Non-Uniform Data Complexity of Query Answering in Description Logics

Carsten Lutz[1] and Frank Wolter[2]

[1] Department of Computer Science, University of Bremen, Germany
[2] Department of Computer Science, University of Liverpool, UK
clu@uni-bremen.de,Wolter@liverpool.ac.uk

1 Introduction

In recent years, the use of ontologies to access instance data has become increasingly popular. The general idea is that an ontology provides a vocabulary or conceptual model for the application domain, which can then be used as an interface for querying instance data and to derive additional facts. In this emerging area, called ontology-based data access (OBDA), 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 [12] and thus intractable (when speaking of complexity, we *always* mean data complexity). 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 and have minimal models that can be exploited for PTIME query answering. Horn DLs in this sense include, for example, 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 [12]. While CQ-answering in Horn-$\mathcal{SHIQ}$ and the $\mathcal{EL}$ family of DLs is also hard for PTIME, the problem has even lower complexity in DL-Lite. In fact, the design goal of DL-Lite was to achieve *FO-rewritability*, i.e., that any CQ q and TBox $\mathcal{T}$ can be rewritten into an FO query q' such that the answers to q w.r.t. $\mathcal{T}$ coincide with the answers that a standard database system produces for q' [6]. Achieving this goal requires CQ-answering to be in AC^0.

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 syntactically through 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 advocate a *non-uniform* study of the complexity of query answering by considering data complexity on the *level of individual TBoxes*. For a TBox $\mathcal{T}$, we say that *CQ-answering w.r.t. $\mathcal{T}$ is in* PTIME if for every CQ q, there is a PTIME algorithm that, given an ABox $\mathcal{A}$, computes the answers to q in $\mathcal{A}$ w.r.t. $\mathcal{T}$. In a similar way, we can define coNP-hardness and FO-rewritability on the TBox level. The non-uniform perspective allows us to investigate more fine-grained questions regarding the data complexity of query answering such as: given an expressive DL $\mathcal{L}$ such as $\mathcal{ALC}$ or $\mathcal{SHIQ}$, how can one characterize those $\mathcal{L}$-TBoxes $\mathcal{T}$ for which CQ-answering is in PTIME? How can we do the same for FO-rewritability? Is there a dichotomy for the complexity of query answering w.r.t. TBoxes formulated in $\mathcal{L}$, such as: for any $\mathcal{L}$-TBox $\mathcal{T}$, CQ-answering w.r.t. $\mathcal{T}$ is either in PTIME or coNP-hard?

In this paper, we consider TBoxes formulated in the expressive DL $\mathcal{ALCFI}$, answer some of the above questions, and take some steps towards others. Our main results are:

1. there is a dichotomy between PTIME and coNP-complete for CQ-answering w.r.t. $\mathcal{ALC}$-TBoxes if, and only if, Feder and Vardi's dichotomy conjecture that "constraint satisfaction problems (CSPs) with finite template are in PTIME or NP-complete" [10] is true; the same holds for $\mathcal{ALCI}$-TBoxes;
2. there is no dichotomy between PTIME and coNP-complete for CQ-answering w.r.t. $\mathcal{ALCF}$-TBoxes, unless PTIME $=$ NP; moreover, PTIME-complexity of CQ answering and many related problems are undecidable for $\mathcal{ALCF}$.
3. there is a dichotomy between PTIME and coNP-complete for CQ-answering w.r.t. $\mathcal{ALCFI}$-TBoxes of depth one, i.e., TBoxes where concepts have role depth ≤ 1;
4. FO-rewritability is decidable for Horn-$\mathcal{ALCFI}$-TBoxes of depth two and all Horn-$\mathcal{ALCF}$-TBoxes;

It should be noted that there has been steady progress regarding the dichotomy conjecture of Feder and Vardi over the last fifteen years and though the problem is still open, a solution does not seem completely out of reach [4, 5]. Our proof of Point 1 is based on a novel connection between CSPs and query answering w.r.t. $\mathcal{ALCI}$-TBoxes that can be exploited to transfer numerous results from the CSP world to query answering w.r.t. $\mathcal{ALCI}$-TBoxes and related problems. For example, together with [16, 5] we obtain the following results on 'FO-rewritability of ABox consistency':

5. Given an $\mathcal{ALCI}$-TBox $\mathcal{T}$, it can be decided in NEXPTIME whether there is an FO-sentence $\varphi_{\mathcal{T}}$ such that for all ABoxes $\mathcal{A}$, $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ iff $\mathcal{A}$ viewed as an FO-structure satisfies $\varphi_{\mathcal{T}}$. Moreover, such a sentence $\varphi_{\mathcal{T}}$ exists iff ABox consistency w.r.t. $\mathcal{T}$ can be decided in non-uniform AC^0. Finally, if no such sentence $\varphi_{\mathcal{T}}$ exists, then ABox consistency w.r.t. $\mathcal{T}$ is LOGSPACE-hard (under FO-reductions).

To prove our results, we introduce some new notions that are relevant for studying the questions raised and prove some additional results of general interest. A central such notion is *materializability* of a TBox $\mathcal{T}$, which formalizes the existence of minimal models as known from Horn-DLs. We show that, in the case of TBoxes of depth one, materializability characterizes PTIME CQ-answering, which allows us to establish Point 2 above. For TBoxes of unrestricted depth, non-materializability still provides a sufficient condition for coNP-hardness of CQ-answering. We also develop the notion of *unraveling tolerance* of a TBox $\mathcal{T}$, which provides a sufficient condition for query answering to be in PTIME. The resulting upper bound strictly generalizes the known result that CQ-answering in Horn-$\mathcal{ALCFI}$ is in PTIME. Our framework also allows to formally establish some common intuitions and beliefs held in the context of CQ-answering in description logics. For example, we show that for any $\mathcal{ALCFI}$-TBox $\mathcal{T}$, CQ-answering is in PTIME iff answering positive existential queries is in PTIME iff answering $\mathcal{ELI}$-instance queries is in PTIME and likewise for FO-rewritability. Another observation in this spirit is that an $\mathcal{ALCFI}$-TBox is materializable (has minimal models) iff it is convex (a notion related to the entailment of disjunctions).

Most proofs in this paper are deferred to the (appendix of the) long version, which is available at http://www.csc.liv.ac.uk/~frank/publ/publ.html.

2 Preliminaries

We use standard notation for the syntax and semantics of $\mathcal{ALCFI}$ and other well-known DLs. Our TBoxes are finite sets of concept inclusions $C \sqsubseteq D$, where C and D are potentially compound concepts, and functionality assertions $\mathsf{func}(r)$, where r is a potentially inverse role. ABoxes are finite sets of assertions $A(a)$ and $r(a,b)$ with A a concept *name* and r a role name. 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 $r(b,a) \in \mathcal{A}$. For the interpretation of individual names, we make the unique name assumption.

A *first-order query (FOQ)* $q(\boldsymbol{x})$ is a first-order formula with free variables $\boldsymbol{x}$ constructed from atoms $A(t)$, $r(t,t')$, and $t = t'$ (where t,t' range over individual names and variables) using negation, conjunction, disjunction, and existential quantification. The variables in $\boldsymbol{x}$ are the *answer variables* of q. A FOQ without answer variables is *Boolean*. We say that a tuple $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$ is an *answer to* $q(\boldsymbol{x})$ *in an interpretation* $\mathcal{I}$ if $\mathcal{I} \models q[\boldsymbol{a}]$, where $q[\boldsymbol{a}]$ results from replacing the answer variables $\boldsymbol{x}$ in $q(\boldsymbol{x})$ with $\boldsymbol{a}$. A tuple $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$ is a *certain answer to* $q(\boldsymbol{x})$ *in* $\mathcal{A}$ *given* $\mathcal{T}$, in symbols $\mathcal{T},\mathcal{A} \models q(\boldsymbol{a})$, if $\mathcal{I} \models q[\boldsymbol{a}]$ for all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$. Set $\mathsf{cert}_{\mathcal{T}}(q,\mathcal{A}) = \{\boldsymbol{a} \mid \mathcal{T},\mathcal{A} \models q(\boldsymbol{a})\}$. A *positive existential query (PEQ)* $q(\boldsymbol{x})$ is a FOQ without negation and equality and a *conjunctive query (CQ)* is a positive existential query without disjunction. If C is an $\mathcal{ELI}$-concept and $a \in \mathsf{N_I}$, then $C(a)$ is an *$\mathcal{ELI}$-query (ELIQ)*. *$\mathcal{EL}$-queries (ELQs)* are defined analogously. Note that $\mathcal{ELI}$-queries and $\mathcal{EL}$-queries are always Boolean. In what follows, we sometimes slightly abuse notation and use FOQ to denote the set of all first-order queries, and likewise for CQ, PEQ, ELIQ, and ELQ.

Definition 1. *Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox. Let $\mathcal{Q} \in \{CQ, PEQ, ELIQ, ELQ\}$. Then*

- $\mathcal{Q}$-answering w.r.t. $\mathcal{T}$ is in PTIME *if for every $q(\boldsymbol{x}) \in \mathcal{Q}$, there is a polytime algorithm that computes, given an ABox $\mathcal{A}$, the answer* $\mathsf{cert}_{\mathcal{T}}(q,\mathcal{A})$*;*
- $\mathcal{Q}$-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 $\mathcal{T},\mathcal{A} \models q$;*
- $\mathcal{T}$ *is* FO-rewritable for $\mathcal{Q}$ *iff for every $q(\boldsymbol{x}) \in \mathcal{Q}$ one can effectively construct an FO-formula $q'(\boldsymbol{x})$ such that for every ABox $\mathcal{A}$,* $\mathsf{cert}_{\mathcal{T}}(q,\mathcal{A}) = \{\boldsymbol{a} \mid \mathcal{I}_{\mathcal{A}} \models q'(\boldsymbol{a})\}$*, where $\mathcal{I}_{\mathcal{A}}$ denotes $\mathcal{A}$ viewed as an interpretation.*

The above notions of complexity are rather robust under changing the query language: as we show next, neither the PTIME bounds nor FO-rewritability depend on whether we consider PEQs, CQs, or ELIQs.

Theorem 1. *For all $\mathcal{ALCFI}$-TBoxes $\mathcal{T}$, the following equivalences hold:*

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 $\mathcal{ALCF}$-TBox, then we can replace ELIQ in Points 1 and 2 with ELQ.

The proof is based on Theorems 2 and 3 below. Theorem 1 allows us to (sometimes) speak of the 'complexity of query answering' without reference to a query language.

3 Materializability

An important tool we use for analyzing the complexity of query answering is the notion of materializability of a TBox $\mathcal{T}$, which means that computing the certain answers to any query q and ABox $\mathcal{A}$ w.r.t. $\mathcal{T}$ reduces to evaluating q in a single model of $\mathcal{A}$ and $\mathcal{T}$.

Definition 2. *Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox and $\mathcal{Q} \in \{CQ, PEQ, ELIQ, ELQ\}$. $\mathcal{T}$ is $\mathcal{Q}$-*materializable *if for every ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$, there exists a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $\mathcal{I} \models q[\boldsymbol{a}]$ iff $\mathcal{T}, \mathcal{A} \models q(\boldsymbol{a})$ for all $q(\boldsymbol{x}) \in \mathcal{Q}$ and $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$.*

We show that PEQ, CQ, and ELIQ-materializability coincide (and for $\mathcal{ALC}$-TBoxes, all these also coincide with ELQ-materializability). Materializability is also equivalent to the following disjunction property (sometimes also called *convexity*): a TBox $\mathcal{T}$ has the *ABox disjunction property* if for all ABoxes $\mathcal{A}$ and ELIQs $C_1(a_1), \ldots, C_n(a_n)$, from $\mathcal{T}, \mathcal{A} \models C_1(a_1) \vee \ldots \vee C_n(a_n)$ it follows that $\mathcal{T}, \mathcal{A} \models C_i(a_i)$, for some $i \leq n$.

Theorem 2. *Let $\mathcal{T}$ be an $\mathcal{ALCFI}$-TBox. The following equivalences hold: $\mathcal{T}$ is PEQ-materializable iff $\mathcal{T}$ is CQ-materializable iff $\mathcal{T}$ is ELIQ-materializable iff $\mathcal{T}$ has the ABox disjunction property.*

If $\mathcal{T}$ is an $\mathcal{ALC}$-TBox, the above are equivalent to ELQ-materializability.

Because of Theorem 2, we sometimes use the term materializability without reference to a query language. We call an interpretation $\mathcal{I}$ that satisfies the condition formulated in Definition 2 for PEQs a *minimal model* of $\mathcal{T}$ and $\mathcal{A}$. Note that in many cases, only an infinite minimal models exists. For example, for $\mathcal{T} = \{A \sqsubseteq \exists r.A\}$ and $\mathcal{A} = \{A(a)\}$ every minimal model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ comprises an infinite r-chain starting at $a^{\mathcal{I}}$. Every TBox that is equivalent to an FO Horn sentence (in the general sense of [7]) is materializable: to construct a minimal model for such a TBox $\mathcal{T}$ and some ABox $\mathcal{A}$, one can take the direct product of all at most countable models of $\mathcal{T}$ and $\mathcal{A}$ (for additional information on direct products in DLs, see [17]). Conversely, however, there are simple materializable TBoxes that are not equivalent to FO Horn sentences.

Example 1. Let $\mathcal{T} = \{\exists r.(A \sqcap \neg B \sqcap \neg E) \sqsubseteq \exists r.(\neg A \sqcap \neg B \sqcap \neg E)\}$. One can easily show that $\mathcal{T}$ is not preserved under direct products; thus, it is not equivalent to a Horn sentence. However, one can construct a minimal model $\mathcal{I}$ for $\mathcal{T}$ and any ABox $\mathcal{A}$ by taking the interpretation $\mathcal{I}_{\mathcal{A}}$ obtained by viewing $\mathcal{A}$ as an interpretation and then adding, for any $a \in \mathsf{Ind}(\mathcal{A})$ with $a \in (\exists r.(A \sqcap \neg B \sqcap \neg E))^{\mathcal{I}_\mathcal{A}}$, a fresh d_a such that $(a, d_a) \in r^\mathcal{I}$ and d_a is not in the extension of any concept name. PEQ-answering w.r.t. $\mathcal{T}$ is FO-rewritable since for any PEQ q, $\mathsf{cert}_\mathcal{T}(q, \mathcal{A})$ consists of precisely the answers to q in $\mathcal{I}_\mathcal{A}$ (i.e., no query rewriting is necessary). Thus, PEQ-answering w.r.t. $\mathcal{T}$ is also in PTIME.

We show that materializability is a necessary condition for query answering being in PTIME.

Theorem 3. *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 uses the violation of the ABox disjunction property stated in Theorem 2 and generalizes the reduction of 2+2-SAT used in [19] to prove that instance checking in a variant of $\mathcal{EL}$ is coNP-hard.

Materializability is not a sufficient condition for query answering to be in PTIME. In fact, we show that for any non-uniform constraint satisfaction problem, there is a materializable $\mathcal{ALC}$-TBox for which Boolean CQ-answering has the same complexity, up to complementation of the complexity class. For two finite relational FO-structures $\mathcal{R}$ and $\mathcal{R}'$ over relation symbols Σ, we write $\mathsf{Hom}(\mathcal{R}', \mathcal{R})$ if there is a homomorphism from $\mathcal{R}'$ to $\mathcal{R}$. The non-uniform constraint satisfaction problem for $\mathcal{R}$, denoted by $\mathsf{CSP}(\mathcal{R})$, is the problem to decide, for every finite $\mathcal{R}'$ over Σ, whether $\mathsf{Hom}(\mathcal{R}', \mathcal{R})$. Numerous algorithmic problems, among them many NP-complete ones such as k-SAT and k-colourability of graphs, can be given in the form $\mathsf{CSP}(\mathcal{R})$. It is known that every problem of the form $\mathsf{CSP}(\mathcal{R})$ is polynomially equivalent to some $\mathsf{CSP}(\mathcal{R}')$ with $\mathcal{R}'$ a digraph [10]. Thus, in what follows we can restrict ourselves to considering CSPs of the form $\mathsf{CSP}(\mathcal{I})$, where $\mathcal{I}$ is a DL interpretation. A *signature* Σ is a set of concept and role names. The signature $\mathsf{sig}(\mathcal{T})$ of a TBox $\mathcal{T}$ is the set of concept and role names that occur in $\mathcal{T}$. A Σ-TBox is a TBox that uses symbols from Σ only. Similar notation is used for ABoxes, concepts, and interpretations. For an ABox $\mathcal{A}$, we denote by $\mathcal{A}^\Sigma$ the subset of $\mathcal{A}$ containing symbols from Σ only. We will often not distinguish between ABoxes and finite interpretations.

Theorem 4. *For every non-uniform constraint satisfaction problem $\mathsf{CSP}(\mathcal{I})$, one can compute in polytime a materializable $\mathcal{ALC}$-TBox $\mathcal{T}$ such that for all ABoxes $\mathcal{A}$,*

1. *$\mathsf{Hom}(\mathcal{A}^\Sigma, \mathcal{I})$, with $\Sigma = \mathsf{sig}(\mathcal{I})$, iff $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$;*
2. *for any Boolean CQ q, answering q w.r.t. $\mathcal{T}$ is polynomially reducible to the complement of $\mathsf{CSP}(\mathcal{I})$.*

The proof Theorem 4 relies on the existence of $\mathcal{ALC}$-concepts H whose value $H^\mathcal{I}$ in interpretations $\mathcal{I}$ cannot be detected directly using CQs, but which can be used in a TBox to influence the values $A^\mathcal{I}$ of concept names A and, therefore, have an indirect effect on the answers to CQs. From the viewpoint of CQ query answering, they thus behave similarly to second-order variables. More precisely, let, for a finite set V of indices, Z_v, r_v, s_v be concept and role names, respectively. Let

$$\mathcal{T}_V = \{\top \sqsubseteq \exists r_v.\top, \top \sqsubseteq \exists s_v.Z_v \mid v \in V\}, \quad H_v = \forall r_v.\exists s_v.\neg Z_v.$$

Lemma 1. *For any ABox $\mathcal{A}$ and sets $I_v \subseteq \mathsf{Ind}(\mathcal{A})$, $v \in V$, one can construct a minimal model $\mathcal{I}$ of $(\mathcal{T}_V, \mathcal{A})$ such that $H_v^\mathcal{I} = I_v$ for all $v \in V$. $\mathcal{T}_V$ is FO-rewritable for PEQ.*

To prove Theorem 4, one extends the TBox $\mathcal{T}_V$. Assume $\mathsf{CSP}(\mathcal{I})$ is given. Let $V = \Delta^{\mathcal{I}}$ and assume, for simplicity, that $\mathsf{sig}(\mathcal{I}) = \{r\}$. Define

$$\mathcal{T} = \mathcal{T}_V \cup \{H_v \sqcap \exists r.H_w \sqsubseteq \bot \mid v, w \in V, (v, w) \notin r^{\mathcal{I}}\} \cup \\ \{H_v \sqcap H_w \sqsubseteq \bot \mid v, w \in V, v \neq w\} \cup \{\bigsqcap_{v \in V} \neg H_v \sqsubseteq \bot\}$$

Based on Lemma 1, it is possible to verify Points 1 and 2 of Theorem 4. For Point 2, it can be seen that for all Boolean CQs q and ABoxes $\mathcal{A}$, $(\mathcal{T}, \mathcal{A}) \models q$ iff $(\mathcal{T}_V, \mathcal{A}) \models q$ or not $\mathsf{Hom}(\mathcal{A}^{\Sigma}, \mathcal{I})$; since $\mathcal{T}_V$ is FO-rewritable, the former can be checked in PTIME.

4 (Towards) Dichotomies

We start with a reduction of Boolean CQ-answering w.r.t. $\mathcal{ALCI}$-TBoxes to CSPs that yields, together with Theorem 4, a proof of Point 1 in the introduction: the dichotomy problem for CSPs is equivalent to the dichotomy problem for CQ answering w.r.t. $\mathcal{ALC}$- (and $\mathcal{ALCI}$-) TBoxes.

Theorem 5. *Let $\mathcal{T}$ be an $\mathcal{ALCI}$-TBox and $C(a)$ an ELIQ. Then one can construct, in time exponential in $|\mathcal{T}| + |C|$,*

1. *a Σ-interpretation $\mathcal{I}$, $\Sigma = (\mathsf{sig}(\mathcal{T}) \cup \mathsf{sig}(C)) \uplus \{P\}$, with P a concept name, such that for all ABoxes $\mathcal{A}$,*
   (a) *there is a polynomial reduction of answering $C(a)$ w.r.t. $\mathcal{T}$ to the complement of $\mathsf{CSP}(\mathcal{I})$;*
   (b) *there is a polynomial reduction from the complement of $\mathsf{CSP}(\mathcal{I})$ to Boolean CQ-answering w.r.t. $\mathcal{T}$;*
2. *a Σ-interpretation $\mathcal{I}$, $\Sigma = \mathsf{sig}(\mathcal{T})$, such that for every ABox $\mathcal{A}$, $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ iff $\mathsf{Hom}(\mathcal{A}^{\Sigma}, \mathcal{I})$.*

For Point 1, $\mathcal{I}$ is in fact the interpretation that is obtained by the standard type elimination procedure for $\mathcal{ALCI}$-TBoxes $\mathcal{T}$ and concepts C. More specifically, let S be the closure under single negation of all subconcepts of $\mathcal{T}$ and C. A *type* t is a maximal subset of S that is satisfiable w.r.t. $\mathcal{T}$. Then $\Delta^{\mathcal{I}}$ is the set of all types, $t \in A^{\mathcal{I}}$ iff $A \in t$, and $(t, t') \in r^{\mathcal{I}}$ iff $\forall r.D \in t$ implies $D \in t'$ and $\forall r^-.D \in t'$ implies $D \in t$. For the special concept name P, set $P^{\mathcal{I}} = \{t \mid C \notin t\}$. With the type elimination algorithm, $\mathcal{I}$ can be constructed in exponential time. The mentioned reductions are then as follows:

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

Result 1 from the introduction can be derived as follows. 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 $\mathcal{T}$ from Theorem 4. By Point 1 of that theorem and since consistency of ABoxes w.r.t. $\mathcal{T}$ can trivially be reduced to the complement of answering Boolean CQs w.r.t. $\mathcal{T}$, 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 either. Conversely, let $\mathcal{T}$ be a TBox for which CQ-answering w.r.t. $\mathcal{T}$ is neither in

PTIME nor coNP-hard. Then by Theorem 1 and since every ELIQ is a CQ, the same holds for ELIQ-answering w.r.t. $\mathcal{T}$. Thus, there is a concrete ELIQ $C(a)$ such that answering $C(a)$ w.r.t. $\mathcal{T}$ is coNP-intermediate. Let $\mathcal{I}$ be the interpretation constructed in Point 1 of Theorem 5 for $\mathcal{T}$ and $C(a)$. By Point 1a, $\mathsf{CSP}(\mathcal{I})$ is not in PTIME; by Point 1b, it is not NP-hard either.

Result 5 from the introduction can be derived as follows. It is proved in [16, 5] that the problem to decide whether the class of structures $\{\mathcal{I}' \mid \mathsf{Hom}(\mathcal{I}', \mathcal{I})\}$ is FO-definable is NP-complete. We obtain a NEXPTIME upper bound since the template $\mathcal{I}$ associated with $\mathcal{T}$ can be constructed in exponential time. The claims for AC^0 and LOGSPACE follow in the same way from other results in [16, 5].

We now develop a condition on TBoxes, called unraveling tolerance, that is sufficient for PTIME CQ-answering and strictly generalizes Horn-$\mathcal{ALCFI}$, the $\mathcal{ALCFI}$-fragment of Horn-$\mathcal{SHIQ}$. For the case of TBoxes of depth one, we obtain a PTIME/coNP dichotomy result. The notion of 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. This is inspired by (i) the observation that, in the proof of Theorem 3, the non-tree-shape of ABoxes is essential; and (ii) by Theorem 5 together with the known fact the non-uniform CSPs are tractable when restricted to tree-shaped input structures. The *unraveling* $\mathcal{A}_u$ of an ABox $\mathcal{A}$ is the following ABox:

- the individual names $\mathsf{Ind}(\mathcal{A}_u)$ of $\mathcal{A}_u$ are sequences $b_0 r_0 b_1 \cdots r_{n-1} b_n$, $b_0, \dots, b_n \in \mathsf{Ind}(\mathcal{A})$ and $r_0, \dots, r_{n-1}$ (possibly inverse) roles such that for all $i < n$, we have $r_i(b_i, b_{i+1}) \in \mathcal{A}$ and $b_{i+1} \neq b_{i-1}$ (whenever $i > 0$);
- for each $C(b) \in \mathcal{A}$ and $\alpha = b_0 r_0 b_1 \cdots r_{n-1} b_n \in \mathsf{Ind}(\mathcal{A}_u)$ with $b_n = b$, we have $C(\alpha) \in \mathcal{A}_u$;
- for each $b_0 r_0 b_1 \cdots r_{n-1} b_n \in \mathsf{Ind}(\mathcal{A}_u)$, we have $r_{n-1}(b_{n-1}, b_n) \in \mathcal{A}_u$.

For all $\beta = b_0 r_0 \cdots r_{n-1} b_n \in \mathsf{Ind}(\mathcal{A}_u)$, we write $\mathsf{tail}(\beta)$ to denote b_n. Note that the condition $b_{i+1} \neq b_{i-1}$ is needed to ensure that functional roles can still be interpreted in a functional way after unraveling, despite the UNA.

Definition 3. *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. We now show that the class of unraveling tolerant $\mathcal{ALCFI}$-TBoxes generalizes Horn-$\mathcal{ALCFI}$. This is based on the original and most general definition of Horn-$\mathcal{SHIQ}$ in [12] and thus also captures weaker variants as used e.g. in [13, 9]. The TBox in Example 1, which is unraveling tolerant but not a Horn-$\mathcal{ALCFI}$-TBox, demonstrates that the generalization is strict.

Lemma 2. *Every Horn-$\mathcal{ALCFI}$-TBox is unraveling tolerant.*

It is interesting to note that unraveling tolerance implies materializability. We shall see that the converse is, in general, not true.

Lemma 3. *Every unraveling-tolerant $\mathcal{ALCFI}$-TBox is materializable.*

We now show that unraveling tolerance yields a class of $\mathcal{ALCFI}$-TBoxes for which query answering is in PTIME. By Lemma 2 and since we actually exhibit a *uniform* algorithm for query answering w.r.t. unraveling tolerant TBoxes, this also reproves the known PTIME upper bound for CQ-answering in Horn-$\mathcal{ALCFI}$ [9]. This result is not a consquence of Theorem 4 and known results for CSPs since we capture full $\mathcal{ALCFI}$.

Theorem 6. *If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is unraveling tolerant, then PEQ-answering w.r.t. $\mathcal{T}$ is in* PTIME.

To see that unraveling tolerance does not capture all $\mathcal{ALCFI}$-TBoxes for which query answering is in PTIME, we can invoke Theorem 4. For example, taking a CSP for 2-colorability, we obtain a TBox $\mathcal{T}$ for which CQ-answering is in PTIME and such that an ABox $\mathcal{A}$ with $\mathsf{sig}(\mathcal{A}) = \{r\}$ is consistent w.r.t. $\mathcal{T}$ iff $\mathcal{A}$ is 2-colorable. Thus, $\mathcal{A}, \mathcal{T} \models X(a)$, X a fresh concept name, iff $\mathcal{A}$ is not 2-colorable. It follows that $\mathcal{T}$ is not unraveling tolerant. We conjecture that it is possible to generalize Theorem 6 to larger classes of TBoxes by relaxing the operation of ABox unraveling such that it yields ABoxes of bounded treewidth instead of tree-shaped ABoxes. Such a generalization would still not capture 2-colorability.

We now turn to TBoxes of depth one. The central observation is that for this special case, we can prove a converse of Lemma 3.

Lemma 4. *Every materializable $\mathcal{ALCFI}$-TBox of depth one is unraveling tolerant.*

This brings us into the position where we can establish the announced dichotomy result for $\mathcal{ALCFI}$-TBoxes of depth one. If such a TBox $\mathcal{T}$ is materializable, then Lemma 4 and Theorem 6 yield that PEQ-answering w.r.t. $\mathcal{T}$ is in PTIME. Otherwise, ELIQ-answering w.r.t. $\mathcal{T}$ is coNP-complete by Theorem 3. We thus obtain the following.

Theorem 7 (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 $\mathcal{Q} \in \{PEQ, CQ, ELIQ\}$.*

5 Deciding FO-Rewritability

The results of this section are based on the observation that for materializable TBoxes of depth one, FO-rewritability for CQ follows from FO-rewritability for *atomic* concepts, i.e., concept names and $\bot$. We say that an atomic concept A is *FO-rewritable w.r.t. a TBox $\mathcal{T}$ and a signature Σ* if there exists an FO-formula φ_A such that for all Σ-ABoxes $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$: $\mathcal{T}, \mathcal{A} \models A(a)$ iff $\mathcal{I}_\mathcal{A} \models \varphi_A[a]$. Clearly, if $\mathcal{T}$ is FO-rewritable for CQ, then every atomic concept is FO-rewritable w.r.t. $\mathcal{T}$ and any signature. For materializable TBoxes of depth one, the converse is also true.

Lemma 5. *A materializable $\mathcal{ALCFI}$-TBox of depth one is FO-rewritable for CQs iff all atomic concepts are FO-rewritable w.r.t. $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T})$.*

Based on Lemma 5, we can use Theorem 5 and results from [16] to obtain the following result, in a similar (but slightly more involved) way as in the proof of Result 5 from the introduction.

Theorem 8. *FO-rewritability for CQs is decidable in* NEXPTIME*, for any of the following classes of TBoxes: materializable $\mathcal{ALCI}$-TBoxes of depth one, Horn-$\mathcal{ALC}$-TBoxes, and Horn-$\mathcal{ALCI}$-TBoxes of depth two.*

Theorem 5 does not apply to DLs with functional roles. To analyze FO-rewritability in the presence of functional roles, we associate with every materializable TBox $\mathcal{T}$ of depth one a monadic datalog program $\Pi_{\mathcal{T}}$ such that $\mathcal{T}$ and $\Pi_{\mathcal{T}}$ give the same answers to queries $A(a)$, A atomic. We then show that $\mathcal{T}$ is FO-rewritable if, and only if, $\Pi_{\mathcal{T}}$ is equivalent to a non-recursive datalog program. The latter property is known as *boundedness* of a datalog program and has been studied extensively for fixpoint logics [3, 18] and datalog programs [8]. Using existing decidability results for boundedness, we can then establish a counterpart of Theorem 8 for the case of $\mathcal{ALCFI}$.

For our purposes, a monadic datalog program Π consists of rules $A(x) \leftarrow X$, where A is a concept name and X is a finite set consisting of assertions of the form $B(x)$, $r(x_1, x_2)$, and inequalities $x_1 \neq x_2$, where B is a concept name, r a role, and x, x_1, x_2 range over variables. Inequalities are required to model functional roles. We also use a special unary predicate $\bot$ and rules $\bot(x) \leftarrow X$ stating that X is inconsistent. For an ABox $\mathcal{A}$, we denote by $\Pi^i(\mathcal{A})$ the set of all assertions $A(a)$ that can be derived using i applications of rules from Π to $\mathcal{A}$. We set $\Pi^\infty(\mathcal{A}) = \bigcup_{i\geq 0} \Pi^i(\mathcal{A})$.

Definition 4 (Boundedness). *Let Π be a datalog program and Σ a signature. An atomic concept A* is bounded in Π for Σ-ABoxes *if there exists a $k > 0$ such that for all Σ-ABoxes $\mathcal{A}$ and all $a \in \mathsf{sig}(\mathcal{A})$: $A(a) \in \Pi^\infty(\mathcal{A})$ iff $A(a) \in \Pi^k(\mathcal{A})$.*

Let $\mathcal{T}$ be a materializable TBox of depth one. A Σ-neighbourhood ABox (Σ-NH) consists of a Σ-ABox $\mathcal{A}$ with a distinguished individual name f such that $\mathcal{A}$ consists of assertions of the form $r(f, a)$ with r a role and $a \neq f$ and $A(b)$ such that

- for each $b \neq f$ with $b \in \mathsf{Ind}(\mathcal{A})$ there is exactly one r such that $r(f, b) \in \mathcal{A}$;
- if $r(f, b_1)$ and $r(f, b_2) \in \mathcal{A}$ and $b_1 \neq b_2$, then there exists $A(b_1) \in \mathcal{A}$ with $A(b_2) \notin \mathcal{A}$ or vice versa.

The ABox $\mathcal{A}$ in which each individual b is replaced by a variable x_b is denoted by $\mathcal{A}^x$. Now define a monadic datalog program associated with $\mathcal{T}$, where $\Sigma = \mathsf{sig}(\mathcal{T})$:

$$\begin{aligned}\Pi_{\mathcal{T}} = \{&A(x_a) \leftarrow \mathcal{A}^x \mid \mathcal{A} \text{ is a } \Sigma\text{-NH}, a \in \mathsf{Ind}(\mathcal{A}), A \in \Sigma, (\mathcal{T}, \mathcal{A}) \models A(a)\} \cup \\ &\{\bot(x) \leftarrow \mathcal{A}^x \mid \mathcal{A} \text{ is a } \Sigma\text{-NH that is not consistent w.r.t. } \mathcal{T}\} \cup \\ &\{\bot(x) \leftarrow r(y, y_1), r(y, y_2), y_1 \neq y_2 \mid \mathsf{func}(r) \in \mathcal{T}\} \cup \\ &\{A(x) \leftarrow \bot(x) \mid A \in \Sigma\}.\end{aligned}$$

The following lemma states that $\Pi_{\mathcal{T}}$ behaves as intended.

Lemma 6. *For every materializable $\mathcal{ALCFI}$-TBox $\mathcal{T}$ of depth one, every $A \in \mathsf{sig}(\mathcal{T})$, every ABox $\mathcal{A}$, and every $a \in \mathsf{Ind}(\mathcal{A})$, $(\mathcal{T}, \mathcal{A}) \models A(a)$ iff $A(a) \in \Pi_{\mathcal{T}}^\infty(\mathcal{A})$. Moreover, $\bot(a) \in \Pi_{\mathcal{T}}^\infty(\mathcal{A})$ iff $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$.*

Using unfolding tolerance of materializable TBoxes of depth one, one can show the following equivalence for FO-rewritability and boundedness.

Lemma 7. *For every materializable $\mathcal{ALCFI}$-TBox $\mathcal{T}$ of depth one and signature Σ: an atomic concept A is bounded in $\Pi_{\mathcal{T}}$ for Σ-ABoxes iff A is FO-rewritable w.r.t. $\mathcal{T}$ and Σ.*

Unfortunately, decidability results for boundedness of monadic datalog programs are not directly applicable to $\Pi_{\mathcal{T}}$ since they assume programs without inequalities [8, 11]. However, using unfolding tolerance, one can employ instead recent decidability results on boundedness of least fixed points over trees [18] to obtain the following theorem.

Theorem 9. *FO-rewritability for CQs is decidable, for any of the following classes of TBoxes: materializable $\mathcal{ALCFI}$-TBoxes of depth one, Horn-$\mathcal{ALCF}$-TBoxes, and Horn-$\mathcal{ALCFI}$-TBoxes of depth two.*

6 Non-Dichotomy and Undecidability in $\mathcal{ALCF}$

The aim of this section is to show that the addition of functional roles significantly complicates the problems studied in the previous sections. More precisely, 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) CQ-answering in PTIME is undecidable for $\mathcal{ALCF}$-TBoxes, and likewise for coNP-hardness, materializability and FO-rewritability. Point (i) is a consequence of the following result.

Theorem 10. *For every language L in* coNP*, there is an $\mathcal{ALCF}$-TBox $\mathcal{T}$ and query* $\mathsf{rej}(a)$*,* rej *a concept name, such that the following holds:*

1. *there exists a polynomial reduction of deciding $v \in L$ 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 deciding $v \in L$.*

Ladners theorem [15] states that unless PTIME = NP, coNP intermediate problems exist. Suppose to the contrary of Point (i) that for every $\mathcal{ALCF}$-TBox $\mathcal{T}$, CQ answering w.r.t. $\mathcal{T}$ is in PTIME or coNP-hard. Take a coNP-intermediate language L and let $\mathcal{T}$ be the TBox from Theorem 10. 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 1 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 10, this is impossible.

The proof of Theorem 10 combines the 'hidden' concepts H_v from the proof of Theorem 4 with ideas from a proof in [1] which establishes undecidability of a certain *query emptiness* problem in $\mathcal{ALCF}$. Using a similar strategy, we establish the undecidability results announced as Point (ii) above, summarized by the following theorem.

Theorem 11. *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.*

In the appendix, we also prove that FO-rewritability for CQ is undecidable in $\mathcal{ALCF}$, for a slightly modified definition of FO-rewritability that only considers *consistent* ABoxes.

7 Conclusions

We have have introduced non-uniform data complexity of query answering w.r.t. description logic TBoxes and proved that it enables a more fine-grained analysis than the standard approach. Many questions remain. In particular, the newly established CSP-connection should be exploited further. We believe that the techniques introduced in this paper can be extended to richer DLs such as $\mathcal{SHIQ}$.

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

References

1. F. Baader, M. Bienvenu, C. Lutz, and F. Wolter. Query and predicate emptiness in description logics. In *Proc. of KR2010*. AAAI Press, 2010.
2. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of IJCAI05*, pages 364–369. Professional Book Center, 2005.
3. J. Barwise and Y. N. Moschovakis. Global inductive definability. *J. Symb. Log.*, 43(3):521–534, 1978.
4. A. A. Bulatov. A dichotomy theorem for constraints on a three-element set. In *Proc. of FOCS02*, pages 649–658, 2002.
5. A. A. Bulatov, A. A. Krokhin, and B. Larose. Dualities for constraint satisfaction problems. In *Complexity of Constraints*, pages 93–124, 2008.
6. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Autom. Reasoning*, 39(3):385–429, 2007.
7. C. C. Chang and H. J. Keisler. *Model Theory*, volume 73 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 1990.
8. S. S. Cosmadakis, H. Gaifman, P. C. Kanellakis, and M. Y. Vardi. Decidable optimization problems for database logic programs (preliminary report). In *Proc. of STOC88*, pages 477–490, 1988.
9. T. Eiter, G. Gottlob, M. Ortiz, and M. Simkus. Query answering in the description logic Horn-$\mathcal{SHIQ}$. In *Proc. of JELIA08*, volume 5293 of LNCS, pages 166–179. Springer, 2008.
10. T. Feder and M. Y. Vardi. Monotone monadic snp and constraint satisfaction. In *Proc. of STOC93*, pages 612–622, 1993.
11. H. Gaifman, H. G. Mairson, Y. Sagiv, and M. Y. Vardi. Undecidable optimization problems for database logic programs. In *Proc. of LICS87*, pages 106–115, 1987.
12. U. Hustadt, B. Motik, and U. Sattler. Reasoning in description logics by a reduction to disjunctive datalog. *J. Autom. Reasoning*, 39(3):351–384, 2007.
13. Y. Kazakov. Consequence-driven reasoning for horn SHIQ ontologies. In *Proc. of IJCAI09*, pages 2040–2045, 2009.
14. M. Krötzsch. Efficient inferencing for OWL EL. In *Proc. of JELIA10*, volume 6341 of LNCS, pages 234–246. Springer, 2010.
15. R. E. Ladner. On the structure of polynomial time reducibility. *JACM*, 22(1):155171, 1975.
16. B. Larose, C. Loten, and C. Tardif. A characterisation of first-order constraint satisfaction problems. *Logical Methods in Computer Science*, 3(4), 2007.
17. C. Lutz, R. Piro, and F. Wolter. Description logic tboxes: Model-theoretic characterizations and rewritability. In *Proc. of IJCAI11*, 2011.
18. M. Otto, A. Blumensath, and M. Weyer. Decidability results for the boundedness problem. Technical report, TU Darmstadt, 2010.
19. A. Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. *J. of Intell. Inf. Sys.*, 2:265–278, 1993.

A Proofs for Section 3

A.1 Proof of Theorem 2

We state the result to be proved again.

Theorem 2 Let $\mathcal{T}$ be a $\mathcal{ALCFI}$ TBox. The following conditions are equivalent:

1. $\mathcal{T}$ is PEQ materializable;
2. $\mathcal{T}$ is CQ materializable;
3. $\mathcal{T}$ is ELIQ materializable;
4. $\mathcal{T}$ has the ABox disjunction property.

To prove Theorem 2, we require the notions of a homomorphism and of finite homomorphic embeddability. Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be interpretations. A function f from $\Delta^{\mathcal{I}_1}$ to $\Delta^{\mathcal{I}_2}$ is called a homomorphism if

- $d \in A^{\mathcal{I}_1}$ implies $f(d) \in A^{\mathcal{I}_2}$, for all $d \in \Delta^{\mathcal{I}_1}$ and all concept names A;
- $(d_1, d_2) \in r^{\mathcal{I}_1}$ implies $(f(d_1, f(d_2)) \in r^{\mathcal{I}_2}$, for all $d_1, d_2 \in \Delta^{\mathcal{I}_1}$ and all role names r;
- $f(a^{\mathcal{I}_1}) = a^{\mathcal{I}_2}$ for all individual names a interpreted in $\mathcal{I}_1$.

Say that an interpretation $\mathcal{I}_1$ is *finitely homomorphically embeddable* into $\mathcal{I}_2$ if for every finite subset of the domain of $\mathcal{I}_1$ there exists a homomorphism from the induced subinterpretation of $\mathcal{I}_1$ to $\mathcal{I}_2$. The following lemma is readily checked:

Lemma 8. *The following conditions are equivalent, for all interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$::*

- *$\mathcal{I}_1 \models q[\boldsymbol{a}]$ implies $\mathcal{I}_2 \models q[\boldsymbol{a}]$, for all CQs $q(\boldsymbol{x})$ and sequences $\boldsymbol{a}$ of individual names interpreted in $\mathcal{I}_1$;*
- *$\mathcal{I}_1$ is finitely homomorphically embeddable into $\mathcal{I}_2$;*
- *$\mathcal{I}_1 \models q[\boldsymbol{a}]$ implies $\mathcal{I}_2 \models q[\boldsymbol{a}]$ for all PEQs $q(\boldsymbol{x})$ and sequences $\boldsymbol{a}$ of individual names interpreted in $\mathcal{I}_1$.*

Using Lemma 8, one can directly prove the following:

Lemma 9. *For every $\mathcal{ALCFI}$-TBox $\mathcal{T}$, the following conditions are equivalent:*

- *$\mathcal{T}$ is CQ materializable;*
- *$\mathcal{T}$ is PEQ materializable;*
- *for every ABox $\mathcal{A}$ such that $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ there exists a model $\mathcal{I}$ of $(\mathcal{T}, \mathcal{A})$ that is finitely homomorphically embeddable into any model $\mathcal{J}$ of $(\mathcal{T}, \mathcal{A})$.*

From Lemma 9, we obtain the equivalence of Points 1 and 2 in Theorem 2. We now give a similar semantic characterization for ELIQs.

Definition 5 (Simulation). *A relation S between interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ is a $\mathcal{ELI}$-simulation if the domain of S coincides with $\Delta^{\mathcal{I}_1}$ and*

- *If $d_1 \in A^{\mathcal{I}_1}$ and $(d_1, d_2) \in S$, then $d_2 \in A^{\mathcal{I}_2}$, for all $d_1 \in \Delta^{\mathcal{I}_1}$;*

– *If* $(d_1, d_2) \in S$ *and* $(d_1, d_1') \in r^{\mathcal{I}_1}$ *for some role* r*, then there exists* $d_2' \in \Delta^{\mathcal{I}_2}$ *such that* $(d_1', d_2') \in S$ *and* $(d_2, d_2') \in r^{\mathcal{I}_2}$*.*
– $(a^{\mathcal{I}_1}, a^{\mathcal{I}_2}) \in S$*, for all individual names* a *interpreted in* $\mathcal{I}_1$*;*

$\mathcal{I}_1$ is *finitely simulated in* $\mathcal{I}_2$ iff for every finite subset of the domain of $\mathcal{I}_1$ there exists a simulation from the induced subinterpretation of $\mathcal{I}_1$ to $\mathcal{I}_2$. Let X be a set of individual names. An interpretation $\mathcal{I}$ is *weakly covered by* X if for all $d \in \Delta^{\mathcal{I}}$ there exist $d_0, \ldots, d_n \in \Delta^{\mathcal{I}}$ with $a^{\mathcal{I}} = d_0$ and $d_n = d$ for some individual name $a \in X$ such that for all $i < n$ there exists a role r with $(d_i, d_{i+1}) \in r^{\mathcal{I}}$.

Lemma 10. *Let* $\mathcal{I}_1, \mathcal{I}_2$ *be interpretations and* X *a set of individual names. Assume that* $\mathcal{I}_1$ *interprets exactly the individual names in* X *and is weakly covered by* X*. Then following conditions are equivalent:*

– $\mathcal{I}_1 \models C(a)$ *implies* $\mathcal{I}_2 \models C(a)$ *for all ELIQs* $C(a)$ *with* $a \in X$*;*
– $\mathcal{I}_1$ *is finitely simulated in* $\mathcal{I}_2$*.*

From Lemma 10, we directly obtain the following

Lemma 11. *For every* $\mathcal{ALCFI}$*-TBox* $\mathcal{T}$*, the following conditions are equivalent:*

– $\mathcal{T}$ *is ELIQ materializable;*
– *for every ABox* $\mathcal{A}$ *such that* $\mathcal{A}$ *is consistent w.r.t.* $\mathcal{T}$ *there exists a model* $\mathcal{I}$ *of* $(\mathcal{T}, \mathcal{A})$ *that is finitely simulated in any model* $\mathcal{J}$ *of* $(\mathcal{T}, \mathcal{A})$*.*

To show the equivalence of Points 1 and 3 of Theorem 2, it remains to relate simulations to homomorphisms. The following result is based on the observation that on tree-like interpretations simulations can be transformed into homomorphisms.

Lemma 12. *For every* $\mathcal{ALCFI}$*-TBox* $\mathcal{T}$*, the following conditions are equivalent:*

– *for every ABox* $\mathcal{A}$ *such that* $\mathcal{A}$ *is consistent w.r.t.* $\mathcal{T}$ *there exists a model* $\mathcal{I}$ *of* $(\mathcal{T}, \mathcal{A})$ *that is finitely simulated in any model* $\mathcal{J}$ *of* $(\mathcal{T}, \mathcal{A})$*.*
– *for every ABox* $\mathcal{A}$ *such that* $\mathcal{A}$ *is consistent w.r.t.* $\mathcal{T}$ *there exists a model* $\mathcal{I}$ *of* $(\mathcal{T}, \mathcal{A})$ *that is finitely homomorphically embeddable in any model* $\mathcal{J}$ *of* $(\mathcal{T}, \mathcal{A})$*.*

Proof. The proof of the direction from Point 2 to Point 1 is trivial. Conversely, assume that $\mathcal{A}_0$ is an ABox and $\mathcal{I}$ is a model of $(\mathcal{T}, \mathcal{A}_0)$ that is finitely simulated in any model $\mathcal{J}$ of $(\mathcal{T}, \mathcal{A}_0)$.

We unfold $\mathcal{I}$ into a forest-like interpretation $\mathcal{I}^*$ as follows. (Observe that the construction of $\mathcal{I}^*$ from $\mathcal{I}$ is very similar to the construction of the unraveling $\mathcal{A}_u$ of an ABox $\mathcal{A}$. The main difference is that we do not modify the part of $\mathcal{I}$ that interprets the individual symbols of the ABox $\mathcal{A}_0$. We therefore obtain a forest-like structure, rather than a tree-like structure.) The domain $\Delta^{\mathcal{I}^*}$ consists of all sequences

$$d_0 r_1 d_1 \ldots r_n d_n$$

with $n \geq 0$ and r_i a role such that

– there exists $a \in \mathsf{Ind}(\mathcal{A})$ such that $d_0 = a^{\mathcal{I}}$;
– $(d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}}$ for all $i < n$;

– if $\mathsf{func}(r_1) \in \mathcal{T}$, then there does not exist any b with $r_1(a, b) \in \mathcal{A}$;
– if $\mathsf{func}(r_{i+1}) \in \mathcal{T}$, then $r_i^- \neq r_{i+1}$.

We set

– for all $A \in \mathsf{N_C}$:
$$A^{\mathcal{I}^*} = \{d_0 \dots d_n \in \Delta^{\mathcal{I}^*} \mid d_n \in A^{\mathcal{I}}\};$$
– for all $r \in \mathsf{N_R}$:
$$r^{\mathcal{I}^*} = \{(\sigma, \sigma r d) \mid \sigma r d \in \Delta^{\mathcal{I}^*} \cup \{(\sigma r^- d, \sigma) \mid \sigma r^- d \in \Delta^{\mathcal{I}^*}\}.$$
– $a^{\mathcal{I}^*} = a^{\mathcal{I}}$, for all $a \in \mathsf{Ind}(\mathcal{A})$.

One can show that, by construction, $\mathcal{I}^*$ is a model of $(\mathcal{T}, \mathcal{A})$ and $\mathcal{I}^*$ is simulated in $\mathcal{I}$ by the simulation
$$S = \{(\sigma r d, d) \mid \sigma r d \in \Delta^{\mathcal{I}^*}\}.$$
Thus, $\mathcal{I}^*$ is finitely simulated in any model of $(\mathcal{T}, \mathcal{A})$. Now let $\mathcal{J}$ be a model of $(\mathcal{T}, \mathcal{A})$. We want to finitely homomorphically embed $\mathcal{I}^*$ into $\mathcal{J}$. But, for any finite subset of the domain of $\mathcal{I}^*$ such an embedding is easily constructed from a simulation in $\mathcal{J}$. ❑

Lemmas 9, 11, and 12 imply the equivalence of Points 1, 2, and 3 in Theorem 2. It remains to show the equivalence of Point 3 and Point 4. Clearly, Point 4 follows from Point 3. Conversely, assume that Point 3 does not hold. Let $\mathcal{A}$ be an ABox that is consistent w.r.t. $\mathcal{T}$ and such that there does not exist any model $\mathcal{I}$ of $(\mathcal{T}, \mathcal{A})$ such that $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff $\mathcal{I} \models C[a]$, for all ELIQ $C(a)$ with $a \in \mathsf{Ind}(\mathcal{A})$. Then $\mathcal{T} \cup \mathcal{A} \cup \Gamma$, where
$$\Gamma = \{\neg C(a) \mid \mathcal{T}, \mathcal{A} \not\models C(a), a \in \mathsf{Ind}(\mathcal{A}), C(a) \text{ an ELIQ}\},$$
is not satisfiable (any model $\mathcal{I}$ satisfying $\mathcal{T} \cup \mathcal{A} \cup \Gamma$ would have the property that $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff $\mathcal{I} \models C[a]$, for all ELIQ $C(a)$ with $a \in \mathsf{Ind}(\mathcal{A})$). By compactness, there is a finite subset Γ' of Γ such that $\mathcal{T} \cup \mathcal{A} \cup \Gamma'$ is not satisfiable. Equivalently,
$$(\mathcal{T}, \mathcal{A}) \models \bigvee_{\neg C(a) \in \Gamma'} C(a).$$
By definition of Γ', $(\mathcal{T}, \mathcal{A}) \not\models C(a)$, for all $\neg C(a) \in \Gamma'$. Thus, $\mathcal{T}$ does not have the ABox disjunction property. This finishes the proof of Theorem 2.

A.2 Proof of Theorem 3

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 [19]. 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 [19] that 2+2-SAT is NP-complete.

Theorem 3. 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 2 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^i_j 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 $C_0(a^i_0)$ is satisfied and as false if any of the ELIQs $C_j(a^i_j)$, $0 < j \leq k$, 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 $b_1^i, \ldots, b_k^i$ reachable via h such that for $1 \leq j \leq k$,

– $\exists r_\ell.C_\ell(b_j^i)$ is satisfied for all $\ell = 1, \ldots, j-1, j+1, \ldots, k$ and
– the assertion $r_j(b_j^i, a_j^i)$ is in $\mathcal{A}_\varphi$.

Thus, $(\exists r_1.C_1 \sqcap \cdots \sqcap \exists r_k.C_k)(b_j^i)$ is satisfied iff $C_j(a_j^i)$ 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 $r_j(q_i, a_j^i)$, $1 \leq j \leq k$, with the following:
  - $h(q_i, b_1^i), \ldots, h(q_i, b_k^i)$ for all $i \leq m$;
  - $r_j(b_j^i, a_j^i), r_1(b_j^i, d_1), \ldots, r_{j-1}(b_j^i, d_{j-1}), r_{j+1}(b_j^i, d_{j+1}), \ldots, r_k(b_j^i, d_k)$ 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. ❑

A.3 Proof of Lemma 1

We show Lemma 1 for singleton sets V. The extension to arbitrary finite V is straightforward. Thus, let Z be a concept name and z_0, z_1 role names. Let

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

Lemma 13. *For any ABox $\mathcal{A}$ and set $I \subseteq \mathsf{Ind}(\mathcal{A})$, one can construct a model $\mathcal{I}$ of $(\mathcal{T}_Z, \mathcal{A})$ such that*

- *$H^{\mathcal{I}} = I$;*
- *for all models $\mathcal{J}$ of $(\mathcal{T}_Z, \mathcal{A})$ there exists a homomorphism from $\mathcal{I}$ to $\mathcal{J}$.*

Thus, any such interpretation $\mathcal{I}$ is a minimal model of $(\mathcal{T}_Z, \mathcal{A})$.

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 $\mathcal{T}_Z$ 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 $\mathcal{T}_Z$ and we have $H^{\mathcal{I}} = I$. Now let $\mathcal{J}$ be a model of $(\mathcal{T}_Z, \mathcal{A})$. 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. ❑

A.4 Remarks on Theorem 4

The proof of Theorem 4 is straightforward, by using Lemma 1. It is of interest to find out whether the reduction still works for TBoxes that are consistent w.r.t. any ABox. We show that this is indeed the case.

Theorem 12. *For every $\mathsf{CSP}(\mathcal{I})$, $\mathcal{I}$ a Σ-interpretation, one can compute a materializable $\mathcal{ALC}$-TBox $\mathcal{T}$ containing a fresh concept name $A \notin \Sigma$ such that all ABoxes are consistent w.r.t. $\mathcal{T}$ and the following are equivalent for all ABoxes $\mathcal{A}$ not containing A:*

- *$(\mathcal{T}, \mathcal{A}) \models \exists v.A(v)$;*
- *not $\mathsf{Hom}(\mathcal{A}^{\Sigma}, \mathcal{I})$.*

Moreover, on connected ABoxes the CQ answering problem w.r.t. $\mathcal{T}$ is polynomially reducible to the complement of $\mathsf{CSP}(\mathcal{I})$.

Proof. Assume that $\mathsf{CSP}(\mathcal{I})$ is given. Let $V = \Delta^{\mathcal{I}}$ and assume, for simplicity, that $\Sigma = \{r\}$. We show that the TBox

$$\begin{aligned}\mathcal{T} = \{H_v \sqcap \exists r.H_w \sqsubseteq A \mid v, w \in V, (v,w) \notin r^{\mathcal{I}}\} \cup \\ \{H_v \sqcap H_w \sqsubseteq A \mid v, w \in V, v \neq w\} \cup \{\bigsqcup_{v \in V} \neg H_v \sqsubseteq A\} \cup \\ \{A \sqsubseteq \forall r.A, \exists r.A \sqsubseteq A\} \cup \mathcal{T}_V\end{aligned}$$

is as required. Consider an ABox $\mathcal{A}$. Let $\mathcal{A}_1, \dots, \mathcal{A}_n$ be the decomposition of $\mathcal{A}$ into maximal connected components. Let I be the set of i such that there is a homomorphism h_i from $\mathcal{A}_i^\Sigma$ to $\mathcal{I}$. Let, for $v \in V = \Delta^\mathcal{I}$,

$$I_v = \bigcup_{1 \leq i \leq n} \{a \in \mathsf{Ind}(\mathcal{A}_i) \mid h_i(a) = v\}$$

By Lemma 1, we can take a minimal model $\mathcal{J}$ of $\mathcal{T}_V$ with $H_v^\mathcal{I} = I_v$ for $v \in V$. Extend $\mathcal{J}$ to model $\mathcal{J}'$ of $(\mathcal{T}, \mathcal{A})$ by setting

$$A^{\mathcal{I}'} = \bigcup_{i \notin I} \{a^\mathcal{I} \mid a \in \mathsf{Ind}(\mathcal{A}_i)\}.$$

It is readily checked that $\mathcal{J}'$ is a minimal model of $(\mathcal{T}, \mathcal{A})$. (Since $\mathcal{A}$ is arbitrary, it follows that all ABoxes are consistent w.r.t. $\mathcal{T}$ and that $\mathcal{T}$ is materializable.) It follows that $(\mathcal{T}, \mathcal{A}) \models \exists v.A(v)$ iff there is no homomorphism from $\mathcal{A}^\Sigma$ to $\mathcal{I}$. This proves the first claim.

We can also set

$$\mathcal{A}' = \mathcal{A} \cup \bigcup_{i \notin I} \{A(a) \mid a \in \mathsf{Ind}(\mathcal{A}_i)\}.$$

Then, for any CQ q, we have $(\mathcal{T}, \mathcal{A}) \models q$ iff $(\mathcal{T}_V, \mathcal{A}') \models q$, and the latter problem is in PTIME. For a polynomial reduction for connected ABoxes, observe that for connected $\mathcal{A}$ computing $\mathcal{A}'$ reduces to checking not $\mathsf{Hom}(\mathcal{A}^\Sigma, \mathcal{I})$. ❑

B Proofs for Section 2

In this section, we prove Theorem 1. Note that in the proofs of Theorems 2 and 3 we did not use Theorem 1. Thus, we can (and will) employ them in the proof below. We formulate Theorem 1 again.

Theorem 1 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.

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 (we can regard ELIQs as CQ). For the proofs of the other directions we can assume that $\mathcal{T}$ is materializable: otherwise, by Theorems 2 and 3, 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 trivial 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 $\boldsymbol{r} = r_1 \cdots r_n$ of roles, we set $\exists \boldsymbol{r}.C = \exists r_1. \cdots \exists r_n.C$. 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 14. *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 $\boldsymbol{r} = r_1 \cdots r_n$ 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 $(\mathcal{T}, \mathcal{A}) \models \exists \boldsymbol{r}.C(a)$.*

Proof. (Sketch) Let $\mathcal{I}$ be a minimal model of $(\mathcal{T}, \mathcal{A})$. We may assume that $\mathcal{I}$ is an unfolded interpretation as described in the proof of Lemma 12. 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 $\mathcal{J} \not\models \exists \boldsymbol{r}.C(a)$. We have derived a contradiction since $a^{\mathcal{I}} \in (\exists \boldsymbol{r}.C)^{\mathcal{I}}$ and therefore, since $\mathcal{I}$ is a minimal model of $(\mathcal{T}, \mathcal{A})$, $(\mathcal{T}, \mathcal{A}) \models (\exists \boldsymbol{r}.C)(a)$. ❑

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

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

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

– for every $y \in \boldsymbol{y} \setminus Y$ 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 $q(\boldsymbol{x})$, we can form

- φ_Y consisting of all $A(t)$ with $t \in \mathsf{Ind}(q) \cup \boldsymbol{x} \cup Y$ and all $r(t, t')$ with $t, t' \in \mathsf{Ind}(q) \cup \boldsymbol{x} \cup Y$;
- 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 $\boldsymbol{x}, \boldsymbol{y}$ in an unfolded minimal model $\mathcal{I}$ 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 14 (for those y that are not reachable in φ from any member of $\mathsf{Ind}(\mathcal{A}) \cup \boldsymbol{x} \cup Y$) 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 minimal models $\mathcal{I}$ 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 $\varphi(\boldsymbol{x})$, 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

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

for all $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$. 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 $q(\boldsymbol{x})$ in polynomial time as follows: to decide whether $(\mathcal{T}, \mathcal{A}) \models q(\boldsymbol{a})$ for a given $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$, go through all splittings $S = (Y, \sim, f)$ of $q(\boldsymbol{x})$ and all assignments $\pi(y) \in \mathsf{Ind}(\mathcal{A})$ for $y \in Y$ and check

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

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 $(\mathcal{T}, \mathcal{A}) \models q(\boldsymbol{a})$; otherwise $(\mathcal{T}, \mathcal{A}) \not\models q(\boldsymbol{a})$. Both conditions can be checked in polynomial time.

C Proofs for Section 4

C.1 Proof of Theorem 5

We prove Theorem 5. To this end, it is sufficient to show the Claims (a) and (b) for the "type-model" $\mathcal{I}$ based on the set of types S.

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

We start by proving (a).

"$\Rightarrow$". Assume $\mathsf{Hom}(\mathcal{A}^{\Sigma}_{P(a)}, \mathcal{I})$. Let $h : \mathcal{A}^{\Sigma}_{P(a)} \to \mathcal{I}$ 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 $\mathcal{A}_{P(a)}$ 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 $r(a, b) \in \mathcal{A}^{\Sigma}$. 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 $h(b) \in C^{\mathcal{I}}$. We obtain that $\mathcal{H}$ is a model of $(\mathcal{T}, \mathcal{A})$. Moreover, $a \notin C^{\mathcal{H}}$ since $h(a) \notin C^{\mathcal{I}}$ and the latter follows because otherwise $h(a) \notin P^{\mathcal{I}}$ and $P(a) \in \mathcal{A}_{P(a)}$ 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 $\mathcal{A}^{\Sigma}_{P(a)}$ to $\mathcal{I}$. We only consider preservation of P. Assume $P(b) \in \mathcal{A}^{\Sigma}_{P(a)}$. Then $a = b$. We have $C \notin t(a^{\mathcal{J}})$. Thus $C \notin h(a)$. Hence $h(a) \in P^{\mathcal{I}}$.

Consider (b). 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 $\mathcal{A}^{\Sigma}$ to $\mathcal{I}$. We only consider preservation of P. Assume $P(b) \in \mathcal{A}^{\Sigma}$. 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 $h(a) \in P^{\mathcal{I}}$.

"$\Rightarrow$". Assume $\mathsf{Hom}(\mathcal{A}^{\Sigma}, \mathcal{I})$. Let $h : \mathcal{A}^{\Sigma} \rightarrow \mathcal{I}$ be a witness homomorphism. For each $b \in \mathsf{Ind}(\mathcal{A})$, let $\mathcal{I}_b$ be a copy of $\mathcal{I}$ (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 $r(a, b) \in \mathcal{A}^{\Sigma}$. 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 $h(b) \in C^{\mathcal{I}}$. Thus, $\mathcal{H}$ is a model of $(\mathcal{T}, \mathcal{A})$. 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, $h(b) \in P^{\mathcal{I}}$. But then $h(b) \notin C^{\mathcal{I}}$. 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.

C.2 Horn-$\mathcal{ALCFI}$-TBoxes

Different versions of Horn-$\mathcal{SHIQ}$ have been proposed in the literature, giving rise to different versions of Horn-$\mathcal{ALCFI}$. The original and most general, but also rather technical definition was given in [12]. Applying some simple transformations, it is easy to show that every Horn-$\mathcal{ALCFI}$-TBox according to [12] is equivalent to a Horn-$\mathcal{ALCFI}$-TBox of the form introduced below.[3] Our result that Horn-$\mathcal{ALCFI}$-TBoxes are unraveling tolerant thus also applies to the original definition from [12].

A Horn-$\mathcal{ALCFI}$-TBox $\mathcal{T}$ consists of functionality assertions $\mathsf{func}(r)$, r a potentially inverse role, and a single concept inclusion of the form $\top \sqsubseteq C_{\mathcal{T}}$, where $C_{\mathcal{T}}$ is built according to the topmost syntax rule in:

$$\begin{aligned} R, R' &::= \top \mid \bot \mid A \mid \neg A \mid R \sqcap R' \mid L \rightarrow R \mid \exists r.R \mid \forall r.R \\ L, L' &::= \top \mid \bot \mid A \mid L \sqcap L' \mid L \sqcup L' \mid \exists r.L \end{aligned}$$

where r ranges over all roles, potentially inverse. Note that the concept $C_{\mathcal{T}}$ is in negation normal form (NNF) and that concepts built according to L are $\mathcal{ELIU}_{\bot}$-concepts. This version of Horn-$\mathcal{ALCFI}$ is a strict generalization of the simplified form of Horn-$\mathcal{ALCFI}$ used e.g. in [13].

Lemma 2. Every Horn-$\mathcal{ALCFI}$-TBox is unraveling tolerant.

Proof. Let $\mathcal{T} = \{\top \sqsubseteq C_{\mathcal{T}}\}$ be a Horn-$\mathcal{ALCFI}$-TBox and $\mathcal{A}$ a potentially infinite ABox. We give a characterization of the entailment of ELIQs by $\mathcal{T}$ and $\mathcal{A}$ that 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. [2, 13, 14]. In this characterization, we use *extended ABoxes*, i.e., finite sets of assertions

[3] Although this is not important here, we note that even a polytime transformation is possible.

$C(a)$ with C a *potentially compound* concept and $r(a, b)$. 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 $\mathcal{A}' \vdash C(b)$.

We produce a sequence of extended ABoxes $\mathcal{A}_0, \mathcal{A}_1, \ldots$, starting with $\mathcal{A}_0 = \mathcal{A} \cup \{\top(a_\top)\}$. Intuitively, $a_\top$ 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 $a \in \mathsf{Ind}(\mathcal{A}_0)$, $r_1, \ldots, r_k$ roles that occur in $\mathcal{T}$, and $C_1, \ldots, C_k \in \mathsf{sub}(\mathcal{T})$. 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 $\mathcal{I}_\mathcal{A} \models C(a)$, 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 $r(a, b) \in \mathcal{A}_i$ (or $r^-(b, a) \in \mathcal{A}_i$), then add $C(b)$.

Let $\mathcal{A}_c = \bigcup_{i \geq 0} \mathcal{A}_i$. 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 $\mathcal{A}_c \vdash \bot(a)$.

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) if $C(a) \in \mathcal{A}_c$, then $h_i(a) \in C^\mathcal{I}$;
(b) 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) and (b) 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 $\mathcal{A}_c \models C(a)$, then preservation of ELIQs under homomorphisms also yields $\mathcal{T}, \mathcal{A} \models C(a)$. For Point 2, assume $\mathcal{A}_c \models C(a_\top)$. 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:

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

It can be shown that $\mathcal{I}$ is a model of $\mathcal{A}_c$ (thus $\mathcal{A}$) and $\mathcal{T}$. Moreover, it can be seen that $\mathcal{I} \models D(b)$ iff $\mathcal{A}_c \vdash D(b)$ for all ELIQs $D(b)$. Thus, $\mathcal{I} \not\models C(a)$, which yields $\mathcal{T}, \mathcal{A} \not\models C(a)$ as required.

We now consider the application of the above 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 $\mathcal{A}_c^u$ 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 $\beta \in \mathsf{Ind}(\mathcal{A}_c^u)$, we write $\alpha \sim \beta$ if

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

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 $\beta \in \mathsf{Ind}(\mathcal{A}_i^u)$ with $\alpha \sim \beta$, we have

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

C.3 Unraveling Tolerance

Lemma 3. Every unraveling-tolerant $\mathcal{ALCFI}$-TBox is materializable.

Proof. To show the contrapositive, assume that the $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is not materializable. By Theorem 2, $\mathcal{T}$ 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 $\mathcal{A}_\vee, \mathcal{A}_0, \dots, \mathcal{A}_k$ share any individual names. Consider the ABox

$$\begin{aligned}\mathcal{A} = {} & \mathcal{A}_\vee \cup \mathcal{A}_0 \cup \dots \cup \mathcal{A}_k \cup \{r(b, b_1), \dots, r(b, b_k)\} \\ & \cup \bigcup_{i \leq k} \{r_0(b_j, b_0), \dots, r_{j-1}(b_j, b_{j-1}), r_{j+1}(b_j, b_{j-1}), \dots, r_k(b_j, b_k)\} \\ & \cup \{r_0(b_0, a_0), \dots, r_k(b_k, a_k)\}\end{aligned}$$

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

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

The idea underlying this ABox and query is very similar to the second step in the proof of Theorem 3, where UELIQs are replaced with ELIQs. It can be shown that $\mathcal{T}, \mathcal{A} \models q$, but $\mathcal{T}, \mathcal{A}_a \not\models q$. ❑

Theorem 6. If an $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is unraveling tolerant, then PEQ-answering w.r.t. $\mathcal{T}$ is in PTIME.

To prove Theorem 6, 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 1, 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 $\mathsf{func}(r) \in \mathcal{T}$.

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 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 6.

Lemma 15. *$\mathcal{T}, \mathcal{A} \models q$ 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 $t^{\mathcal{I}}(a_k^{\mathcal{I}}) \in \pi_i(a_k)$. In particular, this implies that $t^{\mathcal{I}}(a_0) \in \pi(a_0)$.

"$\Rightarrow$". Let $t \in \pi(a_0)$. We build a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}_u$ such that $t^{\mathcal{I}}(a_0^{\mathcal{I}}) = t$, as follows. First, construct a map $\lambda : \mathsf{Ind}(\mathcal{A}_u) \to \mathsf{tp}(\mathcal{T}, q)$ 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 $\pi_0, \pi_1, \ldots$, 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 $\pi_0, \pi_1, \ldots$, 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 $\mathcal{I}_\alpha$, for all $\alpha \in \mathsf{Ind}(\mathcal{A}_u)$;
(ii) whenever $(d_\alpha, e) \in r^{\mathcal{I}}$, $\mathsf{func}(r) \in \mathcal{T}$, and there is an assertion $r(\alpha, \beta) \in \mathcal{A}_u$, then 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 $\pi_0, \pi_1, \ldots$ and Point (iii) in the definition of $\mathcal{T}$, $\mathcal{I}$ 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. ❑

For the proof of Lemma 4, we need a preliminary. An $\mathcal{ALCFI}$-TBox $\mathcal{T}$ is *infinitely materializable* if for all finite *and infinite* ABoxes $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$, there is a minimal model of $\mathcal{T}$ and $\mathcal{A}$, i.e., a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $\mathcal{I} \models q[\boldsymbol{a}]$ iff $\mathcal{T}, \mathcal{A} \models q[\boldsymbol{a}]$ for every ELIQ $q(\boldsymbol{x})$ and all $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$. As for plain materializability, it would be equivalent to define infinite materializability based on CQs or PEQs.

Lemma 16. *An $\mathcal{ALCFI}$-TBox is materializable iff it is infinitely materializable.*

This lemma follows from the observation that the proof of "$3 \Rightarrow 4$" of Theorem 2 goes through also for infinite ABoxes without modification.

Lemma 4. Every materializable $\mathcal{ALCFI}$-TBox of depth one is unraveling tolerant.

Proof. Let $\mathcal{T}$ be a materializable TBox of depth one, $\mathcal{A}$ an ABox, and $q = C_0(a_0)$ an ELIQ with $\mathcal{A}_u, \mathcal{T} \not\models q$. The latter implies that $\mathcal{A}_u$ is consistent w.r.t. $\mathcal{T}$. Thus and since $\mathcal{T}$ is infinitely materializable, there is a minimal model $\mathcal{I}$ for $\mathcal{A}_u$ and $\mathcal{T}$, and we have $\mathcal{I} \not\models q$. We may w.l.o.g. assume that $\mathcal{I}$ has forest-shape, i.e., that $\mathcal{I}$ can be obtained by first taking the disjoint union of tree-shaped models, one for each $\alpha \in \mathsf{Ind}(\mathcal{A}_u)$ with root $\alpha^{\mathcal{I}}$, and then adding role edges $(\alpha^{\mathcal{I}}, \beta^{\mathcal{I}})$ to $r^{\mathcal{I}}$ whenever $r(\alpha, \beta) \in \mathcal{A}_u$. We may also assume w.l.o.g. that $\alpha^{\mathcal{I}} = \alpha$ for all $\alpha \in \mathsf{Ind}(\mathcal{A}_u)$.

We show how to construct a model $\widehat{\mathcal{I}}$ of $\mathcal{T}$ and the original ABox $\mathcal{A}$ such that $\widehat{\mathcal{I}} \not\models q$. For $a \in \mathsf{Ind}(\mathcal{A})$, let $\mathcal{I}|_a$ denote the tree model in $\mathcal{I}$ that is rooted at a (note that $a \in \mathsf{Ind}(\mathcal{A}_u)$). Construct $\widehat{\mathcal{I}}$ as follows:

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

For every $d \in \Delta^{\mathcal{I}}$, let $\mathcal{I}|_d^1$ denote the *1-neighborhood* of d in $\mathcal{I}$, i.e., the restriction of $\mathcal{I}$ to the domain

$$\{d\} \cup \{e \mid (d, e) \in r^{\mathcal{I}} \text{ for some } r \in \mathsf{N_R}\} \cup \{e \mid (e, d) \in r^{\mathcal{I}} \text{ for some } r \in \mathsf{N_R}\}.$$

The following is a central property of $\widehat{\mathcal{I}}$:

Claim 1. For every $d \in \Delta^{\widehat{\mathcal{I}}}$, we have that d in $\widehat{\mathcal{I}}|_d^1$ is bisimilar to d in $\mathcal{I}|_d^1$.

This is trivial for all $d \in \Delta^{\widehat{\mathcal{I}}}$ with $d \notin \mathsf{Ind}(\mathcal{A})$. Thus, let $d = a \in \mathsf{Ind}(\mathcal{A})$. Then a bisimulation $\sim \subseteq \Delta^{\widehat{\mathcal{I}}|_d^1} \times \Delta^{\mathcal{I}|_d^1}$ can be defined as follows:

- $a \sim a$;
- if $e \in \Delta^{\widehat{\mathcal{I}}|_d^1} \setminus \mathsf{Ind}(\mathcal{A})$, then $e \sim e$;
- if $b \in \Delta^{\widehat{\mathcal{I}}|_d^1} \cap \mathsf{Ind}(\mathcal{A})$ and $b \neq a$, then $b \sim arb$ for all roles r with $r(a, b) \in \mathcal{A}$.

This finishes the proof of Claim 1.

Since $\mathcal{I}$ is a model of $\mathcal{T}$ and all concept assertions in $\mathcal{T}$ are of depth one, we obtain by Claim 1 that $\widehat{\mathcal{I}}$ satisfies all concept assertions in $\mathcal{T}$. By construction, $\widehat{\mathcal{I}}$ satisfies all role assertions in $\mathcal{A}$. By construction of $\mathcal{A}_u$ and Claim 1, $\widehat{\mathcal{I}}$ also satisfies all concept assertions of $\mathcal{A}$. Let $\mathsf{func}(r) \in \mathcal{T}$. We show that each $d \in \Delta^{\widehat{\mathcal{I}}}$ has at most one r-successor in $\widehat{\mathcal{I}}$. Distinguish two cases:

- $d \notin \mathsf{Ind}(\mathcal{A})$. Then d has at most one r-successor since $\mathcal{I}$ satisfied $\mathsf{func}(r)$ and by construction of $\widehat{\mathcal{I}}$.
- $d = a \in \mathsf{Ind}(\mathcal{A})$. Since $\mathcal{A}_u$ is consistent w.r.t. $\mathcal{T}$, there is at most one β with $r(a, \beta) \in \mathcal{A}_u$. By definition of $\mathcal{A}_u$, this implies that there is at most one b with $r(a, b) \in \mathcal{A}$. Moreover, if there is a b with $r(a, b) \in \mathcal{A}$, then there is a β with $r(a, \beta) \in \mathcal{A}_u$. Together with the construction of $\widehat{\mathcal{I}}$, these observations imply that d has at most one r-successor in $\widehat{\mathcal{I}}$.

Next, we show that $\widehat{\mathcal{I}} \not\models q = C_0(a_0)$. Assume to the contrary that there is a match π of q in $\widehat{\mathcal{I}}$, i.e., a mapping $\pi : \mathsf{term}(q) \to \Delta^{\widehat{\mathcal{I}}}$ such that $\pi(a_0) = a_0^{\widehat{\mathcal{I}}}$, $A(t) \in q$ implies $\pi(t) \in A^{\widehat{\mathcal{I}}}$, and $r(t, t') \in q$ implies $(\pi(t), \pi(t')) \in r^{\widehat{\mathcal{I}}}$.[4] We prove that this implies the existence of a match π' for q in $\mathcal{I}$, which yields a contradiction to $\mathcal{I} \not\models q$. First, we need the following claim.

Claim 2. For all $\alpha, \beta \in \mathsf{Ind}(\mathcal{A}_u)$ with $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$ and all $\mathcal{ELI}$-concepts C, we have $\alpha \in C^{\mathcal{I}}$ iff $\beta \in C^{\mathcal{I}}$.

Assume to the contrary that $\alpha \in C^{\mathcal{I}}$ and $\beta \notin C^{\mathcal{I}}$. By construction of $\mathcal{A}_u$ and since $\mathsf{tail}(\alpha) = \mathsf{tail}(\beta)$, we find an ABox-isomorphism $\iota : \mathsf{Ind}(\mathcal{A}_u) \to \mathsf{Ind}(\mathcal{A}_u)$ with $\iota(\alpha) = \beta$, i.e., ι satisfies the following properties:

- if $A(\gamma) \in \mathcal{A}_u$, then $A(\iota(\gamma)) \in \mathcal{A}_u$;
- if $r(\gamma, \gamma') \in \mathcal{A}_u$, then $r(\iota(\gamma), \iota(\gamma')) \in \mathcal{A}_u$.

Define a new interpretation $\mathcal{I}'$ as $\mathcal{I}$, but put $\gamma^{\mathcal{I}'} = \iota(\gamma)^{\mathcal{I}}$ for all $\gamma \in \mathsf{Ind}(\mathcal{A}_u)$. Clearly, $\mathcal{I}'$ is a model of both $\mathcal{A}_u$ and $\mathcal{T}$. It follows that $\mathcal{A}_u, \mathcal{T} \not\models C(\alpha)$. This is a contradiction since $\mathcal{I}$ is a minimal model with $\mathcal{I} \models C(\alpha)$, implying $\mathcal{A}_u, \mathcal{T} \models C(\alpha)$, which finishes the proof of the claim.

We start the construction of the match π' of q in $\mathcal{I}$ as follows:

- start with setting $\pi'(a_0) = a_0$;
- if $\pi'(x) = \alpha$ for some $\alpha \in \mathsf{Ind}(\mathcal{A}_u)$, $r(x, y) \in q$, and $\pi(y) = b \in \mathsf{Ind}(\mathcal{A})$, then set $\pi'(y) = \alpha r b$.

It thus remains to define $\pi'(x)$ for all variables x in q such that $\pi(x) \neq a$ for all $a \in \mathsf{Ind}(\mathcal{A})$. To this end, consider a $t \in \mathsf{terms}(q)$ with $\pi'(t)$ already defined. Then $\pi(t) = a$ for some $a \in \mathsf{Ind}(\mathcal{A})$ and $\pi'(t) = \alpha$ for some $\alpha \in \mathsf{Ind}(\mathcal{A}_u)$ with $\mathsf{tail}(\alpha) = a$. Let V be the variables in q such that $\pi(v)$ is an element of $\widehat{\mathcal{I}}|_a$, the restriction of $\widehat{\mathcal{I}}$ to the tree rooted at a. Moreover, let $\mathcal{J}$ be the smallest subtree interpretation of $\widehat{\mathcal{I}}|_a$ that contains $\pi(v)$ for all $v \in V$. Since $\mathcal{J}$ is a finite tree interpretation, there is an $\mathcal{ELI}$-concept $C_{\mathcal{J}}$ that is satisfied at the root a of $\mathcal{J}$ and such that $\mathcal{J}$ is homomorphically embeddable into any model of $C_{\mathcal{J}}$. By construction of $\widehat{\mathcal{I}}$ and of $C_{\mathcal{J}}$, we have $\mathcal{I} \models C_{\mathcal{J}}(a)$. By Claim 1, this yields $\mathcal{I} \models C_{\mathcal{J}}(\alpha)$. Thus, there is a homomorphism h that embeds $\mathcal{J}$ into $\mathcal{I}$ such that $h(a) = \alpha$. To define the match π' for the variables in V, compose π with h. It can be verified that, by applying this construction for all $a \in \mathsf{Ind}(\mathcal{A})$, we ontain a match π' for q in $\mathcal{I}$. ❏

[4] Here, we view q as a tree-shaped CQ whose root is the individual name a_0 and whose non-root nodes are all variables.

D Proofs for Section 5

D.1 FO-rewritability and Locality

In this section, we prove basic results about FO-rewritability. The main result is Theorem 14 in which we characterize, for unraveling tolerant $\mathcal{ALCFI}$-TBoxes and $\mathcal{ELI}_\bot$-concepts C when there exists an FO-formula $\varphi_C(x)$ such that $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff $\mathcal{I}_\mathcal{A} \models \varphi_C[a]$, for all ABoxes $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$.

Definition 6. *Let $\mathcal{T}$ be a $\mathcal{ALCFI}$-TBox, Σ a signature, and C na $\mathcal{ELI}_\bot$-concept. Then C is called FO-rewritable w.r.t. $\mathcal{T}$ and Σ iff there exists an FO-formula $\varphi_C(x)$ such that*

$$(\mathcal{T}, \mathcal{A}) \models C(a) \quad \Leftrightarrow \quad \mathcal{I}_\mathcal{A} \models \varphi_C[a],$$

for all Σ-Aboxes $\mathcal{A}$.

To characterize FO-rewritability, we require the following notions. A *pointed interpretation* is a pair consisting of an interpretation $\mathcal{I}$ and some $d \in \Delta^\mathcal{I}$. We identify finite pointed interpretations with pairs $(\mathcal{A}, a)$ consisting of an ABox $\mathcal{A}$ and a distinguished individual name a. The n*-neighbourhood* $\mathcal{I}^n_d$ *of* d in an interpretation $\mathcal{I}$ is the relativization of $\mathcal{I}$ to the set

$$\{d' \in \Delta^\mathcal{I} \mid \mathsf{dist}_\mathcal{I}(d, d') \leq n\}.$$

A class $\mathcal{K}$ of pointed interpretations $(\mathcal{I}, d)$ is n*-local* if $(\mathcal{I}^n_d, d) \in \mathcal{K}$ whenever $(\mathcal{I}, d) \in \mathcal{K}$. Classes definable by modal logic formulas or, equivalently, $\mathcal{ALC}$-concepts are known to be n-local for some $n > 0$.

We are interested in the following classes of pointed interpretations. For an $\mathcal{ALCFI}$-TBox $\mathcal{T}$, signature Σ, and $\mathcal{ELI}$-concept C let

$$\mathcal{K}_{C,\mathcal{T},\Sigma} = \{(\mathcal{A}, a) \mid (\mathcal{T}, \mathcal{A}) \models C(a),\ \mathcal{A} \text{ a } \Sigma\text{-ABox}\}$$

Call a class $\mathcal{K}$ of pointed Σ-ABoxes *FO-definable* iff there exists an FO-formula $\varphi(x)$ such that

$$\mathcal{K} = \{(\mathcal{A}, a) \mid \mathcal{I}_\mathcal{A} \models \varphi[a],\ \mathcal{A} \text{ a } \Sigma\text{-ABox}\}.$$

Clearly, $\mathcal{K}_{C,\mathcal{T},\Sigma}$ if FO-definable iff C is FO-rewritable w.r.t. $\mathcal{T}$ and Σ. For simplicity, in what follows we mostly consider classes $\mathcal{K}_{A,\mathcal{T},\Sigma}$, where A is an atomic concept. Clearly, $\mathcal{K}_{C,\mathcal{T},\Sigma}$ if FO-definable if $\mathcal{K}_{A,\mathcal{T}',\Sigma}$ with $\mathcal{T}' = \mathcal{T} \cup \{A \equiv C\}$ is FO-definable. In contrast to modal logic, even if $\mathcal{K}_{A,\mathcal{T},\Sigma}$ is FO-definable, it is no necessarily n-local for any n.

Example 2. Let $\mathcal{T} = \{A \sqsubseteq \top, B \sqsubseteq \bot\}$ and $\Sigma = \{A, B\}$. $\mathcal{K}_{A,\mathcal{T},\Sigma}$ is defined by $\varphi_A(x) = A(x) \vee \exists y.B(y)$. Note that $(\mathcal{A}, a) \in \mathcal{K}_{A,\mathcal{T},\Sigma}$ for $\mathcal{A} = \{\top(a), B(b)\}$ because $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$. However, $\mathcal{A}^n_a = \{\top(a)\}$ for all $n > 0$ and so $\mathcal{A}^n_a \notin \mathcal{K}_{A,\mathcal{T},\Sigma}$, for any $n > 0$. Thus, $\mathcal{K}_{A,\mathcal{T},\Sigma}$ is not n-local, for any $n > 0$.

It turns out that, in some sense, inconsistency is the only reason for not being n-local for any n for FO-definable classes of the form $\mathcal{K}_{A,\mathcal{T},\Sigma}$. We now make this claim precise.

A class $\mathcal{K}$ of pointed interpretations $(\mathcal{I}, d)$ is *weakly n-local* if whenever $(\mathcal{I}, d) \in \mathcal{K}$, then $(\bigcup^+_{e \in \Delta^\mathcal{I}} \mathcal{I}^n_e, d) \in \mathcal{K}$, where $\bigcup^+_{e \in \Delta^\mathcal{I}} \mathcal{I}^n_e$ denotes the disjoint union of all $\mathcal{I}^n_e$ with

$e \in \Delta^{\mathcal{I}}$. $\mathcal{K}$ is *monotone* if it is closed under expansions: if $(\mathcal{I}, d) \in \mathcal{K}$ and $\mathcal{I}'$ is an interpretation with $\Delta^{\mathcal{I}'} \supseteq \Delta^{\mathcal{I}}$ and $X^{\mathcal{I}'} \supseteq X^{\mathcal{I}}$ for all $X \in \Sigma$, then $(\mathcal{I}', d) \in \mathcal{K}$. $\mathcal{K}$ *reflects disjoint copies* if whenever $(\mathcal{I}, d) \in \mathcal{K}$ and $\mathcal{I}$ is the disjoint union of isomorphic copies $\mathcal{I}_0, \dots, \mathcal{I}_n$, then $(\mathcal{I}_0, d) \in \mathcal{K}$ for the (unique) $\mathcal{I}_0$ with $d \in \Delta^{\mathcal{I}_0}$.

Lemma 17. *If a FO-definable class of pointed interpreations is monotone and reflects disjoint copies, then it is weakly n-local, for some $n > 0$. More precisely, if $\mathcal{K}$ is* FO-definable *using an FO-formula of quantifier depth q, then $\mathcal{K}$ is weakly n-local, for $n = 2^q - 1$.*

Proof. Assume $(\mathcal{I}, d) \in \mathcal{K}$ and $\varphi(x)$ defines $\mathcal{K}$. Let $(\mathcal{J}_0, d)$ be the disjoint union of $(\mathcal{I}, d)$ and q copies of $\bigcup^{+}_{e \in \Delta^{\mathcal{I}}} \mathcal{I}^n_e$. Let $(\mathcal{J}_1, d)$ be the disjoint union of $(\bigcup^{+}_{e \in \Delta^{\mathcal{I}}} \mathcal{I}^n_e, d)$ and q additional copies of $\bigcup^{+}_{e \in \Delta^{\mathcal{I}}} \mathcal{I}^n_e$. One can show, using Ehrenfeucht-Fraisse games with q rounds, that $\mathcal{J}_0 \models \varphi_A[d]$ iff $\mathcal{J}_1 \models \varphi_A[d]$: A winning strategy for the duplicator is to keep in round m the distance 2^{q-m}: if the spoiler's move (say, e) in round m is within distance 2^{q-m} to a pebbled element, then the duplicator plays according to the "local" isomorphism for all d's within the neighbourhood $\mathcal{I}_e^{2^{q-m}}$. Otherwise, the duplicator responds within a new isomorphic copy.

By monotonicity, we have $(\mathcal{J}_0, d) \in \mathcal{K}$. Thus we obtain $(\mathcal{J}_1, d) \in \mathcal{K}$. Since $\mathcal{K}$ reflects disjoint copies, we obtain that $(\bigcup^{+}_{e \in \Delta^{\mathcal{I}}} \mathcal{I}^n_e, d) \in \mathcal{K}$, as required. ❑

Theorem 13. *Let A be an atomic concept. If A is FO-rewritable w.r.t. $\mathcal{T}$ and Σ, then $\mathcal{K}_{A,\mathcal{T},\Sigma}$ it is weakly n-local, for some $n > 0$.*

Proof. Follows from Lemma 17 and the observation that $\mathcal{K}_{A,\mathcal{T},\Sigma}$ is monotone and reflects disjoint copies. ❑

The converse of Theorem 13 can be proved for unraveling tolerant TBoxes. An interpretation $\mathcal{I}$ is a *tree interpretation* if $(\Delta^{\mathcal{I}}, \bigcup_{r \text{ a role}} r^{\mathcal{I}})$ is a symmetric (possibly infinite) tree and $r^{\mathcal{I}} \cap s^{\mathcal{I}} = \emptyset$ for all distinct roles r and s. Observe that the unraveling $\mathcal{A}_u$ of an ABox $\mathcal{A}$ is the disjoint union of tree interpretations. An interpretation $\mathcal{I}$ is *n-generated by d* if $\mathcal{I}^n_d = \mathcal{I}$. A tree interpretation $\mathcal{I}$ with root ρ has depth n if it is n-generated by ρ. We define the notion of a tree interpretation of depth n without redundancies by induction: A tree interpretation of depth $n+1$ has no redundancies if for any two sons d_1, d_2 of its root ρ with $(\rho, d_i) \in r^{\mathcal{I}}$ for $i = 1, 2$, the subinterpretation $\mathcal{I}_{d_1}$ and $\mathcal{I}_{d_2}$ are not isomorphic and do not have redundancies. Clearly, the number of tree interpretations of depth n without redundancies is finite, up to isomorphisms.

Theorem 14. *Assume $\mathcal{T}$ is an unraveling tolerant $\mathcal{ALCFI}$-TBox, Σ a signature, and A an atomic concept. Then the following conditions are equivalent:*

- *A is FO-rewritable w.r.t. $\mathcal{T}$ and Σ;*
- *$\mathcal{K}_{A,\mathcal{T},\Sigma}$ is weakly n-local for some $n > 0$.*

Proof. (sketch) Let $\mathcal{K}_{A,\mathcal{T},\Sigma}$ be weakly n-local for some $n > 0$. Let $\mathcal{C}_A$ be the finite set of Σ-tree-interpretations with root ρ and of depth n without redundancies such that $\mathcal{I} \models A[\rho]$. We can define the class of all $(\mathcal{A}, a)$ with $\mathcal{A}$ a Σ-ABoxes and a a distinguished element whose unraveling contains some $(\mathcal{I}, d) \in \mathcal{C}_A$ as a subinterpretation using an FO-formula $\varphi_A(x)$. Determine the formula $\varphi_\perp$ in the same way. Now $\varphi_A(x) \vee \exists x \varphi_\perp(x)$ is as required. ❑

D.2 Proofs of Lemma 5 and Theorem 8 (Part 1)

Lemma 5 A materializable $\mathcal{ALCFI}$-TBox of depth one is FO-rewritable for CQs iff all atomic concepts are FO-rewritable w.r.t. $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T})$.

Proof. We prove the direction from right to left. Assume that all atomic concepts are FO-rewritable w.r.t. $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T})$. It is sufficient to show that $\mathcal{T}$ is FO-rewritable for ELIQs. Consider an $\mathcal{ELI}$-concept C. Take FO-formulas $\varphi_A(x)$ for $A \in \mathsf{sig}(\mathcal{T}) \cup \{\bot\}$ such that $(\mathcal{T}, \mathcal{A}) \models A(a)$ iff $\mathcal{I}_\mathcal{A} \models \varphi_A(a)$, for all Σ-ABoxes $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$.

For every subconcept D of C we construct an FO-formula $\varphi_D(x)$ such that

$$(\mathcal{T}, \mathcal{A}) \models D(b) \quad \Leftrightarrow \quad (\mathcal{T}, \mathcal{A}) \models \varphi_D[b],$$

for all Σ-ABoxes $\mathcal{A}$ and $b \in \mathsf{Ind}(\mathcal{A})$. If D is a concept name A from $\mathcal{T}$ or $A = \bot$, we can take the formula $\varphi_A(x)$ from above. Otherwise we set $\varphi_A(x) = A(x) \vee \varphi_\bot(x)$. If $D = D_1 \sqcap D_2$, then we can take $\varphi_{D_1}(x) \wedge \varphi_{D_2}(x)$. Now assume $D = \exists r.D_0$. Let Ξ be the set of $\mathsf{sig}(\mathcal{T}) \cup \mathsf{sig}(C)$-NH ABoxes $\mathcal{A}, f$ such that $(\mathcal{T}, \mathcal{A}) \models \exists r.D_0(f)$. Replace in any $\mathcal{A}^x \in \Xi$, any assertion $B(y)$ by $\varphi_B(y)$ and denote the result by $\mathcal{A}^x_{\mathsf{FO}}$. Then we set

$$\varphi_D(x) = (\exists y.r(x,y) \wedge \varphi_{D_0}(y)) \vee \bigvee_{(\mathcal{A},f)\in\Xi} \exists \boldsymbol{x}. \bigwedge \mathcal{A}^x_{\mathsf{FO}}$$

Using the condition that $\mathcal{T}$ has depth one, one can show that $\varphi_D(a)$ is as required.

❑

Lemma 18. *For every $\mathcal{ALCI}$-TBox $\mathcal{T}$, signature Σ, and $\mathcal{ELI}_\bot$-concept C: C is FO-rewritable w.r.t. $\mathcal{T}$ and Σ iff, for the interpretation $\mathcal{I}$ constructed in the proof of Theorem 5, the class*

$$\mathsf{Hom}_{\Sigma,P}(\mathcal{I}) = \{\mathcal{A} \mid \ \mathcal{A} \text{ a } \Sigma \cup \{P\}\text{-ABox}, \mathsf{Hom}(\mathcal{A}, \mathcal{I})\}$$

is FO-definable.

Proof. Let φ be an FO-sentence defining $\mathsf{Hom}_{\Sigma,P}(\mathcal{I})$. Recall that we have $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff not $\mathsf{Hom}(\mathcal{A}_{P(a)}, \mathcal{I})$, where $\mathcal{A}_{P(a)}$ results from $\mathcal{A}$ by adding $P(a)$ to $\mathcal{A}$ and removing all other assertions using P from $\mathcal{A}$. Thus $(\mathcal{T}, \mathcal{A}) \models C(a)$ iff $\mathcal{A}_{P(a)} \models \neg\varphi$. The latter conndition holds iff $\mathcal{A} \models \neg\varphi[P/a]$, where $\varphi[P/a]$ denotes the result of replacing every occurrence of $P(z)$ by $z = a$. Thus,

$$\mathcal{A} \models \neg\varphi[P/a] \quad \Leftrightarrow \quad (\mathcal{T}, \mathcal{A}) \models C(a)$$

and we have shown FO-rewritability of C w.r.t. $\mathcal{T}$ and Σ.

Conversely, assume that

$$\mathcal{A} \models \varphi_C(a) \quad \Leftrightarrow \quad (\mathcal{T}, \mathcal{A}) \models C(a)$$

holds for all Σ-ABoxes $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$. We have not $\mathsf{Hom}(\mathcal{A}, \mathcal{I})$ iff $(\mathcal{T}, \mathcal{A}) \models \exists v.(P(v) \wedge C(v))$, for all $\Sigma \cup \{P\}$-ABoxes $\mathcal{A}$. Thus $\neg\exists v.(P(v) \wedge \varphi_C(v))$ defines $\mathsf{Hom}_{\Sigma,P}(\mathcal{I})$. ❑

We can now prove the first part of Theorem 8.

Theorem 8 (Part 1) FO-rewritability for CQs is decidable in NEXPTIME, for the class of materializable $\mathcal{ALCI}$-TBoxes of depth one.

Proof. By Lemma 5, it is sufficient to show that FO-rewritability of atomic concepts with respect to $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T})$ is in NEXPTIME. By [16], FO-definability of $\{\mathcal{I}' \mid \mathsf{Hom}(\mathcal{I}', \mathcal{I})\}$ is decidable in NP. By Lemma 18, we obtain a NEXPTIME upper bound since the templates $\mathcal{I}$ can be constructed in exponential time. ❑

D.3 Proof of Lemma 6, Lemma 7 and Theorem 9 (Part 1)

We formulate Lemma 6 again.

Lemma 6 For every materializable $\mathcal{ALCFI}$-TBox $\mathcal{T}$ of depth one, every $A \in \mathsf{sig}(\mathcal{T})$, every ABox $\mathcal{A}$, and every $a \in \mathsf{Ind}(\mathcal{A})$, $(\mathcal{T}, \mathcal{A}) \models A(a)$ iff $A(a) \in \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$. $\bot(a) \in \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$ iff $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$.

Proof. Clearly, the program $\Pi_{\mathcal{T}}$ is sound in the sense that $A(a) \in \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$ implies $(\mathcal{T}, \mathcal{A}) \models A(a)$ and $\bot(a) \in \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$ implies that $\mathcal{A}$ is not consistent w.r.t. $\mathcal{T}$.

Now assume that $A(a_0) \notin \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$ for some $a_0 \in \mathsf{Ind}(\mathcal{A})$ and $A \in \Sigma = \mathsf{sig}(\mathcal{T})$. (The case $A = \bot$ is considered similarly and left to the reader.) Define $\mathcal{A}' = \mathcal{A} \cup \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$. For every $a \in \mathsf{Ind}(\mathcal{A}')$, let $\mathcal{A}_a$ be the (unique up to renaming) maximal Σ NH with distinguished individual name a such that there is an assignment π into $\mathcal{I}_{\mathcal{A}'}$ with $\pi(x_a) = a$ and $\mathcal{I}_{\mathcal{A}'} \models_\pi \mathcal{A}^x_a$. (Equivalently, there in a homomorphism $f_a : \mathcal{I}_{\mathcal{A}_a} \to \mathcal{I}_{\mathcal{A}'}$ with $f(a) = a$.) Observe that there does not exist any $a \in \mathsf{Ind}(\mathcal{A})$ such that $\mathcal{A}_a$ is not consistent w.r.t. $\mathcal{T}$: otherwise, by the definition of $\Pi^{\infty}_{\mathcal{T}}$, $\bot(a) \in \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$ and, therefore, $A(a_0) \in \Pi^{\infty}_{\mathcal{T}}(\mathcal{A})$; so we have derived a contradiction. The following three properties of $\mathcal{A}_a$ are readily checked:

1. $(\mathcal{T}, \mathcal{A}_a) \models B(e)$ iff $B(e) \in \mathcal{A}_a$, for all $e \in \mathsf{Ind}(\mathcal{A}_a)$ and all $B \in \Sigma$;
2. For every $b \in \mathsf{Ind}(\mathcal{A})$ with $r(a, b) \in \mathcal{A}$ for some role r, there exists an $e \in \mathsf{Ind}(\mathcal{A}_a)$ and an assignment π with $\pi(x_a) = a$ and $\pi(x_e) = b$ such that $\mathcal{A}' \models_\pi \mathcal{A}_a$;
3. For every $\mathcal{ELIF}$-concept C, every $e \in \mathsf{Ind}(\mathcal{A}_a)$ and every assignment π such that $\mathcal{A}' \models \mathcal{A}_a$ with $\pi(x_a) = a$: $(\mathcal{T}, \mathcal{A}_a) \models C(e)$ implies $(\mathcal{T}, \mathcal{A}) \models C(\pi(x_e))$.

For every $a \in \mathsf{Ind}(\mathcal{A})$ we can take a minimal interpretation $\mathcal{I}_a$ of $(\mathcal{T}, \mathcal{A}_a)$. By Point 1, we have $e^{\mathcal{I}_a} \in B^{\mathcal{I}_a}$ iff $B(e) \in \mathcal{A}_a$ for all $B \in \Sigma$ and $e \in \mathsf{Ind}(\mathcal{A}_a)$. We can assume that $\mathcal{I}_a$ is an unfolded interpretation as constructed in the proof of Lemma 12. Define an interpretation $\mathcal{I}$ satisfying $\mathcal{A}'$ by hooking to every $a \in \mathsf{Ind}(\mathcal{A})$ the "anonymous tree" $\mathcal{I}'_a$ generated by a in $\mathcal{I}_a$.

We show that $\mathcal{I}$ is a model of $\mathcal{T}$. Suppose that $C \sqsubseteq D \in \mathcal{T}$ is refuted in $\mathcal{I}$. Since $\mathcal{T}$ has depth one, there exists $a \in \mathsf{Ind}(\mathcal{A})$ such that $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and $a^{\mathcal{I}} \notin D^{\mathcal{I}}$. We show that $C \sqsubseteq D$ is refuted in $\mathcal{I}_a$ and, thus, derive a contradiction. It is sufficient to show that $a^{\mathcal{I}} \in C^{\mathcal{I}_a}$ and $a^{\mathcal{I}} \notin D^{\mathcal{I}_a}$. But $a \notin D^{\mathcal{I}_a}$ follows from Point 3 and $a \in C^{\mathcal{I}_a}$ follows from the condition that C has depth one and Points 1 and 2. ❑

We now prove the following extended version of Lemma 7.

Lemma 19. *Let $\mathcal{T}$ be a materializable $\mathcal{ALCFI}$-TBox of depth one, Σ a signature, and A an atomic concept. Then the following conditions are equivalent:*

- *A is FO-rewritable w.r.t. $\mathcal{T}$ and Σ;*
- *$\mathcal{K}_{A,\mathcal{T},\Sigma}$ is weakly n-local for some $n > 0$;*
- *A is bounded in $\Pi_\mathcal{T}$ for Σ-ABoxes.*

Proof. (3) implies (1). Assume A is bounded in $\Pi_\mathcal{T}$ for Σ-ABoxes. Take $k > 0$ such that $A(a) \in \Pi^k_\mathcal{T}(\mathcal{A})$ iff $A(a) \in \Pi^\infty_\mathcal{T}(\mathcal{A})$ for all Σ-ABoxes $\mathcal{A}$. From $\Pi^k_\mathcal{T}$, one can directly construct a FO-formula $\varphi_A(x)$ such that $A(a) \in \Pi^k_\mathcal{T}(\mathcal{A})$ iff $\mathcal{A} \models \varphi_A[a]$ for all Σ-ABoxes $\mathcal{A}$. By Lemma 6, we obtain $(\mathcal{T}, \mathcal{A}) \models A(a)$ iff $\mathcal{A} \models \varphi_A[a]$, as required.

(1) implies (2). This is Theorem 13.

(2) implies (3). Assume $\mathcal{K}_{A,\mathcal{T},\Sigma}$ is weakly n-local. By unraveling tolerance, it is sufficient to show that there exists a $k \geq 1$ such that for every Σ-tree ABox $(\mathcal{A}, a)$ of depth at most n: $A(a) \in \Pi^\infty_\mathcal{T}(\mathcal{A})$ iff $A(a) \in \Pi^k_\mathcal{T}(\mathcal{A})$. But this is trivial since it is sufficient to consider the finite set of ABoxes $(\mathcal{A}, a)$ without redundancies. ❏

The boundedness problem for A in $\Pi_\mathcal{T}$ for Σ-ABoxes can easily be translated into a boundedness problem for monadic least fixed points based on FO over trees in the sense investigated in [18]. Thus we obtain from [18]:

Theorem 15. *For $\mathcal{T}$ a unfolding tolerant (equivalently, materializable) $\mathcal{ALCFI}$-TBox of depth one, Σ a signature, and A an atomic concept, boundedness of A in $\Pi_\mathcal{T}$ for Σ-ABoxes is decidable.*

The first part of Theorem 9 now follows from Lemma 19 and Lemma 5.

Theorem 9 (Part 1) FO-rewritability for CQs is decidable, for the class of materializable $\mathcal{ALCFI}$-TBoxes of depth one.

D.4 Horn-$\mathcal{ALCFI}$

In this section we prove the claims about Horn-TBoxes in Theorems 8 and 9. In fact, the conditions we require to prove decidability (and decidability in NExpTime) are much less restrictive than those given in the theorems.

To begin with, observe that for a Horn-$\mathcal{ALCFI}$-TBox $\mathcal{T}$ and ABox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$ the interpretation $\mathcal{I}$ corresponding to $\mathcal{A}_c$ as constructed above is a minimal model of $\mathcal{T}$. In what follows we call it the *canonical minimal model of* $(\mathcal{T}, \mathcal{A})$ and denote it by $\mathcal{I}_{\mathcal{T},\mathcal{A}}$. Recall that its domain consists of sequences of the form $ar_1C_1 \cdots r_kC_k$ with $a \in \mathsf{Ind}(\mathcal{A})$, $r_1, \dots, r_k$ roles that occur in $\mathcal{T}$, and $C_1, \dots, C_k \in \mathsf{sub}(\mathcal{T})$.

Recall that 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 syntax rule in:

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

Assume a Horn-$\mathcal{ALCFI}$ TBox $\mathcal{T} = \{\top \sqsubseteq C_\mathcal{T}\} \cup \mathcal{F}$ is given. We transform $\mathcal{T}$ into another Horn-$\mathcal{ALCFI}$-TBox $\mathcal{T}^1$ of depth one such that, under certain conditions, $\mathcal{T}$ is FO-rewritable for CQ iff all atomic concepts are FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$-ABoxes. The right-role-depth $d(R)$ of a concept R is defined as follows:

$$
\begin{aligned}
d(\top) = d(\bot) &= 0\\
d(A) = d(\neg A) &= 0\\
d(R \sqcap R') &= \max\{d(R), d(R')\}\\
d(L \rightarrow R) &= d(R)\\
d(\exists r.R) &= d(R) + 1\\
d(\forall r.R) &= d(R) + 1
\end{aligned}
$$

In what follows, we assume that $C_\mathcal{T}$ satisfies the following condition:

- every concept L in $\mathcal{T}$ has role depth one;
- every L in $\mathcal{T}$ that occurs within the scope of $\exists r$ or $\forall r$ within a subconcept R of $C_\mathcal{T}$ is a concept name.

One can easily transform $C_\mathcal{T}$ into such a concept and preserve FO-rewritability for CQ by introducing abbreviations $A \equiv L$. We define a new concept $C^1_\mathcal{T}$ by applying, recursively, the following rule $(*)$ to an outermost occurrence of R in $C_\mathcal{T}$, where R has the form $\mathsf{Op}\ r.R'$, $\mathsf{Op} \in \{\exists, \forall\}$ such that R' has right-role depth at least one:

$(*)$ Replace $\mathsf{Op}\ r.R'$ by $\mathsf{Op}\ r.E$, where E is a fresh concept name, and add $E \rightarrow R'$ as a conjunct to $C_\mathcal{T}$.

We sometimes denote the fresh concept E replacing R by E_R. Observe that $C^1_\mathcal{T}$ has right-role-depth one and is still a Horn-$\mathcal{ALCFI}$-TBox. Note that $C^1_\mathcal{T}$ is of polynomial size. We set $\mathcal{T}^1 = \{\top \sqsubseteq C^1_\mathcal{T}\} \cup \mathcal{F}$. The following lemma is be checked by close inspection of the construction of $\mathcal{A}_c$.

Lemma 20. *For every* $\mathsf{sig}(\mathcal{T})$*-ABox* $\mathcal{A}$*:*

1. *$\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ iff $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}^1$;*
2. *There is bijection $f : \Delta^{\mathcal{I}_{\mathcal{T},\mathcal{A}}} \rightarrow \Delta^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}}$ with*
   - *$f(a) = a$, for all $a \in \mathsf{Ind}(\mathcal{A})$;*
   - *$d \in A^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$ iff $f(d) \in A^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}}$, for all $d \in \Delta^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$ and $A \in \mathsf{sig}(\mathcal{T})$;*
   - *$(d, d') \in r^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$ iff $(f(d), f(d')) \in r^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}}$ for all $d, d' \in \Delta^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$ and $A \in \mathsf{sig}(\mathcal{T})$;*
3. *for every CQ q not containing fresh concepts:*

$$\mathcal{T}, \mathcal{A} \models q(\boldsymbol{a}) \quad \Leftrightarrow \quad \mathcal{T}^1, \mathcal{A} \models q(\boldsymbol{a}).$$

Say that $\mathcal{T} = \{\top \sqsubseteq C_\mathcal{T}\} \cup \mathcal{F}$ contains a *return point* if there exists a subconcept $R = \exists r.R'$ of $C_\mathcal{T}$ such that $\mathsf{func}(r) \notin \mathcal{F}$ and R' has a subconcept $\forall r^-.R''$ with R'' of right-role-depth at least one which is not within the scope of any $\exists s$ or $\forall s$ in R'. Note that if $\mathcal{T}$ is a Horn-$\mathcal{ALCF}$-TBox or has role depth at most two, then $\mathcal{T}$ contains no return point.

Lemma 21. *Let $\mathcal{T}$ be a Horn-$\mathcal{ALCFI}$-TBoxes that has no return point. Then $\mathcal{T}$ is FO-rewritable for CQ iff every atomic concept is FO-rewritable w.r.t. $\mathcal{T}^1$ and* $\mathsf{sig}(\mathcal{T})$.

Proof. Assume $\mathcal{T} = \{\top \sqsubseteq C_{\mathcal{T}}\} \cup \mathcal{F}$, where $\mathcal{F}$ is a set of functionality assertions.

Assume first that every atomic concept is FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$. Recall that $\mathcal{T}^1$ is materializable and has role depth one. Then a straighforward extension of Lemma 5 (to ABoxes of a fixed signature) shows that all CQs are FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$-ABoxes. By Point 3 of Lemma 20, $\mathcal{T}$ is FO-rewritable for CQ.

For the converse direction, assume that $\mathcal{T}$ is FO-rewritable.

Claim 1. For any fresh E_R with R not in the scope of any $\exists r$ with $r \notin \mathsf{func}(r)$ there exists a $\mathcal{ELI}$-concept C such that for every $\mathsf{sig}(\mathcal{T})$-ABox $\mathcal{A}$:

$$E_R^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}} = C^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}}.$$

The proof is by induction: we consider the first step, the induction step is similar and left to the reader. Assume R is replaced by E_R in the situation described in the claim. There are $R_1, \dots, R_{m+1}$ such that R_{m+1} is a conjunct of $C_{\mathcal{T}}$ and

$$R_1 = (L_0 \rightarrow R_0), \quad R_{j+1} = (L_j \rightarrow (R_j \sqcap R_j')),$$

for $1 \leq j \leq m$ and (i) $R_0 = \forall r.R$ or (ii) $R_0 = \exists r.R$ and $\mathsf{func}(r) \in \mathcal{F}$. In both cases, set

$$C = \exists r^-.(L_0 \sqcap \cdots \sqcap L_{m+1}).$$

It is readily checked that $E_R^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}} = C^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}}$, as required.

It follows from Claim 1, Lemma 20, and the condition that $\mathcal{T}$ is FO-rewritable for CQ (and hence for ELIQ) that every fresh E_R with R not in the scope of any $\exists r$ with $r \notin \mathsf{func}(r)$ is FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$-ABoxes. Now, since $\mathcal{T}$ has no return point, it is readily checked that for any fresh E_R with R within the scope of some $\exists r$, $\mathsf{func}(r) \notin \mathcal{F}$, we have $E^{\mathcal{I}_{\mathcal{T}^1,\mathcal{A}}} \cap \mathsf{Ind}(\mathcal{A}) = \emptyset$. Thus, any such E_R is FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$-ABoxes.

We have shown that all fresh E_R are FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$-ABoxes. By Lemma 20, all atomic concepts in $\mathsf{sig}(\mathcal{T})$ are FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$-Aboxes. Thus all atomic concepts are FO-rewritable w.r.t. $\mathcal{T}^1$ and $\mathsf{sig}(\mathcal{T})$-ABoxes, as required. ❑

Observe that no Horn-$\mathcal{ALCF}$-TBox nor Horn-$\mathcal{ALCFI}$-TBox of depth two have a return point. Thus, we obtain the claims about Horn-TBoxes in Theorems 8 and 9 from Lemma 21 and the fact that for materializable $\mathcal{ALCFI}$-TBoxes ($\mathcal{ALCI}$-TBoxes) of depth 1, one can decide (in NEXPTIME) FO-rewritability of atomic concepts for Σ-ABoxes, where Σ is an arbitrary signature.

E Proofs for Section 6

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

Using similar modifications of Definition 1, 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 1 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 11) and then modify the TBoxes constructed in its proof to show the non-dichomy result (Theorem 10). The modified version of Theorem 11 is as follows:

Theorem 16. *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 [1] 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 [1]. 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 $f : \{0, \ldots, n\} \times \{0, \ldots, m\} \rightarrow \mathfrak{T}$ 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 [1], 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 $(T_i, T_\ell) \in V$:

$$
\begin{aligned}
T_{\mathsf{final}} &\sqsubseteq Y \sqcap U \sqcap R \\
\exists x.(U \sqcap Y \sqcap T_j) \sqcap I_x \sqcap T_i &\sqsubseteq U \sqcap Y \\
\exists y.(R \sqcap Y \sqcap T_\ell) \sqcap I_y \sqcap T_i &\sqsubseteq R \sqcap Y \\
\exists x.(T_j \sqcap Y \sqcap \exists y.Y) & \\
\sqcap \exists y.(T_\ell \sqcap Y \sqcap \exists x.Y) & \\
\sqcap I_x \sqcap I_y \sqcap C \sqcap T_i &\sqsubseteq Y \\
Y \sqcap T_{\mathsf{init}} &\sqsubseteq A
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{B}_x \sqcap \exists x.\exists x^-.\mathcal{B}_x &\sqsubseteq I_x \\
\mathcal{B}_y \sqcap \exists y.\exists y^-.\mathcal{B}_y &\sqsubseteq I_y \\
\exists x.\exists y.\mathcal{B}_c \sqcap \exists y.\exists x.\mathcal{B}_c &\sqsubseteq C
\end{aligned}
$$

$$
\begin{aligned}
U &\sqsubseteq \forall y.\bot \\
R &\sqsubseteq \forall x.\bot \\
U &\sqsubseteq \forall x.U \\
R &\sqsubseteq \forall y.R \\
\bigsqcup_{1 \leq s < t \leq p} T_s \sqcap T_t &\sqsubseteq \bot
\end{aligned}
$$

$$
\begin{aligned}
D &\sqsubseteq \forall y^-.\bot \\
L &\sqsubseteq \forall x^-.\bot \\
D &\sqsubseteq \forall x.D \sqcap \forall x^-.D \\
L &\sqsubseteq \forall y.L \sqcap \forall y^-.L \\
Y \sqcap T_{\mathsf{init}} &\sqsubseteq D \sqcap L
\end{aligned}
$$

We note that the final five inclusions (and the concept names L and D) are not used in [1]. 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 [1] (the proof is easily extended to cover the additional concepts for the lower and left border):

Lemma 22. *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 22.

Lemma 23. $\mathfrak{P}$ *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 24. *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 1, we can find a model $\mathcal{I}$ of $(\mathcal{T}_{\mathfrak{P}}', \mathcal{A})$ refuting $\exists v.H_A(v)$. ❑

Lemma 25. *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 23. 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 24. 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 26. *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 25 and 26, we obtain Points 3 and 4 of Theorem 16 as well as Points 1 and 2 for consistent ABoxes. Thus, to prove Theorem 16 it remains to show the following lemma.

Lemma 27. *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 10.

Theorem 10 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 exists a polynomial reduction of deciding $v \in L$ 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 deciding $v \in L$.

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 10.

Lemma 28. *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 28, 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 27.

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.