r.t. $\mathcal{ALC}$- and $\mathcal{ALCI}$-TBoxes has the same computational power as non-uniform CSPs in the following sense: (i) for every CSP, there is an $\mathcal{ALC}$-TBox such that query answering w.r.t. $\mathcal{T}$ is of the same complexity, up to complementation; conversely, (ii) for every $\mathcal{ALCI}$-TBox $\mathcal{T}$ and ELIQ q, there is a CSP that has the same complexity as answering q w.r.t. $\mathcal{T}$, up to complementation. This has many interesting consequences, a main one being that the Feder-Vardi conjecture holds if and only if there is a PTIME/coNP-dichotomy for query answering w.r.t. $\mathcal{ALC}$-TBoxes (equivalently $\mathcal{ALCI}$-TBoxes). All this is true already for *materializable* TBoxes. By Theorem 4 and since we carefully choose the appropriate query language in each technical result below, it is true for any of the languages ELIQ, CQ, and PEQ (and ELQ for $\mathcal{ALC}$-TBoxes).

We begin by introducing non-uniform CSPs. Since every non-uniform CSP is polynomially equivalent to a non-uniform CSP with one binary predicate (Feder and Vardi 1993), we consider CSPs over unary and binary predicates (concept names and role names), only. A *signature* Σ is a finite set of concept and role names. An interpretation $\mathcal{I}$ is a Σ-*interpretation* if $\mathsf{Ind}(\mathcal{I}) = \emptyset$ and $X^\mathcal{I} = \emptyset$ for all $X \in (\mathsf{N_C} \cup \mathsf{N_R}) \setminus \Sigma$. For two finite Σ-interpretations $\mathcal{I}$ and $\mathcal{I}'$, we write $\mathsf{Hom}(\mathcal{I}', \mathcal{I})$ if there is a homomorphism from $\mathcal{I}'$ to $\mathcal{I}$. Any Σ-interpretation $\mathcal{I}$ gives rise to the following non-uniform constraint satisfaction problem in signature Σ, denoted by $\mathsf{CSP}(\mathcal{I})$: given a finite Σ-interpretation $\mathcal{I}'$, decide whether $\mathsf{Hom}(\mathcal{I}', \mathcal{I})$. Numerous algorithmic problems can be given in the form $\mathsf{CSP}(\mathcal{I})$. For example, k-colorability is $\mathsf{CSP}(\mathsf{C}_k)$, where C_k is an $\{r\}$-interpretation defined by setting $\Delta^{\mathsf{C}_k} = \{1, \dots, k\}$ and $r^{\mathsf{C}_k} = \{(i, j) \mid i \neq j\}$.[2]

We first show how to convert a CSP into a (materializable) $\mathcal{ALC}$-TBox. For a Σ-interpretation $\mathcal{I}$, $\mathcal{A}_\mathcal{I}$ denotes $\mathcal{I}$ viewed as an ABox: $\mathcal{A}_\mathcal{I} = \{A(a_d) \mid A \in \Sigma \cap \mathsf{N_C} \wedge d \in A^\mathcal{I}\} \cup \{r(a_d, a_e) \mid r \in \Sigma \cap \mathsf{N_R} \wedge (d, e) \in r^\mathcal{I}\}$.

Theorem 20. *For every non-uniform constraint satisfaction problem $\mathsf{CSP}(\mathcal{I})$ in signature Σ, one can compute (in polytime) a materializable $\mathcal{ALC}$-TBox $\mathcal{T}_\mathcal{I}$ such that*

1. *$\mathsf{Hom}(\mathcal{J}, \mathcal{I})$ iff $\mathcal{A}_\mathcal{J}$ is consistent w.r.t. $\mathcal{T}_\mathcal{I}$, for all Σ-interpretations $\mathcal{J}$;*
2. *for any Boolean PEQ q, answering q w.r.t. $\mathcal{T}_\mathcal{I}$ is polynomially reducible (in fact, FO-reducible) to the complement of $\mathsf{CSP}(\mathcal{I})$.*

Note that $\mathsf{CSP}(\mathcal{I})$ and $\mathcal{T}_\mathcal{I}$ 'have the same complexity' in the following sense: by Point 1 of Theorem 20, $\mathsf{CSP}(\mathcal{I})$ reduces to consistency of ABoxes w.r.t. $\mathcal{T}_\mathcal{I}$; since an ABox $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_\mathcal{I}$ iff $(\mathcal{T}_\mathcal{I}, \mathcal{A}) \not\models A(a)$ with A a fresh concept name and $a \in \mathsf{Ind}(\mathcal{A})$, this also yields a reduction from the complement of $\mathsf{CSP}(\mathcal{I})$ to ELQ-answering w.r.t. $\mathcal{T}_\mathcal{I}$; conversely, Point 2 ensures that (Boolean) PEQ-answering w.r.t. $\mathcal{T}_\mathcal{I}$ reduces to the complement of $\mathsf{CSP}(\mathcal{I})$. All reductions are extremely simple, in polytime and in fact even FO-reductions.

Our approach to proving Theorem 20 is to generalize the reduction of k-colorability to query answering w.r.t. $\mathcal{ALC}$-TBoxes discussed in Examples 1 and 3, where the main challenge is to overcome the observation from Example 3 that PTIME CSPs such as 2-colorability may be translated into coNP-hard TBoxes. Note that this is due to the disjunction in the TBox $\mathcal{T}_k$ of Example 1, which causes non-materializability. Our solution is to replace the concept names $A_1, \dots, A_k$ in $\mathcal{T}_k$ with compound concepts that are 'invisible to the query', behaving essentially like second-order variables. Unlike the original depth one TBox $\mathcal{T}_k$, the resulting TBox is of depth three.

In detail, fix a constraint satisfaction problem $\mathsf{CSP}(\mathcal{I})$, reserve a concept name Z_d and role names r_d, s_d for any $d \in \Delta^\mathcal{I}$, and set

$$\mathcal{T} = \{\top \sqsubseteq \exists r_d.\top, \top \sqsubseteq \exists s_d.Z_d \mid d \in \Delta^\mathcal{I}\}$$
$$H_d = \forall r_d.\exists s_d.\neg Z_d, \quad d \in \Delta^\mathcal{I}$$

[2]Although the input to $\mathsf{CSP}(\mathsf{C}_k)$ formally is a digraph, it is treated like an undirected graph.

The following shows that we can use the concepts H_d as unary predicates to represent the 'values' of $\mathsf{CSP}(\mathcal{I})$ (these values are the domain elements of $\mathcal{I}$).

Lemma 21. *For every ABox $\mathcal{A}$ and family of sets $I_d \subseteq \mathsf{Ind}(\mathcal{A})$, $d \in \Delta^{\mathcal{I}}$, there is a materialization $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $H_d^{\mathcal{J}} = I_d$ for all $d \in \Delta^{\mathcal{I}}$.*

Now, the TBox $\mathcal{T}_{\mathcal{I}}$ for $\mathsf{CSP}(\mathcal{I})$ in signature Σ from Theorem 20 is $\mathcal{T}$ extended with the following CIs:

$$\begin{array}{rl} \top \sqsubseteq \bigsqcup_{d \in \Delta^{\mathcal{I}}} H_d & \\ H_d \sqcap H_e \sqsubseteq \bot & \text{for all } d, e \in \Delta^{\mathcal{I}}, d \neq e \\ H_d \sqcap \exists r.H_e \sqsubseteq \bot & \text{for all } d, e \in \Delta^{\mathcal{I}}, r \in \Sigma, (d,e) \notin r^{\mathcal{I}} \\ H_d \sqcap A \sqsubseteq \bot & \text{for all } d \in \Delta^{\mathcal{I}}, A \in \Sigma, d \notin A^{\mathcal{I}}. \end{array}$$

Based on Lemma 21, it can be verified that $\mathcal{T}_{\mathcal{I}}$ satisfies Conditions 1 and 2 of Theorem 20. For Point 2, we show that for all Boolean PEQs q and ABoxes $\mathcal{A}$, $(\mathcal{T}_{\mathcal{I}}, \mathcal{A}) \models q$ iff $(\mathcal{T}, \mathcal{A}) \models q$ or not $\mathsf{Hom}(\mathcal{I}_{\mathcal{A}}^{\Sigma}, \mathcal{I})$ where $\mathcal{I}_{\mathcal{A}}^{\Sigma}$ is the restriction of $\mathcal{I}_{\mathcal{A}}$ to signature Σ and with $\mathsf{Ind}(\mathcal{I}_{\mathcal{A}}^{\Sigma}) = \emptyset$. Moreover, it is not hard to see that $\mathcal{T}$ is unraveling tolerant, thus $(\mathcal{T}, \mathcal{A}) \models q$ is in PTIME.

We now come to the conversion of an $\mathcal{ALCI}$-TBox and query q into a CSP. We start with considering Boolean CQs of the form $\exists x.C(x)$ with C an $\mathcal{ELI}$-concept, which is not strong enough to obtain the desired dichotomy result, but serves as a warmup that is conceptually cleaner than the version for ELIQs that we present afterwards. We use $\mathsf{sig}(\mathcal{T})$ to denote the signature of the TBox $\mathcal{T}$, and likewise for a CQ q.

Theorem 22. *Let $\mathcal{T}$ be an $\mathcal{ALCI}$-TBox, $q = \exists x.C(x)$ with C an $\mathcal{ELI}$-concept, and $\Sigma = \mathsf{sig}(\mathcal{T}) \cup \mathsf{sig}(q)$. Then one can construct (in time exponential in $|\mathcal{T}| + |C|$) a Σ-interpretation $\mathcal{I}_{\mathcal{T},q}$ such that for all ABoxes $\mathcal{A}$:*

$$(\mathsf{HomDual}) \quad (\mathcal{T}, \mathcal{A}) \models q \text{ iff not } \mathsf{Hom}(\mathcal{I}_{\mathcal{A}}^{\Sigma}, \mathcal{I}_{\mathcal{T},q})$$

Proof.(sketch) The interpretation $\mathcal{I}_{\mathcal{T},q}$ can be obtained using a standard type-based construction. We use the sets $\mathsf{cl}(\mathcal{T}, C)$, $\mathsf{tp}(\mathcal{T}, C)$, and the relation $\rightsquigarrow_r$ between types as defined in the proof of Theorem 16. A $\mathcal{T}$*-type t that omits q* is an element of $\mathsf{tp}(\mathcal{T}, C)$ that is satisfiable in a model $\mathcal{J}$ of $\mathcal{T}$ with $C^{\mathcal{J}} = \emptyset$. Then $\Delta^{\mathcal{I}_{\mathcal{T},q}}$ is the set of all $\mathcal{T}$-types that omit q, $t \in A^{\mathcal{I}_{\mathcal{T},q}}$ iff $A \in t$, for all $A \in \Sigma$, and $(t, t') \in r^{\mathcal{I}_{\mathcal{T},q}}$ iff $t \rightsquigarrow_r t'$, for all $r \in \Sigma$. It is shown in the long version that condition (HomDual) is satisfied. A Pratt-style type elimination algorithm can be used to construct $\mathcal{I}_{\mathcal{T},q}$ in exponential time (Pratt 1979). ❑

Example 23. Let $\mathcal{T} = \{A \sqsubseteq \forall r.B\}$ and define $q = \exists x.B(x)$. Then $\mathcal{I}_{\mathcal{T},q}$ is defined, up to isomorphism, by $\Delta^{\mathcal{I}_{\mathcal{T},q}} = \{a, b, c\}$, $A^{\mathcal{I}_{\mathcal{T},q}} = \{b\}$, $B^{\mathcal{I}_{\mathcal{T},q}} = \emptyset$, and $r^{\mathcal{I}_{\mathcal{T},q}} = \{(a,a), (a,b), (a,c)\}$. ⊣

For ELIQs, the conversion of a TBox and query into a CSP is similar to the construction above, but employs a concept name P that represents the individual name used in the ELIQ.

Theorem 24. *Let $\mathcal{T}$ be an $\mathcal{ALCI}$-TBox, $C(a)$ an ELIQ and $\Sigma = \mathsf{sig}(\mathcal{T}) \cup \mathsf{sig}(C) \cup \{P\}$, where P is a fresh concept name. Then one can construct (in time exponential in $|\mathcal{T}| + |C|$) a Σ-interpretation $\mathcal{I}_{\mathcal{T},q}$ such that for all ABoxes $\mathcal{A}$:*

1. *$(\mathcal{T}, \mathcal{A}) \models C(a)$ iff not $\mathsf{Hom}(\mathcal{I}_{\mathcal{A}'}^{\Sigma}, \mathcal{I}_{\mathcal{T},q})$, where $\mathcal{A}'$ is obtained from $\mathcal{A}$ by adding $P(a)$ and removing all other assertions that use P;*
2. *$(\mathcal{T}, \mathcal{A}) \models \exists x.(P(x) \wedge C(x))$ iff not $\mathsf{Hom}(\mathcal{I}_{\mathcal{A}}^{\Sigma}, \mathcal{I}_{\mathcal{T},q})$.*

As a consequence of Theorems 20 and 24, we obtain:

Theorem 25. *There is a dichotomy between* PTIME *and* CONP *for CQ-answering w.r.t. $\mathcal{ALC}$-TBoxes if and only if the Feder-Vardi conjecture is true.*

The same is true for $\mathcal{ALCI}$-TBoxes, for ELIQs, and PEQs. For $\mathcal{ALC}$-TBoxes, it additionally holds for ELQs.

Proof. Let $\mathsf{CSP}(\mathcal{I})$ be an NP-intermediate CSP, i.e., a CSP that is neither in PTIME nor NP-hard. Take the TBox $\mathcal{T}_{\mathcal{I}}$ from Theorem 20. By Point 1 of that theorem (and the mentioned reduction of ABox consistency to the complement of ELQ-answering), CQ-answering w.r.t. $\mathcal{T}$ is not in PTIME. By Point 2, CQ-answering w.r.t. $\mathcal{T}$ is not CONP-hard.

Conversely, let $\mathcal{T}$ be an $\mathcal{ALC}$-TBox for which CQ-answering w.r.t. $\mathcal{T}$ is neither in PTIME nor CONP-hard. Then by Theorem 4 and since every ELIQ is a CQ, the same holds for ELIQ-answering w.r.t. $\mathcal{T}$. It follows that there is concrete ELIQ q such that answering q w.r.t. $\mathcal{T}$ is CONP-intermediate. Let $\mathcal{I}_{\mathcal{T},q}$ be the interpretation constructed in Point 1 of Theorem 24. By Point 1 of that theorem, $\mathsf{CSP}(\mathcal{I})$ is not in PTIME; by Point 2, it is not NP-hard. ❑

The construction underlying Theorem 24 cannot be generalized from ELIQs to CQs. To discuss this further, let us consider the simpler 'warmup' Theorem 22 instead. We show that it is impossible to construct an interpretation $\mathcal{I}_{\mathcal{T},q}$ which satisfies (HomDual) for Boolean CQs that are not of the simple form $\exists x.C(x)$. This is true even when the TBox is*empty*. It is thus crucial to use ELIQs even when proving the dichotomy result for CQs and PEQs. The following theorem states this more formally.

Theorem 26. *Let q be a Boolean CQ without individual names, $\mathsf{sig}(q) = \Sigma$, and $\mathcal{T}_\emptyset$ the empty TBox. Then there is a Σ-interpretation $\mathcal{I}_{q,\mathcal{T}_\emptyset}$ that satisfies (*HomDual*) iff q is logically equivalent to a CQ of the form $\exists x.C(x)$ with C an $\mathcal{ELI}$-concept.*

Proof. This is a consequence of results on *homomorphism dualities* (Nesetril and Tardif 2000), the problem of constructing, for a given Σ-interpretation $\mathcal{I}$, a Σ-interpretation $\overline{\mathcal{I}}$ such that the following duality holds for all Σ-interpretations $\mathcal{J}$:

$$\mathsf{Hom}(\mathcal{I}, \mathcal{J}) \text{ iff not } \mathsf{Hom}(\mathcal{J}, \overline{\mathcal{I}}).$$

By (Nesetril and Tardif 2000; Nesetril 2009), such an $\overline{\mathcal{I}}$ exists iff the undirected graph induced by $\mathcal{I}$ is a tree. It remains to observe that for any Boolean CQ q without individual names and all Σ-interpretations $\mathcal{J}$, we have $\mathcal{A}_{\mathcal{J}} \models q$ iff $\mathsf{Hom}(\mathcal{I}_q, \mathcal{J})$, where $\mathcal{I}_q$ is the interpretation with $\Delta^{\mathcal{I}_q}$ the variables in q and in which $x \in A^{\mathcal{I}_q}$ (resp. $(x,y) \in r^{\mathcal{I}_q}$) if $A(x)$ (resp. $r(x,y)$) is a conjunct of q. ❑

Interestingly, (Nesetril 2009) presents five constructions of $\overline{\mathcal{I}}$, one of which resembles our type elimination procedure (but, of course, without taking into account TBoxes).

7 Non-Dichotomy in $\mathcal{ALCF}$

We show that the complexity landscape for query answering w.r.t. $\mathcal{ALCF}$-TBoxes is much richer than for $\mathcal{ALC}$ and $\mathcal{ALCI}$. In particular, we show that (i) for CQ-answering w.r.t. $\mathcal{ALCF}$-TBoxes, there is no dichotomy between PTIME and CONP unless PTIME = NP; and (ii) it is undecidable whether CQ-answering is in PTIME for a given $\mathcal{ALCF}$-TBox, and likewise for CONP-hardness and materializability. Point (i) is a consequence of the following, much stronger, result.

Theorem 27. *For every language $L \in$ CONP, there is an $\mathcal{ALCF}$-TBox $\mathcal{T}$ and ELIQ $\mathsf{rej}(a)$, rej a concept name, such that the following holds:*

1. *there is a polynomial reduction of L to answering $\mathsf{rej}(a)$ w.r.t. $\mathcal{T}$;*
2. *for every ELIQ q, answering q w.r.t. $\mathcal{T}$ is polynomially reducible to L.*

We now use Theorem 27 to establish Point (i) from above. Assume to the contrary of what is to be shown that for every $\mathcal{ALCF}$-TBox $\mathcal{T}$, CQ answering w.r.t. $\mathcal{T}$ is in PTIME or CONP-hard. By Ladner's Theorem, there is a CONP-intermediate language L. Let $\mathcal{T}$ be the TBox from Theorem 27. By Point 1 of the theorem, CQ-answering w.r.t. $\mathcal{T}$ is not in PTIME. Thus it must be CONP-hard. By Theorem 4 and since a dichotomy for CQ-answering w.r.t. $\mathcal{T}$ also implies a dichotomy for ELIQ-answering w.r.t. $\mathcal{T}$, ELIQ-answering w.r.t. $\mathcal{T}$ is also CONP-hard. By Point 2 of Theorem 27, this is impossible.

The proof of Theorem 27 combines the 'hidden' concepts H_d from the proof of Theorem 20 with a modification of the TBox constructed in (Baader et al. 2010) to prove the undecidability of *query emptiness* in $\mathcal{ALCF}$. Using a similar strategy, we establish the undecidability results announced as Point (ii) above, summarized by the following theorem.

Theorem 28. *For $\mathcal{ALCF}$-TBoxes $\mathcal{T}$, the following problems are undecidable (Points 1 and 2 are subject to the side condition that* PTIME $\neq$ NP*):*

1. *CQ-answering w.r.t. $\mathcal{T}$ is in* PTIME*;*
2. *CQ-answering w.r.t. $\mathcal{T}$ is* CONP*-hard;*
3. *$\mathcal{T}$ is materializable.*

8 Future Work

Much work remains to be done in order to fully accomplish the general research goal set out in the introduction. We propose four interesting directions.

(1) It would be interesting to consider additional complexity classes such as LOGSPACE, NLOGSPACE, and AC^0. The latter is particularly relevant in the context of FO-rewritability and the implementation of query answering using standard relational database systems, see (Calvanese et al. 2007) for details. Note that even for TBoxes of depth one, the complexity landscape is still rich. Relevant results can be found in (Calvanese et al. 2006): (i) there are $\mathcal{EL}$-TBoxes of depth one for which CQ-answering is PTIME-complete; and (ii) CQ-answering w.r.t. the $\mathcal{EL}$-TBox $\{\exists r.A \sqsubseteq A\}$, which encodes reachability in directed graphs, is NLOGSPACE-complete. We add that CQ-answering w.r.t. the Horn-$\mathcal{ALC}$-TBox $\{\exists r.A \sqsubseteq A, A \sqsubseteq \forall r.A\}$ corresponds to reachability in undirected graphs and can be shown to be LOGSPACE-complete. Also note that every DL-Lite TBox is of depth one, which gives a whole class of TBoxes for which CQ-answering is in AC^0.

(2) We conjecture that for a given $\mathcal{ALCFI}$-TBox of depth one, it is decidable whether CQ-answering is in PTIME, NLOGSPACE, LOGSPACE, and AC^0. A first step towards establishing these result is the observation from (Lutz and Wolter 2011) that FO-rewritability of CQ-answering, which is very closely related to CQ-answering being in AC^0, is decidable for $\mathcal{ALCFI}$-TBoxes of depth one. As an encouraging example of how results from the CSP world can be employed to obtain significant insights into CQ-answering, we note that for $\mathcal{ALCI}$-TBoxes, one can establish this result by using Theorem 24 and the fact that deciding FO-definability of a a CSP is in NP (Larose, Loten, and Tardif 2007). This yields a NEXPTIME upper bound due to the exponential size of the template constructed in the proof of Theorem 24.

(3) To better understand the complexity of TBoxes whose depth is larger than one, it would be interesting to generalize the notion of unraveling tolerance without leaving PTIME. In the CSP world, the corresponding notions of arc consistency and tree obstructions have both been significantly generalized, for example to structures of bounded treewidth (Bulatov, Krokhin, and Larose 2008).

(4) Alternatively to classifying the complexity of TBoxes while quantifying over *all* queries as in our Definition 2, one could also consider pairs $(\mathcal{T}, q)$ and classify the complexity of answering q w.r.t. $\mathcal{T}$, for all such pairs. This setup is significantly different from the one considered in this paper, and will required different techniques.

Acknowledgments. C. Lutz was supported by the DFG SFB/TR 8 "Spatial Cognition".

References

Allender, E.; Bauland, M.; Immerman, N.; Schnoor, H.; and Vollmer, H. 2005. The complexity of satisfiability problems: Refining Schaefer's theorem. In *MFCS*, 71–82.

Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The DL-Lite family and relations. *J. Artif. Intell. Res. (JAIR)* 36:1–69.

Baader, F.; Calvanese, D.; McGuiness, D.; Nardi, D.; and Patel-Schneider, P. 2003. *The Description Logic Handbook: Theory, implementation and applications*. Cambridge University Press.

Baader, F.; Bienvenu, M.; Lutz, C.; and Wolter, F. 2010. Query and predicate emptiness in description logics. In *KR*.

Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ envelope. In *IJCAI*, 364–369. Professional Book Center.

Baget, J.-F.; Mugnier, M.-L.; Rudolph, S.; and Thomazo, M. 2011. Walking the complexity lines for generalized guarded existential rules. In *IJCAI*, 712–717.

Barany, V.; Gottlob, G.; and Otto, M. 2010. Querying the guarded fragment. In *LICS*, 1–10.

Bulatov, A. A.; Jeavons, P.; and Krokhin, A. A. 2005. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.* 34(3):720–742.

Bulatov, A. A.; Krokhin, A. A.; and Larose, B. 2008. Dualities for constraint satisfaction problems. In *Complexity of Constraints*, 93–124.

Bulatov, A. A. 2002. A dichotomy theorem for constraints on a three-element set. In *FOCS*, 649–658.

Bulatov, A. A. 2011. On the CSP dichotomy conjecture. In *CSR*, 331–344.

Calvanese, D.; Giacomo, G. D.; Lenzerini, M.; and Vardi, M. Y. 2000. View-based query processing and constraint satisfaction. In *LICS*, 361–371.

Calvanese, D.; Giacomo, G. D.; Lenzerini, M.; and Vardi, M. Y. 2003a. Reasoning on regular path queries. *SIGMOD Record* 32(4):83–92.

Calvanese, D.; Giacomo, G. D.; Lenzerini, M.; and Vardi, M. Y. 2003b. View-based query containment. In *PODS*, 56–67.

Calvanese, D.; Giacomo, G. D.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2006. Data complexity of query answering in description logics. In *KR*, 260–270.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Autom. Reasoning* 39(3):385–429.

Chang, C. C., and Keisler, H. J. 1990. *Model Theory*, volume 73 of *Studies in Logic and the Foundations of Mathematics*. Elsevier.

Cosmadakis, S. S.; Gaifman, H.; Kanellakis, P. C.; and Vardi, M. Y. 1988. Decidable optimization problems for database logic programs (preliminary report). In *STOC*, 477–490.

Dechter, R. 2003. *Constraint processing*. Elsevier Morgan Kaufmann.

Dolby, J.; Fokoue, A.; Kalyanpur, A.; Ma, L.; Schonberg, E.; Srinivas, K.; and Sun, X. 2008. Scalable grounded conjunctive query evaluation over large and expressive knowledge bases. In *ISWC*, 403–418.

Eiter, T.; Gottlob, G.; Ortiz, M.; and Simkus, M. 2008. Query answering in the description logic Horn-$\mathcal{SHIQ}$. In *JELIA*, 166–179.

Feder, T., and Vardi, M. Y. 1993. Monotone monadic SNP and constraint satisfaction. In *STOC*, 612–622.

Gaifman, H.; Mairson, H. G.; Sagiv, Y.; and Vardi, M. Y. 1987. Undecidable optimization problems for database logic programs. In *LICS*, 106–115.

Glimm, B.; Lutz, C.; Horrocks, I.; and Sattler, U. 2008. Conjunctive query answering for the description logic $\mathcal{SHIQ}$. *JAIR* 31:157–204.

Hell, P., and Nesetril, J. 1990. On the complexity of h-coloring. *J. Comb. Theory, Ser. B* 48(1):92–110.

Hustadt, U.; Motik, B.; and Sattler, U. 2007. Reasoning in description logics by a reduction to disjunctive datalog. *J. Autom. Reasoning* 39(3):351–384.

Kazakov, Y. 2009. Consequence-driven reasoning for Horn-$\mathcal{SHIQ}$ ontologies. In Boutilier, C., ed., *IJCAI*, 2040–2045.

Kontchakov, R.; Lutz, C.; Toman, D.; Wolter, F.; and Zakharyaschev, M. 2010. The combined approach to query answering in DL-Lite. In *KR*.

Krisnadhi, A., and Lutz, C. 2007. Data complexity in the $\mathcal{EL}$ family of description logics. In *LPAR*, 333–347.

Krokhin, A. A. 2010. Tree dualities for constraint satisfaction. In *CSL*, 32–33.

Krötzsch, M.; Rudolph, S.; and Hitzler, P. 2007. Complexity boundaries for Horn description logics. In *AAAI*, 452–457.

Krötzsch, M. 2010. Efficient inferencing for OWL EL. In *JELIA*, 234–246.

Kun, G., and Szegedy, M. 2009. A new line of attack on the dichotomy conjecture. In *STOC*, 725–734.

Ladner, R. E. 1975. On the structure of polynomial time reducibility. *J. ACM* 22(1):155–171.

Larose, B.; Loten, C.; and Tardif, C. 2007. A characterisation of first-order constraint satisfaction problems. *Logical Methods in Computer Science* 3(4).

Lutz, C., and Wolter, F. 2011. Non-uniform data complexity of query answering in description logics. In *Description Logics*.

Lutz, C.; Piro, R.; and Wolter, F. 2011. Description logic tboxes: Model-theoretic characterizations and rewritability. In *IJCAI*.

Lutz, C.; Toman, D.; and Wolter, F. 2009. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *IJCAI*, 2070–2075.

Makowsky, J. A. 1987. Why Horn formulas matter in computer science: Initial structures and generic examples. *J. Comput. Syst. Sci.* 34(2/3):266–292.

Malcev, A. 1971. *The metamathematics of algebraic systems, collected papers:1936-1967*. North-Holland.

Meseguer, J., and Goguen, J. 1985. Initiality, induction, and computability. In *Algebraic Methods in Semantics*. Cambridge University Press. 459–541.

Nesetril, J., and Tardif, C. 2000. Duality theorems for finite structures (characterising gaps and good characterisations). *J. Comb. Theory, Ser. B* 80(1):80–97.

Nesetril, J. 2009. Many facets of dualities. In *Research Trends in Combinatorial Optimization*. 285–302.

Otto, M.; Blumensath, A.; and Weyer, M. 2010. Decidability results for the boundedness problem. Technical report, TU Darmstadt.

Poggi, A.; Lembo, D.; Calvanese, D.; Giacomo, G. D.; Lenzerini, M.; and Rosati, R. 2008. Linking data to ontologies. *J. Data Semantics* 10:133–173.

Pratt, V. R. 1979. Models of program logics. In *FOCS*, 115–122.

Schaefer, T. J. 1978. The complexity of satisfiability problems. In *STOC*, 216–226.

Schaerf, A. 1993. On the complexity of the instance checking problem in concept languages with existential quantification. *J. of Intel. Inf. Systems* 2:265–278.

A Proofs for Section 2

In this section, we prove Theorem 4. Note that in the proofs of Theorems 9 and 11 we do not use Theorem 4. Thus, we can (and will) employ them in the proof below. We formulate Theorem 4 again.

Theorem 4 For all $\mathcal{ALCFI}$-TBoxes $\mathcal{T}$, the following are equivalent:

1. CQ-answering w.r.t. $\mathcal{T}$ is in PTIME iff PEQ-answering w.r.t. $\mathcal{T}$ is in PTIME iff ELIQ-answering w.r.t. $\mathcal{T}$ is in PTIME;
2. $\mathcal{T}$ is FO-rewritable for CQ iff it is FO-rewritable for PEQ iff it is FO-rewritable for ELIQ.

If $\mathcal{T}$ is an $\mathcal{ALCF}$-TBoxes, then we can replace ELIQ in Points 1 and 2 with ELQ.

We start the proof with the observation that the implications

- If PEQ-answering w.r.t. $\mathcal{T}$ is in PTIME, then CQ-answering w.r.t. $\mathcal{T}$ is in PTIME;
- If CQ answering w.r.t. $\mathcal{T}$ is in PTIME, then ELIQ-answering w.r.t. $\mathcal{T}$ is in PTIME;
- If $\mathcal{T}$ is FO-rewritable for PEQ, then $\mathcal{T}$ is FO-rewritable for CQ;
- If $\mathcal{T}$ is FO-rewritable for CQ, then $\mathcal{T}$ is FO-rewritable for ELIQ

are trivial, by the obvious inclusions between the sets of queries considered. For the proofs of the other directions we can assume that $\mathcal{T}$ is materializable: otherwise, by Theorems 9 and 11, ELIQ-answering w.r.t. $\mathcal{T}$ is CONP-hard and the implications are trivial.

For materializable $\mathcal{T}$, the implications

- If CQ-answering w.r.t. $\mathcal{T}$ is in PTIME, then PEQ-answering w.r.t. $\mathcal{T}$ is in PTIME;
- If $\mathcal{T}$ is FO-rewritable for CQ, then $\mathcal{T}$ is FO-rewritable for PEQ;

are obvious since the evaluation of a disjunction in an interpretation reduces to evaluating all its disjuncts. Thus, it remains to show the following two implications:

1. If ELIQ answering w.r.t. $\mathcal{T}$ is in PTIME, then CQ-answering w.r.t. $\mathcal{T}$ is in PTIME;
2. If $\mathcal{T}$ is FO-rewritable for ELIQ, then $\mathcal{T}$ is FO-rewritable for CQ.

To show these implications, we introduce some notation and a lemma. For a sequence $\vec{r} = r_1 \cdots r_n$ of roles, we set $\exists \vec{r}.C = \exists r_1. \cdots \exists r_n.C$. In an interpretation $\mathcal{I}$, the distance $\mathsf{dist}_\mathcal{I}(d_1, d_2)$ between $d_1, d_2 \in \Delta^\mathcal{I}$ is the minimal n such that there are $d_1 = e_0, \ldots, e_n = d_2$ and roles $r_1, \ldots, r_n$ with $(d_i, d_{i+1}) \in r_{i+1}^\mathcal{I}$ for $i < n$.

Lemma 29. *Let C be an $\mathcal{ELI}$-concept and assume that $(\mathcal{T}, \mathcal{A}) \models \exists v.C(v)$. If $\mathcal{T}$ is materializable, then there exists a sequence of roles $\vec{r} = r_1 \cdots r_n$ of length $n \leq 2^{(2(|\mathcal{T}|+|C|)} \times 2|\mathcal{T}||C| + 1$ such that there exists $a \in \mathsf{Ind}(\mathcal{A})$ with $(\mathcal{T}, \mathcal{A}) \models \exists \vec{r}.C(a)$.*

Proof. Let $\mathcal{I}$ be a PEQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. We may assume that $\mathcal{I}$ is i-unfolded. From $(\mathcal{T},\mathcal{A}) \models \exists v.C(v)$, we obtain $C^{\mathcal{I}} \neq \emptyset$. Choose $d \in C^{\mathcal{I}}$ and $a \in \mathsf{Ind}(\mathcal{A})$ such that $n := \mathsf{dist}_{\mathcal{I}}(d, a^{\mathcal{I}})$ is minimal. (We assume, for simplicity, that there is only one such d. The argument is easily generalized.) Assume $n > 2^{(2(|\mathcal{T}|+|C|)} \times 2|\mathcal{T}||C| + 1$.

Let $a^{\mathcal{I}} = d_0, \ldots, d_n = d$ and $(d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}}$ for $i < n$. Let $\mathsf{sub}(\mathcal{T}, C)$ denote the closure under single negation of the set of subconcepts of concepts in $\mathcal{T}$ and C. Set

$$t^{\mathcal{I}}(e) = \{D \in \mathsf{sub}(\mathcal{T}, C) \mid e \in D^{\mathcal{I}}\}.$$

As $n > 2^{(2(|\mathcal{T}|+|C|)} \times 2|\mathcal{T}||C| + 1$, there exist d_i and d_{i+j} with $j > 1$ and $i + j < n$ such that

$$t^{\mathcal{I}}(d_i) = t^{\mathcal{I}}(d_{i+j}), \quad t^{\mathcal{I}}(d_{i+1}) = t^{\mathcal{I}}(d_{i+j+1}), \quad r_{i+1} = r_{i+j+1}$$

Now replace in $\mathcal{I}$ the interpretation induced by the subtree generated by d_{i+j+1} by the interpretation induced by the subtree generated by d_{i+1} and denote the resulting interpretation by $\mathcal{J}$. $\mathcal{J}$ is still a model of $(\mathcal{T},\mathcal{A})$. But now $\mathcal{J} \not\models \exists \vec{r}.C(a)$. We have derived a contradiction since $a^{\mathcal{I}} \in (\exists \vec{r}.C)^{\mathcal{I}}$ and therefore, since $\mathcal{I}$ is a minimal model of $(\mathcal{T},\mathcal{A})$, $(\mathcal{T},\mathcal{A}) \models (\exists \vec{r}.C)(a)$. ❏

Let $q(\vec{x}) = \exists \vec{y}.\varphi(\vec{x}, \vec{y})$ be a CQ with $\vec{x} = x_1, \ldots, x_n$ and $\vec{y} = y_1, \ldots, y_m$. We regard φ as a set of atoms. A *splitting* $S = (Y, \sim, f)$ of $q(\vec{x})$ consists of a subset Y of $\vec{y}$, an equivalence relation $\sim$ on $\mathsf{Ind}(q) \cup \vec{x} \cup Y$ and a mapping f

$$f : \{u_\sim \mid u \in \mathsf{Ind}(q) \cup \vec{x} \cup Y\} \rightarrow 2^{\vec{y} \setminus Y}$$

(we denote by $u_\sim$ the equivalence class of u w.r.t. $\sim$) such that

- for every $y \in \vec{y} \setminus Y$ there exists u with $y \in f(u_\sim)$;
- $f(u_\sim) \cap f(v_\sim) = \emptyset$ whenever $u_\sim \neq v_\sim$.
- if $r(t,t') \in \varphi$ or $r(t',t) \in \varphi$ and $t \in f(u_\sim)$, then $t' \in u_\sim$ or $t' \in f(u_\sim)$.

Let U_S denote the set of all equivalence classes w.r.t. $\sim$. Thus, if $(Y, \sim, f)$ is a splitting of $q(\vec{x})$, we can form

- φ_Y consisting of all $A(t)$ with $t \in \mathsf{Ind}(q) \cup \vec{x} \cup Y$ and all $r(t,t')$ with $t,t' \in \mathsf{Ind}(q) \cup \vec{x} \cup Y$;
- for every $u_\sim \in U_S$, φ_u consisting of all $A(t)$ and $r(t,t')$ with $t,t' \in u_\sim \cup f(u_\sim)$.

Intuitively, splittings describe potential assignments π for the variables in $\vec{x}, \vec{y}$ in an unfolded CQ-materialization $\mathcal{I}$ of $(\mathcal{T},\mathcal{A})$ in which

- all $v \in u_\sim$ receive the same value $\pi(v)$ and this value is in $\mathsf{Ind}(\mathcal{A})$;
- all $y \in f(u_\sim)$ receive values $\pi(y)$ in the "anonymous" tree generated by $\pi(u)$.

Using Lemma 29 (for those y that are not reachable in φ from any member of $\mathsf{Ind}(\mathcal{A}) \cup \vec{x} \cup Y$) one can easily construct, for every $u_\sim \in U_S$ a disjunction $D_u = \bigvee_{i \in I_u} C_i$ of $\mathcal{ELI}$-concepts such that for all CQ-materializations $\mathcal{I}$ of some $(\mathcal{T},\mathcal{A})$ and all $a \in \mathsf{Ind}(\mathcal{A})$, (1.) implies (2.) and (2.) implies (3.), where

1. there exists an assignment π in $\mathcal{I}$ with
   - $\pi(u) = \pi(u') = a^{\mathcal{I}}$ for all $u' \in u_\sim$
   - $\pi(x)$ in the anonymous subtree generated by $a^{\mathcal{I}}$ for all $x \in f(u_\sim)$
   - $\mathcal{I} \models_\pi \varphi_u$.
2. $a^{\mathcal{I}} \in D_u^{\mathcal{I}}$;
3. there exists an assignment π in $\mathcal{I}$ with
   - $\pi(u) = \pi(u') = a^{\mathcal{I}}$ for all $u' \in u_\sim$
   - $\mathcal{I} \models_\pi \varphi_u$.

For every splitting $S = (Y, \sim, f)$ of $\varphi(\vec{x})$, set

$$\chi_S = \varphi_Y \wedge \bigwedge_{u_\sim \in U_S} \bigwedge_{t,t' \in u_\sim} (t = t') \wedge \bigwedge_{u_\sim \in U_S} D_u.$$

To prove the implication (2.), assume that $\mathcal{T}$ is FO-rewritable for ELIQ. By materializability, $\mathcal{T}$ is FO-rewritable for unions of ELIQs. For every $u_\sim \in U_S$, let χ_u be a FOQ with

$$\mathcal{I}_{\mathcal{A}} \models \chi_u[a] \quad \Leftrightarrow \quad (\mathcal{T},\mathcal{A}) \models D_u(a).$$

for all $a \in \mathsf{Ind}(\mathcal{A})$. Let χ_S^* be the FOQ resulting from χ_S by replacing every D_u with χ_u. Then it is readily checked that

$$\mathcal{I}_{\mathcal{A}} \models \bigvee_{S \text{ is a splitting of } q(\vec{x})} \exists \vec{y}.\chi_S^*[\vec{a}] \quad \Leftrightarrow \quad (\mathcal{T},\mathcal{A}) \models q(\vec{a})$$

for all $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$. Thus, $\mathcal{T}$ is FO-rewritable for CQ.

We come to implication (1.). Assume that ELIQ-answering w.r.t. $\mathcal{T}$ is in PTIME. By materializability, unions of ELIQs can be answered w.r.t. $\mathcal{T}$ in PTIME. We can evaluate a CQ $q(\vec{x})$ in polynomial time as follows: to decide whether $(\mathcal{T},\mathcal{A}) \models q(\vec{a})$ for a given $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, go through all splittings $S = (Y, \sim, f)$ of $q(\vec{x})$ and all assignments $\pi(y) \in \mathsf{Ind}(\mathcal{A})$ for $y \in Y$ and check

$$\mathcal{I}_{\mathcal{A}} \models_\pi \varphi_Y \wedge \bigwedge_{u_\sim \in U_S} \bigwedge_{t,t' \in u_\sim} (t = t')[\vec{a}]$$

and

$$(\mathcal{T},\mathcal{A}) \models \bigwedge_{u_\sim \in U_S} D_u(\pi(u)).$$

If both hold for at least one pair S, π, then $(\mathcal{T},\mathcal{A}) \models q(\vec{a})$; otherwise $(\mathcal{T},\mathcal{A}) \not\models q(\vec{a})$. Both conditions can be checked in polynomial time.

B Proofs for Section 3

We introduce some notions and notations. For any interpretation $\mathcal{I}$, we define its *i-unfolding* $\mathcal{I}^*$. The domain $\Delta^{\mathcal{I}^*}$ of $\mathcal{I}^*$ consists of all words $d_0 r_1 \ldots r_n d_n$ with $n \geq 0$, $d_i \in \Delta^{\mathcal{I}}$, and r_i (possibly inverse) roles such that

- there exists $a \in \mathsf{Ind}(\mathcal{A})$ with $d_0 = a^{\mathcal{I}}$;
- for $0 < i \leq n$ there does not exists $a \in \mathsf{Ind}(\mathcal{A})$ such that $d_i = a^{\mathcal{I}}$;
- for $0 \leq i < n$: $(d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}}$ and if $r_i^- = r_{i+1}$, then $d_{i-1} \neq d_{i+1}$.

For $d_0 \cdots d_n \in \Delta^{\mathcal{I}^*}$, we set $\mathsf{tail}(d_0 \cdots d_n) = d_n$. Now set

- for all $A \in \mathsf{N_C}$:

$$A^{\mathcal{I}^*} = \{w \in \Delta^{\mathcal{I}^*} \mid \mathsf{tail}(w) \in A^{\mathcal{I}}\}$$

- for all $r \in \mathsf{N_R}$:

$$\begin{aligned} r^{\mathcal{I}^*} &= \{(\sigma, \sigma r d) \mid, \sigma, \sigma r d \in \Delta^{\mathcal{I}^*} \cup \\ &\quad \{(\sigma r^- d, \sigma) \mid \sigma, \sigma r^- d \in \Delta^{\mathcal{I}^*}\} \end{aligned}$$

- $\mathsf{Ind}(\mathcal{I}^*) = \mathsf{Ind}(\mathcal{I})$ and $a^{\mathcal{I}^*} = a^{\mathcal{I}}$, for all $a \in \mathsf{Ind}(\mathcal{I})$.

We call an interpretation $\mathcal{I}$ *i-unfolded* if it is isomorphic to its own i-unfolding. Clearly, every i-unfolding $\mathcal{I}^*$ of an interpretation $\mathcal{I}$ is i-unfolded.

For $\mathcal{ALCF}$-TBoxes it is not required to unfold along inverse roles. Thus, we define the domain $\Delta^{\mathcal{I}^+}$ of the *unfolding* $\mathcal{I}^+$ *of* $\mathcal{I}$ as the set of all words $d_0 r_1 \dots r_n d_n$ with $n \geq 0$, $d_i \in \Delta^{\mathcal{I}}$, and r_i *role names*. The definition of the interpretation of concept, role and individual names remains the same (but can be simplified). We call an interpretation $\mathcal{I}$ *unfolded* if it is isomorphic to its own unfolding. Every unfolding $\mathcal{I}^+$ of an interpretation $\mathcal{I}$ is unfolded.

Lemma 30. *Let $\mathcal{I}$ be an interpretation.*

- *$f(w) := \mathsf{tail}(w)$, $w \in \Delta^{\mathcal{I}^*}$, is a homomorphism from $\mathcal{I}^*$ to $\mathcal{I}$;*
- *$f(w) := \mathsf{tail}(w)$, $w \in \Delta^{\mathcal{I}^+}$, is a homomorphism from $\mathcal{I}^+$ to $\mathcal{I}$;*
- *for any interpretation $\mathcal{J}$, if there is an i-simulation between $\mathcal{I}$ and $\mathcal{J}$, then there is a homomorphism from $\mathcal{I}^*$ to $\mathcal{J}$;*
- *For any interpretation $\mathcal{J}$, if there is a simulation between $\mathcal{I}$ and $\mathcal{J}$, then there is a homomorphism from $\mathcal{I}^+$ to $\mathcal{J}$;*
- *If $\mathcal{I}$ is a model of $(\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ an $\mathcal{ALCFI}$-TBox, then $\mathcal{I}^*$ is a model of $(\mathcal{T}, \mathcal{A})$;*
- *If $\mathcal{I}$ is a model of $(\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ an $\mathcal{ALCF}$-TBox, then $\mathcal{I}^+$ is a model of $\mathcal{T}, \mathcal{A}$.*

We formulate Lemma 8 again.

Lemma 8 Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox and $\mathcal{A}$ and ABox. A model $\mathcal{I}$ of $\mathcal{T}, \mathcal{A}$ is

1. an ELIQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is i-sim-initial in $\mathsf{Mod}(\mathcal{I}, \mathcal{A})$;
2. a PEQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is hom-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$.

If $\mathcal{T}$ is a $\mathcal{ALCF}$-TBox, then $\mathcal{I}$ is an ELQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff it is sim-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$.

Proof. We apply Lemma 30.

(1) We consider the direction from left to right only. Let $\mathcal{I}$ be an ELIQ-materialization and $\mathcal{J} \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$. Assume first that $\mathcal{J}$ has finite outdegree. For $a \in \mathsf{Ind}(\mathcal{A})$, let $\mathcal{I}^a$ and $\mathcal{J}^a$ denote the interpretations obtained from $\mathcal{I}$ and $\mathcal{J}$ by setting $\mathsf{Ind}(\mathcal{I}^a) = \{a\}$ and $\mathsf{Ind}(\mathcal{J}^a) = \{a\}$, respectively. Thus, the only difference is that only the individual name a is interpreted. Using the condition that $\mathcal{J}$ has finite outdegree, one can readily show (1) $\Rightarrow$ (2), where

1. for all $\mathcal{ELI}$-concepts C: if $\mathcal{I} \models C(a)$, then $\mathcal{J} \models C(a)$;
2. there is an i-simulation S_a between $\mathcal{I}^a$ and $\mathcal{J}^a$.

Now, condition (1) holds for all $a \in \mathsf{Ind}(\mathcal{I})$ since $\mathcal{I}$ is an ELIQ materialization of $\mathcal{T}$ and $\mathcal{A}$. Thus $\bigcup_{a \in \mathsf{Ind}(\mathcal{A})} S_a$ is an i-simulation between $\mathcal{I}$ and $\mathcal{J}$, as required.

Now assume that $\mathcal{J}$ does not have finite outdegree. Construct the i-unfolding $\mathcal{J}^*$ of $\mathcal{J}$. From $\mathcal{J}^*$ we obtain an interpretation $\mathcal{J}^b$ of bounded outdegree by selective filtration as follows: let $S_0 = \mathsf{Ind}(\mathcal{A})$ and assume S_n has been defined. Then define S_{n+1} as the union of S_n and, for every $d \in S_n$ and $\exists r.D \in \mathsf{sub}(\mathcal{T})$ with $d \in (\exists r.D)^{\mathcal{J}^*}$, some witness $d' \in \Delta^{\mathcal{J}^*}$ with $(d, d') \in r^{\mathcal{J}^*}$ and $d' \in D^{\mathcal{J}^*}$ if no such d' exists already in S_n. Let $S = \bigcup_{i \geq 0} S_n$. Let $\mathcal{J}^b$ be the restriction of $\mathcal{J}^*$ to S. The outdegree of $\mathcal{J}^b$ is bounded by $|\mathsf{sub}(\mathcal{T})| + |\mathcal{A}|$, and, therefore, finite. By construction, $\mathcal{J}^b \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$. Since there is a homomorphism from $\mathcal{J}^*$ to $\mathcal{J}$, its restriction f to the domain of $\Delta^{\mathcal{J}^b}$ is a homomorphism from $\mathcal{J}^b$ to $\mathcal{J}$. Since $\mathcal{J}^b$ has finite outdegree, we find an i-simulation D between $\mathcal{I}$ and $\mathcal{J}^b$. The composition of D and f is the required i-simulation between $\mathcal{I}$ and $\mathcal{J}$.

(2) Let $\mathcal{I}$ be countable and generated. It is straightforward to check that if $\mathcal{I}$ is hom-initial, then it is a PEQ-materialization; obviously, if $\mathcal{I}$ is a PEQ-materialization, then it is a CQ-materialization. Thus, it remains to show that if $\mathcal{I}$ is a CQ-materialization, then it is hom-initial. Assume that $\mathcal{I}$ is a CQ-materialization and $\mathcal{J} \in \mathsf{Mod}(\mathcal{T}, \mathcal{A})$. We assume $\mathcal{J}$ has finite outdegree (the infinite outdegree case can be reduced to the finite outdegree case as in (1)). First, for every finite subset X of $\mathcal{I}$ we obtain from the condition that $\mathcal{I}$ is a CQ-materialization that there is a homomorphism h_X from the subinterpretation $\mathcal{I}_{\restriction X}$ of $\mathcal{I}$ induced by X into $\mathcal{J}$. Since $\mathcal{I}$ is countable, we can take an enumeration $d_1, \dots$ of $\Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$. Let $X_0, X_1, \dots$ be a sequence of finite subsets of $\Delta^{\mathcal{I}}$ such that

- $X_0 = \{a^{\mathcal{I}} \mid a \in \mathsf{Ind}(\mathcal{A})\}$
- $X_n \subseteq X_{n+1}$ for $n \geq 0$;
- $X_n \supseteq \mathsf{Ind}(\mathcal{A}) \cup \{d_1, \dots, d_n\}$, for $n \geq 0$;
- for all $d \in X_n$ there exists a path in X_n from some $a \in \mathsf{Ind}(\mathcal{A})$ to d.

Condition 4 can be satisfied since $\mathcal{I}$ is generated. Let h_{X_n} be a homomorphism from $\mathcal{I}_{X_n}$ to $\mathcal{J}$, for $n \geq 0$. We define the required homomorphism as the limit of a sequence of homomorphisms $f_0, f_1, \dots$. Let $f_0 = h_{X_0}$ and assume f_n has been defined. We assume that there is an infinite set $\mathcal{X} \subseteq \mathbb{N}$ such that for all $m \in \mathcal{X}$: if $d \in \mathsf{dom}(f_n)$, then $f_n(d) = h_{X_m}(d)$ for all $m \in \mathcal{X}$. Let i be minimal such that d_i is not in the domain of f_n (if no such i exists, we are done). Let X_m be the minimal m such that $d_i \in X_m$ and $X_m \in \mathcal{X}$. Let k be the length of the shortest path from some $a^{\mathcal{I}}$ with $a \in \mathsf{Ind}(\mathcal{A})$ to d_i in X_m. This is a path in any X_n with $n \geq m$. Thus, for all X_n, $n \geq m$, there exists a path of length $\leq k$ from $h_X(d_i)$ to some $a^{\mathcal{I}}$ in $\mathcal{J}$. Since $\mathcal{J}$ has finite outdegree, there exists an infinite subset $\mathcal{X}'$ of $\mathcal{X}$ such that $h_X(d_i) = h_Y(d_i)$ for all $X, Y \in \mathcal{X}'$. We now set $f_{n+1} = f_n \cup \{(d_i, e)$, where $h_X(d) = e$ for all $X \in \mathcal{X}'$. This finishes the construction of f_{n+1}.

We set $f = \bigcup_{n \geq 0} f_n$. By definition, f is a homomorphism, as required.

The claim for $\mathcal{ALCF}$-TBoxes is proved in the same way as (1). ❏

We show that the generatedness condition cannot be dropped: consider the TBox

$$\mathcal{T} = \{A \sqsubseteq \exists r.A, B \sqsubseteq A\}$$

It is readily seen that $\mathcal{T}$ is PEQ-materializable. Let $\mathcal{A} = \{B(a)\}$. The interpretation $\mathcal{I}$ with

- $\Delta^{\mathcal{I}} = \{a\} \cup \{1, 2 \ldots\}$;
- $A^{\mathcal{I}} = \Delta^{\mathcal{I}}$;
- $B^{\mathcal{I}} = \{a\}$;
- $r^{\mathcal{I}} = \{(a,1)\} \cup \{(n, n+1) \mid n \geq 1\}$

is hom-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$. However, the interpretation $\mathcal{I}'$ defined as the disjoint union of $\mathcal{I}$ and the interpretation $\mathcal{J}$ with

- $\Delta^{\mathcal{J}} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$;
- $r^{\mathcal{J}} = \{(n, n+1) \mid n \in \Delta^{\mathcal{J}}\}$;
- $A^{\mathcal{J}} = \Delta^{\mathcal{J}}$;
- $B^{\mathcal{J}} = \emptyset$

is a PEQ-materialization of $\mathcal{T}$ and $\mathcal{A}$, but it is not hom-initial as there is no homomorphism from $\mathcal{J}$ to $\mathcal{I}$.

Proof of Theorem 9 We apply Lemmas 8 and 30. For Points 1 and 2, assume that $\mathcal{I}$ is i-sim initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$. We have to show that there exists a model $\mathcal{I}'$ of $\mathcal{T}$ and $\mathcal{A}$ that is hom-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$. To construct $\mathcal{I}'$, let $\mathcal{I}_0$ be an at most countable and generated subinterpretation of $\mathcal{I}$ in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$. For example, one can take an elementary subinterpretation of $\mathcal{I}$ and then restrict its domain to the points reachable from $\{a^{\mathcal{I}} \mid a \in \mathsf{Ind}(\mathcal{A})\}$ in $\mathcal{I}$. Clearly, $\mathcal{I}_0$ is still i-sim initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$. Now, the unfolding $\mathcal{I}_0^*$ of $\mathcal{I}_0$ is hom-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$. The final Point of Theorem 9 is proved similarly.

Proof of Theorem 11

The proof is by reduction of 2+2-SAT, a variant of propositional satisfiability that was first introduced by Schaerf as a tool for establishing lower bounds for the data complexity of query answering in a DL context (Schaerf 1993). A *2+2 clause* is of the form $(p_1 \vee p_2 \vee \neg n_1 \vee \neg n_2)$, where each of p_1, p_2, n_1, n_2 is a propositional letter or a truth constant 0, 1. A *2+2 formula* is a finite conjunction of 2+2 clauses. Now, 2+2-SAT is the problem of deciding whether a given 2+2 formula is satisfiable. It is shown in (Schaerf 1993) that 2+2-SAT is NP-complete.

Theorem 11. If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ ($\mathcal{ALCF}$-TBox $\mathcal{T}$) is not materializable, then ELIQ-answering (ELQ-answering) is coNP-hard w.r.t. $\mathcal{T}$.

Proof. We first show that if an $\mathcal{ALCFI}$-$\mathcal{T}$ is not materializable, then Boolean UELIQ-answering w.r.t. $\mathcal{T}$ is coNP-hard, where a Boolean UELIQ is a disjunction $q_1 \vee \cdots \vee q_k$, with each q_i a Boolean ELIQ. We then sketch the modifications necessary to lift the result to Boolean ELIQ-answering w.r.t. $\mathcal{T}$.

Since $\mathcal{T}$ is not materializable, by Theorem 9 it does not have the disjunction property. Thus, there is an ABox $\mathcal{A}_\vee$ and ELIQs $C_0(a_0), \ldots, C_k(a_k)$ such that $\mathcal{T}, \mathcal{A}_\vee \models C_0(a_0) \vee \cdots \vee C_k(a_k)$, but $\mathcal{T}, \mathcal{A}_\vee \not\models C_i(a_i)$ for all $i \leq k$. Assume w.l.o.g. that this sequence is minimal, i.e., $\mathcal{T}, \mathcal{A}_\vee \not\models C_0(a_0) \vee \cdots \vee C_{i-1}(a_{i-1}) \vee C_{i+1}(a_{i+1}) \vee \cdots \vee C_k(a_k)$ for all $i \leq k$. By minimality, we clearly have that

(∗) for all $i \leq k$, there is a model $\mathcal{I}_i$ of $\mathcal{T}$ and $\mathcal{A}_\vee$ with $\mathcal{I} \models C_i(a_i)$ and $\mathcal{I} \not\models C_j(a_j)$ for all $j \neq i$.

We will use $\mathcal{A}_\vee$ and the sequence $C_0(a_0), \ldots, C_k(a_k)$ to generate truth values for variables in the input 2+2 formula.

Let $\varphi = c_0 \wedge \cdots \wedge c_n$ be a 2+2 formula in propositional letters $q_0, \ldots, q_m$, and let $c_i = p_{i,1} \vee p_{i,2} \vee \neg n_{i,1} \vee \neg n_{i,2}$ for all $i \leq n$. Our aim is to define an ABox $\mathcal{A}_\varphi$ and a Boolean UELIQ q such that φ is unsatisfiable iff $\mathcal{T}, \mathcal{A}_\varphi \models q$. To start, we represent the formula φ in the ABox $\mathcal{A}_\varphi$ as follows:

- the individual name f represents the formula φ;
- the individual names $c_0, \ldots, c_n$ represent the clauses of φ;
- the assertions $c(f, c_0), \ldots, c(f, c_n)$, associate f with its clauses, where c is a role name that does not occur in $\mathcal{T}$;
- the individual names $q_0, \ldots, q_m$ represent variables, and the individual names $0, 1$ represent truth constants;
- the assertions
  $$\bigcup_{i \leq n} \{p_1(c_i, p_{i,1}), p_2(c_i, p_{i,2}), n_1(c_i, n_{i,1}), n_2(c_i, n_{i,2})\}$$
  associate each clause with the four variables/truth constants that occur in it, where p_1, p_2, n_1, n_2 are role names that do not occur in $\mathcal{T}$.

We further extend $\mathcal{A}_\varphi$ to enforce a truth value for each of the variables q_i. To this end, add to $\mathcal{A}_\varphi$ copies $\mathcal{A}_0, \ldots, \mathcal{A}_m$ of $\mathcal{A}_\vee$ obtained by renaming individual names such that $\mathsf{Ind}(\mathcal{A}_i) \cap \mathsf{Ind}(\mathcal{A}_j) = \emptyset$ whenever $i \neq j$. As a notational convention, let a_j^i be the name used for the individual name $a_j \in \mathsf{Ind}(\mathcal{A}_\vee)$ in $\mathcal{A}_i$ for all $i \leq m$ and $j \leq k$ (note that a_j comes from the ELIQ $C_j(a_j)$ in the sequence fixed above). Intuitively, the copy $\mathcal{A}_i$ of $\mathcal{A}$ is used to generate a truth value for the variable q_i, where we want to interpret q_i as true if the ELIQ $C_0(a_0^i)$ is satisfied and as false if any of the ELIQs $C_j(a_j^i)$, $0 < j \leq k$, is satisfied. To actually relate each individual name q_i to the associated ABox $\mathcal{A}_i$, we use role names $r_0, \ldots, r_k$ that do not occur in $\mathcal{T}$. More specifically, we extend $\mathcal{A}_\varphi$ as follows:

- link variables q_i to the ABoxes $\mathcal{A}_i$ by adding assertions $r_j(q_i, a_j^i)$ for all $i \leq m$ and $j \leq k$; thus, truth of q_i means that $\exists r_0.C_0(q_i)$ is satisfied and falsity means that $\exists r_j.C_j(q_i)$ is satisfied for some j with $0 < j \leq k$;
- to ensure that 0 and 1 have the expected truth values, add a copy of C_0 viewed as an ABox with root $1'$ and a copy of C_2 viewed as an ABox with root $0'$; add $r_0(1, 1')$ and $r_1(0, 0')$.

Consider the query

$$q_0 = \exists c.(\exists p_1.\mathsf{ff} \sqcap \exists p_2.\mathsf{ff} \sqcap \exists n_1.\mathsf{tt} \sqcap \exists n_2.\mathsf{tt})$$

which describes the existence of a clause with only false literals and thus captures falsity of φ, where tt is an abbreviation for $\exists r_0.C_0$ and ff an abbreviation for the $\mathcal{ELU}$-concept $\exists r_1.C_1 \sqcup \cdots \sqcup \exists r_k.C_k$. It is straightforward to show that φ is unsatisfiable iff $\mathcal{A}, \mathcal{T} \models q_0$. To obtain the desired UELIQ q, it remains to take q and distribute disjunction to the outside.

We now show how to improve the result from UELIQ-answering to ELIQ-answering. Our aim is to change the encoding of falsity of a variable q_i from satisfaction of $\exists r_1.C_1 \sqcup \cdots \sqcup \exists r_k.C_k(q_i)$ to satisfaction of $\exists h.(\exists r_1.C_1 \sqcap \cdots \sqcap \exists r_k.C_k)(q_i)$, where h is an additional role that does not occur in $\mathcal{T}$. We can then replace the concept ff in the query q_0 with $\exists h.(\exists r_1.C_1 \sqcap \cdots \sqcap \exists r_k.C_k)(q_i)$, which directly gives us the desired ELIQ q.

It remains to modify $\mathcal{A}_\varphi$ to support the new encoding of falsity. The basic idea is that each q_i has k successors $b^i_1, \dots, b^i_k$ reachable via h such that for $1 \leq j \leq k$,

- $\exists r_\ell.C_\ell(b^i_j)$ is satisfied for all $\ell = 1, \dots, j-1, j+1, \dots, k$ and
- the assertion $r_j(b^i_j, a^i_j)$ is in $\mathcal{A}_\varphi$.

Thus, $(\exists r_1.C_1 \sqcap \cdots \sqcap \exists r_k.C_k)(b^i_j)$ is satisfied iff $C_j(a^i_j)$ is satisfied, for all j with $1 \leq j \leq k$. In detail, the modification of $\mathcal{A}_\varphi$ is as follows:

- for $1 \leq j \leq k$, add to $\mathcal{A}_\varphi$ a copy of C_j viewed as an ABox, where the root individual name is d_j;
- for all $i \leq m$, replace the assertions $r_j(q_i, a^i_j)$, $1 \leq j \leq k$, with the following:
  - $h(q_i, b^i_1), \dots, h(q_i, b^i_k)$ for all $i \leq m$;
  - $r_j(b^i_j, a^i_j), r_1(b^i_j, d_1), \dots, r_{j-1}(b^i_j, d_{j-1})$, $r_{j+1}(b^i_j, d_{j+1}), \dots, r_k(b^i_j, d_k)$ for all $i \leq m$ and $1 \leq j \leq k$.

This finishes the modified construction. Again, it is not hard to prove correctness.

It remains to note that, when $\mathcal{T}$ is an $\mathcal{ALCF}$-TBox, then the above construction of q yields an ELQ instead of an ELIQ. ❑

C Proofs for Section 4

Lemma 14.
Every Horn-$\mathcal{ALCFI}$-TBox is unraveling tolerant.

Proof. We give a characterization of the entailment of ELIQs in the presence of Horn-$\mathcal{ALCFI}$-TBoxes which is in the spirit of the *rule-based* (sometimes also called *consequence-driven*) algorithms commonly used for Horn description logics such as $\mathcal{EL}$++ and Horn-$\mathcal{SHIQ}$, see e.g. (Baader, Brandt, and Lutz 2005; Kazakov 2009; Krötzsch 2010).

In the characterization, we use *extended ABoxes*, i.e., finite sets of assertions $C(a)$ with C a *potentially compound* concept and $r(a, b)$. An $\mathcal{ELIU}_\bot$-concept is a concept that is formed according to the second syntax rule in the definition of Horn-$\mathcal{ALCFI}$. For an extended ABox $\mathcal{A}'$ and an assertion $C(a)$, C an $\mathcal{ELIU}_\bot$-concept, we write $\mathcal{A}' \vdash C(a)$ if $\mathcal{A}'$ *syntactically entails* $C(a)$, formally:

- $\mathcal{A}' \vdash \top(a)$ is unconditionally true;
- $\mathcal{A}' \vdash \bot(a)$ if $\bot(b) \in \mathcal{A}'$ for some $b \in \mathsf{Ind}(\mathcal{A})$;
- $\mathcal{A}' \vdash A(a)$ if $A(a) \in \mathcal{A}'$;
- $\mathcal{A}' \vdash C \sqcap D(a)$ if $\mathcal{A}' \vdash C(a)$ and $\mathcal{A}' \vdash D(a)$;
- $\mathcal{A}' \vdash C \sqcup D(a)$ if $\mathcal{A}' \vdash C(a)$ or $\mathcal{A}' \vdash D(a)$;
- $\mathcal{A}' \vdash \exists r.C(a)$ if there is an $r(a, b) \in \mathcal{A}'$ such that $\mathcal{A}' \vdash C(b)$.

Now for the characterization. Let $\mathcal{T} = \{\top \sqsubseteq C_\mathcal{T}\}$ be a Horn-$\mathcal{ALCFI}$-TBox and $\mathcal{A}$ a potentially infinite ABox (so that we can also apply the construction to unravelings of ABoxes). We produce a sequence of extended ABoxes $\mathcal{A}_0, \mathcal{A}_1, \dots$, starting with $\mathcal{A}_0 = \mathcal{A} \cup \{\top(a_\top)\}$, where $a_\top$ is a fresh individual which, intuitively, is a representative for all individual names that do not occur in $\mathcal{A}$. In what follows, we use additional individual names of the form $ar_1C_1 \cdots r_kC_k$ with $a \in \mathsf{Ind}(\mathcal{A}_0)$, $r_1, \dots, r_k$ roles that occur in $\mathcal{T}$, and $C_1, \dots, C_k \in \mathsf{sub}(\mathcal{T})$. We assume that $\mathsf{N_I}$ contains such names as needed and use the symbol a also to refer to individual names of this compound form. Each extended ABox $\mathcal{A}_{i+1}$ is obtained from $\mathcal{A}_i$ by applying the following rules:

R1 if $a \in \mathsf{Ind}(\mathcal{A}_i)$, then add $C_\mathcal{T}(a)$.

R2 if $C \sqcap D(a) \in \mathcal{A}_i$, then add $C(a)$ and $D(a)$;

R3 if $C \rightarrow D(a) \in \mathcal{A}_i$ and $\mathcal{A}_i \vdash C(a)$, then add $D(a)$;

R4 if $\exists r.C(a) \in \mathcal{A}_i$ and $\mathsf{func}(r) \notin \mathcal{T}$, then add $r(a, arC)$ and $C(arC)$;

R5 if $\exists r.C(a) \in \mathcal{A}_i$, $\mathsf{func}(r) \in \mathcal{T}$, and $r(a, b) \in \mathcal{A}_i$, then add $C(b)$;

R6 if $\exists r.C(a) \in \mathcal{A}_i$, $\mathsf{func}(r) \in \mathcal{T}$, and there is no $r(a, b) \in \mathcal{A}_i$, then add $r(a, arC)$ and $C(arC)$;

R7 if $\forall r.C(a) \in \mathcal{A}_i$ and $r(a, b) \in \mathcal{A}_i$, then add $C(b)$.

We call $\mathcal{A}_c = \bigcup_{i \geq 0} \mathcal{A}_i$ the *completion* of the original ABox $\mathcal{A}$. Note that $\mathcal{A}_c$ may be infinite even if $\mathcal{A}$ is finite, and that none of the above rules is applicable in $\mathcal{A}_c$. In the following, we write '$\mathcal{A}_c \vdash \bot$' instead of '$\mathcal{A}_c \vdash \bot(a)$ for some $a \in \mathsf{N_C}$'.

Claim 1. For all ELIQs $C(a)$, we have

1. $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff $\mathcal{A}_c \vdash C(a)$ or $\mathcal{A}_c \vdash \bot$;
2. $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff $\mathcal{A}_c \vdash C(a_\top)$ or $\mathcal{A}_c \vdash \bot$ whenever $a \in \mathsf{N_I} \setminus \mathsf{Ind}(\mathcal{A})$.

We only sketch the proof. For the "if" directions, the central observation is that for any model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$, we can construct a homomorphism h from $\mathcal{A}_c$ to $\mathcal{I}$, i.e., h is a map from $\mathsf{Ind}(\mathcal{A}_c)$ to $\Delta^\mathcal{I}$ such that the following conditions are satisfied:

(a) $h(a) = a$ for all $a \in \mathsf{Ind}(\mathcal{A})$;

(b) if $C(a) \in \mathcal{A}_c$, then $h_i(a) \in C^\mathcal{I}$;

(c) if $r(a, b) \in \mathcal{A}_c$, then $(h_i(a), h_i(b)) \in r^\mathcal{I}$.

More specifically, we inductively construct homomorphisms h_i from $\mathcal{A}_i$ to $\mathcal{I}$, that satisfy Conditions (a) to (c) above with $\mathcal{A}_c$ replaced by $\mathcal{A}_i$ and such that $h_0 \subseteq h_1 \subseteq \cdots$. Then $h = \bigcup_{i \geq 0} h_i$ is the required homomorphism from $\mathcal{A}_c$ to $\mathcal{I}$.

Let $C(a)$ be an ELIQ. If $\mathcal{A}_c \vdash \bot$, the existence of a homomorphism h from $\mathcal{A}_c$ into any model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ implies that $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$, whence $(\mathcal{T}, \mathcal{A}) \models C(a)$. If $\mathcal{A}_c \vdash C(a)$, then preservation of ELIQs under homomorphisms also yields $(\mathcal{T}, \mathcal{A}) \models C(a)$. For Point 2, assume $\mathcal{A}_c \vdash C(a_\top)$. We can construct the above homomorphisms h such that $h(a_\top) = a$. Thus, we again obtain $(\mathcal{T}, \mathcal{A}) \models C(a)$.

For the "only if" direction of Point 1, we have to show that if $\mathcal{A}_c \not\vdash C(a)$, where $C(a)$ is an ELIQ, and $\mathcal{A}_c \not\vdash \bot$, then $(\mathcal{T}, \mathcal{A}) \not\models C(a)$ (and similarly for Point 2). Define an interpretation $\mathcal{I}$ as follows:

$$
\begin{array}{rcll}
\Delta^{\mathcal{I}} & = & \mathsf{Ind}(\mathcal{A}_c) & \\
A^{\mathcal{I}} & = & \{a \mid A(a) \in \mathcal{A}_c\} & \text{for all } A \in \mathsf{N_C} \\
r^{\mathcal{I}} & = & \{r(a,b) \mid r(a,b) \in \mathcal{A}_c\} & \text{for all } r \in \mathsf{N_R} \\
a^{\mathcal{I}} & = & a & \text{for all } a \in \mathsf{Ind}(\mathcal{A}) \\
a^{\mathcal{I}} & = & a_\top & \text{for all } a \in \mathsf{N_I} \setminus \mathsf{Ind}(\mathcal{A})
\end{array}
$$

It can be shown that $\mathcal{I}$ is a model of $\mathcal{A}_c$ (thus $\mathcal{A}$) and $\mathcal{T}$ and that $\mathcal{A}_c \not\vdash C(a)$ implies $\mathcal{I} \not\models C(a)$. Thus $(\mathcal{T}, \mathcal{A}) \not\models C(a)$ as required.

We now consider the application of the above completion construction to both the original ABox $\mathcal{A}$ and its unraveling $\mathcal{A}^u$. Recall that individuals in $\mathcal{A}^u$ are of the form $a_0 r_0 a_1 \cdots r_{n-1} a_n$, thus individuals in $\mathcal{A}^u_c$ are of the form $a_0 r_0 a_1 \cdots r_{n-1} a_n s_1 C_1 \cdots s_k C_k$. For $\alpha \in \mathsf{Ind}(\mathcal{A}_c)$ and $\beta \in \mathsf{Ind}(\mathcal{A}^u_c)$, we write $\alpha \sim \beta$ if

$$
\begin{array}{rcl}
\alpha & = & a_n s_1 C_1 \cdots s_k C_k \text{ and} \\
\beta & = & a_0 r_0 a_1 \cdots r_{n-1} a_n s_1 C_1 \cdots s_k C_k
\end{array}
$$

for some $a_0, \ldots, a_n, r_0, \ldots, r_{n-1}, s_1, \ldots, s_k, C_1, \ldots, C_k$. This includes the case where $k = 0$, i.e., the $s_1 C_1 \cdots s_k C_k$ component is empty in both α and β. The following claim can be shown by induction on i.

Claim 2. For all $\alpha \in \mathsf{Ind}(\mathcal{A}_i)$ and $\beta \in \mathsf{Ind}(\mathcal{A}^u_i)$ with $\alpha \sim \beta$, we have

1. $\mathcal{A}_i \vdash C(\alpha)$ iff $\mathcal{A}^u_i \vdash C(\beta)$ for all $\mathcal{ELI}$-concepts C;
2. $C(\alpha) \in \mathcal{A}_i$ iff $C(\beta) \in \mathcal{A}^u_i$ for all $C \in \mathsf{sub}(\mathcal{T})$.

From Claims 1 and 2, we obtain that $\mathcal{A}$ and $\mathcal{A}^u$ entail exactly the same ELIQs. It follows that $\mathcal{T}$ is unraveling tolerant. ❑

Lemma 17. Every unraveling tolerant $\mathcal{ALCFI}$-TBox is materializable.

Proof. We show the contrapositive using a proof strategy that is very similar to the second step in the proof of Theorem 11. Thus, take an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ that is not materializable. By Theorem 9, $\mathcal{T}$ does not have the disjunction property. Thus, there is an ABox $\mathcal{A}_\vee$ and ELIQs $C_0(a_0), \ldots, C_k(a_k)$ such that $(\mathcal{T}, \mathcal{A}_\vee) \models C_0(a_0) \vee \cdots \vee C_k(a_k)$, but $(\mathcal{T}, \mathcal{A}_\vee) \not\models C_i(a_i)$ for all $i \leq k$. Let $\mathcal{A}_i$ be C_i viewed as a tree-shaped ABox with root b_i, for all $i \leq k$. Assume w.l.o.g. that none of the ABoxes $\mathcal{A}_\vee, \mathcal{A}_0, \ldots, \mathcal{A}_k$ share any individual names. Consider the ABox

$$
\begin{array}{rcl}
\mathcal{A} & = & \mathcal{A}_\vee \cup \mathcal{A}_0 \cup \cdots \cup \mathcal{A}_k \cup \{r(b, b_0), \ldots, r(b, b_k)\} \\
& & \cup \{r_0(b_j, b_0), \ldots, r_{j-1}(b_j, b_{j-1}), \\
& & \qquad r_{j+1}(b_j, b_{j+1}), \ldots, r_k(b_j, b_k)\} \\
& & \cup \{r_0(b_0, a_0), \ldots, r_k(b_k, a_k)\}
\end{array}
$$

where b is a fresh individual name and $r, r_0, \ldots, r_k$ do not occur in $\mathcal{T}$, and the ELIQ

$$q = \exists r.(\exists r_0.C_0 \sqcap \cdots \sqcap \exists r_k.C_k)(b).$$

Then we have

Claim. $(\mathcal{T}, \mathcal{A}) \models q$, but $(\mathcal{T}, \mathcal{A}^u) \not\models q$.

Proof. "$(\mathcal{T}, \mathcal{A}) \models q$". Take a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{T}$. By construction of $\mathcal{A}$, we have $b_i^{\mathcal{I}} \in (\exists r_j.C_j)^{\mathcal{I}}$ whenever $i \neq j$. Due to the inclusion of $\mathcal{A}_\vee$ and since $(\mathcal{T}, \mathcal{A}_\vee) \models C_0(a_0) \vee \cdots \vee C_k(a_k)$, we find one b_i such that $b_i^{\mathcal{I}} \in (\exists r_i.C_i)^{\mathcal{I}}$. Consequently, $\mathcal{I} \models q$.

"$(\mathcal{T}, \mathcal{A}^u) \not\models q$" (sketch). Consider the elements $brb_i r_i a_i$ in $\mathcal{A}^u$. Each such element is the root of a copy of the unraveling $\mathcal{A}^u_\vee$ of $\mathcal{A}_\vee$, restricted to those individuals in $\mathcal{A}_\vee$ that are reachable from a_i. Since $(\mathcal{T}, \mathcal{A}_\vee) \not\models C_i(a_i)$, we find a model $\mathcal{I}_i$ of $\mathcal{T}$ and $\mathcal{A}_\vee$ with $a_i^{\mathcal{I}_i} \notin C_i^{\mathcal{I}_i}$. By unraveling $\mathcal{I}$, we obtain a model $\mathcal{I}^u_i$ of $\mathcal{T}$ and $\mathcal{A}^u_\vee$ with $a_i^{\mathcal{I}^u_i} \notin C_i^{\mathcal{I}^u_i}$. By combining the models $\mathcal{I}^u_0, \ldots, \mathcal{I}^u_k$, one can craft a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}^u_\vee$ such that $brb_i r_i a_i^{\mathcal{I}} \notin C_i^{\mathcal{I}}$ for all $i \leq k$. Consequently, $\mathcal{I} \not\models q$.

It follows that $\mathcal{T}$ is not unraveling tolerant. ❑

Theorem 16. If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is unraveling tolerant, then PEQ-answering w.r.t. $\mathcal{T}$ is in PTIME.

To prove Theorem 16, let $\mathcal{T} = \{\top \sqsubseteq C_{\mathcal{T}}\}$ be an unraveling tolerant TBox, where we assume w.l.o.g. that $C_{\mathcal{T}}$ is built from the constructors $\neg$, $\sqcap$, and $\exists r.C$, only. By Theorem 4, it suffices to show that ELIQ-answering w.r.t. $\mathcal{T}$ is in PTIME. Thus, let $q = C_0(a_0)$ be an ELIQ. We use $\mathsf{cl}(\mathcal{T}, q)$ to denote the set of subconcepts of $\mathcal{T}$ and q, closed under single negation. For an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, we use $t^{\mathcal{I}}_{\mathcal{T},q}(d)$ to denote the set of concepts $C \in \mathsf{cl}(\mathcal{T}, q)$ such that $C \in d^{\mathcal{I}}$. A $\mathcal{T}, q$*-type* is a subset $t \subseteq \mathsf{cl}(\mathcal{T}, q)$ such that for some model $\mathcal{I}$ of $\mathcal{T}$, we have $t = t^{\mathcal{I}}_{\mathcal{T},q}(d)$. We use $\mathsf{tp}(\mathcal{T}, q)$ to denote the set of all $\mathcal{T}$,q-types. For $t, t' \in \mathsf{tp}(\mathcal{T}, q)$ and r a role, we write $t \rightsquigarrow_r t'$ if the following conditions are satisfied:

- if $C \in t'$ then $\exists r.C \in t$, for all $\exists r.C \in \mathsf{cl}(\mathcal{T}, q)$;
- if $C \in t$ then $\exists r^-.C \in t'$, for all $\exists r^-.C \in \mathsf{cl}(\mathcal{T}, q)$;
- $\exists r.C \in t$ iff $C \in t'$, for all $\exists r.C \in \mathsf{cl}(\mathcal{T}, q)$ with $\mathsf{func}(r) \in \mathcal{T}$;
- $\exists r^-.C \in t'$ iff $C \in t$, for all $\exists r^-.C \in \mathsf{cl}(\mathcal{T}, q)$ with $\mathsf{func}(r^-) \in \mathcal{T}$.

A *type assignment* is a map $\mathsf{Ind}(\mathcal{A}) \rightarrow 2^{\mathsf{tp}(\mathcal{T},q)}$. The PTime algorithm for checking, given an ABox $\mathcal{A}$, whether $(\mathcal{T}, \mathcal{A}) \models q$ is based on the computation of a sequence of type assigments $\pi_0, \pi_1, \dots$ as follows. For every $a \in \mathsf{Ind}(\mathcal{A})$, $\pi_0(a)$ is the set of all types $t \in \mathsf{tp}(\mathcal{T}, q)$ such that $A(a) \in \mathcal{A}$ implies $A \in t$. Then, $\pi_{i+1}(a)$ is defined as the set of all types $t_a \in \pi_i(a)$ such that for all $r(a, b) \in \mathcal{A}$, r a role name or the inverse thereof, there is a type $t_b \in \pi_i(b)$ such that $t_a \rightsquigarrow_r t_b$.

Clearly, the sequence $\pi_0, \pi_1, \dots$ will stabilize after at most $\mathcal{O}(|\mathcal{A})|$ steps and can be computed in time polynomial in $|\mathcal{A}|$ (since $|\mathcal{T}|$ and thus $|\mathsf{tp}(\mathcal{T}, q)|$ is a constant). Let π be the final type assignment in the sequence. The following yields Theorem 16.

Lemma 31. *$(\mathcal{T}, \mathcal{A}) \models q$ iff $C_0 \in t$ for all $t \in \pi(a_0)$.*

Proof. By unraveling tolerance, we have $(\mathcal{T}, \mathcal{A}) \models q$ iff $(\mathcal{T}, \mathcal{A}^u) \models q$. It thus suffices to show that for all $t \in \mathsf{tp}(\mathcal{T}, q)$, we have $t \in \pi(a_0)$ iff there is a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}^u$ with $\mathsf{tp}^{\mathcal{I}}_{\mathcal{T},q}(a_0^{\mathcal{I}}) = t$.

"$\Leftarrow$". Let $\mathcal{I}$ be a model of $\mathcal{T}$ and $\mathcal{A}^u$ with $\mathsf{tp}^{\mathcal{I}}_{\mathcal{T},q}(a_0^{\mathcal{I}}) = t$. It is not hard to show by induction on i that for all $i \geq 0$ and all $a_0 \cdots a_k \in \mathsf{Ind}(\mathcal{A}^u)$, we have $t^{\mathcal{I}}_{\mathcal{T},q}(a_k^{\mathcal{I}}) \in \pi_i(a_k)$. In particular, this implies that $t^{\mathcal{I}}_{\mathcal{T},q}(a_0) \in \pi(a_0)$.

"$\Rightarrow$". Let $t \in \pi(a_0)$. We build a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}^u$ such that $t^{\mathcal{I}}_{\mathcal{T},q}(a_0^{\mathcal{I}}) = t$, as follows. First, construct a map $\lambda : \mathsf{Ind}(\mathcal{A}^u) \rightarrow \mathsf{tp}(\mathcal{T}, q)$ such that for all $a_0 \cdots a_k \in \mathsf{Ind}(\mathcal{A}^u)$, we have $\lambda(a_0 \cdots a_k) \in \pi(a_k)$. Start with setting $\lambda(a_0) = t$. Then exhaustively apply the following steps, where r is a role name:

- if $\lambda(a_0 \cdots a_k)$ is defined, $r(a_k, a_{k+1}) \in \mathcal{A}$, and $\lambda(a_0 \cdots a_k r a_{k+1})$ is undefined, then by the definition of the sequence $\pi_0, \pi_1, \dots$ and since $\lambda(a_0 \cdots a_k) \in \pi(a_k)$, there is a type $t' \in \pi(a_{k+1})$ such that $\lambda(a_0 \cdots a_k) \rightsquigarrow_r t'$. Set $\lambda(a_0 \cdots r a_{k+1}) = t'$.
- if $\lambda(a_0 \cdots a_k)$ is defined, $r(a_{k+1}, a_k) \in \mathcal{A}$, and $\lambda(a_0 \cdots a_k r^- a_{k+1})$ is undefined, then by the definition of the sequence $\pi_0, \pi_1, \dots$ and since $\lambda(a_0 \cdots a_k) \in \pi(a_k)$, there is a type $t' \in \pi(a_{k+1})$ such that $\lambda(a_0 \cdots a_k) \rightsquigarrow_{r^-} t'$. Set $\lambda(a_0 \cdots a_k r^- a_{k+1}) = t'$.

By definition of types, for each $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$ we find a tree-shaped model $\mathcal{I}_\alpha$ of $\mathcal{T}$ and $\mathcal{A}$ and a $d_\alpha \in \Delta^{\mathcal{I}_\alpha}$ such that $t^{\mathcal{I}_\alpha}_{\mathcal{T},q}(d_\alpha) = \lambda(\alpha)$. Assume w.l.o.g. that the domains of all these models $\Delta^{\mathcal{I}_\alpha}$ are disjoint. Define a new interpretation $\mathcal{I}$ as follows:

(i) take the disjoint union of the models $\mathcal{I}_\alpha$, $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$;

(ii) whenever $(d_\alpha, e) \in r^{\mathcal{I}}$, $\mathsf{func}(r) \in \mathcal{T}$, and there is an assertion $r(\alpha, \beta) \in \mathcal{A}^u$, remove the subtree rooted at e;

(iii) for all $r(\alpha, \beta) \in \mathcal{A}^u$, add (d_α, d_β) to $r^{\mathcal{I}}$;

(iv) set $\alpha^{\mathcal{I}} = d_\alpha$, for all $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$.

We need to show that that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}^u$, and that $t^{\mathcal{I}}(d_{a_0}) = t$. By definition of π_0 in the sequence $\pi_0, \pi_1, \dots$ and Point (iii) in the definition of $\mathcal{I}$, it is clear that $\mathcal{I}$ is a model of $\mathcal{A}$. All functionality statements $\mathsf{func}(r) \in \mathcal{T}$ are satisfied:

Claim 1. If $\mathsf{func}(r) \in \mathcal{T}$, then $r^{\mathcal{I}}$ is a partial function.

Proof of claim. Since $\mathcal{A}$ is a model of $\mathcal{T}$ and by the UNA, for each $a \in \mathsf{Ind}(\mathcal{A})$ there is at most one $b \in \mathsf{Ind}(\mathcal{A})$ with $r(a, b) \in \mathcal{A}$. By definition of the unraveled ABox $\mathcal{A}^u$, it follows that for each $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$ there is at most one $\beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $r(\alpha, \beta) \in \mathcal{A}$. By Points (ii) and (iii) of the definition of $\mathcal{I}$ and since each $\mathcal{I}_\alpha$ is a model of $\mathcal{T}$, $r^{\mathcal{I}}$ is a partial function.

It thus remains to show that $\mathcal{I}$ satisfies all concept inclusions in $\mathcal{T}$ and that $t^{\mathcal{I}}(d_{a_0}) = t$. Both is a consequence of the following.

Claim 2. For all $C \in \mathsf{cl}(\mathcal{T}, q)$ and $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$, we have

1. $d_\alpha \in C^{\mathcal{I}}$ iff $C \in \lambda(\alpha)$
2. $d \in C^{\mathcal{I}}$ iff $d \in C^{\mathcal{I}_\alpha}$, for all $d \in \Delta^{\mathcal{I}_\alpha} \setminus \{d_\alpha\}$.

The proof is by induction on the structure of C. Details are left to the reader. ❑

A finite interpretation $\mathcal{I}$ is a tree interpretation iff

$$(\Delta^{\mathcal{I}}, \bigcup_{r \in \mathsf{N_R}, (d,d') \in r^{\mathcal{I}}} \{d, d'\})$$

is an undirected tree with $r^{\mathcal{I}} \cap s^{\mathcal{I}} = \emptyset$ for any two distinct $r, s \in \mathsf{N_R}$. A non-uniform constraint satisfaction problem CSP($\mathcal{I}$) in Σ has tree obstructions iff there exists a set $\Xi_{\mathcal{I}}$ of Σ tree interpretations such that for all finite Σ-interpretations $\mathcal{J}$:

$$\text{not } \mathsf{Hom}(\mathcal{J}, \mathcal{I}) \quad \Leftrightarrow \quad \exists \mathcal{J}' \in \Xi_{\mathcal{I}} : \mathsf{Hom}(\mathcal{J}', \mathcal{J})$$

Theorem 32. *Let $\mathcal{T}$ be a $\mathcal{ALCFI}$-TBox. Then $\mathcal{T}$ is unraveling tolerant iff all $\mathcal{I}_{\mathcal{T},q}$, q an ELIQ, have tree obstructions.*

Proof. We use the notation from Theorem 24. The main observation is that if there is a homomorphism from a tree interpretation $\mathcal{I}$ to an ABox (regarded as an interpretation), then there is a homomorphism from $\mathcal{I}$ to the unraveling of the ABox (regarded as an interpretation). We now give the details.

Assume $\mathcal{T}$ is unraveling tolerant. Let $q = C(a)$ be an ELIQ. Let $\Sigma = \mathsf{sig}(\mathcal{T}) \cup \mathsf{sig}(C) \cup \{P\}$. For every ABox $\mathcal{A}$, we have $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff $(\mathcal{T}, \mathcal{A}^u) \models C(a)$. By compactness, for every $\mathcal{A}$ with $(\mathcal{T}, \mathcal{A}) \models C(a)$ there exists a finite $\mathcal{A}^f \subseteq \mathcal{A}^u$ such that $(\mathcal{T}, \mathcal{A}^f) \models C(a)$. From $\mathcal{A}^f$ we obtain $\mathcal{A}^{f,\overline{P}}$ by adding $P(a)$ to $\mathcal{A}^f$ and removing all other occurrences of P. Now let Ξ_q denote the set of all $\mathcal{I}^{\Sigma}_{\mathcal{A}^{f,P}}$. We show that Ξ_q satisfies the conditions for tree obstructions for the template $\mathcal{I}_{\mathcal{T},q}$:

Assume not $\mathsf{Hom}(\mathcal{J}, \mathcal{I}_{\mathcal{T},q})$. Let $\mathcal{A}$ be an ABox with $\mathcal{J} = \mathcal{I}^{\Sigma}_{\mathcal{A}}$. Then $(\mathcal{T}, \mathcal{A}) \models \exists x.(P(x) \wedge C(x))$. By materializability, there exists $a \in \mathsf{Ind}(\mathcal{A})$ with $P(a) \in \mathcal{A}$ and $(\mathcal{T}, \mathcal{A}) \models C(a)$. Hence $\mathcal{I}^{\Sigma}_{\mathcal{A}^{f,P}} \in \Xi_q$ and clearly $\mathsf{Hom}(\mathcal{I}^{\Sigma}_{\mathcal{A}^{f,P}}, \mathcal{J})$, as required.

Conversely, assume $\mathsf{Hom}(\mathcal{I}^{\Sigma}_{\mathcal{A}^{f,P}}, \mathcal{J})$ for some $\mathcal{I}^{\Sigma}_{\mathcal{A}^{f,P}} \in \Xi_q$. We have $(\mathcal{T}, \mathcal{A}^f) \models C(a)$. Hence not $\mathsf{Hom}(\mathcal{I}^{\Sigma}_{\mathcal{A}^{f,P}}, \mathcal{I}_{\mathcal{T},q})$. But then $\mathsf{Hom}(\mathcal{J}, \mathcal{I}_{\mathcal{T},q})$, as required.

Now assume that all $\mathcal{I}_{\mathcal{T},q}$, q an ELIQ, have tree obstructions. Fix an ELIQ $C(a)$ and let $\mathcal{A}$ be an ABox with $(\mathcal{T},\mathcal{A}) \models C(a)$. We have to show that $(\mathcal{T},\mathcal{A}^u) \models C(a)$. We do not have $\mathsf{Hom}(\mathcal{I}^\Sigma_{\mathcal{A}'},\mathcal{I}_{\mathcal{T},q})$. By the existence of tree obstructions, there is a Σ tree interpretation $\mathcal{J}$ with $\mathsf{Hom}(\mathcal{J},\mathcal{I}^\Sigma_{\mathcal{A}'})$ and not $\mathsf{Hom}(\mathcal{J},\mathcal{I}_{\mathcal{T},q})$. But then $\mathsf{Hom}(\mathcal{J},\mathcal{I}^\Sigma_{\mathcal{A}'^u})$, by the observation above. Hence there is a finite subset $\mathcal{A}^f$ of $\mathcal{A}'^u$ with $\mathsf{Hom}(\mathcal{J},\mathcal{I}^\Sigma_{\mathcal{A}^f})$. But then not $\mathsf{Hom}(\mathcal{I}^\Sigma_{\mathcal{A}^f},\mathcal{I}_{\mathcal{T},q})$ from which we obtain $(\mathcal{T},\mathcal{A}^f) \models \exists x.(P(x) \wedge C(x))$, and then $(\mathcal{T},\mathcal{A}^f) \models C(a)$, by materializability. ❑

D Proofs for Section 5

Theorem 18. Every materializable $\mathcal{ALCFI}$-TBox of depth one is unraveling tolerant.

For the proof of Theorem 18, we need a preliminary. An $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is *infinitely materializable* if for every finite *and infinite* ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$, there is an ELIQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. As in the case of plain materializability, it would be equivalent to define infinite materializability based on CQs or PEQs.

Lemma 33. *An $\mathcal{ALCFI}$-TBox is materializable iff it is infinitely materializable.*

This lemma follows from the observation that the proof of the "if" direction of Theorem 10 goes through without modification also for infinite ABoxes.

Proof. (of Theorem 18) Let $\mathcal{T}$ be a materializable TBox of depth one, $\mathcal{A}$ an ABox, and $q = C_0(a_0)$ an ELIQ with $(\mathcal{T},\mathcal{A}^u) \not\models q$. We have to show that $(\mathcal{T},\mathcal{A}) \not\models q$. It follows from $(\mathcal{T},\mathcal{A}^u) \not\models q$ that $\mathcal{A}^u$ is consistent w.r.t. $\mathcal{T}$ and by Lemma 33 there is a materialization $\mathcal{I}^u$ for $\mathcal{T}$ and $\mathcal{A}^u$. We have $\mathcal{I}^u \not\models q$ and our aim is to convert $\mathcal{I}^u$ into a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $\mathcal{I} \not\models q$. Before we do this, we first uniformize $\mathcal{I}^u$ in a suitable way, as detailed below.

We assume w.l.o.g. that $\mathcal{I}^u$ has forest-shape, i.e., that $\mathcal{I}^u$ can be constructed by selecting a tree-shaped interpretation $\mathcal{I}_\alpha$ with root α for each $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$, then taking the disjoint union of all these interpretations, and finally adding role edges (α,β) to $r^{\mathcal{I}^u}$ whenever $r(\alpha,\beta) \in \mathcal{A}^u$. In fact, to achieve the desired shape we can simply unravel $\mathcal{I}^u$ starting from the elements $\mathsf{Ind}(\mathcal{A}^u) \subseteq \Delta^{\mathcal{I}^u}$ and then use Point 1 of Lemma 8 and the fact that there is an i-simulation from the unraveling of $\mathcal{I}^u$ to $\mathcal{I}^u$ to show that the obtained model is still a materialization of $\mathcal{T}$ and $\mathcal{A}$, thus still $\mathcal{I}^u \not\models q$. To ease notation, we generally assume that $\mathsf{Ind}(\mathcal{A}^u) \subseteq \Delta^{\mathcal{I}^u}$ and $\alpha^{\mathcal{I}^u} = \alpha$ for all $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$.

We start with exhibiting a self-similarity inside the unraveled ABox $\mathcal{A}^u$.

Claim 1. For all $\alpha,\beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$ and all $\mathcal{ALCFI}$-concepts C, we have $\mathcal{A}^u \models C(\alpha)$ iff $\mathcal{A}^u \models C(\beta)$.

Assume to the contrary that there are $\alpha,\beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$, $\mathcal{A}^u \models C(\alpha)$, and $\mathcal{A}^u \not\models C(\beta)$. Then there is a model $\mathcal{I}$ of $\mathcal{A}^u$ and $\mathcal{T}$ such that $\mathcal{I} \not\models C(\beta)$. We exhibit a model $\mathcal{J}$ of $\mathcal{A}^u$ and $\mathcal{T}$ such that $\mathcal{J} \not\models C(\alpha)$, in contradiction to $\mathcal{A}^u \models C(\alpha)$.

Define a map $\iota : \mathsf{Ind}(\mathcal{A}^u) \to \mathsf{Ind}(\mathcal{A}^u)$ such that $\mathsf{tail}(\gamma) = \mathsf{tail}(\iota(\gamma))$ for all $\gamma \in \mathsf{Ind}(\mathcal{A}^u)$ as follows:

1. Start with setting $\iota(\alpha) = \beta$;
2. if $\iota(\gamma)$ is defined, $\gamma = a_0r_0a_1 \cdots a_{n-1}r_{n-1}a_n$, $\iota(a_0 \cdots a_{n-1})$ is undefined, and $\iota(\gamma)$ is of the form $b_0s_0b_1 \cdots b_{m-1}s_{m-1}b_m$ with $s_{m-1} = r_{n-1}$ and $b_{m-1} = a_{n-1}$, then set $\iota(a_0 \cdots a_{n-1}) = b_0 \cdots b_{m-1}$;
3. if $\iota(\gamma)$ is defined, $\gamma = a_0r_0a_1 \cdots a_{n-1}r_{n-1}a_n$, $\iota(a_0 \cdots a_{n-1})$ is undefined, and $\iota(\gamma)$ is not of the form $b_0s_0b_1 \cdots b_{m-1}s_{m-1}b_m$ with $s_{m-1} = r_{n-1}$ and $b_{m-1} = a_{n-1}$, then $\iota(\gamma)r^-_{n-1}a_{n-1} \in \mathsf{Ind}(\mathcal{A}^u)$ and we use it as the value of $\iota(a_0 \cdots a_{n-1})$;
4. if $\iota(\gamma)$ is defined, $\gamma ra \in \mathsf{Ind}(\mathcal{A}^u)$, $\iota(\gamma ra)$ is undefined, and $\iota(\gamma)$ is of the form $a_0 \cdots a_{n-1}r_{n-1}a_n$ with $r_{n-1} = r^-$ and $a_{n-1} = a$, then set $\iota(\gamma ra) = a_0 \cdots a_{n-1}$;
5. if $\iota(\gamma)$ is defined, $\gamma ra \in \mathsf{Ind}(\mathcal{A}^u)$, $\iota(\gamma ra)$ is undefined, and $\iota(\gamma)$ is of the form $a_0 \cdots a_{n-1}r_{n-1}a_n$ with $r_{n-1} = r^-$ and $a_{n-1} = a$, then $\iota(\gamma)ra \in \mathsf{Ind}(\mathcal{A}^u)$ and we use it as the value of $\iota(\gamma ra)$;
6. if the value of $\iota(\gamma)$ is undefined after exhaustive application of the above rules, set $\iota(\gamma) = \gamma$.

It can be verified that ι is an ABox automorphism, i.e. for all $\gamma \in \mathsf{Ind}(\mathcal{A}^u)$, $A \in \mathsf{N_C}$, and $r \in \mathsf{N_R}$, we have

- $A(\gamma) \in \mathcal{A}^u$ iff $A(\iota(\gamma)) \in \mathcal{A}^u$;
- $r(\gamma,\gamma') \in \mathcal{A}^u$ iff $r(\iota(\gamma),\iota(\gamma')) \in \mathcal{A}^u$.

Let the interpretation $\mathcal{J}$ be defined as $\mathcal{I}$, but put $\gamma^{\mathcal{J}} = \iota(\gamma)^{\mathcal{I}}$ for all $\gamma \in \mathsf{Ind}(\mathcal{A}^u)$. $\mathcal{J}$ is a model of $\mathcal{A}$ since ι is an ABox automorphism and a model of $\mathcal{T}$ since $\mathcal{I}$ is. Moreover, $\mathcal{I} \not\models C(\beta)$ implies $\beta^{\mathcal{I}} \notin C^{\mathcal{I}}$, which implies $\beta^{\mathcal{I}} \notin C^{\mathcal{J}}$ by definition of $\mathcal{J}$. Since $\alpha^{\mathcal{J}} = \beta^{\mathcal{I}}$, we have $\mathcal{J} \not\models C(\alpha)$ as required. This finishes the proof of Claim 1.

Using Claim 1, we exhibit some self-similarity also inside $\mathcal{I}^u$. However, we cannot use $\mathcal{ALCFI}$-concepts here since entailment by $\mathcal{A}^u$ agrees with truth in $\mathcal{I}^u$ only for ELIQs, but not for $\mathcal{ALCFI}$-instance queries. We thus concentrate on $\mathcal{ELI}$-concepts and $\mathcal{BL}$*-concepts*, where the latter are constructed only from concept names and the Boolean operators.

Claim 2. For all $\alpha,\beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$, we have

1. $\alpha \in C^{\mathcal{I}^u}$ iff $\beta \in C^{\mathcal{I}^u}$ for all $\mathcal{ELI}$-concepts C and
2. $\alpha \in C^{\mathcal{I}^u}$ iff $\beta \in C^{\mathcal{I}^u}$ for all $\mathcal{BL}$-concepts C.

Point 1 is an immediate consequence of Claim 1 and the fact that $\mathcal{I}^u$ is an ELIQ-materialization of $\mathcal{A}^u$. For Point 2, note that Point 1 yields $\alpha \in A^{\mathcal{I}_u}$ iff $\beta \in A^{\mathcal{I}_u}$ for all concept names A. Point 2 then follows by a straightforward induction on the structure of C.

Now for the announced uniformization of $\mathcal{I}^u$. What we want to achieve is that for all $\alpha,\beta \in \mathsf{Ind}(\mathcal{A}^u)$, $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$ implies $\mathcal{I}_\alpha = \mathcal{I}_\beta$ (recall that $\mathcal{I}_\alpha$ is the tree component of $\mathcal{I}^u$ rooted at α, and likewise for $\mathcal{I}_\beta$). Construct the interpretation $\mathcal{J}^u$ as follows:

- for each $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = a$, take a copy $\mathcal{J}_\alpha$ of $\mathcal{I}_a$ with root α;
- then $\mathcal{J}^u$ is the disjoint union of all interpretations $\mathcal{J}_\alpha$, $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$, extended with a role edge $(\alpha, \beta) \in r^{\mathcal{J}^u}$ whenever $r(\alpha, \beta) \in \mathcal{A}^u$.

It is straightforward to verfiy that $\mathcal{J}^u$ is a model of $\mathcal{A}^u$: all role assertions are satisfied by construction of $\mathcal{J}^u$; moreover, $A(\alpha) \in \mathcal{A}^u$ implies $A(a) \in \mathcal{A}^u$ where $a = \mathsf{tail}(\alpha)$, thus $a \in A^{\mathcal{I}_u}$; by construction of $\mathcal{J}^u$, this yields $\alpha \in A^{\mathcal{J}_u}$ as required.

Next, we show that $\mathcal{J}^u$ is a model of $\mathcal{T}$. Let $f : \Delta^{\mathcal{J}^u} \to \Delta^{\mathcal{I}^u}$ be a mapping that assigns to each domain element of $\mathcal{J}^u$ the original element in $\mathcal{I}^u$ of which it is a copy.

Claim 3. For every $d \in \Delta^{\mathcal{J}^u}$ and $\mathcal{ALCFI}$-concept C of depth one, we have $d \in C^{\mathcal{J}^u}$ iff $f(d) \in C^{\mathcal{I}^u}$.

The proof of claim 3 is by induction on the structure of C. We assume w.l.o.g. that C is built only from the constructors $\neg$, $\sqcap$, and $\exists r.C$. The base case, where C is a concept name, is an immediate consequence of the definition of $\mathcal{I}$. The case where $C = \neg D$ and $C = D_1 \sqcap D_2$ is routine. Thus we concentrate on the case where $C = \exists r.D$, where $r \in \mathsf{N_R} \cup \mathsf{N_R^-}$.

First let $d \in C^{\mathcal{J}^u}$. Then there is a $(d, e) \in r^{\mathcal{J}^u}$ with $e \in D^{\mathcal{J}^u}$. First assume that the edge (d, e) was added to $r^{\mathcal{J}^u}$ because $d = \alpha$ and $e = \beta$ for some $\alpha, \beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $r(\alpha, \beta) \in \mathcal{A}^u$. Let $\mathsf{tail}(\alpha) = a$ and $\mathsf{tail}(\beta) = b$. Then we have $f(\alpha) = a$ and $f(\beta) = b$. By construction of $\mathcal{A}^u$, $r(\alpha, \beta) \in \mathcal{A}^u$ implies that $\beta = \alpha r b$ or $\alpha = \beta r^- a$. In both cases we have $r(a, b) \in \mathcal{A}$, thus $r(a, arb) \in \mathcal{A}^u$, thus $(a, arb) \in r^{\mathcal{I}^u}$. Since $\beta = e \in D^{\mathcal{J}^u}$, IH yields that $b \in D^{\mathcal{I}^u}$. Since C is of depth one, D is a $\mathcal{BL}$-concept. By Point 2 of Claim 2, $arb \in D^{\mathcal{I}^u}$ and we are done. Now assume that there is an $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$ such that $(d, e) \in \mathcal{J}_\alpha$. By construction of $\mathcal{J}^u$, we then have $(f(d), f(e)) \in r^{\mathcal{I}^u}$ and IH yields $f(e) \in D^{\mathcal{I}^u}$.

Now let $f(d) \in C^{\mathcal{I}^u}$. Then there is a $(f(d), e) \in r^{\mathcal{I}^u}$ with $e \in D^{\mathcal{I}^u}$. First assume that $f(d) = \alpha$ and $e = \beta$ for some $\alpha, \beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $r(\alpha, \beta) \in \mathcal{A}^u$. Since $f(d) \in \mathsf{Ind}(\mathcal{A}^u)$, we must have $d = \gamma \in \mathsf{Ind}(\mathcal{A}^u)$ and $f(d) = a \in \mathsf{Ind}(\mathcal{A})$ with $\mathsf{tail}(\gamma) = a$. By construction of $\mathcal{A}^u$, $r(a, \beta) \in \mathcal{A}^u$ implies that $\beta = arb$, thus $r(a, b) \in \mathcal{A}$, thus $r(\gamma, \delta) \in \mathcal{A}^u$ with (i) $\delta = \gamma r b$ or (ii) $\gamma = \delta r^- a$ and $\mathsf{tail}(\delta) = b$. Since $arb = e \in D^{\mathcal{I}^u}$, Point 2 of Claim 2 yields $b \in D^{\mathcal{I}_u}$. Since $\mathsf{tail}(\delta) = b$ implies $f(\delta) = b$, IH yields $\delta \in D^{\mathcal{J}^u}$ and we are done. Now assume that there is an $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$ such that $(f(d), e) \in \mathcal{I}_\alpha$. By construction of $\mathcal{J}^u$, $f(d)$ being in $\mathcal{I}_\alpha$ implies that $\alpha = a$ for some $a \in \mathsf{Ind}(\mathcal{A})$ and that there is an $\alpha' \in \mathsf{Ind}(\mathcal{A}^u)$ such that d is in $\mathcal{J}_{\alpha'}$ and $\mathsf{tail}(\alpha') = a$. Again by construction of $\mathcal{J}^u$, we thus find an e' in $\mathcal{J}_{\alpha'}$ with $f(e') = e$ and $(d, e') \in r^{\mathcal{J}_{\alpha'}} \subseteq r^{\mathcal{J}^u}$. IH yields $e' \in D^{\mathcal{J}^u}$.

This finishes the proof of Claim 3. We can now show that $\mathcal{J}^u$ is a model of $\mathcal{T}$. First, $\mathcal{J}^u$ satisfies all CIs in $\mathcal{T}$ since $\mathcal{I}^u$ does and by Claim 3. It remains to show that $\mathcal{I}$ satisfies all functionality assertions in $\mathcal{T}$. Thus, let $\mathsf{func}(r) \in \mathcal{T}$. We show that each $d \in \Delta^{\mathcal{J}^u}$ has at most one r-successor in $\mathcal{J}^u$. Distinguish two cases:

- $d \notin \mathsf{Ind}(\mathcal{A}^u)$. Then d has at most one r-successor since $\mathcal{I}^u$ satisfies $\mathsf{func}(r)$ and by construction of $\mathcal{J}^u$.
- $d = \alpha \in \mathsf{Ind}(\mathcal{A}^u)$. Let $\mathsf{tail}(\alpha) = a$. By construction of $\mathcal{J}^u$ and $\mathcal{A}^u$, α has the same number of r-successors in $\mathcal{J}^u$ as a in $\mathcal{I}^u$. Since $\mathcal{I}^u$ satisfies $\mathsf{func}(r)$, α can have at most one r-successor in $\mathcal{J}^u$.

The final condition that $\mathcal{J}^u$ should satisfy is that $\mathcal{J}^u \not\models q = C_0(a_0)$. Assume to the contrary that $\mathcal{J}^u \models q$. We view q as a tree-shaped CQ whose root is the individual name a_0 and whose non-root nodes are variables, thus $\mathcal{J}^u \models q$ means that there is a match π of q in $\mathcal{J}^u$, i.e., a mapping $\pi : \mathsf{term}(q) \to \Delta^{\mathcal{J}^u}$ such that $\pi(a_0) = a_0$, $A(t) \in q$ implies $\pi(t) \in A^{\mathcal{J}^u}$, and $r(t, t') \in q$ implies $(\pi(t), \pi(t')) \in r^{\mathcal{J}^u}$. We prove that this implies the existence of a match τ for q in $\mathcal{I}^u$, which yields a contradiction to $\mathcal{I}^u \not\models q$.

We start the construction of τ by setting $\tau(t) = \pi(t)$ for all $t \in \mathsf{term}(q)$ with $\pi(t) \in \mathsf{Ind}(\mathcal{A}^u)$. It remains to define $\tau(x)$ for all variables $x \in \mathsf{term}(q)$ such that $\pi(x) \neq \alpha$ for all $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$. This is done by applying the following construction, for each $t \in \mathsf{term}(q)$ such that $\pi(t) = \alpha \in \mathsf{Ind}(\mathcal{A}^u)$.

Recall that $\mathcal{J}_\alpha$ is the tree interpretation rooted at α in $\mathcal{J}^u$. Let V be the set of all variables $x \in \mathsf{term}(q)$ such that there is a sequence $r_1(t_1, t_2), \ldots, r_{n-1}(t_{n-1}, t_n) \in q$, $r_i \in \mathsf{N_R} \cup \mathsf{N_R^-}$, such that $t_1 = t$, $t_n = x$, and $\pi(t_i) \in \Delta^{\mathcal{J}_\alpha} \setminus \{\alpha\}$ for $2 \leq i \leq n$. We define $\tau(x)$ for all $x \in V$ simultaneously. To this end, let $\mathcal{J}_\alpha^V$ be the restriction of $\mathcal{J}_\alpha$ to those elements that are in V. It is not hard to verify that $\mathcal{J}_\alpha^V$ is a finite tree and an initial piece of the potentially infinite tree $\mathcal{J}_\alpha$. Let C_V be an $\mathcal{ELI}$-concept that describes $\mathcal{J}_\alpha^V$ up to homomorphisms, i.e., for any interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$ we have $d \in C_V^{\mathcal{I}}$ iff $\mathcal{J}_\alpha^V$ can be embedded into $\mathcal{I}$ with a $\mathsf{sig}(q)$-homomorphism (a homomorphism that ignores all symbols which do not occur in q) h such that $h(\alpha) = d$. Let $\mathsf{tail}(\alpha) = a$. By construction of $\mathcal{J}^u$, the tree component $\mathcal{I}_a$ of $\mathcal{I}^u$ is identical to $\mathcal{J}_\alpha$ and thus has $\mathcal{J}_\alpha^V$ as an initial piece, which implies $a \in C_V^{\mathcal{I}^u}$. Point 1 of Claim 2 yields $\alpha \in C_V^{\mathcal{I}_u}$ and consequently there is a homomorphism h that embeds $\mathcal{J}_\alpha^v$ into $\mathcal{I}^u$ such that $h(\alpha) = \alpha$. To define the match τ for the variables in V, compose π with h.

It can be verified that the overall mapping τ obtain in this way is a match for q in $\mathcal{I}$.

This finishes the construction and analysis of the uniform model $\mathcal{J}^u$. It remains to convert $\mathcal{J}^u$ into a model $\mathcal{I}$ of $\mathcal{T}$ and the original ABox $\mathcal{A}$ such that $\mathcal{I} \not\models q$.

- take the disjoint union of the components $\mathcal{J}_a$ of $\mathcal{J}^u$, for each $a \in \mathsf{Ind}(\mathcal{A})$;
- set $a^{\mathcal{I}} = a$ for all $a \in \mathsf{Ind}(\mathcal{A})$;
- add the edge (a, b) to $r^{\mathcal{I}}$ whenever $r(a, b) \in \mathcal{A}$.

It is straightforward to verfiy that $\mathcal{I}$ is a model of $\mathcal{A}$: all role assertions are satisfied by construction of $\mathcal{I}$; moreover, $A(a) \in \mathcal{A}$ implies $A(a) \in \mathcal{A}^u$, whence $a \in A^{\mathcal{J}^u}$ which in turn implies $a \in A^{\mathcal{I}}$ by construction of $\mathcal{I}$. To show that $\mathcal{I}$ is a model of $\mathcal{T}$, we first note that $\mathcal{J}^u$ is self-similar in a way that parallels Claim 1.

Claim 4. For all $\alpha, \beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$ and all $\mathcal{ALCFI}$-concepts C, we have $\alpha \in C^{\mathcal{J}^u}$ iff $\beta \in C^{\mathcal{J}^u}$,

Proof sketch, The proof parallels the one of Claim 1. This time, we define an automorphism ι on the *model* $\mathcal{J}^u$ instead of on the *ABox* $\mathcal{A}^u$. For the elements $\mathsf{Ind}(\mathcal{A}^u) \subseteq \Delta^{\mathcal{J}^u}$, the construction of ι is exactly as in the proof of Claim 1. We can then extend the initial ι to all non-ABox-elements of $\mathcal{J}^u$ exploiting the uniformity of this interpretation. Details are left to the reader.

Next, we show the following.

Claim 5. For every $d \in \Delta^{\mathcal{I}}$ and $\mathcal{ALCFI}$-concept C, we have $d \in C^{\mathcal{J}^u}$ iff $d \in C^{\mathcal{I}}$.

The proof of Claim 5 is by induction on the structure of C. Again, the only interesting case is $C = \exists r.D$, where $r \in \mathsf{N_R} \cup \mathsf{N_R^-}$.

First assume $d \in C^{\mathcal{J}^u}$. Then there is a $(d, e) \in r^{\mathcal{J}^u}$ with $e \in D^{\mathcal{J}^u}$. First assume that the edge (d, e) was added to $r^{\mathcal{J}^u}$ because $d = \alpha$ and $e = \beta$ for some $\alpha, \beta \in \mathsf{Ind}(\mathcal{A}^u)$ and $r(\alpha, \beta) \in \mathcal{A}^u$. Since $d \in \Delta^{\mathcal{I}}$, we must have $d = \alpha = a \in \mathsf{Ind}(\mathcal{A})$. Let $\mathsf{tail}(\beta) = b$. By construction of $\mathcal{A}^u$, $r(a, \beta) \in \mathcal{A}^u$ thus yields $r(a, b) \in \mathcal{A}$ and hence $(a, b) \in r^{\mathcal{I}}$. We are done since Claim 4 and $\beta \in D^{\mathcal{J}^u}$ yields $b \in D^{\mathcal{J}^u}$, which implies $b \in D^{\mathcal{I}}$ by IH.

Now let $d \in C^{\mathcal{I}}$. Then there is a $(d, e) \in r^{\mathcal{I}}$ with $e \in D^{\mathcal{I}}$. First assume that $d = a$ and $e = b$ with $a, b \in \mathsf{Ind}(\mathcal{A})$ and $r(a, b) \in \mathcal{A}$. By construction of $\mathcal{A}^u$, this implies that $r(a, arb) \in \mathcal{A}^u$. Thus $(a, arb) \in r^{\mathcal{J}^u}$ and we are done since IH yields $b \in D^{\mathcal{J}^u}$ abd thus $arb \in D^{\mathcal{J}^u}$ by Claim 4. Now assume that there is an $a \in \mathsf{Ind}(\mathcal{A})$ such that $(d, e) \in r^{\mathcal{J}_a}$. Then the construction of $\mathcal{I}$ yields $(d, e) \in r^{\mathcal{J}^u}$ and we are done since IH yields $e \in D^{\mathcal{I}}$.

By Claim 4, $\mathcal{I}$ satisfies all CIs in $\mathcal{T}$. To show that $\mathcal{I}$ is a model of $\mathcal{T}$, it remains to show that $\mathcal{I}$ satisfies all functionality assertions in $\mathcal{T}$. Thus, let $\mathsf{func}(r) \in \mathcal{T}$. We show that each $d \in \Delta^{\mathcal{J}^u}$ has at most one r-successor in $\mathcal{J}^u$. Distinguish two cases:

- $d \notin \mathsf{Ind}(\mathcal{A})$. Then d has at most one r-successor since $\mathcal{J}^u$ satisfies $\mathsf{func}(r)$ and by construction of $\mathcal{I}$.
- $d = a \in \mathsf{Ind}(\mathcal{A})$. By construction of $\mathcal{I}$ and $\mathcal{A}^u$, a has the same number of r-successors in $\mathcal{I}$ as in $\mathcal{J}^u$. Since $\mathcal{J}^u$ satisfies $\mathsf{func}(r)$, a can have at most one r-successor in $\mathcal{I}$.

It remains to show that $\mathcal{I} \not\models q$. Assume to the contrary of what is to be shown that $\mathcal{I} \models q$. Let $S \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{J}^u}$ be the of pairs (d, e) such that for some $a \in \mathsf{Ind}(\mathcal{A})$ and $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$ with $\mathsf{tail}(\alpha) = a$, $d \in \Delta^{\mathcal{J}_a}$ is the element in the tree interpretation $\mathcal{J}_a$ that corresponds to $e \in \Delta^{\mathcal{J}_\alpha}$ in the isomorphoc tree interpretation $\mathcal{J}_\alpha$. Using the definition of $\mathcal{A}^u$ and $\mathcal{I}$, it can be verified that S is an i-simulation from $\mathcal{I}$ to $\mathcal{J}^u$. We only prove that when $(a, b) \in r^{\mathcal{I}}$ with $a, b \in \mathsf{Ind}(\mathcal{A})$ and $(a, \alpha) \in S$, then there is a β with $(\alpha, \beta) \in r^{\mathcal{J}_u}$ and $(b, \beta) \in S$. To start, note that, by definition of S, we have $\alpha \in \mathsf{Ind}(\mathcal{A}^u)$ and $\mathsf{tail}(\alpha) = a$. From $(a, b) \in r^{\mathcal{I}}$, we obtain $r(a, b) \in \mathcal{A}$ and thus by construction of $\mathcal{A}^u$ there is a $\beta \in \mathsf{Ind}(\mathcal{A}^u)$ with $r(\alpha, \beta) \in \mathcal{A}^u$ and $\mathsf{tail}(\beta) = b$. From $r(\alpha, \beta) \in \mathcal{A}^u$, we obtain $(\alpha, \beta) \in r^{\mathcal{J}^u}$. From $\mathsf{tail}(\beta) = b$, it follows that $(b, \beta) \in S$ as required.

Since matches of ELIQs are preserved under i-simulations and $(a_0, a_0) \in S$, $\mathcal{I} \models q$ implies $\mathcal{J}^u \models q$, which is a contradiction. ❑

E Proofs for Section 6

Proof of Lemma 21

We show Lemma 21 for singleton sets $\Delta^{\mathcal{I}}$. The extension to arbitrary interpretations is straightforward. Thus, let Z be a concept name and z_0, z_1 role names. Let

$$\mathcal{T} = \{\top \sqsubseteq \exists z_0.\top, \top \sqsubseteq \exists z_1.Z\}, \quad H = \forall z_0.\exists z_1.\neg Z.$$

Lemma 34. *For any ABox $\mathcal{A}$ and set $I \subseteq \mathsf{Ind}(\mathcal{A})$, one can construct a model $\mathcal{I}$ of $(\mathcal{T}, \mathcal{A})$ such that $H^{\mathcal{I}} = I$ and $\mathcal{I}$ is hom-initial in $\mathsf{Mod}(\mathcal{T}, \mathcal{A})$.*

Proof. Assume $\mathcal{A}$ and $I \subseteq \mathsf{Ind}(\mathcal{A})$ are given. Denote by $\mathcal{I}_b$ the interpretation based on a binary tree in which every node has one z_0-son and one z_1-son, and every node reachable with z_1 satisfies Z. More precisely, the domain $\Delta^{\mathcal{I}_b}$ of $\mathcal{I}_b$ is the set of words over $\{0, 1\}$, $(\sigma, \sigma 0) \in z_0^{\mathcal{I}_b}$ for all $\sigma \in \Delta^{\mathcal{I}_b}$, $(\sigma, \sigma 1) \in z_1^{\mathcal{I}_b}$ for all $\sigma \in \Delta^{\mathcal{I}_b}$, and $Z^{\mathcal{I}_b} = \{\sigma 1 \mid \sigma \in \Delta^{\mathcal{I}_b}\}$. Now, hook mutually disjoint copies of $\mathcal{I}_b$ to each $a \in \mathsf{Ind}(\mathcal{A})$ (i.e., we identify the root of the copy of $\mathcal{I}_b$ with $a^{\mathcal{I}}$). The resulting interpretation, call it $\mathcal{I}_0$, satisfies $\mathcal{T}$ and $H^{\mathcal{J}_0} = \emptyset$. To satisfy the condition $H^{\mathcal{I}} = I$, we add for all $a \in I$ and d with $(a^{\mathcal{I}_0}, d) \in z_0^{\mathcal{I}_0}$ a new d' to $\mathcal{I}_0$ with $(d, d') \in z_1^{\mathcal{I}}$ and $d' \notin Z^{\mathcal{I}}$. Also, hook a copy of $\mathcal{I}_b$ to d'. The resulting interpretation, $\mathcal{I}$, satisfies $\mathcal{T}$ and we have $H^{\mathcal{I}} = I$. Now let $\mathcal{J}$ be a model of $(\mathcal{T}, \mathcal{A})$. To construct a homomorphism f, we set $f(a^{\mathcal{I}}) = a^{\mathcal{J}}$ for all $a \in \mathsf{Ind}(\mathcal{A})$. Suppose $d \neq a^{\mathcal{I}}$ for any $a \in \mathsf{Ind}(\mathcal{A})$ and $f(d')$ has been defined for the unique z_0 or z_1-predecessor of d. If $(d', d) \in z_0^{\mathcal{I}}$, by $\top \sqsubseteq \exists z_0.\top$, we find e with $(f(d'), e) \in z_0^{\mathcal{J}}$. Set $f(d) = e$. (Observe that $d \notin Z^{\mathcal{I}}$!). If $(d', d) \in z_1^{\mathcal{I}}$, by $\top \sqsubseteq \exists z_1.Z$, we find $e \in Z^{\mathcal{J}}$ with $(f(d'), e) \in z_1^{\mathcal{J}}$. Set $f(d) = e$. One can show that the resulting function f is a homomorphism. ❑

Proof of Theorem 24

Assume $\mathcal{T}$ and $C(a)$ are given. Similarly to Theorem 22, the interpretation $\mathcal{I}_{\mathcal{T},q}$ can be obtained using a standard type-based construction. We use the sets $\mathsf{cl}(\mathcal{T}, C)$, $\mathsf{tp}(\mathcal{T}, C)$, and the relation $\rightsquigarrow_r$ between types as defined in the proof of Theorem 16. We define $\Delta^{\mathcal{I}_{\mathcal{T},q}}$ as the set of all $t \in \mathsf{tp}(\mathcal{T}, C)$ that are satisfiable w.r.t. $\mathcal{T}$ and let $t \in A^{\mathcal{I}_{\mathcal{T},q}}$ iff $A \in t$, for all $A \in \Sigma$, and $(t, t') \in r^{\mathcal{I}_{\mathcal{T},q}}$ iff $t \rightsquigarrow_r t'$, for all $r \in \Sigma$. Finally, $P^{\mathcal{I}_{\mathcal{T},q}} = \{t \in \Delta^{\mathcal{I}_{\mathcal{T},q}} \mid C \notin t\}$. We now show:

1. $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff not $\mathsf{Hom}(\mathcal{I}^{\Sigma}_{\mathcal{A}'}, \mathcal{I}_{\mathcal{T},q})$, where $\mathcal{A}'$ results from $\mathcal{A}$ by adding $P(a)$ to $\mathcal{A}$ and removing all other assertions using P from $\mathcal{A}$;
2. not $\mathsf{Hom}(\mathcal{I}^{\Sigma}_{\mathcal{A}}, \mathcal{I}_{\mathcal{T},q})$ iff $(\mathcal{T}, \mathcal{A}) \models \exists v.(P(v) \wedge C(v))$.

We start by proving (1).

"$\Rightarrow$". Assume $\mathsf{Hom}(\mathcal{I}^{\Sigma}_{\mathcal{A}'}, \mathcal{I}_{\mathcal{T},q})$. Let $h : \mathcal{I}^{\Sigma}_{\mathcal{A}'} \rightarrow \mathcal{I}$ be a witness homomorphism. For each $b \in \mathsf{Ind}(\mathcal{A})$, let $\mathcal{I}_b$ be a copy of $\mathcal{I}_{\mathcal{T},q}$ (with isomophism $h_b : \mathcal{I}_b \rightarrow \mathcal{I}$). Hook each

$\mathcal{I}_b$ to $\mathcal{A}'$ by identifying b with $h(b)$. The resulting interpretation, $\mathcal{H}$, is the disjoint union of all $\mathcal{I}_b$, $b \in \mathsf{Ind}(\mathcal{A})$ together with $(a,b) \in r^{\mathcal{H}}$ whenever $r(a,b) \in \mathcal{A}$ and $r \in \Sigma$. It is readily checked that

- $\bigcup_{b \in \mathsf{Ind}(\mathcal{A})} h_b$ is a $\Sigma \setminus \{P\}$-bisimulation (two-way!) between $\mathcal{H}$ and $\mathcal{I}$.

Thus, for all subconcepts D of $\mathcal{T}$ and C and all $b \in \mathsf{Ind}(\mathcal{A})$: $b \in C^{\mathcal{H}}$ iff $h(b) \in C^{\mathcal{I}_{\mathcal{T},q}}$. We obtain that $\mathcal{H}$ is a model of $\mathcal{T}$ and $\mathcal{A}$. Moreover, $a \notin C^{\mathcal{H}}$ since $h(a) \notin C^{\mathcal{I}_{\mathcal{T},q}}$ and the latter follows because otherwise $h(a) \notin P^{\mathcal{I}_{\mathcal{T},q}}$ and $P(a) \in \mathcal{A}'$ which would contradict that h is a homomorphism. Thus, $(\mathcal{T},\mathcal{A}) \not\models C(a)$.

"$\Leftarrow$". Assume $(\mathcal{T},\mathcal{A}) \not\models C(a)$. Take a witness interpretation $\mathcal{J}$. The type $t(d)$ of $d \in \Delta^{\mathcal{I}}$ is the set of (negated) subconcepts D of C and $\mathcal{T}$ such that $d \in D^{\mathcal{J}}$. The mapping $h : a \mapsto t(a^{\mathcal{J}})$, for $a \in \mathsf{Ind}(\mathcal{A})$ is a homomorphism from $\mathcal{I}^{\Sigma}_{\mathcal{A}'}$ to $\mathcal{I}$. We only consider preservation of P. Assume $P(b) \in \mathcal{A}'$. Then $a = b$. We have $C \notin t(a^{\mathcal{J}})$. Thus $C \notin h(a)$. Hence $h(a) \in P^{\mathcal{I}_{\mathcal{T},q}}$.

Consider (2). The proof is similar.

"$\Leftarrow$". Assume $(\mathcal{T},\mathcal{A}) \not\models \exists v.(P(v) \wedge C(v))$. Take a witness interpretation $\mathcal{J}$. The type $t(d)$ of $d \in \Delta^{\mathcal{I}}$ is the set of (negated) subconcepts D of C and $\mathcal{T}$ such that $d \in D^{\mathcal{J}}$. The mapping $h : a \mapsto t(a^{\mathcal{J}})$, for $a \in \mathsf{Ind}(\mathcal{A})$ is a homomorphism from $\mathcal{I}^{\Sigma}_{\mathcal{A}}$ to $\mathcal{I}$. We only consider preservation of P. Assume $P(b) \in \mathcal{A}$. Then, since $(\mathcal{T},\mathcal{A}) \not\models \exists v.(P(v) \wedge C(v))$, $C \notin t(b^{\mathcal{J}})$. Then $C \notin h(b)$. Hence $h(a) \in P^{\mathcal{I}_{\mathcal{T},q}}$.

"$\Rightarrow$". Assume $\mathsf{Hom}(\mathcal{I}^{\Sigma}_{\mathcal{A}}, \mathcal{I}_{\mathcal{T},q})$. Let $h : \mathcal{I}^{\Sigma}_{\mathcal{A}} \to \mathcal{I}$ be a witness homomorphism. For each $b \in \mathsf{Ind}(\mathcal{A})$, let $\mathcal{I}_b$ be a copy of $\mathcal{I}$ (with isomophism $h_b : \mathcal{I}_b \to \mathcal{I}$). Hook each $\mathcal{I}_b$ to $\mathcal{A}$ by identifying b with $h(b)$. The resulting interpretation, $\mathcal{H}$, is the disjoint union of all $\mathcal{I}_b$, $b \in \mathsf{Ind}(\mathcal{A})$ together with $(a,b) \in r^{\mathcal{H}}$ whenever $r(a,b) \in \mathcal{A}$ and $r \in \Sigma$. For all concepts X that do not occur in $\mathcal{T}$ or C (including, in particular, P), we set $X^{\mathcal{H}} = \{b \in \mathsf{Ind}(\mathcal{A}) \mid X(b) \in \mathcal{A}\}$. It is readily checked that

- $\bigcup_{b \in \mathsf{Ind}(\mathcal{A})} h_b$ is a $\Sigma \setminus \{P\}$-bisimulation (two-way!) between $\mathcal{H}$ and $\mathcal{I}$.

Thus, for all subconcepts D of $\mathcal{T}$ and C and all $b \in \mathsf{Ind}(\mathcal{A})$: $b \in C^{\mathcal{H}}$ iff $h(b) \in C^{\mathcal{I}_{\mathcal{T},q}}$. Thus, $\mathcal{H}$ is a model of $\mathcal{T}$ and $\mathcal{A}$. Moreover, $P^{\mathcal{H}} \cap C^{\mathcal{H}} = \emptyset$: if $d \in P^{\mathcal{H}}$, then $d = b^{\mathcal{J}}$ for some $b \in \mathsf{Ind}(\mathcal{A})$ with $P(b) \in \mathcal{A}$. Thus, $h(b) \in P^{\mathcal{I}_{\mathcal{T},q}}$. But then $h(b) \notin C^{\mathcal{I}_{\mathcal{T},q}}$. Therefore $b \notin C^{\mathcal{H}}$, as required.

It follows that $(\mathcal{T},\mathcal{A}) \not\models \exists v.(P(v) \wedge C(v))$, as required.

F Proofs for Section 7

To formulate the result for FO-rewritability, we introduce a slightly modified version of FO-rewritability that takes into account only those ABoxes that are consistent w.r.t. the TBox.

Definition 35. Let $\mathcal{T}$ be a $\mathcal{ALCFI}$-TBox. Let $\mathcal{Q} \in \{\text{CQ}, \text{PEQ}, \text{ELIQ}, \text{ELQ}\}$. We say that $\mathcal{T}$ is *FO-rewritable for $\mathcal{Q}$ for consistent ABoxes* iff for every $q(\vec{x}) \in \mathcal{Q}$ one can effectively construct a FOQ $q'(\vec{x})$ such that for every ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$, $\mathsf{cert}_{\mathcal{T}}(q,\mathcal{A}) = \{\vec{a} \mid \mathcal{I}_{\mathcal{A}} \models q'(\vec{a})\}$.

Using similar modifications of Definition 2, one can define the obvious notions of $\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ being in PTIME for consistent ABoxes and $\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ being CONP-hard for consistent ABoxes. Theorem 4 still holds for these modified notions. For simplicity, we state the following result for CQs only.

We first prove an extended version of the undecidability result (Theorem 28) and then modify the TBoxes constructed in its proof to show the non-dichomy result (Theorem 27). The modified version of Theorem 28 is as follows:

Theorem 36. *For $\mathcal{ALCF}$-TBoxes $\mathcal{T}$, the following problems are undecidable (Points 1 and 2 are subject to the side condition that* PTIME $\neq$ NP*):*

1. *CQ-answering w.r.t. $\mathcal{T}$ is in* PTIME *(with and w/o restriction to consistent ABoxes);*
2. *CQ answering w.r.t. $\mathcal{T}$ is* CONP*-hard; (with and w/o restriction to consistent ABoxes);*
3. *$\mathcal{T}$ is materializable;*
4. *$\mathcal{T}$ is FO-rewritable for CQ for consistent ABoxes;*

The proofs employ TBoxes that have been introduced in (Baader et al. 2010) to prove the undecidability of the following *emptiness problem*: given an $\mathcal{ALCF}$-TBox $\mathcal{T}$, a signature Σ with $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$ and a concept name A, does there exist a Σ-ABox $\mathcal{A}$ such that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ and $(\mathcal{T},\mathcal{A}) \models \exists v.A(v)$? Note that this problem is of interest only for $A \notin \Sigma$ because otherwise one could clearly take the ABox $\{A(a)\}$.

We start by defining the TBoxes $\mathcal{T}_{\mathfrak{P}}$ constructed in (Baader et al. 2010). An instance of the *finite rectangle tiling problem (FRTP)* is given by a triple $\mathfrak{P} = (\mathfrak{T}, H, V)$ with $\mathfrak{T}$ a non-empty, finite set of *tile types* including an *initial tile* T_{init} to be placed on the lower left corner and a *final tile* T_{final} to be placed on the upper right corner, $H \subseteq \mathfrak{T} \times \mathfrak{T}$ a *horizontal matching relation*, and $V \subseteq \mathfrak{T} \times \mathfrak{T}$ a *vertical matching relation*. A *tiling* for $(\mathfrak{T}, H, V)$ is a map $f : \{0,\ldots,n\} \times \{0,\ldots,m\} \to \mathfrak{T}$ such that $n,m \geq 0$, $f(0,0) = T_{\mathsf{init}}$, $f(n,m) = T_{\mathsf{final}}$, $(f(i,j), f(i+1,j)) \in H$ for all $i < n$, and $(f(i,j), f(i,j+1)) \in v$ for all $i < m$. It is undecidable whether an instance $\mathfrak{P}$ of the FRTP has a tiling. For simplicity, in the following we fix a set $\mathfrak{T} = \{T_1,\ldots,T_p\}$ of tile types and consider instances of the FRTP over $\mathfrak{T}$ only. It is easy to see that the tiling problem is still undecidable if $\mathfrak{T}$ is sufficiently large.

Now let $\Sigma = \{T_1,\ldots,T_p, x, y, x^-, y^-\}$ be a signature consisting of a set $T_1,\ldots,T_p$ of concept names (identical to the tile types) and role names x, y, x^-, and y^- (we are not assuming that x^- and y^- are interpreted as the inverse of $x$and y, respectively). In (Baader et al. 2010), with any $\mathfrak{P} = (\mathfrak{T}, H, V)$ one associates the $\mathcal{ALCF}$-TBox $\mathcal{T}_{\mathfrak{P}}$ containing

$$\mathcal{F} = \{\mathsf{func}(x), \quad \mathsf{func}(y), \quad \mathsf{func}(x^-), \quad \mathsf{func}(y^-)\}$$

and CIs using additional concept names $U, R, L, D, A, Y, I_x, I_y, C, Z_{c,1}, \; Z_{c,2}, \; Z_{x,1}, \; Z_{x,2}, \; Z_{y,1}$. x and y are used to build the rectangle. U and R mark the upper and right border of the rectangle. L and D (for "down") mark the left and lower border of the rectangle. In the following, for $e \in \{c, x, y\}$, we let $\mathcal{B}_e$ range over all Boolean combinations of the concept names $Z_{e,1}$ and $Z_{e,2}$, i.e., over all concepts $L_1 \sqcap L_2$ where L_i is a literal over $Z_{e,i}$, for $i \in \{1, 2\}$. The TBox $\mathcal{T}_\mathfrak{P}$ is defined as the union of $\mathcal{F}$ and the following CIs, where $(T_i, T_j) \in H$ and $(T_i, T_\ell) \in V$:

$$
\begin{array}{rcl}
T_{\mathsf{final}} & \sqsubseteq & Y \sqcap U \sqcap R \\
\exists x.(U \sqcap Y \sqcap T_j) \sqcap I_x \sqcap T_i & \sqsubseteq & U \sqcap Y \\
\exists y.(R \sqcap Y \sqcap T_\ell) \sqcap I_y \sqcap T_i & \sqsubseteq & R \sqcap Y \\
\exists x.(T_j \sqcap Y \sqcap \exists y.Y) & & \\
\sqcap \exists y.(T_\ell \sqcap Y \sqcap \exists x.Y) & & \\
\sqcap I_x \sqcap I_y \sqcap C \sqcap T_i & \sqsubseteq & Y \\
Y \sqcap T_{\mathsf{init}} & \sqsubseteq & A \\
\mathcal{B}_x \sqcap \exists x.\exists x^-.\mathcal{B}_x & \sqsubseteq & I_x \\
\mathcal{B}_y \sqcap \exists y.\exists y^-.\mathcal{B}_y & \sqsubseteq & I_y \\
\exists x.\exists y.\mathcal{B}_c \sqcap \exists y.\exists x.\mathcal{B}_c & \sqsubseteq & C \\
U & \sqsubseteq & \forall y.\bot \\
R & \sqsubseteq & \forall x.\bot \\
U & \sqsubseteq & \forall x.U \\
R & \sqsubseteq & \forall y.R \\
\bigsqcup_{1 \leq s < t \leq p} T_s \sqcap T_t & \sqsubseteq & \bot \\
D & \sqsubseteq & \forall y^-.\bot \\
L & \sqsubseteq & \forall x^-.\bot \\
D & \sqsubseteq & \forall x.D \sqcap \forall x^-.D \\
L & \sqsubseteq & \forall y.L \sqcap \forall y^-.L \\
Y \sqcap T_{\mathsf{init}} & \sqsubseteq & D \sqcap L
\end{array}
$$

We note that the final five inclusions (and the concept names L and D) are not used in (Baader et al. 2010). We use them here to fix the left and lower border of the rectangle. Those inclusions are not required in the present proof, but are used in the non-dichotomy proof below.

Call an ABox $\mathcal{A}$ a $\mathfrak{P}$*-ABox* (with initial node a) iff there is a tiling f for $\mathfrak{P}$ with domain $\{0, \ldots, n\} \times \{0, \ldots, m\}$ and a bijection $f_\mathfrak{P} : \{0, \ldots, n\} \times \{0, \ldots, m\} \to \mathsf{Ind}(\mathcal{A})$ with $f_\mathfrak{P}(0, 0) = a$ such that

- $T_{\mathsf{init}}(f_\mathfrak{P}(0, 0)) \in \mathcal{A}$;
- $T_{\mathsf{final}}(f_\mathfrak{P}(n, m)) \in \mathcal{A}$;
- $T_i(f_\mathfrak{P}(k, j)) \in \mathcal{A}$ iff $T_i = f(k, j)$;
- $x(b_1, b_2) \in \mathcal{A}$ iff $x^-(b_2, b_1) \in \mathcal{A}$ iff $(b_1, b_2) = (f_\mathfrak{P}(k, j), f_\mathfrak{P}(k + 1, j))$
- $y(b_1, b_2) \in \mathcal{A}$ iff $y^-(b_2, b_1) \in \mathcal{A}$ iff $(b_1, b_2) = (f_\mathfrak{P}(k, j), f_\mathfrak{P}(k, j + 1))$

The following is shown in (Baader et al. 2010) (the proof is easily extended to cover the additional concepts for the lower and left border):

Lemma 37. *For every Σ-ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}_\mathfrak{P}$, the following conditions are equivalent:*

- $(\mathcal{T}_\mathfrak{P}, \mathcal{A}) \models \exists v.A(v)$*;*
- $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$ *for a $\mathfrak{P}$-ABox $\mathcal{A}_0$ and a, possibly empty, ABox $\mathcal{A}_1$ with $\mathsf{Ind}(\mathcal{A}_0) \cap \mathsf{Ind}(\mathcal{A}_1) = \emptyset$.*

Observe that the concept name A used in the CQ occurs only once in the TBox, on the right-hand side of a CI. The CI for C enforces confluence, i.e., C is entailed at an individual name a if there is an individual b that is both an x-y-successor and a y-x-successor of a. This is so because, intuitively, $\mathcal{B}_c$ is universally quantified: if confluence fails, we can interpret $Z_{c,1}$ and $Z_{c,2}$ in a way such that neither of the two conjuncts in the precondition of the CI for C is satisfied. In a similar manner, the CI for I_x (resp. I_y) is used to ensure that x^- (resp. y^-) acts as the inverse of x (resp. y) at all points in the rectangle, which means that x (resp. y) is inverse functional within the rectangle. The following characterization of tilings follows directly from Lemma 37.

Lemma 38. *$\mathfrak{P}$ admits a tiling iff there is a Σ-ABox $\mathcal{A}$ that is consistent with $\mathcal{T}_\mathfrak{P}$ and such that $\mathcal{T}_\mathfrak{P}, \mathcal{A} \models \exists v.A(v)$.*

Set $\overline{\Sigma} = \mathsf{sig}(\mathcal{T}_\mathfrak{P}) \setminus \Sigma$. To construct the TBoxes we use for the reduction, replace within the TBoxes $\mathcal{T}_\mathfrak{P}$ all $B \in \overline{\Sigma}$ by the concepts $H_B = \forall r_B.\exists s_B.\neg Z_B$ and add

$$\mathcal{T}_Z = \{\top \sqsubseteq \exists r_B.\top, \top \sqsubseteq \exists s_B.Z_B \mid B \in \overline{\Sigma}\}$$

to $\mathcal{T}_\mathfrak{P}$. Also, add the inclusion $H_A \sqsubseteq B_1 \sqcup B_2$, where B_1, B_2 are fresh concept names, to $\mathcal{T}_\mathfrak{P}$. Denote the resulting TBox by $\mathcal{T}_\mathfrak{P}^\vee$.

For any ABox $\mathcal{A}$, we denote by $\mathcal{A}^\Sigma$ the subset of $\mathcal{A}$ consisting of all assertions in $\mathcal{A}$ that use only symbols from Σ.

Lemma 39. *For any ABox $\mathcal{A}$, $\mathcal{T}_\mathfrak{P}^\vee, \mathcal{A} \models \exists v.H_A(v)$ iff $\mathcal{T}_\mathfrak{P}, \mathcal{A}^\Sigma \models \exists v.A(v)$.*

Proof. The direction from right to left is trivial. Conversely, suppose $(\mathcal{T}_\mathfrak{P}, \mathcal{A}^{\Sigma_\mathfrak{P}}) \not\models \exists v.A(v)$. Take a model $\mathcal{I}$ of $(\mathcal{T}_\mathfrak{P}, \mathcal{A}^\Sigma)$ such that $A^\mathcal{I} = \emptyset$. Since there are no existential restrictions on the right hand side of CIs, we can assume that $\Delta^\mathcal{I} = \{a^\mathcal{I} \mid a \in \mathsf{Ind}(\mathcal{A})\}$. Now set, for $B \in \overline{\Sigma}$, $I_B = \{a \in \mathsf{Ind}(\mathcal{A}) \mid a^\mathcal{I} \in B^\mathcal{I}\}$. Using Lemma 21, we can find a model $\mathcal{I}$ of $(\mathcal{T}'_\mathfrak{P}, \mathcal{A})$ refuting $\exists v.H_A(v)$. ❑

Lemma 40. *Assume $\mathfrak{P}$ does not admit a tiling. Then $\mathcal{T}_\mathfrak{P}^\vee$ is FO-rewritable for consistent ABoxes. Hence $\mathcal{T}_\mathfrak{P}^\vee$ is materializable and CQ-answering w.r.t. $\mathcal{T}_\mathfrak{P}^\vee$ is in* PTIME.

Proof. If $\mathfrak{P}$ does not admit a tiling, then $(\mathcal{T}_\mathfrak{P}, \mathcal{A}^\Sigma) \not\models \exists v.A(v)$, for any ABox $\mathcal{A}$ such that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_\mathfrak{P}$, by Lemma 38. Thus, $(\mathcal{T}_\mathfrak{P}^\vee, \mathcal{A}) \not\models \exists v.H_A(v)$ for any ABox $\mathcal{A}$ such that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_\mathfrak{P}^\vee$, by Lemma 39. But now one can show for any ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}_\mathfrak{P}^\vee$ and any CQ q,

$$(\mathcal{T}_\mathfrak{P}^\vee, \mathcal{A}) \models q \quad \Leftrightarrow \quad (\mathcal{T}_Z, \mathcal{A}) \models q$$

$\mathcal{T}_Z$ is FO-rewritable. Thus, $\mathcal{T}_\mathfrak{P}^\vee$ is FO-rewritable for consistent ABoxes. ❑

Lemma 41. *Assume $\mathfrak{P}$ admits a tiling. Then $\mathcal{T}_\mathfrak{P}^\vee$ is not materializable. Thus, $\mathcal{T}_\mathfrak{P}^\vee$ is not FO-rewritable for consistent ABoxes and CQ-answering w.r.t. $\mathcal{T}$ is* CONP*-hard.*

Proof. Let $\mathcal{A}$ be a Σ-ABox such that $(\mathcal{T}_{\mathfrak{P}}, \mathcal{A}) \models \exists v.A(v)$ and $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_{\mathfrak{P}}$. Then $(\mathcal{T}_{\mathfrak{P}}^{\vee}, \mathcal{A}) \models \exists v.(B_1(v) \vee B_2(v))$ and $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_{\mathfrak{P}}^{\vee}$. It is readily checked that $(\mathcal{T}_{\mathfrak{P}}^{\vee}, \mathcal{A}) \not\models \exists v.B_1(v)$ and $(\mathcal{T}_{\mathfrak{P}}^{\vee}, \mathcal{A}) \not\models \exists v.B_2(v)$. Thus, $\mathcal{T}_{\mathfrak{P}}^{\vee}$ is not materializable. ❑

From Lemmas 40 and 41, we obtain Points 3 and 4 of Theorem 36 as well as Points 1 and 2 for consistent ABoxes. Thus, to prove Theorem 36 it remains to show the following lemma.

Lemma 42. *Consistency of ABoxes w.r.t. $\mathcal{T}_{\mathfrak{P}}^{\vee}$ can be decided in polynomial time (in the size of the ABox).*

Proof. Assume $\mathcal{A}$ is given. Form $\mathcal{A}^{\Sigma}$ and apply the following rules exhaustively:

- add $I_x(a)$ to $\mathcal{A}^{\Sigma}$ if there exists b with $x(a,b), x^-(b,a) \in \mathcal{A}$;
- add $I_y(a)$ to $\mathcal{A}^{\Sigma}$ if there exists b with $y(a,b), y^-(b,a) \in \mathcal{A}$;
- add $C(a)$ to $\mathcal{A}^{\Sigma}$ if there exist a_1, a_2, b with $x(a,a_1)$, $y(a,a_2)$, $y(a_1,b)$, $x(a_2,b) \in \mathcal{A}$.

Denote the resulting ABox by $\mathcal{A}'$. Now remove the three inclusion schemata involving the Boolean combinations $\mathcal{B}$ from $\mathcal{T}_{\mathfrak{P}}$ and denote by $\mathcal{T}$ the resulting TBox. One can show that $(\mathcal{T}_{\mathfrak{P}}^{\vee}, \mathcal{A})$ is consistent iff $(\mathcal{T}, \mathcal{A}')$ is consistent. The consistency of the latter can be checked in polynomial time since $\mathcal{T}$ is a Horn-$\mathcal{ALCF}$-TBox. ❑

We now come to the proof of Theorem 27.

Theorem 27 For every language $L \in$ coNP there exists a $\mathcal{ALCF}$-TBox $\mathcal{T}$ and query $\mathsf{rej}(a)$, rej a concept name, such that the following holds:

- there is a polynomial reduction of L to answering $\mathsf{rej}(a)$ w.r.t. $\mathcal{T}$;
- for every Boolean ELIQ q, answering q w.r.t. $\mathcal{T}$ is polynomially reducible to L.

Consider a non-deterministic TM $M = (Q, \Sigma, \Delta, q_0, q_a, q_r)$ with Q a finite set of states, Σ a finite alphabet, $q_0 \in Q$ a starting state, $\Delta \subseteq Q \times \Sigma \times Q \times \Sigma \times \{L, R\}$ the a transition relation, and $q_a, q_r \in Q$ the accepting and rejecting states. We assume that for any input $v \in \Sigma^*$, M halts after exactly $|v|^k$ steps in the accepting or rejecting state and that it uses exactly n^k cells for the computation. Denote by $L(M)$ the language accepted by M and assume that $L = \Sigma^* \setminus L(M)$.

The ABoxes we use to simulate input words $v \in \Sigma^*$ are $m_1 \times m_2$ grids in which T_{init} is written in the lower left corner followed by the the word v, T_{final} is written in the upper right corner, and B (for blank) is written everywhere else. In our construction of $\mathcal{T}$ we first build a TBox that "checks" whether the input ABox is of this form.

To define this part of the TBox, we re-use the above TBox $\mathcal{T}_{\mathfrak{P}}$, where $\mathfrak{P} = (\mathfrak{T}, H, V)$ with $\mathfrak{T} = \{B, T_{\mathsf{final}}, T_{\mathsf{init}}\} \cup \Sigma$ and H consisting of all pairs in $\mathfrak{T} \times \mathfrak{T}$ except

- (B, σ) for $\sigma \in \Sigma$,
- $(\sigma, T_{\mathsf{final}})$ for $\sigma \in \Sigma$,
- (T_{final}, T), (T, T_{init}), for $T \in \mathfrak{T}$,

and V consisting of all pairs in $\mathfrak{T} \times \mathfrak{T}$ except

- (B, σ) for $\sigma \in \Sigma$,
- (σ_1, σ_2) for $\sigma_1, \sigma_2 \in \Sigma$,
- $(\sigma, T_{\mathsf{final}})$ for $\sigma \in \Sigma$,
- (T_{final}, T), (T, T_{init}), for $T \in \mathfrak{T}$.

For any $n, m \geq 2$, and any word $v \in L^*$ there is exactly one tiling f for $\mathfrak{P}$. That tiling places T_{init} in the lower left corner followed by the the word v, T_{final} in the upper right corner, and B is written everywhere else. Thus, every $\mathfrak{P}$-ABox $\mathcal{A}$ (with initial node a) is isomorphic to some $n \times m$-grid with a word $T_{\mathsf{init}}v$ ($v \in L^*$) written in the lower left corner. We call this ABox the *grid-ABox for the $n \times m$-rectangle with word v*. Set

$$\mathcal{T}_{\mathsf{grid}} := \mathcal{T}_{\mathfrak{P}}, \quad \mathcal{T}_{\mathsf{grid}}^{\mathsf{SO}} := \mathcal{T}_{\mathfrak{P}}^{\vee} \setminus \{H_A \sqsubseteq B_1 \sqcup B_2\}.$$

Recall that $\mathcal{T}_{\mathsf{grid}}^{\mathsf{SO}}$ contains the inclusions $\mathcal{T}_Z$ for "second-order variables".

To encode the computation of the TM M we use the following set $\mathcal{Z}_M$ of inclusions. Intuitively, assume that a grid-ABox with initial node a for the $n \times m$-rectangle with word v is given. Then $(\mathcal{T}_{\mathsf{grid}}^{\mathsf{SO}}, \mathcal{A}) \models H_A(a)$. We introduce a concept name H_{grid} denoting all individual names in $\mathcal{A}$:

$$H_A \sqsubseteq H_{\mathsf{grid}}, \quad H_{\mathsf{grid}} \sqsubseteq \forall r.H_{\mathsf{grid}}$$

for all $r \in \{x, y, x^-, y^-\}$. The remaining inclusions are all relativized to H_{grid}. The remaining inclusions use

- concept names $q \in Q$ that indicate the state of the TM in the computation;
- concept names $\sigma \in \Sigma$ for the input word;
- concept names A_σ, $\sigma \in \Sigma$, for symbols written during the computation (and as copies of the symbols of the input word);
- a concept name H for the head of the TM.

We simulate the instructions of M by taking for $(q, \sigma, q') \in Q \times \Sigma \times Q$:

$$H_{\mathsf{grid}} \sqcap H \sqcap q \sqcap A_\sigma \sqsubseteq \bigsqcup_{(q,\sigma,q',\sigma',L)\in\Delta} \exists y.(A_{\sigma'} \sqcap q' \sqcap \neg H \sqcap \forall x.\neg H \sqcap \exists x^-.H) \sqcup \bigsqcup_{(q,\sigma,q',\sigma',R)\in\Delta} \exists y.(A_{\sigma'} \sqcap q' \sqcap \neg H \sqcap \forall x^-.\neg H \sqcap \exists x.H)$$

We state that cells can only change where H is:

$$H_{\mathsf{grid}} \sqcap \neg H \sqcap A_\sigma \sqsubseteq \forall y.A_\sigma, \quad H_{\mathsf{grid}} \sqcap \neg H \sqcap \neg A_\sigma \sqsubseteq \forall y.\neg A_\sigma$$

We state that H cannot be introduced without a corresponding computation step:

$$H_{\mathsf{grid}} \sqcap \neg H \sqcap \forall x^-.\neg H \sqcap \forall x.\neg H \sqsubseteq \forall y.\neg H.$$

We state that, when M starts, it is in state q_0 and that the head is at the first cell:

$$T_{\mathsf{init}} \sqcap H_{\mathsf{grid}} \sqsubseteq q_0, \quad T_{\mathsf{init}} \sqcap H_{\mathsf{grid}} \equiv \exists x.H \sqcap \forall y^-.\bot \sqcap H_{\mathsf{grid}}.$$

We state that every state q is uniform over each step of the computation:

$$q \sqcap H_{\mathsf{grid}} \sqsubseteq \forall x.q \sqcap \forall x^-.q.$$

We state that A_σ is true where σ from the input word is true:

$$H_{\mathsf{grid}} \sqcap \sigma \equiv H_{\mathsf{grid}} \sqcap \forall y^-.\bot \sqcap A_\sigma,$$

for $\sigma \in \Sigma$. We close with

$$H_{\mathsf{grid}} \sqcap A_\sigma \sqcap A_{\sigma'} \sqsubseteq \bot, \quad H_{\mathsf{grid}} \sqcap q \sqcap q' \sqsubseteq \bot,$$

for $\sigma \neq \sigma'$ and $q \neq q'$, and the assertion that rej is true everywhere in the ABox if the machine reaches the rejecting state:

$$H_{\mathsf{grid}} \sqcap q_r \sqsubseteq \mathsf{rej}, \quad H_{\mathsf{grid}} \sqcap \mathsf{rej} \sqsubseteq \forall r.\mathsf{rej}$$

for $r \in \{x, y, x^-, y^-\}$. This finishes the definition of $\mathcal{Z}_M$. As before, we replace every concept name

$$B \in X := Q \cup \{A_\sigma \mid \sigma \in \Sigma\} \cup \{H_{\mathsf{grid}}, H\}$$

by $H_B = \forall r_B.\exists s_B.\neg Z_B$, add

$$\mathcal{T}_{Z,1} = \{\top \sqsubseteq \exists r_B.\top, \top \sqsubseteq \exists s_B.Z_B \mid B \in X\}$$

to $\mathcal{Z}_M$ and denote the resulting TBox by $\mathcal{Z}_M^{\mathsf{SO}}$. We set $\mathcal{T}_M^{\mathsf{SO}} = \mathcal{T}_{\mathsf{grid}}^{\mathsf{SO}} \cup \mathcal{Z}_M^{\mathsf{SO}}$. Note that the only "real" concept names in $\mathcal{T}_M^{\mathsf{SO}}$ are $\mathfrak{T}$ and rej. The following lemma is straightforward now and proves Part 1 of Theorem 27.

Lemma 43. *If $\mathcal{A}$ is the grid-ABox for the $m_1 \times m_2$-rectangle with word v and $m_1, m_2 \geq n^k$ for $n = |v|$, then $(\mathcal{T}_M^{\mathsf{SO}}, \mathcal{A}) \models \mathsf{rej}(a)$ iff $v \notin L(M)$.*

By Lemma 43, to check $v \notin L(M)$, it sufficient to construct the grid-ABox for the $n^k \times n^k$-rectangle with word v and then decide $(\mathcal{T}_M^{\mathsf{SO}}, \mathcal{A}) \models \mathsf{rej}(a)$. Thus, we have shown that there exists a polynomial reduction of deciding $v \in L$ to answering $\mathsf{rej}(a)$ w.r.t. $\mathcal{T}_M^{\mathsf{SO}}$.

We now show that for every ELIQ $C(f)$, answering $C(f)$ w.r.t. $\mathcal{T}_M^{\mathsf{SO}}$ can be polynomially reduced to deciding $v \in L$. Assume $C(f)$ is given. Consider an ABox $\mathcal{A}$.

Claim 1. It can be checked in polytime (in the size of $\mathcal{A}$) whether $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_M^{\mathsf{SO}}$.

Observe that $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}_M^{\mathsf{SO}}$ iff

- $\mathcal{A}$ contains a grid-ABox for a $m_1 \times m_2$-rectangle with word v and $m_1 < n^k$ or $m_2 < n^k$ for $n = |v|$; or
- $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}_{\mathsf{grid}}^{\mathsf{SO}}$.

The first condition can clearly be checked in polytime and the latter is in PTIME by Lemma 42.

Now, if $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}_M^{\mathsf{SO}}$, then one of the following two cases applies:

- f is in a grid-ABox for the $m_1 \times m_2$-rectangle with word v and $m_1, m_2 \geq n^k$ for $n = |v|$ (there can be other disjoint components). In that case $(\mathcal{T}_M^{\mathsf{SO}}, \mathcal{A}) \models C(f)$ iff $(\mathcal{T}_V, \mathcal{A}') \models C(f)$, where
  - $\mathcal{A}'$ is defined by setting $\mathcal{A}' = \mathcal{A} \cup \{\mathsf{rej}(b) \mid b \in \mathsf{Ind}(\mathcal{A})\}$ if $v \notin L(M)$; and $\mathcal{A}' := \mathcal{A}$ otherwise.
  - $\mathcal{T}_V = \mathcal{T}_Z \cup \mathcal{T}_{Z,1}$.

  Both conditions can be checked in polytime.
- f is not in a grid-ABox for the $m_1 \times m_2$-rectangle with word v. In that case $(\mathcal{T}_M^{\mathsf{SO}}, \mathcal{A}) \models C(f)$ iff $(\mathcal{T}_Z, \mathcal{A}) \models C(f)$. The latter condition can be checked in polytime.