Reverse Engineering Queries in Ontology-Enriched Systems: The Case of Expressive Horn Description Logic Ontologies

Víctor Gutiérrez-Basulto
Cardiff University, UK
gutierrezbasultov@cardiff.ac.uk

Jean Christoph Jung
University of Bremen, Germany
KU Leuven, Belgium
jeanjung@uni-bremen.de

Leif Sabellek
University of Bremen, Germany
sabellek@uni-bremen.de

Abstract

We introduce the query-by-example (QBE) paradigm for query answering in the presence of ontologies. Intuitively, QBE permits non-expert users to explore the data by providing examples of the information they (do not) want, which the system then generalizes into a query. Formally, we study the following question: given a knowledge base and sets of positive and negative examples, is there a query that returns all positive but none of the negative examples? We focus on description logic knowledge bases with ontologies formulated in Horn-$\mathcal{ALCI}$ and (unions of) conjunctive queries. Our main contributions are characterizations, algorithms and tight complexity bounds for QBE.

1 Introduction

In recent times, ontology-enriched systems (OES) have risen as a prominent technology for data management. The appeal of OES comes from the fact that the ontology provides rich schema information or background knowledge which enriches the answers of queries. The success of this paradigm has led not only to the development of a vast amount of foundational results, but also of optimized systems used in real-life scenarios, see e.g., [Rodriguez-Muro *et al.*, 2013; Kharlamov *et al.*, 2015; Calvanese *et al.*, 2016; Hovland *et al.*, 2017] and references therein. For instance, the OES Ontop is currently being used to access exploration data generated by the petroleum company Statoil [Kharlamov *et al.*, 2015]. In these OES, users access the data through queries usually formulated in powerful query languages such as conjunctive or path queries. Unfortunately, in real life, casual non-expert users are often not able to specify queries using these formalisms (e.g., Statoil geologists [Hovland *et al.*, 2017]), clearly hampering the usability of OES.

In relational databases (witnessing the same problem), an alternative approach for querying was proposed to alleviate this problem: *query-by-example (QBE)*, where roughly, users give positive and negative examples which the system should reverse-engineer into a query conforming with the examples [Zloof, 1975]. Because of 'big data', this querying paradigm has lately gained new interest since even expert users might find it useful to explore the data in this way. As a result, QBE has been investigated for different query languages and data representations, e.g., conjunctive queries over relational data [Tran *et al.*, 2014; ten Cate and Dalmau, 2015; Bonifati *et al.*, 2016; Barceló and Romero, 2017], SPARQL queries over RDF data [Arenas *et al.*, 2016], and path queries over graph databases [Bonifati *et al.*, 2015].

The goal of this paper is two-fold. First, we aim at initiating research on the QBE approach to querying in the context of ontology-enriched systems. We mainly focus on establishing foundational results for QBE over OES with the ontology formulated in description logics (DLs). Formally, we introduce and study the following problem $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ for an ontology language $\mathcal{L}$ and some query language $\mathcal{Q}$: given an $\mathcal{L}$-knowledge base and sets of positive and negative examples, decide whether there is a query $q \in \mathcal{Q}$ such that all positive examples are certain answers to q over $\mathcal{K}$, and none of the negative is. As query language $\mathcal{Q}$, we consider (unions of) conjunctive queries, (U)CQs. We allow for a restricted signature Σ, which is a common feature in many OES. As a simple example, consider the knowledge base consisting of

$$\mathcal{T} = \{\mathsf{Human} \sqsubseteq \mathsf{Vertebrate}, \mathsf{Vertebrate} \sqsubseteq \exists \mathsf{hasPart}.\mathsf{Spine}\},$$
$$\mathcal{A} = \{\mathsf{Human}(ax), \mathsf{hasPart}(an, sp), \mathsf{Spine}(sp), \mathsf{Bug}(bug)\}.$$

If the positive examples are ax, an and the negative example is bug, then $q(x) = \exists y\, \mathsf{hasPart}(x, y) \wedge \mathsf{Spine}(y)$ is a witness query. However, there is no witnessing query for the positive examples an, bug if ax is to be avoided.

The second aim is to continue bridging the gap between DL and machine learning research. Indeed, QBE over knowledge bases can be viewed as an instantiation of the *inductive logic programming (ILP)* framework [Nienhuys-Cheng and de Wolf, 1997]: the background knowledge is given by a DL knowledge base and the learning goal are single rules for $\mathcal{Q} = \mathrm{CQ}$ and sets of rules with the same head for $\mathcal{Q} = \mathrm{UCQ}$, respectively. In this area, the work closest to ours is perhaps [Kietz, 2002].

Our main contributions are characterizations, algorithms, and complexity bounds for $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ for $\mathcal{L}$ an expressive Horn DL $\mathcal{L} \in \{\text{Horn-}\mathcal{ALCI}, \text{Horn-}\mathcal{ALC}\}$ and $\mathcal{Q} \in \{\mathrm{CQ}, \mathrm{UCQ}\}$. In Section 3, we start with providing natural model-theoretic characterizations for $\mathsf{QBE}(\text{Horn-}\mathcal{ALCI}, \mathcal{Q})$ for $\mathcal{Q} \in \{\mathrm{CQ}, \mathrm{UCQ}\}$ by lifting characterizations known from the relational database setting [ten Cate and Dalmau, 2015] by replacing the database with the universal model of the knowledge base. Unfortunately, our characterizations do not give immediate rise to a decision procedure because the universal model is typically infinite. In Section 4, we exploit the regularity of universal models and provide decision procedures running in 2-ExpTime and coNExpTime for Horn-$\mathcal{ALCI}$ and Horn-$\mathcal{ALC}$, respectively. Having these, we prove matching lower bounds, the most challenging one being a 2-ExpTime-lower bound for $\mathsf{QBE}(\text{Horn-}\mathcal{ALCI}, \mathcal{Q})$, $\mathcal{Q} \in \{\text{CQ}, \text{UCQ}\}$. Interestingly, some results depend on restricting the signature, so we consider also the variant QBE_f of QBE with unrestricted signature. The following table summarizes our results.

$\mathcal{L} \rightarrow$	Horn-$\mathcal{ALCI}$	Horn-$\mathcal{ALC}$
$\mathsf{QBE}(\mathcal{L}, \text{CQ})$	2-ExpTime	coNExpTime
$\mathsf{QBE}(\mathcal{L}, \text{UCQ})$	2-ExpTime	ExpTime
$\mathsf{QBE}_f(\mathcal{L}, \text{CQ})$	2-ExpTime	coNExpTime
$\mathsf{QBE}_f(\mathcal{L}, \text{UCQ})$	ExpTime	ExpTime

We obtain the same results for the variant QDEF of QBE, the problem to decide whether some $q \in \mathcal{Q}$ returns *precisely* the positive examples. In Section 5, we investigate the *size of witness queries*. This is of course vital for practical purposes since at the end the user is interested in obtaining a (witness) query to further explore the data. We particularly show that they can be double exponentially large, which is in contrast to the relational database setting. In Section 6, we discuss related work and lay out directions for future work.

An extended version with appendix can be found under www.informatik.uni-bremen.de/tdki/research/papers.html.

2 Preliminaries

Syntax. We introduce the DL Horn-$\mathcal{ALCI}$ [Krötzsch *et al.*, 2013]. Let $\mathsf{N_C}, \mathsf{N_R}, \mathsf{N_I}$ be infinite disjoint sets of *concept, role,* and *individual names*, respectively. The syntax of *Horn-$\mathcal{ALCI}$ concepts* C, D is given by the grammar:

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

where $A \in \mathsf{N_C}$ and $r \in \{s, s^- \mid s \in \mathsf{N_R}\}$ is a *role*. Concepts of the form B are called *basic concepts* and roles of the form r^- *inverse roles.* We identify r^- with $s \in \mathsf{N_R}$ if $r = s^-$.

A *Horn-$\mathcal{ALCI}$ TBox (ontology)* $\mathcal{T}$ is a finite set of *concept inclusions (CIs)* $B \sqsubseteq C$, with B a basic concept and C a Horn-$\mathcal{ALCI}$ concept. An *ABox* $\mathcal{A}$ is a finite set of *concept* and *role assertions* of the form $A(a)$ and $r(a, b)$, where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $a, b \in \mathsf{N_I}$. We write $\mathsf{ind}(\mathcal{A})$ for the set of individuals in $\mathcal{A}$. A *Horn-$\mathcal{ALCI}$ knowledge base (KB)* $\mathcal{K}$ is a pair $(\mathcal{T}, \mathcal{A})$ of a Horn-$\mathcal{ALCI}$ TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. The fragment Horn-$\mathcal{ALC}$ is obtained by disallowing inverse roles; $\mathcal{ELI}$ is the fragment allowing only concept inclusions $C \sqsubseteq D$ with $C, D ::= \top \mid A \mid C \sqcap D \mid \exists r.C$.

Semantics. The semantics is defined in terms of interpretations $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, consisting of a non-empty *domain* $\Delta^\mathcal{I}$ and an *interpretation function* $\cdot^\mathcal{I}$ mapping concept names to subsets of the domain and role names to binary relations over the domain. Further, we adopt the *standard name assumption*, i.e., $a^\mathcal{I} = a$ for all $a \in \mathsf{N_I}$. The interpretation of complex concepts $C^\mathcal{I}$ is defined in the usual way [Baader *et al.*, 2017]. An interpretation $\mathcal{I}$ is a *model of a TBox* $\mathcal{T}$ if $B^\mathcal{I} \subseteq C^\mathcal{I}$ for all CIs $B \sqsubseteq C \in \mathcal{T}$; and it is a *model of an ABox* $\mathcal{A}$ if $(a, b) \in r^\mathcal{I}$ for all $r(a, b) \in \mathcal{A}$ and $a \in A^\mathcal{I}$ for all $A(a) \in \mathcal{A}$. We call a KB $(\mathcal{T}, \mathcal{A})$ *consistent* if $\mathcal{T}$ and $\mathcal{A}$ have a common model.

Queries. A *conjunctive query (CQ)* is an expression of the form $q(\mathbf{x}) = \exists \mathbf{y}\, \varphi(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are tuples of variables and $\varphi(\mathbf{x}, \mathbf{y})$ is a conjunction of *atoms* of the form $A(v)$ or $r(v, w)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $v, w \in \mathbf{x} \cup \mathbf{y}$. We call $\mathbf{x}$ *answer variables* and $\mathbf{y}$ *quantified variables* of q. A *union of conjunctive queries (UCQ)* is an expression of the form $q(\mathbf{x}) = q_1(\mathbf{x}) \vee \ldots \vee q_n(\mathbf{x})$, where each $q_i(\mathbf{x})$ is a CQ with answer variables $\mathbf{x}$. A *match* of a CQ q in an interpretation $\mathcal{I}$ is a function $\pi : \mathbf{x} \cup \mathbf{y} \rightarrow \Delta^\mathcal{I}$ such that $\pi(v) \in A^\mathcal{I}$ for every atom $A(v)$ of q and $(\pi(v), \pi(w)) \in r^\mathcal{I}$ for every atom $r(v, w)$ of q. We write $\mathcal{I} \models q(a_1, \ldots, a_n)$ if there is a match of q in $\mathcal{I}$ with $\pi(x_i) = a_i$, for all $i \leq n$. A tuple $\mathbf{a}$ of elements from $\mathsf{ind}(\mathcal{A})$ is a *certain answer* to q over a KB $(\mathcal{T}, \mathcal{A})$, written $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$, if $\mathcal{I} \models q(\mathbf{a})$ for all models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$.

A *signature* Σ is a set of concept and role names. For a given signature Σ and a query language $\mathcal{Q}$, we denote with $\mathcal{Q}_\Sigma$ the set of all queries in $\mathcal{Q}$ that use only names from Σ. Given an ABox $\mathcal{A}$, S^+ and S^- denote n-ary relations over $\mathsf{ind}(\mathcal{A})$, called *positive* and *negative examples over* $\mathcal{A}$, resp.

Reasoning Problems. We study the following decision problem for some ontology language $\mathcal{L}$ and query language $\mathcal{Q}$:

Problem:	*Query-by-Example* $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$
Input:	$(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$ with $(\mathcal{T}, \mathcal{A})$ an $\mathcal{L}$- KB, S^+ and S^- examples over $\mathcal{A}$ and Σ a signature
Question:	Is there a query $q(\mathbf{x}) \in \mathcal{Q}_\Sigma$ such that • $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$ for all $\mathbf{a} \in S^+$, and • $\mathcal{T}, \mathcal{A} \not\models q(\mathbf{b})$, for all $\mathbf{b} \in S^-$?

A closely related problem is the *query definability problem* $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$ which takes as input a tuple $(\mathcal{T}, \mathcal{A}, S^+, \Sigma)$ and asks whether there is a query $q(\mathbf{x}) \in \mathcal{Q}_\Sigma$ such that an n-tuple $\mathbf{a}$ is a certain answer if, and only if $\mathbf{a} \in S^+$. If a tuple $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$ is a yes-instance of $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$, then we call the query $q(\mathbf{x})$ a *witness*. We further define the variant $\mathsf{QBE}_f(\mathcal{L}, \mathcal{Q})$ (f standing for *full*) as the problem of deciding for a given tuple $(\mathcal{T}, \mathcal{A}, S^+, S^-)$ whether $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma^*) \in \mathsf{QBE}(\mathcal{L}, \mathcal{Q})$, where Σ^* is the set of *all* concept and role names occurring in $(\mathcal{T}, \mathcal{A})$; $\mathsf{QDEF}_f(\mathcal{L})$ is defined analogously. Besides the decision problems, we will also be interested in the size of witness queries (if they exist).

We remark that allowing individual names in witness queries might be desirable in some applications, where the user knows some 'special' individuals which are relevant for her query. We show that our choice of forbidding them is without loss of generality. Let $\mathsf{QBE}_c(\mathcal{L}, \mathcal{Q})$ be the variant of $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ that takes another input $I \subseteq \mathsf{ind}(\mathcal{A})$ and allows the witness query to use constants from I. We then have:

Lemma 1. *$\mathsf{QBE}_c(\mathcal{L}, \mathcal{Q})$ and $\mathsf{QDEF}_c(\mathcal{L}, \mathcal{Q})$ reduce in polynomial time to $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ and $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$, resp., for $\mathcal{Q} \in \{CQ, UCQ\}$, for any $\mathcal{L}$.*

Throughout the paper, we will assume that the input knowledge base $(\mathcal{T}, \mathcal{A})$ is consistent and that S^+ is not empty. Both conditions can be effectively checked and if one of them isn't satisfied the reasoning problems become easier, see appendix.

Moreover, we assume that all TBoxes $\mathcal{T}$ are in $\mathcal{ELI}_\bot$*-normal form*, that is, CIs in $\mathcal{T}$ take one of the following forms:

$$\top \sqsubseteq A \quad A \sqsubseteq \bot \quad A \sqcap A' \sqsubseteq B \quad A \sqsubseteq \exists r.B \quad \exists r.A \sqsubseteq B$$

where A, A', B range over concept names and r ranges over roles. It has been shown that every Horn-$\mathcal{ALCI}$ TBox $\mathcal{T}$ can be transformed in polynomial time to an $\mathcal{ELI}_\bot$ TBox $\mathcal{T}'$ such that $\mathcal{T}'$ is a conservative extension of $\mathcal{T}$ [Bienvenu *et al.*, 2016], and it is easily verified that then $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\text{Horn-}\mathcal{ALCI}, \mathcal{Q})$ iff $(\mathcal{T}', \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\mathcal{ELI}_\bot, \mathcal{Q})$, for $\mathcal{Q} \in \{\text{CQ}, \text{UCQ}\}$.

3 Model-Theoretic Characterizations

In this section, we provide model-theoretic characterizations of QBE and QDEF, setting the foundations for the development of our decision procedures later on. We presume the standard notion of Σ-*homomorphisms* between interpretations (cf. appendix) and write $\mathcal{I} \to_\Sigma \mathcal{J}$ if there is a homomorphism restricted to the signature Σ from $\mathcal{I}$ to $\mathcal{J}$, and $(\mathcal{I}, \mathbf{a}) \to_\Sigma (\mathcal{J}, \mathbf{b})$ if there is such homomorphism that additionally maps the tuple $\mathbf{a}$ from $\Delta^\mathcal{I}$ to $\mathbf{b}$ from $\Delta^\mathcal{J}$. We drop the Σ in case it comprises all relevant names.

Our characterization is based on the notion of direct products. Let $\mathcal{I}, \mathcal{J}$ be interpretations. The *direct product* $\mathcal{I} \otimes \mathcal{J}$ *of* $\mathcal{I}$ *and* $\mathcal{J}$ is the interpretation defined by $\Delta^{\mathcal{I}\otimes\mathcal{J}} = \Delta^\mathcal{I} \times \Delta^\mathcal{J}$, $A^{\mathcal{I}\otimes\mathcal{J}} = A^\mathcal{I} \times A^\mathcal{J}$, and

$$r^{\mathcal{I}\otimes\mathcal{J}} = \{((a_1, b_1), (a_2, b_2)) \mid (a_1, a_2) \in r^\mathcal{I}, (b_1, b_2) \in r^\mathcal{J}\},$$

for all concept names A and role names r. The product $(\mathcal{I}, \mathbf{a}) \otimes (\mathcal{J}, \mathbf{b})$ is defined as $(\mathcal{I} \otimes \mathcal{J}, \mathbf{a} \otimes \mathbf{b})$, where $(a_1, \ldots, a_n) \otimes (b_1, \ldots, b_n) = ((a_1, b_1), \ldots, (a_n, b_n))$. Given Σ, a product $\Pi_{i=1}^n (\mathcal{I}_i, \mathbf{a}_i) = (\mathcal{I}_1, \mathbf{a}_1) \otimes \ldots \otimes (\mathcal{I}_n, \mathbf{a}_n)$ is called Σ-*safe* if every element of the tuple $\mathbf{a}_1 \otimes \ldots \otimes \mathbf{a}_n$ appears in the extension of some concept or role name from Σ in $\Pi_{i=1}^n (\mathcal{I}_i, \mathbf{a}_i)$; again, we drop Σ in case it is trivial.

Let us recall the characterization for QBE with CQs over relational databases [ten Cate and Dalmau, 2015; Barceló and Romero, 2017]. For the sake of simplicity, we state it here in our terminology, that is, consider ABoxes instead of databases. Given an ABox $\mathcal{A}$ and sets S^+, S^- of examples over $\mathcal{A}$, there is a CQ distinguishing S^+ and S^- iff

1. $\Pi_{\mathbf{a} \in S^+} (\mathcal{I}_\mathcal{A}, \mathbf{a})$ is safe, and
2. $\Pi_{\mathbf{a} \in S^+} (\mathcal{I}_\mathcal{A}, \mathbf{a}) \not\to (\mathcal{I}_\mathcal{A}, \mathbf{b})$ for every $\mathbf{b} \in S^-$,

where $\mathcal{I}_\mathcal{A}$ is $\mathcal{A}$ viewed as an interpretation. The intuition behind this characterization is as follows: the constructed product can be viewed as CQ with answer variables $\Pi_{\mathbf{a} \in S^+} \mathbf{a}$; in fact, this CQ is the *least general generalization* of the positive examples. Condition 1 ensures that it is a well-defined CQ by requiring all answer variables to actually appear, and Condition 2 ensures that no negative examples are returned.

We argue, however, that this simple characterization does not apply to the case with ontologies. In fact, the example from the introduction does not satisfy Condition 1, but there exists a witness query. We lift the characterization to take into account non-empty TBoxes using *universal interpretations*.

Universal Interpretations. Let $(\mathcal{T}, \mathcal{A})$ be a consistent Horn-$\mathcal{ALCI}$ KB and $\mathcal{T}$ in $\mathcal{ELI}_\bot$-normal form. A *type for* $\mathcal{T}$ is a subset t of the concept names in $\mathcal{T}$ such that $\mathcal{T} \models \bigsqcap t \sqsubseteq A$ implies $A \in t$ for all concept names A. When $a \in \mathsf{ind}(\mathcal{A})$, t, t' are types for $\mathcal{T}$, and r is a role, we write

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

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

(i) $a \rightsquigarrow_{r_0}^{\mathcal{T},\mathcal{A}} t_1$ and (ii) $t_i \rightsquigarrow_{r_i}^{\mathcal{T}} t_{i+1}$ for every $1 \leq i < n$.

We use $\mathsf{tail}(\pi)$ to denote the last element of a path π. Let Paths be the set of all paths for $\mathcal{A}$ and $\mathcal{T}$. The *universal model* $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ of $(\mathcal{T}, \mathcal{A})$ is defined as follows:

$$\begin{aligned}
\Delta^{\mathcal{U}_{\mathcal{T},\mathcal{A}}} &= \mathsf{Paths} \\
A^{\mathcal{U}_{\mathcal{T},\mathcal{A}}} &= \{a \in \mathsf{ind}(\mathcal{A}) \mid \mathcal{T}, \mathcal{A} \models A(a)\} \cup \\
&\quad\ \{\pi \in \mathsf{Paths} \setminus \mathsf{ind}(\mathcal{A}) \mid A \in \mathsf{tail}(\pi)\} \\
r^{\mathcal{U}_{\mathcal{T},\mathcal{A}}} &= \{(a, b) \in \mathsf{ind}(\mathcal{A})^2 \mid r(a, b) \in \mathcal{A}\} \cup \\
&\quad\ \{(\pi, \pi r t) \mid \pi r t \in \mathsf{Paths}\} \cup \\
&\quad\ \{(\pi r^- t, \pi) \mid \pi r^- t \in \mathsf{Paths}\}
\end{aligned}$$

It is well-known that $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ is *universal* in the sense that $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$ iff $\mathcal{U}_{\mathcal{T},\mathcal{A}} \models q(\mathbf{a})$ for every UCQ $q(\mathbf{x})$ and every tuple $\mathbf{a}$ of individuals [Bienvenu and Ortiz, 2015].

We state now our characterization for $\mathcal{Q} = \text{CQ}$.

Theorem 1. *For every Horn-$\mathcal{ALCI}$ KB $(\mathcal{T}, \mathcal{A})$, all n-ary relations S^+ and S^- over* $\mathsf{ind}(\mathcal{A})$*, and signatures Σ, we have:*

- $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}$(*Horn-$\mathcal{ALCI}$, CQ*) *iff*
  1. $\Pi_{\mathbf{a} \in S^+} (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe, and*
  2. $\Pi_{\mathbf{a} \in S^+} (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ *for all* $\mathbf{b} \in S^-$.
- $(\mathcal{T}, \mathcal{A}, S^+, \Sigma) \in \mathsf{QDEF}$(*Horn-$\mathcal{ALCI}$, CQ*) *iff*

  1.' $\Pi_{\mathbf{a} \in S^+} (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe, and*

  2.' $\Pi_{\mathbf{a} \in S^+} (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ *for all* $\mathbf{b} \in \mathsf{ind}(\mathcal{A})^n \setminus S^+$.

Thus, the characterization is the same as in the database setting with $\mathcal{I}_\mathcal{A}$ replaced by $\mathcal{U}_{\mathcal{T},\mathcal{A}}$. Note that $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ is possibly infinite, so the product is, in contrast to the database case, *not* the witness. In fact, the proof for direction ($\Leftarrow$) merely shows that *there is* a witness, but in a non-constructive way based on the finite outdegree of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$. Hence, Theorem 1 does not give immediate bounds on the size of witness queries.

In case of UCQs the additional expressive power leaves us with a simpler characterization, the product is compensated for by the use of disjunction in the query language and is thus not necessary anymore.

Theorem 2. *For every Horn-$\mathcal{ALCI}$ KB $(\mathcal{T}, \mathcal{A})$, all n-ary relations S^+ and S^- over* $\mathsf{ind}(\mathcal{A})$*, and signatures Σ, we have:*

- $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}$(*Horn-$\mathcal{ALCI}$, UCQ*) *iff* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe and* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$*, for all* $\mathbf{a} \in S^+$ *and* $\mathbf{b} \in S^-$.
- $(\mathcal{T}, \mathcal{A}, S^+, \Sigma) \in \mathsf{QDEF}$(*Horn-$\mathcal{ALCI}$, UCQ*) *iff* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe and* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ *for all* $\mathbf{a} \in S^+$ *and* $\mathbf{b} \in \mathsf{ind}(\mathcal{A})^n \setminus S^+$.

4 Complexity of QBE and QDEF

Based on the characterizations in Theorems 1 and 2, we now pinpoint the precise complexity for the introduced decision problems. We start with observing that Σ-safety (in both theorems) can be checked in exponential time by computing first $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ up to depth 1, computing the product (only in case of Theorem 1), and directly checking the condition.

Lemma 3. *Σ-safety can be decided in* ExpTime.

For Conditions 2 and 2' of Theorem 1 it is sufficient to give an algorithm for deciding $\Pi_{\mathbf{a}\in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{a}) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{b})$ for some $\mathbf{b}$; this algorithm can also be used for the homomorphism checks in Theorem 2 by treating the elements $\mathbf{a}\in S^+$ individually. Note that there is no immediate decision procedure, as the involved interpretations $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ are typically infinite. We can, however, exploit regularity of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$.

Let us fix an input $(\mathcal{T},\mathcal{A},S^+,S^-,\Sigma)$ with $k=|S^+|$, and denote with $\mathcal{U}^k_{\mathcal{T},\mathcal{A}}$ the product $\Pi^k_{i=1}\mathcal{U}_{\mathcal{T},\mathcal{A}}$. Observe first that $\mathcal{U}^k_{\mathcal{T},\mathcal{A}}$ might be disconnected and that for our purposes it suffices to consider the substructure $\mathcal{P}$ of $\mathcal{U}^k_{\mathcal{T},\mathcal{A}}$ containing all elements from $\mathsf{ind}(\mathcal{A})^k$ and everything that is reachable from there; thus, the domain $\Delta^\mathcal{P}$ of $\mathcal{P}$ is the smallest set such that:

- $\mathsf{ind}(\mathcal{A})^k\subseteq\Delta^\mathcal{P}$, and whenever $\mathbf{p}\in\Delta^\mathcal{P}$ and $(\mathbf{p},\mathbf{p}')\in r^{\mathcal{U}^k_{\mathcal{T},\mathcal{A}}}$ or $(\mathbf{p}',\mathbf{p})\in r^{\mathcal{U}^k_{\mathcal{T},\mathcal{A}}}$, then also $\mathbf{p}'\in\Delta^\mathcal{P}$.

It is easy to show that for $\mathbf{a}^*=\Pi_{\mathbf{a}\in S^+}\mathbf{a}$, we have:

Lemma 4. *For every $\mathbf{b}\in S^-$, we have $(\mathcal{U}^k_{\mathcal{T},\mathcal{A}},\mathbf{a}^*)\to_\Sigma(\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{b})$ iff $(\mathcal{P},\mathbf{a}^*)\to_\Sigma(\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{b})$.*

For what follows, it is convenient to characterize $r^\mathcal{P}$ in terms of (tuples of) *types*, similar to the definition of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$. For doing so, let TP be the set of all types for $\mathcal{T}$ and $\Delta=\mathsf{ind}(\mathcal{A})\cup\mathsf{TP}$. Then define, for each role r, a binary relation $\hookrightarrow^{\mathcal{T},\mathcal{A}}_r$ on Δ^k by taking $\mathbf{c}\hookrightarrow^{\mathcal{T},\mathcal{A}}_r\mathbf{d}$ iff $\mathbf{c}=(c_1,\dots,c_k)$ and $\mathbf{d}=(d_1,\dots,d_k)$ and for each $1\le i\le k$ we have:

- if $c_i,d_i\in\mathsf{ind}(\mathcal{A})$, then $r(c_i,d_i)\in\mathcal{A}$ or $r^-(d_i,c_i)\in\mathcal{A}$;
- if $c_i\in\mathsf{ind}(\mathcal{A}),d_i\in\mathsf{TP}$, then $c_i\rightsquigarrow^{\mathcal{T},\mathcal{A}}_r d_i$;
- if $c_i,d_i\in\mathsf{TP}$, then $c_i\rightsquigarrow^{\mathcal{T}}_r d_i$ or $d_i\rightsquigarrow^{\mathcal{T}}_{r^-} c_i$.

For $\mathbf{p}=(\pi_1,\dots,\pi_k)\in\Delta^\mathcal{P}$, denote with $\mathsf{tail}(\mathbf{p})$ the tuple $(\mathsf{tail}(\pi_1),\dots,\mathsf{tail}(\pi_k))$. It should be clear that we have

- $(\mathbf{p},\mathbf{p}')\in r^\mathcal{P}$ iff $\mathsf{tail}(\mathbf{p})\hookrightarrow^{\mathcal{T},\mathcal{A}}_r\mathsf{tail}(\mathbf{p}')$.

We give a characterization for $(\mathcal{P},\mathbf{a}^*)\to_\Sigma(\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{b})$, which will be the basis of our decision procedure. Intuitively, we decompose $\mathcal{P}$ into the non-tree-shaped part with domain $N=\mathsf{ind}(\mathcal{A})^k$ and the tree-shaped subinterpretations below each $\mathbf{a}\in N$ (which are characterized alone by their roots, similar to $\mathcal{U}_{\mathcal{T},\mathcal{A}}$). The latter have to be decomposed again because they might be mapped to different parts of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$. We denote with $\mathcal{P}_\mathbf{c}$ for $\mathbf{c}\in\Delta^k$ the sub-interpretation of $\mathcal{P}$ rooted at some $\mathbf{c}$. Moreover, we use the notation $\mathcal{U}_{\mathcal{T},t}$ for a type t for $\mathcal{T}$ as an abbreviation for $\mathcal{U}_{\mathcal{T},\{A(a_t)\mid A\in t\}}$ and denote with a_t its root. Given some Σ-role r, a tuple $\mathbf{c}\in\Delta^k$, a set $T\subseteq\Delta^k$, and a type $t\in\mathsf{TP}$, we write $(r,\mathbf{c},T,t)\in\mathsf{PHom}$, for *partial homomorphism*, if there is a partial function $g:\Delta^{\mathcal{P}_\mathbf{c}}\to\Delta^{\mathcal{U}_{\mathcal{T},t}}$ satisfying the following conditions:

- g is a homomorphism on its domain;
- $g(\mathbf{c})=\pi$ for some $\pi=a_trt'\in\Delta^{\mathcal{U}_{\mathcal{T},t}}$;
- if $g(\mathbf{p})$ is defined and $(\mathbf{p},\mathbf{p}')\in s^{\mathcal{P}_\mathbf{c}}$ for a Σ-role s, then either $g(\mathbf{p})=a_t$ and $\mathsf{tail}(\mathbf{p})\in T$ or $g(\mathbf{p}')$ is defined.

Intuitively, $(r,\mathbf{c},T,t)$ belongs to PHom if there is a homomorphism from $\mathcal{P}_\mathbf{c}$ to the subtree rooted at some r-successor of the root a_t of $\mathcal{U}_{\mathcal{T},t}$ given that some parts of $\mathcal{P}_\mathbf{c}$ can be 'delayed' to T when they map to a_t. The component T is necessary because of the 'bidirectional nature' of Horn-$\mathcal{ALCI}$. In general, a homomorphism $(\mathcal{P},\mathbf{a}^*)\to_\Sigma(\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{b})$ does not map subtrees of $\mathcal{P}$ to subtrees in $\mathcal{U}_{\mathcal{T},\mathcal{A}}$, and T is used to synchronize between different subtrees of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$, see the characterization below. For $\mathbf{c}\in\Delta^k,t\in\mathsf{TP}$ we write $\mathbf{c}\to_\Sigma t$ if there is a Σ-homomorphism from an element of type $\mathbf{c}$ to an element of type t. We further denote with $\mathsf{tp}_{\mathcal{U}_{\mathcal{T},\mathcal{A}}}(\pi)$ the type of π in $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ and with $\mathcal{P}|_N$ the restriction of $\mathcal{P}$ to domain N. We establish the following characterization.

Lemma 5. *$(\mathcal{P},\mathbf{a}^*)\to_\Sigma(\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{b})$ iff there is a Σ-homomorphism $h:(\mathcal{P}|_N,\mathbf{a}^*)\to_\Sigma(\mathcal{U}_{\mathcal{T},\mathcal{A}},\mathbf{b})$ and a labeling $T(\pi)\subseteq\Delta^k$ for every $\pi\in\mathsf{range}(h)\cup\mathsf{ind}(\mathcal{A})$ such that:*

1. *for every $\mathbf{p}\in N$, we have $\mathbf{p}\in T(h(\mathbf{p}))$;*
2. *for every $\mathbf{c}\in T(\pi)$, we have $\mathbf{c}\to_\Sigma\mathsf{tp}_{\mathcal{U}_{\mathcal{T},\mathcal{A}}}(\pi)$;*
3. *for every $\pi\in\mathsf{range}(h)\cup\mathsf{ind}(\mathcal{A})$, every $\mathbf{c}\in T(\pi)$, and every $\mathbf{d}$ with $\mathbf{c}\hookrightarrow^{\mathcal{T},\mathcal{A}}_r\mathbf{d}$ one of the following is true:*
   (a) there is some $\pi'\in\mathsf{range}(h)\cup\mathsf{ind}(\mathcal{A})$ such that $(\pi,\pi')\in r^{\mathcal{U}_{\mathcal{T},\mathcal{A}}}$ and $\mathbf{d}\in T(\pi')$, or
   (b) $(r,\mathbf{d},T(\pi),\mathsf{tp}_{\mathcal{U}_{\mathcal{T},\mathcal{A}}}(\pi))\in\mathsf{PHom}$.

4.1 Horn-$\mathcal{ALCI}$

We now devise a decision procedure for the criterion in Lemma 5. First observe that there are only double exponentially many mappings h and T, since $\mathbf{a}^*$ is forced to be mapped to $\mathbf{b}$, parts disconnected from $\mathbf{a}^*$ can be neglected, and $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ has bounded outdegree. We can thus enumerate all possible such mappings. Conditions 1, 2, and 3(a) can be checked in double exponential time using straightforward algorithms. For Condition 3(b), we devise a mosaic-based decision procedure similar to an algorithm in [Jung *et al.*, 2017].

A mosaic represents the neighborhood of an element in $\Delta^{\mathcal{U}_{\mathcal{T},\mathcal{A}}}$ of some type $t\in\mathsf{TP}$, together with 'types' of elements from $\Delta^\mathcal{P}$ which can be mapped there. Formally, a *mosaic* is a tuple $M=(t,T,r_0,t_0,T_0,\dots,r_n,t_n,T_n)$ with $n\le|\mathcal{T}|$, $t,t_i\in\mathsf{TP}$, $T,T_i\subseteq\Delta^k\setminus N$, and r_i Σ-roles such that:

(i) $t_0\rightsquigarrow_{r_0}t$ and $t\rightsquigarrow_{r_i}t_i$ for all $1\le i\le n$;

(ii) if $t\rightsquigarrow_r t'$, there is an $1\le i\le n$ with $(r_i,t_i)=(r,t')$;

(iii) $\mathbf{t}\in T$ implies $\mathbf{t}\to_\Sigma t$;

(iv) for every $\mathbf{t}\in T$ and every $\mathbf{t}\hookrightarrow^{\mathcal{T},\mathcal{A}}_r\mathbf{t}'$ for some Σ-role r, we either have $r=r_0^-$ and $\mathbf{t}'\in T_0$ or $r=r_i$ and $\mathbf{t}'\in T_i$, for some $1\le i\le n$.

Intuitively, t_0 is the predecessor type of t and $t_1,\dots,t_n$ are the successors via roles r_i. Condition (iv) ensures that successors of types $\mathbf{t}\in T$ mapped to t can be mapped to either t_0 or some t_i. Given a set $\hat{T}\subseteq\Delta^k\setminus N$, a *root mosaic for $\hat{T}$* is a tuple $M=(t,T,r_0,t_0,T_0,\dots,r_n,t_n,T_n)$ satisfying (i)–(iii) above, $T_0 = \emptyset$, and the variant (iv') of (iv) which is obtained by replacing '$\mathbf{t} \in T$' with '$\mathbf{t} \in T \setminus \hat{T}$'.

We define a *mosaic elimination* procedure as follows. Define a sequence of sets of mosaics by starting with $\mathfrak{M}_0$ as the set of all mosaics and root mosaics, and obtain $\mathfrak{M}_{i+1}$ from $\mathfrak{M}_i$ by removing all (root) mosaics $M = (t, T, r_0, t_0, T_0, \dots, r_n, t_n, T_n)$ from $\mathfrak{M}_i$ violating the following compatibility condition:

(E) for every $1 \leq j \leq n$, there is an $M' = (t', T', r'_0, t'_0, T'_0, \dots, r'_m, t'_m, T'_m) \in \mathfrak{M}_i$ such that $t = t'_0$, $r_j = r'_0$, $t_j = t'$, $T'_0 \subseteq T$, and $T_j \subseteq T'$.

Let $\hat{\mathfrak{M}}$ be where the sequence $\mathfrak{M}_0 \supseteq \mathfrak{M}_1 \supseteq \dots$ stabilizes.

Lemma 6. *$(\hat{r}, \hat{\mathbf{t}}, \hat{T}, \hat{t}) \in \mathsf{PHom}$ iff $\hat{\mathfrak{M}}$ contains a root mosaic $M = (t, T, r_0, t_0, T_0, \dots, r_n, t_n, T_n)$ for $\hat{T}$ with $t = \hat{t}$ and a mosaic $M' = (t', T', r'_0, t'_0, T'_0, \dots, r'_m, t'_m, T'_m)$ with $\hat{\mathbf{t}} \in T'$ and $r_0 = \hat{r}$ such that $t = t'_0$, $r_i = r'_0$, $t_i = t'$, $T'_0 \subseteq T$, and $T_i \subseteq T'$ for some i.*

It remains to discuss the running time of our procedure. The set Δ has size $|\mathcal{A}| + 2^{|\mathcal{T}|}$, thus Δ^k is of size $N_{\mathcal{A},\mathcal{T},k} := (|\mathcal{A}| + 2^{|\mathcal{T}|})^k$. Hence, the number of mosaics is bounded by $2^{N_{\mathcal{A},\mathcal{T},k}}$. In each round of the elimination procedure at least one mosaic is removed, thus the procedure terminates after double exponentially many steps. Finally, the checks in (E) can be implemented in exponential time. We thus conclude:

Corollary 7. *For $\mathcal{L}$ = Horn-$\mathcal{ALCI}$ and $\mathcal{Q} \in \{CQ, UCQ\}$, $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ and $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$ are in* 2-ExpTime.

We show next a matching lower bound.

Lemma 8. *For $\mathcal{L} = \mathcal{ELI}$ and $\mathcal{Q} \in \{CQ, UCQ\}$, $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ and $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$ are 2-ExpTime-hard.*

We reduce the word problem for exponential space bounded alternating Turing machines (ATM) which is 2-ExpTime-hard [Chandra *et al.*, 1981]. Given an ATM M and a word w, we construct a TBox $\mathcal{T}$ such that M accepts w iff $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, a) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, b)$ for $\mathcal{A} = \{A(a), B(b)\}$ and some signature Σ. We assume without loss of generality that instead of halting in the accepting state, M enters an infinite loop of special states without changing the tape anymore. We sketch the main idea by describing the universal model.

Below a, $\mathcal{T}$ enforces the infinite tree that is obtained by repeatedly glueing the pattern in Figure 1(a) to its leaves (as indicated by $\circ$; only Σ-symbols depicted). Note that this pattern (without the outgoing path labeled with α_0, α_1) is in fact the basic one of a computation tree of an ATM: a universal configuration of length 2^n (labeled with U) followed by a branch into two existential configurations of the same length (labeled with E_1, E_2). We call this the *skeleton tree*. Apart from the skeleton tree, for *every* possible choice of α_0, α_1, a path of the shape depicted in the right starts from *every* node of the tree. There, α_0 and α_1 range over all possible triples containing the content of three consecutive tape cells, e.g., $\langle a, b, c\rangle$ or $\langle a, (q, b), c\rangle$.

Below b, $\mathcal{T}$ enforces an infinite tree as illustrated in Figure 1(b), having the following properties:

- It starts with a path of length 2^n labeled with the initial configuration, encoded using triples.

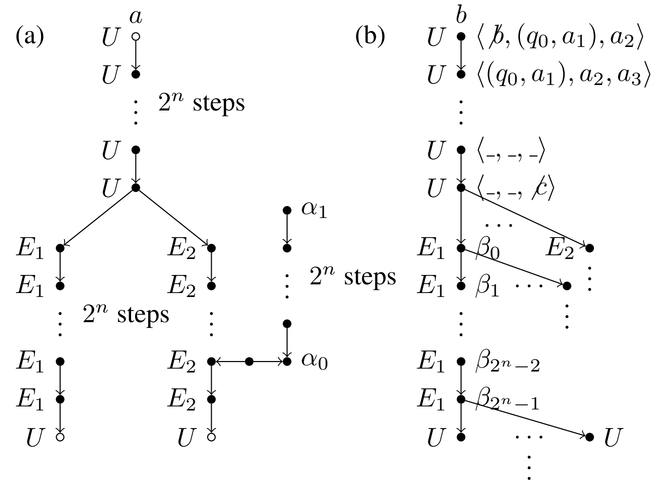


Figure 1: Parts of the universal model enforced in Lemma 8.

- All other nodes are labeled with a pair $\beta = (\alpha, \alpha')$ with α, α' triples as described above. In this case, α (resp., α') is intended to represent the content of the tape cell in the current (resp., previous) configuration.
- Every path of length 2^n (between E_x and U or U and E_x), e.g., $\beta_0, \dots, \beta_{2^n-1}$ in Figure 1, corresponds to the description of a valid configuration of M, and a possible predecessor configuration.
- This is continued infinitely, always switching between universal and existential configurations, as depicted.

It is instructive to consider some homomorphism h_0 of the skeleton tree below a into the tree below b. Informally, h_0 can be thought of as a labeling of the skeleton (and thus of the computation tree) with actual configurations. It remains to ensure that the transition relation of M is obeyed, which is done as follows. Every node in the tree below b has also outgoing paths of the same shape as the one depicted in the left side (for the sake of clarity and space they are not depicted in Fig. 1). However, it has only such paths for every α, α' except the label $\beta = \alpha_0, \alpha_1$ at the current node. Let now v be a node in the skeleton tree and assume its image $v' = h_0(v)$ has label $\beta = (\alpha, \alpha')$. Clearly, h_0 can be extended for all outgoing paths except the one labeled with $\alpha_0 = \alpha$ and $\alpha_1 = \alpha'$. Additionally, by construction, the end of this path can only be mapped to the corresponding cell in the previous configuration. The homomorphism condition ensures that the computation tree obeys the transition relation.

Summarizing, from Lemma 8 and Corollary 7 we obtain:

Theorem 9. *For $\mathcal{L}$ = Horn-$\mathcal{ALCI}$ and $\mathcal{Q} \in \{CQ, UCQ\}$, $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ and $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$ are 2-ExpTime-complete.*

We remark that the hardness proof crucially relies on Σ. For CQs, it is adapted to the unrestricted signature case by adding an assertion $A'(a')$ to the ABox, and enforcing below a' a tree identical (up to Σ-homomorphisms) to the tree below a, by using fresh copies of non-Σ concept names. The product of the trees below a and a' is then as in the proof of Lemma 8.

This approach does not apply to UCQs. In fact, the problem becomes easier with unrestricted signature. To see the reason for this complexity drop, note that, for a TBox $\mathcal{T}$ in normal form, we have $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \to (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ iff $(\mathcal{U}_{\mathcal{T},\mathcal{A}}|_{\mathsf{ind}(\mathcal{A})}, \mathbf{a}) \to (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$. This can be straightforwardly decided in exponential time, see the appendix.

Theorem 10. *For $\mathcal{L}$ = Horn-$\mathcal{ALCI}$, $\mathsf{QBE}_f(\mathcal{L}, UCQ)$ and $\mathsf{QDEF}_f(\mathcal{L}, UCQ)$ are* ExpTime*-complete.*

4.2 Horn-$\mathcal{ALC}$

For Horn-$\mathcal{ALC}$, note that the characterizations in Theorems 1 and 2 and Lemma 5 are still valid since Horn-$\mathcal{ALC}$ is a fragment of Horn-$\mathcal{ALCI}$. However, the absence of inverse roles simplifies the decision procedure of PHom which is the bottleneck for Horn-$\mathcal{ALCI}$. Indeed, both $\mathcal{P}_{\mathbf{c}}$ and the anonymous parts of $\mathcal{U}_{\mathcal{T},t}$ are directed regular trees where all roles point away from the root, so the set T can be ignored and PHom is decided by applying standard techniques for regular trees.

Lemma 11. *Given $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$ with $(\mathcal{T}, \mathcal{A})$ a Horn-$\mathcal{ALC}$ KB, the relation* PHom *can be decided in* ExpTime.

Applying this Lemma, we obtain a coNExpTime upper bound for deciding QBE from the algorithm devised in the previous section. A matching lower bound is inherited from the database setting [ten Cate and Dalmau, 2015]. For UCQs, a careful analysis of Lemma 5 yields an ExpTime upper bound; the matching lower bound is obtained by a reduction from subsumption in Horn-$\mathcal{ALC}$ [Krötzsch *et al.*, 2013].

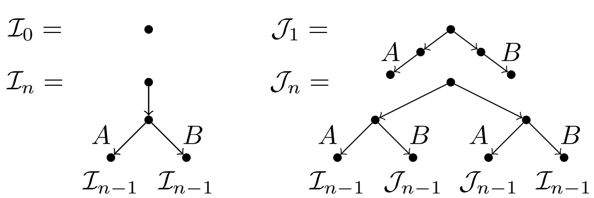
Theorem 12. *For $\mathcal{L}$ = Horn-$\mathcal{ALC}$, $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ and $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$ are* coNExpTime*-complete if $\mathcal{Q}$ = CQ and* ExpTime*-complete if $\mathcal{Q}$ = UCQ. All results also hold with unrestricted signature.*

5 Size of Witness Queries

We finally investigate the size of witness queries. We first establish the following double exponential lower bound.

Lemma 13. *There is a family of Horn-$\mathcal{ALC}$ knowledge bases $(\mathcal{T}_n, \mathcal{A}_n)_{n\geq 1}$, sets of examples S^+ and S^-, a signature Σ, and a polynomial $p(n)$ such that, for all $n \geq 1$, $|\mathcal{T}_n \cup \mathcal{A}_n| \leq p(n)$, $(\mathcal{T}_n, \mathcal{A}_n, S^+, S^-, \Sigma) \in \mathsf{QBE}$(Horn-$\mathcal{ALC}$, (U)CQ) and every (U)CQ witnessing this is of size $\Omega(2^{2^n})$.*

The main idea for the lower bound is to give Horn-$\mathcal{ALC}$ knowledge bases $(\mathcal{T}_n, \mathcal{A}_n)$ over two individuals a, b such that in $\mathcal{U}_{\mathcal{T}_n,\mathcal{A}_n}$ the trees below a and b are Σ-homomorphically equivalent to $\mathcal{I}_{2^n}$ and $\mathcal{J}_{2^n}$, respectively, where $\mathcal{I}_n, \mathcal{J}_n$ are given by the following recursive 'definitions':



It can be shown that $(\mathcal{I}_n, a) \not\to_\Sigma (\mathcal{J}_n, b)$, but that $(\mathcal{I}', a) \to_\Sigma (\mathcal{J}_n, b)$ for any sub-interpretation $\mathcal{I}'$ of $\mathcal{I}_n$. Thus, the smallest Σ-(U)CQ distinguishing between a and b in $(\mathcal{T}_n, \mathcal{A}_n)$ is $(\mathcal{I}_{2^n}, a)$ viewed as CQ, whose size is $\Omega(2^{2^n})$.
For Horn-$\mathcal{ALC}$, a matching upper bound is obtained by an analysis of Lemma 5 and the observations made in the previous Section. For Horn-$\mathcal{ALCI}$, we obtain a four-fold exponential upper bound on the size of the witness query by viewing the check for PHom as a reachability game on pushdown systems and apply known results from there [Kupferman *et al.*, 2010; Carayol and Hague, 2014]. We leave the exact sizes for future work.

Theorem 14. *If $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\mathcal{L}, CQ)$, there is a witness query of at most double (resp., four-fold) exponential size if $\mathcal{L}$ = Horn-$\mathcal{ALC}$ (resp., $\mathcal{L}$ = Horn-$\mathcal{ALCI}$).*

6 Discussion and Future Work

Our investigation opens a new whole research avenue towards improving the usability of ontology-enriched systems. From the theoretical perspective, the most natural next step is to broaden our understanding to different ontology and query languages. Given the state of the art of OES, we are particularly interested in 'lightweight' DLs, such as *DL-Lite* and $\mathcal{EL}$; our results already provide a solid basis for these logics. For non-Horn or Datalog$^\pm$ ontologies it will be more challenging – a good starting point for non-Horn DLs might be [Botoeva *et al.*, 2016b]. As for the query language, we will study regular path queries. From the practical perspective, it suggests itself to develop systems for QBE over KBs which not only implement reverse-engineering algorithms, but also allow interaction with the user, as done e.g., in [Bonifati *et al.*, 2014; Diaz *et al.*, 2016]. Given the high complexity of QBE, it will be also important to design heuristics [Tran *et al.*, 2014; Mottin *et al.*, 2016] or approximations [Barceló and Romero, 2017], as for relational databases. We note that some approximations considered by Barceló and Romero [2017] do not directly lead to better complexity in the context of OES. For example, 2-ExpTime-hardness in Lemma 8 already holds for *tree-shaped* CQs. Another possible approximation is bounding the size of the witness queries.

Related within DL research is the study of *query conservative extensions (QCE)*, where the question is whether two given ontologies or two knowledge bases can be distinguished by a query (without providing examples). Indeed, in the context of QCE, characterizations based on homomorphisms and universal models have been devised and inverse roles also tend to increase the complexity, see [Botoeva *et al.*, 2016a] for a recent survey, and references therein. We are, however, not aware of any direct reductions between QBE and QCE.

Within the broader context of machine learning, we believe that our results lay the foundations for questions related to *learnability* of queries, see [Cohen and Page, 1995] for an overview. In this line, one could investigate an ILP inspired variant: if an instance $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$ of QBE does *not* have a witness, is there an extension $\mathcal{T}' \supseteq \mathcal{T}$ such that there is a witness? In the context of active learning, one would be interested in learning a (conjunctive) query with membership and/or equivalence queries over a DL knowledge base. Finally, it would be interesting to extend the recently introduced framework of learning concepts over background structures of small degree and having only *local access* to the data [Grohe and Ritzert, 2017] with an ontology.

Acknowledgments

The authors were funded by EU's Horizon 2020 programme under the Marie Skłodowska-Curie grant 663830, the Research Foundation - Flanders project G.0428.15, and ERC consolidator grant 647289 CODA, respectively.

References

[Arenas *et al.*, 2016] Marcelo Arenas, Gonzalo I. Diaz, and Egor V. Kostylev. Reverse engineering SPARQL queries. In *Proc. of WWW-16*, pages 239–249, 2016.

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

[Barceló and Romero, 2017] Pablo Barceló and Miguel Romero. The complexity of reverse engineering problems for conjunctive queries. In *Proc. of ICDT-17*, 2017.

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *Proc. of RW-15*, 2015.

[Bienvenu *et al.*, 2016] Meghyn Bienvenu, Peter Hansen, Carsten Lutz, and Frank Wolter. First order-rewritability and containment of conjunctive queries in Horn description logics. In *Proc. of IJCAI-16*, pages 965–971, 2016.

[Bonifati *et al.*, 2014] Angela Bonifati, Radu Ciucanu, and Slawek Staworko. Interactive inference of join queries. In *Proc. of EDBT-14*, pages 451–462, 2014.

[Bonifati *et al.*, 2015] Angela Bonifati, Radu Ciucanu, and Aurélien Lemay. Learning path queries on graph databases. In *Proc. EDBT-15*, pages 109–120, 2015.

[Bonifati *et al.*, 2016] Angela Bonifati, Radu Ciucanu, and Slawek Staworko. Learning join queries from user examples. *ACM Trans. Database Syst.*, 40(4):24:1–24:38, 2016.

[Botoeva *et al.*, 2016a] Elena Botoeva, Boris Konev, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Inseparability and conservative extensions of description logic ontologies: A survey. In *Proc. of RW-16*, 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 *Proc. of IJCAI-16*, 2016.

[Calvanese *et al.*, 2016] Diego Calvanese, Pietro Liuzzo, Alessandro Mosca, José Remesal, Martin Rezk, and Guillem Rull. Ontology-based data integration in epnet: Production and distribution of food during the roman empire. *Eng. Appl. of AI*, 51:212–229, 2016.

[Carayol and Hague, 2014] Arnaud Carayol and Matthew Hague. Regular strategies in pushdown reachability games. In *Proc. of Reachability Problems, RP-14*, 2014.

[Chandra *et al.*, 1981] Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. Alternation. *J. ACM*, 28(1):114–133, January 1981.

[Cohen and Page, 1995] William W. Cohen and C. David Page. Polynomial learnability and inductive logic programming: Methods and results. *New Generation Comput.*, 13(3&4):369–409, 1995.

[Diaz *et al.*, 2016] Gonzalo I. Diaz, Marcelo Arenas, and Michael Benedikt. Sparqlbye: Querying RDF data by example. *PVLDB*, 9(13):1533–1536, 2016.

[Grohe and Ritzert, 2017] Martin Grohe and Martin Ritzert. Learning first-order definable concepts over structures of small degree. In *Proc. of LICS-17*, pages 1–12, 2017.

[Hovland *et al.*, 2017] Dag Hovland, Roman Kontchakov, Martin G. Skjæveland, Arild Waaler, and Michael Zakharyaschev. Ontology-based data access to Slegge. In *Proc. of ISWC-17*, pages 120–129, 2017.

[Jung *et al.*, 2017] Jean Christoph Jung, Carsten Lutz, Mauricio Martel, and Thomas Schneider. Query conservative extensions in horn description logics with inverse roles. In *Proc. of IJCAI-17*, 2017.

[Kharlamov *et al.*, 2015] Evgeny Kharlamov, Dag Hovland, Ernesto Jiménez-Ruiz, Davide Lanti, Hallstein Lie, Christoph Pinkel, Martín Rezk, Martin G. Skjæveland, Evgenij Thorstensen, Guohui Xiao, Dmitriy Zheleznyakov, and Ian Horrocks. Ontology based access to exploration data at Statoil. In *Proc. of ISWC-15*, pages 93–112, 2015.

[Kietz, 2002] Jörg-Uwe Kietz. Learnability of description logic programs. In *Inductive Logic Programming, 12th International Conference, ILP 2002, Sydney, Australia, July 9-11, 2002. Revised Papers*, pages 117–132, 2002.

[Krötzsch *et al.*, 2013] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Complexities of horn description logics. *ACM Trans. Comput. Logic*, 14(1):2:1–2:36, 2013.

[Kupferman *et al.*, 2010] Orna Kupferman, Nir Piterman, and Moshe Y. Vardi. An automata-theoretic approach to infinite-state systems. In *Time for Verification, Essays in Memory of Amir Pnueli*, pages 202–259, 2010.

[Mottin *et al.*, 2016] Davide Mottin, Matteo Lissandrini, Yannis Velegrakis, and Themis Palpanas. Exemplar queries: a new way of searching. *VLDB J.*, 25(6):741–765, 2016.

[Nienhuys-Cheng and de Wolf, 1997] Shan-Hwei Nienhuys-Cheng and Ronald de Wolf, editors. *Foundations of Inductive Logic Programming*, volume 1228 of *Lecture Notes in Computer Science*. Springer, 1997.

[Rodriguez-Muro *et al.*, 2013] Mariano Rodriguez-Muro, Roman Kontchakov, and Michael Zakharyaschev. Ontology-based data access: Ontop of databases. In *Proc. of ISWC-13*, pages 558–573, 2013.

[ten Cate and Dalmau, 2015] Balder ten Cate and Víctor Dalmau. The product homomorphism problem and applications. In *Proc. of ICDT-15*, pages 161–176, 2015.

[Tran *et al.*, 2014] Quoc Trung Tran, Chee Yong Chan, and Srinivasan Parthasarathy. Query reverse engineering. *VLDB J.*, 23(5):721–746, 2014.

[Zloof, 1975] Moshé M. Zloof. Query-by-example: the invocation and definition of tables and forms. In *Proc. of VLDB-75*, pages 1–24, 1975.

APPENDIX

A Additional Preliminaries

Homomorphisms. Let $\mathcal{I}, \mathcal{J}$ be interpretations, and Σ a signature. A Σ-*homomorphism from* $\mathcal{I}$ *to* $\mathcal{J}$ is a mapping $h : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$ such that $a \in A^{\mathcal{I}}$ implies $h(a) \in A^{\mathcal{J}}$ and $(a, b) \in r^{\mathcal{I}}$ implies $(h(a), h(b)) \in r^{\mathcal{J}}$, for all $a, b \in \Delta^{\mathcal{I}}$, concept names $A \in \Sigma$, and role names $r \in \Sigma$. We write $\mathcal{I} \to_\Sigma \mathcal{J}$ in case there exists a Σ-homomorphism from $\mathcal{I}$ to $\mathcal{J}$. If $\mathbf{a}$ and $\mathbf{b}$ are n-tuples from $\Delta^{\mathcal{I}}$ and $\Delta^{\mathcal{J}}$, respectively, we write $(\mathcal{I}, \mathbf{a}) \to_\Sigma (\mathcal{J}, \mathbf{b})$ if there is a Σ-homomorphism $h : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$ with $h(\mathbf{a}) = \mathbf{b}$. If Σ includes all relevant names, we omit it and write just $\mathcal{I} \to \mathcal{J}$.

B Proofs for Section 2

Lemma 1. $\mathsf{QBE}_c(\mathcal{L}, \mathcal{Q})$ *and* $\mathsf{QDEF}_c(\mathcal{L}, \mathcal{Q})$ *reduce in polynomial time to* $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ *and* $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$*, resp., for* $\mathcal{Q} \in \{CQ, UCQ\}$*, for any* $\mathcal{L}$.

Proof. We prove it for $\mathcal{L} = \mathrm{CQ}$, for UCQs it is similar. Let $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma, I)$ be an instance of QBE_c. Define M as the set of all $a \in I$ such that there is some position i such that $a_i = a$ for all $(a_1, \ldots, a_n) \in S^+$. Define $\Sigma' = \Sigma \cup \{X_a \mid a \in I \setminus M\}$ and $\mathcal{A}' = \mathcal{A} \cup \{X_a(a) \mid a \in I \setminus M\}$. It is routine to verify correctness of the reduction:

Claim. $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma, I) \in \mathsf{QBE}_c$ iff $(\mathcal{T}, \mathcal{A}', S^+, S^-, \Sigma') \in \mathsf{QBE}$.

Proof of the Claim. ($\Rightarrow$) Let $q(\mathbf{x})$ be a witness for $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma, I) \in \mathsf{QBE}_c(\mathcal{L})$. Obtain a query $q' \in \mathcal{L}_{\Sigma'}$ from q by processing every individual name a appearing in q as follows:

– if $a \in M$, then let i be a position such that $a_i = a$ for all $\mathbf{a} \in S^+$, and replace all occurrences of a with x_i;

– if $a \in I \setminus M$, then replace all occurrences of a with a fresh quantified variable x_a and add the conjunct $X_a(x_a)$.

It should be clear that $(\mathcal{T}, \mathcal{A}') \models q'(\mathbf{a})$ for all $\mathbf{a} \in S^+$. Assume now that $(\mathcal{T}, \mathcal{A}') \models q'(\mathbf{a})$ for some $\mathbf{a} \in S^-$, and let $\mathcal{I}$ be an arbitrary model of $(\mathcal{T}, \mathcal{A})$. Obviously, the extension $\mathcal{I}'$ of $\mathcal{I}$ interpreting every fresh concept X_a with $X_a^{\mathcal{I}'} = \{a\}$ is a model of $(\mathcal{T}, \mathcal{A}')$. Let π be a match for $q'(\mathbf{a})$ into $\mathcal{I}'$. By construction, π is also a match for $q(\mathbf{a})$ into $\mathcal{I}$. Hence, we obtain $(\mathcal{T}, \mathcal{A}) \models q(\mathbf{a})$, a contradiction.

($\Leftarrow$) Let $q(\mathbf{x})$ be a query witnessing $(\mathcal{T}, \mathcal{A}', S^+, S^-, \Sigma') \in \mathsf{QBE}(\mathcal{L})$. Note that it cannot be the case that there is a quantified variable z such that both $X_a(z)$ and $X_b(z)$ for $a \neq b$ appear in q, since this implies $S^+ = \emptyset$, contradicting our assumption about S^+. Moreover, there cannot be an atom $X_a(x_i)$ with $x_i \in \mathbf{x}$ an answer variable appears in $q(\mathbf{x})$, since this implies $a_i = a$ for all $\mathbf{a} \in S^+$, contradicting the construction of $\mathcal{A}'$. Obtain a query q' from q by replacing every quantified variable z such that $X_a(z)$ appears in q with a. It is routine to verify that $q'(\mathbf{x})$ witnesses that $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma, I) \in \mathsf{QBE}_c(\mathcal{L})$. □

We justify in detail the assumptions that $(\mathcal{T}, \mathcal{A})$ is actually consistent and that $S^+ \neq \emptyset$.

KB Consistency. We consider first the case when $(\mathcal{T}, \mathcal{A})$ happens to be inconsistent. In that case, we have $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$ for every n-ary CQ $q(\mathbf{x})$ and every $\mathbf{a} \in \mathsf{ind}(\mathcal{A})^n$, thus there is a witness if and only if $S^- = \emptyset$ (and then, every n-ary CQ is a witness). Hence, one could check $(\mathcal{T}, \mathcal{A})$ for inconsistency first, which can be done in ExpTime if $\mathcal{T}$ is formulated in Horn-$\mathcal{ALCI}$ [Krötzsch *et al.*, 2013].

No Positive Examples. The second case we discuss is $S^+ = \emptyset$, that is, the question whether there is a $\mathcal{L}_\Sigma$-query $q(\mathbf{x})$ with n answer variables such that $(\mathcal{T}, \mathcal{A}) \not\models q(\mathbf{b})$ for all $\mathbf{b} \in S^-$. A natural candidate for such a query is

$$q(\mathbf{x}) = \bigwedge_{A \in \Sigma} \bigwedge_{i=1}^{n} A(x_i) \wedge \bigwedge_{r \in \Sigma} \bigwedge_{i=1}^{n} \bigwedge_{j=1}^{n} r(x_i, x_j) \ .$$

It is easy to see that $q(\mathbf{x})$ is the most restrictive query in the sense that if $\mathcal{T}, \mathcal{A} \models q(\mathbf{a})$ then $\mathcal{T}, \mathcal{A} \models q'(\mathbf{a})$ for every n-ary query $q'(\mathbf{x})$, so the instance has a witness if and only if $q(\mathbf{x})$ is such a witness. Hence, an algorithm for deciding QBE in this special case has to check whether $(\mathcal{T}, \mathcal{A}) \not\models q(\mathbf{b})$ for all $\mathbf{b} \in S^-$. This can be done in ExpTime.

C Proofs for Section 3

Theorem 1. *For every Horn-$\mathcal{ALCI}$ KB $(\mathcal{T}, \mathcal{A})$, all n-ary relations S^+ and S^- over* $\mathsf{ind}(\mathcal{A})$*, and signatures Σ, we have:*

- $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}$*(Horn-$\mathcal{ALCI}$, CQ) iff*
  1. $\Pi_{\mathbf{a} \in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe, and*
  2. $\Pi_{\mathbf{a} \in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ *for all* $\mathbf{b} \in S^-$.
- $(\mathcal{T}, \mathcal{A}, S^+, \Sigma) \in \mathsf{QDEF}$*(Horn-$\mathcal{ALCI}$, CQ) iff*

  1.' $\Pi_{\mathbf{a} \in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe, and*

  2.' $\Pi_{\mathbf{a} \in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ *for all* $\mathbf{b} \in \mathsf{ind}(\mathcal{A})^n \setminus S^+$.

Proof. We show only the characterization for $\mathsf{QBE}(\mathsf{CQ})$; the proof for $\mathsf{QDEF}(\mathsf{CQ})$ is analogous.

($\Rightarrow$) Let $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\mathsf{CQ})$ with $S^+ = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ and let $q(\mathbf{x})$ be a Σ-CQ witnessing this. By universality, for every i, there is a match π_i from $q(\mathbf{x})$ into $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ with $\pi_i(\mathbf{x}) = \mathbf{a}_i$. Define π by taking $\pi(z) = (\pi_1(z), \ldots, \pi_m(z))$, for every variable z in $q(\mathbf{x})$. By construction, π is a match for $q(\mathbf{x})$ into $\Pi_{\mathbf{a} \in S^+}\mathcal{U}_{\mathcal{T},\mathcal{A}}$ and $\pi(\mathbf{x}) = \mathbf{a}_1 \otimes \ldots \otimes \mathbf{a}_m$, which is thus Σ-safe. Assume that Condition 2 does not hold, that is, there is a $\mathbf{b} \in S^-$ such that there is a homomorphism $h : \Pi_{\mathbf{a} \in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$. Composing π and h yields a match $\hat{\pi}$ from $q(\mathbf{x})$ into $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ with $\hat{\pi}(\mathbf{x}) = \mathbf{b}$. Hence, $\mathcal{T}, \mathcal{A} \models q(\mathbf{b})$, a contradiction.

($\Leftarrow$) For the other direction, let Conditions 1 and 2 be fulfilled. We show that there is a witness for $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\mathsf{CQ})$. Let $(\mathcal{I}, \mathbf{a}^*)$ be the Σ-restriction of $\Pi_{\mathbf{a} \in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$, and let $q(\mathbf{x})$ be $(\mathcal{I}, \mathbf{a}^*)$ viewed as a (possibly infinite) CQ; in particular, $\mathbf{a}^*$ becomes the tuple of answer variables $\mathbf{x}$. By Condition 1, $q(\mathbf{x})$ has the right number of answer variables. Clearly, every $\mathbf{a} \in S^+$ is a certain answer to $q(\mathbf{x})$, using the projection mappings, and none of the $\mathbf{b} \in S^-$ is a certain answer by Condition 2 and universality of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$. If $q(\mathbf{x})$ is finite, we are done. If $q(\mathbf{x})$ is infinite, we show that there is a finite subquery of $q(\mathbf{x})$ which is a witness. Denote with $q_i(\mathbf{x})$, $i \geq 0$, the restriction of $q(\mathbf{x})$ to variables that have distance at most i to the answer variables $\mathbf{x}$ in $q(\mathbf{x})$. By construction, we have that $q_i(\mathbf{x})$ is finite for every $i \geq 1$. To reach a contradiction, assume that q_i is not a witness for every $i \geq 1$, that is, there are matches π_i from $q_i(\mathbf{x})$ to $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ with $\pi_i(\mathbf{x}) = \mathbf{b}_i$ for some $\mathbf{b}_i \in S^-$. Since S^- is finite, there is some $\mathbf{b} \in S^-$ such that $\mathbf{b} = \mathbf{b}_i$ for infinitely many i. Thus, there are matches π_i, $i \geq 1$, from $q_i(\mathbf{x})$ to $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ with $\pi_i(\mathbf{x}) = \mathbf{b}$. We construct a sequence of matches $(\pi'_i)_{i\geq 1}$ such that for all $j \geq 1$, we have

(∗) for all $0 < i < j$, $\pi'_i(z) = \pi'_j(z)$ for all variables z occurring in q_i.

Start with setting $\pi'_1 = \pi_1$, obviously satisfying (∗). To define π'_j, assume that π'_k are defined for all $0 < k < j$. Let V be all variables that occur in q_j but not in q_{j-1}, and define, for all $k \geq j$, τ_k as the restriction of π_k to V. By construction V is finite. Moreover, as $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ has finite outdegree, there are only finitely many different τ_k. Choose some τ such that $\tau = \tau_k$ for infinitely many $k \geq j$. Then obtain a new sequence of matches by dropping all π_k such that $\tau_k \neq \tau$. Setting $\pi'_j = \tau \cup \pi'_{j-1}$ finishes the construction and satisfies (∗).

It remains to note that $\hat{\pi} = \bigcup_{i\geq 0} \pi'_i$ is a match for $q(\mathbf{x})$ into $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ with $\hat{\pi}(\mathbf{x}) = \mathbf{b}$. Thus, $\Pi_{\mathbf{a}\in S^+}(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$, contradicting Condition 2. □

Theorem 2. *For every Horn-$\mathcal{ALCI}$ KB $(\mathcal{T}, \mathcal{A})$, all n-ary relations S^+ and S^- over $\mathsf{ind}(\mathcal{A})$, and signatures Σ, we have:*

- $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\textit{Horn-}\mathcal{ALCI}, \textit{UCQ})$ *iff* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe and* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$, *for all* $\mathbf{a} \in S^+$ *and* $\mathbf{b} \in S^-$.
- $(\mathcal{T}, \mathcal{A}, S^+, \Sigma) \in \mathsf{QDEF}(\textit{Horn-}\mathcal{ALCI}, \textit{UCQ})$ *iff* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ *is Σ-safe and* $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ *for all* $\mathbf{a} \in S^+$ *and* $\mathbf{b} \in \mathsf{ind}(\mathcal{A})^n \setminus S^+$.

Proof. (⇒) Let $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\mathsf{UCQ})$, witnessed by a Σ-UCQ $q(\mathbf{x})$. By universality, for every $\mathbf{a} \in S^+$, there is a disjunct $q'(\mathbf{x})$ of $q(\mathbf{x})$ such that there is a match π of $q'(\mathbf{x})$ into $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ with $\pi(\mathbf{x}) = \mathbf{a}$. Since q' is a Σ-query, $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ is Σ-safe. Suppose now that there are $\mathbf{a} \in S^+$, $\mathbf{b} \in S^-$ such that there is a Σ-homomorphism $h : (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$. As $\mathbf{a} \in S^+$ and $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ is universal, there is a match π from $q(\mathbf{x})$ into $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ with $\pi(\mathbf{x}) = \mathbf{a}$. Composing h and π yields a match for $q(\mathbf{x})$ into $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$, thus $\mathcal{T}, \mathcal{A} \models q(\mathbf{b})$, a contradiction to $\mathbf{b} \in S^-$.

(⇐) For the other direction, suppose that $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ is Σ-safe and $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$, for all $\mathbf{a} \in S^+$ and $\mathbf{b} \in S^-$. We show that there is witness for $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\mathsf{UCQ})$.

For the sake of simplicity, we abbreviate $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ just with $\mathcal{I}$. Further, we denote with $q_{\mathcal{I},\mathbf{a}}(\mathbf{x})$ the interpretation $(\mathcal{I}, \mathbf{a})$ viewed as (possibly infinite) CQ with the answer variables $\mathbf{x}$ being the distinguished tuple $\mathbf{a}$. For a (U)CQ q, denote with $q^\Sigma(\mathbf{x})$ the restriction of $q(\mathbf{x})$ to symbols from Σ, and with $q^i(\mathbf{x})$ the restriction of $q(\mathbf{x})$ to variables that have distance at most i to the answer variables of $q(\mathbf{x})$.

We now define a possibly infinite UCQ $q(\mathbf{x})$ by taking $q(\mathbf{x}) = \bigvee_{\mathbf{a}\in S^+} q^\Sigma_{\mathcal{I},\mathbf{a}}$. By Σ-safety and universality of $\mathcal{I}$, every $\mathbf{a} \in S^+$ is a certain answer to this query and none of the $\mathbf{b} \in S^-$ is a certain answer. Thus, if $q(\mathbf{x})$ is finite, then $q(\mathbf{x})$ is the required witness. If $q(\mathbf{x})$ is infinite, we show that there is a finite subset of $q(\mathbf{x})$ that is a witness for $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$. Assume the opposite, that is, $q^i(\mathbf{x})$ is not a witness for every $i \geq 1$. To reach a contradiction, we will show that there exist $\mathbf{a} \in S^+$, $\mathbf{b} \in S^-$ with $(\mathcal{I}, \mathbf{a}) \to_\Sigma (\mathcal{I}, \mathbf{b})$, contradicting Condition 1. Clearly, For every such $q^i(\mathbf{x})$ we still have $\mathcal{T}, \mathcal{A} \models q^i(\mathbf{a})$ for all $\mathbf{a} \in S^+$. However, by our assumption, for every i, there is a $\mathbf{b}_i \in S^-$ such that $\mathcal{T}, \mathcal{A} \models q^i(\mathbf{b}_i)$. Since S^- is finite and $q(\mathbf{x})$ consists of finitely many disjuncts, there have to be $\mathbf{a} \in S^+$ and $\mathbf{b} \in S^-$ such that, for infinitely many i:

- $\mathbf{b} = \mathbf{b}_i$, and
- the disjunct $p_i = q^{\Sigma,i}_{\mathcal{I},\mathbf{a}}$ of q_i corresponding to $\mathbf{a}$, satisfies $\mathcal{T}, \mathcal{A} \models p_i(\mathbf{b}_i)$.

We can then proceed as in the proof of Theorem 1 and construct a homomorphism $(\mathcal{I}, \mathbf{a}) \to_\Sigma (\mathcal{I}, \mathbf{b})$. □

D Proofs for Section 4

Lemma 3. *Σ-safety can be decided in* ExpTime.

Proof. It suffices to compute $\Pi_{\mathbf{a}\in S^+}\mathcal{U}_{\mathcal{T},\mathcal{A}}$ for $\mathbf{a}^* = \Pi_{\mathbf{a}\in S^+}\mathbf{a}$ and all neighbors $\mathbf{b}$ of $\mathbf{a}$. This can be done by computing the universal model up to depth 1. Thus it is in ExpTime; it is also ExpTime-hard since it is at least as hard as subsumption. By the same approach, we obtain PTime-completeness for data complexity. □

Lemma 4. *For every $\mathbf{b} \in S^-$, we have $(\mathcal{U}^k_{\mathcal{T},\mathcal{A}}, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ iff $(\mathcal{P}, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$.*

Proof. The direction (⇒) is trivial since $\mathcal{P}$ is a sub-interpretation of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$.

For (⇐), let $h : (\mathcal{P}, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$. Note that, by definition, $\mathcal{P}$ is actually the maximal connected sub-interpretation containing $\mathsf{ind}(\mathcal{A})^k$. We can extend the homomorphism h to a connected component $\mathcal{I} \neq \mathcal{P}$ of $\mathcal{U}_{\mathcal{T},\mathcal{A}}$ by taking the projection of $\Delta^\mathcal{I}$ to an arbitrary (but fixed) component. □

Lemma 5. *$(\mathcal{P}, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ iff there is a Σ-homomorphism $h : (\mathcal{P}|_N, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ and a labeling $T(\pi) \subseteq \Delta^k$ for every $\pi \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$ such that:*

1. *for every $\mathbf{p} \in N$, we have $\mathbf{p} \in T(h(\mathbf{p}))$;*
2. *for every $\mathbf{c} \in T(\pi)$, we have $\mathbf{c} \to_\Sigma \mathsf{tp}_{\mathcal{U}_{\mathcal{T},\mathcal{A}}}(\pi)$;*
3. *for every $\pi \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$, every $\mathbf{c} \in T(\pi)$, and every $\mathbf{d}$ with $\mathbf{c} \hookrightarrow^{\mathcal{T},\mathcal{A}}_r \mathbf{d}$ one of the following is true:*
   - *(a) there is some $\pi' \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$ such that $(\pi, \pi') \in r^{\mathcal{U}_{\mathcal{T},\mathcal{A}}}$ and $\mathbf{d} \in T(\pi')$, or*
   - *(b) $(r, \mathbf{d}, T(\pi), \mathsf{tp}_{\mathcal{U}_{\mathcal{T},\mathcal{A}}}(\pi)) \in \mathsf{PHom}$.*

Proof. (⇒) Let $g : (\mathcal{P}, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_\mathcal{K}, \mathbf{b})$. We define $h = g|_N$ and, for $a \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$, we set

$$T(a) = \{\mathsf{tail}(\mathbf{p}) \mid \mathbf{p} \in \Delta^\mathcal{P} \wedge g(\mathbf{p}) = a\}.$$

We verify h and T satisfy Conditions 1–3.

- **Condition 1.** Let $\mathbf{p} \in N$. We have $\mathbf{p} \in h^{-1}(h(\mathbf{p})) \subseteq g^{-1}(h(\mathbf{p}))$, which implies $\mathbf{p} \in T(h(\mathbf{p}))$.
- **Condition 2.** Let $a \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$, $\mathbf{c} \in T(a)$ and $A \in \Sigma$. Since $\mathbf{c} \in T(a)$, there exists a $\mathbf{p} \in \Delta^{\mathcal{P}}$ such that $g(\mathbf{p}) = a$ and $\mathsf{tail}(\mathbf{p}) = \mathbf{c}$. Since $A \in \mathbf{c}$, by the definition of $\mathcal{P}$, we have $\mathbf{p} \in A^{\mathcal{P}}$. Since g is a Σ-homomorphism, $a = g(\mathbf{p}) \in A^{\mathcal{U}_{\mathcal{K}}}$.
- **Condition 3.** Let $a \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$, $\mathbf{c} \in T(a)$ and $\mathbf{d} \in \Delta^k$ with $\mathbf{c} \hookrightarrow_r^k \mathbf{d}$. Since $\mathbf{c} \in T(a)$, there exists a $\mathbf{p} \in \Delta^{\mathcal{P}}$ such that $g(\mathbf{p}) = a$ and $\mathsf{tail}(\mathbf{p}) = \mathbf{c}$. We distinguish two cases, depending on $b := g(\mathbf{p}r\mathbf{d})$:
  - If $b \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$, then it is clear that $(a, b) \in r^{\mathcal{U}_{\mathcal{K}}}$ and $\mathbf{d} \in T(b)$. In this case, Condition 3(a) holds.
  - If $b \notin \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$, we need to show that $(r, \mathbf{d}, T(a), \mathsf{tp}_{\mathcal{U}_{\mathcal{K}}}(a)) \in \mathsf{PHom}$. We define a partial function $\hat{g} : \Delta^{\mathcal{P}_{\mathbf{d}}} \to \Delta^{\mathcal{U}_{\mathcal{T},\mathsf{tp}_{\mathcal{U}_{\mathcal{K}}}(a)}}$ witnessing PHom by setting $\hat{g}(\mathbf{q}) := g(\mathbf{q})$, for every $\mathbf{q} \in \Delta^{\mathcal{P}}$ of the shape $\mathbf{q} = \mathbf{p}r\mathbf{d}\mathbf{p}'$ such that $g(\mathbf{q})$ lies in the subtree rooted at a. We argue that $\hat{g}$ fulfils all the necessary conditions: Since $\hat{g}$ is a restriction of g, it is a homomorphism on its domain. Since $g(\mathbf{p}) = a \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$ but $g(\mathbf{p}r\mathbf{d}) \notin \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$, we know that $g(\mathbf{p}r\mathbf{d})$ is a child of $g(\mathbf{p})$ in the anonymous part of $\mathcal{U}_{\mathcal{K}}$. It follows that $g(\mathbf{p}r\mathbf{d}) \in \Delta^{\mathcal{U}_{\mathcal{T},\mathsf{tp}_{\mathcal{U}_{\mathcal{K}}}(a)}}$ and thus, the second condition for PHom holds. For the last condition, let $\hat{g}(\mathbf{q}) \in \Delta^{\mathcal{P}_{\mathbf{d}}}$ be defined and let $(\mathbf{q}, \mathbf{q}') \in s^{\mathcal{P}_{\mathbf{c}}}$ for a Σ-role s. Now, assume that $\hat{g}(\mathbf{q}')$ is not defined. Since $\hat{g}(\mathbf{q})$ is defined, and from the definition of the domain of $\hat{g}$, we have that $\hat{g}(\mathbf{q}) = a$. This implies $g(\mathbf{q}) = a$, and thus, $\mathsf{tail}(\mathbf{q}) \in T(a)$.

($\Leftarrow$) Let $h : (\mathcal{P}|_N, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_{\mathcal{K}}, \mathbf{b})$ and $T : \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A}) \to 2^{\Delta^k}$ such that Condition 1 to 3 are fulfilled.

For constructing the homomorphism $g : (\mathcal{P}, \mathbf{a}^*) \to_\Sigma (\mathcal{U}_{\mathcal{K}}, \mathbf{b})$, we construct a series $g_0, g_1, \ldots$ of homomorphisms with increasing domains $N = \mathsf{dom}(g_0) \subseteq \mathsf{dom}(g_1) \subseteq \ldots$, such that every g_{i+1} extends g_i and $\bigcup_{i=0}^{\infty} \mathsf{dom}(g_i) = \Delta^{\mathcal{P}}$, and we will then set $g = \bigcup_{i=0}^{\infty} g_i$. While constructing the g_i, we will keep the following invariants about g_i.

- Every element from $\mathsf{dom}(g_i)$ is reachable from an individual in N. So $\mathsf{dom}(g_i) \subseteq \Delta^{\mathcal{P}}$ can be regarded as N plus a set of trees with roots in N.
- If $\mathbf{p} \in \mathsf{dom}(g_i)$ is a leaf in $\mathsf{dom}(g_i)$, but $\mathbf{p}$ has a successor in $\mathcal{P}$, then $\mathbf{p} \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$. So with every step from g_i to g_{i+1}, we define the homomorphism up to the points where it *comes back* to $\mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$.
- g_i is a Σ-homomorphism on its domain.

Set $g_0 = h$ and note that the invariants are fulfilled. If g_i has been defined and $\mathsf{dom}(g_i) \subsetneq \Delta^{\mathcal{P}}$, we define g_{i+1} in the following way: Choose a leaf $\mathbf{pc} \in \mathsf{dom}(g_i)$ such that there exists $\mathbf{pc}r\mathbf{d} \notin \mathsf{dom}(g_i)$ for some role r and some $\mathbf{d} \in \Delta^k$. Denote $g_i(\mathbf{pc}) = a$. From the second invariant, we know that $\mathbf{c} \in T(a)$, so the requirements for Condition 3 are fulfilled and we know that either 3(a) is true or 3(b) is true. If 3(a) is true, then let $\mathsf{dom}(g_{i+1}) = \mathsf{dom}(g_i) \cup \{\mathbf{pc}r\mathbf{d}\}$ and define $g_{i+1}(\mathbf{pc}r\mathbf{d}) = b$, where b is the element from Condition 3 (a). Note that then the invariants hold for g_{i+1}. If 3 (b) is true, then $(r, \mathbf{d}, T(a), \mathsf{tp}_{\mathcal{U}_{\mathcal{K}}}(a)) \in \mathsf{PHom}$, so there is a partial function $h' : \Delta^{\mathcal{P}_{\mathbf{d}}} \to \Delta^{\mathcal{U}_{\mathcal{T},\mathsf{tp}_{\mathcal{U}_{\mathcal{K}}}(a)}}$ fulfilling the conditions of the definition of PHom. In $\mathsf{dom}(h')$, consider the connected component $Z \subseteq \mathsf{dom}(h')$ of $\mathbf{d}$. For every $\mathbf{dq} \in Z$, where $\mathbf{q}$ could possibly be empty, we define $g_{i+1}(\mathbf{pc}r\mathbf{dq}) = h'(\mathbf{dq})$ and we obtain $\mathsf{dom}(g_{i+1}) = \mathsf{dom}(g_i) \cup \{\mathbf{pc}r\mathbf{dq} \mid \mathbf{dq} \in Z\}$. Since Z is connected, the first invariant holds. From the third condition of the definition of PHom, we know that whenever $\mathbf{dq}$ is a leaf in Z that has a successor in $\mathcal{P}$, then $g_{i+1}(\mathbf{dq}) = a$ and $\mathsf{tail}(\mathbf{pc}r\mathbf{dq}) \in T(a)$, so the second invariant holds. Since g_i is a Σ-homomorphism on its domain and h' is a Σ-homomorphism on Z, it follows that g_{i+1} is a Σ-homomorphism on its domain, aside from possibly the leafs in Z. But from Condition 2, it follows that g_{i+1} is also a homomorphism on the leafs, so the third invariant holds.

If the leafs $\mathbf{pc}$ are chosen in a fair way, i.e. ensuring that every such leaf gets chosen at some point, it follows that $\bigcup_{i=0}^{\infty} \mathsf{dom}(g_i) = \Delta^{\mathcal{P}}$. From the third invariant, it follows then that g is a Σ-homomorphism from $(\mathcal{P}, \mathbf{a}^*)$ to $(\mathcal{U}_{\mathcal{K}}, \mathbf{b})$. ☐

Lemma 6. $(\hat{r}, \hat{\mathbf{t}}, \hat{T}, \hat{t}) \in \mathsf{PHom}$ *iff* $\hat{\mathfrak{M}}$ *contains a root mosaic* $M = (t, T, r_0, t_0, T_0, \ldots, r_n, t_n, T_n)$ *for* $\hat{T}$ *with* $t = \hat{t}$ *and a mosaic* $M' = (t', T', r_0', t_0', T_0', \ldots, r_m', t_m', T_m')$ *with* $\hat{\mathbf{t}} \in T'$ *and* $r_0 = \hat{r}$ *such that* $t = t_0'$, $r_i = r_0'$, $t_i = t'$, $T_0' \subseteq T$, *and* $T_i \subseteq T'$ *for some* i.

Proof. For the direction ($\Rightarrow$), let g be the function witnessing $(\hat{r}, \hat{\mathbf{t}}, \hat{T}, \hat{t}) \in \mathsf{PHom}$. We define a set $\mathfrak{M}^*$ of mosaics by including a mosaic $M(\pi) = (t, T, r_0, t_0, T_0, \ldots, r_n, t_n, T_n)$ for every $\pi \in \Delta^{\mathcal{U}_{\mathcal{T},\hat{t}}}$. There, $M(\pi)$ is given by:

- $t = \mathsf{tail}(\pi)$;
- $T = \{\mathsf{tail}(\mathbf{p}) \mid g(\mathbf{p}) = \pi\}$.
- r_0 is arbitrary and $T_0 = \emptyset$, if $\pi = a_{\hat{t}}$; otherwise let $\pi = \pi' rt$ and define $r_0 = r$ and $T_0 = \{\mathsf{tail}(\mathbf{p}) \mid g(\mathbf{p}) = \pi'\}$;
- r_i, t_i for $1 \le i \le n$ are determined by requiring that $\pi r_1 t_1, \ldots, \pi r_n t_n$ be all successors of π;
- T_i for $1 \le i \le n$ is given by $T_i = \{\mathsf{tail}(\mathbf{p}) \mid g(\mathbf{p}) = \pi r_i t_i\}$.

By construction, $M(a_{\hat{t}})$ is a root mosaic for $\hat{T}$, and all other $M(\pi)$ are mosaics. Moreover, there is a successor of the shape $\pi = a_{\hat{t}}\hat{t} \in \Delta^{\mathcal{U}_{\mathcal{T},\hat{t}}}$, such that $g(\hat{\mathbf{t}}) = \pi$. Define $M = M_{\hat{t}}$ and $M' = M_\pi$. By construction, they satisfy the conditions from the lemma. Moreover, using the homomorphism condition of g, it is routine to verify that $\mathfrak{M}^* \subseteq \mathfrak{M}_i$ for all $i \ge 0$.

For the direction ($\Leftarrow$), let $\hat{\mathfrak{M}}$ be the result of mosaic elimination procedure, and let M and M' be the mosaics that are guaranteed to exist by assumption. Before defining the function g, let us define a map $\tau : \Delta^{\mathcal{U}_{\mathcal{T},\hat{t}}} \to \hat{\mathfrak{M}}$ which assigns every node in $\Delta^{\mathcal{U}_{\mathcal{T},\hat{t}}}$ a mosaic. Start with $\tau(a_{\hat{t}}) = M$ and $\tau(a_{\hat{t}}\hat{r}t') = M'$. By Condition (E), we can continue defining $\tau(\pi st)$ whenever $\tau(\pi)$ is defined.

We define now the homomorphism g inductively, maintaining the following invariants:

- $(*)$ if $g(\mathbf{p}) = \pi$ is defined and $\tau(\pi) = (t, T, r_0, t_0, T_0, \ldots, r_n, t_n, T_n)$, then $\mathsf{tail}(\mathbf{p}) \in T$, and
- $(**)$ g is a homomorphism on its domain.

Start with setting $g(\hat{\mathbf{t}})$ to the successor $\pi = \hat{\pi}\hat{r}t$ that exists because of Condition 2 of strategy tree, obviously satisfying the invariant. To extend g, let $g(\mathbf{p}) = \pi$ defined, $(\mathbf{p}, \mathbf{p}') \in r^{\mathcal{P}_c}$ for some Σ-role r. If $\pi = \hat{\pi}$ and $\mathsf{tail}(\mathbf{p}) \in \hat{T}$, we do nothing. Otherwise, suppose $\tau(\pi) = (t, T, r_0, t_0, T_0, \ldots, r_n, t_n, T_n)$. By $(*)$, we have that $\mathsf{tail}(\mathbf{p}) \in T$. Moreover, we know that $\mathsf{tail}(\mathbf{p}) \hookrightarrow_r^{\mathcal{T},\mathcal{A}} \mathsf{tail}(\mathbf{p}')$. Condition (vi) (resp., (vi')) yields that $\mathbf{t} \in T_i$ for some i and either $i = 0$ and $r = r_0^-$ or $i > 0$ and $r = r_i$. Extend g by setting $g(\mathbf{p}')$ to the predecessor of π if $i = 0$ and to $\pi r_i t_i$ if $i > 0$. □

Lemma 8. *For $\mathcal{L} = \mathcal{ELI}$ and $\mathcal{Q} \in \{CQ, UCQ\}$, $\mathsf{QBE}(\mathcal{L},\mathcal{Q})$ and $\mathsf{QDEF}(\mathcal{L},\mathcal{Q})$ are 2-ExpTime-hard.*

Proof. We reduce the word problem for the following variant of exponential space bounded alternating Turing machines (ATMs) which is easily seen to be equivalent to the standard model. An *alternating Turing machine* is a tuple $M = (Q, \Theta, \Gamma, q_0, \Delta)$ where $Q = Q_\exists \uplus Q_\forall \uplus \{q_a, q_a^1, q_a^2\}$ is the set of states, consisting of *existential states* $Q_\exists$, *universal states* $Q_\forall$, and some accepting states q_a, q_a^1, q_a^2. Further, Θ is the input alphabet and Γ is the tape alphabet, $q_0 \in Q_\exists \cup Q_\forall$ is the *starting state*, and the *transition relation* Δ is of the form

$$\Delta \subseteq Q \times \Gamma \times Q \times \Gamma \times \{L, N, R\}.$$

We write $\Delta(q, \sigma)$ to denote the set $\{(q', \sigma', M) \mid (q, \sigma, q', \sigma', M) \in \Delta\}$ and assume w.l.o.g that the state q_0 is universal and that every set $\Delta(q, \sigma)$ contains exactly two elements when q is universal; moreover, we assume a fixed order on these two elements, speaking about the *first* and the *second possible successor*. Finally, we assume that $\Delta(q_a, \sigma) = \{(q_a^1, \sigma, N), (q_a^2, \sigma, N)\}$ and $\Delta(q_a^i, \sigma) = \{(q_a, \sigma, N)\}$ for all $\sigma \in \Gamma$.

A *configuration* of an ATM is a word wqw' with $w, w' \in \Gamma^*$ and $q \in Q$. The *successor configurations* of a configuration wqw' are defined in the usual way.

A *computation tree* of an ATM M on input w is a tree whose nodes are labeled with configurations of M such that

- the root is labeled with the initial configuration q_0w, and
- the descendants of any inner node which is labeled by a universal (resp., existential) configuartion include all (resp., one) of the successors of that configuration.

A computation tree is called *accepting* if on all paths in the tree, the sequence of visited states is of the shape $u \cdot (q_a q_a^1)^\omega$ or $u \cdot (q_a q_a^2)^\omega$, that is, eventually M loops through the accepting states. An ATM M accepts an input w if there is a computation tree of M on w.

It is well-known that there is a fixed exponentially space bounded M whose word problem is 2-ExpTime-hard such that every configuration wqw' appearing in some computation tree actually satisfies $|ww'| \leq 2^n$ with n the length of the input [Chandra *et al.*, 1981]. Moreover, we assume that Γ contains special symbols $\not{b}, \not{e}$ which delimit the input word from left and right and are never overwritten, and further the *blank symbol* $\_$.

We are going to construct a TBox $\mathcal{T}$ including two special concept names A, B, and an alphabet Σ such that

$$(\mathcal{T}, \{A(a), B(b)\}, \{a\}, \{b\}, \Sigma) \in \mathsf{QBE} \quad \text{iff} \quad M \text{ accepts } w.$$

We enforce an exponential counter using concept names $X_1, \ldots, X_n$, $\overline{X}_1, \ldots, \overline{X}_n$ and $C_1, \ldots, C_n$, $\overline{C}_n, \ldots, \overline{C}_n$ as follows:

$$\begin{aligned}
\exists r^-.X_1 &\sqsubseteq \overline{X}_1 \sqcap C_1 \\
\exists r^-.\overline{X}_1 &\sqsubseteq X_1 \sqcap \overline{C}_1 \\
C_{i-1} \sqcap \exists r^-.X_i &\sqsubseteq \overline{X}_i \sqcap C_i, \quad \text{for all } 2 \leq i \leq n \\
C_{i-1} \sqcap \exists r^- \overline{X}_i &\sqsubseteq X_i \sqcap \overline{C}_i, \quad \text{for all } 2 \leq i \leq n \\
\overline{C}_{i-1} \sqcap \exists r^-.X_i &\sqsubseteq X_i \sqcap \overline{C}_i, \quad \text{for all } 2 \leq i \leq n \\
\overline{C}_{i-1} \sqcap \exists r^-.\overline{X}_i &\sqsubseteq \overline{X}_i \sqcap \overline{C}_i, \quad \text{for all } 2 \leq i \leq n
\end{aligned}$$

Moreover, we include the following CIs to have convenient abbreviations available:

$$\begin{aligned}
X_1 \sqcap \ldots \sqcap X_n &\sqsubseteq \mathsf{Max} \\
\overline{X}_i &\sqsubseteq \overline{\mathsf{Max}} \quad \text{for all } 1 \leq i \leq n
\end{aligned}$$

We start with CIs that enforce the skeleton tree below a, as described in the main part:

$$\begin{aligned}
A &\sqsubseteq \overline{X}_1 \sqcap \ldots \sqcap \overline{X}_n \sqcap U \\
U \sqcap \overline{\mathsf{Max}} &\sqsubseteq \exists r.U \\
U \sqcap \mathsf{Max} &\sqsubseteq \exists r.E_1 \sqcap \exists r.E_2 \\
E_j \sqcap \overline{\mathsf{Max}} &\sqsubseteq \exists r.E_j, \quad \text{for } j \in \{1, 2\} \\
E \sqcap \mathsf{Max} &\sqsubseteq \exists r.U
\end{aligned}$$

For enforcing the paths which starting from every element, let $\mathbf{X}$ denote the set of all concept names of the shape $\alpha_{\langle x,y,z \rangle}$ with $x, y, z \in \Gamma \cup (\Gamma \times Q)$ such that at most one of x, y, z is actually from $(\Gamma \times Q)$. Intuitively, $\alpha_{\langle x,y,z \rangle}$ describes three consecutive tape cells in a configuration. We include the following transition for all $\alpha, \alpha' \in \mathbf{X}$ and $j \in \{1, 2\}$:[1]

$$E_j \sqsubseteq \exists r.\exists r^-.(\alpha \sqcap \exists (r^-)^{2^n}.\alpha') \tag{1}$$

$$U \sqsubseteq \exists r.\exists r^-.(\alpha \sqcap \exists (r^-)^{2^n}.\alpha')) \tag{2}$$

This finishes the first part, that is, the tree below a.

For the second part, we use the same counters and enforce that every model of B contains in some sense *all possible computation trees*. For this purpose, let c_0 be the initial configuration of M on input w, and let α_i be $X_{\langle x,y,z \rangle}$ if the three cells $i-1$, i, $i+1$ are labeled with x, y, z, respectively. To enforce the initial configuration, we include the following CI:

$$\begin{aligned}
B &\sqsubseteq U \sqcap \overline{X}_1 \sqcap \ldots \overline{X}_n \sqcap \mathsf{Init} \\
\mathsf{Init} \sqcap (X = i) &\sqsubseteq \exists r.(U \sqcap \mathsf{Init} \sqcap \alpha_i) \quad \text{for all } i \leq n \\
\mathsf{Init} \sqcap (X > n) \sqcap \overline{\mathsf{Max}} &\sqsubseteq \exists r.(U \sqcap \mathsf{Init} \sqcap \alpha_{\langle _,_,\not{e} \rangle})
\end{aligned}$$

[1] $\exists r^k.C$ denotes $\exists r. \ldots \exists r.C$ with a sequence of k times $\exists r$.

To enforce the remainder of the tree, let $\mathbf{X}_E$ (resp., $\mathbf{X}_{U_1}$ and $\mathbf{X}_{U_2}$) as the set of all pairs (α_0, α_1') such that $\alpha_0, \alpha_1 \in \mathbf{X}$, and there are two configurations c, c' of M such that c is existential (resp., universal), c' is a successor configuration of c (where the first, resp., second possibility from the universal transition was chosen), some cell in c is labeled with α, and the same cell in c' is labeled with α'. Moreover, denote with $\mathbf{X}_*^{\not y}$ (resp. $\mathbf{X}_*^{\not q}$) the subset of $\mathbf{X}_*$ that contains only (α_0, α_1') with α of the shape $X_{\langle \not y, y, z\rangle}$ (resp., $X_{\langle x, y, \not q\rangle}$). Additionally, introduce concept names S_j, $j \in \{0, 1, 2\}$ where, intuitively, the subscript j denotes the number of states seen so far, and a function $st : \mathbf{X} \to \{0, 1\}$ defined as follows: $st(\alpha_{\langle x,y,z\rangle}) = 1$ if y contains a state, and 0 otherwise. We include the following CIs to start:

$$\begin{aligned}\mathsf{Init} \sqcap \mathsf{Max} \sqsubseteq & \bigsqcap_{(\alpha_0,\alpha_1')\in\mathbf{X}^0_{U_1}} \exists r.(E_1 \sqcap \alpha_0 \sqcap \alpha_1' \sqcap S_{st(\alpha)}) \sqcap \\ & \bigsqcap_{(\alpha_0,\alpha_1')\in\mathbf{X}^0_{U_2}} \exists r.(E_2 \sqcap \alpha_0 \sqcap \alpha_1' \sqcap S_{st(\alpha)})\end{aligned}$$

For $j \in \{1, 2\}$ and $i \in \{0, 1\}$ we include the following CIs:

$$\begin{aligned}E_j \sqcap S_i \sqcap \overline{\mathsf{Max}} \sqcap \alpha_0 \sqcap \alpha_1' \sqsubseteq & \\ & \bigsqcap_{\substack{(\beta_0,\beta_1')\in\mathbf{X}_{U_j}\\(\alpha_0,\alpha_1')\sim(\beta_0,\beta_1')}} \exists r.(E_j \sqcap \beta_0 \sqcap \beta_1' \sqcap S_{i+st(\beta_0)}) \\ U \sqcap S_i \sqcap \overline{\mathsf{Max}} \sqcap \alpha_0 \sqcap \alpha_1 \sqsubseteq & \\ & \bigsqcap_{\substack{(\beta_0,\beta_1')\in\mathbf{X}_{E}\\(\alpha_0,\alpha_1')\sim(\beta_0,\beta_1')}} \exists r.(E_j \sqcap \beta_0 \sqcap \beta_1' \sqcap S_{i+st(\beta_0)}),\end{aligned}$$

where $(\alpha_0, \alpha_1') \sim (\beta_0, \beta_1')$ if $\alpha_0 = \alpha_{\langle x,y,z\rangle}$, $\beta_0 = \alpha_{\langle y,z,z'\rangle}$, $\alpha_1 = \alpha_{\langle x',y',z'\rangle}$, and $\beta_1 = \alpha_{\langle y',z',z''\rangle}$, that is, both α_0, β_0 and α_1, β_1 can be labels of adjacent cells in a configuration. Note that there are no CIs with S_2 on the left-hand side because there would be too many states.

Finally, to realize the change between existential and universal configurations, we include the following CIs for every $(\alpha_0, \alpha_1) \in \mathbf{X}_*^{\not q}$ and $j \in \{1, 2\}$.

$$\begin{aligned}E_j \sqcap S_1 \sqcap \mathsf{Max} \sqcap \alpha_0 \sqcap \alpha_1' \sqsubseteq & \\ & \bigsqcap_{(\beta_0,\beta_1')\in\mathbf{X}^0_{E}} \exists r.(U \sqcap \beta_0 \sqcap \beta_1' \sqcap S_{st(\beta_0)}) \\ U \sqcap S_1 \sqcap \mathsf{Max} \sqcap \alpha_0 \sqcap \alpha_1' \sqsubseteq & \\ & \bigsqcap_{(\beta_0,\beta_1')\in\mathbf{X}^0_{U_1}} \exists r.(E_1 \sqcap \beta_0 \sqcap \beta_1' \sqcap S_{st(\beta_0)}) \sqcap \\ & \bigsqcap_{(\beta_0,\beta_1')\in\mathbf{X}^0_{U_2}} \exists r.(E_2 \sqcap \beta_0 \sqcap \beta_1' \sqcap S_{st(\beta_0)})\end{aligned}$$

Here, we include only transitions for S_1 since a valid configuration includes *exactly one* state. To generate the outgoing paths, we include the following CIs for every $(\alpha_0, \alpha_1') \in \mathbf{X}$:

$$\begin{aligned}\alpha_0 \sqcap \alpha_1' \sqsubseteq & \bigsqcap_{(\beta_0,\beta_1)\in\mathbf{X}^2\setminus\{(\alpha_0,\alpha_1)\}} \exists r.\exists r^-.(\beta_0 \sqcap \exists (r^-)^{2^n}.\beta_1) \\ \mathsf{Init} \sqsubseteq & \bigsqcap_{(\beta_0,\beta_1)\in\mathbf{X}^2} \exists r.\exists r^-.(\beta_0 \sqcap \exists (r^-)^{2^n}.\beta_1)\end{aligned}$$

Note that, at non-Init-nodes, the same outgoing paths are created as in (1) and (2), except for the current labeling. This is then responsible for verifying that the configuration has changed accordingly.

Using the given intuitions, it is now routine to verify the following claim for $\Sigma = \{r, U, E_1, E_2\} \cup \mathbf{X}$.

Claim.

$(\mathcal{T}, \{A(a), B(b)\}, \{a\}, \{b\}, \Sigma) \in \mathsf{QBE}$ iff M accepts w.

Since the instance $(\mathcal{T}, \{A(a), B(b)\}, \{a\}, \{b\})$, in particular $\mathcal{T}$, can be constructed in polynomial time from M, we obtain the desired result. □

Theorem 10. *For $\mathcal{L}$ = Horn-$\mathcal{ALCI}$, $\mathsf{QBE}_f(\mathcal{L}, UCQ)$ and $\mathsf{QDEF}_f(\mathcal{L}, UCQ)$ are* ExpTime*-complete.*

Proof. The lower bound follows from Lemma 17 below. For the upper bound, we consider only TBoxes in normal form. For TBoxes not in normal form, the definition of universal model has to adapted, but the proof principle remains the same.

As mentioned in the main part, we have $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \to (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ iff $(\mathcal{U}_{\mathcal{T},\mathcal{A}}|_{\mathsf{ind}(\mathcal{A})}, \mathbf{a}) \to (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$. To decide the latter, we proceed as follows. First, compute $\mathcal{U}_{\mathcal{T},\mathcal{A}}|_{\mathsf{ind}(\mathcal{A})}$. Then enumerate all possible maps h from $\mathsf{ind}(\mathcal{A})$ to $\Delta^{\mathcal{U}_{\mathcal{T},\mathcal{A}}}$ with $h(\mathbf{a}) = \mathbf{b}$ (domain elements disconnected from $\mathbf{a}$ can be ignored, as for such elements the identity establishes a homomorphism). We accept if one of the maps is actually a homomorphism. It remains to note that the computation of the universal interpretation can be done in exponential time, there are only exponentially many such homomorphisms, and the checks can be done in exponential time. □

D.1 Proofs for Horn-$\mathcal{ALC}$

Lemma 11. *Given $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$ with $(\mathcal{T}, \mathcal{A})$ a Horn-$\mathcal{ALC}$ KB, the relation* PHom *can be decided in* ExpTime.

We show that PHom can be decided using the following relation $\mathsf{Hom} \subseteq \Delta^k \times \mathsf{TP}$. For a pair $(\mathbf{c}, t) \in \Delta^k \times \mathsf{TP}$ we have $(\mathbf{c}, t) \in \mathsf{Hom}$ if there is a Σ-homomorphism $(\mathcal{P}_\mathbf{c}, \mathbf{c}) \to_\Sigma (\mathcal{U}_{\mathcal{T},t(a)}, a)$. By $\mathcal{U}_{\mathcal{T},t(a)}$ we denote the structure $\mathcal{U}_{\mathcal{T},t}$ where the root is called a.

Lemma 11 follows then from the following two Lemmas.

Lemma 15. *Let r be a Σ-role, $\mathbf{c} \in \Delta^k$, $T \subseteq \Delta^k$ and $t \in \mathsf{TP}$. Then we have $(r, \mathbf{c}, T, t) \in \mathsf{PHom}$ iff there exists a $t' \in \mathsf{TP}$ such that $t \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} t'$ and $(\mathbf{c}, t') \in \mathsf{Hom}$.*

Proof. For the first direction, let $(r, \mathbf{c}, T, t) \in \mathsf{PHom}$ and let $g : \Delta^{\mathcal{P}_\mathbf{c}} \to \Delta^{\mathcal{U}_{\mathcal{T},t}}$ a partial Σ-homomorphism witnessing that. From the second condition for PHom it follows that the root of $\mathcal{U}_{\mathcal{T},t}$ has an r-successor b such that $g(\mathbf{c}) = b$. Let t' be the type of b. Since $\mathcal{T}$ is formulated in Horn-$\mathcal{ALC}$, all roles in both $\mathcal{P}_\mathbf{c}$ and $\mathcal{U}_{\mathcal{T},t}$ are pointing away from the root, and it can be shown by induction on the length on the path $\mathbf{p}$, that $g(\mathbf{p})$ is defined for all $\mathbf{p} \in \Delta^{\mathcal{P}_\mathbf{c}}$. Thus, g is a Σ-homomorphism from $\mathcal{P}_\mathbf{c}$ to $\mathcal{U}_{\mathcal{T},t'(b)}$ which maps $\mathbf{c}$ to b, and hence, $(\mathbf{c}, t') \in \mathsf{Hom}$.

For the other direction, consider any tuple $(r, \mathbf{c}, T, t)$ and let there be a $t' \in \mathsf{TP}$ such that $t \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} t'$ and $\mathsf{Hom}(\mathbf{c}, t')$. We aim to show that $(r, \mathbf{c}, T, t) \in \mathsf{PHom}$. Since $\mathsf{Hom}(\mathbf{c}, t')$, there exists $h : (\mathcal{P}_\mathbf{c}, \mathbf{c}) \to_\Sigma (\mathcal{U}_{\mathcal{T},t'(a)}, a)$. This h is a witness for $(r, \mathbf{c}, T, t) \in \mathsf{PHom}$. □

Lemma 16. *Given $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$, where $\mathcal{T}$ is formulated in Horn-$\mathcal{ALC}$, the problem of deciding* Hom *is in* ExpTime.

Proof. We give an ExpTime procedure for computing the relation Hom. We define a sequence of relations $\mathsf{Hom}_i \subseteq \Delta^k \times \mathsf{TP}$ that approximates Hom from above. Let $\mathbf{c} = (\pi_1, \dots, \pi_k)$. By $\mathsf{tp}(\pi_i)$ we denote the set of concept names that are true at π_i, i.e. if $\pi_i = a$ for some $a \in \mathsf{ind}(\mathcal{A})$, then $\mathsf{tp}(\pi_i) = \mathsf{tp}_{\mathcal{T},\mathcal{A}}(a)$ and otherwise, $\mathsf{tp}(\pi_i) = \mathsf{tail}(\pi_i)$. Let Hom_0 contain all pairs $(\mathbf{c}, t)$ such that $\bigcap_{i=1}^k \mathsf{tp}(\pi_i) \subseteq t$. A pair $(\mathbf{c}, t)$ is in Hom_{i+1} if it is in Hom_i and for every $\mathbf{d} \in \Delta^k$ and $r \in \Sigma$ with $\mathbf{c} \hookrightarrow_r^{\mathcal{T},\mathcal{A}} \mathbf{d}$ there exists a type t' such that $t \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} t'$ and $(\mathbf{d}, t') \in \mathsf{Hom}_i$.

Claim: $\mathsf{Hom} = \bigcap_{i=0}^\infty \mathsf{Hom}_i$.

Proof of the Claim. ($\subseteq$): First we show by induction on i:

$$\mathsf{Hom} \cap \mathsf{Hom}_i \subseteq \mathsf{Hom}_{i+1} \qquad (*)$$

– We begin by showing $\mathsf{Hom} \subseteq \mathsf{Hom}_0 \cap \mathsf{Hom}_1$, which implies the induction start. Let $(\mathbf{c}, t) \in \mathsf{Hom}$, so there exists $h : (\mathcal{P}_\mathbf{c}, \mathbf{c}) \to_\Sigma (\mathcal{U}_{\mathcal{T},t(a)}, a)$. Since $h(\mathbf{c}) = t$, we have for every concept name $A \in \Sigma$ that $\mathbf{c} \in A^{\mathcal{P}_\mathbf{c}}$ implies $A \in t$ and it follows that $(\mathbf{c}, t) \in \mathsf{Hom}_0$. Let $r \in \Sigma$ and $\mathbf{d} \in \Delta^k$ such that $\mathbf{c} \hookrightarrow_r^{\mathcal{T},\mathcal{A}} \mathbf{d}$. Since h is defined for $\mathbf{d}$, it follows that there is a $t' \in \mathsf{TP}$ with $t \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} t'$ and it follows that $(\mathbf{d}, t') \in \mathsf{Hom}_0$. Therefore, $(\mathbf{c}, t) \in \mathsf{Hom}_1$.

– For the induction step, assume $\mathsf{Hom} \cap \mathsf{Hom}_i \subseteq \mathsf{Hom}_{i+1}$. We need to show that $\mathsf{Hom} \cap \mathsf{Hom}_{i+1} \subseteq \mathsf{Hom}_{i+2}$. Let $(\mathbf{c}, t) \in \mathsf{Hom} \cap \mathsf{Hom}_{i+1}$. Thus, for every $r \in \Sigma$ and $\mathbf{d} \in \Delta^k$ with $\mathbf{c} \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} \mathbf{d}$, there exists a $t_\mathbf{d} \in \mathsf{TP}$ such that $t \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} t_\mathbf{d}$ and $(\mathbf{d}, t_\mathbf{d}) \in \mathsf{Hom}_i$. But we also have $(\mathcal{P}_\mathbf{d}, \mathbf{d}) \to_\Sigma (\mathcal{U}_{\mathcal{T},t_\mathbf{d}(a)}, a)$ by restricting h to the subtree rooted at $\mathbf{d}$, and thus, $(\mathbf{d}, t_\mathbf{d}) \in \mathsf{Hom}$. Using the induction hypothesis, we conclude that $(\mathbf{d}, t_\mathbf{d}) \in \mathsf{Hom}_{i+1}$. Hence, $(\mathbf{c}, t) \in \mathsf{Hom}_{i+2}$, which finishes the proof of (*).

Now we can show that $\mathsf{Hom} \subseteq \mathsf{Hom}_i$ for all i, again by induction on i. The case $i = 0$ has been shown above, and the induction step follows from (*).

($\supseteq$): Let $(\mathbf{c}, t) \in \bigcap_{i=0}^\infty \mathsf{Hom}_i$. We construct a homomorphism $h : (\mathcal{P}_\mathbf{c}, \mathbf{c}) \to_\Sigma (\mathcal{U}_{\mathcal{T},t(a)}, a)$ level by level, i.e. we inductively define a sequence $h_j : (\mathcal{P}_\mathbf{c}|_j, \mathbf{c}) \to_\Sigma (\mathcal{U}_{\mathcal{T},t(a)}, a)$, where $\mathcal{P}_\mathbf{c}|_j$ denotes the restriction of $\mathcal{P}_\mathbf{c}$ to the first j levels. While constructing the h_j, we will keep the following invariants:

- h_j is a Σ-homomorphism on its domain.
- For every leaf $\mathbf{p}$ of $\mathcal{P}_\mathbf{c}|_j$ we have $(\mathsf{tail}(\mathbf{p}), \mathsf{tp}(h_j(\mathbf{p})) \in \bigcap_{i=0}^\infty \mathsf{Hom}_i$.

We have $\mathsf{dom}(h_0) = \{\mathbf{c}\}$ and set $h_0(\mathbf{c}) = a$. Note that h_0 is a homomorphism, since $(\mathbf{c}, t) \in \mathsf{Hom}_0$ and the second invariant is also true. Now assume h_j has been defined. To define h_{j+1}, consider any leaf $\mathbf{p}r\mathbf{d} \in \Delta^{\mathcal{P}_\mathbf{c}|_{j+1}}$. From the second invariant we know that $(\mathbf{p}, \mathsf{tp}(h_j(\mathbf{p}))) \in \mathsf{Hom}_i$ for all i. So for every i we have a type t_i such that $(\mathbf{d}, t_i) \in \mathsf{Hom}_i$. Since there are only finitely many types, there must be a type t' that appears infinitely many times among the t_i. Clearly $t \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} t'$, so we can choose this type as the image of $\mathbf{d}$, i.e. we set $h_{j+1}(\mathbf{p}r\mathbf{d}) = h_j(\mathbf{p})rt'$. Since t' appeared infinitely many times in the sequence t_i, and since the sequence Hom_i is descending, we have $(\mathbf{d}, t') \in \bigcap_{i=0}^\infty \mathsf{Hom}_i$, i.e. the second invariant holds for h_{j+1}. The first invariant follows from the fact that $(\mathbf{d}, t') \in \mathsf{Hom}_0$ and because h_j is a Σ-homomorphism on its domain.

Finally, we set $h = \bigcup_{i=0}^\infty h_j$. Since $\mathcal{P}_\mathbf{c}$ is connected, we have $\mathsf{dom}(h) = \Delta^{\mathcal{P}_\mathbf{c}}$ and the first invariant assures that h is a Σ-homomorphism $(\mathcal{P}_\mathbf{c}, \mathbf{c}) \to_\Sigma (\mathcal{U}_{\mathcal{T},t(a)}, a)$, so $(\mathbf{c}, t) \in \mathsf{Hom}$. This finishes the proof of the claim.

Now we argue the time complexity of computing Hom. Since $\mathsf{Hom}_{i+1} \supseteq \mathsf{Hom}_i$ for all i and since Hom_0 contains at most $|\Delta^k \times \mathsf{TP}|$ many elements, which is single exponential measured in the size of $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma)$, the sequence Hom_i stabilizes after exponentially many steps. Computing Hom_0 and computing Hom_{i+1} from Hom_i can also be done in exponential time. Hence, Hom can be decided in ExpTime. □

Theorem 12. *For $\mathcal{L}$ = Horn-$\mathcal{ALC}$,* $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ *and* $\mathsf{QDEF}(\mathcal{L}, \mathcal{Q})$ *are* coNExpTime*-complete if $\mathcal{Q}$ = CQ and* ExpTime*-complete if $\mathcal{Q}$ = UCQ. All results also hold with unrestricted signature.*

The missing statement is the ExpTime-completeness, which follows from the following two lemmas.

Lemma 17. *Both* QBE_f*(Horn-$\mathcal{ALC}$, UCQ) and* QDEF_f*(Horn-$\mathcal{ALC}$, UCQ) are* ExpTime*-hard.*

Proof. We reduce from (the complement of) the subsumption problem in Horn-$\mathcal{ALC}$, which is the problem of deciding, given a Horn-$\mathcal{ALC}$-TBox $\mathcal{T}$ and concept names A, B, whether $\mathcal{T} \not\models A \sqsubseteq B$. The problem is ExpTime-complete [Krötzsch *et al.*, 2013].

Given a TBox $\mathcal{T}$ and concept names A, B, define an input $(\mathcal{T}, \mathcal{A}, S^+, S^-)$ for QBE_f, by taking $\mathcal{A} = \{A(a), B(b)\}$, $S^+ = \{b\}$ and $S^- = \{a\}$. We now argue correctness of the reduction. It is easy to see that $\mathcal{T} \models A \sqsubseteq B$ iff there is a homomorphism $h : \mathcal{U}_{\mathcal{T},\{B(b)\}} \to \mathcal{U}_{\mathcal{T},\{A(a)\}}$ with $h(b) = a$. By Theorem 2, this is that case iff $(\mathcal{T}, \mathcal{A}, S^+, S^-) \notin \mathsf{QBE}_f$. (Note that the constructed instance is always safe and that both S^+ and S^- contain only one individual each.) Thus, we have $\mathcal{T} \not\models A \sqsubseteq B$ iff $(\mathcal{T}, \mathcal{A}, S^+, S^-) \in \mathsf{QBE}_f$.

The ExpTime-hardness of QDEF_f can be achieved by the same reduction, just ignoring S^-. □

Lemma 18. *Both* QBE*(Horn-$\mathcal{ALC}$, UCQ) and* QDEF*(Horn-$\mathcal{ALC}$, UCQ) are in* ExpTime.

Proof. We only describe an exponential time algorithm for QBE(Horn-$\mathcal{ALC}$, UCQ); the arguments for QDEF are essentially the same. By Theorem 1, we need to check Σ-safety of $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a})$ for every $\mathbf{a} \in S^+$ and whether $(\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{a}) \not\to (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ for all $\mathbf{a} \in S^+$ and $b \in S^-$. The former can be checked in ExpTime by Lemma 3. For the latter, we loop over all (polynomially many) pairs $(\mathbf{a}, \mathbf{b}) \in S^+ \times S^-$ and use Lemma 5 for every such pair individually, that is, in the definition of $\mathcal{P}$ we use $k = 1$. In this case, there are only single exponentially many candidates for the homomorphism h and the labeling T, so we can loop over all possible h and T and check Conditions 1 to 3. Clearly, Conditions 1, 2 and

3a can be checked in EXPTIME, whereas 3b can be checked in EXPTIME by Lemma 11. $\square$

E Proofs for Section 5

Lemma 13. *There is a family of Horn-$\mathcal{ALC}$ knowledge bases $(\mathcal{T}_n, \mathcal{A}_n)_{n\geq 1}$, sets of examples S^+ and S^-, a signature Σ, and a polynomial $p(n)$ such that, for all $n \geq 1$, $|\mathcal{T}_n \cup \mathcal{A}_n| \leq p(n)$, $(\mathcal{T}_n, \mathcal{A}_n, S^+, S^-, \Sigma) \in \mathsf{QBE}($Horn-$\mathcal{ALC}$, (U)CQ) and every (U)CQ witnessing this is of size $\Omega(2^{2^n})$.*

Proof. Given some $n \geq 1$, we construct a Horn-$\mathcal{ALC}$ KB $(\mathcal{T}_n, \mathcal{A}_n)$ using concept names $A, B, G, H, U, X_1, \overline{X}_1, C_1, \overline{C}_1, \dots, X_n, \overline{X}_n, \underline{C_1}, \overline{C}_n$ and a single role name r. The concept names $X_i, \overline{X}_i$ are used to implement an exponential counter. The ABox $\mathcal{A}_n$ is given by

$$\mathcal{A}_n = \{G(a), \overline{X}_1(a), \dots, \overline{X}_n(a)\} \cup \{H(b), \overline{X}_1(b), \dots, \overline{X}_n(b)\}$$

For the construction of the TBox, we start with including the following CIs:

$$\begin{aligned}
\overline{X}_i &\sqsubseteq U, \quad \text{for all } 1 \leq i \leq n \\
\overline{X}_i &\sqsubseteq U', \quad \text{for all } 1 \leq i < n \\
G \sqcap U &\sqsubseteq \exists r.\, (\exists r.(G \sqcap A) \sqcap \exists r.(G \sqcap B)) \\
H \sqcap U' &\sqsubseteq \exists r.\, (\exists r.(G \sqcap A) \sqcap \exists r.(H \sqcap B)) \sqcap \\
&\qquad \exists r.\, (\exists r.(G \sqcap B) \sqcap \exists r.(H \sqcap A)) \\
H \sqcap U &\sqsubseteq \exists r.\exists r.A \sqcap \exists r.\exists r.B
\end{aligned}$$

Note that U and U' are enforced if the counter value is smaller than $2^n - 1$ and $2^n - 2$, respectively. It remains to implement the counter:

$$\begin{aligned}
X_1 &\sqsubseteq \forall r.\forall r.(\overline{X}_1 \sqcap C_1) \\
\overline{X}_1 &\sqsubseteq \forall r.\forall r.(X_1 \sqcap \overline{C}_1) \\
X_i &\sqsubseteq \forall r.\forall r.(\neg C_{i-1} \sqcup (\overline{X}_i \sqcap C_i)), \quad \text{for all } 2 \leq i \leq n \\
\overline{X}_i &\sqsubseteq \forall r.\forall r.(\neg C_{i-1} \sqcup (X_i \sqcap \overline{C}_i)), \quad \text{for all } 2 \leq i \leq n \\
X_i &\sqsubseteq \forall r.\forall r.(\neg \overline{C}_{i-1} \sqcup (X_i \sqcap \overline{C}_i)), \quad \text{for all } 2 \leq i \leq n \\
\overline{X}_i &\sqsubseteq \forall r.\forall r.(\neg \overline{C}_{i-1} \sqcup (\overline{X}_i \sqcap \overline{C}_i)), \quad \text{for all } 2 \leq i \leq n
\end{aligned}$$

Obviously, the size of both $\mathcal{A}_n$ and $\mathcal{T}_n$ is bounded by some polynomial in n. Let $\mathcal{I}$ denote the subtree of $\mathcal{U}_{\mathcal{T}_n,\mathcal{A}_n}$ starting at a (that is, with domain $\{\pi \in \Delta^{\mathcal{U}_{\mathcal{T}_n,\mathcal{A}_n}} \mid \pi \text{ starts with } a\}$) restricted to signature $\Sigma = \{r, A, B\}$. Then, define $q_n(x)$ as $(\mathcal{I}, a)$, viewed as CQ. By construction, q_n is a finite tree of exponential depth, thus of double exponential size. We verify the following claim:

Claim. $(\mathcal{T}_n, \mathcal{A}_n) \not\models q_n(b)$, but $(\mathcal{T}_n, \mathcal{A}_n) \models q'(b)$ for every subquery q' of q_n.

Proof of the Claim. For the first part of the claim it suffices to note that, by construction, $(\mathcal{U}_{\mathcal{T}_n,\mathcal{A}_n}, a) \not\to_\Sigma (\mathcal{U}_{\mathcal{T}_n,\mathcal{A}_n}, b)$. For the second part, note that $q_n(x)$ is minimal, that is, there is no subquery that is equivalent. Let $q'(x)$ be obtained from $q_n(x)$ by removing an atom. We distinguish cases.

- If some atom $A(z)$ or $B(z)$ is removed, then $q'(x)$ is not minimal anymore. In fact, one can remove the whole subtree below z from the query and obtain an equivalent query $\hat{q}(x)$. It is then easy to construct a match π from $\hat{q}(x)$ to $\mathcal{U}_{\mathcal{T}_n,\mathcal{A}_n}$ with $\pi(x) = b$, and thus, $\mathcal{T}_n, \mathcal{A}_n \models q'(b)$.
- If some atom $r(z, z')$ is removed, we can drop the disconnected part from $q'(x)$ and obtain an equivalent query $\hat{q}(x)$. Again, we can construct a match π from $\hat{q}(x)$ to $\mathcal{U}_{\mathcal{T}_n,\mathcal{A}_n}$ with $\pi(x) = b$.

The lemma then follows: set $S^+ = \{a\}$ and $S^- = \{b\}$. $\square$

Theorem 14. *If $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\mathcal{L}, CQ)$, there is a witness query of at most double (resp., four-fold) exponential size if $\mathcal{L}$ = Horn-$\mathcal{ALC}$ (resp., $\mathcal{L}$ = Horn-$\mathcal{ALCI}$).*

We prove Theorem 14 only for $\mathcal{L} = \text{CQ}$, it is completely analogous for UCQs.

Suppose that $(\mathcal{T}, \mathcal{A}, S^+, S^-, \Sigma) \in \mathsf{QBE}(\text{CQ})$. By Theorem 1, we know that both Condition 1 and Condition 2 are satisfied; thus, we also have $(\mathcal{P}, \mathbf{a}^*) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ for all $\mathbf{b} \in S^-$. To derive an upper bound on size of witness queries, we view the characterization in Lemma 5 as a *game*, where the second Player A tries to show the existence of the homomorphism and the first one, Player E, tries to refute that. The game proceeds as follows:

- Player A chooses a homomorphism from $(\mathcal{P}|_N, \mathbf{a}^*) \to (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$ and a labeling $T(\pi) \subseteq \Delta^k$ for every $\pi \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$.
- Player E wins immediately if there is no such homomorphism or T does not satisfy Point 2 in Lemma 5.
- After that initial phase, Players E and A take turns according to Point 3 in Lemma 5. Player E starts with choosing $\mathbf{a} \in N^k$ and sets $\pi := h(\mathbf{a})$ and $\mathbf{c} := \mathbf{a}$. Then:
  - Player E chooses $\mathbf{d} \in \Delta^k$ with $\mathbf{c} \hookrightarrow_r^{\mathcal{T},\mathcal{A}} \mathbf{d}$.
  - Player A responds with one of the following moves:
    * He chooses some $\pi' \in \mathsf{range}(h) \cup \mathsf{ind}(\mathcal{A})$ such that there is $(\pi, \pi') \in r^{\mathcal{U}_{\mathcal{T},\mathcal{A}}}$ and $\mathbf{d} \in T(\pi')$.
    * He chooses $T \subseteq T(\pi)$ and promises that $(r, \mathbf{d}, T, \mathsf{tp}_{\mathcal{U}_{\mathcal{T},\mathcal{A}}}(\pi)) \in \mathsf{PHom}$.

  In the first case, the game continues with $\mathbf{c} := \mathbf{d}$. In the second case, Player E can either challenge the promise or choose some new $\mathbf{c} \in T(\pi)$.
- Player E wins if, at some point, she can successfully challenge the promise or Player A cannot move anymore. Player A wins otherwise.

It is not hard to show that Player E has a winning strategy in this game iff $(\mathcal{P}, \mathbf{a}^*) \not\to_\Sigma (\mathcal{U}_{\mathcal{T},\mathcal{A}}, \mathbf{b})$. Since we know that the latter is the case, we know that Player E has a winning strategy. The witness that we are going to construct consists of $\mathcal{P}|_N$ together with initial parts of the subtrees rooted at each $\mathbf{a} \in N$ (again viewed as CQ with answer variables $\mathbf{a}^*$). The depth of these initial parts is related to the length of the plays. Thus, let us determine the length of the maximal play in the above game.

We can assume without loss of generality that Player E never visits a pair $(\pi, \mathbf{c})$ twice (otherwise she could play her reply of the second visit already in her first opportunity and win earlier). Thus, there are at most $|\Delta^k|^2$ many moves in the game.

In case of Horn-$\mathcal{ALC}$, we can always assume that the $T \subseteq T(\pi)$ chosen by Player A is $T = \emptyset$, due to the absence of inverse roles. Thus, he chooses the promise move at most once during the game, and if it is chosen once, the challenge of this move by Player E is successful. Now, note that a successful challenge of Player A's promise about PHom can be witnessed by an initial part of $\mathcal{P}_{\mathbf{d}}$ which is of depth $|\Delta^k \times \mathsf{TP}|$ (this can be extracted from the proof of Lemma 16 above). Thus, summarizing, the depth of the tree-shaped parts in the constructed CQ is at most $|\Delta^k|^2 + |\Delta^k| \times |\mathsf{TP}|$, which is exponential in the input size. Thus, the size of the overall query is doubly exponential in the input size.

For Horn-$\mathcal{ALCI}$, the argumentation is much more involved. We provide a characterization of PHom via reduction to pushdown reachability games and then apply results from this field.

A *pushdown reachability game* is a tuple $G = (Q, Q_I, Q_{II}, \Omega, \mathcal{R}, F)$ where Q is a finite set of control states partitioned into two sets Q_I, Q_{II}, Ω is a finite stack alphabet, and $\mathcal{R} \subseteq (Q \times \Sigma) \times (Q \times \Sigma^{\leq 2})$ is the set of transitions, and F is a set of target configurations. There a *configuration* is a pair (w, q) with $w \in \Omega^*$ is the stack content and $q \in Q$ is a control state. We write $(q, w) \to (q', w')$ if there is a one step transition from configuration (q, w) to configuration (q', w').

A *play* of a pushdown game is a sequence $(q_0, w_0), (q_1, w_1), \ldots,$ where (q_0, w_0) is some starting configuration and $(p_i, w_i) \to (p_{i+1}, w_{i+1})$ for all $i \geq 0$. Such a play is *winning for player I* if there is some i such that $(q_i, w_i) \in F$. A *strategy ξ for player I* is a function assigning to every finite sequence $(q_0, w_0), \ldots, (q_n, w_n)$ of configurations a configuration $\xi(\mathbf{v})$ such that $(q_n, w_n) \to \xi(\mathbf{v})$). A strategy is *winning* if it guarantees a win for player I whenever she follows the strategy.

For deciding $(\hat{r}, \hat{\mathbf{t}}, \hat{T}, \hat{t}) \in \mathsf{PHom}$, we view the homomorphism test as a game between Player I (trying to show that there is no homomorphism) and Player II (trying to show that there is a homomorphism). We model this game using pushdown systems, where intuitively, the state captures the current node in $\mathcal{P}$ and the stack contains the path from the root to the current node in $\mathcal{U}_{\mathcal{T},t}$. Thus, the pushdown game is based on the following states and stack alphabet:

$$\begin{aligned} Q_I &= \Delta^k \cup \{q_I, q_{II}\} \\ Q_{II} &= \{(\mathbf{t}, r, \mathbf{t}') \mid \mathbf{t} \hookrightarrow_r \mathbf{t}'\} \cup \{q_0, q_I', q_{II}'\} \\ \Omega &= \{(r, t) \mid r \text{ a } \Sigma\text{-role}, t \in \mathsf{TP}\} \cup \{(\bot, \hat{t})\} \end{aligned}$$

States q_I, q_I' and q_{II}, q_{II}' are entered when player I and player II, respectively, has won the homomorphism game. In $\mathcal{R}$, we include the following transitions:

1. $(q_0, (\bot, \hat{t})) \to (\hat{\mathbf{t}}, (\bot, \hat{t}) \cdot (\hat{r}, t))$ for all $t \in \mathsf{TP}$ with $\hat{t} \rightsquigarrow_r t$;
2. $(\mathbf{t}, a) \to ((\mathbf{t}, r, \mathbf{t}'), a)$ for all $(\mathbf{t}, r, \mathbf{t}') \in Q_{II}$ and $a \in \Omega$;
3. $((\mathbf{t}, r, \mathbf{t}'), (s, t)) \to (\mathbf{t}', (s, t) \cdot (r, t'))$ for all $t \rightsquigarrow_r t'$;
4. $((\mathbf{t}, r, \mathbf{t}'), (r^-, t)) \to (\mathbf{t}', \varepsilon)$;
5. $(\mathbf{t}, (r, t)) \to (q_I, \varepsilon)$ for all $\mathbf{t}, t$ with $\mathbf{t} \not\to_\Sigma t$;
6. $((\mathbf{t}, r, \mathbf{t}'), a) \to (q_I, \varepsilon)$ for all $a \in \Omega$;
7. $(q_I, a) \to (q_I', \varepsilon)$ and $(q_I', a) \to (q_I, a)$ for $a \in \Omega$;
8. $((\mathbf{t}, r, \mathbf{t}'), (\bot, \hat{t})) \to (q_{II}, (\bot, \hat{t}))$ for all $\mathbf{t} \in \hat{T}$;
9. $(q_{II}, a) \to (q_{II}', a)$ and $(q_{II}', a) \to (q_{II}, a)$ for $a \in \Omega$.

The intuition behind these rules is as follows. In the first rule, Player II chooses the successor of $\hat{t}$ from the second point of PHom. After that, players take turns. In rule 2, Player I chooses a possible successor of $\mathbf{t}$. Player II responds by either going to a successor of t (written on the stack), see rule 3, or going to the predecessor of the current node (by deleting the top-most symbol from the stack), see rule 4. Player I wins the game if at some point the game is at a state where the types are not compatible, see rule 5, or Player II is forced to enter q_I, see rule 6. In states q_I, q_I' players reduce the stack stepwise, until the target configuration $F = \{(q_I', \varepsilon)\}$ is reached. Player II wins if the game reaches the root and $\mathbf{t} \in \hat{T}$ (rule 8), because the game enters an infinite loop (rule 9). Based on these intuitions it is straightforward to show that:

Lemma 19. *Player I has a winning strategy from $(q_0, (\bot, \hat{t}))$ iff $(\hat{r}, \hat{\mathbf{t}}, \hat{T}, \hat{t}) \notin \mathsf{PHom}$.*

It is known that, if there is a winning strategy, then there is a *regular* one. To make this formal, define a *strategy automaton for G* to be a finite deterministic automaton $\mathfrak{D} = (S, \Omega, s_0, \delta, (\mathsf{Out}_s)_{s \in S})$ where S is a finite set of states, Ω is the underlying alphabet (coinciding with the one of the game), $\delta : S \times \Omega \to S$ is the transition function, and for every $s \in S$, Out_s is a function that takes as input a state $q \in Q_I$ and returns a move for player I. Intuitively, $\mathfrak{D}$ processes a configuration (q, w) by reading first the stack content from bottom to top and then returning a move for q. In order to be a strategy automaton, $\mathfrak{D}$ has to satisfy the property that, when the game is in a configuration (q, w), $q \in Q_I$ that is winning for Player I, then $\mathfrak{D}$ outputs a move that is winning for Player I. In other words, $\mathfrak{D}$ specifies a winning strategy for Player I.

It has been shown in [Kupferman *et al.*, 2010] and more explicitely in [Carayol and Hague, 2014] that, if there is a winning strategy for Player I in G, then there is a strategy automaton $\mathfrak{D}$ for G of size exponential in Q. Let us assume that Player I has a winning strategy for the homomorphism game, that is, no matter the moves of player II, after finitely many steps the game reaches a situation where II can only go to state q_{err}. Let $\mathfrak{D}$ be the strategy automaton which exists because of the winning strategy, and let $N = |S|$ be the number of its states. Moreover, let $M = |\Omega|$ be the size of the underlying alphabet. We can then show:

Lemma 20. *The game played according to the strategy automaton does not reach a configuration (q, w) with $w \geq MN + 1$.*

Proof. Assume to the contrary that such a configuration is reached via play $(q_0, w_0), \ldots, (q_n, w_n) = (q, w)$. For each $i \in \{1, \ldots, MN+1\}$, define n_i to be the minimal j such that $q_j \in Q_I$ and $|w_k| > i$ for all $k > j$. Consider the sequence $(p_1, s_1), \ldots, (p_{MN+1}, s_{MN+1})$ where, for all i, $p_i = q_{n_i}$ and s_{n_i} is the state of $\mathfrak{D}$ after reading w_{n_i}. By choice of M, N, there are $i < j$ such that $(p_i, s_i) = (p_j, s_j)$. But this means that Player II has found a way to force Player I to visit the same pair (p, s) twice *without decreasing the stack*. Hence, using the same strategy, Player II can extend the game to infinity, contradicting the assumption that $\mathfrak{D}$ is a strategy automaton. $\square$

Note further that a winning game for Player I never reaches the same configuration (w, q) twice. Since there are only $|\Omega|^{MN} \times |Q|$ configurations with stack height at most MN, any winning game has length at most $|M|^{MN} \times |Q|$. It remains to note that $M = O(2^{p(|\mathcal{T}|)})$, $|Q| = O((2^{p(|\mathcal{T}|)} + |\mathcal{A}|)^{p(k)})$, and $N = O(2^{p(|Q|)})$ to derive that a winning game has at most

$$n_0 = O(2^{2^{2^{p(|\mathcal{T}|k)} + |\mathcal{A}|^{p(k)}}})$$

many rounds, for some polynomial p.

Summing up. We have shown above that, if Player I successfully challenges the promise about PHom at some point, this can be witnessed by a CQ of depth n_0, that is, triple exponential depth. Now, if during the (outside) game Player A always chooses the step corresponding to 3(a) of Lemma 5 and only finally the promise step corresponding to 3(b), we are done: the depth of the tree-shaped parts is bounded triple exponentially in the input size. In case there are several (unchallenged) moves corresponding to 3(b), one can show with a similarly defined reachability game that the types in T can be 'enforced' in a game of at most n_0 rounds. The overall size of the witness CQ that we obtain is then four-fold exponential.