Mixing Open and Closed World Assumption in Ontology-Based Data Access: Non-Uniform Data Complexity

Carsten Lutz[1], İnanç Seylan[1], and Frank Wolter[2]

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

Abstract. When using ontologies to access instance data, it can be useful to make a closed world assumption (CWA) for some predicates and an open world assumption (OWA) for others. The main problem with such a setup is that conjunctive query (CQ) answering becomes intractable already for inexpressive description logics such as DL-Lite and $\mathcal{EL}$. We take a closer look at this situation and carry out a fine-grained complexity analysis by considering the complexity of CQ answering w.r.t. individual TBoxes. Our main results are a dichotomy between AC^0 and coNP for TBoxes formulated in DL-Lite and a dichotomy between PTIME and coNP for $\mathcal{EL}$-TBoxes. In each tractable case, CQ answering coincides with CQ answering under pure OWA; the CWA might still be useful as it allows queries that are more expressive than CQs.

1 Introduction

Description logics (DLs) increasingly find application in ontology-based data access (OBDA), where an ontology is used to enrich instance data and the chief aim 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, for example 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 instance data is taken from a relational database, then the CWA can be appropriate for the data predicates while additional predicates in the ontology should always be interpreted under the OWA (this is the very idea of OBDA). As a concrete example, consider geographical databases such as OpenStreetMap which contain pure geographical data as well as rich annotations, stating for example that a certain polygon describes a 'popular Thai restaurant'. As argued in [11, 7], it is useful to pursue an OBDA approach to take full advantage of the annotations, where one would naturally interpret the geographical data under the CWA and the annotations under the OWA.

In the DL literature, there are a variety of approaches to adopting the CWA, often based on epistemic operators or rules [6, 8, 10, 19, 22]. In this paper, we adopt the standard semantics from relational databases, which is natural, and straightforward: CWA predicates have to be interpreted exactly as described in the data, assuming standard (and thus unique) names 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 [23]. In fact, the setup considered in this paper generalizes both standard OBDA (only OWA predicates permitted) and DBoxes (only CWA predicates permitted in data) by allowing to freely mix OWA and CWA predicates both in the TBox and in the data. For readability, we will from now on speak of open and closed predicates rather than OWA and CWA predicates.

A major problem in admitting closed predicates is that query answering easily becomes intractable regarding data complexity (where the TBox and query are assumed to be fixed and thus of constant size). In fact, this is true already for instance queries (IQs), when only closed predicates are admitted in the data, and for TBoxes formulated in inexpressive DLs such as the core dialect of DL-Lite [5] and $\mathcal{EL}$ [3]; this is shown for conjunctive queries (CQs) and DL-Lite in [9], can easily be transfered to $\mathcal{EL}$, and strengthened to IQs by adapting a well-known reduction of Schaerf [21]. While this is a relevant and interesting first step, it was recently demonstrated in [15, 16] in the context of standard OBDA with more expressive DLs that a more fine grained, 'non-uniform' analysis is possible by studying data complexity on the level of individual TBoxes instead of on the level of logics. In our context, we work with TBoxes of the form $(\mathcal{T}, \Sigma)$, where $\mathcal{T}$ is a set of TBox statements as usual and Σ is a set of predicates (concept and role names) that are declared to be closed. We say that CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ *is in* PTIME if for every CQ $q(x)$, 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 complexities are defined analogously. The main aim of this paper is to carry out a non-uniform analysis of data complexity for query answering with closed predicates in DL-Lite and $\mathcal{EL}$.

Our main results are a dichotomy between AC^0 and CONP for TBoxes formulated in DL-Lite and a dichotomy between PTIME and CONP for $\mathcal{EL}$-TBoxes. In each case, we provide a transparent characterization that separates the easy cases from the hard cases. These results are interesting when contrasted with query answering w.r.t. TBoxes that are formulated in the expressive DLs $\mathcal{ALC}$ and $\mathcal{ALCI}$, where the data complexity is also between AC^0 and CONP, but where 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 non-uniform constraint satisfaction problems [16]. We also show that when CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in PTIME, then the certain answers to any CQ q in any ABox $\mathcal{A}$ given $(\mathcal{T}, \Sigma)$ (which respect the closed-world declarations in Σ) coincide with the open world answers to q in $\mathcal{A}$ given $\mathcal{T}$—for ABoxes that are satisfiable w.r.t. $\mathcal{T}$. In a sense, we thus show that CQ answering with closed predicates is *inherently intractable*: in all the tractable and consistent cases, the declaration of closed predicates does not have any impact on query answers.

While this sounds discouraging, there is still a potential benefit of closed predicates in tractable cases: for the 'closed part' of the signature, we can go beyond conjunctive queries and admit (almost) full first-order queries without becoming undecidable and, indeed, without any negative impact on data complexity. We propose a concrete query language that implements this idea and show that AC^0 data complexity is preserved for DL-Lite TBoxes and PTIME data complexity is preserved for $\mathcal{EL}$-TBoxes when CQs are replaced with queries formulated in the extended language.

2 Preliminaries

We use standard notation from description logic [4]. 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 $r \in \mathsf{N_R}$. A *DL-Lite-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 *DL-Lite-TBox* is a finite set of DL-Lite-inclusions. In the literature, this version of DL-Lite is often called DL-Lite$_{\mathsf{core}}$. *$\mathcal{EL}$-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 *$\mathcal{EL}$-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 $\mathcal{EL}$-inclusions. $\mathcal{ELI}$ extends $\mathcal{EL}$ with the constructor $\exists r^-.C$. *ABoxes* are finite sets of assertions $A(a)$ and $r(a, b)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and a, b individual names. We use $\mathsf{Ind}(\mathcal{A})$ to denote the set of individual names used in the ABox $\mathcal{A}$ and write $r^-(a, b) \in \mathcal{A}$ instead of $r(b, a) \in \mathcal{A}$. We sometimes also use *infinite* ABoxes, but this will be stated explicitly. Interpretations $\mathcal{I}$ are defined as usual, where for the interpretation of individual names we make the standard name assumption (SNA), i.e., $a^{\mathcal{I}} = a$. Note that this implies the unique name assumption (UNA); to avoid enforcing infinite models, we assume that interpretations need not interpret all individual names and use $\mathsf{Ind}(\mathcal{I})$ to denote the individual names interpreted by $\mathcal{I}$. 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 $C^{\mathcal{I}} \neq \emptyset$ (that satisfies $\mathcal{A}$, respectively).

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 $\Delta^{\mathcal{I}}$.

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, $\mathsf{sig}(\mathcal{T})$ of a TBox $\mathcal{T}$, and $\mathsf{sig}(\mathcal{A})$ of an ABox $\mathcal{A}$, is the set of predicates occurring in C, $\mathcal{T}$, and $\mathcal{A}$, 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 $\mathcal{T}$ 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 $\mathsf{Ind}(\mathcal{A}) \subseteq \mathsf{Ind}(\mathcal{I})$ that satisfies $\mathcal{T}$ and $\mathcal{A}$ and such that

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

Note that TBox statements that only involve closed predicates are effectively integrity constraints in the standard database sense [1]. In a DL context, integrity constraints are discussed for example in [6, 8, 17–19].

A *first-order query (FOQ)* $q(\boldsymbol{x})$ is a first-order formula constructed from atoms $A(t)$, $r(t, t')$, and $t = t'$, where t, t' range over individual names and variables and $\boldsymbol{x} =$

$x_1, \ldots, x_k$ contains all free variables of q. We call $\boldsymbol{x}$ the *answer variables of* $q(\boldsymbol{x})$. A *conjunctive query (CQ)* $q(\boldsymbol{x})$ is a FOQ using conjunction and existential quantification, only. A tuple $\boldsymbol{a} = a_1, \ldots, a_k \subseteq \mathsf{Ind}(\mathcal{A})$ *is a certain answer to* $q(\boldsymbol{x})$ *in* $\mathcal{A}$ *given* $(\mathcal{T}, \Sigma)$, in symbols $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\boldsymbol{a})$, if $\mathcal{I} \models q[a_1, \ldots, a_k]$ for all models $\mathcal{I}$ of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$. When computing certain answers we assume that all individual names in the query occur in the ABox. If $\Sigma = \emptyset$, then we simply omit Σ and write $\mathcal{T}, \mathcal{A} \models q(\boldsymbol{a})$ instead of $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\boldsymbol{a})$. If C is an $\mathcal{ELI}$-concept, then the CQ corresponding to $C(a)$ is defined in the usual way and called an $\mathcal{ELI}$-instance query. $\mathcal{EL}$-instance queries are defined analogously.

The following definition generalizes the definition of non-uniform data complexity introduced in [16] to TBoxes with closed predicates.

Definition 1. *Let* $(\mathcal{T}, \Sigma)$ *be a TBox with closed predicates. Then*

- CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ *is in* PTIME *if for every CQ* $q(\boldsymbol{x})$ *there is a polytime algorithm that computes, for a given ABox* $\mathcal{A}$*, all* $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$ *with* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\boldsymbol{a})$*;*
- 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 FOQs and $\mathcal{ELI}$-instance queries, analogous notions can be defined. It is known that for $\Sigma = \emptyset$, CQ answering is in PTIME for $\mathcal{EL}$-TBoxes [5, 13] and in AC^0 for DL-Lite [5, 2]. FOQ-answering is undecidable even for the empty TBox, due to the OWA.

The following property, plays a central role in our analysis; see [16] which also makes intensive use of this notion (there called *ABox* disjunction property).

Definition 2 (Disjunction property). *A TBox with closed predicates* $(\mathcal{T}, \Sigma)$ *has the* disjunction property *if for all ABoxes* $\mathcal{A}$ *and* $\mathcal{ELI}$*-instance queries* $C_1(a)$ *and* $C_2(a)$*,* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} C_1(a) \vee C_2(a)$ *implies* $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} C_i(a)$ *for some* $i \in \{1, 2\}$.

It is standard to show that DL-Lite and $\mathcal{EL}$ TBoxes *without* closed predicates have the disjunction property [16].

3 Main Results and Illustrating Examples

We first formulate the dichotomy result for DL-Lite. The next definition introduces a class of TBoxes with closed predicates that turn out to be exactly the TBoxes for which query answering is in AC^0.

Definition 3. *A DL-Lite TBox* $\mathcal{T}$ *with closed predicates* Σ *is* safe *if there are no DL-Lite-concepts* B_1, B_2 *and role* r *such that*

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

[3] In other words, r is a role name and $r \notin \Sigma$ or $r = s^-$ for a role name $s \notin \Sigma$.

Note that it is easy to check in PTIME whether a given DL-Lite TBox is safe since subsumption in DL-Lite can be decided in PTime [5]. Our results concerning DL-Lite are summarized by the following theorem.

Theorem 1 (DL-Lite dichotomy). *Let $(\mathcal{T}, \Sigma)$ be a DL-Lite-TBox with closed predicates. Then the following holds:*

1. *If $(\mathcal{T}, \Sigma)$ is not safe, then the disjunction property fails and there is an $\mathcal{ELI}$-instance query $C(a)$ such that answering $C(a)$ w.r.t. $(\mathcal{T}, \Sigma)$ is co*NP*-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. $(\mathcal{T}, \Sigma)$, i.e., for every CQ $q(\boldsymbol{x})$ and $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\boldsymbol{a})$ iff $\mathcal{T}, \mathcal{A} \models q(\boldsymbol{a})$.*
   (b) *CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in AC^0 and $(\mathcal{T}, \Sigma)$ has the disjunction property.*

The following example illustrates Theorem 1.

Example 1. (a) Let $\mathcal{T} = \{A \sqsubseteq \exists r.\top, \exists r^-.\top \sqsubseteq B\}$ and $\Sigma = \{B\}$. $(\mathcal{T}, \Sigma)$ is not safe. The disjunction property can be refuted as follows. Let $\mathcal{A} = \{A(a), B(b_1), A_1(b_1), B(b_2), A_2(b_2)\}$, where A_1, A_2 are fresh concept names. Then

1. $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \exists r.(A_1 \sqcap B)(a) \vee \exists r.(A_2 \sqcap B)(a)$;
2. $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} \exists r.(A_i \sqcap B)(a)$ for $i = 1, 2$.

Point 1 should be clear since in any model $\mathcal{I}$ of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$ one has to link a with r to b_1 or to b_2 to satisfy $\mathcal{T}$. For Point 2 and $i \in \{1, 2\}$, consider the model $\mathcal{I}_i$ that corresponds to $\mathcal{A}$ expanded with $r(a, b_i)$. Then $\mathcal{I}_i$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}$ (note that $r \notin \Sigma$) but $a \notin (\exists r.(A_{\bar{i}} \sqcap B))^{\mathcal{I}_i}$ where $\bar{1} = 2$ and $\bar{2} = 1$. Thus $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} \exists r.(A_i \sqcap B)(a)$. When we add any of A, A_1, A_2 to Σ, all statements are still true.

(b) The failure of the disjunction property for the ABox $\mathcal{A}$ and TBox $\mathcal{T}$ results in a choice that enables a coNP-hardness proof by reduction of 2+2-SAT, a variant of propositional satisfiability where each clause contains precisely two positive literals and two negative literals [21]. For this reduction, it suffices to use an $\mathcal{EL}$-query that uses the above queries $\exists r.(A_i \sqcap B)(a)$, $i = 1, 2$, as subqueries for encoding truth values.

The proof of Theorem 1 is given in the next section. We now come to the case of TBoxes formulated in $\mathcal{EL}$, where we start with examples. Observe that we can find an $\mathcal{EL}$-TBox without the disjunction property by a reformulation of the DL-Lite TBox from Example 1. Let $\mathcal{T}' = \{A \sqsubseteq \exists r.B\}$ with $\Sigma = \{B\}$. Then, in the same way as for $(\mathcal{T}, \Sigma)$ one can show that the disjunction property fails for $(\mathcal{T}', \Sigma)$ and that CQ answering is coNP-hard. In $\mathcal{EL}$, however, there is an additional (and more subtle) cause for non-tractability, which we discuss in the following example.

Example 2. Consider again $\mathcal{T}' = \{A \sqsubseteq \exists r.B\}$, but now set $\Sigma' = \{r\}$. We first show that $(\mathcal{T}', \Sigma')$ does not have the disjunction property. Let $\mathcal{A}' = \{A(a), r(a, b_1), A_1(b_1), r(a, b_2), A_2(b_2)\}$, where A_1, A_2 are fresh concept names. Then one can easily show that the disjunction property fails:

- $\mathcal{T}', \mathcal{A}' \models_{c(\Sigma')} \exists r.(A_1 \sqcap B)(a) \vee \exists r.(A_2 \sqcap B)(a)$;

– $\mathcal{T}', \mathcal{A}' \not\models_{c(\Sigma')} \exists r.(A_i \sqcap B)(a)$, for $i = 1, 2$.

It is crucial that $B \notin \Sigma'$. The proof of coNP-hardness is very similar to the proof mentioned in Example 1.

Observe that one cannot reproduce this example in DL-Lite: for the TBox $\mathcal{T} = \{A \sqsubseteq \exists r.\top, \exists r^-.\top \sqsubseteq B\}$ with $\Sigma' = \{r\}$, we have $\mathcal{T}, \mathcal{A}' \models_{c(\Sigma')} B(b_i)$ for $i = 1, 2$ and, therefore, $\mathcal{T}, \mathcal{A}' \models_{c(\Sigma')} \exists r.(A_i \sqcap B)(a)$, for $i = 1, 2$. Thus, the disjunction property is not violated.

We now identify a class of $\mathcal{EL}$-TBoxes with closed predicates that turn out to be exactly the TBoxes for which CQ answering is in PTIME. We call a concept E a *top-level conjunct (tlc) of* C if C is of the form $D_1 \sqcap \cdots \sqcap D_n$ with $n \geq 1$ and $E = D_i$ for some i.

Definition 4. *Let $(\mathcal{T}, \Sigma)$ be 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 $\mathcal{T} \models E \sqsubseteq D$.*

Note that Condition 3(s1) of Definition 4 is similar to the definition of safety for DL-Lite. Example 2 shows why Condition 3(s2) is needed. The following example illustrates the additional requirement of 3(s2) that no "interpolating" Σ-concept E exists.

Example 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 safety because E can be used as a 'Σ-interpolant'. Note that the ABox $\mathcal{A}'$ from Example 2, which we used to refute the disjunction property in a very similar situation, is simply unsatisfiable w.r.t. $(\mathcal{T}, \Sigma)$ because E has to interpreted as the empty set. Indeed, it can be shown that $(\mathcal{T}, \Sigma)$ is safe.

Note that, unlike the DL-Lite case, the definition of safety of $\mathcal{EL}$-TBoxes with closed predicates does not immediately suggest a decision procedure since there are infinitely many candidates for the concepts C, D, and E. We conjecture that safety is decidable in PTIME, but pursuing this further is left for future work.

Theorem 2 **($\mathcal{EL}$-dichotomy).** *Let $(\mathcal{T}, \Sigma)$ be an $\mathcal{EL}$-TBox with closed predicates. Then the following holds:*

1. *If $(\mathcal{T}, \Sigma)$ is not safe, then the disjunction property fails and there exists an $\mathcal{EL}$-instance query $C(a)$ such that answering $C(a)$ w.r.t. $(\mathcal{T}, \Sigma)$ is co*NP*-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. $(\mathcal{T}, \Sigma)$.*

(b) CQ answering w.r.t. $(\mathcal{T}, \Sigma)$ is in PTIME *and $(\mathcal{T}, \Sigma)$ has the disjunction property.*

As noted in the introduction, Theorems 1 and 2 essentially show that CQ answering with closed predicates is inherently intractable. Note, though, that Points 2(a) of these theorems refer only to satisfiable ABoxes. In fact, TBox statements that refer only to closed predicates act as integrity constraints also in the tractable cases. Moreover, safe TBoxes admit *any* integrity constraint that can be formulated in the DL at hand, i.e., if a TBox $\mathcal{T}$ formulated in DL-Lite or $\mathcal{EL}$ is safe, then it is still safe after adding any concept inclusions that refers only to closed predicates. In the appendix, we show that checking satisfiability of ABoxes w.r.t. safe TBoxes is in AC^0 for DL-Lite and in PTIME for $\mathcal{EL}$ (for data complexity).

Another way of taking advantage of closed predicates without losing tractability is to admit more expressive query languages. 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 mentioned in the introduction. In such a setup, one would typically not want to give up FOQs (SQL queries) available in the relational system to accomodate open world predicates. We propose a query language that combines, in a straightforward way, FOQs for closed predicates with CQs for open (and closed) predicates. We then show that, 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). Formally, a FOQ $q(\boldsymbol{x})$ is *domain-independent* if for all interpretations $\mathcal{I}$ and $\mathcal{J}$ such that $P^{\mathcal{I}} = P^{\mathcal{J}}$ for all $P \in \mathsf{sig}(q(\boldsymbol{x}))$, we have $\mathcal{I} \models q[\boldsymbol{d}]$ iff $\mathcal{J} \models q[\boldsymbol{d}]$ for all tuples $\boldsymbol{d} \subseteq \Delta^{\mathcal{I}} \cup \Delta^{\mathcal{J}}$. Intuitively, the truth value of a domain-independent FOQ depends only on the interpretation of the data 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. In our query language, we allow domain-independent FOQs over closed predicates as atoms in CQs.

Definition 5. *Let Σ be a signature that declares closed predicates. A* conjunctive query with FO(Σ) plugins (abbreviated $\mathrm{CQ}^{\mathrm{FO}(\Sigma)}$) *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 either a CQ or a domain-independent FOQ whose signature is included in Σ.*

The subsequent theorem shows that switching from CQs to $\mathrm{CQ}^{\mathrm{FO}(\Sigma)}$s does not increase data complexity.

Theorem 3.

1. *$CQ^{FO(\Sigma)}$-answering w.r.t. safe DL-Lite-TBoxes with closed predicates is in AC^0. More precisely, for every such TBox $(\mathcal{T}, \Sigma)$ and $CQ^{FO(\Sigma)}$ $q(\boldsymbol{x})$, there exists a FOQ $q'(\boldsymbol{x})$ such that for all $\mathcal{A}$ and $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$: $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\boldsymbol{a})$ iff $\mathcal{I}_{\mathcal{A}} \models q'(\boldsymbol{a})$.*
2. *$CQ^{FO(\Sigma)}$-answering w.r.t. safe $\mathcal{EL}$-TBoxes is in* PTIME.

Query languages between CQs and FOQs have been studied before. In the standard setup where all predicates are open, it was shown in [20] that extending CQs with union and atomic negation results in coNP-hardness both in DL-Lite and in $\mathcal{EL}$, and that extending CQs with union and inequality results in undecidability in $\mathcal{EL}$. In our query language $\text{CQ}^{\text{FO}(\Sigma)}$, we avoid these problems by allowing only CQs for the open predicates while restricting the expressive power of FOQs (which admits disjunction, full negation, and (in)equality) to closed predicates. The language EQL-Lite(CQ) proposed in [6] can, in some sense, be viewed as a fragment of our language that admits the full expressivity of FOQs, but in which only closed predicates are admitted. Note though, that the EQL-Lite approach closes predicates only for querying while all predicates are interpreted as open world for TBox reasoning.

4 Proof Sketches

We sketch proofs of Theorems 1 and 2. We begin with the first part of Theorem 1.

Lemma 1. *If a DL-Lite-TBox $\mathcal{T}$ with closed predicates Σ is not safe, then the disjunction property fails and there exists an $\mathcal{ELI}$-instance query $C(a)$ such that answering $C(a)$ w.r.t. $(\mathcal{T},\Sigma)$ is co*NP*-hard.*

Proof. Assume that $B_1 \sqsubseteq \exists r.\top, \exists r^-.\top \sqsubseteq B_2$ satisfy the conditions of Definition 3. Take a finite model $\mathcal{I}$ of $\mathcal{T}$ with $a_0 \in B_1^{\mathcal{I}}$; such a model exists since B_1 is satisfiable w.r.t. $\mathcal{T}$ and DL-Lite has the finite model property. Let $S = \{b \in \Delta^{\mathcal{I}} \mid (a_0,b) \in r^{\mathcal{I}}\}$. Since $B_1 \neq \exists r.\top$, we have $a_0 \in B_1^{\mathcal{I}_S}$ for the interpretation $\mathcal{I}_S$ obtained from $\mathcal{I}$ by removing all pairs (a_0,b) with $b \in S$ from $r^{\mathcal{I}}$. Take the ABox $\mathcal{A}_S$ corresponding to $\mathcal{I}_S$ and let $\mathcal{A}$ be the disjoint union of two copies of $\mathcal{A}_S$. 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

$$\mathcal{A}' = \mathcal{A} \cup \{A_1(b,1) \mid b \in B_2^{\mathcal{I}}\} \cup \{A_2(b,2) \mid b \in B_2^{\mathcal{I}}\},$$

where A_1 and A_2 are fresh concept names. Now one can show that the disjunction property fails: $(\mathcal{T},\mathcal{A}') \models_{c(\Sigma)} \exists r.(A_1 \sqcap B_2)(a_0,1) \vee \exists r.(A_2 \sqcap B_2)(a_0,1)$ and $(\mathcal{T},\mathcal{A}') \not\models_{c(\Sigma)} \exists r.(A_i \sqcap B_2)(a_0,1)$ for $i=1,2$.

The coNP-hardness proof is now similar to the proof for Example 1 given in the appendix. ❑

For the second part of Theorem 1, we first prove (a):

Lemma 2. *Let $(\mathcal{T},\Sigma)$ be a DL-Lite 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. $\mathcal{T}$ without closed predicates 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!) [12]. $\mathcal{I}$ 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. Let i be minimal such that there exists $B_1 \sqsubseteq B_2 \in \mathcal{T}$ with $\mathcal{A}_j \models B_1(a_i)$ but $\mathcal{A}_j \not\models B_2(a_i)$ (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:

1. For all CQs $q(\boldsymbol{x})$ and $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$: $\mathcal{T}, \mathcal{A} \models q(\boldsymbol{a})$ iff $\mathcal{J} \models q[\boldsymbol{a}]$.
2. 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 $\mathcal{A}$ we prove the following

Claim 1. For all $i \geq 0$ and for all $a \in \mathsf{Ind}(\mathcal{A}_i)$, if $B_1 \sqsubseteq B_2 \in \mathcal{T}$ with $\mathcal{A}_i \models B_1(a)$ but $\mathcal{A}_i \not\models B_2(a)$, then $\mathsf{sig}(B_2) \cap \Sigma = \emptyset$.

Claim 1 holds for all $\mathcal{A}_i$, $i \geq 0$, and all $a \in \mathsf{Ind}(\mathcal{A})$: otherwise, $\mathcal{A}_{i+1}$ is unsatisfiable, in contradiction to Point 1. It follows that Claim 1 holds for $i = 0$. We proceed by induction, assuming that Claim 1 has been proved for $\mathcal{A}_i$, but that to the contrary of what is to be shown there are $B_1 \sqsubseteq B_2 \in \mathcal{T}$ with $\mathcal{A}_{i+1} \models B_1(a)$, $\mathcal{A}_{i+1} \not\models B_2(a)$, and $\mathsf{sig}(B_2) \cap \Sigma \neq \emptyset$, i.e., $\mathsf{sig}(B_2) \subseteq \Sigma$. By what was said above, we have $a \notin \mathsf{Ind}(\mathcal{A})$ and thus a was introduced as a witness for some $\exists s.\top$. By IH, $\mathsf{sig}(s) \cap \Sigma = \emptyset$ and there exists $B_1 \neq \exists s.\top$ with $B_1 \sqsubseteq \exists s.\top \in \mathcal{T}$. Point 2 yields $\mathcal{T} \models \exists s^-.\top \sqsubseteq B_2$, in contrary to $(\mathcal{T}, \Sigma)$ being safe. This finishes the proof of Claim 1.

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

To obtain a proof of Part 2 of Theorem 1, it remains to show that, for safe $(\mathcal{T}, \Sigma)$, it is in AC^0 to decide whether an ABox is satisfiable w.r.t. $(\mathcal{T}, \Sigma)$. To this end, it is readily checked that an ABox $\mathcal{A}$ which is satisfiable w.r.t. $\mathcal{T}$ is satisfiable w.r.t. a safe $(\mathcal{T}, \Sigma)$ iff $\mathcal{T}, \mathcal{A} \models B(a)$ implies $\mathcal{A} \models B(a)$ for all DL-Lite concepts B over Σ. To see that this condition is in AC^0, let $\varphi_B(x)$ be an FO query 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 $\mathcal{A}$. Then $\mathcal{A}$ is not satisfiable w.r.t. $(\mathcal{T}, \Sigma)$ iff $\mathcal{I}_{\mathcal{A}} \models \exists x \bigvee_{B \in X} (\varphi_B(x) \wedge \neg B(x))$ where X denotes the set of all DL-Lite concepts over Σ.

We now come to Theorem 2. With the exception of proof steps involving Condition 3(s2) of Definition 4, the proof technique for Theorem 2 extends the proof technique introduced for DL-Lite. We therefore focus on 3(s2) and refer the reader to the appendix for the full proof. 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 [23, 24]. Note that it is different from the interpolation property investigated in [14], which requires the interpolant to be a TBox instead of a concept.

Lemma 3 (Interpolation). *Let $\mathcal{T}_1, \mathcal{T}_2$ be $\mathcal{EL}$-TBoxes, $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq D_1$ with $\mathsf{sig}(\mathcal{T}_1, D_0) \cap \mathsf{sig}(\mathcal{T}_2, D_1) \subseteq \Sigma$. Then there exists a Σ-concept F such that $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq F$ and $\mathcal{T}_1 \cup \mathcal{T}_2 \models F \sqsubseteq D_1$.*

The following lemma is the crucial step for proving Part 1 of Theorem 2 if 3(s2) applies.

Lemma 4. *Let $(\mathcal{T}, \Sigma)$ be an $\mathcal{EL}$-TBox with closed predicates such that safety is violated by $C \sqsubseteq \exists r.D \in \mathcal{T}$ because 3(s2) holds. Then the disjunction property fails.*

Proof. Assume $C \sqsubseteq \exists r.D$ is given. Take the canonical model $\mathcal{I}_{\mathcal{T},C}$ of $\mathcal{T}$ and C as defined in [14] (its domain $\Delta^{\mathcal{I}_{\mathcal{T},C}}$ consists of names a_F, F a subconcept of $\mathcal{T}$ or C, and $a_F \in G^{\mathcal{I}_{\mathcal{T},C}}$ iff $\mathcal{T} \models F \sqsubseteq G$, for all $\mathcal{EL}$-concepts G). Assume for simplicity that $(a_F, a_C) \notin r^{\mathcal{I}_{\mathcal{T},C}}$ for any $a_F \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$. Let

$$S = \{a_G \in \Delta^{\mathcal{I}_{\mathcal{T},C}} \mid (a_C, a_G) \in r^{\mathcal{I}_{\mathcal{T},C}}, \exists r.G \text{ is not a top level conjunct of } C\}$$

and 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 $r^{\mathcal{I}_{\mathcal{T},C}}$. We have $a_C \in C^{\mathcal{I}_S}$. Let $\mathcal{A}_S$ be the ABox corresponding to $\mathcal{I}_S$ and

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

Since there is no tlc C' of C with $\mathcal{T} \models C' \sqsubseteq D$, by a result of [14] (Lemma 16), there exists $G \in K$ with $\mathcal{T} \models G \sqsubseteq D$.

Introduce copies X^0 and X^1 of any non-Σ-predicate X. Denote by E^0 and E^1 the resulting concept if each non-Σ predicate X in E is replaced by X^0 and, respectively, X^1. Similarly, denote by $\mathcal{T}^0$ and $\mathcal{T}^1$ the TBoxes obtained from $\mathcal{T}$ by replacing all concepts E in $\mathcal{T}$ by E^0 and E^1, respectively. The following can be proved using Lemma 3:

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

Now one 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 $\mathcal{J}_G$ and 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 $G \in K$. Introduce two copies $\mathcal{A}^1_{G,\Sigma}$ and $\mathcal{A}^2_{G,\Sigma}$ 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 $a \in \mathsf{Ind}(\mathcal{A}_{G,\Sigma})$. Define the ABox $\mathcal{A}$ by taking two fresh concept names A_1 and A_2 and the union of $\mathcal{A}_S \cup \bigcup_{G \in K} \mathcal{A}^1_{G,\Sigma} \cup \mathcal{A}^2_{G,\Sigma}$ and the assertions $r(a_C, (a_G, 1)), r(a_C, (a_G, 2))$, $A_1(a_G, 1)$, and $A_2(a_G, 2)$, for every $G \in K$ and $A_1(a_{D'})$, for every tlc $\exists r.D'$ of C. One can show that the disjunction property is violated: $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \exists r.(A_1 \sqcap D)(a_C) \vee \exists r.(A_2 \sqcap D)(a_C)$ and $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} \exists r.(A_i \sqcap D)(a_C)$, for $i = 1, 2$. ❑

5 Future Work

We have presented first results regarding the non-uniform complexity of query answering in the presence of open and closed world predicates. We expect the proposed extension of CQs with FO-plugins to be useful in practical applications where closed predicates are important and safe TBoxes suffice. As future work, we plan to extend our investigation to more expressive DLs. For example, we conjecture that transparent dichotomy results can still be obtained for the extensions of DL-Lite$_{core}$ and $\mathcal{EL}$ with role hierarchies.

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

References

1. S. Abiteboul, R. Hull, and V. Vianu. Foundations of Databases. Addison-Wesley, 1995.
2. A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. The DL-Lite family and relations. Journal of Artificial Intelligence Research. (JAIR), 36:1–69, 2009.
3. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In IJCAI, pages 364–369, 2005.
4. F. Baader, D. Calvanese, D. McGuiness, D. Nardi, and P. Patel-Schneider. The Description Logic Handbook: Theory, implementation and applications. Cambridge Univ. Press, 2003.
5. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. Journal of Automated reasoning, 39(3):385–429, 2007.
6. D. Calvanese, G. D. Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. EQL-Lite: Effective first-order query processing in description logics. In IJCAI, pages 274–279, 2007.
7. M. Codescu, G. Horsinka, O. Kutz, T. Mossakowski, and R. Rau. DO-ROAM: Activity-oriented search and navigation with OpenStreetMap. In GeoSpatial Semantics, volume 6631 of LNCS, pages 88–107. Springer, 2011.
8. F. M. Donini, D. Nardi, and R. Rosati. Description logics of minimal knowledge and negation as failure. ACM Transactions on Computational Logic, 3(2):177–225, 2002.
9. E. Franconi, Y. A. Ibáñez-García, and I. Seylan. Query answering with DBoxes is hard. Electronic Notes on Theoretical Computer Science, 278:71–84, 2011.
10. S. Grimm and B. Motik. Closed world reasoning in the semantic web through epistemic operators. In OWLED, volume 188 of CEUR Workshop Proceedings. CEUR-WS.org, 2005.
11. S. Hübner, R. Spittel, U. Visser, and T. J. Vögele. Ontology-based search for interactive digital maps. IEEE Intelligent Systems, 19(3):80–86, 2004.
12. R. Kontchakov, C. Lutz, D. Toman, F. Wolter, and M. Zakharyaschev. The combined approach to query answering in DL-Lite. In KR, 2010.
13. C. Lutz, D. Toman, and F. Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In IJCAI, pages 2070–2075, 2009.
14. C. Lutz and F. Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. Journal of Symbolic Computation, 45(2):194–228, 2010.
15. C. Lutz and F. Wolter. Non-uniform data complexity of query answering in description logics. In Description Logics, 2011.
16. C. Lutz and F. Wolter. Non-uniform data complexity of query answering in description logics. In KR, 2012.
17. A. Mehdi, S. Rudolph, and S. Grimm. Epistemic querying of OWL knowledge bases. In ESWC, volume 6643 of LNCS, pages 397–409. Springer, 2011.
18. B. Motik, I. Horrocks, and U. Sattler. Bridging the gap between OWL and relational databases. Journal of Web Semantics, 7(2):74–89, 2009.
19. B. Motik and R. Rosati. Reconciling description logics and rules. J. ACM, 57(5), 2010.
20. R. Rosati. The limits of querying ontologies. In ICDT, pages 164–178, 2007.
21. A. Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. Journal of Intelligent Information Systems, 2:265–278, 1993.
22. K. Sengupta, A. A. Krisnadhi, and P. Hitzler. Local closed world semantics: Grounded circumscription for OWL. In ISWC, volume 7031 of LNCS, pages 617–632. Springer, 2011.
23. I. Seylan, E. Franconi, and J. de Bruijn. Effective query rewriting with ontologies over DBoxes. In IJCAI, pages 923–929, 2009.
24. B. ten Cate, E. Franconi, and I. Seylan. Beth definability in expressive description logics. In IJCAI, pages 1099–1106, 2011.

A Proofs for Section 3

We prove coNP-hardness for Example 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 [21]. 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 [21] that 2+2-SAT is NP-complete.

Let $\varphi = c_0 \wedge \cdots \wedge c_n$ be a 2+2 formula in propositional letters $q_0, \ldots, q_m$, and let $c_i = p_{i,1} \vee p_{i,2} \vee \neg n_{i,1} \vee \neg n_{i,2}$ for all $i \leq n$. Our aim is to define an ABox $\mathcal{A}_\varphi$ and 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 $q_0, \ldots, q_m$ represent variables, and the individual names $0, 1$ represent truth constants;
- the assertions

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

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

We further extend $\mathcal{A}_\varphi$ to enforce a truth value for each of the variables q_i and the truth-constants $0, 1$. To this end, add to $\mathcal{A}_\varphi$ copies $\mathcal{A}_0, \ldots, \mathcal{A}_m$ of $\mathcal{A}$ 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 q_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 q_i, where we want to interpret q_i as true if the query $\exists r.(A_1 \sqcap B)(q_i)$ is satisfied and as false if the query $\exists r.(A_2 \sqcap B)(q_i)$ is satisfied.

To ensure that 0 and 1 have the expected truth values, add the ABoxes $\mathcal{A}(1) = \{r(1, c_1), A_1(c_1), B(c_1)\}$ and $\mathcal{A}(0) = \{r(0, c_2), A_2(c_2), B(c_2)\}$. Let $\mathcal{B}$ be the resulting ABox.

Consider the query

$$q_0 = \exists c.(\exists p_1.\mathsf{ff} \sqcap \exists p_2.\mathsf{ff} \sqcap \exists n_1.\mathsf{tt} \sqcap \exists n_2.\mathsf{tt})$$

which describes the existence of a clause with only false literals and thus captures falsity of φ, where tt is an abbreviation for $\exists r.(A_1 \sqcap B)$ and ff an abbreviation for $\exists r.(A_2 \sqcap B)$. It is straightforward to show that φ is unsatisfiable iff $\mathcal{T}, \mathcal{B} \models_{c(\Sigma)} q_0$.

B Proofs for Section 4

We first provide the missing step for the proof of Lemma 1.

Lemma 5. *With the definitions from the proof of Lemma 1:*
(1) $(\mathcal{T}, \mathcal{A}') \models_{c(\Sigma)} \exists r.(A_1 \sqcap B_2)(a_0, 1) \vee \exists r.(A_2 \sqcap B_2)(a_0, 1)$
(2) $(\mathcal{T}, \mathcal{A}') \not\models_{c(\Sigma)} \exists r.(A_i \sqcap B_2)(a_0, 1)$ *for* $i = 1, 2$.

Proof. (1) Let $\mathcal{J}$ be a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}'$. We have $(a_0, 1) \in B_1^{\mathcal{J}}$. Since $\mathcal{J}$ is a model of $\mathcal{T}$, 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.

(2) 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

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

It is readily checked that $\mathcal{J}$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}'$. Moreover, $(a_0, 1) \notin (\exists r.(A_1 \sqcap B_2))^{\mathcal{J}}$. 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 $\mathcal{I}_S$: in this case, $\mathcal{J}$ is defined as the interpretation corresponding to the ABox $\mathcal{A}'$ extended by

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

Again $\mathcal{J}$ is a model of $(\mathcal{T}, \Sigma)$ and $\mathcal{A}'$, and $(a_0, 1) \notin (\exists r.(A_2 \sqcap B_2))^{\mathcal{J}}$. ❑

We now prove Theorem 2. We split Part 1 of Theorem 2 into two parts, and begin with the case in which condition 3(s1) for non-safety is satisfied.

Lemma 6. *Let* $(\mathcal{T}, \Sigma)$ *be a* $\mathcal{EL}$*-TBox with closed predicates such that safety 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 the disjunction property fails and there exists an* $\mathcal{EL}$*-instance query* $C(a)$ *such that answering* $C(a)$ *w.r.t.* $(\mathcal{T}, \Sigma)$ *is co*NP*-hard.*

Proof. Assume $C \sqsubseteq \exists r.D$ with the properties of Lemma 6 is given. We remind the reader of the canonical model $\mathcal{I}_{\mathcal{T},C}$ of a $\mathcal{EL}$-TBox $\mathcal{T}$ and $\mathcal{EL}$-concept C [14]. 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). The *canonical model* $\mathcal{I}_{C,\mathcal{T}} = (\Delta^{C,\mathcal{T}}, \cdot^{C,\mathcal{T}})$ of C and $\mathcal{T}$ is defined as follows:

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

$\mathcal{I}_{\mathcal{T},C}$ is a model of $\mathcal{T}$ and

Fact 1. The following conditions are equivalent for all D_0 with $a_{D_0} \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$ and all $\mathcal{EL}$-concepts D_1:

1. $\mathcal{T} \models D_0 \sqsubseteq D_1$;
2. $a_{D_0} \in D_1^{\mathcal{I}_{\mathcal{T},C}}$.

Note that by our assumptions there is no $a_D \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$ with $(a_D, a_C) \in s^{\mathcal{I}_{\mathcal{T},C}}$ for any role name s. Let

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

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 Fact 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 safety 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 exactly the same as in Example 1. ❏

For proving that also violation of condition 3(s2) for non-safety gives rise to coNP-hardness, we introduce some preliminaries including the well-known tree-shaped canonical models for $\mathcal{EL}$. Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and $\mathcal{A}$ a (possibly infinite) ABox. In the construction, we use *extended ABoxes*, i.e., sets of assertions $C(a)$ with C a *potentially compound* concept and $r(a,b)$. 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 $ar_1C_1 \cdots r_kC_k$ 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 $r(a, arC)$ and $C(arC)$ 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 7. *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 $q(\boldsymbol{x})$ and $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$: $\mathcal{T}, \mathcal{A} \models q(\boldsymbol{a})$ iff $\mathcal{J}_{\mathcal{T},\mathcal{A}} \models q[\boldsymbol{a}]$.*

We now construct the tree-shaped canonical model of a TBox and a 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 as

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

Let $\mathcal{J}_{\mathcal{T},C} := \mathcal{J}_{\mathcal{T},\mathcal{A}_C}$. Then the following is straightforward:

Lemma 8. *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 $\mathcal{T} \models \mathsf{tail}(p) \sqsubseteq D$;*

We also require a lemma on the connection of reasoning with concepts and reasoning with 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 [14] (Lemma 22).

Lemma 9. *For all $\mathcal{EL}$-TBoxes $\mathcal{T}$, ABoxes $\mathcal{A}$ and $\mathcal{EL}$-concepts C:*

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

We are in the position now to prove Lemma 3.

Lemma 3 Let $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq D_1$ with $\mathsf{sig}(\mathcal{T}_1, D_0) \cap \mathsf{sig}(\mathcal{T}_2, D_1) \subseteq \Sigma$. Then there exists a Σ-concept F such that $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_0 \sqsubseteq F$ and $\mathcal{T}_1 \cup \mathcal{T}_2 \models F \sqsubseteq D_1$.

Proof. 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 canonical tree-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 1. $\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}_\Sigma \not\models D_1(a_{D_0})$.

To see this, assume that $\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}_\Sigma \models D_1(a_{D_0})$. By Lemma 9, 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 8 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 $D_0 \sqsubseteq D_1$. ❑

We now provide a full proof of Lemma 4 (and, thus, finish the proof of Part 1 of Theorem 2).

Lemma 4 Let $(\mathcal{T}, \Sigma)$ be an $\mathcal{EL}$-TBox with closed predicates such that safety is violated by the inclusion $C \sqsubseteq \exists r.D$ because 3(s2) holds. Then the disjunction property fails and there exists an $\mathcal{EL}$-instance query $C(a)$ such that answering $C(a)$ w.r.t. $(\mathcal{T}, \Sigma)$ is coNP-hard.

Proof. Consider the interpretation $\mathcal{I}_S$ from the the proof of Lemma 6 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 [14] (Lemma 16), there exists $G \in K$ with $\mathcal{T} \models G \sqsubseteq D$.

Introduce copies X^0 and X^1 of any non-Σ-predicate X. Denote by G^0 and G^1 the resulting concept if each non-Σ predicate X in G is replaced by X^0 and, respectively, X^1. Similarly, denote by $\mathcal{T}^0$ and $\mathcal{T}^1$ the TBoxes obtained from $\mathcal{T}$ by replacing all concepts G in $\mathcal{T}$ by G^0 and G^1, respectively.

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 3, 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 1. ❑

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

Lemma 10. *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 7).

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 7 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 9, 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 7, $\mathcal{T}, \mathcal{A} \models \exists r_1.(C_1 \sqcap \cdots \exists r._k.(C_k \sqcap F))$. 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 2 it now suffices to show:

Lemma 11. *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 $(\mathcal{T}, \Sigma)$.*

To show Lemma 11, we first show:

Lemma 12. *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 $\mathsf{sig}(D) \cap \Sigma = \emptyset$.*

Proof. 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. ❑

Now Lemma 11 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).

C Proof of Theorem 3

Let $(\mathcal{T}, \Sigma)$ be a safe DL-Lite-TBox or a safe $\mathcal{EL}$-TBox with closed predicates and let $\vartheta(\boldsymbol{x})$ be a $\mathrm{CQ}^{\mathrm{FO}(\Sigma)}$. 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. the proofs of Theorem 1 and Theorem 2).

Now the case where $\vartheta(\boldsymbol{x})$ is a CQ and $(\mathcal{T}, \Sigma)$ is a DL-Lite-TBox is covered by Theorem 1; and the case where $\vartheta(\boldsymbol{x})$ is a CQ and $(\mathcal{T}, \Sigma)$ is an $\mathcal{EL}$-TBox is covered by Theorem 2. Thus, suppose that $\vartheta(\boldsymbol{x})$ is a $\mathrm{CQ}^{\mathrm{FO}(\Sigma)}$ that is not a CQ. This means that there is some conjunct φ of $\vartheta(\boldsymbol{x})$ 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 $\vartheta(\boldsymbol{x})$; because if this is not the case then we can reorder the conjuncts of $\vartheta(\boldsymbol{x})$ 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 $\vartheta(\boldsymbol{x}) = \varphi$ 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 $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$, we have

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

where $\mathcal{I}_{\mathcal{A}}$ is the interpretation corresponding to $\mathcal{A}$. Thus, suppose that $\vartheta(\boldsymbol{x}) = \exists y_1 \dots \exists y_n(\varphi \wedge \psi_1 \wedge \dots \wedge \psi_m)$, where $n \geq 0$, $m \geq 1$, and $\mathsf{sig}(\psi_i) \cap \Sigma = \emptyset$ for all $i \in \{1, \dots, m\}$. Note that (i) some of the variables in $\{y_1, \dots, y_n\}$ may not occur free in φ and (ii) some others from the same set may not occur free in any one of ψ_i. W.l.o.g. let $\{y_1, \dots, y_k\}$ be the set of variables of type (i), $\{y_{k+1}, \dots, y_j\}$ be the set of variables of type (ii), and $\{y_{j+1}, \dots, y_n\}$ be the remaining set of variables, where $k \leq j \leq n$. We rewrite $\vartheta(\boldsymbol{x})$ to obtain the formula

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

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

Thus, we can assume w.l.o.g. that $\vartheta(\boldsymbol{x})$ is of the form $\exists y_1 \dots \exists y_n(\varphi \wedge \psi)$, where the occurrence of each y_i is free in both φ and ψ, 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 $\boldsymbol{a}$ be a tuple of individual names from $\mathsf{Ind}(\mathcal{A})$ that is of the same length as $\boldsymbol{x}$. Denote by $\varphi(\boldsymbol{a})$ $(\psi(\boldsymbol{a}))$ the formula obtained from φ (respectively ψ) by substituting the occurrence of each free variable from $\boldsymbol{x}$ by the corresponding individual name from $\boldsymbol{a}$. $\varphi(\boldsymbol{a})$ and $\psi(\boldsymbol{a})$ may have more free variables and these free variables are exactly $y_1, \ldots, y_n$ in both of these formulae. Using this fact and the domain-independence of $\varphi(\boldsymbol{a})$, we conclude

$$\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \vartheta(\boldsymbol{a}) \text{ iff } \exists \boldsymbol{b} \subseteq \mathsf{Ind}(\mathcal{A}) \text{ such that } \mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\boldsymbol{a}, \boldsymbol{b}) \wedge \psi(\boldsymbol{a}, \boldsymbol{b}), \quad (1)$$

where $\varphi(\boldsymbol{a}, \boldsymbol{b})$ and $\psi(\boldsymbol{a}, \boldsymbol{b})$ are obtained from $\varphi(\boldsymbol{a})$ and $\psi(\boldsymbol{a})$ respectively by substituting $\boldsymbol{b}$ for $y_1, \ldots, y_n$. Obviously,

$$\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\boldsymbol{a}, \boldsymbol{b}) \wedge \psi(\boldsymbol{a}, \boldsymbol{b}) \text{ iff } \mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\boldsymbol{a}, \boldsymbol{b}) \text{ and } \mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \psi(\boldsymbol{a}, \boldsymbol{b}). \quad (2)$$

Now by the domain-independence of $\varphi(\boldsymbol{a}, \boldsymbol{b})$, $\mathsf{sig}(\varphi) \subseteq \Sigma$, 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

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

So far we have assumed that $\mathcal{T}$ is either a DL-Lite-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).

C.1 DL-Lite

We know by Lemma 2 and [5] that there is some domain-independent first-order query ψ' such that for every satisfiable ABox $\mathcal{A}$ w.r.t. $(\mathcal{T}, \Sigma)$ and every $\boldsymbol{a} \subseteq \mathsf{Ind}(\mathcal{A})$, we have

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

Then by (1), (2), (3), and (4), we obtain

$$\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} \varphi(\boldsymbol{a}, \boldsymbol{b}) \wedge \psi(\boldsymbol{a}, \boldsymbol{b}) \text{ iff } \mathcal{I}_{\mathcal{A}} \models \varphi(\boldsymbol{a}, \boldsymbol{b}) \wedge \psi'(\boldsymbol{a}, \boldsymbol{b}). \quad (5)$$

Obviously,

$$\mathcal{I}_{\mathcal{A}} \models \varphi(\boldsymbol{a}, \boldsymbol{b}) \wedge \psi'(\boldsymbol{a}, \boldsymbol{b}) \text{ iff } \mathcal{I}_{\mathcal{A}} \models \exists y_1, \ldots \exists y_n (\varphi \wedge \psi')[\boldsymbol{a}]. \quad (6)$$

(1), (5), and (6) now imply the desired result for DL-Lite.

C.2 $\mathcal{EL}$

By (1), (2), and (3), we have

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

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