Horn-Rewritability vs PTime Query Evaluation in Ontology-Mediated Querying

André Hernich[1], **Carsten Lutz**[2], **Fabio Papacchini**[1], **Frank Wolter**[1]
[1] University of Liverpool, UK
[2] University of Bremen, Germany
{hernich,papacchf,wolter}@liverpool.ac.uk, clu@uni-bremen.de

Abstract

In ontology-mediated querying with an expressive description logic (DL) $\mathcal{L}$, two desirable properties of a TBox $\mathcal{T}$ are (1) being able to replace $\mathcal{T}$ with a TBox formulated in the Horn-fragment of $\mathcal{L}$ without affecting the answers to conjunctive queries (CQs) and (2) that every CQ can be evaluated in PTIME w.r.t. $\mathcal{T}$. We investigate in which cases (1) and (2) are equivalent, finding that the answer depends on whether the unique name assumption (UNA) is made, on the DL under consideration, and on the nesting depth of quantifiers in the TBox. We also clarify the relation between query evaluation with and without UNA and consider natural variations of property (1).

1 Introduction

In ontology-mediated querying, description logic (DL) TBoxes are used to enrich incomplete data with domain knowledge, enabling more complete answers to queries [Poggi *et al.*, 2008; Bienvenu and Ortiz, 2015; Kontchakov and Zakharyaschev, 2014]. For expressive DLs such as $\mathcal{ALC}$ or $\mathcal{SHIQ}$, this results in query evaluation to be CONP-hard (in data complexity) [Schaerf, 1993; Hustadt *et al.*, 2007; Krisnadhi and Lutz, 2007]. Consequently, identifying computationally more well-behaved setups has been an important goal of research [Calvanese *et al.*, 2013]. In particular, this has led to the introduction of Horn-DLs, syntactically defined fragments of expressive DLs that fall within the Horn-fragment of first-order logic and enable polynomial time ontology-mediated querying, examples include Horn-$\mathcal{ALC}$ and Horn-$\mathcal{SHIQ}$ [Hustadt *et al.*, 2007; Eiter *et al.*, 2008; Ortiz *et al.*, 2010; 2011]. On top of enjoying lower data complexity, Horn-DLs come with several techniques that facilitate efficient query evaluation in practice such as the chase, query rewriting, and deterministic materialization [Bienvenu and Ortiz, 2015; Kontchakov and Zakharyaschev, 2014].

In this paper, we ask the converse question: *Assume that a TBox $\mathcal{T}$ is formulated in an expressive DL $\mathcal{L}$ and admits* PTIME *query evaluation. Does it follow that $\mathcal{T}$ can be replaced by a TBox $\mathcal{T}'$ formulated in the corresponding Horn-DL without affecting the answers to queries?* Let us make this more precise. We concentrate on queries that are conjunctive queries (CQ) since these are widely used in ontology-mediated querying and require $\mathcal{T}$ and $\mathcal{T}'$ to be *CQ-inseparable*, that is, to give exactly the same answers to any CQ on any ABox, see [Lutz and Wolter, 2010; Botoeva *et al.*, 2016a; 2016b]. We say that an $\mathcal{L}$ TBox $\mathcal{T}$ is *CQ-Horn-rewritable* if there is a TBox $\mathcal{T}'$ formulated in Horn-$\mathcal{L}$ that is CQ-inseparable from $\mathcal{T}$. The main property of an expressive DL $\mathcal{L}$ that we are interested in is then whether *CQ-Horn-rewritability captures* PTIME *query evaluation*, that is, whether every $\mathcal{L}$ TBox that enjoys PTIME CQ-evaluation is CQ-Horn-rewritable. Note that when $\mathcal{L}$ satisfies this property, then for any $\mathcal{L}$ TBox $\mathcal{T}$ that enjoys PTIME CQ-evaluation one can take advantage of the algorithms available for CQ-evaluation w.r.t. Horn-$\mathcal{L}$ TBoxes, via the CQ-inseparable Horn TBox.

Seemingly natural variations of the above are obtained by requiring that $\mathcal{T}'$ is logically equivalent to $\mathcal{T}$ rather than CQ-inseparable or that it is a model-theoretic conservative extension of $\mathcal{T}$. Then, however, rewritability into a Horn TBox fails already for very simple TBoxes that admit CQ-evaluation in PTIME. For example, it can be shown that the TBox $\mathcal{T}_1$ which states that every author is the author of a novel or a short story or of non fiction, in symbols

$$\exists \mathsf{author}.\top \sqsubseteq \exists \mathsf{author}.\mathsf{Novel} \sqcup \exists \mathsf{author}.\mathsf{Short\_Story} \sqcup \exists \mathsf{author}.\neg \mathsf{Fiction},$$

has no conservative extension that is a Horn TBox, but nevertheless enjoys CQ-evaluation in PTIME. In fact, $\mathcal{T}_1$ is CQ-inseparable from the empty TBox, which is a Horn TBox.

It turns out that whether CQ-Horn-rewritability captures PTIME query evaluation depends on various factors, in particular on whether or not the unique name assumption (UNA) is made, on the DL under consideration, and on the nesting depth of quantifiers in TBoxes. Regarding the UNA, recall that answers to ontology-mediated queries depend on whether the UNA is made whenever a DL is used that admits a form of counting such as number restrictions and functional roles. To illustrate this, consider the following TBox $\mathcal{T}_2$ stating that everybody who authored at least 200 publications is a prolific author:

$$\mathcal{T}_2 = \{(\geqslant 200\, \mathsf{author}\, \top) \sqsubseteq \mathsf{ProlificAuthor}\}$$

Consider the ABox

$$\mathcal{A}_2 = \{\mathsf{author}(\mathsf{bob}, \mathsf{book}_i) \mid 1 \leq i \leq 200\}.$$

Then, with the UNA, it clearly follows that Bob is a prolific author. Without the UNA, however, some of the individual names book_i might denote the same individual, and so it does not follow that Bob is a prolific author.

Regarding the impact of the UNA on CQ-Horn-rewritability and PTIME CQ-evaluation, we first make the following fundamental observations for the expressive DL $\mathcal{ALCHIQ}$, which is the main DL considered in this paper:

1. PTIME CQ-evaluation without UNA implies PTIME CQ-evaluation with UNA; the converse does not hold with $\mathcal{T}_2$ being a counterexample: one can show that CQ-evaluation w.r.t. $\mathcal{T}_2$ is in PTIME with UNA, but CONP-hard without UNA.
2. CQ-Horn-rewritability (and, in fact, whether a given TBox is a CQ-Horn-rewriting) does not depend on the UNA; we can thus speak about CQ-Horn-rewritability independently from the UNA.

As stated in Point 1, $\mathcal{T}_2$ admits PTIME CQ-evaluation with the UNA while it is CONP-hard without. Unless PTIME = NP, $\mathcal{T}_2$ is thus not CQ-Horn-rewritable without the UNA. Consequently, with the UNA CQ-Horn-rewritability does not capture PTIME CQ-evaluation for $\mathcal{ALCQ}$-TBoxes without quantifier nesting (depth 1 TBoxes, for short). Interestingly, concept inclusions (CIs) of the form used in $\mathcal{T}_2$ are very common in real-world ontologies: we analyzed the Bioportal and ORE repositories [Whetzel *et al.*, 2011; Parsia *et al.*, 2017] and found a total of 5081 (respectively, 6958) CIs of depth 1 that contain number restrictions of which 2066 (respectively, 1720) are provably not CQ-Horn-rewritable but enjoy PTIME CQ-evaluation with the UNA. Such CIs occur in 41 (from a total of 97) and 186 (from a total of 414) ontologies with number restrictions in the Bioportal and ORE repositories.

The situation is very different without the UNA: in this case, we prove that CQ-Horn-rewritability captures PTIME query evaluation for all $\mathcal{ALCHIQ}$ TBoxes of depth 1. We show this by constructing from a TBox $\mathcal{T}$ of depth 1 a canonical Horn-TBox $\mathcal{T}_{\text{horn}}$ such that $\mathcal{T}_{\text{horn}}$ is a CQ-inseparable rewriting of $\mathcal{T}$ if and only if CQ-evaluation w.r.t. $\mathcal{T}$ without UNA is in PTIME. We also show that deciding whether an $\mathcal{ALCHIQ}$ TBox of depth 1 is CQ-Horn-rewritable is EXPTIME-complete. Observe that in practice the restriction to TBoxes of depth 1 is a rather minor one (more than 95% of all ontologies on the Bioportal and ORE repositories have depth 1, sometimes modulo a straightforward reformulation). In theory, however, the restriction is crucial: we show that for $\mathcal{ALC}$ TBoxes of depth 3, CQ-Horn-rewritability does not capture PTIME query evaluation and that for $\mathcal{ALCF}$ TBoxes of depth 3 CQ-Horn-rewritability is undecidable.

Finally, we return to CQ-evaluation with the UNA and show that TBoxes in the fragment $\mathcal{ALCHIF}^{\not\sqsubseteq f}$ of $\mathcal{ALCHIF}$ in which no functional role includes another role enjoy PTIME CQ-evaluation with the UNA iff they enjoy PTIME CQ-evaluation without the UNA and that without this condition the equivalence fails already for TBoxes of depth 1. We thus determine a 'maximal' fragment of $\mathcal{ALCHIF}$ in which CQ-Horn-rewritability captures PTIME query evaluation with the UNA for TBoxes of depth 1.

Related Work. Rewritability into tractable languages has been studied extensively in description logic. A large body of work investigates rewritability of ontology-mediated queries (OMQs) into FO or Datalog queries giving the same answers on all ABoxes [Bienvenu *et al.*, 2014; 2016; Feier *et al.*, 2017]. The main difference to the work presented in this paper is that both the TBox and the CQ are given as input whereas in this paper we quantify over all CQs. In [Kaminski *et al.*, 2016; Kaminski and Grau, 2015; Carral *et al.*, 2014], the authors consider Horn-DL and $\mathcal{EL}$ rewritability of OMQs with atomic queries. Rewritability of TBoxes in an expressive DL into logically equivalent TBoxes or conservative extensions in a weaker DLs has been investigated in [Lutz *et al.*, 2011; Konev *et al.*, 2016].

2 Preliminaries

We use standard notation for DLs [Baader *et al.*, 2017]. Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be countably infinite sets of concept, role, and individual names. A *role* is a role name or the *inverse* r^- of a role name r. $\mathcal{ALCIQ}$*-concepts* are formed according to the rule

$$C, D := \top \mid A \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid (\geqslant n\, r\, C) \mid (\leqslant n\, r\, C)$$

where $A \in \mathsf{N_C}$, r is a role, and n is a non-negative integer. Concepts of the form $(\geqslant n\, r\, C)$ and $(\leqslant n\, r\, C)$ are called *qualified number restrictions*. An $\mathcal{ALCIQ}$ *concept inclusion (CI)* takes the form $C \sqsubseteq D$, where C and D are $\mathcal{ALCIQ}$-concepts. An $\mathcal{ALCIQ}$ *TBox* is a finite set of $\mathcal{ALCIQ}$ CIs. A *role inclusion (RI)* takes the form $r \sqsubseteq s$, where r and s are roles. An $\mathcal{ALCHIQ}$ *TBox* $\mathcal{T}$ is a finite set of $\mathcal{ALCIQ}$ CIs and RIs. We also consider various DLs contained in $\mathcal{ALCHIQ}$. $\mathcal{ALCHI}$ is obtained from $\mathcal{ALCHIQ}$ by restricting the qualified number restrictions to concepts of the form $(\geqslant 1\, r\, C)$ (also written $\exists r.C$) and $(\leqslant 0\, r\, C)$ (also written $\forall r.\neg C$), and $\mathcal{ALCHIF}$ is the extension of $\mathcal{ALCHI}$ with *functionality assertion* taking the form $\top \sqsubseteq (\leqslant 1\, r\, \top)$. We also use $\mathcal{ELI}$ *concepts* which are constructed using $\top$, concept names, $\sqcap$, and $\exists r.C$ with r a role. For any concept, CI, or TBox $\mathcal{T}$, we use $|\mathcal{T}|$ to denote the number of symbols needed to write $\mathcal{T}$ assuming that numbers in number restrictions are coded in unary.

An *ABox* $\mathcal{A}$ is a non-empty finite set of assertions of the form $A(a)$ and $r(a,b)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. We denote by $\mathsf{ind}(\mathcal{A})$ the set of individual names occurring in $\mathcal{A}$.

Interpretations $\mathcal{I}$ take the form $(\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is the non-empty *domain* of $\mathcal{I}$ and $\cdot^\mathcal{I}$ interprets every concept name A as a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$, role name r as a binary relation $r^\mathcal{I}$ in $\Delta^\mathcal{I}$, and individual name a as an element $a^\mathcal{I}$ of $\Delta^\mathcal{I}$. The extension $C^\mathcal{I}$ of a concept C in $\mathcal{I}$ is defined as usual. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, an RI $r \sqsubseteq s$ if $r^\mathcal{I} \subseteq s^\mathcal{I}$, an assertion $A(a)$ if $a^\mathcal{I} \in A^\mathcal{I}$, and an assertion $r(a,b)$ if $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$. We say that $\mathcal{I}$ satisfies the *unique name assumption (UNA)* if $a^\mathcal{I} \neq b^\mathcal{I}$ for all individual names $a \neq b$. An interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ if it satisfies all CIs and RIs in $\mathcal{T}$ and $\mathcal{I}$ is a *model* of an ABox $\mathcal{A}$ if it satisfies all assertions in $\mathcal{A}$. We call an ABox $\mathcal{A}$ *satisfiable w.r.t. a TBox* $\mathcal{T}$ *(with UNA)* if $\mathcal{A}$ and $\mathcal{T}$ have a common model (satisfying the UNA).

The *depth* of an $\mathcal{ALCIQ}$ concept is the maximal number of nestings of the qualified number restrictions in it; thus $(\geq 5r.A)$ has depth 1 and $(\geqslant 5\, r\, (\geqslant 4\, r\, A))$ has depth 2. The *depth* of a TBox, which will play an important role in this paper, is the maximal depth of the concepts that occur in it. For deciding satisfiability and subsumption, TBoxes are often normalized to depth 1 in a pre-processing step [Kazakov, 2009; Kaminski *et al.*, 2016]. This does not work for the questions studied in this paper since normalization can change the complexity of the TBox, see [Lutz and Wolter, 2017; Hernich *et al.*, 2017].

A *Horn-$\mathcal{ALCIQ}$ CI* takes the form $L \sqsubseteq R$, where L and R are built according to the following syntax rules

$$L, L' ::= \top \mid \bot \mid A \mid L \sqcap L' \mid L \sqcup L' \mid \exists r.L$$
$$R, R' ::= \top \mid \bot \mid A \mid \neg A \mid R \sqcap R' \mid \neg L \sqcup R \mid (\geqslant n\, r\, R) \mid \forall r.R \mid (\leqslant 1\, r\, L)$$

A Horn-$\mathcal{ALCHIQ}$ TBox is a finite set of Horn-$\mathcal{ALCIQ}$ CIs and RIs. Note that there are several alternative ways to define Horn-DLs [Hustadt *et al.*, 2007; Krötzsch *et al.*, 2007; Eiter *et al.*, 2008; Kazakov, 2009]. The results in this paper apply to all these definitions: whenever we claim that a sentence cannot be expressed using a Horn-TBox, the proof establishes failure of preservation under direct products which shows that the sentence cannot be expressed in FO-Horn [Chang and Keisler, 1990; Lutz *et al.*, 2011], and if we rewrite into a Horn-TBox we always rewrite into a TBox of depth 1 in which case all definitions of Horn-TBoxes coincide.

A conjunctive query (CQ) $q(\vec{x})$ is an FO-formula of the form $\exists \vec{y}\, \varphi(\vec{x}, \vec{y})$, where $\varphi(\vec{x}, \vec{y})$ is a conjunction of atoms of the form $A(x)$, $r(x, y)$, and $x = y$. Every $\mathcal{ELI}$ concept C defines in the natural way a tree-shaped CQ with one free variable, written $C(x)$ [Lutz and Wolter, 2017]. Let ELIQ denote the class of all such CQs, and let ELIQ$^=$ denote the union of ELIQ and the set of all equalities $x = y$. We say that a tuple $\vec{a}$ of individuals in an ABox $\mathcal{A}$ is a *certain answer to the CQ $q(\vec{x})$ over $\mathcal{A}$ w.r.t. a TBox $\mathcal{T}$*, in symbols $\mathcal{T}, \mathcal{A} \models q(\vec{a})$ if $\mathcal{I} \models q(\vec{a})$ holds for all models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$. The *query evaluation problem for $\mathcal{T}$ and CQ q* is the problem to decide for a given ABox $\mathcal{A}$ and tuple $\vec{a}$ of individuals from $\mathcal{A}$, whether $\mathcal{T}, \mathcal{A} \models q(\vec{a})$. We say that *the CQ-evaluation problem for $\mathcal{T}$ is in* PTIME if the query evaluation problem for $\mathcal{T}$ and q is in PTIME for every CQ q. Note that our default assumption when speaking about query evaluation is that we do not make the UNA. If we do, then we shall always explicitly say so. We write $\mathcal{T}, \mathcal{A} \models_{\text{UNA}} q(\vec{a})$ if $\mathcal{I} \models q(\vec{a})$ holds for all models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ satisfying the UNA and the *query evaluation problem for $\mathcal{T}$ and CQ q with the UNA* is the problem to decide $\mathcal{T}, \mathcal{A} \models_{\text{UNA}} q(\vec{a})$. If we want to emphasize that we do not make the UNA, we write $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q(\vec{a})$ instead of $\mathcal{T}, \mathcal{A} \models q(\vec{a})$. The relationship between certain answers with and without the UNA can be expressed using the following equivalence:

$$\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q(\vec{a}) \vee \bigvee_{a \neq b \in \mathsf{ind}(\mathcal{A})} (a = b) \ \Leftrightarrow\ \mathcal{T}, \mathcal{A} \models_{\text{UNA}} q(\vec{a}). \quad (1)$$

It is well known that for DLs that do not admit any forms of counting the UNA does not affect the certain answers to CQs. Thus, if $\mathcal{T}$ is an $\mathcal{ALCHI}$ TBox, then $\mathcal{T}, \mathcal{A} \models_{\text{UNA}} q(\vec{a})$ iff $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q(\vec{a})$.

In this paper, we aim to understand whether and when a TBox formulated in an expressive DL can be replaced with a TBox formulated in the corresponding Horn-DL without changing the answers to CQs. Following [Lutz and Wolter, 2010; Botoeva *et al.*, 2016a; 2016b], TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ are *CQ-inseparable* if for all CQs q, all ABoxes $\mathcal{A}$, and all tuples $\vec{a}$ of individual names in $\mathcal{A}$, the following equivalence holds:

$$\mathcal{T}_1, \mathcal{A} \models_{\text{nUNA}} q(\vec{a}) \quad \Leftrightarrow \quad \mathcal{T}_2, \mathcal{A} \models_{\text{nUNA}} q(\vec{a}).$$

If $\mathcal{T}_2$ is a Horn TBox, then we call $\mathcal{T}_2$ a *CQ-Horn-rewriting of* $\mathcal{T}_1$. A TBox $\mathcal{T}$ in a DL $\mathcal{L}$ is *CQ-Horn-rewritable* if there exists a CQ-Horn-rewriting of $\mathcal{T}$ in Horn-$\mathcal{L}$. We further say that *CQ-Horn-rewritability captures* PTIME *query evaluation for $\mathcal{L}$* if every TBox in $\mathcal{L}$ is CQ-Horn-rewritable. Thus, as before, by default we do not make the UNA. The notions introduced above are modified in the obvious way if one makes the UNA and we will always make this explicit.

3 Transfer between UNA and non-UNA

We investigate the influence of the UNA on CQ-Horn-rewritability and the complexity of CQ-evaluation. We show that for $\mathcal{ALCHIQ}$ TBoxes CQ-Horn-rewritability does not depend on the UNA, but that for PTIME CQ-evaluation only one direction holds: if CQ-evaluation is in PTIME without UNA, then it is in PTIME with UNA. In the proof we use a disjunction property of TBoxes and show that it is a necessary condition for CQ-evaluation to be in PTIME, with and without UNA (unless PTIME equals CONP).

For an ABox $\mathcal{A}$, CQs $q_1(\vec{x}_1), \ldots, q_n(\vec{x}_n)$, and tuples $\vec{a}_1, \ldots, \vec{a}_n$ in $\mathcal{A}$, we write $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \leq i \leq n} q_i(\vec{a}_i)$ if for every model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ there is $i \in \{1, \ldots, n\}$ with $\mathcal{I} \models q_i(\vec{a}_i)$, and we define $\mathcal{T}, \mathcal{A} \models_{\text{UNA}} \bigvee_{1 \leq i \leq n} q_i(\vec{a}_i)$ accordingly based on models that satisfy the UNA. Let $\mathcal{Q}$ be a class of CQs. A TBox $\mathcal{T}$ has the *$\mathcal{Q}$-disjunction property without UNA* if for all ABoxes $\mathcal{A}$, CQs $q_1(\vec{x}_1), \ldots, q_n(\vec{x}_n) \in \mathcal{Q}$ and tuples $\vec{a}_1, \ldots, \vec{a}_n$ in $\mathcal{A}$ with $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \leq i \leq n} q_i(\vec{a}_i)$ there exists $i \in \{1, \ldots, n\}$ such that $\mathcal{T}, \mathcal{A} \models q_i(\vec{a}_i)$. The *$\mathcal{Q}$-disjunction property with UNA* is defined accordingly.

Example 1. The TBox $\mathcal{T}_2$ from the introduction does not enjoy the ELIQ$^=$-disjunction property without UNA, but enjoys it with UNA. To show the first claim note that for the ABox $\mathcal{A}_2$ from the introduction

$$\mathcal{T}_2, \mathcal{A}_2 \models_{\text{nUNA}} \bigvee_{i \neq j} (\mathsf{book}_i = \mathsf{book}_j) \vee \mathsf{ProlificAuthor}(\mathsf{bob})$$

but no disjunct is entailed without UNA. To show the second claim observe that $\mathcal{T}_2, \mathcal{A} \models_{\text{UNA}} q(\vec{a})$ iff $\emptyset, \mathcal{A}' \models q(\vec{a})$ holds for every ABox $\mathcal{A}$, any CQ q, and for $\mathcal{A}'$ obtained from $\mathcal{A}$ by adding the assertions $\mathsf{ProlificAuthor}(b)$ for any b such that $\mathsf{author}(b, c) \in \mathcal{A}$ for at least 200 distinct c. It follows immediately that $\mathcal{T}_2$ has the ELIQ$^=$-disjunction property with UNA and enjoys PTIME CQ-evaluation with UNA.

We need the following technical lemma.

Lemma 1. *If $\mathcal{T}$ is an $\mathcal{ALCHIQ}$ TBox, then $\mathcal{T}$ has the ELIQ-disjunction property iff $\mathcal{T}$ has the ELIQ$^=$-disjunction property* *iff $\mathcal{T}$ has the CQ-disjunction property. The equivalences hold both with and without UNA.*

Proof (sketch). We prove the case without UNA of which the case with UNA is a special case. The direction from CQ to ELIQ is trivial. For the direction from ELIQ to ELIQ$^=$, we simulate equalities between distinct individual names in an ABox $\mathcal{A}$ by ELIQs as follows. Given an equality $(a = b)$ with $a \neq b \in \mathsf{ind}(\mathcal{A})$, we first extend $\mathcal{A}$ by a new assertion $A_a(a)$, where A_a is a fresh concept name, and then replace $(a = b)$ by $A_a(b)$. Note that the corresponding query $A_a(x)$ is an ELIQ. The direction from ELIQ$^=$ to CQ is similar to the proof of Theorem 16 in [Lutz and Wolter, 2017]. ☐

For an ABox $\mathcal{A}$ and an equivalence relation $\sim$ on $\mathsf{ind}(\mathcal{A})$, the *quotient ABox* $\mathcal{A}/_\sim$ of $\mathcal{A}$ is defined by replacing each individual a in $\mathcal{A}$ with the equivalence class $a/_\sim$ of a w.r.t. $\sim$. Given a tuple $\vec{a} = (a_1, \ldots, a_k)$ in $\mathcal{A}$ we denote by $\vec{a}/_\sim$ the tuple $(a_1/_\sim, \ldots, a_k/_\sim)$.

Theorem 1. *A Horn-$\mathcal{ALCHIQ}$ TBox $\mathcal{T}'$ is a CQ-Horn-rewriting of an $\mathcal{ALCHIQ}$ TBox $\mathcal{T}$ without UNA iff it is a CQ-Horn-rewriting of $\mathcal{T}$ with UNA.*

Proof (sketch). For the direction from left to right, let $\mathcal{A}$ be an ABox. We first establish that for all CQs $q_1(\vec{x}_1), \ldots, q_n(\vec{x}_n)$ and tuples $\vec{a}_1, \ldots, \vec{a}_n$ in $\mathcal{A}$:

$$\mathcal{T}, \mathcal{A} \models_{\mathrm{nUNA}} \bigvee_{1 \leq i \leq n} q_i(\vec{a}_i) \;\Leftrightarrow\; \mathcal{T}', \mathcal{A} \models_{\mathrm{nUNA}} \bigvee_{1 \leq i \leq n} q_i(\vec{a}_i). \quad (2)$$

For the proof, we may assume that the $q_i(\vec{x}_i)$ are ELIQs (by Lemma 1). We then simulate disjunctions of ELIQs by single ELIQs (see Theorem 18 in [Lutz and Wolter, 2017] for a similar construction) and use that $\mathcal{T}'$ is a CQ-Horn-rewriting of $\mathcal{T}$ without UNA.

Now let $q(\vec{x})$ be a CQ and $\vec{a}$ a tuple in $\mathcal{A}$. Then, (1) and (2) imply $\mathcal{T}, \mathcal{A} \models_{\mathrm{UNA}} q(\vec{a})$ iff $\mathcal{T}', \mathcal{A} \models_{\mathrm{UNA}} q(\vec{a})$.

For the converse, we first establish that $\mathcal{T}, \mathcal{A} \models_{\mathrm{nUNA}} q(\vec{a})$ iff $\mathcal{T}, \mathcal{A}/_\sim \models_{\mathrm{UNA}} q(\vec{a}/_\sim)$ for all equivalence relations $\sim$ on $\mathsf{ind}(\mathcal{A})$, where $\mathcal{A}$, $q(\vec{x})$, and $\vec{a}$ are as above. The same holds if we substitute $\mathcal{T}'$ for $\mathcal{T}$. Since the right hand side of this equivalence holds for $\mathcal{T}$ iff it holds for $\mathcal{T}'$, it follows that $\mathcal{T}, \mathcal{A} \models_{\mathrm{nUNA}} q(\vec{a})$ iff $\mathcal{T}', \mathcal{A} \models_{\mathrm{nUNA}} q(\vec{a})$. ☐

We now turn to CQ-evaluation w.r.t. $\mathcal{ALCHIQ}$ TBoxes. As shown in [Hernich *et al.*, 2017], the ELIQ$^=$-disjunction property implies that CQ-evaluation with UNA is in PTIME. The proof can be generalized to the case without UNA.

Lemma 2. *Let $\mathcal{T}$ be a $\mathcal{ALCHIQ}$ TBox. If $\mathcal{T}$ does not have the ELIQ$^=$-disjunction property, then ELIQ$^=$-evaluation for $\mathcal{T}$ is* CONP*-hard. This holds both with and without UNA.*

Example 2. As the TBox $\mathcal{T}_2$ from Example 1 does not enjoy the ELIQ$^=$-disjunction property without UNA, CQ-evaluation w.r.t. $\mathcal{T}_2$ is CONP-hard without UNA.

As a consequence of Lemma 2 we obtain that tractability of CQ-evaluation without UNA implies tractability of CQ-evaluation with UNA.

Theorem 2. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox and suppose that CQ-evaluation w.r.t. $\mathcal{T}$ without UNA is in* PTIME*. Then, CQ-evaluation w.r.t. $\mathcal{T}$ with UNA is in* PTIME.

Proof. We reduce the UNA case to the non-UNA case. Let $\mathcal{A}$ be an ABox, $q(\vec{x})$ a CQ, and $\vec{a}$ a tuple in $\mathcal{A}$. By Lemma 2 and Lemma 1, $\mathcal{T}$ has the CQ-disjunction property (unless we are in the trivial case where PTIME = CONP). Now, (1) implies that $\mathcal{T}, \mathcal{A} \models_{\mathrm{UNA}} q(\vec{a})$ iff either $\mathcal{T}, \mathcal{A} \models_{\mathrm{nUNA}} q(\vec{a})$ or there exist $a \neq b \in \mathsf{ind}(\mathcal{A})$ with $\mathcal{T}, \mathcal{A} \models_{\mathrm{nUNA}} (a = b)$. ☐

4 CQ-Horn-Rewritability vs PTIME w/o UNA

We show that, without the UNA, CQ-Horn-rewritability captures PTIME query evaluation for $\mathcal{ALCHIQ}$ TBoxes of depth 1. We also show that the meta problem of deciding CQ-Horn-rewritability is EXPTIME complete for such TBoxes. To prove these results, we first show how to equivalently transform a TBox of depth 1 into a certain normal form. From the resulting TBox $\mathcal{T}$ we construct a Horn TBox $\mathcal{T}_{\mathrm{horn}}$ which we show to be a CQ-Horn-rewriting of $\mathcal{T}$ if and only if CQ-evaluation for $\mathcal{T}$ is in PTIME.

We start with the normal form. A *literal* is a concept name or a negation thereof. A CI $C \sqsubseteq D$ is in *normal form* if

1. C is a conjunction of concept names and concepts of the form $(\geqslant n\, r\, E)$ with E a conjunction of concept names;
2. D is a disjunction of
   - concept names;
   - concepts $(\geqslant n\, r\, E)$ with E a conjunction of literals;
   - concepts $(\leqslant n\, r\, E)$ with E a conjunction of literals that contains at least one negative literal.

We set $C = \top$ if C is the empty conjunction and $D = \bot := \neg\top$ if D is the empty disjunction. An $\mathcal{ALCHIQ}$ TBox $\mathcal{T}$ is in *normal form* if all CIs in $\mathcal{T}$ are in normal form.

Lemma 3. *Every $\mathcal{ALCHIQ}$ TBox $\mathcal{T}$ of depth 1 can be converted into a logically equivalent $\mathcal{ALCHIQ}$ TBox $\mathcal{T}'$ in normal form.*

In the worst case, $\mathcal{T}'$ is of size double exponential in the size $\mathcal{T}$. From now on, we assume that $\mathcal{T}$ is fixed and in normal form. Using $\mathcal{T}$, we define a Horn TBox $\mathcal{T}_{\mathrm{horn}}$. For any conjunction or disjunction of literals E, we use $\mathsf{pos}(E)$ to denote the conjunction of all concept names A in E and $\mathsf{neg}(E)$ to denote the conjunction of all concept names A such that $\neg A$ is in E. We use $L_\mathcal{T}$ to denote the set of

- concept names or concepts of the form $(\geqslant n\, r\, E)$ occurring as top-level conjuncts in C in some CI $C \sqsubseteq D \in \mathcal{T}$;
- concepts $(\geqslant n + 1\, r\, \mathsf{pos}(E))$ such that there is a CI $C \sqsubseteq D \in \mathcal{T}$ such that $(\leqslant n\, r\, E)$ is a disjunct of D.

A set $S \subseteq L_\mathcal{T}$ is a *trigger for* a CI $C \sqsubseteq D \in \mathcal{T}$ if S contains all top-level conjuncts of C and all $(\geqslant n{+}1\, r\, \mathsf{pos}(E))$ with $(\leqslant n\, r\, E)$ a disjunct of D. For a trigger S, we denote by C_S the conjunction of all concepts in S and by $C_S^{\leq 1}$ the $\mathcal{ELI}$ concept obtained from C_S by replacing every $(\geqslant n\, r\, E)$ with $n \geq 2$ by $(\geqslant 1\, r\, E)$. For a concept $(\leqslant n\, r\, E)$ with E a conjunction of literals that contains at least one negative literal, we call $\forall r.E'$ a *Horn specialization* of $(\leqslant n\, r\, E)$ if E' is obtained from E by dropping all but one negative literal. We sometimes write Horn specializations in the form $\forall r.(A_1 \sqcap \cdots \sqcap A_n \rightarrow A)$ where $C \rightarrow D$ stands for $\neg C \sqcup D$.

For each CI $C \sqsubseteq D \in \mathcal{T}$ and trigger S for it we define a set $\mathsf{Horn}(C \sqsubseteq D, S)$ of Horn-$\mathcal{ALCIQ}$-CIs. In the special case that $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq \bot$ we set $\mathsf{Horn}(C \sqsubseteq D, S) = \{C_S^{\leq 1} \sqsubseteq \bot\}$. Otherwise $\mathsf{Horn}(C \sqsubseteq D, S)$ contains the following CIs whenever they are a consequence of $\mathcal{T}$:

- $C_S^{\leq 1} \sqsubseteq (\leqslant 1\, r\, E)$ if $(\geqslant n\, r\, E) \in S$ for some $n \geq 2$;
- $C_S^{\leq 1} \sqsubseteq A$ if $A \in \mathsf{N_C}$ is a top-level disjunct of D;
- $C_S^{\leq 1} \sqsubseteq R$ if $R = \forall r.(A_1 \sqcap \cdots \sqcap A_n \rightarrow A)$ is a Horn specialization of some disjunct $(\leqslant n\, r\, E)$ of D;
- $C_S^{\leq 1} \sqsubseteq (\geqslant 1\, r\, \mathsf{pos}(E))$ if $(\geqslant m\, r\, E)$ is a disjunct of D such that $\mathcal{T} \not\models C_S^{\leq 1} \sqsubseteq \neg(\geqslant m\, r\, E)$.

Now the Horn-$\mathcal{ALCHIQ}$ TBox $\mathcal{T}_{\text{horn}}$ is defined as the union of all RIs in $\mathcal{T}$ and

$$\bigcup_{C \sqsubseteq D \in \mathcal{T},\ S \text{ trigger for } C \sqsubseteq D} \mathsf{Horn}(C \sqsubseteq D, S)$$

It can be verified that, by construction, $\mathcal{T} \models \mathcal{T}_{\text{horn}}$. The following lemma is the main step towards the capturing result.

Lemma 4. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox in normal form. Then the following conditions are equivalent:*

1. *$\mathcal{T}$ has the ELIQ$^=$-disjunction property without UNA;*
2. *for every $C \sqsubseteq D \in \mathcal{T}$ and trigger S for $C \sqsubseteq D$, $\mathsf{Horn}(C \sqsubseteq D, S) \neq \emptyset$;*
3. *$\mathcal{T}$ and $\mathcal{T}_{horn}$ are CQ-inseparable without UNA.*

The following examples illustrate this lemma.

Example 3. (1) Reconsider the TBox $\mathcal{T}_1$ from the introduction, which contains the only CI

$$\exists \mathsf{author}.\top \sqsubseteq \exists \mathsf{author.Novel} \sqcup \exists \mathsf{author.Short\_Story} \sqcup \exists \mathsf{author}.\neg\mathsf{Fiction}$$

that we abbreviate by α. Then $S = \{\exists \mathsf{author}.\top\}$ is the only trigger for α. We have $\mathcal{T}_{1_{\text{horn}}} = \mathsf{Horn}(\alpha, S) = \{\exists \mathsf{author}.\top \sqsubseteq \exists \mathsf{author}.\top\}$ since $\mathsf{pos}(\neg \mathsf{Fiction}) = \top$. Thus, $\mathsf{Horn}(\alpha, S) \neq \emptyset$ and, by Lemma 4, $\mathcal{T}_{1_{\text{horn}}}$ is a CQ-Horn-rewriting of $\mathcal{T}_1$ (equivalent to the empty TBox) .

Define $\mathcal{T}'$ by adding to $\mathcal{T}_1$ the CI $\mathsf{Novelist} \sqsubseteq \forall \mathsf{author.Fiction}$. Then $S = \{\exists \mathsf{author}.\top, \mathsf{Novelist}\}$ is a trigger for α and now $\mathsf{Horn}(\alpha, S) = \emptyset$. Thus, by Point 2, $\mathcal{T}'_{\text{horn}}$ is not a CQ-Horn-rewriting of $\mathcal{T}'$.

(2) Consider the TBox $\mathcal{T}_2$ from the introduction containing

$$\beta = (\geqslant 200\, \mathsf{author}\, \top) \sqsubseteq \mathsf{ProlificAuthor}$$

Then $S = \{(\geqslant 200\, \mathsf{author}\, \top)\}$ is the only trigger for β. We have $C_S^{\leq 1} = \exists \mathsf{author}.\top$ and it is readily checked that $\mathcal{T}_{2_{\text{horn}}} = \mathsf{Horn}(\beta, S) = \emptyset$. By Point 2, $\mathcal{T}_{2_{\text{horn}}}$ is not a CQ-Horn-rewriting of $\mathcal{T}_2$.

(3) Observe that for any TBox $\mathcal{T}$, $0, 1$ are the only numbers used in $\mathcal{T}_{\text{horn}}$. Consider, for example, $\mathcal{T} = \{\mathsf{ProlificScientist} \sqsubseteq (\geqslant 200\, \mathsf{author}\, \neg \mathsf{Fiction})\}$. Then $\mathcal{T}_{\text{horn}} = \{\mathsf{ProlificScientist} \sqsubseteq (\geqslant 1\, \mathsf{author}\, \top)\}$ is a CQ-Horn-rewriting of $\mathcal{T}$.

We give a brief description of the proof of Lemma 4. For the proof of (1) $\Rightarrow$ (2) one constructs under the assumption that (2) does not hold for $C \sqsubseteq D$ and trigger S the tree-shaped ABox $\mathcal{A}_S$ corresponding to the concept C_S and a disjunction of queries in ELIQ$^=$ which refutes the ELIQ$^=$-disjunction property if $\mathsf{Horn}(C \sqsubseteq D, S) = \emptyset$. For (2) $\Rightarrow$ (3) one defines a chase procedure which constructs, if (2) holds, for every ABox $\mathcal{A}$ satisfiable w.r.t. $\mathcal{T}_{\text{horn}}$ a universal model of $\mathcal{A}$ and $\mathcal{T}_{\text{horn}}$ which is also a model of $\mathcal{T}$. For (3) $\Rightarrow$ (1) assume that (3) holds and let $\mathcal{A}$ be an ABox, $q_1(\vec{x}_1), \ldots, q_n(\vec{x}_n)$ CQs, and $\vec{a}_1, \ldots, \vec{a}_n$ tuples in $\mathcal{A}$ with $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \leq i \leq n} q_i(\vec{a}_i)$. By (2) from the proof of Theorem 1, $\mathcal{T}_{\text{horn}}, \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \leq i \leq n} q_i(\vec{a}_i)$. But then there exists i such that $\mathcal{T}_{\text{horn}}, \mathcal{A} \models_{\text{nUNA}} q_i(\vec{a}_i)$ and by Point 3, $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q_i(\vec{a}_i)$, as required.

The following main result of this section now follows from Lemmas 4 and 2.

Theorem 3. *CQ-Horn-rewritability captures* PTIME *CQ-evaluation without UNA for $\mathcal{ALCHIQ}$ TBoxes of depth 1 (unless* PTIME *equals* CONP*).*

Observe that we also obtain a PTIME/CONP dichotomy for CQ-evaluation w.r.t. $\mathcal{ALCHIQ}$ TBoxes of depth 1, without the UNA: for any such TBox $\mathcal{T}$, CQ-evaluation is in PTIME for all CQs or there exists a CQ for which query evaluation is CONP-hard w.r.t. $\mathcal{T}$. Results of this form have so far only been obtained for query evaluation with UNA [Lutz and Wolter, 2017; Hernich *et al.*, 2017].

Point 2 of Lemma 4 provides an effective algorithm for checking CQ-Horn-rewritability. Note, however, that because of the double exponential blow-up in the normalization step for TBoxes and the potentially exponential number of triggers, its worst-case complexity is triple exponential. Using a model-theoretic approach, we improve this to a single-exponential upper bound, and thus deciding CQ-Horn-rewritability is not harder than satisfiability.

Theorem 4. *Deciding CQ-Horn-rewritability of $\mathcal{ALCHIQ}$ TBoxes of depth 1 is* EXPTIME*-complete.*

The lower bound is proved by a polynomial reduction of the satisfiability of $\mathcal{ALCHIQ}$ TBoxes. For the upper bound, one decides the ELIQ$^=$-disjunction property without UNA. Using a model-theoretic reformulation one can show that a TBox $\mathcal{T}$ has the ELIQ$^=$-disjunction property without UNA iff it has the $\mathcal{Q}$-disjunction property without UNA for ABoxes that have the shape of a tree of depth 1 and of outdegree bounded by $|\mathcal{T}|$, where $\mathcal{Q}$ is the class of ELIQ$^=$s of depth 1 and of outdegree bounded by $|\mathcal{T}|$, and where both the ABox and the queries use concept and role names from $\mathcal{T}$ only. The latter condition can be reduced to satisfiability in $\mathcal{ALCHIQ}$.

5 CQ-Horn-Rewritability vs PTIME with UNA

As shown in Examples 1 and 2, CQ-Horn-rewritability does not capture PTIME query evaluation with UNA for very simple $\mathcal{ALCQ}$-TBoxes of depth 1 (unless PTIME equals CONP). The experiments reported in the introduction further show that the CIs occurring in these TBoxes are very common in practice. The following example shows that when number restrictions are restricted to global functionality assertions, then there are still TBoxes of depth 1 for which CQ-evaluation is in PTIME with UNA but which are not CQ-Horn-rewritable.

Example 4. Let $\mathcal{T}$ be the $\mathcal{ALCHIF}$ TBox stating that role names s_1 and s_2 are functional and containing the RIs $r \sqsubseteq s_1$ and $r \sqsubseteq s_2$ and the CIs

$$\begin{array}{lcl} \exists s_1.(B_1 \sqcap B_2) & \sqsubseteq & \exists r.\top \\ \exists s_1.\top \sqcap \exists s_2.\top & \sqsubseteq & \forall s_1.B_1 \sqcap \forall s_2.B_2 \\ \exists s_1.\top \sqcap \exists s_2.\top & \sqsubseteq & B \sqcup \exists r.\top \end{array}$$

One can show that $\mathcal{T}$ has the CQ-disjunction property with UNA but not without UNA. Thus, CQ-evaluation w.r.t. $\mathcal{T}$ with UNA is in PTIME [Hernich *et al.*, 2017] and $\mathcal{T}$ is not CQ-Horn-rewritable. To refute the CQ-disjunction property without UNA, let $\mathcal{A} = \{s_1(a, b_1), s_2(a, b_2)\}$. Then $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} B(a) \vee \exists r.\top(a)$ but $\mathcal{T}, \mathcal{A} \not\models_{\text{nUNA}} B(a)$ since by identifying b_1 and b_2 and adding (a, b_i) to the extension of r and b_i to B_1 and B_2 one can define a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $a^{\mathcal{I}} \notin B^{\mathcal{I}}$; and $\mathcal{T}, \mathcal{A} \not\models_{\text{nUNA}} \exists r.\top(a)$ since by adding a to the extension of B, b_1 to B_1, and b_2 to B_2 one can define a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $a^{\mathcal{I}} \notin (\exists r.\top)^{\mathcal{I}}$. To show the CQ-disjunction property with UNA, one can construct for any ABox $\mathcal{A}$ satisfiable w.r.t. $\mathcal{T}$ with UNA a model $\mathcal{I}$ which maps homomorphically into any model of $\mathcal{A}$ and $\mathcal{T}$ with UNA.

We now show that the interaction between functionality assertions and RIs exploited in Example 4 is needed to construct TBoxes in $\mathcal{ALCHIF}$ which are not CQ-Horn-rewritable but for which CQ-evaluation is in PTIME with UNA. An $\mathcal{ALCHIF}^{\sqsubseteq f}$ TBox is an $\mathcal{ALCHIF}$ TBox $\mathcal{T}$ such that whenever $r \sqsubseteq s \in \mathcal{T}$, then neither s nor s^- are functional in $\mathcal{T}$.

Theorem 5. *Let $\mathcal{T}$ be an $\mathcal{ALCHIF}^{\sqsubseteq f}$ TBox. Then CQ-evaluation w.r.t. $\mathcal{T}$ without UNA is in* PTIME *iff CQ-evaluation w.r.t. $\mathcal{T}$ with UNA is in* PTIME.

Proof (sketch). The direction ($\Rightarrow$) is Theorem 2. Conversely, assume that CQ-evaluation with UNA is in PTIME. Let $\mathcal{A}$ be an ABox. Let $\sim$ be the smallest equivalence relation on $\mathsf{ind}(\mathcal{A})$ such that if $a \sim b$ and $r(a, a'), r(b, b') \in \mathcal{A}$ and $\top \sqsubseteq (\leqslant 1\, r\, \top) \in \mathcal{T}$, then $a' \sim b'$. One can show that $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q(\vec{a})$ iff $\mathcal{T}, \mathcal{A}/_\sim \models_{\text{UNA}} q(\vec{a}/_\sim)$, for every CQ q and tuple $\vec{a}$ in $\mathsf{ind}(\mathcal{A})$. It follows that CQ-evaluation without UNA is in PTIME since $\sim$ can be computed in polynomial time. □

The following is now a consequence of Theorems 5 and 3.

Theorem 6. *CQ-Horn-rewritability captures* PTIME *query evaluation with UNA for all $\mathcal{ALCHIF}^{\sqsubseteq f}$ TBoxes of depth 1 (unless* PTIME *equals* CONP*).*

So far, we have investigated the relationship between PTIME CQ-evaluation and CQ-Horn-rewritability mainly for TBoxes of depth 1. In fact, our results for depth 1 TBoxes do not generalize to arbitrary depth.

Theorem 7. *CQ-Horn-rewritability does not capture* PTIME *query evaluation for $\mathcal{ALC}$ TBoxes of depth 3 (with and without UNA).*

Proof. According to Theorem 6.8 in [Lutz and Wolter, 2017] there are $\mathcal{ALC}$ TBoxes $\mathcal{T}$ of depth 3 such that CQ-evaluation w.r.t. $\mathcal{T}$ is in PTIME but such that some CQs q are not Datalog-rewritable w.r.t. $\mathcal{T}$. Such a TBox cannot be CQ-Horn-rewritable since every CQ is Datalog-rewritable w.r.t. any Horn-$\mathcal{ALC}$ TBox [Lutz and Wolter, 2017]. □

The question whether CQ-Horn-rewritability captures PTIME query evaluation for $\mathcal{ALC}$ TBoxes of depth 2 is open. Decidability of CQ-Horn-rewritability for $\mathcal{ALC}$ TBoxes of arbitrary depth is also open. For $\mathcal{ALCF}$, however, one can easily extend Theorem 7.3 in [Lutz and Wolter, 2017] and show that CQ-Horn-rewritability of $\mathcal{ALCF}$ TBoxes of depth 3 is undecidable.

6 Discussion

We briefly discuss alternative approaches to rewritability into Horn TBoxes. From a logical viewpoint, it is natural to demand that the rewriting $\mathcal{T}'$ should not only give the same answers to CQs as $\mathcal{T}$, but be logically equivalent to $\mathcal{T}$, or at least a conservative extension. Here, $\mathcal{T}'$ is called a *conservative extension* of $\mathcal{T}$ if $\mathcal{T}' \models \alpha$ for every $\alpha \in \mathcal{T}$ and for every model of $\mathcal{T}$ there exists a model of $\mathcal{T}'$ which coincides with $\mathcal{T}$ regarding its domain and the interpretation of the concept and role names from $\mathcal{T}$. Unfortunately, this approach is extremely restrictive. We have seen that the TBox $\mathcal{T}_1$ from the introduction is trivial from the viewpoint of answering CQs (it is CQ-inseparable from the empty TBox), but nevertheless there is no conservative extension of $\mathcal{T}_1$ which is also a Horn TBox. One can show this by proving that no conservative extension of $\mathcal{T}_1$ is preserved under direct products.

In some applications of ontology-mediated querying the user knows in advance signatures (finite sets of concept and role names) Σ_1 and Σ_2 such that all relevant ABoxes and CQs use symbols from Σ_1 and, respectively, Σ_2 only. Then, rather than admitting arbitrary ABoxes and CQs in the definition of CQ-Horn-rewritings, it is natural to consider *CQ-Horn-rewritings w.r.t* (Σ_1, Σ_2) in the sense that $\mathcal{T}$ and $\mathcal{T}'$ give exactly the same answers to all CQs in Σ_1 on all ABoxes in Σ_2. The corresponding notion of (Σ_1, Σ_2)-inseparability has been considered in [Botoeva *et al.*, 2016b]. This relaxation leads to undecidability of CQ-Horn-rewritability as one can reduce the corresponding undecidable CQ-inseparability problem.

Theorem 8. *For $\mathcal{ALC}$ TBoxes of depth 1 there is no algorithm that decides CQ-Horn-rewritability w.r.t.* (Σ_1, Σ_2) *and outputs such a rewriting in case it exists.*

7 Conclusion

We have investigated whether CQ-Horn-rewritability captures PTIME query evaluation, with particular focus on the influence of the UNA and the depth of TBoxes. From a practical viewpoint it would be of interest to investigate query answering algorithms covering the CIs which are in PTIME but cannot be captured using Horn-CIs discussed in the introduction. It would also be of interest to investigate the succinctness of CQ-Horn-rewritings. The normal form of a given TBox is of double exponential size (in the worst case) and our CQ-inseparable rewritings are of exponential size in the size of the TBox in normal form. It is open whether this is optimal.

Acknowledgements

André Hernich, Fabio Papacchini, and Frank Wolter were supported by EPSRC UK grant EP/M012646/1 and Carsten Lutz was supported by the DFG-funded Collaborative Research Center EASE.

References

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

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

[Bienvenu *et al.*, 2014] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive datalog, CSP, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33:1–33:44, 2014.

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

[Botoeva *et al.*, 2016a] Elena Botoeva, Boris Konev, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Inseparability and conservative extensions of description logic ontologies: A survey. In *Reasoning Web*, pages 27–89, 2016.

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

[Calvanese *et al.*, 2013] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Data complexity of query answering in description logics. *Artificial Intelligence*, 195:335–360, 2013.

[Carral *et al.*, 2014] David Carral, Cristina Feier, Bernardo Cuenca Grau, Pascal Hitzler, and Ian Horrocks. *EL*-ifying ontologies. In *IJCAR*, pages 464–479, 2014.

[Chang and Keisler, 1990] C. C. Chang and H. Jerome Keisler. *Model Theory*. Elsevier, 1990.

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

[Feier *et al.*, 2017] Cristina Feier, Carsten Lutz, and Antti Kuusisto. Rewritability in monadic disjunctive datalog, mmsnp, and expressive description logics. In *ICDT*, 2017.

[Hernich *et al.*, 2017] Andre Hernich, Carsten Lutz, Fabio Papacchini, and Frank Wolter. Dichotomies in ontology-mediated querying with the guarded fragment. In *PODS*, 2017.

[Hustadt *et al.*, 2007] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Reasoning in description logics by a reduction to disjunctive datalog. *J. Autom. Reasoning*, 39(3):351–384, 2007.

[Kaminski and Grau, 2015] Mark Kaminski and Bernardo Cuenca Grau. Computing Horn rewritings of description logics ontologies. In *IJCAI*, pages 3091–3097, 2015.

[Kaminski *et al.*, 2016] Mark Kaminski, Yavor Nenov, and Bernardo Cuenca Grau. Datalog rewritability of disjunctive datalog programs and non-Horn ontologies. *Artif. Intell.*, 236:90–118, 2016.

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

[Konev *et al.*, 2016] Boris Konev, Carsten Lutz, Frank Wolter, and Michael Zakharyaschev. Conservative rewritability of description logic TBoxes. In *IJCAI*, pages 1153–1159, 2016.

[Kontchakov and Zakharyaschev, 2014] Roman Kontchakov and Michael Zakharyaschev. An introduction to description logics and query rewriting. In *Proc. of Reasoning Web*, pages 195–244, 2014.

[Krisnadhi and Lutz, 2007] Adila Krisnadhi and Carsten Lutz. Data complexity in the EL family of DLs. In *Proc. of DL*, 2007.

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

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

[Lutz and Wolter, 2017] Carsten Lutz and Frank Wolter. The Data Complexity of Description Logic Ontologies. *Logical Methods in Computer Science*, Volume 13, Issue 4, November 2017.

[Lutz *et al.*, 2011] Carsten Lutz, Robert Piro, and Frank Wolter. Description logic TBoxes: Model-theoretic characterizations and rewritability. In *IJCAI*, 2011.

[Ortiz *et al.*, 2010] Magdalena Ortiz, Sebastian Rudolph, and Mantas Simkus. Worst-case optimal reasoning for the horn-dl fragments of OWL 1 and 2. In *KR*, 2010.

[Ortiz *et al.*, 2011] Magdalena Ortiz, Sebastian Rudolph, and Mantas Simkus. Query answering in the Horn fragments of the description logics $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$. In *IJCAI*, pages 1039–1044, 2011.

[Parsia *et al.*, 2017] Bijan Parsia, Nicolas Matentzoglu, Rafael S. Gonçalves, Birte Glimm, and Andreas Steigmiller. The OWL reasoner evaluation (ORE) 2015 competition report. *J. Autom. Reasoning*, 59(4):455–482, 2017.

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

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

[Whetzel *et al.*, 2011] Patricia L Whetzel, Natalya F Noy, Nigam H Shah, Paul R Alexander, Csongor Nyulas, Tania Tudorache, and Mark A Musen. Bioportal. *Nucleic acids research*, 39(suppl 2):W541–W545, 2011.