Ontology-Based Data Access with Closed Predicates is Inherently Intractable (Sometimes)

Carsten Lutz and **İnanç Seylan**
Fachbereich Informatik
Universität Bremen, Germany

Frank Wolter
Department of Computer Science
University of Liverpool, UK

Abstract

When answering queries in the presence of ontologies, adopting the closed world assumption for some predicates easily results in intractability. We analyze this situation on the level of individual ontologies formulated in the description logics DL-Lite and $\mathcal{EL}$ and show that in all cases where answering conjunctive queries (CQs) with (open and) closed predicates is tractable, it coincides with answering CQs with all predicates assumed open. In this sense, CQ answering with closed predicates is inherently intractable. Our analysis also yields a dichotomy between AC^0 and coNP for CQ answering w.r.t. ontologies formulated in DL-Lite and a dichotomy between PTime and coNP for $\mathcal{EL}$. Interestingly, the situation is less dramatic in the more expressive description logic $\mathcal{ELI}$, where we find ontologies for which CQ answering is in PTime, but does not coincide with CQ answering where all predicates are open.

1 Introduction

Description logics (DLs) increasingly find application in ontology-based data access (OBDA), where an ontology is used to enrich instance data and a chief problem is to provide efficient query answering services. In this context, it is common to make the open world assumption (OWA). Indeed, there are applications where the data is inherently incomplete and the OWA is semantically adequate, such as when the data is extracted from the web. In other applications, however, it is more reasonable to make a closed world assumption (CWA) for some predicates in the data. In particular, when the data is taken from a relational database, then the CWA can be appropriate for some of the data predicates. As a concrete example, consider a geographical database such as OpenStreetMap which contains pure geographical data as well as rich annotations, stating for example that a certain spatial area is the location of a 'popular Thai restaurant'. As argued in [Hübner *et al.*, 2004; Codescu *et al.*, 2011], it is useful to pursue an OBDA approach to take full advantage of the annotations, where one would naturally interpret the geographical data as closed and the annotations as open.

In the DL literature, there are a variety of approaches to imposing a partial CWA, often based on epistemic operators or rules [Calvanese *et al.*, 2007b; Donini *et al.*, 2002; Grimm and Motik, 2005; Motik and Rosati, 2010; Sengupta *et al.*, 2011]. In this paper, we adopt the standard semantics from relational databases, which is both natural and straightforward: CWA predicates have to be interpreted exactly as described in the data, making the standard names assumption for data constants; for example, when A is a closed concept name and $\mathcal{A}$ an ABox, then in any model $\mathcal{I}$ of $\mathcal{A}$ we must have $A^{\mathcal{I}} = \{a \mid A(a) \in \mathcal{A}\}$. Note that this semantics is also used in the recently proposed DBoxes [Seylan *et al.*, 2009]. In fact, the setup considered in this paper generalizes both standard OBDA (only open predicates permitted) and DBoxes (only closed predicates permitted in data) by allowing to freely mix open and closed predicates both in the ontology and in the data.

A major problem in admitting closed predicates in OBDA is that query answering easily becomes intractable regarding data complexity, where the ontology and query are assumed to be fixed and thus of constant size. In fact, answering conjunctive queries (CQs) is coNP-hard already when ontologies are formulated in inexpressive DLs such as DL-Lite and $\mathcal{EL}$ [Franconi *et al.*, 2011]. While this is an interesting first step, it was recently demonstrated in [Lutz and Wolter, 2012] in the context of OBDA with more expressive DLs that a 'non-uniform' analysis of data complexity, which considers individual ontologies instead of entire logics, can reveal a much more detailed and subtle picture. In our context, we work with ontologies of the form $(\mathcal{T}, \Sigma)$, where $\mathcal{T}$ is a DL TBox and Σ a set of predicates (concept and role names) declared to be closed. We say that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ *is in* PTime if for every CQ $q(\vec{x})$, there exists a polytime algorithm that computes for a given ABox $\mathcal{A}$ the certain answers to q in $\mathcal{A}$ given $(\mathcal{T}, \Sigma)$; *CQ answering w.r.t.* $(\mathcal{T}, \Sigma)$ *is* coNP-*hard* if there is a Boolean CQ q such that, given an ABox $\mathcal{A}$, it is coNP-hard to decide whether q is entailed by $\mathcal{A}$ given $\mathcal{T}$. Other complexity classes are defined analogously.

The aim of this paper is to carry out a non-uniform analysis of the data complexity of query answering with closed predicates, when TBoxes are formulated in the DLs DL-Lite$_{\mathcal{R}}$ and $\mathcal{EL}$, underpinning the OWL 2 profiles OWL 2 QL and OWL 2 EL, respectively [Calvanese *et al.*, 2007a; Artale *et al.*, 2009; Baader *et al.*, 2005]. Our main results are (i) characterizations that separate the tractable cases from the intractable ones and map out the frontier of tractability in a transparent way; (ii) a proof that, for every tractable case $(\mathcal{T}, \Sigma)$, CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ coincides with CQ answering w.r.t. the ontology $(\mathcal{T}, \emptyset)$ that treats all predicates as open, for ABoxes that are satisfiable w.r.t. $(\mathcal{T}, \Sigma)$; (iii) a dichotomy for the data complexity of CQ answering between AC^0 and CONP for TBoxes formulated in DL-Lite$_\mathcal{R}$, and between PTIME and CONP for $\mathcal{EL}$; and (iv) algorithms for deciding in PTIME whether a given $(\mathcal{T}, \Sigma)$ admits tractable CQ-answering or not.

Point (ii) can be interpreted as showing that OBDA with closed predicates is inherently intractable since, in all tractable cases, the declaration of closed predicates does not have an impact on query answers (it only results in imposing integrity constraints on the ABox). This rather negative result is relativized by the observation made at the end of the paper that inherent intractability does not transfer to more expressive description logics such as $\mathcal{ELI}$, which is essentially the union of DL-Lite and $\mathcal{EL}$: there are TBoxes with closed predicates $(\mathcal{T}, \Sigma)$ with $\mathcal{T}$ formulated in $\mathcal{ELI}$ such that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is tractable, but does *not* coincide with CQ answering w.r.t. $(\mathcal{T}, \emptyset)$. Further exploring this encouraging observation is left as future work. We also note that, even in cases where no additional answers to CQs are obtained, there is a certain benefit of closed predicates: we can go beyond CQs and admit full first-order expressive power for the 'closed part' of queries without increasing the data complexity. We propose a concrete query language that implements this idea and show that AC^0 data complexity is preserved for DL-Lite$_\mathcal{R}$ TBoxes and PTIME data complexity is preserved for $\mathcal{EL}$-TBoxes.

Point (iii) is interesting when contrasted with CQ answering w.r.t. TBoxes that are formulated in the expressive DLs $\mathcal{ALC}$ and $\mathcal{ALCI}$, without closed predicates. There, the data complexity is also between AC^0 and CONP, but the existence of a dichotomy between PTIME and CONP is a deep open question that is equivalent to the Feder-Vardi conjecture for the existence of a dichotomy between PTIME and NP in constraint satisfaction problems [Lutz and Wolter, 2012]. In this sense, the space of ontologies $(\mathcal{T}, \Sigma)$ studied in this paper is more well-behaved than the space of all $\mathcal{ALC}$-ontologies.

Some proof details are deferred to the appendix of the long version, http://cgi.csc.liv.ac.uk/~frank/publ/publ.html.

2 Preliminaries

We use standard notation from description logic [Baader *et al.*, 2003]. Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be countably infinite sets of *concept* and *role names*. A *DL-Lite concept* is either a concept name from $\mathsf{N_C}$ or a concept of the form $\exists r.\top$ or $\exists r^-.\top$, where $r \in \mathsf{N_R}$. We call r^- an *inverse role* and set $s^- = r$ if $s = r^-$ and $r \in \mathsf{N_R}$. A *role* is of the form r or r^-, with $r \in \mathsf{N_R}$. A *DL-Lite concept inclusion* is an expression of the form $B_1 \sqsubseteq B_2$ or $B_1 \sqsubseteq \neg B_2$, where B_1, B_2 are DL-Lite concepts. A *role inclusion* is an expression of the form $r \sqsubseteq s$, where r, s are roles. A *DL-Lite$_{core}$ TBox* is a finite set of DL-Lite concept inclusions and a DL-Lite$_\mathcal{R}$ TBox is a finite set of DL-Lite concept inclusions and role inclusions.

$\mathcal{EL}$ concepts are constructed according to the rule $C, D := \top \mid A \mid C \sqcap D \mid \exists r.C$, where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. An *$\mathcal{EL}$ concept inclusion* is an expression of the form $C \sqsubseteq D$, where C, D are $\mathcal{EL}$ concepts. An *$\mathcal{EL}$ TBox* is a finite set of $\mathcal{EL}$ concept inclusions. $\mathcal{ELI}$ is the extension of $\mathcal{EL}$ with existential restrictions $\exists r^-.C$, where r^- is an inverse role.

An *ABox* is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a, b)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and a, b individual names from a countably infinite set $\mathsf{N_I}$. We use $\mathsf{Ind}(\mathcal{A})$ to denote the set of individual names used in the ABox $\mathcal{A}$ and take the freedom to write $r^-(a, b) \in \mathcal{A}$ instead of $r(b, a) \in \mathcal{A}$.

An interpretation $\mathcal{I}$ is a pair $(\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where $\Delta^\mathcal{I}$ is a non-empty set called the *domain* of $\mathcal{I}$ and $\cdot^\mathcal{I}$ maps each concept name A to a subset $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ and each role name r to a binary relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The function $\cdot^\mathcal{I}$ is extended to compound concepts in the usual way. An interpretation $\mathcal{I}$ *satisfies* a concept inclusion $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, a role inclusion $r \sqsubseteq s$ if $r^\mathcal{I} \subseteq s^\mathcal{I}$, a concept assertion $A(a)$ if $a \in A^\mathcal{I}$ and a role assertion $r(a, b)$ if $(a, b) \in r^\mathcal{I}$. Note that this interpretation of ABox assertions corresponds to making the standard names assumption (SNA), which stipulates that every ABox individual is interpreted as itself; the SNA implies the unique name assumption (UNA). An interpretation is a *model* of a TBox $\mathcal{T}$ if it satisfies all inclusions in $\mathcal{T}$ and a *model* of an ABox $\mathcal{A}$ if it satisfies all assertions in $\mathcal{A}$. A concept C (ABox $\mathcal{A}$) is *satisfiable w.r.t. a TBox $\mathcal{T}$* if there exists a model $\mathcal{I}$ of $\mathcal{T}$ with $C^\mathcal{I} \neq \emptyset$ (and of $\mathcal{A}$).

A *predicate* is a concept or role name. A *signature* Σ is a finite set of predicates. The signature $\mathsf{sig}(C)$ of a concept C, $\mathsf{sig}(r)$ of a role r, and $\mathsf{sig}(\mathcal{T})$ of a TBox $\mathcal{T}$, is the set of predicates that occur in C, r, and $\mathcal{T}$, respectively.

For being able to declare predicates as closed, we add an additional component to TBoxes. A pair $(\mathcal{T}, \Sigma)$ with $\mathcal{T}$ a TBox and Σ a signature is *a TBox with closed predicates*. For any ABox $\mathcal{A}$, a *model $\mathcal{I}$ of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$* is an interpretation $\mathcal{I}$ with $\mathsf{Ind}(\mathcal{A}) \subseteq \Delta^\mathcal{I}$ that satisfies $\mathcal{T}$ and $\mathcal{A}$ and such that the extension of all closed predicates agrees with what is explicitly stated in the ABox, that is,

$$\begin{array}{lcll} A^\mathcal{I} & = & \{a \mid A(a) \in \mathcal{A}\} & \text{for all } A \in \Sigma \cap \mathsf{N_C} \\ r^\mathcal{I} & = & \{(a, b) \mid r(a, b) \in \mathcal{A}\} & \text{for all } r \in \Sigma \cap \mathsf{N_R}. \end{array}$$

An ABox $\mathcal{A}$ is *satisfiable w.r.t.* $(\mathcal{T}, \Sigma)$ if there is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$.

Example 2.1. *In a geographical database, complete information is typically available for predicates that are tied closely to geographical location and do not change frequently, such as the concept name* $\mathsf{ScandinavianCountry}$ *used to identify regions that describe the spatial extension of a scandinavian country and the role name* $\mathsf{neighbor}$ *used to relate regions that describe neighboring countries. These predicates should therefore be treated as closed. For other predicates, especially those that are less intimately linked to geographical location, complete information is often not available. Examples include concept names such as* $\mathsf{OilExportingCountry}$ *or roles such as* $\mathsf{tradingPartner}$.

Fix a countably infinite set of variables V. A *first-order query (FOQ)* $q(\vec{x})$ is a first-order formula constructed from atoms $A(x)$, $r(x, y)$, and $x = y$, where x, y range over V and $\vec{x} = x_1, \ldots, x_k$ contains all free variables of q. We call $\vec{x}$

the *answer variables of* $q(\vec{x})$ and say that $q(\vec{x})$ is *Boolean* if it has no answer variables. A *conjunctive query (CQ)* $q(\vec{x})$ is a FOQ using conjunction and existential quantification, only. A tuple $\vec{a} = a_1, \ldots, a_k \subseteq \mathsf{Ind}(\mathcal{A})$ *is a certain answer to* $q(\vec{x})$ *in* $\mathcal{A}$ *given* $(\mathcal{T}, \Sigma)$, in symbols $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\vec{a})$, if $\mathcal{I} \models q[a_1, \ldots, a_k]$ for all models $\mathcal{I}$ of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$. If $\Sigma = \emptyset$, then we simply omit Σ, speak of *certain answers to* $q(\vec{x})$ *in* $\mathcal{A}$ *given* $\mathcal{T}$, and write $\mathcal{T}, \mathcal{A} \models q[\vec{a}]$. A CQ $q(x)$ with one answer variable x is a *directed tree CQ* if it is tree-shaped with root x when viewed as a directed graph and a *tree CQ* if the same is true when q is viewed as an undirected graph.

The following example shows that closing predicates can result in more complete query answers.

Example 2.2. *The TBox* $\mathcal{T}$ *consists of the inclusion*

$$\mathsf{ScandComp} \sqsubseteq \exists \mathsf{based\_in}.\mathsf{ScandCountry}$$

where ScandComp *and* ScandCountry *are short for* ScandinavianCompany *and* ScandinavianCountry. *The ABox* $\mathcal{A}$ *consists of the assertions*

$$\begin{aligned}&\mathsf{ScandComp}(cp), \mathsf{ScandCountry}(denmark),\\&\mathsf{ScandCountry}(norway), \mathsf{ScandCountry}(sweden),\\&\mathsf{TimberExporter}(denmark), \mathsf{TimberExporter}(norway)\\&\mathsf{TimberExporter}(sweden).\end{aligned}$$

Note that there is no information in $\mathcal{A}$ *about the concrete scandinavian country in which the company cp is based. For*

$$q = \exists y\, \mathsf{based\_in}(x, y) \;\wedge \mathsf{TimberExporter}(y),$$

cp is not a certain answer to $q(x)$ *in* $\mathcal{A}$ *given* $\mathcal{T}$. *In contrast, when closing* ScandCountry *by setting* $\Sigma = \{\mathsf{ScandCountry}\}$, *we have* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q[cp]$.

As illustrated by Example 2.2, we are interested in reasoning with a mix of closed predicates and open predicates. Note that TBox statements which *only* involve closed predicates act as integrity constraints in the standard database sense [Abiteboul *et al.*, 1995]. As an example, consider $\mathcal{T} = \{A \sqsubseteq B\}$ and $\Sigma = \{A, B\}$. Then $(\mathcal{T}, \Sigma)$ imposes the integrity constraint that if $A(a)$ is contained in an ABox, then so must be $B(a)$. In particular, an ABox $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$ iff $\mathcal{A}$ satisfies this integrity constraint. For ABoxes $\mathcal{A}$ that are satisfiable w.r.t. $\mathcal{T}$, $(\mathcal{T}, \Sigma)$ has no further effect on query answers. In a DL context, integrity constraints are discussed in [Calvanese *et al.*, 2007b; Donini *et al.*, 2002; Mehdi *et al.*, 2011; Motik *et al.*, 2009; Motik and Rosati, 2010]. We now fix the relevant notions of complexity, inspired by [Lutz and Wolter, 2012]. When speaking of complexity, we *always* mean data complexity.

Definition 2.3. *For* $(\mathcal{T}, \Sigma)$ *a TBox with closed predicates,*

- CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in PTIME *if for every CQ* $q(\vec{x})$ *there is a polytime algorithm that computes, for a given ABox* $\mathcal{A}$*, all* $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$ *with* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\vec{a})$*;*
- CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is CONP-hard *if there is a Boolean CQ* q *such that it is* CONP*-hard to decide, given an ABox* $\mathcal{A}$*, whether* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q$.

For other classes of queries such as tree-shaped CQs, analogous notions can be defined. It is shown in [Franconi *et al.*, 2011] that there are DL-Lite$_{\text{core}}$ TBoxes with closed predicates $(\mathcal{T}, \Sigma)$ such that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is CONP-hard. The proof is easily strengthened to directed tree CQs and adapted to $\mathcal{EL}$. CQ answering w.r.t. both DL-Lite$_\mathcal{R}$ and $\mathcal{EL}$-TBoxes is known to be in CONP. Without closed predicates (that is, when $\Sigma = \emptyset$), CQ answering is in PTIME for $\mathcal{EL}$ TBoxes [Calvanese *et al.*, 2007a; Lutz *et al.*, 2009] and in AC^0 for DL-Lite$_\mathcal{R}$ TBoxes [Calvanese *et al.*, 2007a; Artale *et al.*, 2009].

The following property plays a central role in our complexity analysis as it turns out to identify the borderline between tractability and CONP-hardness of CQ answering. It is also studied intensively in [Lutz and Wolter, 2012], where convexity is called the ABox disjunction property.

Definition 2.4. *A TBox with closed predicates* $(\mathcal{T}, \Sigma)$ *is* convex *if for all ABoxes* $\mathcal{A}$ *and tree CQs* $q_1(x)$, $q_2(x)$, $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q_1 \vee q_2[a]$ *implies* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q_i[a]$ *for some* $i \in \{1, 2\}$.

It is well-known that, without closed predicates, every TBox formulated in DL-Lite$_\mathcal{R}$ or $\mathcal{EL}$ is convex [Lutz and Wolter, 2012]. On the other hand, it can be shown that the TBox $(\mathcal{T}, \Sigma)$ from Example 2.2 is not convex.

Example 2.5. *Let* $\mathcal{A}'$ *be the extension of the ABox* $\mathcal{A}$ *from Example 2.2 with the assertions* $\mathsf{MilkExporter}(sweden)$*,* $\mathsf{MilkExporter}(denmark)$*, and* $\mathsf{OilExporter}(norway)$. *Set* $\Sigma = \{\mathsf{ScandCountry}\}$ *and take the tree CQs*

$$\begin{aligned}q_1(x) &= \exists y\, \mathsf{based\_in}(x, y) \wedge \mathsf{MilkExporter}(y)\\q_2(x) &= \exists y\, \mathsf{based\_in}(x, y) \wedge \mathsf{OilExporter}(y).\end{aligned}$$

Then $\mathcal{T}, \mathcal{A}' \models_{c(\Sigma)} q_1 \vee q_2[cp]$*, but* $\mathcal{T}, \mathcal{A}' \not\models_{c(\Sigma)} q_i[cp]$ *for any* $i \in \{1, 2\}$. *The former is a consequence of the fact that, in any model* $\mathcal{I}$ *of* $(\mathcal{T}, \Sigma)$ *and* $\mathcal{A}'$*, at least one of* $(cp, denmark)$*,* $(cp, sweden)$*,* $(cp, norway)$ *must be in* $\mathsf{based\_in}^\mathcal{I}$. *To see that* $\mathcal{T}, \mathcal{A}' \not\models_{c(\Sigma)} q_1[cp]$*, note that it is possible to obtain a model of* $(\mathcal{T}, \Sigma)$ *and* $\mathcal{A}'$ *by viewing* $\mathcal{A}'$ *as an interpretation and adding* $(cp, norway)$ *to the extension of* based_in. *For* $\mathcal{T}, \mathcal{A}' \not\models_{c(\Sigma)} q_2[cp]$*, add* $(cp, denmark)$.

We use *tree* CQs in Definition 2.4 as this allows us to derive stronger lower bounds, which refer to this more restricted class of queries. Note that tree CQs are also known as $\mathcal{ELI}$ *instance queries* and directed tree CQs as $\mathcal{EL}$ *instance queries*, both common in OBDA. All our results remain true when tree CQs are replaced with CQs in Definition 2.4.

3 Results for DL-Lite

We start with an example of a DL-Lite$_{\text{core}}$ TBox that is not convex, essentially by recasting Example 2.5, which is based on an $\mathcal{EL}$ TBox, in this language.

Example 3.1. *Let*

$$\begin{aligned}\mathcal{T} &= \{A \sqsubseteq \exists r.\top, \exists r^-.\top \sqsubseteq B\} \text{ and } \Sigma = \{B\}\\\mathcal{A} &= \{A(a), B(b_1), A_1(b_1), B(b_2), A_2(b_2)\}\\q_i &= \exists y\, r(x, y) \wedge A_i(y) \text{ for } i \in \{1, 2\}.\end{aligned}$$

Then $(\mathcal{T}, \Sigma)$ *is not convex because* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q_1 \vee q_2[a]$*, whereas* $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} q_i[a]$ *for any* $i \in \{1, 2\}$.

The failure of convexity for the TBox $(\mathcal{T}, \Sigma)$ in Example 3.1 results in a choice which can be used to prove that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is coNP-hard. Specifically, the proof is by reduction of 2+2-SAT, a variant of propositional satisfiability where each clause contains precisely two positive literals and two negative literals [Schaerf, 1993]. The queries q_1 and q_2 from the example are used as subqueries of the query constructed in the reduction, where they serve the purpose of distinguishing truth values of propositional variables. The CQ used in the reduction is actually a tree CQ.

It turns out that this proof of coNP-hardness can be adapted to *any* non-convex DL-Lite$_\mathcal{R}$ TBox. Conversely, we will show that convex DL-Lite$_\mathcal{R}$ TBoxes admit CQ answering in AC^0, thus identifying convexity as the borderline between tractability and intractability of CQ answering, and establishing a dichotomy between AC^0 and coNP for CQ answering w.r.t. DL-Lite$_\mathcal{R}$ TBoxes with closed predicates.

Since analyzing DL-Lite$_\mathcal{R}$ TBoxes turns out to be somewhat more technical than analyzing DL-Lite$_{\text{core}}$ TBoxes, we start with the latter as a warmup. The following definition introduces a property of DL-Lite$_{\text{core}}$ TBoxes with closed predicates that we will prove to coincide with convexity, but which is much more concrete.

Definition 3.2 (Safe DL-Lite$_{\text{core}}$ TBox)**.** *A DL-Lite$_{core}$ TBox with closed predicates $(\mathcal{T}, \Sigma)$ is* safe *if there are no DL-Lite concepts B_1, B_2 and role r such that the following conditions are satisfied:*

1. *B_1 is satisfiable w.r.t. $\mathcal{T}$;*
2. *$\mathcal{T} \models B_1 \sqsubseteq \exists r.\top$ and $\mathcal{T} \models \exists r^-.\top \sqsubseteq B_2$;*
3. *$B_1 \neq \exists r.\top$;*
4. *$\mathsf{sig}(B_2) \subseteq \Sigma$ and $\mathsf{sig}(r) \cap \Sigma = \emptyset$.*

Note that Definition 3.2 is essentially a slight generalization of Example 3.1. In particular, the pattern in Point 2 of Definition 3.2 can be found in Example 3.1 (where it is crucial that $r \notin \Sigma$ and $B \in \Sigma$). The following theorem summarizes our results for DL-Lite$_{\text{core}}$.

Theorem 3.3 (Results for DL-Lite$_{\text{core}}$)**.** *Let $(\mathcal{T}, \Sigma)$ be a DL-Lite$_{core}$ TBox with closed predicates. Then*

1. *If $(\mathcal{T}, \Sigma)$ is not safe, then $(\mathcal{T}, \Sigma)$ is not convex and answering tree CQs w.r.t. $(\mathcal{T}, \Sigma)$ is* coNP*-hard.*
2. *If $(\mathcal{T}, \Sigma)$ is safe, then*
   - *(a) CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ coincides with CQ answering w.r.t. $(\mathcal{T}, \emptyset)$ for all ABoxes that are satisfiable w.r.t. $(\mathcal{T}, \Sigma)$, and $(\mathcal{T}, \Sigma)$ is convex;*
   - *(b) CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in* AC^0.

In a sense, Theorem 3.3 shows that CQ answering in DL-Lite$_{\text{core}}$ with closed predicates is inherently intractable: in all cases where closing predicates results in additional answers to queries (on satisfiable ABoxes), CQ answering is coNP-hard. In all tractable cases, the only effect that closing predicates can thus have is to act as integrity constraints on the ABox (but see Section 5 for another virtue of closing predicates). Note that all TBoxes that refer *only* to closed predicates (thus express only integrity constaints) are safe.

It is also interesting to note that Theorem 3.3 establishes a dichotomy between AC^0 and coNP for CQ answering w.r.t. DL-Lite$_{\text{core}}$ TBoxes with closed predicates, that is, there is no such TBox whose complexity is truely between AC^0 and coNP. As noted in the introduction, this is in stark contrast to results recently established in [Lutz and Wolter, 2012] in the context of more expressive DLs without closed predicates.

To prove Point 1 of Theorem 3.3, one shows that non-safeness implies non-convexity by constructing an appropriate ABox. coNP-hardness can then be proved by reduction from 2+2-SAT, generalizing the coNP-hardness proof for Example 3.1. The proof of Point 2(a) relies on canonical models for DL-Lite$_{\text{core}}$ TBoxes $\mathcal{T}$ without closed predicates. Specifically, for every ABox $\mathcal{A}$ that is satsfiable w.r.t. $\mathcal{T}$, there is a model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ such that for all CQs q and potential answers $\vec{a}$, we have $\mathcal{T}, \mathcal{A} \models q[\vec{a}]$ iff $\mathcal{I} \models q[\vec{a}]$. To establish Point 2(a), it suffices to show that, when $(\mathcal{T}, \Sigma)$ is safe, then $\mathcal{I}$ is also a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$. Consequently and since closing predicates can only result in additional answers, but not in invalidating answers, CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ coincides with CQ answering w.r.t. $(\mathcal{T}, \emptyset)$ and it remains to recall that DL-Lite$_{\text{core}}$ TBoxes without closed predicates are convex. For Point 2(b), it suffices to show that satisfiability of ABoxes w.r.t. $(\mathcal{T}, \Sigma)$ is in AC^0 when $(\mathcal{T}, \Sigma)$ is safe, which is a consequence of the fact that ABox satisfiability and CQ answering in DL-Lite$_{\text{core}}$ without closed predicates are in AC^0. Specifically, we observe that whenever an ABox $\mathcal{A}$ is satisfiable w.r.t. $\mathcal{T}$, then $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$ iff $\mathcal{T}, \mathcal{A} \models B(a)$ implies $B(a) \in \mathcal{A}$ for all DL-Lite concepts B with $\mathsf{sig}(B) \subseteq \Sigma$ and $\mathcal{T}, \mathcal{A} \models r(a, b)$ implies $r(a, b) \in \mathcal{A}$ for all role names r from Σ. Proof details for Theorem 3.3 are skipped as we provide them for the strictly stronger DL-Lite$_\mathcal{R}$ version of this theorem, which is given below.

We now extend Definition 3.2 to DL-Lite$_\mathcal{R}$.

Definition 3.4 (Safe DL-Lite$_\mathcal{R}$ TBox)**.** *A DL-Lite$_\mathcal{R}$ TBox with closed predicates $(\mathcal{T}, \Sigma)$ is* safe *if there are no DL-Lite concepts B_1, B_2 and role r such that the following conditions are satisfied:*

1. *B_1 is satisfiable w.r.t. $\mathcal{T}$;*
2. *$\mathcal{T} \models B_1 \sqsubseteq \exists r.\top$ and $\mathcal{T} \models \exists r^-.\top \sqsubseteq B_2$;*
3. *$B_1 \neq \exists r'.\top$ for any role r' such that $\mathcal{T} \models r' \sqsubseteq r$;*
4. *$\mathsf{sig}(B_2) \subseteq \Sigma$ and $\mathsf{sig}(r') \cap \Sigma = \emptyset$ for any role r' such that $\mathcal{T} \models B_1 \sqsubseteq \exists r'.\top$ and $\mathcal{T} \models r' \sqsubseteq r$.*

Note that the conditions in Definition 3.4 generalize the corresponding ones in Definition 3.2 and in this sense, the addition of role hierarchies does not introduce unexpected ways to cause non-convexity and coNP-hardness.

Example 3.5. *Let $\mathcal{T} = \{A \sqsubseteq \exists r.\top, r \sqsubseteq s\}$ and $\Sigma = \{s\}$. Then $(\mathcal{T}, \Sigma)$ is not safe, which is witnessed by the concepts $B_1 = A$, $B_2 = \exists s^-.\top$, and the role r. Indeed, $(\mathcal{T}, \Sigma)$ is not convex, witnessed for example by the ABox $\{A(a), s(a, b_1), A_1(b_1), s(a, b_2), A_2(b_2)\}$ and the queries $q_i = \exists y.r(x, y) \wedge A_i(y)$ for $i \in \{1, 2\}$.*

Now, Theorem 3.3 generalizes to DL-Lite$_\mathcal{R}$.

Theorem 3.6 (Results for DL-Lite$_\mathcal{R}$)**.** *All statements in Theorem 3.3 are still true if DL-Lite$_{core}$ is replaced with DL-Lite$_\mathcal{R}$.*

The proof strategy for Theorem 3.3 is exactly the one described above for DL-Lite$_{\text{core}}$.

Note that it is easy to check in PTIME whether a given DL-Lite$_{\mathcal{R}}$ TBox with closed predicates $(\mathcal{T}, \Sigma)$ is safe (consequently: whether CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in AC^0) since it suffices to consider DL-Lite concepts B_1, B_2 and roles r from the signature of $\mathcal{T}$ (of which there are only polynomially many) and subsumption in DL-Lite can be decided in AC^0 [Calvanese *et al.*, 2007a].

4 Results for $\mathcal{EL}$

As illustrated by Example 2.5, the effect that causes non-convexity and thus coNP-hardness of DL-Lite TBoxes with closed predicates can also be observed in $\mathcal{EL}$. In the simplest form, this is shown by the TBox with closed predicates $(\mathcal{T}, \Sigma)$ with $\mathcal{T} = \{A \sqsubseteq \exists r.B\}$ and $\Sigma = \{B\}$, which is not convex. However, in $\mathcal{EL}$ there is an additional (and more subtle) cause for non-tractability. The simplest illustrating example uses exactly the same TBox $\mathcal{T}$, but swaps the Σ-memberships of r and B.

Example 4.1. *Let*

$$\begin{aligned}
\mathcal{T} &= \{A \sqsubseteq \exists r.B\} \text{ and } \Sigma = \{r\} \\
\mathcal{A} &= \{A(a), r(a, b_1), A_1(b_1), r(a, b_2), A_2(b_2)\} \\
q_i &= \exists y\, r(x, y) \wedge A_i(y)
\end{aligned}$$

Then $(\mathcal{T}, \Sigma)$ is not convex because $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q_1 \vee q_2[a]$, whereas $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} q_i[a]$ for any $i \in \{1, 2\}$.

We now give a definition of safeness of $\mathcal{EL}$-TBoxes with closed predicates that captures both causes of non-convexity and, as in the DL-Lite case, coincides both with convexity and with tractability of CQ-answering. We call a concept E a *top-level conjunct (tlc) of* an $\mathcal{EL}$ concept C if C is of the form $D_1 \sqcap \cdots \sqcap D_n$ and $E = D_i$ for some i.

Definition 4.2 (Safe $\mathcal{EL}$ TBox)**.** *An $\mathcal{EL}$ TBox with closed predicates $(\mathcal{T}, \Sigma)$ is* safe *if there exists no $\mathcal{EL}$ inclusion $C \sqsubseteq \exists r.D$ such that*

1. *$\mathcal{T} \models C \sqsubseteq \exists r.D$;*
2. *there does not exist a tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$;*
3. *one of the following is true:*
   - *(s1) $r \notin \Sigma$ and $\mathsf{sig}(D) \cap \Sigma \neq \emptyset$;*
   - *(s2) $r \in \Sigma$, $\mathsf{sig}(D) \not\subseteq \Sigma$ and there is no Σ-concept E with $\mathcal{T} \models C \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq D$.*

Conditions 3(s1) and 3(s2) reflect the two causes of non-convexity in $\mathcal{EL}$ with closed predicates. The following example illustrates the requirement in Condition 3(s2) that no "interpolating" Σ-concept E exists.

Example 4.3. *Let $\mathcal{T} = \{A \sqsubseteq \exists r.E, E \sqsubseteq B\}$ and first assume that $\Sigma = \{r\}$. Then the inclusion $A \sqsubseteq \exists r.B$ satisfies Condition 3(s2) and thus $(\mathcal{T}, \Sigma)$ is not safe. Now assume $\Sigma = \{r, E\}$. Then, the inclusion $A \sqsubseteq \exists r.B$ does not violate safeness because E can be used as a 'Σ-interpolant'. Note that the ABox $\mathcal{A}$ from Example 4.1, which we used to refute convexity in a very similar situation, is simply unsatisfiable w.r.t. $(\mathcal{T}, \Sigma)$. Indeed, it can be shown that $(\mathcal{T}, \Sigma)$ is safe.*

The following theorem summarizes our main results for $\mathcal{EL}$.

Theorem 4.4 (Main Results for $\mathcal{EL}$)**.** *Let $(\mathcal{T}, \Sigma)$ be an $\mathcal{EL}$ TBox with closed predicates. Then*

1. *If $(\mathcal{T}, \Sigma)$ is not safe, then $(\mathcal{T}, \Sigma)$ is not convex and answering directed tree CQs w.r.t. $(\mathcal{T}, \Sigma)$ is* coNP*-hard.*
2. *If $(\mathcal{T}, \Sigma)$ is safe, then*
   - *(a) CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ coincides with CQ answering w.r.t. $(\mathcal{T}, \emptyset)$ for all ABoxes that are satisfiable w.r.t. $(\mathcal{T}, \Sigma)$, and $(\mathcal{T}, \Sigma)$ is convex;*
   - *(b) CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in* PTIME.

As mentioned before, directed tree CQs are also called $\mathcal{EL}$ instance queries in the literature. The mention of directed tree CQs in Point 1 of Theorem 4.4 thus implies that our results hold for CQs and $\mathcal{EL}$ instance queries alike.

Point 1 of Theorem 4.4 is proved by showing that non-safeness implies non-convexity, which involves two separate constructions that address Cases (s1) and (s2) from Definition 4.2. The proof of Point 2(a) of Theorem 4.4 is again via canonical models, which have to be defined in a rather careful way to make the proof go through. Establishing Point 2(b) involves showing that satisfiability of ABoxes w.r.t. safe $\mathcal{EL}$ TBoxes with closed predicates can be decided in PTIME.

Whereas it is obvious how to check the safeness of a DL-Lite TBox with closed predicates, this is not the case for $\mathcal{EL}$ TBoxes since Definition 4.2 quantifies over all concepts C, D, and E, of which there are infinitely many. In the following, we show that, nevertheless, safeness of an $\mathcal{EL}$-TBox with closed predicates $(\mathcal{T}, \Sigma)$ can be decided in PTIME. The first step is to convert $\mathcal{T}$ into a TBox $\mathcal{T}^*$ that is normalized in the sense that it satisfies the following properties:

(t1) $\mathcal{T}^*$ contains no CI of the form $C \sqsubseteq D_1 \sqcap D_2$;

(t2) if $C \sqsubseteq \exists r.D \in \mathcal{T}^*$, then there is no tlc $\exists r.C'$ of C with $\mathcal{T}^* \models C' \sqsubseteq D$.

Specifically, $\mathcal{T}^*$ can be produced by exhaustively replacing each CI $C \sqsubseteq D_1 \sqcap D_2$ with the two CIs $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$, and each CI $C \sqcap \exists r.C' \sqsubseteq \exists r.D$ where the TBox entails $C' \sqsubseteq D$ with the CI $C' \sqsubseteq D$. It is easy to see that the conversion takes only polynomial time (since subsumption in $\mathcal{EL}$ can be decided in PTIME) and that $\mathcal{T}^*$ is equivalent to $\mathcal{T}$, thus $\mathcal{T}^*$ is safe iff $\mathcal{T}$ is.

It thus suffices to consider $\mathcal{EL}$ TBoxes $(\mathcal{T}, \Sigma)$ where $\mathcal{T}$ is normalized. In this case, the following, stronger version of safeness is equivalent to the original version.

Definition 4.5. *An $\mathcal{EL}$ TBox with closed predicates $(\mathcal{T}, \Sigma)$ is* strongly safe *if there exists no $\mathcal{EL}$ inclusion $C \sqsubseteq \exists r.D \in \mathcal{T}$ such that one of the following is true:*

(c1) $r \notin \Sigma$ and there is some concept E such that $\mathcal{T} \models D \sqsubseteq E$ and $\mathsf{sig}(E) \cap \Sigma \neq \emptyset$;

(c2) $r \in \Sigma$, $\mathsf{sig}(D) \not\subseteq \Sigma$, and there is no Σ-concept E with $\mathcal{T} \models C \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq D$.

Note that, in Definition 4.5, the concepts C and D are now restricted to subconcepts of $\mathcal{T}$ (whereas E can still be any concept). The following is proved in the long version.

Lemma 4.6. *If $\mathcal{T}$ satisfies Conditions (t1) and (t2), then $(\mathcal{T}, \Sigma)$ is safe iff it is strongly safe.*

It thus remains to deal with the quantification over the concept E in Conditions (c1) and (c2). In the long version, we show that Condition (c1) can be checked in PTIME by carrying out a reachability test in a suitable canonical model of $\mathcal{T}$, and Condition (c2) can be checked in PTIME by executing a polynomial number of subsumption tests (the correctness of the latter relies on $\mathcal{EL}$ having a certain interpolation property). In summary, we obtain the following result.

Theorem 4.7. *Deciding safeness of $\mathcal{EL}$ TBox with closed predicates is* PTIME*-complete.*

5 First-Order Queries over Closed Predicates

As observed in [Reiter, 1992; Calvanese *et al.*, 2007b], closing predicates allows to use more expressive query languages without increasing the complexity of query answering. Indeed, mixing open and closed predicates seems particularly useful when large parts of the data stem from a relational database, as in the geographical database application from Section 2. In such a setup, one would typically like to use full FOQs or, in other words, SQL queries. We consider a query language that combines FOQs for closed predicates with CQs for open predicates. For safe TBoxes with closed predicates, such queries can be answered as efficiently as CQs both in the case of DL-Lite and of $\mathcal{EL}$.

As in the relational database setting, we allow only FOQs that are domain-independent and thus correspond to expressions of relational algebra (and SQL queries), see [Abiteboul *et al.*, 1995].

Definition 5.1 ($\text{CQ}^{\text{FO}(\Sigma)}$ queries)**.** *Let Σ be a signature that declares closed predicates. A* conjunctive query with FO(Σ) plugins (abbreviated $\text{CQ}^{\text{FO}(\Sigma)}$) *is of the form $\exists x_1 \cdots \exists x_n(\varphi_1 \wedge \ldots \wedge \varphi_m)$, where $n \geq 0$, $m \geq 1$, and each φ_i is an atom or a domain-independent FOQ with $\mathsf{sig}(\varphi_i) \subseteq \Sigma$.*

The next theorem shows that, for safe TBoxes with closed predicates, switching from CQs to $\text{CQ}^{\text{FO}(\Sigma)}$s does not increase data complexity. Thus, in addition to enforcing integrity constraints, such TBoxes have the virtue of admitting more expressive queries without an increase in complexity.

Theorem 5.2.

1. *For safe DL-Lite$_{\mathcal{R}}$ TBoxes with closed predicates $(\mathcal{T}, \Sigma)$, $CQ^{FO(\Sigma)}$-answering w.r.t. $(\mathcal{T}, \Sigma)$ is in* AC^0.
2. *For safe $\mathcal{EL}$ TBoxes with closed predicates $(\mathcal{T}, \Sigma)$, $CQ^{FO(\Sigma)}$-answering w.r.t. $(\mathcal{T}, \Sigma)$ is in* PTIME.

While the proof of Theorem 5.2 is not intricate, we believe that $\text{CQ}^{\text{FO}(\Sigma)}$ can be very useful for applications. Note that the query language EQL-Lite(CQ) from [Calvanese *et al.*, 2007b] can be viewed as a fragment of $\text{CQ}^{\text{FO}(\Sigma)}$ in which *only* closed predicates are admitted.

6 The Case of $\mathcal{ELI}$

We consider TBoxes formulated in $\mathcal{ELI}$, the extension of $\mathcal{EL}$ with inverse roles. $\mathcal{ELI}$ can be regarded as the logical core of expressive Horn DLs such as Horn-$\mathcal{SHIQ}$ [Hustadt *et al.*, 2007; Eiter *et al.*, 2008]. In contrast to the cases of DL-Lite$_{\mathcal{R}}$ and $\mathcal{EL}$, CQ answering with closed predicates turns out to not be inherently intractable in $\mathcal{ELI}$: there are $\mathcal{ELI}$ TBoxes with closed predicates $(\mathcal{T}, \Sigma)$ such that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in PTIME, but does *not* coincide with CQ answering w.r.t. $(\mathcal{T}, \emptyset)$ for all ABoxes that are satisfiable w.r.t. $(\mathcal{T}, \Sigma)$. We use the $\mathcal{ELI}$-TBox

$$\mathcal{T} = \{ \top \sqsubseteq \exists r.A \qquad \exists r^-.A \sqsubseteq B \qquad \exists r.(A \sqcap B) \sqsubseteq A \}$$

and the signature $\Sigma = \{r, B\}$. It is not hard to see that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ does not coincide with CQ answering w.r.t. $(\mathcal{T}, \emptyset)$. In particular, for

$$\mathcal{A} = \{r(a,a), B(a)\} \quad \text{and} \quad q() = \exists x\, r(x,x) \wedge A(x)$$

one can verify that $\mathcal{T}, \mathcal{A} \not\models q()$, but $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q()$; moreover, it is straightforward to construct a model which shows that $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$.

To prove our claim, it thus remains to show that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in PTIME. Let $\mathcal{A}$ be an ABox. The interpretation $\mathcal{I}$ is defined as follows:

(a) start with $\mathcal{A}$ viewed as an interpretation;

(b) add $a \in \mathsf{Ind}(\mathcal{A})$ to $A^{\mathcal{I}}$ if $r(a,b) \in \mathcal{A}$ implies $B(b) \in \mathcal{A}$;

(c) add $a \in \mathsf{Ind}(\mathcal{A})$ to $A^{\mathcal{I}}$ when $a \in (\exists r.(A \sqcap B))^{\mathcal{I}}$, repeat exhaustively.

Clearly, $\mathcal{I}$ can be constructed in polynomial time. The following lemma thus shows that CQ-answering w.r.t. $(\mathcal{T}, \Sigma)$ is in PTIME, and so is satisfiability of ABoxes w.r.t. $(\mathcal{T}, \Sigma)$. Note that $\mathcal{I}$ can be viewed as a canonical model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$.

Lemma 6.1.

1. *$\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$ iff $\mathcal{I}$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$;*
2. *if $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$, then for all CQs q and $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q[\vec{a}]$ iff $\mathcal{I} \models q[\vec{a}]$.*

We have thus shown that $\mathcal{ELI}$ behaves differently from DL-Lite$_{\mathcal{R}}$ and $\mathcal{EL}$. This raises a number of questions, dicussed in the next section.

7 Future Work

We have observed that, for simple DLs such as DL-Lite$_{\text{core}}$, DL-Lite$_{\mathcal{R}}$, and $\mathcal{EL}$, CQ answering with closed predicates is inherently intractable, while this is not the case for more expressive DLs such as $\mathcal{ELI}$. It would be interesting to conduct a broader study to fully understand this phenomenon, including additional TBox languages such as other extensions of DL-Lite, versions of Horn-$\mathcal{SHIQ}$, and possibly even members of the Datalog$^{\pm}$ family of ontology languages [Calì *et al.*, 2012].

Concerning the concrete case of $\mathcal{ELI}$, the observation presented in Section 6 raises the question whether there is a dichotomy between PTIME and CONP for CQ answering w.r.t. $\mathcal{ELI}$-TBoxes with closed predicates, and how the PTIME cases can be characterized. It also asks for a characterization of those $\mathcal{ELI}$-TBoxes with closed predicates $(\mathcal{T}, \Sigma)$ for which CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ coincides with CQ answering w.r.t. $(\mathcal{T}, \emptyset)$. We leave these questions as interesting future work.

Acknowledgments. Carsten Lutz and İnanç Seylan were supported by the DFG SFB/TR 8 "Spatial Cognition".

References

[Abiteboul *et al.*, 1995] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995.

[Artale *et al.*, 2009] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The DL-Lite family and relations. *Journal of Artificial Intelligence Research*, 36:1–69, 2009.

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

[Baader *et al.*, 2005] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In *IJCAI*, 2005.

[Calì *et al.*, 2012] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *Journal of Web Semantics*, 14:57–83, 2012.

[Calvanese *et al.*, 2007a] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *Journal of Automated Reasoning*, 39(3):385–429, 2007.

[Calvanese *et al.*, 2007b] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. EQL-Lite: Effective first-order query processing in description logics. In *IJCAI*, 2007.

[Codescu *et al.*, 2011] Mihai Codescu, Gregor Horsinka, Oliver Kutz, Till Mossakowski, and Rafaela Rau. DO-ROAM: Activity-oriented search and navigation with OpenStreetMap. In *GeoSpatial Semantics*, 2011.

[Donini *et al.*, 2002] Francesco M. Donini, Daniele Nardi, and Riccardo Rosati. Description logics of minimal knowledge and negation as failure. *ACM Transactions on Computational Logic*, 3(2):177–225, 2002.

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

[Franconi *et al.*, 2011] Enrico Franconi, Yazmin Angélica Ibáñez-García, and Inanç Seylan. Query answering with DBoxes is hard. *Electronic Notes in Theoretical Computer Science*, 278:71–84, 2011.

[Grimm and Motik, 2005] Stephan Grimm and Boris Motik. Closed world reasoning in the semantic web through epistemic operators. In *OWLED*, 2005.

[Hübner *et al.*, 2004] Sebastian Hübner, Rainer Spittel, Ubbo Visser, and Thomas J. Vögele. Ontology-based search for interactive digital maps. *IEEE Intelligent Systems*, 19(3):80–86, 2004.

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

[Kontchakov *et al.*, 2010] Roman Kontchakov, Carsten Lutz, David Toman, Frank Wolter, and Michael Zakharyaschev. The combined approach to query answering in DL-Lite. In *KR*, 2010.

[Lutz and Wolter, 2010] Carsten Lutz and Frank Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *Journal of Symbolic Computation*, 45(2):194–228, 2010.

[Lutz and Wolter, 2012] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *KR*, 2012.

[Lutz *et al.*, 2009] Carsten Lutz, David Toman, and Frank Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *IJCAI*, 2009.

[Mehdi *et al.*, 2011] Anees Mehdi, Sebastian Rudolph, and Stephan Grimm. Epistemic querying of OWL knowledge bases. In *ESWC*, 2011.

[Motik and Rosati, 2010] Boris Motik and Riccardo Rosati. Reconciling description logics and rules. *Journal of the ACM*, 57(5):1–62, 2010.

[Motik *et al.*, 2009] Boris Motik, Ian Horrocks, and Ulrike Sattler. Bridging the gap between OWL and relational databases. *Journal of Web Semantics*, 7(2):74–89, 2009.

[Reiter, 1992] Raymond Reiter. What should a database know? *Journal of Logic Programming*, 14(1&2):127–153, 1992.

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

[Sengupta *et al.*, 2011] Kunal Sengupta, Adila Alfa Krisnadhi, and Pascal Hitzler. Local closed world semantics: Grounded circumscription for OWL. In *ISWC*, 2011.

[Seylan *et al.*, 2009] Inanç Seylan, Enrico Franconi, and Jos de Bruijn. Effective query rewriting with ontologies over DBoxes. In *IJCAI*, 2009.

[ten Cate *et al.*, 2011] Balder ten Cate, Enrico Franconi, and Inanç Seylan. Beth definability in expressive description logics. In *IJCAI*, 2011.

A Additional Preliminaries

A CQ $q(x)$ with one answer variable x is a *directed tree CQ* if it satisfies the following conditions:

1. the directed graph $G_q = (V_q, E_q)$ is a tree with root x, where V_q is the set of terms used in q and E_q contains an edge (t_1, t_2) whenever there is an atom $r(t_1, t_2)$ in q;
2. if $r(x, y), s(x, y)$ are conjuncts of $q(x)$ then $r = s$

A CQ $q(x)$ with one answer variable x is a *tree CQ* if it satisfies the following conditions:

1. G_q is a tree when viewed as an *undirected* graph;
2. if $r(x, y), s(x, y)$ are conjuncts of $q(x)$ then $r = s$;
3. there are no conjuncts $r(x, y), s(y, x)$ in $q(x)$.

Every ABox $\mathcal{A}$ corresponds to an interpretation $\mathcal{I}_\mathcal{A}$ whose domain is $\mathsf{Ind}(\mathcal{A})$ and in which $a \in A^{\mathcal{I}_\mathcal{A}}$ iff $A(a) \in \mathcal{A}$, for all $A \in \mathsf{N_C}$ and $a \in \mathsf{Ind}(\mathcal{A})$, and similarly for role names. Conversely, every interpretation $\mathcal{I}$ corresponds to a (possibly infinite) ABox $\mathcal{A}_\mathcal{I}$ whose individual names are $\Delta^\mathcal{I}$.

B Proofs for Section 3

We prove coNP-hardness for Example 3.1. Recall that $\mathcal{T} = \{A \sqsubseteq \exists r.\top, \exists r^-.\top \sqsubseteq B\}$ and $\Sigma = \{r\}$. The coNP-hardness proof is by reduction of 2+2-SAT, a variant of propositional satisfiability that was first introduced by Schaerf as a tool for establishing lower bounds for the data complexity of query answering in a DL context [Schaerf, 1993]. In fact, our proof is very similar to Schaerf's original proof. A *2+2 clause* is of the form $(p_1 \vee p_2 \vee \neg n_1 \vee \neg n_2)$, where each of p_1, p_2, n_1, n_2 is a propositional letter or a truth constant 0, 1. A *2+2 formula* is a finite conjunction of 2+2 clauses. Now, 2+2-SAT is the problem of deciding whether a given 2+2 formula is satisfiable. It is shown in [Schaerf, 1993] that 2+2-SAT is NP-complete.

Let $\varphi = c_0 \wedge \cdots \wedge c_n$ be a 2+2 formula in propositional letters $w_0, \ldots, w_m$, and let $c_i = p_{i,1} \vee p_{i,2} \vee \neg n_{i,1} \vee \neg n_{i,2}$ for all $i \leq n$. Our aim is to define an ABox $\mathcal{A}_\varphi$ and an instance query $C(a)$ such that φ is unsatisfiable iff $\mathcal{T}, \mathcal{A}_\varphi \models_{c(\Sigma)} q$. To start, we represent the formula φ in the ABox $\mathcal{A}_\varphi$ as follows:

- the individual name f represents the formula φ;
- the individual names $c_0, \ldots, c_n$ represent the clauses of φ;
- the assertions $c(f, c_0), \ldots, c(f, c_n)$, associate f with its clauses, where c is a role name that does not occur in $\mathcal{T}$;
- the individual names $w_0, \ldots, w_m$ represent variables, and the individual names $0, 1$ represent truth constants;
- the assertions

  $$\bigcup_{i \leq n} \{p_1(c_i, p_{i,1}), p_2(c_i, p_{i,2}), n_1(c_i, n_{i,1}), n_2(c_i, n_{i,2})\}$$

  associate each clause with the four variables/truth constants that occur in it, where p_1, p_2, n_1, n_2 are role names that do not occur in $\mathcal{T}$.

We further extend $\mathcal{A}_\varphi$ to enforce a truth value for each of the variables w_i and the truth-constants $0, 1$. To this end, add to $\mathcal{A}_\varphi$ copies $\mathcal{A}_0, \ldots, \mathcal{A}_m$ of the ABox $\mathcal{A}$ from Example 3.1 obtained by renaming individual names such that $\mathsf{Ind}(\mathcal{A}_i) \cap \mathsf{Ind}(\mathcal{A}_j) = \emptyset$ whenever $i \neq j$. Moreover, assume that a_i coincides with the ith copy of a. Intuitively, the copy $\mathcal{A}_i$ of $\mathcal{A}$ is used to generate a truth value for the variable w_i, where we want to interpret w_i as true in an interpretation $\mathcal{I}$ if $\mathcal{I} \models q_1[w_i]$ and as false if $\mathcal{I} \models q_2[w_i]$, where q_1 and q_2 are the queries from Example 3.1.

To ensure that 0 and 1 have the expected truth values, add the ABoxes

$$\begin{aligned} \mathcal{A}(1) &= \{r(1, c_1), A_1(c_1), B(c_1)\} \text{ and} \\ \mathcal{A}(0) &= \{r(0, c_2), A_2(c_2), B(c_2)\}. \end{aligned}$$

Let $\mathcal{B}$ be the resulting ABox. Consider the query

$$\begin{aligned} q_0 \;=\; \exists y \exists z_1 \exists z_2 \exists z_3 \exists z_4\, c(x, y) &\wedge p_1(y, z_1) \wedge \mathsf{ff}(z_1) \\ &\wedge p_2(y, z_2) \wedge \mathsf{ff}(z_2) \\ &\wedge n_1(y, z_3) \wedge \mathsf{tt}(z_3) \\ &\wedge n_2(y, z_4) \wedge \mathsf{tt}(z_4) \end{aligned}$$

which describes the existence of a clause with only false literals and thus captures falsity of φ, where $\mathsf{tt}(z_i)$ is an abbreviation for $q_1(x)$ with the free variable x replaced by z_i and ff is an abbreviation for $q_2(x)$ with the free variable x replaced by z_i. It is straightforward to show that φ is unsatisfiable iff $\mathcal{T}, \mathcal{B} \models_{c(\Sigma)} q_0$.

We now establish Theorem 3.6, starting with Point 1. We write tree CQs in the form $C(x)$ with C an $\mathcal{ELI}$ concept.

Lemma B.1. *Let $(\mathcal{T}, \Sigma)$ be a DL-Lite$_\mathcal{R}$ TBox with closed predicates. If $(\mathcal{T}, \Sigma)$ is not safe, then $(\mathcal{T}, \Sigma)$ is not convex and there exists a tree CQ $C(x)$ such that answering $C(x)$ w.r.t. $(\mathcal{T}, \Sigma)$ is co*NP*-hard.*

Proof. Assume that B_1, B_2 and role r show that $(\mathcal{T}, \Sigma)$ is not safe and satisfy Points 1 to 4 of Definition 3.4. By satisfiability of B_1 w.r.t. $\mathcal{T}$, Point 4, and the finite model property of DL-Lite$_\mathcal{R}$ one can find a finite model $\mathcal{I}$ of $\mathcal{T}$ with $a_0 \in B_1^\mathcal{I}$ such that $a_0 \notin (\exists r'.\top)^\mathcal{I}$ for any role r' with $\mathsf{sig}(r') \subseteq \Sigma$ and $\mathcal{T} \models r' \sqsubseteq r$. Let $\mathcal{I}_r$ be the interpretation obtained from $\mathcal{I}$ by removing all pairs (a_0, b) from any $r'^\mathcal{I}$ with $\mathcal{T} \models r' \sqsubseteq r$. By Point 3, $B_1 \neq \exists r'.\top$ for any such r'. Thus we have $a_0 \in B_1^{\mathcal{I}_r}$. Take the ABox $\mathcal{A}_r$ corresponding to $\mathcal{I}_r$ and let $\mathcal{A}$ be the disjoint union of two copies of $\mathcal{A}_r$. We denote the individual names of the first copy by $(b, 1)$, $b \in \Delta^\mathcal{I}$, and the elements of the second copy by $(b, 2)$, $b \in \Delta^\mathcal{I}$. Let $\mathcal{A}'$ be

$$\begin{aligned} &\mathcal{A} \cup \\ &\{A_1(b, 1) \mid b \in B_2^\mathcal{I}\} \cup \{A_2(b, 2) \mid b \in B_2^\mathcal{I}\} \cup \\ &\{r'((a_0, i), (b, j)) \mid (a_0, b) \in r'^\mathcal{I}, \mathcal{T} \not\models r' \sqsubseteq r, \mathsf{sig}(r') \subseteq \Sigma, \\ &\qquad i, j \in \{1, 2\}\} \end{aligned}$$

where A_1 and A_2 are fresh concept names.

Now one can show that convexity fails, that is, for the tree CQs

$$q_i = \exists y\, r(x, y) \wedge A_i(y) \wedge B_2(y),$$

and with $\mathcal{A}'$ the ABox defined above, we have

(a) $(\mathcal{T},\mathcal{A}') \models_{c(\Sigma)} q_1 \vee q_2[a_0,1]$,

(b) $(\mathcal{T},\mathcal{A}') \not\models_{c(\Sigma)} q_i[a_0,1]$ for $i=1,2$.

The coNP-hardness proof is now similar to the proof for Example 3.1 given above and omitted.

(a) Let $\mathcal{J}$ be a model of $(\mathcal{T},\Sigma)$ and $\mathcal{A}'$. We have $(a_0,1) \in B_1^{\mathcal{J}}$ (since, by Point 3, $B_1 \neq \exists r'.\top$ for every r' with $\mathcal{T} \models r' \sqsubseteq r$). Since $\mathcal{J}$ is a model of $\mathcal{T}$, $\mathcal{T} \models B_1 \sqsubseteq \exists r.\top$ and $\mathcal{T} \models \exists r^-.\top \sqsubseteq B_2$, there exists $e \in \Delta^{\mathcal{J}}$ with $((a_0,1),e) \in r^{\mathcal{J}}$ and $e \in B_2^{\mathcal{J}}$. Since $\mathsf{sig}(B_2) \subseteq \Sigma$, and by the definition of $\mathcal{A}'$, e is of the form (e',i) with $e' \in B_2^{\mathcal{I}}$ and $i \in \{1,2\}$. If $i=1$, we have $A_1(e',1) \in \mathcal{A}'$ and so $(a_0,1) \in \exists r.(A_1 \sqcap B_2)^{\mathcal{J}}$, as required. If $i=2$, we have $A_2(e',2) \in \mathcal{A}'$ and so $(a_0,1) \in \exists r.(A_2 \sqcap B_2)^{\mathcal{J}}$, as required.

(b) We construct a model $\mathcal{J}$ of $(\mathcal{T},\Sigma)$ and $\mathcal{A}'$ with $(a_0,1) \notin (\exists r.(A_1 \sqcap B_2))^{\mathcal{J}}$. $\mathcal{J}$ is defined as the interpretation corresponding to the ABox $\mathcal{A}'$ extended by

$$\{r'((a_0,1),(e,2)) \mid (a_0,e) \in r'^{\mathcal{I}}\} \cup$$
$$\{r'((a_0,2)),(e,1)) \mid (a_0,e) \in r'^{\mathcal{I}}\},$$

for all roles r' such that $\mathsf{sig}(r') \cap \Sigma = \emptyset$ and $\mathcal{T} \models r' \sqsubseteq r$, and

$$\{r'((a_0,i),(e,j)) \mid (a_0,e) \in r'^{\mathcal{I}}, i,j \in \{1,2\}\},$$

for all roles r' with $\mathsf{sig}(r') \cap \Sigma = \emptyset$ and $\mathcal{T} \not\models r' \sqsubseteq r$.

Clearly $(a_0,1) \notin (\exists r.(A_1 \sqcap B_2))^{\mathcal{J}}$. Thus it is remains to show that $\mathcal{J}$ is a model of $(\mathcal{T},\Sigma)$ and $\mathcal{A}'$. Since no symbol from Σ has changed its interpretation, it is sufficient to show that $\mathcal{J}$ satisfies all concept and role inclusions in $\mathcal{T}$.

Let $s \sqsubseteq s'$ be a role inclusion in $\mathcal{T}$. Since $\mathcal{I}$ is a model of $\mathcal{T}$, the only pairs where $s \sqsubseteq s'$ can possible be refuted are of the form $((a_0,i),(b,j))$ with $i,j \in \{1,2\}$. Assume $((a_0,i),(b,j)) \in s^{\mathcal{J}}$. Then, by definition, $(a_0,b) \in s^{\mathcal{I}}$ and so $(a_0,b) \in s'^{\mathcal{I}}$ because $\mathcal{I}$ is a model of $\mathcal{T}$. We distinguish the following cases:

- $\mathcal{T} \not\models s' \sqsubseteq r$. Then, by definition of $\mathcal{J}$, $((a_0,i),(b,j)) \in s'^{\mathcal{J}}$ since $((a_0,i'),(b,j')) \in s'^{\mathcal{J}}$ for all $i',j' \in \{1,2\}$.
- $\mathcal{T} \models s' \sqsubseteq r$. Then $\mathcal{T} \models s \sqsubseteq r$. Note that, by construction of $\mathcal{I}$, $\mathsf{sig}(s) \cap \Sigma = \emptyset$ and $\mathsf{sig}(s') \cap \Sigma = \emptyset$. Hence, by construction of $\mathcal{J}$, $(i,j)=(1,2)$ or $(i,j)=(2,1)$. In both cases we have $((a_0,i),(b,j)) \in s'^{\mathcal{J}}$ as well.

To prove that all concept inclusions of $\mathcal{T}$ are satisfied in $\mathcal{J}$ observe that $B^{\mathcal{J}} = (B^{\mathcal{I}} \times \{1\}) \cup (B^{\mathcal{I}} \times \{2\})$ holds for all DL-Lite concepts B. Thus, $\mathcal{J}$ satisfies all concept inclusions satisfied in $\mathcal{I}$ and, therefore, is a model of any concept inclusion in $\mathcal{T}$, as required.

To construct a model $\mathcal{J}$ of $(\mathcal{T},\Sigma)$ and $\mathcal{A}'$ with $(a_0,1) \notin (\exists r.(A_2 \sqcap B_2))^{\mathcal{J}}$ swap the roles of the two copies of $\mathcal{I}_r$. $\square$

Since it is known that DL-Lite$_{\mathcal{R}}$ is convex, the following Lemma proves Point 2(a) of Theorem 3.6.

Lemma B.2. *Let $(\mathcal{T},\Sigma)$ be a DL-Lite$_{\mathcal{R}}$ TBox with closed predicates. If $(\mathcal{T},\Sigma)$ is safe, then CQ answering w.r.t. $(\mathcal{T},\Sigma)$ coincides with CQ answering w.r.t. $(\mathcal{T},\emptyset)$ for ABoxes that are satisfiable w.r.t. $(\mathcal{T},\Sigma)$.*

Proof. Let $(\mathcal{T},\Sigma)$ be safe and assume that $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T},\Sigma)$. We remind the reader of the construction of a canonical model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ (without closed predicates!) [Kontchakov *et al.*, 2010]. $\mathcal{I}$ is the interpretation corresponding to an ABox $\mathcal{A}_c$ that is the limit of a sequence of ABoxes $\mathcal{A}_0,\mathcal{A}_1,\ldots$. Let $\mathcal{A}_0 = \mathcal{A}$ and assume $a_0,\ldots$ is an infinite list of individual names such that $\mathsf{Ind}(\mathcal{A}_0) = \{a_0,\ldots,a_k\}$. Assume $\mathcal{A}_j$ has been defined already. If there exist i,ℓ such that there exists r,s with

- $\mathcal{T} \models r \sqsubseteq s$;
- $\mathcal{A}_j \models r(a_i,a_\ell)$; and
- $\mathcal{A}_j \not\models s(a_i,a_\ell)$;

then set $\mathcal{A}_{j+1} = \mathcal{A}_j \cup \{s(a_i,a_\ell)\}$. Otherwise let i be minimal such that there exist B_1,B_2 with

- $\mathcal{T} \models B_1 \sqsubseteq B_2$;
- $\mathcal{A}_j \models B_1(a_i)$; and
- $\mathcal{A}_j \not\models B_2(a_i)$;

(if no such i exists, then set $\mathcal{A}_c := \mathcal{A}_j$). Then

- if B_2 is a concept name, let $\mathcal{A}_{j+1} = \mathcal{A}_j \cup \{B_2(a_i)\}$;
- if $B_2 = \exists s.\top$, then take a fresh individual $b_{a_i,s}$ and set $\mathcal{A}_{j+1} = \mathcal{A}_j \cup \{s(a_i,b_{a_i,s})\}$.

Now let $\mathcal{J}$ be the interpretation corresponding to the ABox $\mathcal{A}_c = \bigcup_{i \geq 0} \mathcal{A}_i$. It is known that $\mathcal{J}$ is a model of $(\mathcal{T},\mathcal{A})$ with the following properties:

(a) For all CQs $q(\vec{x})$ and $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$: $\mathcal{T},\mathcal{A} \models q(\vec{a})$ iff $\mathcal{J} \models q[\vec{a}]$.

(b) For any individual $b_{a_i,s} \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A})$ introduced as a witness for some $B_2 = \exists s.\top$, we have $B(b_{a_i,s}) \in \mathcal{A}_c$ iff $\mathcal{T} \models \exists s^-.\top \sqsubseteq B$, for every DL-Lite concept B.

To show that $\mathcal{J}$ is a model of $(\mathcal{T},\Sigma)$ and $\mathcal{A}$ it is sufficient to prove that every assertion using Σ-symbols in $\mathcal{A}_c$ is contained in $\mathcal{A}$. This follows from Claim 1 and Claim 2 below:

Claim 1. For all $a,b \in \mathsf{Ind}(\mathcal{A}_0)$,

- if $B(a) \in \mathcal{A}_c$ and $\mathsf{sig}(B) \subseteq \Sigma$, then $B(a) \in \mathcal{A}_0$; and
- if $r(a,b) \in \mathcal{A}^c$ and $\mathsf{sig}(r) \subseteq \Sigma$, then $r(a,b) \in \mathcal{A}_0$.

Claim 1 follows from Point (a) above and the assumption that $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T},\Sigma)$. For assume that $B(a) \in \mathcal{A}_c$ but $B(a) \notin \mathcal{A}_0 = \mathcal{A}$. Then $\mathcal{T},\mathcal{A} \models B(a)$ and, therefore, $\mathcal{T},\mathcal{A} \models_{c(\Sigma)} B(a)$. But then satisfiability of $\mathcal{A}$ w.r.t. $(\mathcal{T},\Sigma)$ implies $\mathsf{sig}(B) \cap \Sigma = \emptyset$. The argument for role assertions $r(a,b)$ is similar and omitted.

Claim 2. For any $a \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A}_0)$ there does not exist any DL-Lite concept B with $\mathsf{sig}(B) \subseteq \Sigma$ such that $\mathcal{A}_c \models B(a)$.

For a proof by contradiction assume that there exist an $a \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A}_0)$ and DL-Lite concept B with $\mathsf{sig}(B) \subseteq \Sigma$ such that $\mathcal{A}_c \models B(a)$. Let a be the first such individual introduced in the construction of $\mathcal{A}_c$. By Point (b) above and the construction of $\mathcal{A}_c$ there exist B_1, r, a' and $i \geq 0$ such that

- $\mathcal{T} \models B_1 \sqsubseteq \exists r.\top$;
- $\mathcal{A}_i \models B_1(a')$;

- $\mathcal{A}_i \not\models \exists r.\top(a')$;
- $r(a', b_{a',r}) \in \mathcal{A}_{i+1}$ for $a = b_{a',r}$;
- $\mathcal{T} \models \exists r^-.\top \sqsubseteq B_2$, where $B_2 := B$.

We show that B_1, B_2, and r satisfy Conditions 1 to 4 from Definition 3.4 and thus derive a contradiction to the assumption that $(\mathcal{T}, \Sigma)$ is safe. Points 1 and 2 are clear. For Point 3, assume that $B_1 = \exists r'.\top$ for some r' such that $\mathcal{T} \models r' \sqsubseteq r$. Then $r'(a', e) \in \mathcal{A}_i$ for some e. But then, since witnesses for role inclusions are added to $\mathcal{A}_i$ before witnesses for concept inclusions are added to $\mathcal{A}_i$, we have $r(a', e) \in \mathcal{A}_i$ which contradicts $\mathcal{A}_i \not\models \exists r.\top(a')$. For Point 4 assume that $\mathcal{T} \models B_1 \sqsubseteq \exists r'.\top$ for some role r' such that $\mathsf{sig}(r') \subseteq \Sigma$ and $\mathcal{T} \models r' \sqsubseteq r$. Then $\mathcal{A}_c \models \exists r'.\top(a')$. Then $a' \in \mathsf{Ind}(\mathcal{A})$ because otherwise a' is an individual introduced before a such that $\mathcal{A}_c \models B'(a')$ for some B' with $\mathsf{sig}(B') \subseteq \Sigma$, and we have derived a contradiction. Then by Claim 1 and since $\mathsf{sig}(r') \subseteq \Sigma$, we have $r'(a', e) \in \mathcal{A}$ for some e. But then, again since witnesses for role inclusions are added to $\mathcal{A}_i$ before witnesses for concept inclusions are added to $\mathcal{A}_i$, $r(a', e) \in \mathcal{A}_i$ which contradicts $\mathcal{A}_i \not\models \exists r.\top(a')$.

It follows from Claims 1 and 2 that $\mathcal{J}$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$. Thus, we have for all CQs $q(\vec{x})$ and $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$: if $\mathcal{T}, \mathcal{A} \not\models q(\vec{a})$, then $\mathcal{J} \not\models q[\vec{a}]$, and so $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} q(\vec{a})$, as required. □

To obtain a proof of Point 2(b) of Theorem 3.6, it remains to show the following.

Lemma B.3. *Let $(\mathcal{T}, \Sigma)$ be a safe DL-Lite$_\mathcal{R}$ TBox with closed predicates. Then given an ABox $\mathcal{A}$, the satisfiability of $\mathcal{A}$ w.r.t. $(\mathcal{T}, \Sigma)$ is in* AC^0.

Proof (sketch). We claim that an ABox $\mathcal{A}$ is satisfiable w.r.t. a safe $(\mathcal{T}, \Sigma)$ iff (i) $\mathcal{A}$ is satisfiable w.r.t. $\mathcal{T}$, (ii) $\mathcal{T}, \mathcal{A} \models B(a)$ implies $\mathcal{A} \models B(a)$ for all DL-Lite concepts B over Σ, and (iii) $\mathcal{T}, \mathcal{A} \models r(a, b)$ implies $\mathcal{A} \models r(a, b)$ for all roles r over Σ.

Now we show that the three conditions above are in AC^0. (i) is already known to be in AC^0 [Calvanese *et al.*, 2007a]. To see that the other two conditions are in AC^0, let $\varphi_B(x)$ be a FOQ with $\mathcal{T}, \mathcal{A} \models B(a)$ iff $\mathcal{I}_\mathcal{A} \models \varphi_B(a)$, where $\mathcal{I}_\mathcal{A}$ is the interpretation corresponding to $\mathcal{A}$; and let $\varphi_r = \bigvee_{\mathcal{T} \models s \sqsubseteq r} s(x, y)$. Note that such a $\varphi_B(x)$, whose size depends only on $|\mathcal{T}|$ and $|B|$, always exists in DL-Lite$_\mathcal{R}$ [Calvanese *et al.*, 2007a]. Then (ii) or (iii) is not satisfied iff

$$\mathcal{I}_\mathcal{A} \models \varphi_1 \vee \varphi_2,$$

where

$$\varphi_1 = \exists x \bigvee_{B \in X} (\varphi_B(x) \wedge \neg B(x))$$

and X is the set of all DL-Lite concepts over Σ and

$$\varphi_2 = \exists x \exists y \bigvee_{r \in Y} (\varphi_r(x, y) \wedge \neg r(x, y))$$

where Y denotes the set of all role names in Σ. It is not hard to see that $\mathcal{I}_\mathcal{A} \models \varphi_1 \vee \varphi_2$ is in AC^0. □

C Some Preliminaries for $\mathcal{EL}$

We present the canonical model construction for $\mathcal{EL}$ and show an interpolation property.

C.1 Canonical Models for $\mathcal{EL}$

We start by introducing canonical models of $\mathcal{EL}$ TBoxes and concepts that were first introduced in [Lutz and Wolter, 2010]. Canonical models are finite and of polynomial size in the size of the input TBox and concept. After presenting some lemmas from [Lutz and Wolter, 2010], we define tree-shaped canonical models. Intuitively, tree-shaped canonical models correspond to tree unfoldings of standard, finite, canonical models; however, for our purposes it is more useful to give a syntactic construction.

The *canonical model* $\mathcal{I}_{\mathcal{T},C} = (\Delta^{\mathcal{I}_{\mathcal{T},C}}, \cdot^{\mathcal{I}_{\mathcal{T},C}})$ of $\mathcal{T}$ and C is defined as follows:

- $\Delta^{\mathcal{I}_{\mathcal{T},C}} = \{a_C\} \cup \{a_{C'} \mid \exists r.C' \in \mathsf{sub}(C) \cup \mathsf{sub}(\mathcal{T})\}$;
- $a_{D_0} \in A^{\mathcal{I}_{\mathcal{T},C}}$ iff $\mathcal{T} \models D_0 \sqsubseteq A$, for all $A \in \mathsf{N_C}$ and $a_{D_0} \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$;
- $(a_{D_0}, a_{D_1}) \in r^{\mathcal{I}_{\mathcal{T},C}}$ iff $\mathcal{T} \models D_0 \sqsubseteq \exists r.D_1$ and $\exists r.D_1 \in \mathsf{sub}(\mathcal{T})$ or $\exists r.D_1$ is a tlc of D_0, for all $a_{D_0}, a_{D_1} \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$ and $r \in \mathsf{N_R}$.

Since concept subsumption in $\mathcal{EL}$ is in PTIME, $\mathcal{I}_{\mathcal{T},C}$ can be constructed in time polynomial in $\mathcal{T}$ and C. The following result was shown in [Lutz and Wolter, 2010] (Lemma 13).

Lemma C.1. *Let C be an $\mathcal{EL}$ concept and $\mathcal{T}$ an $\mathcal{EL}$ TBox. Then*

- *$\mathcal{I}_{\mathcal{T},C}$ is a model of $\mathcal{T}$;*
- *for all D_0 with $a_{D_0} \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$ and all $\mathcal{EL}$ concepts D_1: $\mathcal{T} \models D_0 \sqsubseteq D_1$ iff $a_{D_0} \in D_1^{\mathcal{I}_{\mathcal{T},C}}$*

The following result is shown in [Lutz and Wolter, 2010] (Lemma 16). It follows from the definition of canonical models by Lemma C.1.

Lemma C.2. *Suppose $\mathcal{T} \models C \sqsubseteq \exists r.D$, where C, D are $\mathcal{EL}$ concepts and $\mathcal{T}$ is an $\mathcal{EL}$ TBox. Then one of the following holds:*

- *there is a tlc $\exists r.C'$ of C such that $\mathcal{T} \models C' \sqsubseteq D$;*
- *there is a $\exists r.C' \in \mathsf{sub}(\mathcal{T})$ such that $\mathcal{T} \models C \sqsubseteq \exists r.C'$ and $\mathcal{T} \models C' \sqsubseteq D$.*

Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $\mathcal{A}$ a (possibly infinite) ABox. In the construction of the tree-shaped canonical model for $(\mathcal{T}, \mathcal{A})$, we use *extended ABoxes*, i.e., sets of assertions of the form $r(a, b)$ and $C(a)$, where r is a role name and C a *possibly compound* $\mathcal{EL}$ concept. We produce a sequence of extended ABoxes $\mathcal{A}_0, \mathcal{A}_1, \ldots$, starting with $\mathcal{A}_0 = \mathcal{A}$. In what follows, we use additional individual names of the form $a \cdot r_1 \cdot C_1 \cdots r_k \cdot C_k$ with $a \in \mathsf{Ind}(\mathcal{A}_0)$, $r_1, \ldots, r_k$ role names that occur in $\mathcal{T}$, and $C_1, \ldots, C_k \in \mathsf{sub}(\mathcal{T})$. Each extended ABox $\mathcal{A}_{i+1}$ is obtained from $\mathcal{A}_i$ by applying the following rules:

R1 if $C \sqcap D(a) \in \mathcal{A}_i$, then add $C(a)$ and $D(a)$ to $\mathcal{A}_i$;

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

R3 if $\exists r.C(a) \in \mathcal{A}_i$ and there exist $b \in \mathcal{A}_i$ with $r(a,b) \in \mathcal{A}_i$ and $\mathcal{T}, \mathcal{A}_i \models C(b)$, then add $C(b)$ to $\mathcal{A}_i$; otherwise add $r(a, a \cdot r \cdot C)$ and $C(a \cdot r \cdot C)$ to $\mathcal{A}_i$.

Let $\mathcal{A}_c = \bigcup_{i \geq 0} \mathcal{A}_i$. Note that $\mathcal{A}_c$ may be infinite even if $\mathcal{A}$ is finite, and that none of the above rules adds anything to $\mathcal{A}_c$. Denote by $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ the interpretation corresponding to $\mathcal{A}_c$. The following lemma is standard:

Lemma C.3. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $\mathcal{A}$ a possibly infinite ABox. Then*

- *$\mathcal{J}_{\mathcal{T},\mathcal{A}}$ is a model of $\mathcal{T}$ and $\mathcal{A}$;*
- *for all $p \in \Delta^{\mathcal{J}_{\mathcal{T},\mathcal{A}}} \setminus \mathsf{Ind}(\mathcal{A})$ and all $\mathcal{EL}$ concepts D: $p \in D^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$ iff $\mathcal{T} \models \mathsf{tail}(p) \sqsubseteq D$;*
- *for all CQs $q(\vec{x})$ and $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$: $\mathcal{T}, \mathcal{A} \models q(\vec{a})$ iff $\mathcal{J}_{\mathcal{T},\mathcal{A}} \models q[\vec{a}]$.*

We now construct the tree-shaped canonical model of an $\mathcal{EL}$ TBox $\mathcal{T}$ and a $\mathcal{EL}$ concept C. A *path* in a concept C is a finite sequence $C_0 \cdot r_1 \cdot C_1 \cdots r_n \cdot C_n$, where $C_0 = C$, $n \geq 0$, and $\exists r_{i+1}.C_{i+1}$ is a tlc of C_i, for $0 \leq i < n$. We use $\mathsf{paths}(C)$ to denote the set of paths in C. If $p \in \mathsf{paths}(C)$, then $\mathsf{tail}(p)$ denotes the last element of p. The *canonical ABox* $\mathcal{A}_C$ associated with C is defined by setting

$$\begin{aligned}\mathcal{A}_C \quad = \quad & \{r(p,q) \mid p,q \in \mathsf{paths}(C); q = p \cdot r \cdot C'\} \\ & \{A(p) \mid A \text{ a tlc of } \mathsf{tail}(p), p \in \mathsf{paths}(C)\}\end{aligned}$$

Noe let the tree-shaped canonincal model $\mathcal{J}_{\mathcal{T},C}$ be defined as $\mathcal{J}_{\mathcal{T},C} = \mathcal{J}_{\mathcal{T},\mathcal{A}_C}$. The following can be proved in a straightforward way.

Lemma C.4. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and C a concept. Then*

- *$\mathcal{J}_{\mathcal{T},C}$ is a model of $\mathcal{T}$;*
- *for all $p \in \Delta^{\mathcal{J}_{\mathcal{T},C}}$ and all $\mathcal{EL}$ concepts D: $p \in D^{\mathcal{J}_{\mathcal{T},C}}$ iff $\mathcal{T} \models \mathsf{tail}(p) \sqsubseteq D$.*

C.2 Interpolation for $\mathcal{EL}$

We require a certain interpolation property. This interpolation property has been studied before for $\mathcal{ALC}$ and several of its extensions in the context of query rewriting for DBoxes and Beth definability [Seylan *et al.*, 2009; ten Cate *et al.*, 2011]. Note that it is different from the interpolation property investigated in [Lutz and Wolter, 2010], which requires the interpolant to be a TBox instead of a concept.

Lemma C.5 ($\mathcal{EL}$ Interpolation)**.** *Let $\mathcal{T}_1, \mathcal{T}_2$ be $\mathcal{EL}$ TBoxes and let D_0, D_1 be $\mathcal{EL}$ concepts. Assume $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq D_1$ with $\mathsf{sig}(\mathcal{T}_1, D_0) \cap \mathsf{sig}(\mathcal{T}_2, D_1) \subseteq \Sigma$. Then there exists an $\mathcal{EL}$ concept F in Σ such that $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq F$ and $\mathcal{T}_1 \cup \mathcal{T}_2 \models F \sqsubseteq D_1$.*

To prove Lemma C.5, we require a lemma connecting concepts and ABoxes. Let $\mathcal{A}$ be an ABox. For $a \in \mathsf{Ind}(\mathcal{A})$ we define a concept C_a^m by "unfolding" $\mathcal{A}$ at a up to depth m:

$$C_a^0 = (\bigsqcap_{A(a)\in\mathcal{A}} A), \quad C_a^{m+1} = (\bigsqcap_{A(a)\in\mathcal{A}} A) \sqcap (\bigsqcap_{r(a,b)\in\mathcal{A}} \exists r.C_b^m)$$

The following is shown in [Lutz and Wolter, 2010] (Lemma 22).

Lemma C.6. *For all $\mathcal{EL}$ TBoxes $\mathcal{T}$, ABoxes $\mathcal{A}$, and $\mathcal{EL}$ concepts C:*

$$\mathcal{T}, \mathcal{A} \models C(a) \quad \Leftrightarrow \quad \exists m: \quad \mathcal{T} \models C_a^m \sqsubseteq C$$

Proof of Lemma C.5. Let $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq D_1$ with $\mathsf{sig}(\mathcal{T}_1, D_0) \cap \mathsf{sig}(\mathcal{T}_2, D_1) \subseteq \Sigma$. Assume that the required Σ-concept F does not exist. Consider the tree-shaped canonical model $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_0}$. Denote by $\mathcal{A}_\Sigma$ the ABox corresponding to the Σ-reduct of $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_0}$. For the sake of readability, denote the individual names in $\mathcal{A}_\Sigma$ by a_p instread of by p.

Claim. $\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}_\Sigma \not\models D_1(a_{D_0})$.

Proof of claim. To see this, assume that $\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}_\Sigma \models D_1(a_{D_0})$. By Lemma C.6, there is a Σ-concept F such that $\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}_\Sigma \models F(a_{D_0})$ and $\mathcal{T}_1 \cup \mathcal{T}_2 \models F \sqsubseteq D_1$; the former yields $a_{D_0} \in F^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_0}}$ and thus by Lemma C.4 we obtain $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq F$. This is in contradiction to our assumption that no such concept F exists. ⊣

Consider the canonical tree model $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}$ and let $\mathcal{J}$ be the union of the $\mathsf{sig}(\mathcal{T}_1, D_0)$-reduct of $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_0}$ and of $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}$. Note that $\Delta^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_0}} \subseteq \Delta^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}}$ and $\mathcal{J}$ can be constructed by starting with the interpretation $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}$ and then expanding some $X^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}}$, for $X \in \mathsf{sig}(\mathcal{T}_1, D_0) \setminus \Sigma$. $\mathcal{J}$ satisfies $\mathcal{T}_1 \cup \mathcal{T}_2$, but refutes $D_0 \sqsubseteq D_1$. □

D Proofs for Section 4

We now prove Theorem 4.4. We split Part 1 of Theorem 4.4 into two parts, and begin with the case in which condition 3(s1) for non-safeness is satisfied.

Lemma D.1. *Let $(\mathcal{T}, \Sigma)$ be a $\mathcal{EL}$ TBox with closed predicates such that safeness is violated by the inclusion $C \sqsubseteq \exists r.D$ because 3(s1) holds: $r \notin \Sigma$ and $\mathsf{sig}(D) \cap \Sigma \neq \emptyset$. Then convexity fails and there exists a directed tree CQ $q(a)$ such that answering $q(a)$ w.r.t. $(\mathcal{T}, \Sigma)$ is co*NP*-hard.*

Proof. Assume $C \sqsubseteq \exists r.D$ with the properties of Lemma D.1 is given. Consider the canonical model $\mathcal{I}_{\mathcal{T},C}$ of $\mathcal{T}$ and C (see Section C.1). Assume w.l.o.g. that C does not occur in $\mathcal{T}$ (if it does, replace C by $A \sqcap C$ for a fresh concept name A). Note that by our assumptions there is no $a_E \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$ with $(a_E, a_C) \in s^{\mathcal{I}_{\mathcal{T},C}}$ for any role name s. Let

$$S = \{a_E \in \Delta^{\mathcal{I}_{\mathcal{T},C}} \mid (a_C, a_E) \in r^{\mathcal{I}_{\mathcal{T},C}}, \exists r.E \text{ is not a tlc of } C\}$$

Let $\mathcal{I}_S$ be the interpretation obtained from $\mathcal{I}_{\mathcal{T},C}$ by removing all pairs (d, d') with $d' \in S$ from $r^{\mathcal{I}_{\mathcal{T},C}}$. Observe that $a_C \in C^{\mathcal{I}_S}$. Let $\mathcal{A}_S$ be the ABox corresponding to $\mathcal{I}_S$ and let $\mathcal{A}$ be the disjoint union of two copies of $\mathcal{A}_S$. We denote the elements of the first copy by $(d, 1)$ for $d \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$ and the elements of the second copy by $(d, 2)$, for $d \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$. Let A_1 and A_2 be fresh concept names and

$$\mathcal{A}' = \mathcal{A} \cup \{A_1(d,1) \mid d \in \Delta^{\mathcal{I}_{\mathcal{T},C}}\} \cup \{A_2(d,2) \mid d \in \Delta^{\mathcal{I}_{\mathcal{T},C}}\}$$

If some concept name $E \in \Sigma$ occurs in D, then fix one such E and denote by D_i the resulting concept after one occurrence of E is replaced by $A_i \sqcap E$. Similarly, if no concept name from Σ occurs in D, then let $s \in \Sigma$ be such that a concept of the form $\exists s.G$ occurs in D. Denote by D_i the resulting concept after one occurrence of $\exists s.G$ is replaced by $A_i \sqcap \exists s.G$.

Claim 1.
(1) $(\mathcal{T},\mathcal{A}') \models_{c(\Sigma)} \exists r.D_1(a_C,1) \vee \exists r.D_2(a_C,1)$
(2) $(\mathcal{T},\mathcal{A}') \not\models_{c(\Sigma)} \exists r.D_i(a_C,1)$ for $i=1,2$.

(1) is straightforward using the condition that $\mathcal{T} \models C \sqsubseteq \exists r.D$. (2) We construct a model $\mathcal{J}$ of $(\mathcal{T},\Sigma)$ and $\mathcal{A}'$ with $(a_C,1) \notin (\exists r.D_1)^{\mathcal{J}}$. $\mathcal{J}$ is defined as the interpretation corresponding to the ABox $\mathcal{A}'$ extended by

$$\{r((a_C,1),(e,2)) \mid e \in S\} \cup \{r((a_C,2)),(e,1)) \mid e \in S\}$$

Since $\mathcal{I}_{\mathcal{T},C}$ is a model of $\mathcal{T}$ it is readily checked that $\mathcal{J}$ is a model of $(\mathcal{T},\Sigma)$ and $\mathcal{A}'$. Moreover, $(a_C,1) \notin (\exists r.D_1)^{\mathcal{J}}$. To prove this assume $(a_C,1) \in (\exists r.D_1)^{\mathcal{J}}$. Then one of the following two conditions holds:

- there exists a tlc $\exists r.C'$ of C such that $(a_{C'},1) \in D_1^{\mathcal{J}}$;
- there exists $a_{C'}$ with $(a_C,a_{C'}) \in r^{\mathcal{I}_{\mathcal{T},C}}$ such that $(a_{C'},2) \in D_1^{\mathcal{J}}$.

The first condition leads to a contradiction since it implies, by Lemma C.1, that $\mathcal{T} \models C' \sqsubseteq D$ for a tlc $\exists r.C'$ of C. Hence $C \sqsubseteq \exists r.D$ does not violate safeness of $(\mathcal{T},\Sigma)$. The second condition cannot hold since no point $(a_{C'},2)$ can reach along a role-path in $\mathcal{J}$ any point in the first copy of $\mathcal{A}_S$ and A_1 applies only to points in the first copy (here we need that a_C is not reachable).

The construction of a model $\mathcal{J}$ of $(\mathcal{T},\Sigma)$ and $\mathcal{A}'$ with $(a_C,1) \notin (\exists r.D_2)^{\mathcal{J}}$ is similar and left to the reader.

The coNP-hardness proof is now exactly the same as in Example 3.1. □

Lemma D.2. *Let $(\mathcal{T},\Sigma)$ be an $\mathcal{EL}$ TBox with closed predicates such that safeness is violated by $C \sqsubseteq \exists r.D$ because 3(s2) holds. Then convexity fails and there exists a directed tree CQ $q(a)$ such that answering $q(a)$ w.r.t. $(\mathcal{T},\Sigma)$ is co*NP-*hard.*

Proof. Consider the interpretation $\mathcal{I}_S$ from the the proof of Lemma D.1 and let $\mathcal{A}_S$ be the corresponding ABox. Consider

$$K = \{G \mid \exists r.G \in \mathsf{sub}(\mathcal{T}), \mathcal{T} \models C \sqsubseteq \exists r.G\}$$

Since there is no tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$, by a result of [Lutz and Wolter, 2010] (Lemma 16), there exists $G \in K$ with $\mathcal{T} \models G \sqsubseteq D$.

Claim 1. For all $G \in K$: $\mathcal{T}^0 \cup \mathcal{T}^1 \not\models G^0 \sqsubseteq D^1$.

Assume Claim 1 does not hold. Let $G \in K$ with $\mathcal{T}^0 \cup \mathcal{T}^1 \models G^0 \sqsubseteq D^1$. By Lemma C.5, there exists a Σ-concept F such that $\mathcal{T}^0 \cup \mathcal{T}^1 \models G^0 \sqsubseteq F$ and $\mathcal{T}^0 \cup \mathcal{T}^1 \models F \sqsubseteq D^1$. Then $\mathcal{T} \models G \sqsubseteq F$ and $\mathcal{T} \models F \sqsubseteq D$. We have $\mathcal{T} \models C \sqsubseteq \exists r.G$. Hence $\mathcal{T} \models C \sqsubseteq \exists r.F$ and we have derived a contradiction to Condition 3(s2).

By Claim 1 we can take the canonical models $\mathcal{J}_G := \mathcal{I}_{\mathcal{T}^0\cup\mathcal{T}^1,G^0}$ for any $G \in K$ and obtain for $a_G := a_{G^0}$ that $a_G \notin (D^1)^{\mathcal{J}_G}$. Let $\mathcal{A}_{G,\Sigma}$ be the Σ-reduct of the ABox corresponding to $\mathcal{J}_G$. We assume that the $\mathsf{Ind}(\mathcal{A}_{G,\Sigma})$ are mutually disjoint, for $G \in K$, and that $a_G \in \mathsf{Ind}(\mathcal{A}_{G,\Sigma})$, for all $G \in K$.

Claim 2. For every $G \in K$, there exist
(1) a model $\mathcal{I}_G^1$ of $(\mathcal{T},\Sigma)$ and $\mathcal{A}_{G,\Sigma}$ whose domain coincides with $\mathsf{Ind}(\mathcal{A}_{G,\Sigma})$ and for which $a_G \in G^{\mathcal{I}_G^1}$ and $a_G \in H^{\mathcal{I}_G^1}$ implies $\mathcal{T} \models G \sqsubseteq H$, for all $\mathcal{EL}$ concepts H with $\mathsf{sig}(H) \subseteq \mathsf{sig}(\mathcal{T},C,D)$;
(2) a model $\mathcal{I}_G^2$ of $(\mathcal{T},\Sigma)$ and $\mathcal{A}_{G,\Sigma}$ whose domain coincides with $\mathsf{Ind}(\mathcal{A}_{G,\Sigma})$ such that $a_G \notin D^{\mathcal{I}_G^2}$ and $a_G \in H^{\mathcal{I}_G^2}$ implies $\mathcal{T} \models G \sqsubseteq H$, for all $\mathcal{EL}$ concepts H with $\mathsf{sig}(H) \subseteq \mathsf{sig}(\mathcal{T},C,D)$.

The interpretation $\mathcal{I}_G^1$ is obtained from $\mathcal{J}_G$ by interpreting all non-Σ-symbols $X \in \mathsf{sig}(\mathcal{T},C,D)$ as $X^{\mathcal{I}_G^1} := (X^0)^{\mathcal{J}_G}$. The interpretation $\mathcal{I}_G^2$ is obtained from $\mathcal{J}_G$ by interpreting all non-Σ-symbols $X \in \mathsf{sig}(\mathcal{T},C,D)$ as $X^{\mathcal{I}_G^2} := (X^1)^{\mathcal{J}_G}$.

Introduce two copies $\mathcal{A}_{G,\Sigma}^1$ and $\mathcal{A}_{G,\Sigma}^2$ of $\mathcal{A}_{G,\Sigma}$, for $G \in K$. We denote the elements of the first copy by $(a,1)$, for $a \in \mathsf{Ind}(\mathcal{A}_{G,\Sigma})$ and the elements of the second copy by $(a,2)$, for $a \in \mathsf{Ind}(\mathcal{A}_{G,\Sigma})$. Now define the ABox $\mathcal{A}$ by taking two fresh concept names A_1 and A_2 and the union

$$\mathcal{A}_S \cup \bigcup_{G\in K} \mathcal{A}_{G,\Sigma}^1 \cup \mathcal{A}_{G,\Sigma}^2$$

and the additional assertions

- $r(a_C,(a_G,1)), r(a_C,(a_G,2))$, for every $G \in K$;
- $A_1(a_G,1)$, for every $G \in K$;
- $A_1(a_{D'})$, for every tlc $\exists r.D'$ of C;
- $A_2(a_G,2)$, for every $G \in K$.

Claim 3.
(1) $\mathcal{T},\mathcal{A} \models_{c(\Sigma)} \exists r.(A_1 \sqcap D)(a_C) \vee \exists r.(A_2 \sqcap D)(a_C)$.
(2) $\mathcal{T},\mathcal{A} \not\models_{c(\Sigma)} \exists r.(A_i \sqcap D)(a_C)$, for $i=1,2$.

(1) is straightforward since $\mathcal{T} \models C \sqsubseteq \exists r.D$.

(2) We first show $\mathcal{T},\mathcal{A} \not\models_{c(\Sigma)} \exists r.(A_2 \sqcap D)(a_C)$. The interpretation $\mathcal{J}$ showing this is obtained by expanding all $\mathcal{A}_{G,\Sigma}^2$, $G \in K$, to $\mathcal{I}_G^2$ and all $\mathcal{A}_{G,\Sigma}^1$, $G \in K$, to $\mathcal{I}_G^1$. The ABox $\mathcal{A}_S$ is transformed into the interpretation $\mathcal{I}_S$. Using the properties of $\mathcal{I}_G^1$ and $\mathcal{I}_G^2$ from Claim 2, it is readily checked that $\mathcal{J}$ is a model of $(\mathcal{T},\Sigma)$ and $\mathcal{A}$. Moreover, $a_C \notin (\exists r.(A_2 \sqcap D))^{\mathcal{J}}$ since $(a_G,2) \notin D^{\mathcal{J}}$ for any $G \in K$ (by the properties of $\mathcal{I}_G^2$ from Claim 2).

We now show $\mathcal{T},\mathcal{A} \not\models_{\Sigma} \exists r.(A_1 \sqcap D)(a_C)$. The interpretation $\mathcal{J}$ showing this is obtained by expanding all $\mathcal{A}_{G,\Sigma}^2$, $G \in K$, to $\mathcal{I}_G^1$ and all $\mathcal{A}_{G,\Sigma}^1$, $G \in K$, to $\mathcal{I}_G^2$. The ABox $\mathcal{A}_S$ is again transformed into $\mathcal{I}_S$. Using the properties of $\mathcal{I}_G^1$ and $\mathcal{I}_G^2$ from Claim 2, it is readily checked that $\mathcal{J}$ is a model of $(\mathcal{T},\Sigma)$ and $\mathcal{A}$. Moreover, $a_C \notin (\exists r.(A_1 \sqcap D))^{\mathcal{J}}$ since $a_{C'} \notin D^{\mathcal{J}}$ for any tlc $\exists r.C'$ of C and since $(a_G,1) \notin D^{\mathcal{J}}$ for any $G \in K$.

The coNP-hardness proof is exactly the same as in Example 3.1. □

We come to the proof of Part 2 of Theorem 4.4. We first show (a):

Lemma D.3. *Let $(\mathcal{T}, \Sigma)$ be safe. Then CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ coincides with CQ answering w.r.t. $\mathcal{T}$ without closed predicates for ABoxes that are satisfiable w.r.t. $(\mathcal{T}, \Sigma)$.*

Proof. Let $(\mathcal{T}, \Sigma)$ be safe. Consider an ABox $\mathcal{A}$ that is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$.

We show that $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$ (from which the lemma follows by Lemma C.3).

To show this, it is sufficient to observe

- if $a \in A^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$ for some $a \in \mathsf{Ind}(\mathcal{A})$ and $A \in \Sigma$, then $A(a) \in \mathcal{A}$.
- if $a \in (\exists r.\top)^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$ for some $a \in \mathsf{Ind}(\mathcal{A})$ and $r \in \Sigma$, then there exists $b \in \mathsf{Ind}(\mathcal{A})$ with $r(a, b) \in \mathcal{A}$.
- if $p \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A})$, then there is no Σ-concept $F \neq \top$ such that $p \in F^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$.

Point 1 follows from Lemma C.3 since $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$. For Point 2, assume this is not the case. Then $\mathcal{T}, \mathcal{A} \models \exists r.C(a)$ for some C such that there does not exist $b \in \mathcal{A}$ with $r(a, b) \in \mathcal{A}$ and $\mathcal{T}, \mathcal{A} \models C(b)$. But then, by Lemma C.6, there exists m such that $\mathcal{T} \models C_a^m \sqsubseteq \exists r.C$ and there is no tlc $\exists r.C_b^{m-1}$ of C_a^m with $\mathcal{T} \models C_b^{m-1} \sqsubseteq C$. If $\mathsf{sig}(\exists r.C) \subseteq \Sigma$ we have a contradiction to the condition that $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$. Otherwise, $\mathsf{sig}(C) \not\subseteq \Sigma$ and we have a contradiction to the assumption that $(\mathcal{T}, \Sigma)$ is safe.

To show Point 3, assume such p and F exist. Then $p = ar_1C_1 \cdots r_kC_k$ for some $a \in \mathsf{Ind}(\mathcal{A})$. We assume that no example shorter than p exists. Then $r_1 \notin \Sigma$. By Lemma C.3, $\mathcal{T}, \mathcal{A} \models \exists r_1.(C_1 \sqcap \cdots \exists r_{.k}.(C_k \sqcap F))$. By construction of $\mathcal{J}_{\mathcal{T},\mathcal{A}}$, there is no b with $r_1(a, b) \in \mathcal{A}$ such that $\mathcal{T}, \mathcal{A} \models C_1(b)$. From

$$\mathcal{T}, \mathcal{A} \models \exists r_1.(C_1 \sqcap \cdots \exists r_{.k}.(C_k \sqcap F))(a)$$

we obtain that there exists m with $\mathcal{T} \models C_a^m \sqsubseteq \exists r_1.(C_1 \sqcap \cdots \exists r_{.k}.(C_k \sqcap F))$. Moreover, there exists no tlc C' of C_a^m with $\mathcal{T} \models C' \sqsubseteq (C_1 \sqcap \cdots \exists r_{.k}.(C_k \sqcap F))$. We thus have derived a contradiction to $(\mathcal{T}, \Sigma)$ being safe. ☐

To show Condition (b) for Theorem 4.4 it now suffices to show:

Lemma D.4. *Let $(\mathcal{T}, \Sigma)$ be safe. Then it can be decided in polytime (data complexity) whether an ABox $\mathcal{A}$ is satisfiable w.r.t a safe $(\mathcal{T}, \Sigma)$.*

Proof. We first the show the following

Claim 1. If $(\mathcal{T}, \Sigma)$ is safe, then there exists an $\mathcal{EL}$ TBox $\mathcal{T}'$ that is equivalent to $\mathcal{T}$ such that for any $C \sqsubseteq D \in \mathcal{T}'$, $\mathsf{sig}(D) \subseteq \Sigma$ or $\mathsf{sig}(D) \cap \Sigma = \emptyset$.

To prove Claim 1 we modify the TBox $\mathcal{T}$ as follows: first, replace any $C \sqsubseteq D$ with D a proper conjunction of concepts by the set of $C \sqsubseteq D'$ with D' a tlc of D. Second, replace recursively,

- any $C \sqsubseteq \exists r.D$ such that $\mathsf{sig}(\exists r.D) \not\subseteq \Sigma$ for which exists a tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$ by the inclusions $C' \sqsubseteq D'$ with D' a tlc of D;
- any $C \sqsubseteq \exists r.D$ with $r \in \Sigma$ and $\mathsf{sig}(D) \not\subseteq \Sigma$ by $C \sqsubseteq \exists r.F$ and $F \sqsubseteq D'$ for every tlc D' of D, where F is a Σ-concept with $\mathcal{T} \models C \sqsubseteq \exists r.F$ and $\mathcal{T} \models F \sqsubseteq D$. Such a Σ-concept F exists by Condition 3(s2).

The resulting TBox $\mathcal{T}'$ is as required and Claim 1 is proved.

Now Lemma D.4 follows from the observation that $\mathcal{A}$ is satisfiable w.r.t. a safe $(\mathcal{T}, \Sigma)$ iff, for $\mathcal{T}'$ of the form above, whenever $\mathcal{T}, \mathcal{A} \models F(a)$ for some $C \sqsubseteq F \in \mathcal{T}'$ with $\mathsf{sig}(F) \subseteq \Sigma$, then $\mathcal{A} \models F(a)$. This condition can be checked in polytime (data complexity). ☐

We now turn to the proof of Theorem 4.7.

Lemma 4.6. *If $\mathcal{T}$ satisfies Conditions (t1) and (t2), then $(\mathcal{T}, \Sigma)$ is safe iff it is strongly safe.*

Proof. Suppose that $\mathcal{T}$ satisfies Conditions (t1) and (t2).
($\Rightarrow$) Suppose that $(\mathcal{T}, \Sigma)$ is not strongly safe, that is, there is some $C \sqsubseteq \exists r.D \in \mathcal{T}$ satisfying (c1) or (c2). Then $\mathsf{sig}(\exists r.D) \not\subseteq \Sigma$ and Condition (t2) yields

$$\text{there is no tlc } \exists r.C' \text{ of } C \text{ with } \mathcal{T} \models C' \sqsubseteq D. \qquad (*)$$

If $C \sqsubseteq \exists r.D$ satisfies Condition (c1), then $r \notin \Sigma$ and there is some concept E such that $\mathcal{T} \models D \sqsubseteq E$ and $\mathsf{sig}(E) \cap \Sigma \neq \emptyset$. Now, $C \sqsubseteq \exists r.(D \sqcap E)$ makes $\mathcal{T}$ unsafe via Condition 3(s1):

1. $\mathcal{T} \models C \sqsubseteq \exists r.(D \sqcap E)$ since $C \sqsubseteq \exists r.D \in \mathcal{T}$ and $\mathcal{T} \models D \sqsubseteq E$.
2. there is no tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D \sqcap E$; this follows from $(*)$.
3. $r \notin \Sigma$ and $\mathsf{sig}(D \sqcap E) \cap \Sigma \neq \emptyset$ since $\mathsf{sig}(E) \cap \Sigma \neq \emptyset$.

If $C \sqsubseteq \exists r.D$ satisfies Condition (c2), then it is easily shown that $C \sqsubseteq \exists r.D$ makes $\mathcal{T}$ unsafe via Condition 3(s2).

($\Leftarrow$) Suppose that $\mathcal{T}$ is not safe. Then there is some $\mathcal{EL}$ inclusion $C \sqsubseteq \exists r.D$ such that $\mathcal{T} \models C \sqsubseteq \exists r.D$,

$$\text{there is no tlc } \exists r.C' \text{ of } C \text{ with } \mathcal{T} \models C' \sqsubseteq D, \qquad (**)$$

and one of Conditions 3(s1) and 3(s2) is satisfied. In the following, we use the tree-shaped canonical model $\mathcal{J}_{\mathcal{T},C}$ defined in Section C.1. Note that $C \in \Delta^{\mathcal{J}_{\mathcal{T},C}}$ and Lemma C.4 yields $C \in (\exists r.D)^{\mathcal{J}_{\mathcal{T},C}}$. Thus there is some $d \in \Delta^{\mathcal{J}_{\mathcal{T},C}}$ such that $(C, d) \in r^{\mathcal{J}_{\mathcal{T},C}}$ and $d \in D^{\mathcal{J}_{\mathcal{T},C}}$. By definition of $\mathcal{J}_{\mathcal{T},C}$, $d = C \cdot r \cdot E$ for some $\mathcal{EL}$ concept $E \in \mathsf{sub}(C) \cup \mathsf{sub}(\mathcal{T})$. By Lemma C.4 and $d \in D^{\mathcal{J}_{\mathcal{T},C}}$, we have $\mathcal{T} \models E \sqsubseteq D$.

Let $\mathcal{A}_C = \mathcal{A}_0, \mathcal{A}_1, \ldots$ be the ABoxes used in the construction of $\mathcal{J}_{\mathcal{T},C}$. By definition of $\mathcal{A}_C$ and by $(**)$, we have $C \cdot r \cdot E \notin \mathsf{paths}(C)$, that is, $d = C \cdot r \cdot E$ must have been generated by $\mathsf{R3}$. Consequently, there is an $i \in \mathbb{N}$ such that $\exists r.E(C) \in \mathcal{A}_i$, and $\mathcal{A}_{i+1} = \mathcal{A}_i \cup \{r(C, C \cdot r \cdot E), E(C \cdot r \cdot E)\}$. We aim at showing that the assertion $\exists r.E(C)$ was generated by an application of $\mathsf{R2}$. This is essentially a consequence of the following.

Claim. *For all $i \geq 0$: $\mathcal{A}_i$ does not contain assertions of the form $C_1 \sqcap C_2(C)$.*

Proof of claim. The proof is by induction on i. The base case is trivial by definition of $\mathcal{A}_C$. For the inductive step, we make a case distinction according to the rule applied:

- R1: By the inductive hypothesis, there is no concept assertion of the form $C_1 \sqcap C_2(C) \in \mathcal{A}_i$. This means that R1 was applied to some $C_1 \sqcap C_2(p)$, where $p \neq C$. Hence the inductive hypothesis holds for $\mathcal{A}_{i+1}$.
- R2: In this case, $\mathcal{A}_{i+1} = \mathcal{A}_i \cup \{D_2(p)\}$ for some $D_1 \sqsubseteq D_2 \in \mathcal{T}$. By Condition (t1), we know that D_2 is not of the form $C_1 \sqcap C_2$. Hence the inductive hypothesis holds for $\mathcal{A}_{i+1}$.
- R3: Since C is the root of $\mathcal{A}_i$, this rule never adds any concept assertions of the form $C'(C)$. Hence the inductive hypothesis holds for $\mathcal{A}_{i+1}$.

$\dashv$

Since $\mathcal{A}_0$ contains no concept assertions of the form $\exists r.C'(p)$, there is some $j \in \{0, \ldots, i-1\}$ such that $\exists r.E(C) \in \mathcal{A}_{j+1} \setminus \mathcal{A}_j$. By the claim, this addition is due to R2. Thus there is some $C' \sqsubseteq \exists r.E \in \mathcal{T}$ with $\mathcal{A}_j \models C'(C)$. From the latter, we obtain $C \in (C')^{\mathcal{J}_{\mathcal{T},C}}$ and this implies by Lemma C.4 that $\mathcal{T} \models C \sqsubseteq C'$. Since $C \sqsubseteq \exists r.D$ makes $\mathcal{T}$ unsafe, one of the following cases applies:

- $C \sqsubseteq \exists r.D$ satisfies Condition 3(s1).
  Then $r \notin \Sigma$ and $\mathsf{sig}(D) \cap \Sigma \neq \emptyset$. Since $\mathcal{T} \models E \sqsubseteq D$, we thus have that $C' \sqsubseteq \exists r.E \in \mathcal{T}$ satisfies Condition (c1), therefore $\mathcal{T}$ is not strongly safe.
- $C \sqsubseteq \exists r.D$ satisfies Condition 3(s2).
  Then $r \in \Sigma$, $\mathsf{sig}(D) \not\subseteq \Sigma$, and
  (†) there is no Σ-concept F with $\mathcal{T} \models C \sqsubseteq \exists r.F$ and $\mathcal{T} \models F \sqsubseteq D$.

  We aim at showing that $C' \sqsubseteq \exists r.E \in \mathcal{T}$ satisfies Condition (c2). We already know that $r \in \Sigma$. Since $\mathcal{T} \models C \sqsubseteq C' \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq D$, (†) yields $\mathsf{sig}(E) \not\subseteq \Sigma$. It thus remains to show that there is no Σ-concept F with $\mathcal{T} \models C' \sqsubseteq \exists r.F$ and $\mathcal{T} \models F \sqsubseteq E$. This, however, is a consequence of (†) and the facts that $\mathcal{T} \models C \sqsubseteq C'$ and $\mathcal{T} \models E \sqsubseteq D$.

$\square$

We now show how to check Conditions (c1) and (c2) for strong safeness in PTIME, when the TBox $\mathcal{T}$ satisfies (t1) and (t2). We start with the former. Consider the canonical model $\mathcal{I}_{\mathcal{T},\top}$. We define the notion of a *marked node* in $\mathcal{I}_{\mathcal{T},\top}$ inductively as follows:

- Every $a_C \in \Delta^{\mathcal{I}_{\mathcal{T},\top}}$ with
  - $a_C \in A^{\mathcal{I}_{\mathcal{T},\top}}$ for some $A \in \Sigma$; or
  - $(a_C, a_D) \in r^{\mathcal{I}_{\mathcal{T},\top}}$ for some $a_D \in \Delta^{\mathcal{I}_{\mathcal{T},\top}}$ and $r \in \Sigma$

  is a marked node.
- If $(a_C, a_D) \in r^{\mathcal{I}_{\mathcal{T},\top}}$ and a_D is a marked node then a_C is also a marked node.

Since the size of $\mathcal{I}_{\mathcal{T},\top}$ is polynomial in the size of $\mathcal{T}$, the following lemma shows that Condition (c1) can be checked in PTIME.

Lemma D.5. *Let $C \sqsubseteq \exists r.D \in \mathcal{T}$. Then $C \sqsubseteq \exists r.D$ satisfies (c1) if and only if $r \notin \Sigma$ and a_D is a marked node in $\mathcal{I}_{\mathcal{T},\top}$.*

Proof. ($\Rightarrow$) Suppose $C \sqsubseteq \exists r.D \in \mathcal{T}$ satisfies (c1), i.e., $r \notin \Sigma$ and there is some concept E such that $\mathcal{T} \models D \sqsubseteq E$ and $\mathsf{sig}(E) \cap \Sigma \neq \emptyset$. We need to show that a_D is a marked node. By the definition of $\mathcal{I}_{\mathcal{T},\top}$, we have $a_D \in \Delta^{\mathcal{I}_{\mathcal{T},\top}}$. Then by $\mathcal{T} \models D \sqsubseteq E$ and Lemma C.1, $a_D \in E^{\mathcal{I}_{\mathcal{T},\top}}$. It thus suffices to show the following.

Claim. *For all $d \in \Delta^{\mathcal{I}_{\mathcal{T},\top}}$ and all $\mathcal{EL}$ concepts C with $\mathsf{sig}(C) \cap \Sigma \neq \emptyset$, if $d \in C^{\mathcal{I}_{\mathcal{T},\top}}$, then d is a marked node.*

Proof of claim. Suppose $d \in C^{\mathcal{I}_{\mathcal{T},\top}}$. The proof is by induction on the structure of C.

- $C = A$ for some $A \in \mathsf{N_C}$. Then $A \in \Sigma$ and thus d is a marked node.
- C is a conjunction of concepts. Since $\mathsf{sig}(C) \cap \Sigma \neq \emptyset$, there is some tlc C' of C with $\mathsf{sig}(C') \cap \Sigma \neq \emptyset$. Then by the inductive hypothesis, d is a marked node.
- $C = \exists s.C'$. Since $d \in C^{\mathcal{I}_{\mathcal{T},\top}}$, there is some $d' \in \Delta^{\mathcal{I}_{\mathcal{T},\top}}$ such that $(d, d') \in s^{\mathcal{I}_{\mathcal{T},\top}}$ and $d' \in (C')^{\mathcal{I}_{\mathcal{T},\top}}$. Since $\mathsf{sig}(C) \cap \Sigma \neq \emptyset$, we have $s \in \Sigma$ or $\mathsf{sig}(C') \cap \Sigma \neq \emptyset$. If $s \in \Sigma$, then by the base case in the definition of a marked node, d is a marked node. If $\mathsf{sig}(C') \in \Sigma$, then by the inductive hypothesis, d' is a marked node and thus d is also a marked node.

$\dashv$

($\Leftarrow$) Let $C \sqsubseteq \exists r.D \in \mathcal{T}$ with $r \notin \Sigma$ and a_D a marked node. By the inductive definition of a marked node, we can construct a concept

$$E = \exists r_1. \cdots \exists r_n.C'$$

with $n \geq 0$ such that $(a_D, d_1) \in r_1^{\mathcal{I}_{\mathcal{T},\top}}, \ldots, (d_{n-1}, d_n) \in r_n^{\mathcal{I}_{\mathcal{T},\top}}$, d_i for $i \in \{1, \ldots, n\}$ is a marked node, $d_n \in (C')^{\mathcal{I}_{\mathcal{T},\top}}$ and

- $C' = A$ for some $A \in \Sigma$; or
- $C' = \exists s.D'$, where $(d_n, d_{D'}) \in s^{\mathcal{I}_{\mathcal{T},\top}}$ and $s \in \Sigma$.

By construction, $a_D \in E^{\mathcal{I}_{\mathcal{T},\top}}$ and $\mathsf{sig}(E) \cap \Sigma \neq \emptyset$. From the former and Lemma C.1, we have $\mathcal{T} \models D \sqsubseteq E$. Thus $C \sqsubseteq \exists r.D \in \mathcal{T}$ satisfies (c1). $\square$

It is an immediate consequence of the following lemma that Condition (c2) can also be checked in PTIME.

Lemma D.6. *Let $C \sqsubseteq \exists r.D \in \mathcal{T}$. Then the following are equivalent:*

1. *$C \sqsubseteq \exists r.D$ satisfies (c2)*
2. *$r \in \Sigma$, $\mathsf{sig}(D) \not\subseteq \Sigma$, and there is no $\exists r.G \in \mathsf{sub}(\mathcal{T})$ such that*

$$\mathcal{T} \models C \sqsubseteq \exists r.G \text{ and } (\mathcal{T})^0 \cup (\mathcal{T})^1 \models G^0 \sqsubseteq D^1.$$

Proof. (2 $\Rightarrow$ 1). Suppose $C \sqsubseteq \exists r.D \in \mathcal{T}$ does not satisfy (c2). Then one of the following holds:

1. $r \notin \Sigma$;
2. $\mathsf{sig}(D) \subseteq \Sigma$; or

3. $r \in \Sigma$, $\mathsf{sig}(D) \not\subseteq \Sigma$, and there exists a Σ-concept E with $\mathcal{T} \models C \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq D$.

If item 1 or item 2 holds then we are done. Otherwise, item 3 holds.

By $\mathcal{T} \models C \sqsubseteq \exists r.E$, $C \in \mathsf{sub}(\mathcal{T})$, and Lemma C.2, there is some $\exists r.G \in \mathsf{sub}(\mathcal{T})$ such that $\mathcal{T} \models C \sqsubseteq \exists r.G$ and $\mathcal{T} \models G \sqsubseteq E$. It remains to show $(\mathcal{T})^0 \cup (\mathcal{T})^1 \models G^0 \sqsubseteq D^1$.

Claim. $(\mathcal{T})^0 \models G^0 \sqsubseteq E$.

Proof of claim. Let $\mathcal{I}$ be a model of $(\mathcal{T})^0$. We show that $\mathcal{I} \models G^0 \sqsubseteq E$, which implies the claim. Define the interpretation $\mathcal{J}$ as follows:

- $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$;
- for all $P \in \mathsf{N_C} \cup \mathsf{N_R}$,

$$P^{\mathcal{J}} = \begin{cases} P^{\mathcal{I}} & \text{if } P \in \Sigma \\ (P^0)^{\mathcal{I}} & \text{if } P \in \mathsf{sig}(G, \mathcal{T}) \setminus \Sigma \\ \emptyset & \text{otherwise} \end{cases}$$

Now for all $d \in \Delta^{\mathcal{I}} = \Delta^{\mathcal{J}}$ and $\mathcal{EL}$ concepts C with $\mathsf{sig}(C) \subseteq \mathsf{sig}(G, \mathcal{T}) \cup \Sigma$, we have

$$d \in (C^0)^{\mathcal{I}} \text{ iff } d \in C^{\mathcal{J}}. \tag{*}$$

This implies that $\mathcal{J}$ is a model of $\mathcal{T}$. Then by $\mathcal{T} \models G \sqsubseteq E$, we obtain $\mathcal{J} \models G \sqsubseteq E$. Then using (*) again, we obtain that $\mathcal{I} \models G^0 \sqsubseteq E$, which is what we wanted to show. $\dashv$

Claim. $(\mathcal{T})^1 \models E \sqsubseteq D^1$.

Proof of claim. Let $\mathcal{I}$ be a model of $(\mathcal{T})^1$. We show that $\mathcal{I} \models E \sqsubseteq D^1$, which implies the claim. Define the interpretation $\mathcal{J}$ as follows:

- $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$;
- for all $P \in \mathsf{N_C} \cup \mathsf{N_R}$,

$$P^{\mathcal{J}} = \begin{cases} P^{\mathcal{I}} & \text{if } P \in \Sigma \\ (P^1)^{\mathcal{I}} & \text{if } P \in \mathsf{sig}(D, \mathcal{T}) \setminus \Sigma \\ \emptyset & \text{otherwise} \end{cases}$$

Now for all $d \in \Delta^{\mathcal{I}} = \Delta^{\mathcal{J}}$ and $\mathcal{EL}$ concepts C with $\mathsf{sig}(C) \subseteq \mathsf{sig}(D, \mathcal{T}) \cup \Sigma$, we have

$$d \in (C^1)^{\mathcal{I}} \text{ iff } d \in C^{\mathcal{J}}. \tag{†}$$

This implies that $\mathcal{J}$ is a model of $\mathcal{T}$. Then by $\mathcal{T} \models E \sqsubseteq D$, we obtain $\mathcal{J} \models E \sqsubseteq D$. Then using (†) again, we obtain that $\mathcal{I} \models E \sqsubseteq D^1$, which is what we wanted to show. $\dashv$

Now by the last two claims, we have $(\mathcal{T})^0 \models G^0 \sqsubseteq E$ and $(\mathcal{T})^1 \models E \sqsubseteq D^1$. These immediately yield $(\mathcal{T})^0 \cup (\mathcal{T})^1 \models G^0 \sqsubseteq D^1$, which is what we wanted to show.

($1 \Rightarrow 2$) Suppose that item 2 in the lemma does not hold. Then one of the following is true:

1. $r \notin \Sigma$;
2. $\mathsf{sig}(D) \subseteq \Sigma$; or
3. $r \in \Sigma$, $\mathsf{sig}(D) \not\subseteq \Sigma$, and there is some $\exists r.G \in \mathsf{sub}(\mathcal{T})$ such that $\mathcal{T} \models C \sqsubseteq \exists r.G$ and $(\mathcal{T})^0 \cup (\mathcal{T})^1 \models G^0 \sqsubseteq D^1$.

Each of the first two items above immediately implies that $C \sqsubseteq \exists r.D$ does not satisfy (c2). Hence it remains to consider item 3. Suppose that item 3 holds. It suffices to show that there is some Σ-concept E with $\mathcal{T} \models C \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq D$.

By $(\mathcal{T})^0 \cup (\mathcal{T})^1 \models G^0 \sqsubseteq D^1$ and $\mathsf{sig}(G^0, (\mathcal{T})^0) \cap \mathsf{sig}(D^1, (\mathcal{T})^1) \subseteq \Sigma$ we obtain by Lemma C.5 that there is some Σ-concept E such that $(\mathcal{T})^0 \cup (\mathcal{T})^1 \models G^0 \sqsubseteq E$ and $(\mathcal{T})^0 \cup (\mathcal{T})^1 \models E \sqsubseteq D^1$.

Claim. $\mathcal{T} \models C \sqsubseteq \exists r.E$.

Proof of claim. Since $\mathcal{T} \models C \sqsubseteq \exists r.G$, it is enough to show that $\mathcal{T} \models G \sqsubseteq E$. Suppose for contradiction that $\mathcal{T} \not\models G \sqsubseteq E$. Then there is some model $\mathcal{I}$ of $\mathcal{T}$ such that $\mathcal{I} \not\models G \sqsubseteq E$, i.e., there is some $d_0 \in \Delta^{\mathcal{I}}$ such that $d_0 \in G^{\mathcal{I}}$ and $d_0 \notin E^{\mathcal{I}}$. Define the interpretation $\mathcal{J}$ as follows:

- $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$
- for all $P \in \mathsf{N_C} \cup \mathsf{N_R}$,

$$P^{\mathcal{J}} = \begin{cases} X^{\mathcal{I}} & \text{if } P = X^0 \text{ for some } X \in \mathsf{sig}(\mathcal{T}) \setminus \Sigma \\ X^{\mathcal{I}} & \text{if } P = X^1 \text{ for some } X \in \mathsf{sig}(\mathcal{T}) \setminus \Sigma \\ P^{\mathcal{I}} & \text{otherwise} \end{cases}$$

Now for all $d \in \Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$ and all $\mathcal{EL}$ concepts $C \in \mathsf{sub}(\mathcal{T})$, we have that $d \in C^{\mathcal{I}}$ iff $d \in (C^0)^{\mathcal{J}}$ iff $d \in (C^1)^{\mathcal{J}}$. This implies $\mathcal{J} \models (\mathcal{T})^0 \cup (\mathcal{T})^1$; and by $G \in \mathsf{sub}(\mathcal{T})$, it implies $d_0 \in (G^0)^{\mathcal{J}}$. Since the interpretation of the symbols in Σ did not change, we still have $d_0 \notin E^{\mathcal{J}}$. But then $\mathcal{T}^0 \cup \mathcal{T}^1 \not\models G^0 \sqsubseteq E$, which is a contradiction. $\dashv$

The missing fact that $\mathcal{T} \models E \sqsubseteq D$ can be shown analogously to the previous claim. $\square$

It remains to prove that checking safeness of $\mathcal{EL}$-TBoxes is PTIME-hard. This is done by reduction from subsumption in $\mathcal{EL}$, which is a known PTIME-hard problem. Thus, let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and A, B concept names. We may assume that $\mathcal{T}$ satisfies conditions (t1) and (t2). Let

$$\mathcal{T}' = \mathcal{T} \cup \{A' \sqsubseteq A, E \sqsubseteq \exists r.B, E \sqsubseteq \exists r.A'\}$$

where E, A', and r are fresh predicates, and set $\Sigma = \{A', r\}$.

Lemma D.7. *$\mathcal{T} \models A \sqsubseteq B$ iff $(\mathcal{T}', \Sigma)$ is safe.*

Proof. "if". Assume that $\mathcal{T}'$ is not safe. Let $C \sqsubseteq \exists s.D \in \mathcal{T}'$ be a witness satisfying Condition (c1) or (c2). Assume first that Condition (c1) holds. Thus, $s \notin \Sigma$ and $\mathsf{sig}(D) \cap \Sigma \neq \emptyset$. We have $r \neq s$ and so $C \sqsubseteq \exists s.D \in \mathcal{T}$. Either r or A' occur in D and so we have derived a contradiction since they are fresh symbols not in $\mathcal{T}$. Now assume Condition (c2) holds. Then $s = r$ and since r is fresh the inclusion $C \sqsubseteq \exists r.D$ coincides with $E \sqsubseteq \exists r.B$. But then $\mathcal{T} \not\models A \sqsubseteq B$ since otherwise there exists a Σ-concept (namely A') such that $\mathcal{T}' \models E \sqsubseteq \exists r.A'$ and $\mathcal{T}' \models A' \sqsubseteq B$.

"only if". Assume $\mathcal{T} \not\models A \sqsubseteq B$. We have $\mathcal{T}' \models E \sqsubseteq \exists r.B$ and there is no Σ-concept E' with $\mathcal{T}' \models E \sqsubseteq \exists r.E'$ and $\mathcal{T}' \models E' \sqsubseteq B$. Thus $\mathcal{T}'$ is not safe. $\square$

E Proof for Section 5

Formally, a FOQ $q(\vec{x})$ is *domain-independent* if for all interpretations $\mathcal{I}$ and $\mathcal{J}$ such that $P^{\mathcal{I}} = P^{\mathcal{J}}$ for all $P \in \mathsf{sig}(q(\vec{x}))$, we have $\mathcal{I} \models q[\vec{d}]$ iff $\mathcal{J} \models q[\vec{d}]$ for all tuples $\vec{d} \subseteq \Delta^{\mathcal{I}} \cup \Delta^{\mathcal{J}}$. Intuitively, the truth value of a domain-independent FOQ depends only on the interpretation of the predicates, but not on the actual domain of the interpretation. For example, $\neg A(x)$, is not domain-independent whereas $B(x) \wedge \neg A(x)$ is domain-independent.

Let $(\mathcal{T}, \Sigma)$ be a safe DL-Lite$_\mathcal{R}$ TBox or a safe $\mathcal{EL}$ TBox with closed predicates and let $\vartheta(\vec{x})$ be a CQ$^{\mathrm{FO}(\Sigma)}$. In the following, we will only consider satisfiable ABoxes w.r.t. $(\mathcal{T}, \Sigma)$. This is w.l.o.g. because unsatisfiable ABoxes do not affect the results we want to show (cf. Lemma B.3 and Lemma D.4).

Now the case where $\vartheta(\vec{x})$ is a CQ and $(\mathcal{T}, \Sigma)$ is a DL-Lite$_\mathcal{R}$ TBox is covered by Theorem 3.6; and the case where $\vartheta(\vec{x})$ is a CQ and $(\mathcal{T}, \Sigma)$ is an $\mathcal{EL}$ TBox is covered by Theorem 4.4. Thus, suppose that $\vartheta(\vec{x})$ is a CQ$^{\mathrm{FO}(\Sigma)}$ that is not a CQ. This means that there is some conjunct φ of $\vartheta(\vec{x})$ that is a complex, i.e., not of the form $A(t)$ or $r(t, t')$, domain-independent first-order formula over Σ. W.l.o.g. we assume that φ is the only domain-independent first-order formula over Σ in $\vartheta(\vec{x})$; because if this is not the case then we can reorder the conjuncts of $\vartheta(\vec{x})$ so that domain-independent first-order formula over Σ come before other formulae meaning that the conjunction of initial formulae over Σ is now a domain-independent first-order formula over Σ itself.

Now if $\vartheta = \varphi$ then by the domain-independence of φ and $\mathsf{sig}(\varphi) \subseteq \Sigma$, it immediately follows that for every satisfiable ABox $\mathcal{A}$ w.r.t. $(\mathcal{T}, \Sigma)$ and every $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, we have

$$\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \vartheta(\vec{a}) \text{ iff } \mathcal{I}_\mathcal{A} \models \varphi(\vec{a}),$$

where $\mathcal{I}_\mathcal{A}$ is the interpretation corresponding to $\mathcal{A}$. Thus, suppose that $\vartheta(\vec{x}) = \exists y_1 \ldots \exists y_n(\varphi \wedge \psi_1 \wedge \ldots \wedge \psi_m)$, where $n \geq 0$, $m \geq 1$, and $\mathsf{sig}(\psi_i) \cap \Sigma = \emptyset$ for all $i \in \{1, \ldots, m\}$. Note that

(i) some of the variables in $y_1, \ldots, y_n$ may not occur free in φ, and

(ii) some others from the same sequence may not occur free in any one of ψ_i.

W.l.o.g. let $y_1, \ldots, y_k$ be the sequence of variables of type (i), $y_{k+1}, \ldots, y_j$ be the sequence of variables of type (ii), and $y_{j+1}, \ldots, y_n$ be the remaining sequence of variables, where $k \leq j \leq n$. We rewrite $\vartheta(\vec{x})$ to obtain the formula

$$\exists y_{j+1} \ldots \exists y_n[\exists y_{k+1} \ldots \exists y_j \varphi \wedge \exists y_1 \ldots \exists y_k(\psi_1 \wedge \ldots \wedge \psi_m)].$$

Obviously this formula is equivalent to $\vartheta(\vec{x})$, and $\exists y_1 \ldots \exists y_k(\psi_1 \wedge \ldots \wedge \psi_m)$ is a CQ.

Thus, we can assume w.l.o.g. that

$$\vartheta(\vec{x}) = \exists \vec{y}(\varphi(\vec{x}, \vec{y}) \wedge \psi(\vec{x}, \vec{y}))$$

where φ is a domain-independent first-order formula with $\mathsf{sig}(\varphi) \subseteq \Sigma$, and ψ is a CQ with $\mathsf{sig}(\psi) \cap \Sigma = \emptyset$.

Let $\mathcal{A}$ be a satisfiable ABox w.r.t. $(\mathcal{T}, \Sigma)$ and let $\vec{a}$ be a tuple of individual names from $\mathsf{Ind}(\mathcal{A})$ that is of the same length as $\vec{x}$. Using the domain-independence of φ, $\mathsf{sig}(\varphi) \subseteq \Sigma$, and the fact that every model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$ agrees on the extension of predicates in Σ with $\mathcal{I}_\mathcal{A}$ we conclude

$$\begin{aligned}&\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \vartheta(\vec{a}) \text{ iff}\\ &\exists \vec{b} \subseteq \mathsf{Ind}(\mathcal{A}) \text{ such that } \mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\vec{a}, \vec{b}) \wedge \psi(\vec{a}, \vec{b}),\end{aligned} \tag{1}$$

where $\vec{b}$ is of the same length as $\vec{y}$. Obviously, for all $\vec{c} \subseteq \mathsf{Ind}(\mathcal{A})$,

$$\begin{aligned}&\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\vec{a}, \vec{c}) \wedge \psi(\vec{a}, \vec{c}) \text{ iff}\\ &\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\vec{a}, \vec{c}) \text{ and } \mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \psi(\vec{a}, \vec{c}).\end{aligned} \tag{2}$$

Now by the domain-independence of φ, $\mathsf{sig}(\varphi) \subseteq \Sigma$, and the fact that any model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$ agrees on the extension of predicates in Σ with $\mathcal{I}_\mathcal{A}$ we obtain for all $\vec{c} \subseteq \mathsf{Ind}(\mathcal{A})$

$$\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\vec{a}, \vec{c}) \text{ iff } \mathcal{I}_\mathcal{A} \models \varphi(\vec{a}, \vec{c}). \tag{3}$$

So far we have assumed that $\mathcal{T}$ is either a DL-Lite$_\mathcal{R}$ TBox or an $\mathcal{EL}$ TBox. In the rest of the proof we will distinguish between these two cases to show the desired results. In both cases though, we make use of (1), (2), and (3).

DL-Lite$_\mathcal{R}$

We know by Lemma B.2 and [Calvanese *et al.*, 2007a] that there is some domain-independent FOQ ψ' such that for every satisfiable ABox $\mathcal{A}$ w.r.t. $(\mathcal{T}, \Sigma)$ and every $\vec{c} \subseteq \mathsf{Ind}(\mathcal{A})$, we have

$$\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \psi(\vec{a}, \vec{c}) \text{ iff } \mathcal{I}_\mathcal{A} \models \psi'[\vec{a}, \vec{c}]. \tag{4}$$

We have enough results to show the theorem for DL-Lite$_\mathcal{R}$. The following are equivalent:

- $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \vartheta(\vec{a})$
- $\exists \vec{b} \subseteq \mathsf{Ind}(\mathcal{A})$ such that $\mathcal{I}_\mathcal{A} \models \varphi(\vec{a}, \vec{b}) \wedge \psi'(\vec{a}, \vec{b})$ (by (1), (2), (3), and (4))
- $\mathcal{I}_\mathcal{A} \models \exists \vec{y}(\varphi(\vec{a}, \vec{y}) \wedge \psi'(\vec{a}, \vec{y}))$.

Hence CQ$^{\mathrm{FO}(\Sigma)}$-answering w.r.t. safe DL-Lite$_\mathcal{R}$ TBoxes with closed predicates is in AC^0.

$\mathcal{EL}$

By (1), (2), and (3), the following are equivalent

- $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \vartheta(\vec{a})$
- $\exists \vec{b} \subseteq \mathsf{Ind}(\mathcal{A})$ such that $\mathcal{I}_\mathcal{A} \models_{c(\Sigma)} \varphi(\vec{a}, \vec{b})$ and $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \psi(\vec{a}, \vec{b})$.

This suggests an algorithm for CQ$^{\mathrm{FO}(\Sigma)}$-answering in $\mathcal{EL}$. In detail, the algorithm goes through all tuples $\vec{b} \subseteq \mathsf{Ind}(\mathcal{A})$ until one that satisfies $\mathcal{I}_\mathcal{A} \models \varphi(\vec{a}, \vec{b})$ and $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \psi(\vec{a}, \vec{b})$ can be found. There are polynomially many such tuples in the size of the data since $|\vec{b}|$ is fixed, $\mathcal{I}_\mathcal{A} \models \varphi(\vec{a}, \vec{b})$ can be checked in AC^0, and $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \psi(\vec{a}, \vec{b})$ is standard CQ answering in $\mathcal{EL}$, which can be done in PTIME. Hence for safe $\mathcal{EL}$ TBoxes with closed predicates, CQ$^{\mathrm{FO}(\Sigma)}$-answering is in PTIME.

F Proof for Section 6

Lemma 6.1.

1. $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$ iff $\mathcal{I}$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$;
2. if $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$, then for all CQs q and $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q[\vec{a}]$ iff $\mathcal{I} \models q[\vec{a}]$.

Proof. As a preliminary, we note the following, which is immediate in view of the construction of $\mathcal{I}$ (no matter whether or not $\mathcal{A}$ is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$).

Claim. If $a \in A^{\mathcal{I}}$, then $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} A(a)$.

For Point 1, it suffices to prove the contrapositive of the "if" direction. Thus assume that $\mathcal{I}$ is not a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$. Since $\mathcal{I}$ is by definition a model of $\mathcal{A}$ and $(\emptyset, \Sigma)$, it must violate one of the CIs in $\mathcal{T}$. By Point (c) of the construction of $\mathcal{I}$, the last CI $\exists r.(A \sqcap B) \sqsubseteq A$ in $\mathcal{T}$ is not violated. This leaves us with two cases:

- $\top \sqsubseteq \exists r.A$ is violated.

  By Point (b) of the construction of $\mathcal{I}$, this means that there is an $a_0 \in \mathsf{Ind}(\mathcal{A})$ such that whenever $r(a_0, a) \in \mathcal{A}$, then there is an $r(a, b) \in \mathcal{A}$ with $B(b) \notin \mathcal{A}$. Then $\mathcal{A}$ is unsatisfiable w.r.t. $(\mathcal{T}, \Sigma)$ since any model $\mathcal{J}$ of $\mathcal{A}$ and $(\mathcal{T}, \Sigma)$ has to make A true at some $a \in \mathsf{Ind}(\mathcal{A})$ with $r(a_0, a) \in \mathcal{T}$ (by the first CI in $\mathcal{T}$ and since $r \in \Sigma$), but this means (by the second CI in $\mathcal{T}$ and since $B \in \Sigma$) that whenever $r(a, b) \in \mathcal{A}$, then $B(b) \in \mathcal{A}$.

- $\exists r^-.A \sqsubseteq B$ is violated.

  Then there are $a, b \in \mathsf{Ind}(\mathcal{A})$ such that $(a, b) \in r^{\mathcal{I}}$ and $a \in A^{\mathcal{I}}$ and $b \notin B^{\mathcal{I}}$. By the claim, any model $\mathcal{J}$ of $\mathcal{A}$ and $(\mathcal{T}, \Sigma)$ has to make A true at a. Moreover, $b \notin B^{\mathcal{I}}$ implies $B(b) \notin \mathcal{A}$, and thus $B \in \Sigma$ implies that $\mathcal{J}$ has to make B false at b, thus also violates the second CI in $\mathcal{T}$. Consequently, $\mathcal{A}$ is not satisfiable w.r.t. $(\mathcal{T}, \Sigma)$.

Now for Point 2. Let $\mathcal{A}$ be satisfiable w.r.t. $(\mathcal{T}, \Sigma)$ and take a CQ q and $\vec{a} \subseteq \mathsf{Ind}(\mathcal{A})$. By Point 1, $\mathcal{I}$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$. Thus, the "only if" direction is immediate. For the "if" direction, assume that $\mathcal{I} \models q[\vec{a}]$. By the claim and by construction of $\mathcal{I}$, we can find $\mathcal{I}$ in any model $\mathcal{J}$ of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$, that is, $p^{\mathcal{I}} \subseteq p^{\mathcal{J}}$ for all predicates p. Thus, $\mathcal{J} \models q[\vec{a}]$ for all these $\mathcal{J}$. $\square$