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
{firstname.surname}@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.

Detailed proofs are at https://anon.to/XwVG7c.

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

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

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 $\mathcal{A}$:

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

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 $\mathcal{T}, \mathcal{A} \models_{\text{UNA}} q(\vec{a})$ iff $\mathcal{T}', \mathcal{A} \models_{\text{UNA}} q(\vec{a})$.

For the converse, we first establish 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 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 $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q(\vec{a})$ iff $\mathcal{T}', \mathcal{A} \models_{\text{nUNA}} q(\vec{a})$. □

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 $\mathcal{T}, \mathcal{A} \models_{\text{UNA}} q(\vec{a})$ iff either $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q(\vec{a})$ or there exist $a \neq b \in \mathsf{ind}(\mathcal{A})$ with $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} (a = b)$. □

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 $\mathcal{T}_{\text{horn}}$ 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 $\mathcal{T}_{\text{horn}}$. 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

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

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 $\mathsf{Novelist} \sqsubseteq \forall \mathsf{author}.\mathsf{Fiction}$. 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 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

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

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 a $\mathcal{ALCHIF}^{\sqsubseteq f}$ 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 $a' \sim b'$. Then 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 CO-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.

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.

A Basic Results and Notation

We have introduced the disjunction property of TBoxes. For some proofs it will be more convenient to work with the equivalent notion of materializability.

Definition 1 (Materializability)**.** Let $\mathcal{T}$ be a TBox, $\mathcal{Q}$ a class of CQs, and $\mathcal{M}$ a class of ABoxes. Then

- an interpretation $\mathcal{I}$ is a $\mathcal{Q}$*-materialization* of $\mathcal{T}$ and an ABox $\mathcal{A}$ if it is a model of $\mathcal{T}$ and $\mathcal{A}$ and for all $q(\vec{x}) \in \mathcal{Q}$ and $\vec{a}$ in $\mathsf{ind}(\mathcal{A})$, $\mathcal{I} \models q(\vec{a})$ iff $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} q(\vec{a})$.
- $\mathcal{T}$ is $\mathcal{Q}$*-materializable for* $\mathcal{M}$ if for every ABox $\mathcal{A} \in \mathcal{M}$ that is satisfiable w.r.t. $\mathcal{T}$ there is a $\mathcal{Q}$-materialization of $\mathcal{T}$ and $\mathcal{A}$.

If $\mathcal{M}$ is the class of all ABoxes, we simply speak of $\mathcal{Q}$-materializability of $\mathcal{T}$.

In the definition above, we define materializations and materializability without UNA. If we require materializations $\mathcal{I}$ to be models of $\mathcal{A}$ and $\mathcal{T}$ with UNA such that $\mathcal{I} \models q(\vec{a})$ iff $\mathcal{T}, \mathcal{A} \models_{\text{UNA}} q(\vec{a})$, then we speak about $\mathcal{Q}$-materializations with UNA and $\mathcal{Q}$-materializability for $\mathcal{M}$ with UNA. The equivalence of $\mathcal{Q}$-materializability and the $\mathcal{Q}$-disjunction property with and without UNA can be proved in a straightforward way. It has been shown with UNA for $\mathcal{ALCIF}$ in [Lutz and Wolter, 2017].

Lemma 5. *Let $\mathcal{Q}$ be a class of CQs and $\mathcal{T}$ an $\mathcal{ALCHIQ}$ TBox. Then $\mathcal{T}$ is $\mathcal{Q}$-materializable iff $\mathcal{T}$ has the $\mathcal{Q}$-disjunction property. This equivalence holds with and without the UNA.*

An interpretation $\mathcal{I}$ is a *tree interpretation* if the undirected graph $G_\mathcal{I} = (\Delta^\mathcal{I}, \bigcup_{r \in \mathsf{N_R}}(r^\mathcal{I} \cup (r^-)^\mathcal{I}))$ is a tree without self-loops. An interpretation $\mathcal{I}$ is a *closed model* of an ABox $\mathcal{A}$ if it is a model of $\mathcal{A}$ and $\Delta^\mathcal{I} = \{a^\mathcal{I} \mid a \in \mathsf{ind}(\mathcal{A})\}$. Let $\mathcal{J}$ be an interpretation and $\mathcal{I}$ be a tree interpretation with $\Delta^\mathcal{I} \cap \Delta^\mathcal{J} = \{d\}$. Then we say that the interpretation $\mathcal{J}'$ defined by setting

- $\Delta^{\mathcal{J}'} = \Delta^\mathcal{J} \cup \Delta^\mathcal{I}$;
- $A^{\mathcal{J}'} = A^\mathcal{J} \cup A^\mathcal{I}$ for all concept names A;
- $r^{\mathcal{J}'} = r^\mathcal{J} \cup r^\mathcal{I}$ for all role names r;
- $a^{\mathcal{J}'} = a^\mathcal{J}$ for all individual names a

is obtained from $\mathcal{J}$ by *hooking $\mathcal{I}$ to $\mathcal{J}$ at d*. Call a model $\mathcal{I}$ of $\mathcal{A}$ a *forest model of $\mathcal{A}$* if $\mathcal{I}$ is obtained from a closed model $\mathcal{J}$ of $\mathcal{A}$ by hooking tree interpretations to $\mathcal{J}$ for every $d \in \Delta^\mathcal{J}$.

A *homomorphism* h between interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ is a mapping from $\Delta^{\mathcal{I}_1}$ to $\Delta^{\mathcal{I}_2}$ such that

- if $d \in A^{\mathcal{I}_1}$, then $h(d) \in A^{\mathcal{I}_2}$ for all concept names A;
- if $(d, d') \in r^{\mathcal{I}_1}$, then $(h(d), h(d')) \in r^{\mathcal{I}_2}$ for all role names r;
- $h(a^{\mathcal{I}_1}) = a^{\mathcal{I}_2}$ for all individual names a.

The following result is folklore [Hernich *et al.*, 2017].

Lemma 6. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox and $\mathcal{A}$ an ABox satisfiable w.r.t. $\mathcal{T}$. Then $\mathcal{T}, \mathcal{A} \models q(\vec{a})$ for a CQ $q(\vec{x})$ and tuple $\vec{a}$ in $\mathsf{ind}(\mathcal{A})$ if $\mathcal{I} \models q(\vec{a})$ for all forest models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$. This holds with and without the UNA.*

If $\mathcal{T}$ is CQ-materializable, then there exists a forest model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ such that for every model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ there exists a homomorphism from $\mathcal{I}$ to $\mathcal{J}$. $\mathcal{I}$ is a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. This holds with and without the UNA.

We give a more detailed definition of the *quotient ABox* and also the *quotient interpretation*. For an ABox $\mathcal{A}$ and an equivalence relation $\sim$ on $\mathsf{ind}(\mathcal{A})$, we obtain the *quotient ABox* $\mathcal{A}/_\sim$ of $\mathcal{A}$ as follows: we regard the equivalence classes $a/_\sim$ of individual names a w.r.t. $\sim$ as individual names and include in $\mathcal{A}/_\sim$ the assertion

- $A(a/_\sim)$ if there exists $b \sim a$ such that $A(b) \in \mathcal{A}$;
- $r(a/_\sim, b/_\sim)$ if there exist $a' \sim a$ and $b' \sim b$ such that $r(a', b') \in \mathcal{A}$.

Let $\mathcal{I}$ be an interpretation and $\sim$ an equivalence relation on $\Delta^\mathcal{I}$. Then we define the *quotient interpretation* $\mathcal{J} = \mathcal{I}/_\sim$ as follows:

- $\Delta^\mathcal{J} = \{d/_\sim \mid d \in \Delta^\mathcal{I}\}$;
- for every individual name a: $a^\mathcal{J} = a^\mathcal{I}/_\sim$;
- for every concept name A: $A^\mathcal{J} = \{d/_\sim \mid d \in A^\mathcal{I}\}$;
- for every role name r: $r^\mathcal{J} = \{(d/_\sim, d'/_\sim) \mid (d, d') \in r^\mathcal{I}\}$.

For an ABox $\mathcal{A}$, we denote by $\mathcal{I}_\mathcal{A}$ the interpretation obtained from $\mathcal{A}$ in the obvious way by setting

- $\Delta^{\mathcal{I}_\mathcal{A}} = \mathsf{ind}(\mathcal{A})$;
- $a \in A^{\mathcal{I}_\mathcal{A}}$ if $A(a) \in \mathcal{A}$;
- $(a, b) \in r^{\mathcal{I}_\mathcal{A}}$ if $r(a, b) \in \mathcal{A}$;
- $a^{\mathcal{I}_\mathcal{A}} = a$ for all individual names a.

Let C be a concept of depth 1 constructed from concept names and the concept $\top$ using the constructors $\sqcap$ and $(\geq nr.C)$. Then the ABox $\mathcal{A}_C$ *corresponding to* C with root a_0 contains

- $A(a_0)$, for every top-level conjunct $A \in \mathsf{N_C}$ of C;
- $r(a_0, a_i)$ and $A(a_i)$ for $1 \leq i \leq n$, for every top-level conjunct $(\geq nr.C)$ of C and conjunct A of C, where we assume that the sets $\{a_1, \dots, a_n\}$ of individual names have cardinality n and are mutually disjoint for every concept $(\geqslant n\, r\, C)$.

Here and in what follows we use $r(a_0, a_i)$ as a shorthand for $s(a_i, a_0)$ if r is an inverse role with $r = s^-$.

B Experiments Reported in Section 1

We ran experiments to understand how many CIs α with proper number restrictions (i.e., not equivalent to existential or universal restrictions) and of depth 1 from real life ontologies can be easily classified according to whether

- CQ-evaluation w.r.t. the TBox $\mathcal{T}_\alpha = \{\alpha\}$ is in PTIME with UNA;
- $\mathcal{T}_\alpha$ is CQ-Horn-rewritable.

We devised seven schemata of CIs such that for all CIs α belonging to any of the schemata CQ-evaluation w.r.t. $\mathcal{T}_\alpha$ is in PTIME with UNA and for all except the schema **Horn**, the TBox $\mathcal{T}_\alpha$ is not CQ-Horn-rewritable. The schemata are as follows, where E is a potentially empty (i.e., $\top$) conjunction of concept names, F a potentially empty conjunction of the form $E \sqcap \bigsqcap(\geqslant n_i\, r_i\, E_i)$, and for any role r and any two distinct restrictions $(\mathcal{Q}\ n_i\, r\, E_i)$ and $(\mathcal{Q}'\ n_j\, r\, E_j)$ with $\mathcal{Q}, \mathcal{Q}' \in \{\geq, =, \leq\}$ neither $E_i \subseteq E_j$ nor $E_j \subseteq E_i$.

- **Horn**: the CI is a Horn-CI;
- **A_exact**: CIs of the form $A \sqsubseteq F \sqcap \bigsqcap(= n_j\, r_j\, E_j)$;
- **A_max**: CIs of the form $A \sqsubseteq F \sqcap \bigsqcap(\leqslant n_j\, r_j\, E_j)$;
- **equiv_exact**: CIs of the form $A \equiv F \sqcap \bigsqcap(= n_j\, r_j\, E_j)$;
- **equiv_min**: CIs of the form $A \equiv F$ with F containing at least one number restriction;
- **exact_A**: CIs of the form $F \sqcap \bigsqcap(= n_j\, r_j\, E_j) \sqsubseteq A$;
- **min_A**: CIs of the form $F \sqsubseteq A$ with F containing at least one number restriction;

Schemata	BioPortal	ORE
Horn	2911	5156
A_exact	1390	1446
A_max	6	15
equiv_exact	642	234
equiv_min	23	20
exact_A	2	2
min_A	3	3
unknown	104	82
Total	5081	6958

Table 1: CI classification

The dataset is composed of a snapshot of BioPortal[1] and the corpus used for the ORE 2015 competition[2]. For each parsable ontology in the dataset, all CIs with at least one number restriction not equivalent to either an existential or a universal restriction were collected, and a simple normalization was performed in order to avoid tautological CIs and false negatives. The resulting dataset is composed of 97 ontologies (5081 axioms) from BioPortal and 447 ontologies (6958 axioms) from the ORE corpus. All the experiments were ran on a machine equipped with an Intel Core i7-2600 CPU with 3.40GHz and 16GB of RAM. Each CI is classified in one of the seven schemata or as 'unknown' if it does not belong to any of them. Note that for a CI α classified as unknown the CQ-evaluation with UNA might still be in PTIME, but it simply cannot be classified according to our schemata. Tables 1 summarizes the classifications of the CIs, and Table 2 summarizes the occurrence of CI schemata in the ontologies under consideration without counting those containing only Horn-CIs.

From Table 1, only 104 out of 5081 (2.05%) of the CIs from BioPortal, and 82 out of 6958 (1.18%) of the CIs from the ORE corpus are not classifiable. Many ontologies, however, contain only Horn CIs. Specifically, 55 out of 97 ontologies from BioPortal, and 259 out of 447 ontologies from the ORE corpus. After removing such ontologies the total number of

[1] https://zenodo.org/record/439510
[2] https://zenodo.org/record/18578

Schemata	BioPortal	ORE
Horn	33	155
A_exact	22	139
A_max	6	13
equiv_exact	20	52
equiv_min	11	16
exact_A	1	1
min_A	1	1
unknown	22	20
# Ontologies	42	188

Table 2: Occurrence of CI schemata

axioms drops to 3420 and 4243 respectively, but the non-classifiable CIs are still just 3.04% and 1.93%, respectively. Table 2 shows the distribution of occurrences of schemata in the ontologies without counting those containing only Horn CIs. For Table 2, it is worth pointing out that only one ontology in the BioPortal dataset and only two ontologies in the ORE dataset contain only CIs classified as 'unknown', resulting in 41 out of 42 ontologies with at least one non-Horn CI classified as not CQ-Horn-rewritable for BioPortal, and 186 out of 188 for ORE.

From the tables it is evident that two of the patterns do not occur often, namely, the "exact_A" and "min_A" schemata. The reason being that common ontology editors (e.g., Protégé) discourage the creation of this form of CIs, while encouraging CIs of the form $A \sqsubseteq C$ or $A \equiv C$.

C Proofs for Section 3

Lemma 1 (restated). *If $\mathcal{T}$ is an $\mathcal{ALCHIQ}$ TBox, then the following are equivalent:*

1. *$\mathcal{T}$ has the ELIQ-disjunction property;*
2. *$\mathcal{T}$ has the ELIQ$^=$-disjunction property;*
3. *$\mathcal{T}$ has the CQ-disjunction property.*

These equivalences hold both with and without UNA.

Proof. We prove the case without UNA of which the case with UNA is a special case. The implication $3 \Rightarrow 1$ is trivial. In what follows, we prove $1 \Rightarrow 2$ and $2 \Rightarrow 3$.

To prove the implication $1 \Rightarrow 2$, assume that $\mathcal{T}$ has the ELIQ-disjunction property without UNA. Suppose there is an ABox $\mathcal{A}$, queries $q_1(\vec{x}_1), \dots, q_n(\vec{x}_n)$ in ELIQ$^=$, and tuples $\vec{a}_1, \dots, \vec{a}_n$ in $\mathcal{A}$ such that $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \le i \le n} q_i(\vec{a}_i)$ and $\mathcal{T}, \mathcal{A} \not\models_{\text{nUNA}} q_i(\vec{a}_i)$ for all $1 \le i \le n$. Let $\mathcal{A}'$ be a new ABox obtained from $\mathcal{A}$ by adding an assertion $A_a(a)$ for every individual a that occurs in some of the tuples $\vec{a}_i$, where A_a is a fresh concept name. For every $i \in \{1, \dots, n\}$, let $q'_i(\vec{a}_i)$ denote $q_i(\vec{a}_i)$ if $q_i(\vec{a}_i)$ is an ELIQ, and let $q'_i(\vec{a}_i) := A_{a_1}(a_1) \wedge A_{a_2}(a_1)$ if $q_i(a_1, a_2)$ is an equality $(a_1 = a_2)$. Then $\mathcal{T}, \mathcal{A}' \models_{\text{nUNA}} \bigvee_{1 \le i \le n} q'_i(\vec{a}_i)$ but $\mathcal{T}, \mathcal{A}' \not\models_{\text{nUNA}} q'_i(\vec{a}_i)$ for all $1 \le i \le n$, which contradicts the fact that $\mathcal{T}$ has the ELIQ-disjunction property without UNA.

For the implication $2 \Rightarrow 3$, it suffices to prove, by Lemma 5, that if $\mathcal{T}$ is ELIQ$^=$-materializable, then $\mathcal{T}$ is CQ-materializable. A similar statement, that ELIQ-materializability implies CQ-materializability for $\mathcal{ALCIF}$ TBoxes, was proved in [Lutz and Wolter, 2017, Theorem 16]. In the following, we generalize this result.

Assume that $\mathcal{T}$ is ELIQ$^=$-materializable. To show that $\mathcal{T}$ is CQ-materializable, let $\mathcal{A}$ be an ABox that is satisfiable w.r.t. $\mathcal{T}$ and let $\mathcal{I}$ be an ELIQ$^=$-materialization of $\mathcal{T}$ and $\mathcal{A}$. By the following claim, we can assume, without loss of generality, that $\mathcal{I}$ is a forest model of $\mathcal{T}$ and $\mathcal{A}$.

CLAIM 1. *For every model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ there exists a forest model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ and a homomorphism from $\mathcal{J}$ to $\mathcal{I}$.*

Proof. This was proved in [Hernich *et al.*, 2017, Lemma 1] with UNA in a more general context, and readily carries over to the case without UNA. ⌟

Since $\mathcal{I}$ is a forest model of $\mathcal{T}$ and $\mathcal{A}$, it is obtained from a closed model $\mathcal{I}_0$ of $\mathcal{A}$ by hooking tree interpretations $\mathcal{I}_d$ to $\mathcal{I}_0$ for every $d \in \Delta^{\mathcal{I}_0}$:

$$\mathcal{I} = \mathcal{I}_0 \cup \bigcup_{d \in \Delta^{\mathcal{I}_0}} \mathcal{I}_d.$$

Here, we have $\Delta^{\mathcal{J}_d} \cap \Delta^{\mathcal{J}_0} = \{d\}$ and $\Delta^{\mathcal{J}_d} \cap \Delta^{\mathcal{J}_{d'}} = \emptyset$ for all $d, d' \in \Delta^{\mathcal{J}_0}$ with $d \neq d'$.

We show that $\mathcal{I}$ is a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. To this end, it suffices to show that for every finite subset $Q \subseteq \mathcal{I}$ and every model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ there exists a homomorphism from Q to $\mathcal{J}$.

Let Q be a finite subset of $\mathcal{I}$ and let $\mathcal{J}$ be a model of $\mathcal{T}$ and $\mathcal{A}$. Without loss of generality, we may assume that $\Delta^{\mathcal{I}_0} \subseteq \Delta^Q$ and that $Q \cap \mathcal{I}_d$ is connected for every $d \in \Delta^{\mathcal{I}_0}$. We will need the following result:

CLAIM 2. *For every $d \in \Delta^{\mathcal{I}_0}$ there exists $\mathcal{I}'_d \subseteq \mathcal{I}_d$ such that*

- *$\Delta^{\mathcal{I}_d} = \Delta^{\mathcal{I}'_d}$;*
- *$A^{\mathcal{I}_d} = A^{\mathcal{I}'_d}$ for all concept names A;*
- *$r^{\mathcal{I}_d} = \{(e, e') \mid (e, e') \in s^{\mathcal{I}'_d}, s \text{ is a role}, \mathcal{T} \models s \sqsubseteq r\}$ for all role names r; and*
- *$r^{\mathcal{I}'_d} \cap s^{\mathcal{I}'_d} = \emptyset$ for all distinct roles r and s.*

Proof. We obtain $\mathcal{I}'_d$ from $\mathcal{I}_d$ by repeatedly applying the following procedure: if there is a role name r and a role $s \neq r$ with $\mathcal{T} \models s \sqsubseteq r$ and $r^{\mathcal{I}_d} \cap s^{\mathcal{I}_d} \neq \emptyset$, then remove from $r^{\mathcal{I}_d}$ all pairs in $r^{\mathcal{I}_d} \cap s^{\mathcal{I}_d}$. ⌟

Let $d \in \Delta^{\mathcal{I}_0}$ and let $\mathcal{I}'_d$ be as in Claim 2. We view $Q \cap \mathcal{I}'_d$ as an ELIQ $C_d(x)$ with the property that for all interpretations $\mathcal{K}$ with $d \in \Delta^{\mathcal{K}}$ we have $\mathcal{K} \models C_d(d)$ iff there exists a homomorphism from $Q \cap \mathcal{I}'_d$ to $\mathcal{K}$. Since $\mathcal{I} \models C_d(d)$ and $\mathcal{I}$ is an ELIQ$^=$-materialization of $\mathcal{T}$ and $\mathcal{A}$, we have $\mathcal{J} \models C_d(d)$. Hence, there exists a homomorphism h_d from $Q \cap \mathcal{I}'_d$ to $\mathcal{J}$, and by our choice of $\mathcal{I}'_d$ this homomorphism is also a homomorphism from $Q \cap \mathcal{I}_d$ to $\mathcal{J}$.

To conclude the proof, we note that the union of all homomorphisms h_d, for $d \in \Delta^{\mathcal{I}_0}$, is a homomorphism from Q to $\mathcal{J}$, as desired. □

The following lemma plays an important role in the proof of Theorem 1. It holds with and without UNA, but we restrict our attention to the variant that is used for Theorem 1.

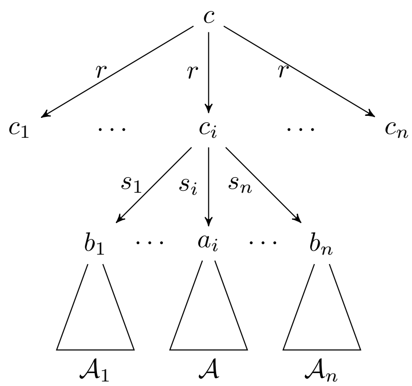

Figure 1: Illustration of the ABox $\mathcal{A}'$ in the proof of Lemma 7.

Lemma 7. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox, let $\mathcal{T}'$ be a CQ-Horn-rewriting of $\mathcal{T}$, and let $q_1(\vec{x}_1), \ldots, q_n(\vec{x}_n)$ be CQs. Then for all ABoxes $\mathcal{A}$ and tuples $\vec{a}_1, \ldots, \vec{a}_n$ in $\mathcal{A}$:*

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

Proof. It suffices to focus on the case that each query $q_i(\vec{x}_i)$ is an ELIQ. Indeed, we know that $\mathcal{T}'$ has the ELIQ-disjunction property (because it is a Horn-$\mathcal{ALCHIQ}$ TBox), so if the lemma holds for ELIQs, then the ELIQ-disjunction property carries over to $\mathcal{T}$ (because $\mathcal{T}'$ is a CQ-Horn-rewriting of $\mathcal{T}$). By Lemma 1, $\mathcal{T}$ and $\mathcal{T}'$ have the CQ-disjunction property, which implies the lemma.

To prove the lemma for ELIQs, we simulate disjunctions of ELIQs by single ELIQs. To this end, let $\mathcal{A}$ be an ABox, let $C_1(x), \ldots, C_n(x)$ be ELIQs, and let $a_1, \ldots, a_n \in \mathsf{ind}(\mathcal{A})$. For each $i \in \{1, \ldots, n\}$, let $\mathcal{A}_i$ be an ABox that is satisfiable w.r.t. $\mathcal{T}$ and satisfies $\mathsf{ind}(\mathcal{A}_i) \cap \mathsf{ind}(\mathcal{A}) = \emptyset$, $\mathsf{ind}(\mathcal{A}_i) \cap \mathsf{ind}(\mathcal{A}_j) = \emptyset$ for all $j \neq i$, and $\mathcal{A}_i \models C_i(b_i)$ for some individual b_i. Define a new ABox $\mathcal{A}'$ as follows:

- initialize $\mathcal{A}'$ to $\mathcal{A} \cup \mathcal{A}_1 \cup \ldots \cup \mathcal{A}_n$;
- pick fresh individuals $c, c_1, \ldots, c_n$ and a fresh role name r and add the assertions $r(c, c_1), \ldots, r(c, c_n)$ to $\mathcal{A}'$;
- pick fresh role names $s_1, \ldots, s_n$ and add the assertion $s_j(c_i, d_j)$ to $\mathcal{A}'$ for all $1 \leq i, j \leq n$, where $d_j = a_i$ if $j = i$ and $d_j = b_j$ for all $j \neq i$.

See Figure 1 for an illustration. It is straightforward to check that $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \leq i \leq n} C_i(a_i)$ iff $\mathcal{T}, \mathcal{A}' \models_{\text{nUNA}} C(c)$, where $C(x)$ is the ELIQ defined by $C = \exists r. \bigsqcap_{1 \leq i \leq n} \exists s_i.C_i$, and accordingly for $\mathcal{T}'$. Since $\mathcal{T}'$ is a CQ-Horn-rewriting of $\mathcal{T}$, this equivalence implies that $\mathcal{T}, \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \leq i \leq n} C_i(a_i)$ iff $\mathcal{T}', \mathcal{A} \models_{\text{nUNA}} \bigvee_{1 \leq i \leq n} C_i(a_i)$. □

Lemma 2 (restated). *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.*

Proof. The result with UNA follows from Theorem 3 in [Hernich *et al.*, 2017], and it remains to prove the result without UNA. Assume that $\mathcal{T}$ does not have the ELIQ$^=$-disjunction property without UNA. By Lemma 1, $\mathcal{T}$ also does not have the ELIQ-disjunction property without UNA. We show that ELIQ-evaluation for $\mathcal{T}$ without UNA is CONP-hard. As ELIQ is a subset of ELIQ$^=$, this implies that ELIQ$^=$-evaluation for $\mathcal{T}$ without UNA is CONP-hard.

We use the same construction as in the proof of Theorem 18 in [Lutz and Wolter, 2017], where CONP-hardness of ELIQ-evaluation with UNA was shown for $\mathcal{ALCIF}$ TBoxes that lack the ELIQ-disjunction property with UNA, but we have to argue more carefully that it works as intended. More precisely, we give a reduction from 2+2 SAT. The input to 2+2 SAT is a propositional formula in conjunctive normal form, where each clause has the form $(p_1 \vee p_2 \vee \neg n_1 \vee \neg n_2)$ and each of p_1, p_2, n_1, n_2 is a variable or one of the truth constants $0, 1$. Given such a formula, we want to decide if it is satisfiable. This problem was shown to be NP-complete in [Schaerf, 1993].

Since $\mathcal{T}$ does not have the ELIQ-disjunction property without UNA, there is an ABox $\mathcal{A}_\vee$ and a minimal sequence $C_0(x), \ldots, C_k(x)$ of ELIQs and individual names $a_0, \ldots, a_k$ in $\mathsf{ind}(\mathcal{A}_\vee)$ such that:

$$\mathcal{T}, \mathcal{A}_\vee \models_{\text{nUNA}} \bigvee_{0 \leq i \leq k} C_i(a_i);$$

$$\mathcal{T}, \mathcal{A}_\vee \not\models_{\text{nUNA}} C_i(a_i) \quad \text{for all } i \in \{0, 1, \ldots, k\}.$$

We use $\mathcal{A}_\vee$, $C_0(x), \ldots, C_k(x)$, and $a_0, \ldots, a_k$ to encode truth values for the variables in a given 2+2 SAT instance.

In what follows, we first consider the case that $k = 1$, and then show how to generalize the construction to larger k.

Let $\varphi = \varphi_1 \wedge \cdots \wedge \varphi_m$ be a 2+2 SAT instance with variables $x_1, \ldots, x_n$. We represent φ by an ABox $\mathcal{A}_\varphi$ that we construct as follows:

- We use an individual name f to represent φ, for each $i \in \{1, \ldots, m\}$ we use an individual name c_i to represent the clause φ_i, for each $j \in \{1, \ldots, n\}$ we view the variable x_j as an individual name that represents x_j itself, and we use the truth constants 0 and 1 as individual names representing themselves.
- We use role names c, p_1, p_2, n_1, n_2 that do not occur in $\mathcal{T}$ or $\mathcal{A}_\vee$ to connect the above individual names with each other. More precisely, for each $i \in \{1, \ldots, m\}$ we add to $\mathcal{A}_\varphi$ the assertion $c(f, c_i)$; and assuming that $\varphi_i = (p_1^i \vee p_2^i \vee \neg n_1^i \vee \neg n_2^i)$, we also add the assertions $p_1(c_i, p_1^i)$, $p_2(c_i, p_2^i)$, $n_1(c_i, n_1^i)$, and $n_2(c_i, n_2^i)$.

We further extend $\mathcal{A}_\varphi$ to enforce a truth value for each of the variables x_i. We first add n disjoint copies $\mathcal{A}_1, \ldots, \mathcal{A}_n$ of $\mathcal{A}_\vee$ to $\mathcal{A}_\varphi$. For all $i \in \{1, \ldots, n\}$ and $j \in \{0, \ldots, k\}$, let $a_{i,j}$ denote the name of the copy of $a_j \in \mathsf{ind}(\mathcal{A}_\vee)$ in $\mathcal{A}_i$. We then use fresh role names $r_0, \ldots, r_k$ to connect each variable x_i to $\mathcal{A}_i$ as follows:

- For all $i \in \{1, \ldots, n\}$ and $j \in \{0, \ldots, k\}$, we add the assertion $r_j(x_i, a_{i,j})$. We will interpret the fact that

  $$\mathsf{tt}(x) = \exists y \, (r_0(x, y) \wedge C_0(y))$$

  is true at x_i as "x_i is true" and the fact that

  $$\mathsf{ff}(x) = \exists y \, (r_1(x, y) \wedge C_1(y))$$

  is true at x_i as "x_i is false".

- To ensure that 0 and 1 have the expected truth values, we also add a copy of C_0 viewed as an ABox with root $0'$ and a copy of C_1 viewed as an ABox with root $1'$. We also add assertions $r_0(0, 0')$ and $r_1(1, 1')$.

Now, the ELIQ

$$q(x) = \exists y\big(c(x, y) \wedge \bigwedge_{i\in\{1,2\}} \exists z(p_i(y, z) \wedge \mathsf{ff}(z)) \wedge \bigwedge_{i\in\{1,2\}} \exists z(n_i(y, z) \wedge \mathsf{tt}(z))\big)$$

states the existence of a clause in φ with only false literals and thus captures falsity of φ.

CLAIM. *φ is satisfiable iff $\mathcal{T}, \mathcal{A}_\varphi \not\models_{\text{nUNA}} q(f)$.*

Proof. The "only if" direction can be dealt with as in the case with UNA. Assume that φ is satisfiable. Then there exists a truth assignment $\tau\colon \{x_1, \dots, x_n\} \to \{t, f\}$ such that every clause of φ is true under τ. For each $i \in \{1, \dots, n\}$, let $\mathcal{I}_i$ be a model of $\mathcal{T}$ and $\mathcal{A}_i$ such that

- $\mathcal{I}_i \models C_0(a^{\mathcal{I}}_{i,0})$ and $\mathcal{I}_i \not\models C_1(a^{\mathcal{I}}_{i,1})$ if $\tau(x_i) = t$;
- $\mathcal{I}_i \not\models C_0(a^{\mathcal{I}}_{i,0})$ and $\mathcal{I}_i \models C_1(a^{\mathcal{I}}_{i,1})$ if $\tau(x_i) = f$.

Using $\mathcal{I}_1, \dots, \mathcal{I}_n$ it is now easy to construct a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}_\varphi$ with $\mathcal{I} \not\models q(f^{\mathcal{I}})$.

For the converse, let φ be unsatisfiable and let $\mathcal{I}$ be a model of $\mathcal{T}$ and $\mathcal{A}_\varphi$. We define a truth assignment τ for the variables in φ as follows. For each $j \in \{1, \dots, n\}$, we set $\tau(x_j) = t$ if $\mathcal{I} \models C_0(a^{\mathcal{I}}_{j,0})$, and we set $\tau(x_j) = f$ otherwise. Note that $\mathcal{I} \models C_1(a^{\mathcal{I}}_{j,1})$ if $\tau(x_j) = f$. Since φ is unsatisfiable, there exists a clause $\varphi_i = (p^i_1 \vee p^i_2 \vee \neg n^i_1 \vee \neg n^i_2)$ in φ such that $\tau(p^i_1) = \tau(p^i_2) = f$ and $\tau(n^i_1) = \tau(n^i_2) = t$. This implies that the subquery

$$q_0(y) = \bigwedge_{i\in\{1,2\}} \exists z(p_i(y, z) \wedge \mathsf{ff}(z)) \wedge \bigwedge_{i\in\{1,2\}} \exists z(n_i(y, z) \wedge \mathsf{tt}(z))$$

of $q(x)$ is true at $c^{\mathcal{I}}_i$ in $\mathcal{I}$. This establishes that $\mathcal{I} \models q(f^{\mathcal{I}})$. ⌟

If $k > 1$, then we use

$$\mathsf{ff}(x) = \bigvee_{1\leq j\leq k} \exists y\,(r_j(x, y) \wedge C_j(y))$$

to express that "x is false". By distributing disjunctions in $q(x)$ to the outside, we then obtain a union of ELIQs. Using a simulation of disjunctions of ELIQs by single ELIQs similar to the one in the proof of Lemma 7 we can transform $\mathcal{A}_\varphi$ into an ABox $\mathcal{A}'_\varphi$ and $q(x)$ into an ELIQ $q'(x)$ such that φ is unsatisfiable iff $\mathcal{T}, \mathcal{A}'_\varphi \models_{\text{nUNA}} q'(f)$. Details can be found in the proof of Theorem 18 in [Lutz and Wolter, 2017]. □

D Proofs for Section 4

We state the result to be proved again.

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

Proof. The normalization is performed in two steps. First, each concept of the form $(\mathcal{Q}\, n\, r\, C)$, with $\mathcal{Q} \in \{\leq, \geq\}$ and C a Boolean combination of concept names, is transformed into an equivalent concept which is a Boolean combination of concepts of the form $(\mathcal{Q}\, n\, r\, E)$ with E a conjunction of literals. Second, each CI $C \sqsubseteq D$ is split into several CIs in normal form. All steps are equivalence preserving steps. The first normalization step is defined as follows.

1. For each concept $C = (\mathcal{Q}\, n\, r\, C')$ with $\mathcal{Q} \in \{\leq, \geq\}$ and $n > 1$ if $\mathcal{Q}$ is $\geq$, and $n > 0$ otherwise
   - let $C_1 \sqcup \ldots \sqcup C_m = \mathsf{DNF}(C')$, $A_1, \dots, A_h$ be the concept names occurring in C', $A'_j := A_j \mid \neg A_j$ with $1 \leq j \leq h$;
   - let $S_{C_i} = \{C_i \sqcap \bigsqcap A'_j \mid A_j \notin C_i\}$ for each disjunct C_i of $\mathsf{DNF}(C')$, and let $S = \bigcup_{i=1}^m S_{C_i}$;
   - let $k = |S_C| \leq 2^h$ and $l = \binom{k+n-1}{n}$ where h is the number of concept names occurring in C'. We assume an enumeration of the k elements $D_i \in S$. Replace C with the equivalent concept
   $$\bigsqcup_{j=1}^{l} \bigsqcap_{i=1}^{k} (\mathcal{Q}\, n^j_i\, r\, D_i)$$
   where $n^j_i \in \{0, 1, \dots, n\}$ and $\Sigma_{i=1}^k n^j_i = n$ for each j. The resulting disjunction represents all possible combination for which C can be satisfiable.
2. for each concept $C = (\geqslant 0\, r\, C')$, replace C with $\top$;
3. for each concept $C = (\geqslant 1\, r\, C')$ with $C_1 \sqcup \ldots \sqcup C_m = \mathsf{DNF}(C')$, distribute the number restriction over the disjuncts (i.e., $(\geqslant 1\, r\, C_1) \sqcup \ldots \sqcup (\geqslant 1\, r\, C_m)$);
4. for each concept $C = (\leqslant 0\, r\, C')$ with $C_1 \sqcup \ldots \sqcup C_m = \mathsf{DNF}(C')$, distribute the number restriction over the disjuncts (i.e., $(\leqslant 0\, r\, C_1) \sqcap \ldots \sqcap (\leqslant 0\, r\, C_m)$).

The second normalization step is performed as follows. For any CIs $C \sqsubseteq D \in \mathcal{T}$:

- convert the left-hand side to DNF, the right-hand side to CNF, remove disjunctions on the left and conjunctions on the right by splitting into multiple CIs; the left-hand side is now a conjunction on top-level, and the right-hand side is a disjunction on top-level;
- move negative literals left to right and vice versa (negating them), move any $(\geqslant n\, r\, E)$ on the left-hand side with at least one negative literal in E to the right-hand side while computing its negation (i.e., $(\leqslant n-1\, r\, E)$), move any $(\leqslant n\, r\, E)$ on the right-hand side with no negative literal in E to the left-hand side while computing its negation (i.e., $(\geqslant n+1\, r\, E)$), move any $(\leqslant n\, r\, E)$ on the left-hand side to the right-hand side while computing its negation.

As both normal form transformation steps can result in an exponential blowup, and both perform only equivalence preserving transformations; the resulting TBox is of size at most double exponential in $|\mathcal{T}|$, equivalent to $\mathcal{T}$ and in normal form. □

Lemma 4 (restated). *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.*

Proof. (3.) $\Rightarrow$ (1.) was proved in detail in the main text.

(1.) $\Rightarrow$ (2.). Assume $\mathcal{T}$ has the ELIQ$^=$-disjunction property and that there are $C \sqsubseteq D \in \mathcal{T}$ and a trigger S for $C \sqsubseteq D$ such that $\mathsf{Horn}(C \sqsubseteq D, S) = \emptyset$. We derive a contradiction. Let

- $A_1, \dots, A_k$ be the disjuncts in D in $\mathsf{N_C}$,
- $(\geqslant n_1\, r_1\, E_1), \dots, (\geqslant n_n\, r_n\, E_n)$ be the disjuncts in D with $\mathcal{T} \not\models C_S^{\leq 1} \sqsubseteq \neg(\geqslant n_i\, r_i\, E_i)$ for $1 \leq i \leq n$, and
- let $(\leqslant m_1\, s_1\, F_1), \dots, (\leqslant m_m\, s_m\, F_m)$ be the remaining disjuncts of D.

Let $(\geqslant k_1\, t_1\, G_1), \dots, (\geqslant k_k\, t_k\, G_k)$ be the number restrictions in S not introduced due to the disjuncts $(\leqslant m_1\, s_1\, F_1), \dots, (\leqslant m_m\, s_m\, F_m)$ of D. Let $\mathcal{A}_S$ be the ABox corresponding to C_S with root a_0. Thus, we have individual names

- $b_1^1, \dots b_1^{k_1}, \dots, b_k^1, \dots, b_k^{k_k}$ in $\mathcal{A}_S$ introduced for the concepts $(\geqslant k_i\, t_i\, G_i) \in S$ with $t_i(a_0, b_i^j) \in \mathcal{A}_S$ for $1 \leq i \leq k$ and $1 \leq j \leq k_i$; and
- $a_1^1, \dots, a_1^{m_1+1}, \dots, a_m^1, \dots, a_m^{m_m+1}$ in $\mathcal{A}_S$ introduced for the concepts $(\geqslant m_i + 1\, s_i\, \mathsf{pos}(F_i)) \in S$ with $s_i(a_0, a_i^j) \in \mathcal{A}_S$ for $1 \leq i \leq m$ and $1 \leq j \leq m_i + 1$.

It follows from $\mathcal{T} \models C_S \sqsubseteq D$ that

$$\begin{aligned}
\mathcal{T}, \mathcal{A}_S \quad \models \quad & \bigvee_{i=1..k} A_i(a_0) \vee \bigvee_{i=1..n} \exists r_i.\mathsf{pos}(E_i)(a_0) \\
& \vee \bigvee_{i=1..m}\ \bigvee_{j=1..m_i+1, A \in \mathsf{neg}(F_i)} A(a_i^j) \\
& \vee \bigvee_{i=1..k} \bigvee_{j \neq j'} (b_i^j = b_i^{j'}) \vee \bigvee_{i=1..m} \bigvee_{j \neq j'} (a_i^j = a_i^{j'})
\end{aligned}$$

By the ELIQ$^=$-disjunction property of $\mathcal{T}$ we have $\mathcal{T}, \mathcal{A}_S \models F$ for a disjunct F. If $\mathcal{A}_S$ is not satisfiable w.r.t. $\mathcal{T}$, then $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq \bot$ and so $\mathsf{Horn}(C \sqsubseteq D, S) \neq \emptyset$, which contradicts our assumption. Thus, assume $\mathcal{A}_S$ is satisfiable w.r.t. $\mathcal{T}$. Now observe that

- if $F = A_i(a_0)$ then $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq A_i$ and so $C_S^{\leq 1} \sqsubseteq A_i$ is in $\mathsf{Horn}(C \sqsubseteq D, S)$;
- if $F = A(a_i)$ for some $A \in \mathsf{neg}(F_i)$ then $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq \forall s_i.(\mathsf{pos}(F_i) \to A)$ and so $C_S^{\leq 1} \sqsubseteq \forall s_i.(\mathsf{pos}(F_i) \to A)$ is in $\mathsf{Horn}(C \sqsubseteq D, S)$;
- if $F = \exists r_i.\mathsf{pos}(E_i)(a_0)$, then $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq \exists r_i.\mathsf{pos}(E_i)$ and $\mathcal{T} \not\models C_S^{\leq 1} \sqsubseteq \neg(\geqslant n_i\, r_i\, E_i)$; thus $C_S^{\leq 1} \sqsubseteq \exists r_i.\mathsf{pos}(E_i)$ is in $\mathsf{Horn}(C \sqsubseteq D, S)$.
- if $F = (c = d)$ for some $c \neq d$, then $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq (\leqslant 1\, r\, E)$, for some $(\geqslant n\, r\, E) \in S$ with $n \geq 2$. Then $C_S^{\leq 1} \sqsubseteq (\leqslant 1\, r\, E)$ is in $\mathsf{Horn}(C \sqsubseteq D, S)$.

In each case, we obtain $\mathsf{Horn}(C \sqsubseteq D, S) \neq \emptyset$ and thus derive a contradiction.

(2.) $\Rightarrow$ (3.). It suffices to show that for every ABox $\mathcal{A}$ the following holds: if $\mathcal{A}$ is satisfiable relative to $\mathcal{T}_{\text{horn}}$, then there exists a CQ-materialization $\mathcal{U}$ of $\mathcal{A}$ and $\mathcal{T}_{\text{horn}}$ that is a model of $\mathcal{T}$. $\mathcal{U}$ is constructed using a set $R_{\mathcal{T}}$ of chase rules. The rules take as input an interpretation $\mathcal{I}$ and construct a new interpretation $\mathcal{J}$ which is obtained from $\mathcal{I}$ by either adding new facts of the form $A(d)$ and $r(d, d')$ to $\mathcal{I}$ or by 'identifying' two elements d, d' of $\Delta^{\mathcal{I}}$. To ensure that a model of $\mathcal{T}$ is constructed, our rules are slightly different from the standard chase for Horn-DLs, however. At each stage of the construction we have a model $\mathcal{I}$ of $\mathcal{A}$ in which to every $a^{\mathcal{I}}$ with $a \in \mathsf{ind}(\mathcal{A})$ a tree interpretation $\mathcal{I}_a$ is hooked. We use standard terminology to speak about $\mathcal{I}_a$. The rules are now as follows:

1. if $(d, e) \in r^{\mathcal{I}}$ and $r \sqsubseteq s \in \mathcal{T}$, then obtain $\mathcal{J}$ from $\mathcal{I}$ by adding (d, e) to $s^{\mathcal{I}}$;
2. if $d \in C^{\mathcal{I}}$ and $C \sqsubseteq A \in \mathcal{T}_{\text{horn}}$, then obtain $\mathcal{J}$ from $\mathcal{I}$ by adding d to $A^{\mathcal{I}}$;
3. if $d \in C^{\mathcal{I}}$ and $C \sqsubseteq R \in \mathcal{T}_{\text{horn}}$ for $R = \forall r.(A_1 \sqcap \dots \sqcap A_n \to A)$, then add e to $A^{\mathcal{I}}$ whenever $(d, e) \in r^{\mathcal{I}}$ and $e \in (A_1 \sqcap \dots \sqcap A_n)^{\mathcal{I}}$.
4. if $d \in C^{\mathcal{I}}$ and $C \sqsubseteq (\leqslant 1\, r\, E) \in \mathcal{T}_{\text{horn}}$ and $d_1 \neq d_2$ with $(d, d_1), (d, d_2) \in r^{\mathcal{I}}$ and $d_1, d_2 \in E^{\mathcal{I}}$, then construct $\mathcal{J}$ as follows:
   - if there are $a, b \in \mathsf{ind}(\mathcal{A})$ with $a^{\mathcal{I}} = d_1$ and $b^{\mathcal{I}} = d_2$, then let $\mathcal{J}$ be the quotient of $\mathcal{I}$ by $\sim$, where $\sim$ is the smallest equivalence relation on $\Delta^{\mathcal{I}}$ with $d_1 \sim d_2$. To keep track of the equivalence class of d_i in $\mathcal{J}$ we introduce two binary relations, $\prec$ and $\prec_c$. In this case, there is no difference between d_1 and d_2 and we set $d_i \prec \{d_1, d_2\}$ and $d_i \prec_c \{d_1, d_2\}$ for $i = 1, 2$.
   - otherwise assume w.l.o.g. that d_1 is a descendant of d in some tree interpretation $\mathcal{I}_a$ hooked to some $a^{\mathcal{I}}$ with $a \in \mathsf{ind}(\mathcal{A})$ in $\mathcal{I}$. Then let $\mathcal{I}'$ be the result of removing from $\mathcal{I}$ all descendants of d_1 and let $\mathcal{J}$ be the quotient of $\mathcal{I}'$ by $\sim$, where $\sim$ is the smallest equivalence relation on $\Delta^{\mathcal{I}'}$ with $d_1 \sim d_2$. We set $d_2 \prec \{d_1, d_2\}$ and $d_i \prec_c \{d_1, d_2\}$ for $i = 1, 2$.
5. if $d \in C^{\mathcal{I}}$ and $C \sqsubseteq (\geqslant 1\, r\, E) \in \mathcal{T}_{\text{horn}}$ then take fresh $e_1, \dots, e_n$ with n the largest number occurring in number restrictions in $\mathcal{T}$ and define $\mathcal{J}$ by adding (d, e_i) to $r^{\mathcal{I}}$ and e_i to $F^{\mathcal{I}}$ for all concept names F which are conjuncts of E and $1 \leq i \leq n$. If this rule has been applied to d for $(\geqslant 1\, r\, E)$, then it is not applied again to any d' with $d \prec^* d'$, where $\prec^*$ is the transitive reflexive closure of the relation $\prec$ introduced in the previous rule. The elements $e_1, \dots, e_n$ introduced by this rule are called the *witnesses for* $(\geqslant 1\, r\, E)$ *at* d.

The interpretation $\mathcal{U}$ is the limit of the sequence $\mathcal{I}_0, \mathcal{I}_1, \dots$ obtained from the interpretation $\mathcal{I}_0 := \mathcal{I}_{\mathcal{A}}$ corresponding to $\mathcal{A}$ by applying the rules in $R_{\mathcal{T}}$. We assume that the rules are applied in a fair way: if a rule is applicable then it is eventually applied. Observe that it follows from the condition that Rule 5 is never again applied to a $\prec^*$-successor of a node to which it has been applied that for every n there exists an m such that the restrictions of $\mathcal{I}_m, \mathcal{I}_{m+1}, \dots$ to nodes of depth $\leq n$ in the tree interpretations hooked to the $a^{\mathcal{I}_m}$, $a \in \mathsf{ind}(\mathcal{A})$, coincide. Thus, the limit interpretation is well defined. Using the fact that we do not make the UNA it is straightforward to prove that $\mathcal{U}$ is a CQ-materialization of $\mathcal{A}$ and $\mathcal{T}_{\text{horn}}$.

Our aim now is to prove that $\mathcal{U}$ is a model of $\mathcal{T}$. We proceed in two steps. Let $d \in \Delta^{\mathcal{U}}$ and assume that the chase has introduced $e_1, \dots, e_n$ as witnesses for $(\geqslant 1\, r\, E)$ in $\mathcal{I}_m$ at some $d_0 \prec^* d$. Assume that $E = \mathsf{pos}(F)$ for some $(\geqslant m\, r\, F)$. We say that these witnesses have been *invalidated for* $(\geqslant m\, r\, F)$ if

1. F contains a negative literal and there exists $e \in \Delta^{\mathcal{U}}$ and $1 \leq i \leq n$ with $e_i \prec_c^* e$ such that $e \notin F^{\mathcal{U}}$ or
2. $m \geq 2$ and there are distinct e_i, e_j such that there exists $e \in \Delta^{\mathcal{U}}$ with $e_i \prec_c^* e$ and $e_j \prec_c^* e$.

For an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$ we denote by $\mathsf{postp}_{\mathcal{U}}^{\mathcal{T}}(d)$ the set of all $F \in L_{\mathcal{T}}$ such that $d \in F^{\mathcal{I}}$.

Claim 1. Let $d \in \Delta^{\mathcal{U}}$ and assume that witnesses $e_1, \dots, e_n$ for $(\geqslant 1\, r\, E)$ in $\mathcal{I}_m$ at some $d_0 \prec^* d$ have been invalidated for $(\geqslant m\, r\, F)$. Then $\mathcal{T}_{\text{horn}} \models C_S^{\leq 1} \sqsubseteq \neg(\geqslant m\, r\, F)$ for $S = \mathsf{postp}_{\mathcal{U}}^{\mathcal{T}}(d)$.

For the proof of Claim 1, assume first that there exists $a \in \mathsf{ind}(\mathcal{A})$ with $a^{\mathcal{I}} \prec^* d$. Denote by $\mathcal{A}_{S^{\leq 1}}$ the ABox with root ρ corresponding to $C_S^{\leq 1}$. Denote by $\mathcal{A}_{S^{\leq 1}}^{(\geq mr.F)}$ the extension of $\mathcal{A}_{S^{\leq 1}}$ with fresh individuals $e_1, \dots, e_m$ and the assertions $r(\rho, e_i)$ and $A(e_i)$ for all $1 \leq i \leq m$ and all concept names A in E. We apply the chase procedure for $\mathcal{T}_{\text{horn}}$ to $\mathcal{A}_{S^{\leq 1}}^{(\geq mr.F)}$ and obtain a CQ-materialization $\mathcal{U}'$ of $\mathcal{A}_{S^{\leq 1}}^{(\geq mr.F)}$ and $\mathcal{T}_{\text{horn}}$. It suffices to show that

1. either F contains a negative literal and there exists $e \in \Delta^{\mathcal{U}'}$ and $1 \leq i \leq m$ with $e_i \prec_c^* e$ such that $e \notin F^{\mathcal{U}'}$ or
2. $m \geq 2$ and there are distinct e_i, e_j such that there exists $e \in \Delta^{\mathcal{U}'}$ with $e_i \prec_c^* e$ and $e_j \prec_c^* e$.

But this can be proved using the fact that the CIs of $\mathcal{T}_{\text{horn}}$ have depth 1 and the condition that witnesses $e_1, \dots, e_n$ for $(\geqslant 1\, r\, E)$ in $\mathcal{I}_m$ at some $d_0 \prec^* d$ have been invalidated for $(\geqslant m\, r\, F)$ in the chase applied to $\mathcal{A}$. The proof for the case that there is no $a \in \mathsf{ind}(\mathcal{A})$ with $a^{\mathcal{I}} \prec^* d$ is similar.

Claim 2. $\mathcal{U}$ is a model of $\mathcal{T}$.

$\mathcal{U}$ satisfies all RIs in $\mathcal{T}$ by Rule 1. Now let $C \sqsubseteq D \in \mathcal{T}$ and $d \in C^{\mathcal{U}}$. We show that $d \in D^{\mathcal{U}}$. For a proof by contradiction assume that this is not the case. Then the set $S = \mathsf{postp}_{\mathcal{U}}^{\mathcal{T}}(d)$ is a trigger for $C \sqsubseteq D$. By (2.) there exists $\alpha \in \text{Horn}(C \sqsubseteq D, S)$. We obtain that at least one of the following holds:

1. there is a concept name A in D such that $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq A$. Then $C_S^{\leq 1} \sqsubseteq A \in \mathcal{T}_{\text{horn}}$ and so $d \in A^{\mathcal{U}}$ by Rule 2. Thus $d \in D^{\mathcal{U}}$ and we have derived a contradiction.
2. there is a universal restriction $R = \forall r.(A_1 \sqcap \cdots \sqcap A_n \rightarrow A)$ which is a Horn restriction of some $(\leqslant m\, r\, E)$ in D such that $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq R$. Then $C_S^{\leq 1} \sqsubseteq R \in \mathcal{T}_{\text{horn}}$ and so $d \in R^{\mathcal{U}}$ by Rule 3. Thus $d \in D^{\mathcal{U}}$ and we have derived a contradiction.
3. there is an $(\geqslant m\, r\, E)$ in D such that $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq (\geqslant 1\, r\, \mathsf{pos}(E))$ and $\mathcal{T} \not\models C_S^{\leq 1} \sqsubseteq \neg(\geqslant m\, r\, E)$. By Rule 5, there exists m and $d' \prec^* d$ in $\Delta^{\mathcal{I}_m}$ with $d' \in (C_S^{\leq 1})^{\mathcal{I}_m}$ such that there are distinct witnesses $e_1, \dots, e_m \in \Delta^{\mathcal{I}_m}$ for $(\geqslant 1\, r\, \mathsf{pos}(E))$ at d'. By Claim 1, the witnesses $e_1, \dots, e_m$ are not invalidated for $(\geqslant m\, r\, E)$ at d. Thus there are at least m distinct $e_1', \dots, e_m' \in \Delta^{\mathcal{U}}$ such that $(d, e_i') \in r^{\mathcal{U}}$ and $e_i' \in E^{\mathcal{U}}$ for $1 \leq i \leq m$. Thus $d \in (\geqslant m\, r\, E)^{\mathcal{U}}$. Hence $d \in D^{\mathcal{U}}$ and we have derived a contradiction.
4. there is a concept $(\geqslant n\, r\, E)$ in S with $n \geq 2$ such that $\mathcal{T} \models C_S^{\leq 1} \sqsubseteq (\leqslant 1\, r\, E)$. But then $\mathcal{U}$ is not a model of $\mathcal{T}_{\text{horn}}$ and we have derived a contradiction.

This finishes the proof of Lemma 4. □

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

We start with the rather straightforward lower bound proof (which goes through for $\mathcal{ALC}$ TBoxes of depth 1 already). We give a polynomial reduction of the unsatisfiability problem for $\mathcal{ALC}$ TBoxes of depth 1 which is known to be EXPTIME-hard. Given an $\mathcal{ALC}$ TBox $\mathcal{T}$ of depth 1, let $\mathcal{T}' = \mathcal{T} \cup \{\top \sqsubseteq B_1 \sqcup B_2\}$, where B_1, B_2 are fresh concept names. Then $\mathcal{T}$ is not satisfiable iff $\mathcal{T}'$ has the CQ-disjunction property iff $\mathcal{T}$ is CQ-Horn-rewritable.

The proof of the EXPTIME upper bound proceeds through a series of lemmas. The algorithm decides CQ-materializability (recall that CQ-materializability is equivalent to the CQ-disjunction property which is equivalent to being CQ-Horn-rewritable, for $\mathcal{ALCHIQ}$ TBoxes of depth 1). The first insight underlying the proof of the EXPTIME upper bound is that for TBoxes in $\mathcal{ALCHIQ}$ of depth 1 CQ-materializability already follows from the existence of CQ-materializations for tree ABoxes of depth 1. We start by introducing the basic notions used in the proof. An ABox $\mathcal{A}$ is $\mathcal{T}$*-saturated* if

- $\mathcal{T}, \mathcal{A} \models (a = b)$ implies $a = b$;
- $\mathcal{T} \models A(a)$ implies $A(a) \in \mathcal{A}$;
- $\mathcal{T}, \mathcal{A} \models r(a, b)$ implies $r(a, b) \in \mathcal{A}$.

For every $\mathcal{T}$ and ABox $\mathcal{A}$, let $\sim_{\mathcal{T}}$ be the equivalence relation on $\mathsf{ind}(\mathcal{A})$ defined by setting $a \sim_{\mathcal{T}} b$ if $\mathcal{T}, \mathcal{A} \models (a = b)$. Let $a/_{\sim_{\mathcal{T}}}$ denote the equivalence class of a w.r.t. $\sim_{\mathcal{T}}$. We regard the elements of $\{a/_{\sim_{\mathcal{T}}} \mid a \in \mathsf{ind}(\mathcal{A})\}$ as individual names. Define a new ABox $\mathcal{A}^{\mathcal{T}}$, the $\mathcal{T}$*-saturation of* $\mathcal{A}$, by setting

- $r(a/_{\sim_{\mathcal{T}}}, b/_{\sim_{\mathcal{T}}}) \in \mathcal{A}^{\mathcal{T}}$ iff there are $a' \in a/_{\sim_{\mathcal{T}}}$ and $b' \in b/_{\sim_{\mathcal{T}}}$ such that $\mathcal{T}, \mathcal{A} \models r(a', b')$;
- $A(a/_{\sim_{\mathcal{T}}}) \in \mathcal{A}^{\mathcal{T}}$ iff there exists $a' \in a/_{\sim_{\mathcal{T}}}$ with $\mathcal{T}, \mathcal{A} \models A(a')$.

The mapping $h : \mathcal{A} \rightarrow \mathcal{A}^{\mathcal{T}}$ mapping a to $a/_{\sim_{\mathcal{T}}}$ is an isomorphism iff $\mathcal{A}$ is $\mathcal{T}$-saturated. The following lemma lists the basic properties of $\mathcal{T}$-saturated ABoxes.

Lemma 8. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox. Let $\mathcal{A} \subseteq \mathcal{A}'$ be ABoxes with $\mathcal{A}'_{|\mathsf{ind}(\mathcal{A})} = \mathcal{A}$. Assume $\mathcal{A}'$ is satisfiable w.r.t. $\mathcal{T}$. Then the following hold:*

(a) There exists a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ iff there exists a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}^{\mathcal{T}}$;

(b) If $\mathcal{I}$ is a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ and $\mathcal{A}$ is $\mathcal{T}$-saturated, then the restriction of $\mathcal{I}$ to $\{a^{\mathcal{I}} \mid a \in \mathsf{ind}(\mathcal{A})\}$ coincides (up to isomorphism) with $\mathcal{I}_{\mathcal{A}}$;

(c) If $\mathcal{A}'$ is $\mathcal{T}$-saturated, then $\mathcal{A}$ is $\mathcal{T}$-saturated.

We introduce more notation. The *outdegree* of a tree interpretation $\mathcal{I}$ is the outdegree of $G_{\mathcal{I}}$. An ABox $\mathcal{A}$ is a *tree ABox* if $\mathcal{I}_{\mathcal{A}}$ is a tree interpretation. The *outdegree* of a tree ABox $\mathcal{A}$ is the outdegree of $\mathcal{I}_{\mathcal{A}}$. For any interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, denote by $\mathcal{I}_d^{\leq 1}$ the *one-unfolding* of $\mathcal{I}$ at $d \in \Delta^{\mathcal{I}}$. It is the standard unfolding at d, but only up to depth 1. In detail, the domain $\Delta^{\mathcal{I}_d^{\leq 1}}$ of $\mathcal{I}_d^{\leq 1}$ contains d and all pairs (d, d') such that there exists a role r with $(d, d') \in r^{\mathcal{I}}$. Then set

$$\begin{aligned} A^{\mathcal{I}_d^{\leq 1}} &= \{d \mid d \in A^{\mathcal{I}}\} \cup \{(d, d') \mid d' \in A^{\mathcal{I}}\} \\ r^{\mathcal{I}_d^{\leq 1}} &= \{(d, (d, d')) \mid (d, d') \in r^{\mathcal{I}}\} \text{ for all roles } r. \end{aligned}$$

If $\mathcal{I}$ is a tree interpretation already, then we can (and will) identify $\mathcal{I}_d^{\leq 1}$ with the restriction $\mathcal{I}_{|X}$ of $\mathcal{I}$ to X, where X contains d and all d' such that there exists a role r with $(d, d') \in r^{\mathcal{I}}$. We then call $\mathcal{I}_d^{\leq 1}$ the *1-neighborhood* of $\mathcal{I}$ at d. We say that $\mathcal{I}$ a *bouquet with root* d if $\mathcal{I}_d^{\leq 1} = \mathcal{I}$. We use the same notation for ABoxes, in the obvious way. Thus, an ABox $\mathcal{A}$ is a *bouquet with root* a_0 if there exists an individual name a_0 such that all role assertions of $\mathcal{A}$ take the form $r(a_0, a)$ where $a \neq a_0$ and r is a role. Clearly, the outdegree $\mathcal{A}$ is then cardinality of $\{a \mid a_0 \neq a \in \mathsf{ind}(\mathcal{A})\}$.

Lemma 9. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox of depth 1. Then $\mathcal{T}$ is CQ-materializable if $\mathcal{T}$ is CQ-materializable for all bouquet ABoxes.*

Proof. Assume that $\mathcal{T}$ is CQ-materializable for all bouquet ABoxes. Let $\mathcal{A}$ be an ABox satisfiable w.r.t. $\mathcal{T}$. We show that a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$ exists. By Lemma 8 (Point (a)), it suffices to prove that $\mathcal{T}$ and $\mathcal{A}^{\mathcal{T}}$ have a CQ-materialization. Thus, we may assume that $\mathcal{A}$ is $\mathcal{T}$-saturated. Let for any $a \in \mathsf{ind}(\mathcal{A})$, $\mathcal{A}_a^{\leq 1}$ be the one-unfolding of $\mathcal{A}$ at a. Let $\mathcal{I}_a$ be a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}_a^{\leq 1}$ ($\mathcal{I}_a$ exists since $\mathcal{A}_a^{\leq 1}$ is a bouquet ABox). We may assume that $\mathcal{I}_a$ is a forest model of $\mathcal{T}$ and $\mathcal{A}_a^{\leq 1}$. Using Lemma 8 (Point (c)) and a straightforward unfolding argument, one can show that $\mathcal{A}_a^{\leq 1}$ is $\mathcal{T}$-saturated. Thus, by Point (b), the restriction of $\mathcal{I}_a$ to $\{b^{\mathcal{I}_a} \mid b \in \mathsf{ind}(\mathcal{A}_a^{\leq 1}\}$ coincides (up to isomorphism) with $\mathcal{I}_{\mathcal{A}_a^{\leq 1}}$. Define a model $\mathcal{I}$ of $\mathcal{A}$ by

- hooking to $\mathcal{A}$ at every $a \in \mathsf{Ind}(\mathcal{A})$ the interpretation hooked to $\mathcal{A}_a^{\leq 1}$ at $a^{\mathcal{I}_a}$ in $\mathcal{I}_a$ and
- adding (a, b) to $r^{\mathcal{I}}$ for all $r(a, b) \in \mathcal{A}$ and
- setting $a^{\mathcal{I}} = a$ for $a \in \mathsf{ind}(\mathcal{A})$.

Using the condition that $\mathcal{T}$ has depth 1 it is straightforward to prove that $\mathcal{I}$ is CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. □

The signature $\mathsf{sig}(\mathcal{A})$ of an ABox $\mathcal{A}$ is the set of concept and role names occurring in $\mathcal{A}$. The signature $\mathsf{sig}(\mathcal{T})$ of a TBox $\mathcal{T}$ is the set of concept and role names occurring in $\mathcal{T}$.

Lemma 10. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox of depth 1. Then $\mathcal{T}$ is CQ-materializable iff $\mathcal{T}$ is CQ-materializable for bouquet ABoxes $\mathcal{A}$ of outdegree $\leq |\mathcal{T}|$ satisfying $sig(\mathcal{A}) \subseteq sig(\mathcal{T})$.*

Proof. Assume $\mathcal{T}$ is given. Let $\mathcal{A}$ be a bouquet ABox with root a_0 of minimal outdegree such that there is no CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. We show that the outdegree of $\mathcal{A}$ does not exceed $|\mathcal{T}|$. We may assume that the outdegree of $\mathcal{T}$ is at least three (otherwise we are done). By Lemma 8 (Point (a)), we may assume that $\mathcal{A}$ is $\mathcal{T}$-saturated. Take for any subconcept D of the form $(\geq nr.C)$ or $(\leq nr.C)$ occurring in a concept in $\mathcal{T}$ the set

$$S_D = \{b \neq a_0 \mid \mathcal{I}_{\mathcal{A}} \models r(a_0, b) \wedge C(b)\}$$

Let $S'_D = S_D$ if $|S_D| \leq n + 1$; otherwise let S'_D be a subset of S_D of cardinality $n + 1$. Let $\mathcal{A}'$ be the restriction $\mathcal{A}_{|S}$ of $\mathcal{A}$ to the union S of all S'_D and $\{a\}$. We show that there exists no CQ-materialization of $\mathcal{T}$ and $\mathcal{A}'$. Assume for a proof by contradiction that there is a CQ-materialization $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}'$. $\mathcal{A}'$ is $\mathcal{T}$-saturated and so its restriction to $\{a^{\mathcal{I}} \mid a \in \mathsf{ind}(\mathcal{A}')\}$ coincides with $\mathcal{I}_{\mathcal{A}'}$. We may assume that $a^{\mathcal{I}} = a^{\mathcal{I}_{\mathcal{A}}} = a$ for all $a \in \mathsf{Ind}(\mathcal{A}')$. Define an interpretation $\mathcal{I}'$ by

- taking the union of $\mathcal{I}_{\mathcal{A}}$ and $\mathcal{I}$ and
- hooking to the resulting interpretation at every b with $b \in \mathsf{ind}(\mathcal{A}) \setminus \mathsf{ind}(\mathcal{A}')$ the model $\mathcal{I}_b$ hooked to $\mathcal{A}_{|\{a_0, b\}}$ at b in a forest model CQ-materialization of $\mathcal{T}$ and $\mathcal{A}_{|\{a_0, b\}}$.

One can show that $\mathcal{I}'$ is a CQ-materialization of $\mathcal{A}$ and $\mathcal{T}$ (and thus derive a contradiction): using the condition that $\mathcal{A}$ is $\mathcal{T}$-saturated and Points (b) and (c), one can show that the restriction $\mathcal{I}'_{|\mathsf{ind}(\mathcal{A})}$ of $\mathcal{I}'$ to $\mathsf{ind}(\mathcal{A})$ coincides with $\mathcal{A}$. Using the condition that $\mathcal{T}$ has depth 1 it is easy to show that $\mathcal{I}'$ is a model of $\mathcal{T}$. It is a CQ-materialization of $\mathcal{A}$ and $\mathcal{T}$ as it consists of CQ-materializations of sub-ABoxes of $\mathcal{A}$ and $\mathcal{T}$. This finishes the proof. □

Let $\mathcal{A}$ be a $\mathcal{T}$-saturated bouquet ABox with root a_0. A forest model $\mathcal{I}$ of $\mathcal{A}$ is a *1-materialization of $\mathcal{T}$ and $\mathcal{A}$ with root* a_0 if it is a bouquet with root $a_0^{\mathcal{I}} = a_0$ and

1. there exists a forest model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ such that $\mathcal{I} = \mathcal{J}_{a_0}^{\leq 1}$;
2. for any model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ there exists a homomorphism from $\mathcal{I}$ to $\mathcal{J}$.

An ABox $\mathcal{A}$ is called $\mathcal{T}$*-simple* if it is a bouquet ABox, satisfiable w.r.t. $\mathcal{T}$, $\mathcal{T}$-saturated, of outdegree at most $|\mathcal{T}|$, and satisfies $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$. We show that when checking CQ-materializability of $\mathcal{ALCHIQ}$ TBoxes of depth 1, not only is it sufficient to consider $\mathcal{T}$-simple bouquet ABoxes instead of unrestricted bouquet ABoxes, but additionally one can concentrate on 1-materializations of such bouquet ABoxes.

Lemma 11. *Let $\mathcal{T}$ be an $\mathcal{ALCHIQ}$ TBox of depth 1. Then $\mathcal{T}$ is CQ-materializable iff for all $\mathcal{T}$-simple bouquet ABoxes $\mathcal{A}$ there is a 1-materialization of $\mathcal{T}$ and $\mathcal{A}$.*

Proof. Let $\mathcal{A}$ be a $\mathcal{T}$-simple bouquet ABox with root a_0. Assume that for all $\mathcal{T}$-simple bouquet ABoxes $\mathcal{B}$ with root b there exists a 1-materialization $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{B}$ with root b. Call such a triple $(\mathcal{B}, b, \mathcal{J})$ a *1-materializability witness*. It suffices to prove that there exists a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. We construct the desired CQ-materialization step-by-step using these 1-materializability witnesses and also memorizing sets of frontier elements that have to be expanded in the next step. We start with a 1-materializability witness $(\mathcal{A}, a_0, \mathcal{I}^0)$ and set $F_0 = \Delta^{\mathcal{I}^0} \setminus \{a_0^{\mathcal{I}^0}\}$. Then we construct a sequence of tree interpretations $\mathcal{I}^0 \subseteq \mathcal{I}^1 \subseteq \ldots$ and frontier sets $F_{i+1} \subseteq \Delta^{\mathcal{I}^{i+1}} \setminus \Delta^{\mathcal{I}^i}$ inductively as follows: given $\mathcal{I}^i$ and F_i, take for any $b \in F_i$ its predecessor b' in $\mathcal{I}^i$ and a 1-materializability witness $(\mathcal{I}^i_{|\{b',b\}}, b, \mathcal{I}_b)$ and set

$$\mathcal{I}^{i+1} := \mathcal{I}^i \cup \bigcup_{b \in F_i} \mathcal{I}_b \qquad F_{i+1} := \bigcup_{b \in F_i} \Delta^{\mathcal{I}_b} \setminus \{b\}$$

Let $\mathcal{I}^* := \bigcup_{i \geq 0} \mathcal{I}^i$. We show that $\mathcal{I}^*$ is a CQ-materialization of $\mathcal{T}$ and $\mathcal{A}$. $\mathcal{I}^*$ is a model of $\mathcal{T}$ by construction since $\mathcal{T}$ is an $\mathcal{ALCHIQ}$ TBox of depth 1.

We show that for every model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ there exists a homomorphism from $\mathcal{I}^*$ to $\mathcal{J}$. Consider a model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$. We may assume that $\mathcal{J}$ is a forest model and irreflexive in the sense that $(d, d') \in r^{\mathcal{J}}$ implies $d \neq d'$ for all $d, d' \in \Delta^{\mathcal{J}}$. We construct h as the limit of a sequence $h_0 \subseteq h_1 \subseteq \ldots$ of homomorphisms h_i from $\mathcal{I}^i$ to $\mathcal{J}$. By definition, there exists a homomorphism h_0 from $\mathcal{I}^0$ to $\mathcal{J}^{\leq 1}_{a_0}$ mapping $a^{\mathcal{I}^0}$ to $a^{\mathcal{J}}$ for every $a \in \mathsf{ind}(\mathcal{A})$. Now, inductively, assume that h_i is a homomorphism from $\mathcal{I}^i$ to $\mathcal{J}$. Assume c has been added to $\mathcal{I}^i$ in the construction of $\mathcal{I}^{i+1}$. Then there exists $b \in F_i$ and its predecessor b' in $\mathcal{I}^i$ such that $c \in \Delta^{\mathcal{I}_b} \setminus \{b\}$, where $\mathcal{I}_b$ is the irreflexive tree interpretation that has been added to $\mathcal{I}^i$ as the last component of the 1-materializability witness $(\mathcal{I}^i_{|\{b',b\}}, b, \mathcal{I}_b)$. But then, as $\mathcal{I}_b$ is a 1-materialization of $\mathcal{I}^i_{|\{b',b\}}$ and h_i is injective on $\mathcal{I}^i_{|\{b',b\}}$ (since $\mathcal{J}$ is irreflexive), we can expand the homomorphism h_i to a homomorphism from $\Delta^{\mathcal{I}^i} \cup \{c\}$ into $\mathcal{J}$. Thus, we can expand h_i to a homomorphism from $\mathcal{I}^{i+1}$ to $\mathcal{J}$. $\square$

Lemma 11 implies that an $\mathcal{ALCHIQ}$ TBox $\mathcal{T}$ of depth 1 enjoys CQ-Horn-rewritability iff for all $\mathcal{T}$-simple bouquet ABoxes $\mathcal{A}$ there exists a 1-materialization of $\mathcal{T}$ and $\mathcal{A}$. The latter condition can be checked in deterministic exponential time using satisfiability checks for ABoxes w.r.t. $\mathcal{ALCHIQ}$ TBoxes.

Observe that if $\mathcal{T}$ is a $\mathcal{ALCHIF}$ TBox in normal form, then the TBox $\mathcal{T}_{\text{horn}}$ is easily transformed into an equivalent Horn-$\mathcal{ALCHIF}$ TBox: the only CIs which are not already Horn-$\mathcal{ALCHIF}$ CIs take the form $C^{\leq 1}_{S} \sqsubseteq (\leqslant 1\, r\, E)$ with $(\geqslant m\, r\, E) \in S$. But if $\mathcal{T} \not\models C^{\leq 1}_{S} \sqsubseteq \bot$, then $\mathcal{T} \models C^{\leq 1}_{S} \sqsubseteq (\leqslant 1\, r\, E)$ can only hold for an $\mathcal{ALCHIF}$ TBox $\mathcal{T}$ if $\mathcal{T} \models \top \sqsubseteq (\leqslant 1\, r\, \top)$ and thus we can replace $C^{\leq 1}_{S} \sqsubseteq (\leqslant 1\, r\, E)$ by $\top \sqsubseteq (\leqslant 1\, r\, \top)$ in $\mathcal{T}_{\text{horn}}$.

E Proofs for Section 5

We first verify the properties claimed for the TBox $\mathcal{T}$ constructed in Example 4. Recall that the TBox $\mathcal{T}$ states that role names s_1 and s_2 are functional and contains the RIs $r \sqsubseteq s_1$ and $r \sqsubseteq s_2$ and the CIs

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

We show that $\mathcal{T}$ has the CQ-disjunction with UNA by constructing for any ABox $\mathcal{A}$ satisfiable w.r.t. $\mathcal{T}$ with UNA a CQ-materialization $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ with UNA. To construct the CQ-materialization one has to distinguish two kinds of individual names a. Informally,

- if a has distinct s_1- and s_2-successors in $\mathcal{A}$, then because of the UNA a having an r-successor contradicts the role inclusions and s_1, s_2 being functional and thus $B(a)$ is entailed and added to the extension of B;
- if a has a common s_1- and s_2-successor, then $\exists r.\top$ is entailed at a and we add an r-successor of a.

Formally, we construct the CQ-materialization $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ as follows. We start with $\mathcal{I} := \mathcal{I}_{\mathcal{A}}$. Whenever an individual name a has both an s_1-successor and an s_2-successor in $\mathcal{I}$, then add all its s_i-successors, $i \in \{1, 2\}$ to $B_i^{\mathcal{I}}$. Next, for any a that has an $s_1^{\mathcal{I}}$-successor b in $(B_1 \sqcap B_2)^{\mathcal{I}}$, add (a, b) to $s_1^{\mathcal{I}}$ and (a, b) to $s_1^{\mathcal{I}}$. If $s_1^{\mathcal{I}}$ or $s_2^{\mathcal{I}}$ are not functional, then $\mathcal{A}$ is not satisfiable w.r.t. $\mathcal{T}$ with UNA. Otherwise, a CQ-materialization is obtained by adding a to $B^{\mathcal{I}}$ whenever a has distinct s_1- and s_2-successors in $\mathcal{I}$.

Theorem 5 (restated). *Let $\mathcal{T}$ be a $\mathcal{ALCHIF}^{\sqsubseteq f}$ 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 $\mathsf{func}(r) \in \mathcal{T}$, then $a' \sim b'$. We 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 $\mathsf{ind}(\mathcal{A})$.

Clearly, $a \sim b$ implies $a^{\mathcal{I}} = b^{\mathcal{I}}$ for every model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$. Thus, it suffices to construct for every model $\mathcal{I}$ of $\mathcal{A}$ a model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}/_{\sim}$ with UNA such that there is a homomorphism from $\mathcal{J}$ to $\mathcal{I}$ mapping the equivalence class $a/_{\sim}$ of $a \in \mathsf{ind}(\mathcal{A})$ w.r.t. $\sim$ to $a^{\mathcal{I}}$. For the construction of $\mathcal{J}$ assume $a \in \mathsf{ind}(\mathcal{A})$ is given. The *$\mathcal{I}$-unfolding* $\mathcal{I}_a$ of $\mathcal{I}$ at $a^{\mathcal{I}}$ w.r.t. $\mathcal{T}$ is defined as follows. The domain $\Delta^{\mathcal{I}_a}$ of $\mathcal{J}$ consists of all words $d_0 r_1 \ldots r_n d_n$ with $n \geq 0$, $a^{\mathcal{I}} = d_0$, each d_i from $\Delta^{\mathcal{I}}$ and each r_i a role such that

(a) $(d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}}$ for $0 \leq i < n$;

(b) if $\top \sqsubseteq (\leqslant 1\, r_{i+1}\, \top) \in \mathcal{T}$ and $r_i^- = r_{i+1}$, then $d_{i-1} \neq d_{i+1}$ for $0 < i < n$.

(c) if $\top \sqsubseteq (\leqslant 1\, r_1\, \top) \in \mathcal{T}$, then there does not exist an $a' \sim a$ and b such that $r(a', b) \in \mathcal{A}$.

For $d_0 \cdots d_n \in \Delta^{\mathcal{I}_a}$, we set $\mathsf{tail}(d_0 \cdots d_n) = d_n$ and let $w \in A^{\mathcal{I}_a}$ if $\mathsf{tail}(w) \in A^{\mathcal{I}}$ and $(w, wrd) \in s^{\mathcal{I}_a}$ if $\mathcal{T} \models r \sqsubseteq s$ (where s, r are roles).

Now define $\mathcal{J}$ by setting $a^{\mathcal{J}} = a/_\sim$ for all $a \in \mathsf{ind}(\mathcal{A})$ and $(a/_\sim, b/_\sim) \in r^{\mathcal{J}}$ if there exist $a' \sim a$, $b' \sim b$, and a role s such that $\mathcal{T} \models s \sqsubseteq r$ and $s(a', b') \in \mathcal{A}$ and by then hooking the interpretation $\mathcal{I}_a$ to $a/_\sim$ by identifying $a/_\sim$ and the root $a^{\mathcal{I}}$ of $\mathcal{I}_a$. Using the condition that no functional role occurs in the right-hand side of an RI in $\mathcal{T}$ one can easily show that $\mathcal{J}$ is a model of $\mathcal{T}$ and $\mathcal{A}$. Moreover, by construction there is a homomorphism from $\mathcal{J}$ to $\mathcal{I}$ mapping every $a/_\sim$ to $a^{\mathcal{I}}$. □

F Proofs for Section 6

We require the direct product of two interpretations, see [Chang and Keisler, 1990; Lutz *et al.*, 2011]. Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be interpretations. Then the *direct product* $\mathcal{I}_1 \times \mathcal{I}_2$ of $\mathcal{I}_1$ and $\mathcal{I}_2$ is defined by setting

$$\begin{aligned} \Delta^{\mathcal{I}_1 \times \mathcal{I}_2} &= \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2} \\ A^{\mathcal{I}_1 \times \mathcal{I}_2} &= \{(d_1, d_2) \mid d_1 \in A^{\mathcal{I}_1}, d_2 \in A^{\mathcal{I}_2}\} \\ r^{\mathcal{I}_1 \times \mathcal{I}_2} &= \{((d_1, d_2), (e_1, e_2)) \mid (d_1, e_1) \in r^{\mathcal{I}_1}, \\ &\qquad (d_2, e_2) \in r^{\mathcal{I}_2}\} \\ a^{\mathcal{I}_1 \times \mathcal{I}_2} &= (a^{\mathcal{I}_1}, a^{\mathcal{I}_2}) \end{aligned}$$

It is well know (and easy to see) that if $\mathcal{T}'$ is a Horn TBox and a conservative extension of a TBox $\mathcal{T}$, then $\mathcal{T}$ is preserved under direct products; that is, if $\mathcal{I}_1$ and $\mathcal{I}_2$ are models of $\mathcal{T}$, then $\mathcal{I}_1 \times \mathcal{I}_2$ is a model of $\mathcal{T}$. Recall that $\mathcal{T}_1$ contains a single CI:

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

Lemma 12. *$\mathcal{T}_1$ is not preserved under direct products.*

Proof. Consider the interpretation $\mathcal{I}_1$ with

- $\Delta^{\mathcal{I}_1} = \{a, b\}$;
- $\mathsf{author}^{\mathcal{I}_1} = \{(a, b)\}$;
- $\mathsf{Novel}^{\mathcal{I}_1} = \{b\}$;
- $\mathsf{Sort\_Story}^{\mathcal{I}_1} = \emptyset$;
- $\mathsf{Fiction}^{\mathcal{I}_1} = \{b\}$.

$\mathcal{I}_1$ is a model of $\mathcal{T}_1$. Consider the interpretation $\mathcal{I}_2$ defined by

- $\Delta^{\mathcal{I}_2} = \{a, b\}$;
- $\mathsf{author}^{\mathcal{I}_2} = \{(a, b)\}$;
- $\mathsf{Novel}^{\mathcal{I}_2} = \emptyset$;
- $\mathsf{Sort\_Story}^{\mathcal{I}_2} = \{b\}$;
- $\mathsf{Fiction}^{\mathcal{I}_2} = \{b\}$

Then $\mathcal{I}_2$ is a model of $\mathcal{T}_1$. It is easy to see that $\mathcal{I}_1 \times \mathcal{I}_2$ is not a model of $\mathcal{T}_1$. □

Theorem 8 (restated). *For $\mathcal{ALC}$ TBoxes of depth 1 there is not algorithm that decides CQ-Horn-rewritability for (Σ_1, Σ_2) and outputs such a rewriting in case it exists.*

Proof. In [Botoeva *et al.*, 2016b], the authors construct a sequence of $\mathcal{ALC}$ TBoxes $\mathcal{T}_i$ and $\mathcal{EL}$ TBoxes $\mathcal{T}_i'$ such that it is undecidable whether $\mathcal{T}_i$ and $\mathcal{T}_i'$ are CQ-inseparable w.r.t. a signature pair (Σ_1^i, Σ_2^i), $i \in \mathbb{N}$. Moreover, it is shown that for TBoxes in Horn-$\mathcal{ALC}$ CQ-inseparability w.r.t. signature pairs is decidable. Now assume for a proof by contradiction that the theorem is false: we show that one can then decide whether $\mathcal{T}_i$ and $\mathcal{T}_i'$ are CQ-inseparable w.r.t. (Σ_1^i, Σ_2^i). Given $\mathcal{T}_i$, decide whether there is a CQ-Horn-rewriting of $\mathcal{T}_i$ w.r.t. (Σ_1^i, Σ_2^i). If not, output that $\mathcal{T}_i$ and $\mathcal{T}_i'$ are not CQ-inseparable w.r.t. (Σ_1^i, Σ_2^i). If yes, then compute a CQ-Horn rewriting $\mathcal{T}_i''$ of $\mathcal{T}_i$ w.r.t. (Σ_1^i, Σ_2^i). Then check whether $\mathcal{T}_i''$ and $\mathcal{T}_i'$ are CQ-inseparable w.r.t. (Σ_1^i, Σ_2^i). If not, output that $\mathcal{T}_i$ and $\mathcal{T}_i'$ are not CQ-inseparable w.r.t. (Σ_1^i, Σ_2^i). If yes, output that $\mathcal{T}_i$ and $\mathcal{T}_i'$ are CQ-inseparable w.r.t. (Σ_1^i, Σ_2^i). □