Ontology-Mediated Queries with Closed Predicates

Carsten Lutz[1], **İnanç Seylan**[1], and **Frank Wolter**[2]
[1]University of Bremen
{clu, seylan}@informatik.uni-bremen.de
[2]University of Liverpool, UK
wolter@liverpool.ac.uk

Abstract

In the context of ontology-based data access with description logics (DLs), we study ontology-mediated queries in which selected predicates can be closed (OMQCs). In particular, we contribute to the classification of the data complexity of such queries in several relevant DLs. For the case where only concept names can be closed, we tightly link this question to the complexity of surjective CSPs. When also role names can be closed, we show that a full complexity classification is equivalent to classifying the complexity of all problems in coNP, thus currently out of reach. We also identify a class of OMQCs based on ontologies formulated in DL-Lite$_R$ that are guaranteed to be tractable and even FO-rewritable.

1 Introduction

The aim of ontology-based data access (OBDA) is to facilitate querying of data that is significantly incomplete and heterogeneous. To account for the incompleteness, OBDA formalisms typically adopt the open world assumption (OWA). In some applications, though, there are selected parts of the data for which the closed world assumption (CWA) is more appropriate. As an example, consider a large-scale data integration application where parts of the data are extracted from the web and are significantly incomplete, thus justifying the OWA, while other parts of the data come from curated relational database systems and are known to be complete, thus justifying the CWA. Another example is given in [Lutz *et al.*, 2013], namely querying geo-databases such as OpenStreetMap in which the geo-data is typically assumed to be complete, thus justifying the CWA, while annotations are significantly incomplete and thus require the OWA.

In this paper, we consider OBDA formalisms where the ontology is formulated in a description logic (DL). Several approaches have been proposed to implement a partial CWA in OBDA and in other forms of DL reasoning [Calvanese *et al.*, 2007b; Donini *et al.*, 2002; Grimm and Motik, 2005; Motik and Rosati, 2010; Sengupta *et al.*, 2011]. A particularly simple and natural one is to distinguish between OWA predicates and CWA predicates (a predicate is a concept name or a role name) and to adopt the standard semantics from relational databases for the latter: the interpretation of CWA predicates is fixed to what is explicitly stated in the data while OWA predicates can be interpreted as any extension thereof [Lutz *et al.*, 2013]. This semantics generalizes both ABoxes as used in conventional OBDA (all predicates OWA) and so-called DBoxes (all predicates CWA) [Seylan *et al.*, 2009].

Closing predicates has a strong effect on the complexity of query answering. In this paper, we concentrate on data complexity, see [Ngo *et al.*, 2015] for an analysis of combined complexity in the presence of closed predicates. The (data) complexity of answering conjunctive queries (CQs) becomes coNP-hard already when ontologies are formulated in inexpressive DLs such as DL-Lite and $\mathcal{EL}$ [Franconi *et al.*, 2011] while CQ answering without closed predicates is in AC^0 for DL-Lite and in PTime for $\mathcal{EL}$ [Calvanese *et al.*, 2007a; Artale *et al.*, 2009; Hustadt *et al.*, 2005]. Since intractability comes so quickly, from a user's perspective it is not very helpful to analyze complexity on the level of logics, as in the complexity statements just made; instead, one would like to know whether adopting the CWA results in intractability for the *concrete ontology and query used in an application*. If it does not, there can be considerable benefit in adopting the CWA since it potentially results in additional (that is, more complete) answers and allows to use full first-order (FO) queries for the closed part of the vocabulary (which otherwise leads to undecidability). If adopting the CWA results in intractability, this is important information and the user can decide whether (s)he wants to resort to OWA as an approximation semantics or pay (in terms of complexity) for adopting the (partial) CWA.

Such a *non-uniform analysis* has been carried out in [Lutz and Wolter, 2012] and in [Bienvenu *et al.*, 2014] for classical OBDA (that is, no CWA predicates) with expressive DLs such as $\mathcal{ALC}$. The former reference aims to classify the complexity of ontologies, *quantifying over the actual query*: query answering for an ontology $\mathcal{O}$ is in PTime if every CQ can be answered in PTime w.r.t. $\mathcal{O}$ and it is coNP-hard if there is at least one Boolean CQ that is coNP-hard to answer w.r.t. $\mathcal{O}$. In the latter reference, an even more fine-grained approach is taken where the query is not quantified away. It aims to classify the complexity of *ontology-mediated queries (OMQs)*, that is, triples $(\mathcal{O}, \Sigma_{\mathsf{A}}, q)$ where $\mathcal{O}$ is an ontology, Σ_{A} a data vocabulary, and q an actual query. In both cases, a close connection to the complexity of (fixed-template) constraint satisfaction problems (CSPs) is identified, an active field of research that brings together algebra, graph theory, and logic [Feder and Vardi, 1993; Kun and Szegedy, 2009; Bulatov, 2011]. Such a connection is interesting for at least two reasons. First, it clarifies how difficult it is to attain a full complexity classification of relevant classes of ontologies/OMQs; in fact, there is a large body of literature on classifying the complexity of CSPs that revolves around the Feder-Vardi conjecture which states that every CSP is in PTIME or NP-hard [Feder and Vardi, 1993]. And second, it allows to transfer the technically deep results that have been obtained for CSPs in the last years to the world of OBDA.

For OBDA with closed predicates, the case of quantified queries has been analyzed in [Lutz *et al.*, 2013]. The main finding is that there are transparent and PTIME decidable syntactic conditions that separate the easy cases from the hard cases for ontologies formulated in DL-Lite and in $\mathcal{EL}$ (thus the complexity classification is much easier than for CSPs). However, it is also shown that the PTIME cases are exactly those where adopting the CWA does not result in returning additional answers and thus being able to use FO queries on the closed part of the vocabulary is the only benefit. This suggests that an analysis which quantifies over the queries is still too coarse to be practically useful. In the present paper, we therefore consider a complexity analysis on the level of *ontology-mediated queries with closed predicates (OMQCs)*, which are quadruples $(\mathcal{O}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ where $\mathcal{O}$, Σ_A, and q are as above and $\Sigma_\mathsf{C} \subseteq \Sigma_\mathsf{A}$ is a set of closed predicates.

Our main finding is that while classifying the complexity of OMQs with expressive ontologies corresponds to classifying CSPs, classifying OMQCs is tightly linked to classifying *surjective CSPs*. The latter are defined exactly like standard CSPs (with fixed template) except that homomorphisms are required to be surjective. What might sound like a minor change actually makes complexity analyses dramatically more difficult. In fact, there are concrete surjective CSPs of size 6 whose complexity is not understood [Bodirsky *et al.*, 2012] while there are no such open cases for standard CSPs. Like standard CSPs, the complexity of surjective CSPs is currently subject to significant research activities [Bodirsky *et al.*, 2012; Chen, 2014]. Unlike for standard CSPs, though, we are not aware of dichotomy conjectures for surjective CSPs, this kind of question appears to be wide open.

Our connection to surjective CSPs concerns OMQCs where the ontology is formulated in any DL between DL-Lite$_{\text{core}}$ and $\mathcal{ALCHI}$ or between $\mathcal{EL}$ and $\mathcal{ALCHI}$, where only concept names (unary predicates) can be closed, and where the actual queries are tree-shaped unions of conjunctive queries (tUCQs). We find it remarkable that there is no difference between classifying OMQCs based on extremely simple DLs such as DL-Lite$_{\text{core}}$ and rather expressive ones such as $\mathcal{ALCHI}$. For the case where also role names (binary predicates) can be closed, we show that for every NP Turing machine M, there is an OMQC that is polynomially equivalent to the complement of M's word problem and where the ontology can either be formulated in DL-Lite or in $\mathcal{EL}$ (and queries are tUCQs). In the case of closed role names, there is thus no dichotomy between PTIME and CONP (unless PTIME = NP) and a full complexity classification does thus not appear feasible with today's knowledge in complexity theory.

We start in Sections 2 and 3 with formally introducing our framework and establishing some preliminary results. In Section 4, we identify a large and practically useful class of OMQCs that are tractable and even FO-rewritable; ontologies in these OMQCs are formulated in DL-Lite$_\mathcal{R}$, both concept and role names can be closed, and queries are quantifier-free UCQs. In Section 5, we establish the connection to surjective CSPs for the case where only concept names can be closed (and where quantifiers in the query are allowed) and in Section 6 we establish the connection to Turing machines when also role names can be closed.

Proof details can be found in the appendix which is available at http://informatik.uni-bremen.de/tdki/p.html.

2 Preliminaries

Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be countably infinite sets of *concept*, *role*, and *individual names*. A *DL-Lite concept* is either a concept name or a concept of the form $\exists r$ or $\exists r^-$ with $r \in \mathsf{N_R}$. 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 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 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 might additionally contain role inclusions [Calvanese *et al.*, 2007a; Artale *et al.*, 2009]. As usual in DLs, we use the terms *TBox* and *ontology* interchangeably.

$\mathcal{EL}$ 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 $r \in \mathsf{N_R}$. $\mathcal{ELI}$ concepts extend $\mathcal{EL}$ concepts by adding existential restrictions $\exists r^-.C$, where r^- is an inverse role, and $\mathcal{ALCI}$ further allows negation. For $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}, \mathcal{ALCI}\}$, an *$\mathcal{L}$ concept inclusion* is of the form $C \sqsubseteq D$, where C, D are $\mathcal{L}$ concepts. An *$\mathcal{EL}$ TBox* is a finite set of $\mathcal{EL}$ concept inclusions, and $\mathcal{ELI}$ TBoxes are defined accordingly. An $\mathcal{ALCHI}$ TBox consists of $\mathcal{ALCI}$ concept inclusions and role inclusions. 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 $a, b \in \mathsf{N_I}$. We use $\mathsf{Ind}(\mathcal{A})$ to denote the set of individuals used in the ABox $\mathcal{A}$.

An interpretation $\mathcal{I}$ (defined as usual) *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 adopts the standard names assumption (SNA) which requires that $a^\mathcal{I} = a$ for all $a \in \mathsf{N_I}$ and 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 $\mathcal{A}$. As usual, we write $\mathcal{T} \models r \sqsubseteq s$ if every model of $\mathcal{T}$ satisfies the CI $r \sqsubseteq s$ (which can be checked in polytime).

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. An ABox is a *Σ-ABox* if it only uses predicates from Σ.

In this paper, we combine ontologies (that is, TBoxes) with database queries. A *conjunctive query (CQ)* takes the form $q(\vec{x}) = \exists \vec{y}\, \varphi(\vec{x}, \vec{y})$ where $\varphi(\vec{x}, \vec{y})$ is a conjunction of atoms of the form $A(x)$ and $r(x, y)$ with $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. We call $\vec{x}$ the *answer variables* of $q(\vec{x})$. A *tree CQ (tCQ)* is a CQ that is a tree when viewed as a directed graph (multi-edges disallowed) with the root the only answer variable. An *atomic query (AQ)* is a CQ of the form $A(x)$. A CQ is *Boolean* when it has no answer variables. We use *BtCQ* to refer to Boolean tCQs. A *Boolean AQ (BAQ)* has the form $\exists x\, A(x)$. A *union of conjunctive queries (UCQ)* is a disjunction of CQs. Additional query classes such as tUCQ and BtUCQ are then defined in the obvious way, demanding that every CQ in the UCQ is of the expected form. The *length* $|q|$ of a UCQ q is the number of its variables.

In this paper, the general type of query that we are interested in are *ontology-mediated queries with closed predicates (OMQCs)* that consist of a TBox $\mathcal{T}$, a set Σ_A of predicates that can occur in the ABox, a set of closed predicates $\Sigma_\mathsf{C} \subseteq \Sigma_\mathsf{A}$, and an actual query q (such as a UCQ) that is to be answered. The semantics of such queries is as follows. A model $\mathcal{I}$ of an ABox $\mathcal{A}$ *respects closed predicates* Σ_C if the extension of these predicates agrees with what is explicitly stated in the ABox, that is,

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

Let $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ be an OMQC and $\mathcal{A}$ a Σ_A-ABox. A tuple $\vec{a} \in \mathsf{Ind}(\mathcal{A})$ is a *certain answer to Q on $\mathcal{A}$*, written $\mathcal{T}, \mathcal{A} \models_{c(\Sigma_\mathsf{C})} q(\vec{a})$, if $\mathcal{I} \models q[\vec{a}]$ for all models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respect Σ_C.

Example 1. *Assume an automobile company (say Skoda) wants to monitor the model ranges of international automobile manufacturers. It integrates its company databases storing information about its own products with information about other manufacturers extracted from the web. An ontology is used to provide a unifying vocabulary to be used in queries and to add background knowledge such as*

$$\mathsf{SkodaModel} \sqsubseteq \exists \mathsf{has\_engine}.\mathsf{SkodaEngine}$$
$$\mathsf{TeslaModel} \sqsubseteq \exists \mathsf{has\_engine}.\mathsf{TeslaEngine}$$
$$\mathsf{DieselEngine} \sqsubseteq \mathsf{ICEngine} \qquad \mathsf{PetrolEngine} \sqsubseteq \mathsf{ICEngine}$$

where IC stands for internal combustion. Assume that all data is stored in an ABox, which contains information about Skoda models:

$$\begin{array}{l} \mathsf{SkodaModel}(sm_1), \mathsf{SkodaModel}(sm_2), \ldots \\ \mathsf{SkodaEngine}(se_1), \mathsf{SkodaEngine}(se_2) \ldots \\ \mathsf{has\_engine}(sm_1, se_1), \ldots \\ \mathsf{DieselEngine}(se_1), \mathsf{PetrolEngine}(se_2), \ldots \end{array}$$

and about models by other manufacturers:

$$\begin{array}{l} \mathsf{TeslaModel}(tm_1), \ldots \\ \mathsf{TeslaEngine}(te_1), \mathsf{ElectEngine}(te_1), \ldots \end{array}$$

Skoda is sure that its own models and engines are in the database, therefore the concept names SkodaModel *and* SkodaEngine *are closed. As information about other manufacturers is taken from the web, it is assumed to be incomplete. To illustrate the effect of closing these predicates, consider the following query $q_1(x)$:*

$$\exists y (\mathsf{SkodaModel}(x) \wedge \mathsf{has\_engine}(x,y) \wedge \mathsf{ICEngine}(y)).$$

Assume that sm_{17} is a new Skoda model for which an existing engine will be used, but it is not yet decided which one. Thus the data only contains $\mathsf{SkodaModel}(sm_{17})$*, but no other assertions mentioning sm_{17}. Note that Skoda offers only petrol and diesel engines and that Tesla offers only electric engines which is both reflected in the data (e.g, for every assertion* $\mathsf{TeslaEngine}(te_i)$*, there is an assertion* $\mathsf{ElecEngine}(te_i)$*). Due to the knowledge in the ontology and since* SkodaEngine *is closed, sm_{17} is returned as an answer to q_1. This is not the case without closed predicates and it is in this sense that closed predicates can result in more complete answers. In particular, the query*

$$\exists y (\mathsf{TeslaModel}(x) \wedge \mathsf{has\_engine}(x,y) \wedge \mathsf{ElectEngine}(y))$$

does not return tm_4 if the ABox only contains $\mathsf{TeslaModel}(tm_4)$*, but does not associate tm_4 with any specific engine.*

An *OBDA language* is constituted by a triple $(\mathcal{L}, \Sigma, \mathcal{Q})$ that consists of a TBox language (such as DL-Lite$_\text{core}$, $\mathcal{EL}$, or $\mathcal{ALCHI}$), a set of predicates Σ (such as $\mathsf{N_C} \cup \mathsf{N_R}$ or $\mathsf{N_C}$) and a query language $\mathcal{Q}$ (such as UCQ or CQ). It comprises all OMQCs $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ such that $\mathcal{T} \in \mathcal{L}$, $\Sigma_\mathsf{C} \subseteq \Sigma$, and $q \in \mathcal{Q}$. Examples of OBDA languages considered in this paper include, for example, $(\text{DL-Lite}_\mathcal{R}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BtUCQ})$ and $(\mathcal{ALCHI}, \mathsf{N_C}, \text{BtUCQ})$.

For an ABox $\mathcal{A}$, we denote by $\mathcal{I}_\mathcal{A}$ the interpretation corresponding to $\mathcal{A}$, which satisfies $\Delta^{\mathcal{I}_\mathcal{A}} = \mathsf{Ind}(\mathcal{A})$ and is defined in the obvious way. An OMQC $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ is *FO-rewritable* if there is an FO-query $\varphi(\vec{x})$ such that for all Σ_A-ABoxes $\mathcal{A}$ and all tuples $\vec{a}$ of individuals from $\mathsf{Ind}(\mathcal{A})$, we have $\mathcal{I}_\mathcal{A} \models \varphi(\vec{a})$ iff $\mathcal{T}, \mathcal{A} \models_{c(\Sigma_\mathsf{C})} q(\vec{a})$. Here and throughout the paper, we assume that FO-queries use only atoms of the form $A(x)$, $r(x,y)$, and $x = y$ where A is a concept name and r a role name.

In many applications, it is useful to identify and report inconsistencies of the data with the ontology. An ABox $\mathcal{A}$ is *consistent w.r.t. $\mathcal{T}$ and closed* Σ_C if there is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects Σ_C. We say that *ABox consistency is FO-rewritable for an OMQC* $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ if there is a Boolean FO-query φ such that for all Σ_A-ABoxes $\mathcal{A}$, we have $\mathcal{I}_\mathcal{A} \models \varphi$ iff $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ and closed Σ_C.

3 Basic Results

We establish some basic results that set the stage for the rest of the paper. The first one is a general CONP upper bound (in data complexity) that encompasses all OMQ languages studied in this paper. Note that this bound is not a consequence of results on OBDA querying with nominals [Ortiz *et al.*, 2008] because nominals are part of the TBox and thus their number is a constant while closing a predicate corresponds to considering a set of individuals whose number is bounded by the size of the ABox (the input size). The proof uses a decomposition of countermodels (models that demonstrate query non-entailment) into mosaics and then relies on a guess-and-check algorithm for finding such decompositions.

Theorem 2. *Every OMQC in $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, UCQ)$ is in* CONP.

Our next result concerns the relationship between ABox signatures and closed predicates. In the cases relevant for us, we can assume w.l.o.g. that the ABox signature and the set of closed predicates coincide. This setup was called *DBoxes* in [Seylan *et al.*, 2009; Franconi *et al.*, 2011]. In the remainder of the paper, we are thus free to assume that OMQCs $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ satisfy $\Sigma_\mathsf{A} = \Sigma_\mathsf{C}$ whenever convenient. We denote such OMQCs as a triple $(\mathcal{T}, \Sigma, q)$ meaning that Σ serves both as Σ_A and Σ_C. Two Σ-queries Q_1 and Q_2 are *equivalent* if $Q_1(\mathcal{A}) = Q_2(\mathcal{A})$ for all Σ-ABoxes $\mathcal{A}$. A class $\mathcal{Q}$ of queries is called *canonical* if it is closed under replacing a concept or role atom in a query with an atom of the same kind. All classes of queries considered in this paper are canonical.

Theorem 3. *Let $\mathcal{L} \in \{$DL-Lite$_\mathcal{R}, \mathcal{ALCHI}\}$ and $\mathcal{Q}$ be a canonical class of UCQs, or let $\mathcal{L} \in \{$DL-Lite$_{core}, \mathcal{EL}\}$ and $\mathcal{Q}$ be a canonical class of UCQs closed under forming unions of queries. Then for every OMQC $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ from $(\mathcal{L}, \mathsf{N_C} \cup \mathsf{N_R}, \mathcal{Q})$, one can construct in polynomial time an equivalent query OMQC $Q' = (\mathcal{T}', \Sigma_\mathsf{A}, \Sigma_\mathsf{A}, q')$ with $\mathcal{T}' \in \mathcal{L}$ and $q' \in \mathcal{Q}$.*

As noted in the introduction, full first-order queries can be used for the closed predicates. This simple observation was already made in [Lutz *et al.*, 2013] in a related but slightly different setup, and we repeat it here for the setup considered in the present paper. Let Σ_C be a signature that declares closed predicates and let $q = \exists \vec{y}\, \varphi(\vec{x}, \vec{y})$ be a CQ. An *FO(Σ_C)-extension of* q is a query of the form $\exists \vec{y} \exists \vec{z}\, \varphi(\vec{x}, \vec{y}) \wedge \psi_1(\vec{z}_1), \dots, \psi_n(\vec{z}_n)$ where $\psi_1(\vec{z}_1), \dots, \psi_n(\vec{z}_n)$ are FO-queries with $\mathsf{sig}(\psi_i) \subseteq \Sigma_\mathsf{C}$ and $\vec{z}_1 \cup \cdots \cup \vec{z}_n \subseteq \vec{x} \cup \vec{y} \cup \vec{z}$.

Theorem 4. *If an OMQC $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ is FO-rewritable (in* PTIME*), then every $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q')$ with q' an FO(Σ_C)-extension of q is also FO-rewritable (in* PTIME*).*

4 Quantifier-Free UCQs and FO-Rewritability

The aim of this section is to identify useful OBDA languages whose UCQs are guaranteed to be FO-rewritable. It turns out that FO-rewritability can be achieved by combining lightweight DLs with quantifier-free queries. We use qfUCQ to denote the class of *quantifier-free* UCQs, that is, none of the CQs is allowed to contain a quantified variable.

Our main result is that all OMQCs from the OBDA language (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, qfUCQ) are FO-rewritable under the mild restriction that there is no role inclusion which requires an open role to be contained in a closed one. We believe that this class of queries is potentially relevant for practical applications. In fact, DL-Lite$_\mathcal{R}$ is very popular as an ontology language in large-scale OBDA and, in practice, many queries turn out to be quantifier-free. Note that the query language SPARQL, which is used in many web applications, is closely related to qfUCQs and, in fact, does not admit existential quantification under its standard entailment regimes [Glimm and Krötzsch, 2010]. We remark that our result also covers the OBDA language (DL-Lite$_\mathsf{core}$, $\mathsf{N_C} \cup \mathsf{N_R}$, qfUCQ) without further restrictions.

Theorem 5. *Every OMQC $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ from (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, qfUCQ) such that $\mathcal{T}$ contains no role inclusion of the form $s \sqsubseteq r$ with $s \notin \Sigma_\mathsf{C}$ and $r \in \Sigma_\mathsf{C}$ is FO-rewritable.*

We first show that ABox consistency is FO-rewritable for each of the OMQCs $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ covered by Theorem 5. On inconcistent ABoxes, every query returns all possible answers, but in most practical cases it is more useful to detect and point out the inconsistency instead. We assume w.l.o.g. that $\Sigma_\mathsf{A} = \Sigma_\mathsf{C}$ and denote Σ_A by Σ. Let $\mathsf{sub}(\mathcal{T})$ be the set of all concept names in $\mathcal{T}$, their negations, and all concepts $\exists r, \exists r^-$ such that r is a role name that occurs in $\mathcal{T}$. A *$\mathcal{T}$-type* is a set $t \subseteq \mathsf{sub}(\mathcal{T})$ such that for all $B_1, B_2 \in \mathsf{sub}(\mathcal{T})$:

- if $B_1 \in t$ and $\mathcal{T} \models B_1 \sqsubseteq B_2$, then $B_2 \in t$;
- if $B_1 \in t$ and $\mathcal{T} \models B_1 \sqsubseteq \neg B_2$, then $B_2 \notin t$.

A *$\mathcal{T}$-typing* is a set T of $\mathcal{T}$-types. A *path in* T is a sequence $t, r_1, \dots, r_n$ where $t \in T$, $\exists r_1, \dots, \exists r_n \in \mathsf{sub}(\mathcal{T})$ use no symbols from Σ, $\exists r_1 \in t$ and for $i \in \{1, \dots, n-1\}$, $\mathcal{T} \models \exists r_i^- \sqsubseteq \exists r_{i+1}$ and $r_i^- \neq r_{i+1}$. The path is *Σ-participating* if for all $i \in \{1, \dots, n-1\}$, there is no $B \in \mathsf{sub}(\mathcal{T})$ with $\mathsf{sig}(B) \subseteq \Sigma$ and $\mathcal{T} \models \exists r_i^- \sqsubseteq B$ while there is such a B for $i = n$. A $\mathcal{T}$-typing T is *Σ-realizable* if for every Σ-participating path $t, r_1, \dots, r_n$ in T, there is some $u \in T$ such that $\{B \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models \exists r_n^- \sqsubseteq B\} \subseteq u$.

A $\mathcal{T}$-typing T provides an abstraction of a model of $\mathcal{T}$ and a Σ-ABox $\mathcal{A}$, where T contains the types that are realized by ABox elements. Σ-realizability ensures that we can build from T a model that respects the closed predicates in Σ. To make this more precise, define a *$\mathcal{T}$-decoration of* a Σ-ABox $\mathcal{A}$ to be a mapping f that assigns to each $a \in \mathsf{Ind}(\mathcal{A})$ a $\mathcal{T}$-type $f(a)$ such that $f(a)|_\Sigma = t^a_\mathcal{A}|_\Sigma$ where $t^a_\mathcal{A} = \{B \in \mathsf{sub}(\mathcal{T}) \mid a \in B^{\mathcal{I}_\mathcal{A}}\}$ and $S|_\Sigma$ denotes the restriction of the set S of concept to those member that only use symbols from Σ. For brevity, let $R_\Sigma = \{s \sqsubseteq r \mid \mathcal{T} \models s \sqsubseteq r \text{ and } \mathsf{sig}(s \sqsubseteq r) \subseteq \Sigma\}$.

Lemma 6. *A Σ-ABox $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ and closed Σ iff*

1. *$\mathcal{A}$ has a $\mathcal{T}$-decoration f whose image is a Σ-realizable $\mathcal{T}$-typing and*
2. *$s(a, b) \in \mathcal{A}$ and $s \sqsubseteq r \in R_\Sigma$ implies $r(a, b) \in \mathcal{A}$.*

We now construct the required FO-query. For all role names r and variables x, y, define $\psi_r(x, y) = r(x, y)$ and $\psi_{r^-}(x, y) = r(y, x)$. For all concept names A and roles r, define $\psi_A(x) = A(x)$ and $\psi_{\exists r}(x) = \exists y\, \psi_r(x, y)$. For each $\mathcal{T}$-type t, set

$$\psi_t(x) = \bigwedge_{B \in \mathsf{sub}(\mathcal{T}) \setminus t \text{ with } \mathsf{sig}(B) \subseteq \Sigma} \neg \psi_B(x) \wedge \bigwedge_{B \in t \text{ with } \mathsf{sig}(B) \subseteq \Sigma} \psi_B(x)$$

and for each $\mathcal{T}$-typing $T = \{t_1, \dots, t_n\}$, set

$$\psi_T = \forall x \bigvee_{t \in T} \psi_t(x) \wedge \exists x_1 \cdots \exists x_n (\bigwedge_{i \neq j} x_i \neq x_j \wedge \bigwedge_i \psi_{t_i}(x_i)).$$

Let $\mathcal{R}$ be the set of all Σ-realizable typings and set

$$\Psi_{\mathcal{T},\Sigma} = \bigvee_{T \in \mathcal{R}} \psi_T \wedge \bigwedge_{s \sqsubseteq r \in R_\Sigma} \forall x \forall y (\psi_s(x, y) \rightarrow \psi_r(x, y)).$$

Note that the two conjuncts of $\Psi_{\mathcal{T},\Sigma}$ express exactly Points 1 and 2 of Lemma 6. We have thus shown FO-rewritability of ABox consistency for Q.

Proposition 7. *A Σ-ABox $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ and closed Σ iff $\mathcal{I}_\mathcal{A} \models \Psi_{\mathcal{T},\Sigma}$.*

The next step is to construct an FO-rewriting of Q over ABoxes that are consistent w.r.t. $\mathcal{T}$ and closed Σ. If desired, this rewriting can be combined with the one for ABox consistency given above. Due to space limitations and the slightly intricate details of the construction, we only give a rough intuition and defer the details to the appendix.

Whereas the FO-query $\Psi_{\mathcal{T},\Sigma}$ above is Boolean and identifies ABoxes that have a common model with $\mathcal{T}$, we now aim to construct an FO-formula $\varphi(\vec{x})$ (where $\vec{x}$ are the answer variables of the actual query q from Q) such that for all Σ-ABoxes $\mathcal{A}$ and $\vec{a} \in \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{I}_{\mathcal{A}} \models \varphi[\vec{a}]$ iff there is a common model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ such that there is no homomorphism from some CQ in q to $\mathcal{I}$ that takes $\vec{x}$ to $\vec{a}$. The desired FO-rewriting $\Phi_Q(\vec{x})$ of Q over consistent ABoxes is then simply the negation of $\varphi(\vec{x})$. The construction of $\varphi(\vec{x})$ is based on an extended notion of $\mathcal{T}$-typing called $\mathcal{T}, q$*-typing* that provides an abstraction of a model of $\mathcal{T}$ and a Σ-ABox $\mathcal{A}$ that avoids a homomorphism from $\vec{x}$ to certain individuals $\vec{a}$. This finishes the proof of Theorem 5.

We now show that without the restriction on role inclusions adopted in Theorem 5, OMQCs from $(\text{DL-Lite}_{\mathcal{R}}, \mathsf{N_C} \cup \mathsf{N_R}, \text{qfUCQ})$ are no longer FO-rewritable. In fact, we prove the following, slightly stronger result by reduction from propositional satisfiability.

Theorem 8. *There is a DL-Lite$_{\mathcal{R}}$ TBox $\mathcal{T}$ and set of predicates Σ_C such that ABox consistency w.r.t. $\mathcal{T}$ and closed Σ_C is* NP*-complete.*

We close this section with noting that, for the case of $\mathcal{EL}$ and its extensions, quantifier-free queries are computationally no more well-behaved than unrestricted queries; in fact, all hardness results established in the remainder of this paper for $\mathcal{EL}$ and its extensions can be adapted to the case of quantifier-free queries.

5 Closing Concept Names: Connection to CSP

We consider OBDA languages that only allow to close concept names, but not role names. Unlike in the previous section, queries admit unrestricted existential quantification. Our main contribution here is to establish a close connection between such OBDA languages based on a wide range of DLs and surjective constraint satisfaction problems. This result implies that a full complexity classification of these two problem classes is intimately related. In fact, a full complexity classification of surjective CSPs is a very difficult, ongoing research effort. As pointed out in the introduction, there are even concrete surjective CSPs whose complexity is unknown and, via the established connection, these problems can be used to derive concrete OMQCs whose computational properties are currently not understood.

We start with introducing CSPs. An interpretation $\mathcal{I}$ is a Σ*-interpretation* if it only interprets symbols in Σ, that is, all other symbols are interpreted as empty. Every finite Σ-interpretation $\mathcal{I}$ defines a constraint satisfaction problem $\mathsf{CSP}(\mathcal{I})$ in signature Σ: given a finite Σ-interpretation $\mathcal{I}'$, decide whether there is a homomorphism from $\mathcal{I}'$ to $\mathcal{I}$, i.e., a mapping $h : \Delta^{\mathcal{I}'} \to \Delta^{\mathcal{I}}$ such that $d \in A^{\mathcal{I}'}$ implies $h(d) \in A^{\mathcal{I}}$ and $(d,e) \in r^{\mathcal{I}'}$ implies $(h(d), h(e)) \in r^{\mathcal{I}}$. The problem $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ is the variant of $\mathsf{CSP}(\mathcal{I})$ where we require h to be surjective. Note that we do not consider CSPs with relations of arity larger than two.

We first show that for every problem $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$, there is an OMQC Q from $(\text{DL-Lite}_{\text{core}}, \mathsf{N_C}, \text{BtUCQ})$ that has the same complexity as the complement of $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$, up to polynomial time reductions. Here, the complexity of an OMQC $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ is the complexity to decide, given a Σ_A-ABox $\mathcal{A}$, whether $\mathcal{T}, \mathcal{A} \models_{c(\Sigma_\mathsf{C})} q$.

Consider $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ in signature Σ. We may assume w.l.o.g. that there is at least one Σ-interpretation $\mathcal{J}$ that does *not* homomorphically map to $\mathcal{I}$.[1] Define the OMQC $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ as follows:

$$
\begin{aligned}
\mathcal{T} &= \{A \sqsubseteq \exists \mathsf{val},\ \exists \mathsf{val}^- \sqsubseteq V\} \cup \\
&\quad\ \{A \sqsubseteq \exists \mathsf{aux}_d,\ \exists \mathsf{aux}_d^- \sqsubseteq V \sqcap V_d \mid d \in \Delta^{\mathcal{I}}\} \cup \\
&\quad\ \{A \sqsubseteq \exists \mathsf{force}_d,\ \exists \mathsf{force}_d^- \sqsubseteq A \mid d \in \Delta^{\mathcal{I}}\} \\
\Sigma_\mathsf{C} &= \{A, V\} \cup \{V_d \mid d \in \Delta^{\mathcal{I}}\} \\
\Sigma_\mathsf{A} &= \Sigma \cup \Sigma_\mathsf{C} \\
q &= q_1 \vee q_2 \vee q_3
\end{aligned}
$$

where

$$
\begin{aligned}
q_1 = \bigvee_{d,d' \in \Delta^{\mathcal{I}} \mid d \neq d'} & \exists x \exists y_1 \exists y_2\, A(x) \wedge \mathsf{val}(x,y_1) \wedge \\
& \mathsf{val}(x,y_2) \wedge V_d(y_1) \wedge V_{d'}(y_2) \\
q_2 = \bigvee_{d,d' \in \Delta^{\mathcal{I}}, r \in \Sigma \mid (d,d') \notin r^{\mathcal{I}}} & \exists x \exists y \exists x_1 \exists y_1\, A(x) \wedge A(y) \wedge r(x,y) \wedge \\
& \mathsf{val}(x,x_1) \wedge \mathsf{val}(y,y_1) \wedge \\
& V_d(x_1) \wedge V_{d'}(y_1) \\
q_3 = \bigvee_{d,d' \in \Delta^{\mathcal{I}} \mid d \neq d'} & \exists x \exists y \exists z\, A(x) \wedge \mathsf{force}_d(z,x) \wedge \\
& \mathsf{val}(x,y) \wedge V_{d'}(y).
\end{aligned}
$$

To understand the construction, it is useful to consider the reduction of (the complement of) $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ to $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$. Given a Σ-interpretation $\mathcal{J}$ that is an input to $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$, we construct a Σ_A-ABox $\mathcal{A}$ as an input to $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ as

$$\mathcal{A}_{\mathcal{J}} \cup \{A(d) \mid d \in \Delta^{\mathcal{J}}\} \cup \{V(a_d), V_d(a_d) \mid d \in \Delta^{\mathcal{I}}\}$$

where $\mathcal{A}_{\mathcal{J}}$ is $\mathcal{J}$ viewed as an ABox (with the elements of $\mathcal{J}$ serving as ABox individuals) and where a_d is a fresh individual name for each $d \in \Delta^{\mathcal{I}}$. We show in the appendix that $\mathcal{J} \in \mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ iff $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$. For the "if" direction, we extract a homomorphism h from a model $\mathcal{I}'$ of $\mathcal{T}$ and $\mathcal{A}$ that respects Σ_C and satisfies $\mathcal{I} \not\models q$ by setting $h(d) = e$ when $(d, a_e) \in \mathsf{val}^{\mathcal{I}'}$. The first line of $\mathcal{T}$ and the closing of V thus ensure that $h(d)$ is defined for every $d \in \Delta^{\mathcal{J}}$ and q_1 ensures that the value of $h(d)$ is unique; q_2 ensures that h is a homomorphism and Line 3 of $\mathcal{T}$ together with the closing of A and q_3 guarantees that it is surjective. Line 2 of $\mathcal{T}$ is only needed for the converse reduction to go through, ensuring that we always consider homomorphisms onto $\mathcal{I}$. We say that two decision problems P_1 and P_2 are *polynomially equivalent* if P_1 polynomially reduces to P_2 and vice versa.

[1]Otherwise we can simply use as Q any OMQC $(\mathcal{T}, \Sigma, \Sigma_\mathsf{C}, q)$ such that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$ for all Σ-ABoxes $\mathcal{A}$. Then $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ is exactly the complement problem of Q.

Lemma 9. *The complement of* $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ *is polynomially equivalent to* $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$.

Note that the same reduction works when DL-Lite$_{\text{core}}$ is replaced with $\mathcal{EL}$. For example, the first line of $\mathcal{T}$ then reads $A \sqsubseteq \exists \mathsf{val}.V$. Since all disjuncts of q are tree-shaped, we can view them as $\mathcal{EL}$-concepts, extend $\mathcal{T}$ with $q' \sqsubseteq A_0$ for every disjunct q' of q, and replace q with the BAQ $\exists x\, A_0(x)$. We have thus established the following result.

Theorem 10. *For every* $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$*, there is an OMQC* Q *from* $(\text{DL-Lite}_{core}, \mathsf{N_C}, \text{BtUCQ})$ *such that the complement of* $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ *has the same complexity as* Q*, up to polytime reductions. The same holds for* $(\mathcal{EL}, \mathsf{N_C}, \text{BAQ})$.

We remark that, as can easily be verified by checking the constructions in the proof of Lemma 9, the complement of $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ and Q actually have the same complexity up to *FO reductions* [Immerman, 1999]. This links the complexity of the two problems even closer. For example, if one is complete for LOGSPACE or in AC_0, then so is the other.

We now establish (almost) a converse of Theorem 10 by showing that for every OMQC Q from $(\mathcal{ALCHI}, \mathsf{N_C}, \text{BtUCQ})$, there is a *generalized* surjective CSP that has the same complexity as the complement of Q, up to polytime reductions. A generalized surjective CSP in signature Σ is characterized by a *finite set* Γ of finite Σ-interpretations instead of a single such interpretation, denoted $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$. The problem is to decide, given a Σ-interpretation $\mathcal{I}'$, whether there is a homomorphism from $\mathcal{I}'$ to some interpretation in Γ. Note that, in the non-surjective case, every generalized CSP can be translated into an equivalent non-generalized CSP [Foniok *et al.*, 2008]. In the surjective case, such a translation is not known, so there remains a gap in our analysis. Closing this gap is primarily a (difficult!) CSP question rather than an OBDA question.

Let $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ be an OMQC from $(\mathcal{ALCI}, \mathsf{N_C}, \text{BtUCQ})$. We assume w.l.o.g. that $\mathcal{T}$ contains only a single concept inclusion $\top \sqsubseteq C_\mathcal{T}$ (plus role inclusions) where $C_\mathcal{T}$ uses only the constructors $\neg$, $\sqcap$, and $\exists$, and that q has the form $\exists x\, A_0(x)$ with A_0 occurring in $\mathcal{T}$. The latter is possible since every tCQ q can be expressed as an $\mathcal{ALCI}$ concept C_q in an obvious way, and thus when the BtUCQ is $q = \bigvee_i \exists x_i\, q_i$, then we can add $C_{q_i} \sqsubseteq A_0$ to the TBox with A_0 a fresh concept name and replace q with $\exists x\, A_0(x)$. We use $\mathsf{cl}(\mathcal{T})$ to denote the set of subconcepts of $C_\mathcal{T}$, extended with all concepts $\exists s.C$ such that $\exists r.C$ is a subconcept of $C_\mathcal{T}$ and $r \sqsubseteq s \in \mathcal{T}$, as well as the negations of these concepts. A *Q-type* is a subset $t \subseteq \mathsf{cl}(\mathcal{T})$ such that for some model $\mathcal{I}$ of $\mathcal{T}$ and some $d \in \Delta^\mathcal{I}$, we have $t = \mathsf{tp}_\mathcal{I}(d) := \{C \in \mathsf{cl}(\mathcal{T}) \mid d \in C^\mathcal{I}\}$. Let $\mathsf{TP}(\mathcal{T})$ denote the set of all types for Q. For $t, t' \in \mathsf{TP}(\mathcal{T})$ and a role r, we write $t \rightsquigarrow_r t'$ if for all roles r with $\mathcal{T} \models r \sqsubseteq s$, we have

- $C \in t'$ implies $\exists s.C \in t$, for all $\exists s.C \in \mathsf{cl}(\mathcal{T})$ and
- $C \in t$ implies $\exists s^-.C \in t'$, for all $\exists s^-.C \in \mathsf{cl}(\mathcal{T})$.

A subset $T \subseteq \mathsf{TP}(\mathcal{T})$ is *realizable in a countermodel* if there is a Σ_A-ABox $\mathcal{A}$ and model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C such that $\mathcal{I} \not\models q$ and

$$T = \{\mathsf{tp}_\mathcal{I}(a) \mid a \in \mathsf{Ind}(\mathcal{A})\}.$$

We define the desired surjective generalized CSP by taking one template for each $T \subseteq \mathsf{TP}(\mathcal{T})$ that is realizable in a countermodel. The signature Σ of the CSP comprises the symbols in Σ_A, one concept name $\overline{A}$ for each concept name in Σ_C, and the concept name A_0 from q. We assume w.l.o.g. that there is at least one concept name in Σ_C and at least one concept name $A_{\mathsf{open}} \in \Sigma_\mathsf{A} \setminus \Sigma_\mathsf{C}$.

Each $T \subseteq \mathsf{TP}(\mathcal{T})$ realizable in a countermodel gives raise to a template $\mathcal{I}_T$, defined as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_T} &= T \uplus \{d_A \mid A \in \Sigma_\mathsf{C}\} \\
A^{\mathcal{I}_T} &= \{t \in T \mid A \in t\} \cup \{d_B \mid B \neq A\} \\
\overline{A}^{\mathcal{I}_T} &= \{t \in T \mid A \notin t\} \cup \{d_B \mid B \neq A\} \\
r^{\mathcal{I}_T} &= \{(t, t') \in T \times T \mid t \rightsquigarrow_r t'\} \cup \\
& \{(d, d') \in \Delta^{\mathcal{I}_T} \times \Delta^{\mathcal{I}_T} \mid \{d, d'\} \setminus T \neq \emptyset\}.
\end{aligned}$$

Note that, in $\mathcal{I}_T$ restricted to domain T, $\overline{A}$ is interpreted as the complement of A. At each element d_A, all concept names except A and $\overline{A}$ are true, and these elements are connected to all elements with all roles. Intuitively, we need the concept names $\overline{A}$ to ensure that when an assertion $A(a)$ is missing in an ABox $\mathcal{A}$ with A closed, then a can only be mapped to a template element that does not make A true; this is done by extending $\mathcal{A}$ with $\overline{A}(a)$ and exploiting that $\overline{A}$ is essentially the complement of A in each $\mathcal{I}_T$. The elements d_A are then needed to deal with inputs to the CSP where some point satisfies neither A nor $\overline{A}$. Let Γ be the set of all interpretations $\mathcal{I}_T$ obtained in the described way.

Lemma 11. $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ *is polynomially equivalent to the complement of* $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$.

We have thus established the following result

Theorem 12. *For every OMQC* Q *from* $(\mathcal{ALCHI}, \mathsf{N_C}, \text{BtUCQ})$*, there is a (generalized)* $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$ *in binary signature such that* Q *has the same complexity as the complement of* $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$*, up to polytime reductions.*

Again, the theorem can actually be strengthened to state the same complexity up to FO reductions. Note that the DL $\mathcal{ALCHI}$ used in Theorem 12 is a significant extension of the DLs referred to in Theorem 10 and thus our results apply to a remarkable range of DLs: all DLs between DL-Lite$_{\text{core}}$ and $\mathcal{ALCHI}$ as well as all DLs between $\mathcal{EL}$ and $\mathcal{ALCHI}$.

6 Closing Role Names: TM Equivalence

We generalize the setup from the previous section by allowing also role names to be closed. Our main result is that for every NP Turing machine M, there is an OMQC in $(\text{DL-Lite}_\mathcal{R}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BtUCQ})$ that is polynomially equivalent to the complement of M's word problem, and the same is true for $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BAQ})$. By Ladner's theorem, it follows that there are CONP-intermediate OMQCs in both of the mentioned OBDA languages (unless P = NP) and that a full complexity classification of the queries in these languages is currently far beyond reach.

To establish the above, we utilize a result from [Lutz and Wolter, 2012; Bienvenu *et al.*, 2014] which states that for every NP Turing machine M, there is a monadic disjunctive datalog program of a certain restricted shape that is polynomially equivalent to the complement of M's word problem. It thus suffices to show that for every such datalog program, there is a polynomially equivalent OMQC of the required form. The actual reduction then uses similar ideas as the proof of Theorem 10.

Theorem 13. *For every simple disjunctive datalog program* Π*, there exists an OMQC in* $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \mathit{BAQ})$ *that is polynomially equivalent to* Π*. The same is true for* $(\mathit{DL\text{-}Lite}_{\mathcal{R}}, \mathsf{N_C} \cup \mathsf{N_R}, \mathit{BtUCQ})$*.*

The computational status of $(\text{DL-Lite}_{\text{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BtUCQ})$ remains open. In particular, it is open whether Theorem 13 can be strengthened to this case. We are, however, able to clarify the status of $(\text{DL-Lite}_{\text{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BtCQ})$, where BtUCQs are replaced with BtCQs. In fact, we show that every OMQC in this language is polynomially equivalent to an OMQC that is formulated in the same language but does *not* use closed roles. Via the results in Section 5, OMQCs in $(\text{DL-Lite}_{\text{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BtCQ})$ are thus linked to surjective CSPs.

Theorem 14. *For every OMQC in* $(\mathit{DL\text{-}Lite}_{\mathsf{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \mathit{BtCQ})$ *there exists a polynomially equivalent OMCQ in* $(\mathit{DL\text{-}Lite}_{\mathsf{core}}, \mathsf{N_C}, \mathit{BtCQ})$*.*

Interestingly, Theorem 14 can be extended from BtCQs to BtUCQs for the case of ABoxes in which there is at least one assertion for each closed role name. The unrestricted case of $(\text{DL-Lite}_{\text{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BtUCQ})$, though, remains open.

7 CONCLUSION

Admitting closed predicates in OBDA has significant advantages as it can result in more complete answers and allows to use FO queries for the closed part of the vocabulary. The main results of this paper characterize the computational challenges: for a wide range of DLs, closing concept names corresponds to moving from CSPs to surjective CSPs, while closing role names yields the full computational power of NP. As future work, it would be interesting to exploit closed predicates in OBDA practice. Our results on FO-rewritability of quantifier-free UCQs provide a starting point.

References

[Artale *et al.*, 2009] A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. The DL-Lite family and relations. *JAIR*, 36:1–69, 2009.

[Bienvenu *et al.*, 2014] M. Bienvenu, B. ten Cate, C. Lutz, and F. Wolter. Ontology-based data access: A study through disjunctive datalog, CSP, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33, 2014.

[Bodirsky *et al.*, 2012] M. Bodirsky, J. Kára, and B. Martin. The complexity of surjective homomorphism problems–a survey. *Disc. Appl. Math.*, 160(12):1680–1690, 2012.

[Bulatov, 2011] A.A. Bulatov. On the CSP dichotomy conjecture. In *CSR*, 2011.

[Calvanese *et al.*, 2007a] 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. *J. of Auto. Reas.*, 39(3):385–429, 2007.

[Calvanese *et al.*, 2007b] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. EQL-Lite: Effective first-order query processing in description logics. In *IJCAI*, 2007.

[Chen, 2014] H. Chen. An algebraic hardness criterion for surjective constraint satisfaction. *Algebra universalis*, 72(4):393–401, 2014.

[Donini *et al.*, 2002] F.M. Donini, D. Nardi, and R. Rosati. Description logics of minimal knowledge and negation as failure. *ACM Trans. on Comp. Logic*, 3(2):177–225, 2002.

[Feder and Vardi, 1993] T. Feder and M.Y. Vardi. Monotone monadic SNP and constraint satisfaction. In *STOC*, 1993.

[Foniok *et al.*, 2008] J. Foniok, J. Nesetril, and C. Tardif. Generalised dualities and maximal finite antichains in the homomorphism order of relational structures. *Eur. J. Comb.*, 29(4):881–899, 2008.

[Franconi *et al.*, 2011] E. Franconi, Y. Angélica Ibáñez-García, and İ. Seylan. Query answering with DBoxes is hard. *El. Notes in Theo. Comp. Sci.*, 278:71–84, 2011.

[Glimm and Krötzsch, 2010] B. Glimm and M. Krötzsch. SPARQL beyond subgraph matching. In *ISWC*, 2010.

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

[Hustadt *et al.*, 2005] U. Hustadt, B. Motik, and U. Sattler. Data complexity of reasoning in very expressive description logics. In *IJCAI*, 2005.

[Immerman, 1999] N. Immerman. *Descriptive Complexity*. Springer, 1999.

[Kun and Szegedy, 2009] G. Kun and M. Szegedy. A new line of attack on the dichotomy conjecture. In *STOC*, 2009.

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

[Lutz *et al.*, 2013] C. Lutz, İ. Seylan, and F. Wolter. Ontology-based data access with closed predicates is inherently intractable (sometimes). In *IJCAI*, 2013.

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

[Ngo *et al.*, 2015] N. Ngo, M. Ortiz, and M. Simkus. The combined complexity of reasoning with closed predicates in description logics. In *DL*, 2015.

[Ortiz *et al.*, 2008] M. Ortiz, D. Calvanese, and T. Eiter. Data complexity of query answering in expressive description logics via tableaux. *J. of Auto. Reas.*, 41(1):61–98, 2008.

[Sengupta *et al.*, 2011] K. Sengupta, A.A. Krisnadhi, and P. Hitzler. Local closed world semantics: Grounded circumscription for OWL. In *ISWC*, 2011.

[Seylan *et al.*, 2009] İ. Seylan, E. Franconi, and J. de Bruijn. Effective query rewriting with ontologies over DBoxes. In *IJCAI*, 2009.

Appendix

A Proof of Theorem 2

We provide the proof of Theorem 2 for Boolean UCQs rather than arbitrary UCQs as this makes the exposition more transparent and intuitive. The extension of the proof to arbitrary UCQs is straightforward.

A *forest over an alphabet* S is a prefix-closed set of words over $S^* \setminus \{\varepsilon\}$, where ε denotes the empty word. A *successor of* w in F is a $v \in F$ of the form $v = w \cdot x$, where $x \in S$. For a $k \in \mathbb{N}$, F is called *k-ary*, if for all $w \in F$, we have that the number of successors of w is at most k. A *root of* F is a word in F of length one. A *tree* is a forest that has exactly one root. We do not mention the alphabet of a forest if it is not important.

Definition 15. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is *forest-shaped* if $\Delta^{\mathcal{I}}$ is a forest and for all $(d, e) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and $r \in \mathsf{N_R}$, if $(d, e) \in r^{\mathcal{I}}$, then

- d or e is a root of $\Delta^{\mathcal{I}}$,
- e is a successor of d, or
- d is a successor of e.

In the following proof we use $\mathsf{cl}(\mathcal{T})$ to denote the set of subconcepts of concepts in a TBox $\mathcal{T}$.

Lemma 16. *Let $\mathcal{A}$ be a Σ_{A}-ABox and $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q) \in (\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, BUCQ)$. Then the following are equivalent:*

1. *$\mathcal{T}, \mathcal{A} \models_{c(\Sigma_{\mathsf{C}})} q$*
2. *for all forest-shaped models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respect closed predicates Σ_{C}, if $\Delta^{\mathcal{I}}$ is $|\mathcal{T}|$-ary and its roots are $\mathsf{Ind}(\mathcal{A})$, then we have $\mathcal{I} \models q$.*

Proof. $(1 \Rightarrow 2)$ is trivial.

$(2 \Rightarrow 1)$ The proof is indirect. Suppose $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_{\mathsf{C}})} q$. Then there is some model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C} such that $\mathcal{J} \not\models q$. We construct, by induction, a sequence of pairs $(\mathcal{I}_0, h_0), (\mathcal{I}_1, h_1), \ldots$, where each $\mathcal{I}_i$ is an interpretation and h_i is a homomorphism from $\mathcal{I}_i$ to $\mathcal{J}$ such that $h(d) \in A^{\mathcal{J}}$ implies $d \in A^{\mathcal{I}_i}$ for all $d \in \Delta^{\mathcal{I}_i}$.

For $i = 0$, we define $\mathcal{I}_0$ as the restriction of $\mathcal{J}$ to $\mathsf{Ind}(\mathcal{A})$. Moreover, we set $h_0(a) = a$, for all $a \in \mathsf{Ind}(\mathcal{A})$. Clearly h_0 is as required.

For $i \geq 0$, assume $(\mathcal{I}_i, h_i)$ is given. Let $d \in \Delta^{\mathcal{I}_i}$ and $\exists r.C \in \mathsf{cl}(\mathcal{T})$ such that $d \notin (\exists r.C)^{\mathcal{I}_i}$, $h_i(d) \in (\exists r.C)^{\mathcal{J}}$, and for all roles s with $\mathcal{T} \models r \sqsubseteq s$, we have $\mathsf{sig}(s) \not\subseteq \Sigma_{\mathsf{C}}$. $h_i(d) \in (\exists r.C)^{\mathcal{J}}$ implies that there is some $e \in \Delta^{\mathcal{J}}$ such that $(h_i(d), e) \in r^{\mathcal{J}}$ and $e \in C^{\mathcal{J}}$. Assume first that $e \notin \mathsf{Ind}(\mathcal{A})$. We extend $\mathcal{I}_i$ to $\mathcal{I}_{i+1}$ by adding

- $d \cdot e$ to $\Delta^{\mathcal{I}_i}$;
- the tuple $(d, d \cdot e)$ to $s^{\mathcal{I}_i}$, for every role s such that $\mathcal{T} \models r \sqsubseteq s$;
- the individual $d \cdot e$ to $A^{\mathcal{I}_i}$, for every $A \in \mathsf{cl}(\mathcal{T}) \cap \mathsf{N_C}$ with $e \in A^{\mathcal{J}}$.

Moreover, set $h_{i+1} = h_i \cup \{d \cdot e \mapsto e\}$.

Suppose now that $e = a$ for some $a \in \mathsf{Ind}(\mathcal{A})$. In this case, we extend $\mathcal{I}_i$ to $\mathcal{I}_{i+1}$ by adding the tuple (d, a) to $s^{\mathcal{I}_i}$, for every role s such that $\mathcal{T} \models r \sqsubseteq s$. h_{i+1} is simply equal to h_i.

In both of the cases h_{i+1} has the properties required. Now define the interpretation $\mathcal{I}$ as follows:

- $\Delta^{\mathcal{I}} = \bigcup_{i \geq 0} \Delta^{\mathcal{I}_i}$;
- $P^{\mathcal{I}} = \bigcup_{i \geq 0} P^{\mathcal{I}_i}$, for all $P \in \mathsf{N_C} \cup \mathsf{N_R}$.

Moreover, define $h = \bigcup_{i \geq 0} h_i$.

Claim. $\mathcal{I}$ is a forest-shaped model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C} and such that $\Delta^{\mathcal{I}}$ is $|\mathcal{T}|$-ary and its roots are $\mathsf{Ind}(\mathcal{A})$.

Proof of claim. It is clear that $\mathcal{I}$ is a forest-shaped interpretation with $\Delta^{\mathcal{I}}$ a $|\mathcal{T}|$-ary forest having precisely $\mathsf{Ind}(\mathcal{A})$ as its roots. Thus, in the remainder of the proof, we show that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C}. That $\mathcal{I}$ is a model of $\mathcal{A}$ is an easy consequence of the facts that the restrictions of $\mathcal{I}$ and $\mathcal{J}$ to $\mathsf{Ind}(\mathcal{A})$ are identical and $\mathcal{J}$ is a model of $\mathcal{A}$. That $P^{\mathcal{I}} = \{\vec{a} \mid P(\vec{a}) \in \mathcal{A}\}$, for all $P \in (\mathsf{N_C} \cup \mathsf{N_R}) \cap \Sigma_{\mathsf{C}}$, is an easy consequence of the facts that $\mathcal{J}$ satisfies this property (since it is a model of $\mathcal{A}$ that respects closed predicates Σ_{C}) and $P^{\mathcal{I}} = P^{\mathcal{J}}$ for all $P \in (\mathsf{N_C} \cup \mathsf{N_R}) \cap \Sigma_{\mathsf{C}}$. The latter follows from the construction of $\mathcal{I}$. So it remains to show that $\mathcal{I}$ is a model of $\mathcal{T}$. To this aim, we show $(*)$ for all $d \in \Delta^{\mathcal{I}}$ and $C \in \mathsf{cl}(\mathcal{T})$, $d \in C^{\mathcal{I}}$ iff $h(d) \in C^{\mathcal{J}}$. The proof is by structural induction. The base case follows immediately by the construction of $\mathcal{I}$ and the boolean cases are trivial. Therefore, we only consider the case where $C = \exists r.D$.

Suppose $d \in (\exists r.D)^{\mathcal{I}}$. Then there is some $e \in \Delta^{\mathcal{I}}$ such that $(d, e) \in r^{\mathcal{I}}$ and $e \in D^{\mathcal{I}}$. By the former and the construction of $\mathcal{I}$, we have $(h(d), h(e)) \in r^{\mathcal{J}}$; and by the latter and the induction hypothesis, we have $h(e) \in D^{\mathcal{J}}$. Hence $h(d) \in (\exists r.D)^{\mathcal{J}}$, as required. For the other direction, suppose $h(d) \in (\exists r.D)^{\mathcal{J}}$. Then there is some $e \in \Delta^{\mathcal{J}}$ such that $(h(d), e) \in r^{\mathcal{J}}$ and $e \in D^{\mathcal{J}}$. We distinguish between the cases that there is some role s such that $\mathcal{T} \models r \sqsubseteq s$ and $\mathsf{sig}(s) \subseteq \Sigma_{\mathsf{C}}$ and its complement.

Suppose first the former holds. Since $\mathcal{J}$ satisfies $r \sqsubseteq s$ and $h(d) \in (\exists r.D)^{\mathcal{J}}$, we have $(h(d), e) \in r^{\mathcal{J}} \cap s^{\mathcal{J}}$ and $e \in D^{\mathcal{J}}$, for some $e \in \Delta^{\mathcal{J}}$. By the fact that $s^{\mathcal{J}} = \{(a, b) \mid s(a, b) \in \mathcal{A}\}$, we then obtain $h(d) = a$ and $e = b$, for some $a, b \in \mathsf{Ind}(\mathcal{A})$. By the construction of $\mathcal{I}$, a is the only individual in $\Delta^{\mathcal{I}}$ with $h(a) = a$ and thus, $d = a$. To summarize, we have $d = a$, $e = b$, and $(a, b) \in r^{\mathcal{J}}$. Since the restrictions of $\mathcal{I}$ and $\mathcal{J}$ to $\mathsf{Ind}(\mathcal{A})$ are identical, we obtain $(a, b) \in r^{\mathcal{I}}$. Combining this with the fact that $b \in D^{\mathcal{I}}$ (by the induction hypothesis), we obtain $d \in (\exists r.D)^{\mathcal{I}}$, as required.

Suppose now that there is no role s such that $\mathcal{T} \models r \sqsubseteq s$ and $\mathsf{sig}(s) \subseteq \Sigma_{\mathsf{C}}$. Since $\exists r.D \in \mathsf{cl}(\mathcal{T})$ and $h(d) \in (\exists r.D)^{\mathcal{J}}$, the construction of $\mathcal{I}$ guarantees that there are $e \in \Delta^{\mathcal{J}}$ and $e' \in \Delta^{\mathcal{I}}$ such that $(h(d), e) \in r^{\mathcal{J}}$, $e \in D^{\mathcal{J}}$, $h(e') = e$, and $(d, e') \in r^{\mathcal{I}}$. The induction hypothesis then yields $e' \in D^{\mathcal{I}}$ and thus by $(d, e') \in r^{\mathcal{I}}$, we obtain $d \in (\exists r.D)^{\mathcal{I}}$, as required.

The fact that $\mathcal{J}$ is a model of $\mathcal{T}$ and condition $(*)$ imply that $\mathcal{I}$ is a model of every concept inclusion in $\mathcal{T}$. That $\mathcal{I}$ is a model of every role inclusion in $\mathcal{T}$ is obvious by the construction of $\mathcal{I}$. Hence we conclude that $\mathcal{I}$ is a model of $\mathcal{T}$. $\dashv$

It remains to show that $\mathcal{I} \not\models q$. For a proof by contradiction suppose that $\mathcal{I} \models q$. Then for some disjunct q' of q, there is a match π of q' in $\mathcal{I}$. Moreover, we know that h is a homomorphism from $\mathcal{I}$ to $\mathcal{J}$. But then the function $h \circ \pi'$ is a match of q' in $\mathcal{J}$ and so $\mathcal{J} \models q$, which is a contradiction. ❏

Let $\mathcal{T}$ be an $\mathcal{ALCHI}$-TBox. For an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, let

$$\mathsf{tp}_{\mathcal{I}}(d) = \{C \in \mathsf{cl}(\mathcal{T}) \mid d \in C^{\mathcal{I}}\}.$$

A $\mathcal{T}$*-type* is a set $t \subseteq \mathsf{cl}(\mathcal{T})$ such that for some model $\mathcal{I}$ of $\mathcal{T}$ and some $d \in \Delta^{\mathcal{I}}$, we have $t = \mathsf{tp}_{\mathcal{I}}(d)$. For two $\mathcal{T}$-types t, t' and a role r, we write $t \rightsquigarrow_r t'$ if there is some model $\mathcal{I}$ of $\mathcal{T}$ and $d, e \in \Delta^{\mathcal{I}}$ such that $(d, e) \in r^{\mathcal{I}}$, $t = \mathsf{tp}_{\mathcal{I}}(d)$, and $t' = \mathsf{tp}_{\mathcal{I}}(e)$. By $\mathsf{TP}(\mathcal{T})$, we denote the set of all types for $\mathcal{T}$. Note that these notions are also used in the investigation of the relationship between OMQCs and surjective CSPs in the main paper.

Definition 17. Let $\mathcal{A}$ be a Σ_{A}-ABox and $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q) \in (\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BUCQ})$. A *mosaic for* $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ and $\mathcal{A}$ is a pair $(\mathcal{I}, \tau)$, where $\mathcal{I}$ is a forest-shaped interpretation and $\tau : \Delta^{\mathcal{I}} \to \mathsf{TP}(\mathcal{T})$, satisfying the following properties:

1. $\Delta^{\mathcal{I}} \cap \mathsf{N_I} = \mathsf{Ind}(\mathcal{A})$;
2. $\Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$ is a $|\mathcal{T}|$-ary tree of height at most $|q|$;
3. for all $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, the cardinality of $\{a \in \mathsf{Ind}(\mathcal{A}) \mid (d, a) \in r^{\mathcal{I}}$ for some role $r\}$ is at most $|\mathcal{T}|$;
4. for all $d \in \Delta^{\mathcal{I}}$ and $A \in \mathsf{N_C} \cap \mathsf{cl}(\mathcal{T})$, $d \in A^{\mathcal{I}}$ iff $A \in \tau(d)$;
5. for all $(d, e) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and roles r, if $(d, e) \in r^{\mathcal{I}}$ then $\tau(d) \rightsquigarrow_r \tau(e)$;
6. for all $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$ of depth at most $|q| - 1$, if $\exists r.C \in \tau(d)$, then there is some $e \in \Delta^{\mathcal{I}}$ such that $(d, e) \in r^{\mathcal{I}}$ and $C \in \tau(e)$;
7. $\mathcal{I} \models \mathcal{A}$
8. for all $r \sqsubseteq s \in \mathcal{T}$, $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$;
9. for all $A \in \Sigma_{\mathsf{C}}$ and all A that do not occur in $\mathcal{T}$, $A^{\mathcal{I}} = \{a \mid A(a) \in \mathcal{A}\}$ and for all $r \in \Sigma_{\mathsf{C}}$ and all r that do not occur in $\mathcal{T}$, $r^{\mathcal{I}} = \{(a, b) \mid r(a, b) \in \mathcal{A}\}$.

Two interpretations $\mathcal{I}, \mathcal{J}$ are called *isomorphic* if there is a bijective function $f : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$ such that

- for all $a \in \Delta^{\mathcal{I}} \cap \mathsf{N_I}$, $f(a) = a$;
- for all $d \in \Delta^{\mathcal{I}}$ and $A \in \mathsf{N_C}$, $d \in A^{\mathcal{I}}$ iff $f(d) \in A^{\mathcal{J}}$;
- for all $(d, e) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and $r \in \mathsf{N_C}$, $(d.e) \in r^{\mathcal{I}}$ iff $(f(d), f(e)) \in r^{\mathcal{J}}$.

In this case, f is called an *isomorphism* from $\mathcal{I}$ to $\mathcal{J}$. Two mosaics $(\mathcal{I}, \tau)$ and $(\mathcal{I}', \tau')$ for $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ and $\mathcal{A}$ are *isomorphic* if there is an isomorphism f from $\mathcal{I}$ to $\mathcal{I}'$ such that for all $d \in \Delta^{\mathcal{I}}$, we have $\tau(d) = \tau'(f(d))$.

For a forest F, a $w \in F$, and $n \geq 0$, we denote by $F_{w,n}$ the set of all words $w' \in F$ such that w' begins with w and $|w'| \leq |w| + n$.

Definition 18. A set M of mosaics for $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ and $\mathcal{A}$ is *coherent* if the following conditions are satisfied:

- for all $(\mathcal{I}, \tau), (\mathcal{I}', \tau') \in M$,
$$(\mathcal{I}, \tau)|_{\mathsf{Ind}(\mathcal{A})} = (\mathcal{I}', \tau')|_{\mathsf{Ind}(\mathcal{A})};$$
- for all $(\mathcal{I}, \tau) \in M$, $a \in \mathsf{Ind}(\mathcal{A})$, and $\exists r.C \in \mathsf{cl}(\mathcal{T})$, if $\exists r.C \in \tau(a)$, then there is some $(\mathcal{I}', \tau') \in M$ and $d \in \Delta^{\mathcal{I}'}$ such that $(a, d) \in r^{\mathcal{I}'}$ and $C \in \tau'(d)$, where d is either the root of $\Delta^{\mathcal{I}'} \setminus \mathsf{Ind}(\mathcal{A})$ or $d \in \mathsf{Ind}(\mathcal{A})$;
- for each $(\mathcal{I}, \tau) \in M$ and each $d \in \Delta^{\mathcal{I}}$ that is a successor of the root in $\Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, there is some $(\mathcal{I}', \tau') \in M$ with $e \in \Delta^{\mathcal{I}'}$ the root of $\Delta^{\mathcal{I}'} \setminus \mathsf{Ind}(\mathcal{A})$ such that $(\mathcal{I}, \tau)|_{\Delta^{\mathcal{I}}_{d,|q|-1} \cup \mathsf{Ind}(\mathcal{A})}$ is isomorphic to $(\mathcal{I}', \tau')|_{\Delta^{\mathcal{I}'}_{e,|q|-1} \cup \mathsf{Ind}(\mathcal{A})}$.

We write $M \vdash q$ if $\biguplus_{(\mathcal{I},\tau) \in M} \mathcal{I} \models q$, where here and in what follows $\biguplus$ denotes a disjoint union that only makes the elements that are not in $\mathsf{Ind}(\mathcal{A})$ disjoint.

Lemma 19. *Let $\mathcal{A}$ be a Σ_{A}-ABox and $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q) \in (\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, BUCQ)$. Then the following are equivalent:*

1. *$\mathcal{T}, \mathcal{A} \models_{c(\Sigma_{\mathsf{C}})} q$;*
2. *$M \vdash q$, for all coherent sets M of mosaics for Q and $\mathcal{A}$.*

Proof. $(2 \Rightarrow 1)$ Suppose $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_{\mathsf{C}})} q$. By Lemma 16, there is a forest-shaped model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C} such that $\Delta^{\mathcal{I}}$ is $|\mathcal{T}|$-ary and its roots are $\mathsf{Ind}(\mathcal{A})$, and $\mathcal{I} \not\models q$. For each $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, let $\mathcal{I}_d = \mathcal{I}|_{\Delta^{\mathcal{I}}_{d,|q|} \cup \mathsf{Ind}(\mathcal{A})}$ and $\tau_d = \bigcup_{e \in \Delta^{\mathcal{I}_d}} \{e \mapsto \mathsf{tp}_{\mathcal{I}}(e)\}$. Now set $M = \{(\mathcal{I}_d, \tau_d) \mid d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})\}$ if $\Delta^{\mathcal{I}} \neq \mathsf{Ind}(\mathcal{A})$; and set $M = \{(\mathcal{I}, \tau)\}$ with $\tau = \bigcup_{a \in \mathsf{Ind}(\mathcal{A})} a \mapsto \mathsf{tp}_{\mathcal{I}}(a)$ if $\Delta^{\mathcal{I}} = \mathsf{Ind}(\mathcal{A})$. It is not hard to see that M is a coherent set of mosaics for $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ and $\mathcal{A}$ (to satisfy Condition 9 for mosaics not only for concept names $A \in \Sigma_{\mathsf{C}}$ and role names $r \in \Sigma_{\mathsf{C}}$, we can clearly assume that $A^{\mathcal{I}} = \{a \mid A(a) \in \mathcal{A}\}$ for all A that do not occur in $\mathcal{T}$, and $r^{\mathcal{I}} = \{(a, b) \mid r(a, b) \in \mathcal{A}\}$ for all r that do not occur in $\mathcal{T}$). It remains to show that $M \not\vdash q$. The proof is by contradiction, so suppose that $M \vdash q$. Then $\mathcal{I}' \models q$, where $\mathcal{I}' = \biguplus_{(\mathcal{J},\tau) \in M} \mathcal{J}$. Let π be a match witnessing $\mathcal{I}' \models q$. One can now easily construct a homomorphism g from $\mathcal{I}'$ to $\mathcal{I}$. But then $g \circ \pi$ is a match for q in $\mathcal{I}$. Thus $\mathcal{I} \models q$, and we have obtained a contradiction.

$(1 \Rightarrow 2)$ Suppose there is a coherent set M of mosaics for Q and $\mathcal{A}$ with $M \not\vdash q$. We construct, by induction, a sequence of pairs $(\mathcal{I}_0, \tau_0), (\mathcal{I}_1, , \tau_1), \ldots$, where every $\mathcal{I}_i$ is a forest-shaped interpretation and $\tau_i : \Delta^{\mathcal{I}_i} \to \mathsf{TP}(\mathcal{T})$ such that every individual $d \in \Delta^{\mathcal{I}_i} \setminus \mathsf{Ind}(\mathcal{A})$ of depth $\leq i$ is associated with a mosaic $(\mathcal{I}_d, \tau_d) = (\mathcal{I}_i, \tau_i)|_{\Delta^{\mathcal{I}_i}_{d,|q|} \cup \mathsf{Ind}(\mathcal{A})}$ that is isomorphic to a mosaic in M.

For $i = 0$, let M_0 be the set of all $(\mathcal{J}, \tau) \in M$ such that there are $a \in \mathsf{Ind}(\mathcal{A})$, $d \in \Delta^{\mathcal{J}}$, and $\exists r.C \in \mathsf{cl}(\mathcal{T})$ with $\exists r.C \in \tau(a)$, $C \in \tau(d)$, $(a, d) \in r^{\mathcal{J}}$, and d is either the root of $\Delta^{\mathcal{J}} \setminus \mathsf{Ind}(\mathcal{A})$ or $d \in \mathsf{Ind}(\mathcal{A})$. Define

$$\mathcal{I}_0 = \biguplus_{(\mathcal{J},\tau) \in M_0} \mathcal{J}, \quad \tau_0 = \biguplus_{(\mathcal{J},\tau) \in M_0} \tau$$

It is easy to see that $(\mathcal{I}_0, \tau_0)$ satisfies the conditions above.

For $i > 0$, let $d' \in \Delta^{\mathcal{I}_i} \setminus \mathsf{Ind}(\mathcal{A})$ be of depth i and let d be the unique individual in $\Delta^{\mathcal{I}_i} \setminus \mathsf{Ind}(\mathcal{A})$ of depth $i-1$ such that d' is the successor of d. By the induction hypothesis and coherency of M, there is some $(\mathcal{J}, \tau) \in M$ with $e \in \Delta^{\mathcal{J}}$ the root of $\Delta^{\mathcal{J}} \setminus \mathsf{Ind}(\mathcal{A})$ such that $(\mathcal{I}_d, \tau_d)|_{\Delta^{\mathcal{I}_d}_{d',|q|-1} \cup \mathsf{Ind}(\mathcal{A})}$ is isomorphic to $(\mathcal{J}, \tau)|_{\Delta^{\mathcal{J}}_{e,|q|-1} \cup \mathsf{Ind}(\mathcal{A})}$. W.l.o.g. we assume that $\Delta^{\mathcal{I}_d}_{d',|q|-1} = \Delta^{\mathcal{J}}_{e,|q|-1}$; if this is not the case, we can always rename the individuals in the latter without destroying the isomorphism. Set $(\mathcal{I}_{d'}, \tau_{d'}) = (\mathcal{J}, \tau)$ and assume that the points in $\Delta^{\mathcal{I}_{d'}} \setminus \Delta^{\mathcal{I}_d}_{d',|q|-1}$ are fresh. Set

$$(\mathcal{I}_{i+1}, \tau_{i+1}) = (\mathcal{I}_i, \tau_i) \cup \bigcup_{d' \in \Delta^{\mathcal{I}_i} \setminus \mathsf{Ind}(\mathcal{A}) \text{ of depth } i} (\mathcal{I}_{d'}, \tau_{d'})$$

Now define the interpretation $\mathcal{I}$ as follows:

- $\Delta^{\mathcal{I}} = \bigcup_{i \geq 0} \Delta^{\mathcal{I}_i}$;
- $P^{\mathcal{I}} = \bigcup_{i \geq 0} P^{\mathcal{I}_i}$, for all $P \in \mathsf{N_C} \cup \mathsf{N_R}$.

Claim. $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C.

Proof of claim. The following conditions follow directly from the construction of $\mathcal{I}$ and the conditions on mosaics:

- $\mathcal{I}$ is a model of $\mathcal{A}$;
- $\mathcal{I}$ is a model of every role inclusion in $\mathcal{T}$;
- $P^{\mathcal{I}} = \{\vec{a} \mid P(\vec{a}) \in \mathcal{A}\}$, for all $P \in (\mathsf{N_C} \cup \mathsf{N_R}) \cap \Sigma_\mathsf{C}$

It remains to show that $\mathcal{I}$ is a model of every concept inclusion in $\mathcal{T}$. Define for every $d \in \Delta^{\mathcal{I}}$, a $\mathcal{T}$-type t_d as follows.

- if $d \in \mathsf{Ind}(\mathcal{A})$, then let $t_d = \tau(d)$ for some $(\mathcal{J}, \tau) \in M$;
- if $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, then $t_d = \tau_d(d)$.

To prove the that $\mathcal{I}$ is a model of $\mathcal{T}$ it is now sufficient to show the following: for all $d \in \Delta^{\mathcal{I}}$ and $C \in \mathsf{cl}(\mathcal{T})$, $d \in C^{\mathcal{I}}$ iff $C \in t_d$. The proof is by structural induction.

Let $C = A \in \mathsf{N_C}$. If $d \in \mathsf{Ind}(\mathcal{A})$, let $(\mathcal{J}, \tau)$ be any mosaic in M; and if $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, then let $(\mathcal{J}, \tau) = (\mathcal{I}_d, \tau_d)$. We have (i) $d \in B^{\mathcal{I}}$ iff $d \in B^{\mathcal{J}}$ for all $B \in \mathsf{N_C} \cap \mathsf{cl}(\mathcal{T})$ and (ii) $\tau(d) = t_d$. But then $d \in A^{\mathcal{I}}$ iff $d \in A^{\mathcal{J}}$ (by (i)) iff $A \in \tau(d)$ (by the definition of a mosaic) iff $A \in t_d$ (by (ii)).

The boolean cases follow easily by the induction hypothesis and the fact that t_d is a $\mathcal{T}$-type.

Let $C = \exists r.D$. For the direction from left to right, suppose $d \in (\exists r.D)^{\mathcal{I}}$. Then there is some $e \in \Delta^{\mathcal{I}}$ such that $(d, e) \in r^{\mathcal{I}}$ and $e \in D^{\mathcal{I}}$. If $d, e \in \mathsf{Ind}(\mathcal{A})$, let $(\mathcal{J}, \tau)$ be any mosaic in M; if $d, e \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, let $(\mathcal{J}, \tau) = (\mathcal{I}_{d'}, \tau_{d'})$, where d' is the individual in $\{d, e\}$ that has the smaller depth in $\Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$; otherwise let $(\mathcal{J}, \tau) = (\mathcal{I}_{d'}, \tau_{d'})$, where d' is the only individual in $(\Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})) \cap \{d, e\}$. Observe that $(d, e) \in r^{\mathcal{J}}$, $\tau(d) = t_d$, and $\tau(e) = t_e$. By $(d, e) \in r^{\mathcal{J}}$ and the definition of a mosaic, we obtain $\tau(d) \rightsquigarrow_r \tau(e)$ and by the induction hypothesis and $\tau(e) = t_e$, we obtain $D \in \tau(e)$. But then $\exists r.D \in \tau(d)$ and thus, $\exists r.D \in t_d$, which is what we wanted to show.

For the direction from right to left, suppose $\exists r.D \in t_d$. We distinguish between $d \in \mathsf{Ind}(\mathcal{A})$ or not. For the former case, we find by the coherency of M a $(\mathcal{J}, \tau) \in M$ such that for some $e \in \Delta^{\mathcal{J}}$ we have $(d, e) \in r^{\mathcal{J}}$ and $C \in \tau(e)$; for the latter case, we have by the definition of a mosaic and $|q| \geq 1$ that there is some $e \in \Delta^{\mathcal{I}_d}$ with $(d, e) \in r^{\mathcal{I}_d}$ and $C \in \tau_d(e)$. In both cases, we have by the construction of $\mathcal{I}$ that $(d, e) \in r^{\mathcal{I}}$ and by definition that $C \in t_e$. By the latter, the induction hypothesis yields $e \in C^{\mathcal{I}}$. Hence, $d \in (\exists r.D)^{\mathcal{I}}$, as required.

$\dashv$

It remains to show that $\mathcal{I} \not\models q$. We proceed towards a contradiction, thus, suppose that $\mathcal{I} \models q$. Then for some disjunct q' of q, there is a match π of q' in $\mathcal{I}$. Let $F = \{\pi(x) \mid \pi(x) \notin \mathsf{Ind}(\mathcal{A})\}$. Observe that F is a forest and finite. Let $T_1, \ldots, T_n$ denote the set of all maximal and pairwise disjoint trees in F. Fix an $i \in \{1, \ldots, n\}$. Let d be the root of T_i. By the construction of $\mathcal{I}$, $(\mathcal{I}_d, \tau_d)$ is isomorphic to a $(\mathcal{J}, \tau) \in M$. Denote by f_i this isomorphism, let π_i be the restriction of π to those variables that are mapped to T_i, and let $\pi_{\mathcal{A}}$ be the restriction of π to those variables that are mapped to $\mathsf{Ind}(\mathcal{A})$. Define $\pi'_i = f_i \circ \pi_i$ and then

$$\pi' = \bigcup_{i=1}^{n} \pi_i \cup \pi_{\mathcal{A}}.$$

π' is a match for q' in $\biguplus_{(\mathcal{J},\tau) \in M} \mathcal{J}$ and so we have found a contradiction. ❏

Lemma 20. *Let $\mathcal{A}$ be a Σ_A-ABox and $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q) \in (\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, BUCQ)$. Then, up to isomorphisms, the size of any coherent set M of mosaics for Q and $\mathcal{A}$ is bounded by $(2|\mathcal{A}|)^{|\mathcal{T}|^{|q|+3}}$.*

Proof. The bound follows from Conditions 1, 2, 3, and 9 on mosaics and the first condition on coherent sets of mosaics. Note, in particular, that by the first condition on coherent sets M of mosaics the restriction to $\mathsf{Ind}(\mathcal{A})$ coincides for all mosaics in M and that by Condition 3 on mosaics for any $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$ the number of distinct $a \in \mathsf{Ind}(\mathcal{A})$ with $(d, a) \in r^{\mathcal{I}}$ for some role r is bounded by $|\mathcal{T}|$ for any mosaic $(\mathcal{I}, \tau)$. ❏

We are now in the position to prove Theorem 2 for Boolean UCQs.

Theorem 2. Every OMQC in $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BUCQ})$ is in coNP.

Proof. Fix an OMQC $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ in $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BUCQ})$. We show that given a Σ_A-ABox $\mathcal{A}$, deciding $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$ is in NP. Assume $\mathcal{A}$ is given. By Lemmas 19 and 20, $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$ iff there exists a coherent set M of mosaics for Q and $\mathcal{A}$ such that $|M| \leq (2|\mathcal{A}|)^{|\mathcal{T}|^{|q|+3}}$ and $M \not\models q$. Thus, it is sufficient to show that it can be decided in polynomial time in the size $|\mathcal{A}|$ of $\mathcal{A}$ whether M is a coherent set of mosaics for Q and $\mathcal{A}$ and whether $M \not\models q$. The first condition is clear. For the second condition, observe that $\mathcal{J} = \biguplus_{(\mathcal{I},\tau) \in M} \mathcal{I}$ can be constructed in polynomial time (in $|\mathcal{A}|$) and that checking if $\mathcal{J} \models q$ is again in polynomial time (in $|\mathcal{A}|$). ❏

B Other Proofs for Section 3

We split the proof of Theorem 3 into two parts, its claim for OMQCs with role inclusions and its claim for OMQCs without role inclusions.

Proposition 21. *Let $\mathcal{L} \in \{DL\text{-}Lite_{\mathcal{R}}, \mathcal{ALCHI}\}$ and $\mathcal{Q}$ be any canonical class of UCQs. For every OMQC $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ from $(\mathcal{L}, \mathsf{N_C} \cup \mathsf{N_R}, \mathcal{Q})$, one can construct in polynomial time an equivalent query OMQC $Q' = (\mathcal{T}', \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{A}}, q')$ with $\mathcal{T} \in \mathcal{L}$ and $q' \in \mathcal{Q}$.*

Proof. Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ be an OMQC with $\mathcal{T} \in \mathcal{L}$ and $q \in \mathcal{Q}$. For every $X \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$, we take a fresh predicate X' of the same sort (if X is a concept name, then X' is a concept name, and if X is a role name, then X' is a role name). Let $\mathcal{T}'$ be the resulting TBox when all $X \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ are replaced by X' and the (concept or role) inclusion $X \sqsubseteq X'$ is added, for each $X \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$. Denote by q' the resulting query when every $X \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ in q is replaced by X'. We show that $Q' = (\mathcal{T}', \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{A}}, q')$ is equivalent to $\mathcal{Q}$.

($\Rightarrow$) Let $\mathcal{A}$ be a Σ_{A}-ABox with $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_{\mathsf{C}})} q(\vec{a})$. Then there is a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C} and such that $\mathcal{I} \not\models q(\vec{a})$. Define an interpretation $\mathcal{I}'$ by setting

$$\begin{aligned}
\Delta^{\mathcal{I}'} &= \Delta^{\mathcal{I}}\\
A^{\mathcal{I}'} &= \{a \mid A(a) \in \mathcal{A}\} \text{ for all } A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
r^{\mathcal{I}'} &= \{(a,b) \mid r(a,b) \in \mathcal{A}\} \text{ for all } r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
A'^{\mathcal{I}'} &= A^{\mathcal{I}} \text{ for all } A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
r'^{\mathcal{I}'} &= r^{\mathcal{I}} \text{ for all } r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}
\end{aligned}$$

and leaving the interpretation of the remaining symbols unchanged. It can be verified that $\mathcal{I}'$ is a model of $\mathcal{T}'$ and $\mathcal{A}$ that respects closed predicates Σ_{A} and such that $\mathcal{I}' \not\models q'(\vec{a})$.

($\Leftarrow$) Let $\mathcal{A}$ be a Σ_{A}-ABox such that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_{\mathsf{A}})} q'(\vec{a})$. Let $\mathcal{I}'$ be a model of $\mathcal{T}'$ and $\mathcal{A}$ that respects closed predicates Σ_{A} and such that $\mathcal{I}' \not\models q'(\vec{a})$. Define an interpretation $\mathcal{I}$ by setting

$$\begin{aligned}
\Delta^{\mathcal{I}} &= \Delta^{\mathcal{I}'}\\
A^{\mathcal{I}} &= A'^{\mathcal{I}'} \text{ for all concept names } A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
r^{\mathcal{I}} &= r'^{\mathcal{I}'} \text{ for all roles names } r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}
\end{aligned}$$

and leaving the interpretation of the remaining symbols unchanged. It is readily checked that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C} and such that $\mathcal{I} \not\models q(\vec{a})$. ❏

Proposition 22. *Let $\mathcal{L} \in \{DL\text{-}Lite_{core}, \mathcal{EL}\}$ and $\mathcal{Q}$ be any canonical class of UCQs preserved under forming unions of queries. For every OMQC $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ from $(\mathcal{L}, \mathsf{N_C} \cup \mathsf{N_R}, \mathcal{Q})$, one can construct in polynomial time an equivalent query OMQC $Q' = (\mathcal{T}', \Sigma_{\mathsf{A}}, q')$ with $\mathcal{T} \in \mathcal{L}$ and $q' \in \mathcal{Q}$.*

Proof. Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ be an OMQC with $\mathcal{T} \in \mathcal{L}$ and $q \in \mathcal{Q}$. As in the proof above, we take for every $X \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ a fresh predicate X' of the same arity as X. Concept names $A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ are dealt with in the same way as before: replace every concept name $A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ in $\mathcal{T}$ by A' and add the inclusions $A \sqsubseteq A'$, A a concept name in $\Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ to $\mathcal{T}$. Call the resulting TBox $\mathcal{T}_0$. Next replace in $\mathcal{T}_0$ in every inclusion $C \sqsubseteq D$ all occurences of roles r in D from $\Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ by r'. Call the resulting TBox $\mathcal{T}_1$. Finally, $\mathcal{T}'$ results from $\mathcal{T}_1$ by replacing in any inclusion $C \sqsubseteq D \in \mathcal{T}_1$ every subset of the set of occurences of roles $r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ in C by r'. For example, $\exists r.A \sqcap \exists r.B \sqsubseteq E$ gives rise to the inclusions $\exists r.A \sqcap \exists r'.B \sqsubseteq E$, $\exists r'.A \sqcap \exists r.B \sqsubseteq E$, $\exists r'.A \sqcap \exists r'.B \sqsubseteq E$. Let q' be the query that is obtained from q by taking the disjunction over all queries that result from q when some subset of the set of occurrences of role names $r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ is replaced by r' and all occurences of concepts names $A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ are replaced by A'.

We show that $Q' = (\mathcal{T}', \Sigma_{\mathsf{A}}, q')$ is equivalent to $\mathcal{Q}$.

($\Rightarrow$) The proof is the same as above. Assume a Σ_{A}-ABox $\mathcal{A}$ is given and that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_{\mathsf{C}})} q(\vec{a})$. Let $\mathcal{I}$ be a model of $\mathcal{A}$ and $\mathcal{T}$ that respects closed predicates Σ_{C} such that $\mathcal{I} \not\models q(\vec{a})$. Define an interpretation $\mathcal{I}'$ by setting

$$\begin{aligned}
\Delta^{\mathcal{I}'} &= \Delta^{\mathcal{I}}\\
A^{\mathcal{I}'} &= \{a \mid A(a) \in \mathcal{A}\} \text{ for all } A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
r^{\mathcal{I}'} &= \{(a,b) \mid r(a,b) \in \mathcal{A}\} \text{ for all } r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
A'^{\mathcal{I}'} &= A^{\mathcal{I}} \text{ for all } A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
r'^{\mathcal{I}'} &= r^{\mathcal{I}} \text{ for all } r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}
\end{aligned}$$

and leaving the interpretation of the remaining symbols unchanged. Using monotonicity of $\mathcal{EL}$ and DL-Lite concepts (i.e., if $\mathcal{I}$ and $\mathcal{J}$ are interpretations such that $X^{\mathcal{I}} \subseteq X^{\mathcal{J}}$ for all symbols X, then $C^{\mathcal{I}} \subseteq C^{\mathcal{J}}$ for all concepts C) it is readily checked that $\mathcal{I}'$ is a model of $\mathcal{A}$ and and $\mathcal{T}'$ that respects closed predicates Σ_{A} such that $\mathcal{I}' \not\models q'(\vec{a})$.

($\Leftarrow$) Assume a Σ_{A}-ABox $\mathcal{A}$ is given and that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_{\mathsf{A}})} q'(\vec{a})$. Let $\mathcal{I}'$ be a model of $\mathcal{A}$ and $\mathcal{T}'$ that respects closed predicates Σ_{A} such that $\mathcal{I}' \not\models q'(\vec{a})$. Define an interpretation $\mathcal{I}$ by setting $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$,

$$\begin{aligned}
\Delta^{\mathcal{I}} &= \Delta^{\mathcal{I}'}\\
A^{\mathcal{I}} &= A'^{\mathcal{I}'} \text{ for all concept names } A \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}\\
r^{\mathcal{I}} &= r'^{\mathcal{I}'} \cup r^{\mathcal{I}} \text{ for all roles names } r \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}
\end{aligned}$$

and leaving the interpretation of the remaining symbols unchanged. It is readily checked that $\mathcal{I}$ is a model of $\mathcal{A}$ and $\mathcal{T}$ that respects closed predicates Σ_{C} such that $\mathcal{I} \not\models q(\vec{a})$. ❏

Theorem 4. If an OMQC $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ is FO-rewritable (resp. in PTIME), then every $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q')$ with q' an FO(Σ_{C})-extension of q is also FO-rewritable (resp. in PTIME).

Proof. Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ be an FO-rewritable OMQC and let q' be the FO(Σ_{C})-extension $\exists \vec{y} \exists \vec{z}\, \varphi(\vec{x}, \vec{y}) \wedge \psi_1(\vec{z}_1), \ldots, \psi_n(\vec{z}_n)$ of q. It is readily checked that if φ_q is an FO-rewriting of Q, then $\varphi_q \wedge \exists \vec{z}\, \psi_1(\vec{z}_1), \ldots, \psi_n(\vec{z}_n)$ is an FO-rewriting of $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q')$. Likewise, Q being in PTIME clearly implies that $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q')$ is in PTIME: given an ABox $\mathcal{A}$, one can first compute the answers for Q and for each FO-query $\psi_i(\vec{z}_i)$ in PTIME and then combine them together into the answers for q' in a straightforward way. ❏

C Proofs for Section 4

Lemma 6. A Σ-Box $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ with closed Σ iff

1. $\mathcal{A}$ has a $\mathcal{T}$-decoration f whose image is a Σ-realizable $\mathcal{T}$-typing and
2. $s(a,b) \in \mathcal{A}$, $\mathsf{sig}(s \sqsubseteq r) \subseteq \Sigma$, and $\mathcal{T} \models s \sqsubseteq r$ implies $r(a,b) \in \mathcal{A}$.

Proof. ($\Rightarrow$) Let $\mathcal{I}$ be a model of $\mathcal{A}$ and $\mathcal{T}$ that respects closed predicates Σ. For each $d \in \Delta^{\mathcal{I}}$, let $t_{\mathcal{I}}^{d} = \{B \in \mathsf{sub}(\mathcal{T}) \mid d \in B^{\mathcal{I}}\}$ and let $T_{\mathcal{I}} = \{t_{\mathcal{I}}^{a} \mid a \in \mathsf{Ind}(\mathcal{A})\}$. Since $\mathcal{I}$ is a model of $\mathcal{T}$, $T_{\mathcal{I}}$ is a $\mathcal{T}$-typing. We next show that it is Σ-realizable. Let $t_{\mathcal{I}}^{a}, r_1, \ldots, r_n$ be a Σ-participating path in $T_{\mathcal{I}}$. Using $\mathcal{I}$, we find a mapping $g : \{0, \ldots, n\} \to \Delta^{\mathcal{I}}$ such that $g(0) = a$ and for each $i \in \{1, \ldots, n\}$, we have

1. $(g(i-1), g(i)) \in {r_i}^{\mathcal{I}}$,
2. $g(i) \in B^{\mathcal{I}}$ for all $B \in \mathsf{sub}(\mathcal{T})$ with $\mathcal{T} \models \exists r_i^{-} \sqsubseteq B$.

By definition of Σ-participating paths, there is some $B^\star \in \mathsf{sub}(\mathcal{T})$ with $\mathsf{sig}(B^\star) \subseteq \Sigma$ such that $\mathcal{T} \models \exists r_n^{-} \sqsubseteq B^\star$. By Point 2, we obtain $g(n) \in B^{\star\mathcal{I}}$. Since $\mathcal{I}$ is a model of $\mathcal{A}$ and $\mathcal{T}$ that respects closed predicates Σ, we have $g(n) = b$ for some $b \in \mathsf{Ind}(\mathcal{A})$. By Point 2, $\mathcal{T} \models \exists r_n^{-} \sqsubseteq B$ implies $B \in t_{\mathcal{I}}^{b}$ for any $B \in \mathsf{sub}(\mathcal{T})$. Thus, $T_{\mathcal{I}}$ is Σ-realizable. Let $f(a) = t_{\mathcal{I}}^{a}$ for all $a \in \mathsf{Ind}(\mathcal{A})$. It is clear that f is a $\mathcal{T}$-decoration of $\mathcal{A}$. The image of f is $T_{\mathcal{I}}$, thus a Σ-realizable $\mathcal{T}$-typing. Hence we conclude that $\mathcal{A}$ satisfies Point 1 from Lemma 6. Point 2 holds by the fact that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ.

($\Leftarrow$) Suppose that $\mathcal{A}$ satisfies Points 1 and 2 from Lemma 6 and let f be a $\mathcal{T}$-decoration of $\mathcal{A}$ whose image T is a Σ-realizable $\mathcal{T}$-typing. Our goal is to construct a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ as the limit of a sequence of interpretations $\mathcal{I}_0, \mathcal{I}_1, \ldots$. The domains of these interpretations consist of the individuals from $\mathsf{Ind}(\mathcal{A})$ and of paths in T that are not Σ-participating. The construction will ensure that for all i, we have

1 for all $a \in \mathsf{Ind}(\mathcal{A})$, we have $t_{\mathcal{I}_i}^{a} \subseteq f(a)$;

2 for all $p = t, r_1, \ldots, r_n \in \Delta^{\mathcal{I}_i}$, we have $t_{\mathcal{I}_i}^{p} \subseteq \{B \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models \exists r_n^{-} \sqsubseteq B\}$.

Define $\mathcal{I}_0 = (\Delta^{\mathcal{I}_0}, \cdot^{\mathcal{I}_0})$ where

$$\begin{aligned}
\Delta^{\mathcal{I}_0} &= \mathsf{Ind}(\mathcal{A}) \\
r^{\mathcal{I}_0} &= \{(a,b) \mid s(a,b) \in \mathcal{A} \text{ and } \mathcal{T} \models s \sqsubseteq r\} \\
A^{\mathcal{I}_0} &= \{a \mid A \in f(a)\}
\end{aligned}$$

To construct $\mathcal{I}_{i+1}$ from $\mathcal{I}_i$, choose $d \in \Delta^{\mathcal{I}_i}$ and $\exists s \in \mathsf{sub}(\mathcal{T})$ such that $\mathsf{sig}(s) \cap \Sigma = \emptyset$, $\mathcal{T} \models \bigsqcap t_{\mathcal{I}_i}^{d} \sqsubseteq \exists s$ and there is no $(d,e) \in s^{\mathcal{I}_i}$. Let $q = f(a), s$ if $d = a \in \mathsf{Ind}(\mathcal{A})$ and $q = d, s$ otherwise. Using Conditions 1 and 2, it is easy to verify that q is a path in T. If q is not Σ-participating, then define $\mathcal{I}_{i+1}$ as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_{i+1}} &= \Delta^{\mathcal{I}_i} \uplus \{q\} \\
r^{\mathcal{I}_{i+1}} &= \begin{cases} r^{\mathcal{I}_i} \cup \{(d,q)\} & \text{if } \mathcal{T} \models s \sqsubseteq r \\ r^{\mathcal{I}_i} & \text{otherwise} \end{cases} \\
A^{\mathcal{I}_{i+1}} &= \begin{cases} A^{\mathcal{I}_i} \cup \{q\} & \text{if } \mathcal{T} \models \exists s^{-} \sqsubseteq A \\ A^{\mathcal{I}_i} & \text{otherwise.} \end{cases}
\end{aligned}$$

If q is Σ-participating, then by the fact that T is Σ-realizable, there is some $t \in T$ such that $\{B \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models \exists s^{-} \sqsubseteq B\} \subseteq t$. We find a $b \in \mathsf{Ind}(\mathcal{A})$ with $t = f(b)$. Define $\mathcal{I}_{i+1}$ as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_{i+1}} &= \Delta^{\mathcal{I}_i} \\
r^{\mathcal{I}_{i+1}} &= \begin{cases} r^{\mathcal{I}_i} \cup \{(d,b)\} & \text{if } \mathcal{T} \models s \sqsubseteq r \\ r^{\mathcal{I}_i} & \text{otherwise} \end{cases} \\
A^{\mathcal{I}_{i+1}} &= A^{\mathcal{I}_i}.
\end{aligned}$$

Assume that the choice of $d \in \Delta^{\mathcal{I}_i}$ and $\exists s \in \mathsf{sub}(\mathcal{T})$ is fair so that every possible combination of d and $\exists s$ is eventually chosen. Let $\mathcal{I} = \bigcup_{i \geq 0} \mathcal{I}_i$ be the limit of $\mathcal{I}_0, \mathcal{I}_1, \ldots$ (defined in the obvious way). We claim that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ. By definition of $\mathcal{I}_0$ and of $\mathcal{T}$-decorations, it is straightforward to see that $\mathcal{I} \models \mathcal{A}$. Moreover, the role inclusions in $\mathcal{T}$ are clearly satisfied. To show that the concept inclusions are satisfied as well, it is straightforward to first establish the following strengthenings of Conditions 1 and 2 above (details omitted):

$1'$ for all $a \in \mathsf{Ind}(\mathcal{A})$, we have $t_{\mathcal{I}}^{a} = f(a)$;

$2'$ for all $p = t, r_1 \ldots, r_n \in \Delta^{\mathcal{I}}$, we have $t_{\mathcal{I}_i}^{p} = \{B \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models \exists r_n^{-} \sqsubseteq B\}$.

Let $a \in \mathsf{Ind}(\mathcal{A})$, $a \in {B_1}^{\mathcal{I}}$, and $B_1 \sqsubseteq B_2 \in \mathcal{T}$ (or $B_1 \sqsubseteq \neg B_2 \in \mathcal{T}$). Then by Condition $1'$ and since $f(a)$ is a $\mathcal{T}$-type, we have $a \in {B_2}^{\mathcal{I}}$ (resp. $a \notin {B_2}^{\mathcal{I}}$). Now let $d = t, r_1, \ldots, r_n$ be a path. First suppose $d \in {B_1}^{\mathcal{I}}$ and $B_1 \sqsubseteq B_2 \in \mathcal{T}$. By Condition $2'$, we conclude that $\mathcal{T} \models \exists r_n^{-} \sqsubseteq B_1$. Since $B_1 \sqsubseteq B_2 \in \mathcal{T}$, it follows that $\mathcal{T} \models \exists r_n^{-} \sqsubseteq B_2$ and thus again by the property above, $d \in {B_2}^{\mathcal{I}}$. Finally, suppose $d \in {B_1}^{\mathcal{I}}$ and $B_1 \sqsubseteq \neg B_2 \in \mathcal{T}$. By Condition $2'$ and $B_1 \sqsubseteq \neg B_2 \in \mathcal{T}$, we conclude $\mathcal{T} \models \exists r_n^{-} \sqsubseteq \neg B_2$. For a proof by contradiction assume that $d \in {B_2}^{\mathcal{I}}$ and thus $\mathcal{T} \models \exists r_n^{-} \sqsubseteq B_2$ and we already have $\mathcal{T} \models \exists r_n^{-} \sqsubseteq \neg B_2$. Hence $\mathcal{T} \models \exists r_n^{-} \sqsubseteq \bot$. But then $\mathcal{T} \models \exists r_n \sqsubseteq \bot$. It follows that $\mathcal{T} \models \exists r_1 \sqsubseteq \bot$. This implies in particular $\mathcal{T} \models \exists r_1 \sqsubseteq \exists r_1$ and $\mathcal{T} \models \exists r_1 \sqsubseteq \neg \exists r_1$. By definition we have $\exists r_1 \in f(a)$ and by $\mathcal{T} \models \exists r_1 \sqsubseteq \neg \exists r_1$ and the fact $f(a)$ is a $\mathcal{T}$-type, we obtain $\exists r_1 \notin f(a)$, i.e., a contradiction. Hence $d \notin {B_2}^{\mathcal{I}}$ which finishes the proof that $\mathcal{I} \models \mathcal{T}$.

What remains to be shown are the following properties:

- for all $A \in \Sigma$, $A^{\mathcal{I}} = \{a \mid A(a) \in \mathcal{A}\}$;
- for all $r \in \Sigma$, $r^{\mathcal{I}} = \{(a,b) \mid r(a,b) \in \mathcal{A}\}$.

We show for each $i \geq 0$ that $\mathcal{I}_i$ satisfies the properties above.

Suppose $i = 0$. First let $A(a) \in \mathcal{A}$ with $A \in \Sigma$. Then $a \in A^{\mathcal{I}_0}$ by definition of $\mathcal{I}_0$. For the other direction, let $a \in A^{\mathcal{I}_0}$ for an $A \in \Sigma$. Then $A \in f(a)$. The definition of $\mathcal{T}$-decorations yields $A \in t_{\mathcal{A}}^{a}$, and thus $A(a) \in \mathcal{A}$. Now let $r(a,b) \in \mathcal{A}$ with $r \in \Sigma$. Then $(a,b) \in r^{\mathcal{I}_0}$ by definition of $\mathcal{I}_0$. For the other direction, let $(a,b) \in r^{\mathcal{I}_0}$ for some $r \in \Sigma$. Then there is some role s such that $s(a,b) \in \mathcal{A}$ and $\mathcal{T} \models s \sqsubseteq r$. By the adopted restriction on the allowed role inclusions, it follows that $\mathsf{sig}(s) \subseteq \Sigma$. This yields $r(a,b) \in \mathcal{A}$ since $\mathcal{A}$ satisfies Point 2 of Lemma 6.

For $i > 0$, we show that the extension of Σ-predicates is not modified when constructing $\mathcal{I}_{i+1}$ from $\mathcal{I}_i$. Indeed, assume that $\mathcal{I}_{i+1}$ was obtained from $\mathcal{I}_i$ by choosing $d \in \Delta^{\mathcal{I}_i}$ and $\exists s \in \mathsf{sub}(\mathcal{T})$ and let $q = f(a), s$ if $d = a \in \mathsf{Ind}(\mathcal{A})$ and $q = d, s$ otherwise. Then $\mathsf{sig}(s) \cap \Sigma = \emptyset$ and by the restriction on role inclusions, $\mathsf{sig}(r) \cap \Sigma = \emptyset$ for any role r with $\mathcal{T} \models s \sqsubseteq r$. Consequently, none of the role names modified in the construction of $\mathcal{I}_{i+1}$ is from Σ (no matter whether q is Σ-participating or not). In the case where q is Σ-participating, there is nothing else to show. If q is not Σ-participating, then each concept name A with $\mathcal{T} \models \exists s^- \sqsubseteq A$ is not from Σ. Thus also none of the concept names modified in the construction of $\mathcal{I}_{i+1}$ is from Σ. ❑

We now complete the proof of Theorem 5. We show how to construct an FO-rewriting of a Q over ABoxes that are consistent w.r.t. $\mathcal{T}$ and closed Σ. Let $q = \bigvee_{i \in I} q_i$ with answer variables $\vec{x} = x_1, \ldots, x_n$. A *$\mathcal{T}, q$-typing* T is a quadruple $(\sim, f_0, \Gamma, \Delta)$ where

- $\sim$ is an equivalence relation on $\{x_1, \ldots, x_n\}$;
- f_0 is a function that assigns a $\mathcal{T}$-type $f_0(x_i)$ to each x_i, $1 \leq i \leq n$, such that $f_0(x_i) = f_0(x_j)$ when $x_i \sim x_j$;
- Γ is a $\mathcal{T}$-typing;
- Δ is a set of atoms $s(x_i, x_j)$, $s \in \Sigma$, such that $s(x_i, x_j) \in \Delta$ iff $s(x_i', x_j') \in \Delta$ when $x_i \sim x_i'$ and $x_j \sim x_j'$.

Intuitively, $\sim$ describes the answer variables that are identified by a match π of q in an ABox $\mathcal{A}$, $f_0(x_i)$ describes the $\mathcal{T}$-type of the ABox individual $\pi(x_i)$, Γ describes the $\mathcal{T}$-types of ABox individuals that are not in the range of π, and Δ fixes role relationships that do *not* hold between the $\pi(x_i)$. Let $X = \{\alpha_i \mid i \in I\}$ be a set of atoms with α_i in q_i for all $i \in I$. Then T *avoids* X if the following conditions hold:

1. if $A \in f_0(x)$, then $A(x) \notin X$ for any atom $A(x)$;

2. if $\exists s \in f_0(x)$, then (i) $S = \{B \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models \exists s^- \sqsubseteq B\}$ contains no predicate from Σ or (ii) there is a $u \in \Gamma$ such that $S \subseteq u$ or (iii) there is a y such that $S \subseteq f_0(y)$ and there are no $x' \sim x$ and $y' \sim y$ such that $r(x', y') \in X$ or $r(y', x') \in X$ and $\mathcal{T} \models s \sqsubseteq r$ or $\mathcal{T} \models s \sqsubseteq r^-$, respectively;

3. if $r(x, y) \in X$, then Δ contains all $s(x, y)$ with $s \in \Sigma$ and $\mathcal{T} \models s \sqsubseteq r$ and all $s(y, x)$ with $s \in \Sigma$ and $\mathcal{T} \models s^- \sqsubseteq r$.

T *avoids* q if it avoids some set $\{\alpha_i \mid i \in I\}$ with $\alpha_i \in q_i$ for all $i \in I$. We use $\mathsf{tp}(T)$ to denote the $\mathcal{T}$-typing Γ extended with all $\mathcal{T}$-types in the range of f_0. Let $\mathcal{A}$ be a Σ-ABox and let π assign ABox individuals $\pi(x_i)$ to x_i, $1 \leq i \leq n$, such that $\pi(x_i) = \pi(x_j)$ iff $x_i \sim x_j$. A $\mathcal{T}$-decoration f of $\mathcal{A}$ *realizes* $T = (\sim, f_0, \Gamma, \Delta)$ *using* π iff $\mathsf{tp}(T)$ is the range of f, $f_0(x_i) = f(\pi(x_i))$ for $1 \leq i \leq n$, and $r(\pi(x_i), \pi(x_j)) \notin \mathcal{A}$ if $r(x_i, x_j) \in \Delta$ for $1 \leq i, j \leq n$ and all $r \in \Sigma$. $\mathcal{A}$ *realizes* T using π if there exists a $\mathcal{T}$-decoration f that realizes T using π.

Lemma 23. *Let $\mathcal{A}$ be a Σ-ABox that is consistent w.r.t. $\mathcal{T}$ and closed Σ. Then $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} q[\pi(x_1), \ldots, \pi(x_n)]$ iff $\mathcal{A}$ realizes some $\mathcal{T}, q$-typing T using π that avoids q and such that $\mathsf{tp}(T)$ is Σ-realizable.*

Proof.(sketch) The proof is a modification of the proof of Lemma 6. We only sketch the differences.

($\Rightarrow$) Let $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} q(\pi(x_1), \ldots, \pi(x_n))$. We start with a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ such that $\mathcal{I} \not\models q(\pi(x_1), \ldots, \pi(x_n))$. Read off a $\mathcal{T}, q$-typing

$$T_{\mathcal{I}} = (\sim, f_0, \Gamma, \Delta)$$

from $\mathcal{I}$ by setting

- $x_i \sim x_j$ iff $\pi(x_i) = \pi(x_j)$;
- $f_0(x_i) = t_{\mathcal{I}}^{\pi(x_i)}$ for all $1 \leq i \leq n$;
- $\Gamma = \{t_{\mathcal{I}}^a \mid a \in \mathsf{Ind}(\mathcal{A})\} \setminus \{\pi(x_1), \ldots, \pi(x_n)\}$;
- $\Delta = \{r(x_i, x_j) \mid r \in \Sigma, r(\pi(x_i), \pi(x_j)) \notin \mathcal{A}\}$.

We show that $T_{\mathcal{I}}$ avoids $q = \bigvee_{i \in I} q_i$. Since $\mathcal{I} \not\models q(\pi(x_1), \ldots, \pi(x_n))$ we find for every $i \in I$ an atom $\alpha_i \in q_i$ such that $\mathcal{I} \not\models \alpha_i(\pi(x_1), \ldots, \pi(x_n))$. We show that $T_{\mathcal{I}}$ avoids $X = \{\alpha_i \mid i \in I\}$. We distinguish the following cases:

- Let $A(x) \in X$. Then $A \notin t_{\mathcal{I}}^{\pi(x)}$ and so $A \notin f_0(x)$, as required.
- Let $\exists s \in f_0(x)$. Then $\exists s \in t_{\mathcal{I}}^{\pi(x)}$. Thus, there exists $d \in \Delta^{\mathcal{I}}$ such that $(\pi(x), d) \in s^{\mathcal{I}}$. If $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, then $\mathsf{sig}(B) \cap \Sigma = \emptyset$ for all $B \in t_{\mathcal{I}}^d$. Thus (i) holds. If $d \in \mathsf{Ind}(\mathcal{A}) \setminus \{\pi(x_1), \ldots, \pi(x_n)\}$, then (ii) holds. Now assume that $d = \pi(y)$ for some $y \in \{\pi(x_1), \ldots, \pi(x_n)\}$. then y satisfies the conditions for (iii).
- Let $r(x, y) \in X$. Then $(\pi(x), \pi(y)) \notin r^{\mathcal{I}}$. Hence $(\pi(x), \pi(y)) \notin s^{\mathcal{I}}$ for any $s \in \Sigma$ with $\mathcal{T} \models s \sqsubseteq r$. Thus $s(x, y) \in \Delta$ for any such s. Moreover, $(\pi(y), \pi(x)) \notin s^{\mathcal{I}}$ for any $s \in \Sigma$ with $\mathcal{T} \models s^- \sqsubseteq r$. Thus $s(y, x) \in \Delta$ for any such s.

($\Leftarrow$) Assume that a Σ-Abox $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$ with closed Σ realizes some $\mathcal{T}, q$-typing $T = (\sim, f_0, \Gamma, \Delta)$ using π that avoids q. Assume f is a $\mathcal{T}, q$-decoration of $\mathcal{A}$ that realizes T using π. Let $X = \{\alpha_i \mid i \in I\}$ with $\alpha_i \in q_i$ such that T avoids X using π. We construct a model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ that respects closed predicates Σ such that $\mathcal{I} \not\models \alpha_i[\pi(x_1), \ldots, \pi(x_n)]$ for $i \in I$. We build $\mathcal{I}$ as in the proof of Lemma 6 based on $\mathsf{tp}(T)$. Some care is required in the construction of $\mathcal{I}_{i+1}$. Assume $\mathcal{I}_i$ has been constructed. Choose $d \in \Delta^{\mathcal{I}_i}$ and $\exists s \in \mathsf{sub}(\mathcal{T})$ such that $\mathsf{sig}(s) \cap \Sigma = \emptyset$, $\mathcal{T} \models \bigsqcap t_{\mathcal{I}_i}^d \sqsubseteq \exists s$ and there is no $(d, e) \in s^{\mathcal{I}_i}$. If $d \notin \{\pi(x_1), \ldots, \pi(x_n)\}$ or $\{B \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models \exists s^- \sqsubseteq B\}$ does not contain a B with $\mathsf{sig}(B) \subseteq \Sigma$ proceed as in the proof of Lemma 6. Now assume that $d = \pi(x)$. In the proof of Lemma 6 we chose an *arbitrary* $b \in \mathsf{Ind}(\mathcal{A})$ with $\{B \in \mathsf{sub}(\mathcal{T}) \mid \exists s^- \sqsubseteq B\} \subseteq t$ and $t = f(b)$ and added (a, b) to $r^{\mathcal{I}_{i+1}}$ whenever $\mathcal{T} \models s \sqsubseteq r$. Since we want to refute all atoms $\alpha_i(\pi(x_1), \ldots, \pi(x_n))$ with $i \in I$, we now have to choose b more carefully. If there exists $b \in \mathsf{Ind}(\mathcal{A}) \setminus \{\pi(x_1), \ldots, \pi(x_n)\}$ with $\{B \in \mathsf{sub}(\mathcal{T}) \mid \exists s^- \sqsubseteq B\} \subseteq t$ and $t = f(b)$, then we choose such a b and proceed as in Lemma 6. Otherwise, since f is a $\mathcal{T}, q$-decoration of $\mathcal{A}$ that realizes T using π and avoids X, there is y such that $\{B \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models \exists s^- \sqsubseteq B\} \subseteq f_0(y)$ such that there is no $\alpha_i \in X$ of the form $t(x', y')$ or $t(y', x')$ with

$x' \sim x$ and $y' \sim y$ such that $\mathcal{T} \models s \sqsubseteq t$ or $\mathcal{T} \models s \sqsubseteq t^-$, respectively. We set $b = \pi(y)$ and proceed as in the proof of Lemma 6.

The resulting interpretation $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ. Moreover $\mathcal{I} \not\models \alpha_i(\pi(x_1), \dots, \pi(x_n))$ for all $i \in I$. Thus, $\mathcal{I} \not\models q(\pi(x_1), \dots, \pi(x_n))$, as required. ❏

Proposition 24 Let $\mathcal{A}$ be a Σ-ABox that is consistent w.r.t. $\mathcal{T}$ with closed Σ. Then $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(a_1, \dots, a_n)$ iff $\mathcal{I}_\mathcal{A} \models \Phi_Q[a_1, \dots, a_n]$.

Proof. Assume $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} q(a_1, \dots, a_n)$. Let $\pi(x_i) = a_i$ for $1 \leq i \leq n$. By Lemma 23, $\mathcal{A}$ realizes some $\mathcal{T}, q$-typing T using π that avoids q such that $\mathsf{tp}(T)$ is Σ-realizable. It is readily checked that $\mathcal{I}_\mathcal{A} \models \Psi_T(\pi_1(x_1), \dots, \pi(x_n))$. Thus, $\mathcal{I}_\mathcal{A} \not\models \Phi_Q[a_1, \dots, a_n]$

Conversely, assume that $\mathcal{I}_\mathcal{A} \not\models \Phi_Q[a_1, \dots, a_n]$. Let $\pi(x_i) = a_i$ for $1 \leq i \leq n$. Take a $\mathcal{T}, q$-typing T that avoids q such that $\mathsf{tp}(T)$ is Σ-realizable and $\mathcal{I}_\mathcal{A} \models \Psi_T[a_1, \dots, a_n]$. Let $\pi(x_i) = a_i$ for $1 \leq i \leq n$. It is readily checked that $\mathcal{A}$ realizes T using π. Thus $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma)} q(a_1, \dots, a_n)$, by Lemma 23. ❏

We now construct the actual rewriting $\Phi_Q(\vec{x})$. For every $\mathcal{T}, q$-typing $T = (\sim, f_0, \Gamma, \Delta)$ with $\Gamma = \{t_1, \dots, t_k\}$ let $\Psi_T(\vec{x})$ be the conjunction of the following:

$$\bigwedge_{1 \leq i \leq n} \psi_{f_0(x_i)}(x_i) \wedge \bigwedge_{x_i \sim x_j} (x_i = x_j) \wedge \bigwedge_{x_i \not\sim x_j} (x_i \neq x_j)$$

$$\bigwedge_{r(x_i, x_j) \in \Delta} \neg r(x_i, x_j) \wedge \forall y (\bigwedge_{1 \leq i \leq n} (y \neq x_i) \rightarrow \bigvee_{t \in \Gamma} \psi_t(y))$$

$$\exists y_1 \cdots \exists y_k (\bigwedge_{j \neq i} y_j \neq y_i \wedge \bigwedge_{j \leq k, i \leq n} x_i \neq y_j \wedge \bigwedge_{j \leq k} \psi_{t_j}(y_j))$$

Then $\Phi_Q(\vec{x})$ is the conjunction over all $\neg \Psi_T(\vec{x})$ such that T avoids q and $\mathsf{tp}(T)$ is Σ-realizable.

Proposition 24. *Let $\mathcal{A}$ be a Σ-ABox that is consistent w.r.t. $\mathcal{T}$ and closed Σ. Then $\mathcal{T}, \mathcal{A} \models_{c(\Sigma)} q(\vec{a})$ iff $\mathcal{I}_\mathcal{A} \models \Phi_Q[\vec{a}]$ for all tuples $\vec{a}$ from $\mathsf{Ind}(\mathcal{A})$.*

Theorem 8. There is a DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$ and set of predicats Σ_C such that consistency w.r.t. $\mathcal{T}$ and closed Σ_C is NP-complete.

Proof. The proof is by reduction of the satisfiability problem for propositional formulas in conjunctive normal form (CNF). Consider a propositional formula in CNF $\varphi = c_1 \wedge \cdots \wedge c_n$, where each c_i is a disjunction of literals. We write $\ell \in c_i$ if ℓ is a disjunct in c_i. Let $x_1, \dots, x_m$ be the propositional variables in φ. Define an ABox $\mathcal{A}_\varphi$ with individual names $c_1, \dots, c_n$ and $x_i^\top, x_i^\perp, x_i^{\mathsf{aux}}$ for $1 \leq i \leq m$, a concept name A, and role names r, r' as the following set of assertions:

- $r(c_i, x_j^\top)$, for all $x_j \in c_i$ and $1 \leq i \leq n$;
- $r(c_i, x_j^\perp)$, for all $\neg x_j \in c_i$ and $1 \leq i \leq n$;
- $r'(x_j^\top, x_j^\perp)$, $r'(x_j^\perp, x_j^{\mathsf{aux}})$, for $1 \leq j \leq m$;
- $A(c_i)$, for $1 \leq i \leq n$.

Let s and s' be additional role names and let $\mathcal{T}$ consist of the following inclusions:

- $s \sqsubseteq r$ and $A \sqsubseteq \exists s$;
- $\exists s^- \sqsubseteq \exists s'$, $s' \sqsubseteq r'$, and $\exists s'^- \sqcap \exists s^- \sqsubseteq \perp$.

Let $\Sigma_\mathsf{C} = \{r, r'\}$. We show that $\mathcal{A}_\varphi$ is consistent w.r.t. $\mathcal{T}$ and closed Σ_C iff φ is satisfiable. Assume first that $\mathcal{A}_\varphi$ is consistent w.r.t. $\mathcal{T}$ and closed Σ_C. Let $\mathcal{I}$ be a model of $\mathcal{T}$ and $\mathcal{A}_\varphi$ that respects closed predicates Σ_C. Define a propositional valuation v by setting $v(x_j) = 1$ if there exists i such that $s(c_i, x_j^\top) \in \mathcal{I}$ and set $v(x_j) = 0$ if there exists i such that $s(c_i, x_j^\perp) \in \mathcal{I}$. Observe that v is well-defined since if $s(c_i, x_j^\top), s(c_k, x_j^\perp) \in \mathcal{I}$, then $s'(x_j^\top, x_j^\perp) \in \mathcal{I}$ and so $x_j^\perp \in (\exists s'^- \sqcap \exists s^-)^\mathcal{I}$ which contradicts the assumption that $\mathcal{I}$ satisfies $\exists s'^- \sqcap \exists s^- \sqsubseteq \perp$. Next observe that for every c_i there exists a disjunct $\ell \in c_i$ such that $s(c_i, x_j^\top) \in \mathcal{I}$ if $\ell = x_j$ and $s(c_i, x_j^\perp) \in \mathcal{I}$ if $\ell = \neg x_j$. Thus, $v(\varphi) = 1$ and φ is satisfiable.

Conversely, assume that φ is satisfiable and let v be an assignment with $v(\varphi) = 1$. Define an interpretation $\mathcal{I}$ as follows: $\Delta^\mathcal{I}$ is the set of indivdual names in $\mathcal{A}_\varphi$; define the interpretation of r, r' and A_i exactly as in $\mathcal{A}_\varphi$; set

$$\begin{aligned} s^\mathcal{I} &= \{(c_i, x_j^\top)\} \mid x_j \in c_i, v(x_j) = 1, i \leq n\} \cup \\ &\quad \{(c_i, x_j^\perp) \mid \neg x_j \in c_i, v(x_j) = 0, i \leq n\} \\ s'^\mathcal{I} &= \{(x_j^\top, x_j^\perp) \mid v(x_j) = 1\} \cup \\ &\quad \{(x_j^\perp, x_j^{\mathsf{aux}}) \mid v(x_j) = 0\} \end{aligned}$$

It is readily checked that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}_\varphi$ that respects closed predicates Σ_C. ❏

D Proofs for Section 5

Lemma 25. *The complement of $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ can be reduced in polynomial time to answering $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$.*

Proof. Let $\mathcal{J}$ be a Σ-interpretation that is an input of $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$. Let $\mathcal{A}_\mathcal{J}$ be the ABox that corresponds to $\mathcal{J}$, that is, $\mathcal{A}_\mathcal{J}$ consists of the assertion $A(d)$ for each $d \in A^\mathcal{J}$ and $r(d, e)$ for all $(d, e) \in r^\mathcal{J}$ where the elements of $\mathcal{J}$ serve as ABox individuals. For each $d \in \Delta^\mathcal{I}$, introduce a fresh individual name a_d and let the ABox $\mathcal{A}$ be defined as

$$\mathcal{A}_\mathcal{J} \cup \{A(d) \mid d \in \Delta^\mathcal{J}\} \cup \{V(a_d), V_d(a_d) \mid d \in \Delta^\mathcal{I}\}.$$

Obviously, $\mathcal{A}$ can be constructed in polynomial time. We claim $\mathcal{J} \in \mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ iff $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$.

($\Rightarrow$) Suppose that there is a surjective homomorphism h

from $\mathcal{J}$ to $\mathcal{I}$. Define the interpretation $\mathcal{I}'$ as follows:

$$\begin{array}{rcl}
\Delta^{\mathcal{I}'} & = & \mathsf{Ind}(\mathcal{A}) \\
A^{\mathcal{I}'} & = & \mathsf{Ind}(\mathcal{A}_{\mathcal{J}}) \\
V^{\mathcal{I}'} & = & \Delta^{\mathcal{I}} \\
V_d^{\mathcal{I}'} & = & \{a_d\}, \text{ for all } d \in \Delta^{\mathcal{I}} \\
\mathsf{val}^{\mathcal{I}'} & = & \{(a, a_{h(a)}) \mid a \in \mathsf{Ind}(\mathcal{A}_{\mathcal{J}})\} \\
\mathsf{aux}_d^{\mathcal{I}'} & = & \{(a, a_d) \mid a \in \mathsf{Ind}(\mathcal{A}_{\mathcal{J}})\}, \text{ for all } d \in \Delta^{\mathcal{I}} \\
\mathsf{force}_d^{\mathcal{I}'} & = & \{(a, a') \in \mathsf{Ind}(\mathcal{A}_{\mathcal{J}}) \times \mathsf{Ind}(\mathcal{A}_{\mathcal{J}}) \mid h(a') = d\} \\
P^{\mathcal{I}'} & = & P^{\mathcal{J}}, \text{ for all } P \in (\mathsf{N_C} \cup \mathsf{N_R}) \text{ not in} \\
& & (\{A, V, \mathsf{val}\} \cup \{V_d, \mathsf{aux}_d, \mathsf{force}_d \mid d \in \Delta^{\mathcal{I}}\})
\end{array}$$

One can now verify that $\mathcal{I}'$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C, and that $\mathcal{I}' \not\models q$.

($\Leftarrow$) Suppose $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$. Then there is a model $\mathcal{I}'$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C and such that $\mathcal{I}' \not\models q$. Define $h = \{(d, a_e) \in \mathsf{val}^{\mathcal{I}'} \mid d \in \Delta^{\mathcal{J}}\}$. We show that h is a surjective homomorphism from $\mathcal{J}$ to $\mathcal{I}$.

We first show that the relation h is a function. Assume that this is not the case, that is, there are $d \in \Delta^{\mathcal{J}}$ and $e_1, e_2 \in \Delta^{\mathcal{I}}$ such that $e_1 \neq e_2$ and $(d, a_{e_i}) \in \mathsf{val}^{\mathcal{I}'}$ for $i \in \{1, 2\}$. Note that $a_{e_i} \in V_{e_i}^{\mathcal{I}'}$. Thus we get $\mathcal{I}' \models q_1$, which is a contradiction against our choice of $\mathcal{I}'$. To show that h is total, take some $d \in \Delta^{\mathcal{J}}$. Then $d \in A^{\mathcal{I}'}$ and thus the first line of $\mathcal{T}$ yields an $f \in V^{\mathcal{J}}$ with $(d, f) \in \mathsf{val}^{\mathcal{I}'}$. Since V is closed, we must have $f = a_e$ for some e, and thus $h(a_e) = f$.

Now we show that h is a homomorphism. Thus assume for a contradiction that there is $(d, e) \in r^{\mathcal{J}}$ with $(h(d), h(e)) \notin r^{\mathcal{I}}$. The latter implies that the following is a disjunct of q_2:

$$\exists x \exists y \exists x_1 \exists y_1\, A(x) \wedge A(y) \wedge r(x, y) \wedge \mathsf{val}(x, x_1) \wedge \mathsf{val}(y, y_1) \wedge V_{h(d)}(x_1) \wedge V_{h(e)}(y_1).$$

Note that $d, e \in A^{\mathcal{I}'}$, $(d, a_{h(a)}), (e, a_{h(e)}) \in \mathsf{val}^{\mathcal{I}'}$, $a_{h(d)} \in V_{h(d)}^{\mathcal{I}'}$, and $a_{h(e)} \in V_{h(e)}^{\mathcal{I}'}$. Thus the above disjunct of q_2 has a match in $\mathcal{I}'$, a contradiction to our choice of $\mathcal{I}'$.

It remains to show that h is surjective. Fix a $d \in \Delta^{\mathcal{I}}$. We have to show that there is an $e \in \Delta^{\mathcal{J}}$ with $h(e) = d$. Take some $f \in \Delta^{\mathcal{J}}$. Then by the third line of $\mathcal{T}$ and since A is closed, there is some $e \in \Delta^{\mathcal{J}}$ such that $(f, e) \in \mathsf{force}_d^{\mathcal{I}'}$. We show that e is as required. Assume to the contrary that $h(e) \neq d$. Then the following is a disjunct of q_3:

$$A(x) \wedge \mathsf{force}_d(z, x) \wedge \mathsf{val}(x, y) \wedge V_{h(e)}(y).$$

Note that $f \in A^{\mathcal{I}'}$, $(e, a_{h(e)}) \in \mathsf{val}^{\mathcal{I}'}$, and $a_{h(e)} \in V_{h(e)}^{\mathcal{I}'}$. Thus the above disjunct of q_3 has a match in $\mathcal{I}'$, in contradiction to our choice of $\mathcal{I}'$. ❏

Lemma 26. *Answering $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ can be reduced in polynomial time to the complement of $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$.*

Proof. Let $\mathcal{A}$ be a Σ_A-ABox that is an input to $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$. We start with the following:

1. If $\mathcal{A}$ does not contain any assertion of the form $A(a)$, then $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$. In fact, let $\mathcal{I}_\mathcal{A}$ be $\mathcal{A}$ viewed as an interpretation in the obvious way. Then $\mathcal{I}_\mathcal{A}$ is a model of $\mathcal{A}$ that respects closed predicates Σ_C. Since $\mathcal{A}$ does not contain any assertion of the form $A(a)$, $\mathcal{I}_\mathcal{A}$ is also a model of $\mathcal{T}$ and satisfies $\mathcal{I}_\mathcal{A} \not\models q$ (note that each disjunct of q demands the existence of an instance of A). Thus answer 'no'.
2. Otherwise, if $\mathcal{A}$ does not contain for each $d \in \Delta^{\mathcal{I}}$ an element a that satisfies $V(a)$ and $V_d(a)$, then $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ and closed Σ_C. Thus answer 'yes'.
3. Otherwise, if $\mathcal{A}$ contains an element a that satisfies $V(a)$ but not $V_d(a)$ for any $d \in \Delta^{\mathcal{I}}$, then $\mathcal{T}, \mathcal{A} \models_{c(\Sigma_\mathsf{C})} q$ iff $\mathcal{I}_\mathcal{A} \models q$. In fact, Line 1 of $\mathcal{T}$ can be satisfied by linking every element to a via val; Line 2 can be satisfied since Case 2 above does not apply; Line 3 can be satisfied since Case 1 above does not apply. If an interpretation $\mathcal{I}$ is build in this way (taking all remaining choices in an arbitrary way), then it can be verified that $\mathcal{T}, \mathcal{A} \models_{c(\Sigma_\mathsf{C})} q$ iff $\mathcal{I} \models q$ iff $\mathcal{I}_\mathcal{A} \models q$. Thus check in polynomial time whether $\mathcal{I}_\mathcal{A} \models q$ and answer accordingly.

If none of the above cases applies, let $\mathcal{A}'$ be the restriction of $\mathcal{A}$ to all elements a such that $A(a) \in \mathcal{A}$. Since Case 1 above does not apply, $\mathcal{A}'$ is non-empty.

Claim. $\mathcal{A}' \in \mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ iff $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$.

($\Leftarrow$). Assume that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_\mathsf{C})} q$. Then there is a model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C and such that $\mathcal{J} \not\models q$. By the first line of $\mathcal{T}$, since V is closed, Case 3 does not apply, and by the first line of q, for each $a \in \mathsf{Ind}(\mathcal{A})$ there is exactly one $d \in \Delta^{\mathcal{I}}$ such that $a \in (\exists \mathsf{val}.V_d)^{\mathcal{J}}$. Define a homomorphism $h : \mathcal{A}' \to \mathcal{I}$ by mapping each a to the value $d \in \Delta^{\mathcal{I}}$ thus determined. By the second line of q, h is indeed a homomorphism. By the third line of $\mathcal{T}$ and the third line of q and since A is closed, h must be surjective.

($\Rightarrow$). Assume that $\mathcal{A}' \in \mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$, and let h be a surjective homomorphism from $\mathcal{A}'$ to $\mathcal{I}$. Build an interpretation $\mathcal{J}$ as follows. Start by setting $\mathcal{J} = \mathcal{I}_\mathcal{A}$. Since Case 2 above does not apply, for each $d \in \Delta^{\mathcal{I}}$ we can select an element a_d of $\mathcal{A}$ such that $V(a)$ and $V_d(a)$ are in $\mathcal{A}$. For each element a in $\mathcal{A}'$, extend $\mathcal{J}$ by adding $(a, a_{h(a)})$ to $\mathsf{val}^{\mathcal{J}}$ and (a, a_d) to $\mathsf{aux}_d^{\mathcal{J}}$ for each $d \in \Delta^{\mathcal{I}}$. Since h is surjective, for each $d \in \Delta^{\mathcal{I}}$ there must be an element a_d of $\mathcal{A}'$ with $h(a_d) = d$. Further extend $\mathcal{J}$ by adding (a, a_d) to $\mathsf{force}_d^{\mathcal{J}}$ for all $a \in \mathsf{Ind}(\mathcal{A})$ and all $d \in \Delta^{\mathcal{I}}$. It is readily checked that $\mathcal{J}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C, and that $\mathcal{J} \not\models q$. ❏

Lemma 27. *$(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ can be reduced in polynomial time to the complement of $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$.*

Proof. Let $\mathcal{A}$ be a Σ_A-ABox that is an input for $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ and let $\mathcal{A}'$ be its extension with

- all assertions $\overline{A}(a)$ such that $A \in \Sigma_\mathsf{C}$ and $A(a) \notin \mathcal{A}$;
- an assertion $A_{\mathsf{open}}(a_B)$ for each $B \in \Sigma_\mathsf{C}$.

We claim that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_C)} q$ iff there is an $\mathcal{I}_T \in \Gamma$ such that there exists a surjective homomorphism from $\mathcal{A}'$ to $\mathcal{I}_T$.

($\Leftarrow$). Let $\mathcal{I}_T \in \Gamma$ and let h be a surjective homomorphism from $\mathcal{A}'$ to $\mathcal{I}_T$. Note that each element a of $\mathcal{A}$ is mapped by h to some element $t \in T$ of $\mathcal{I}_T$ because $A(a) \in \mathcal{A}'$ or $\overline{A}(a) \in \mathcal{A}'$ for every $A \in \Sigma_C$ (which is non-empty). Since $\mathcal{I}_T \in \Gamma$, there is a Σ_A-ABox $\mathcal{B}$ and model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{B}$ that respects closed predicates Σ_C such that $\mathcal{I} \not\models q$ and $T = \{\mathsf{tp}_\mathcal{I}(a) \mid a \in \mathsf{Ind}(\mathcal{B})\}$. For each $a \in \mathsf{Ind}(\mathcal{A})$, set $t_a = h(a) \in T$ and for each $d \in \Delta^\mathcal{I}$, set $t_d = \mathsf{tp}_\mathcal{I}(d)$. Construct an interpretation $\mathcal{J}$ as follows:

$$\begin{aligned} \Delta^\mathcal{J} &= \mathsf{Ind}(\mathcal{A}) \cup (\Delta^\mathcal{I} \setminus \mathsf{Ind}(\mathcal{B})) \\ A^\mathcal{J} &= \{d \in \Delta^\mathcal{J} \mid A \in t_d\} \\ r^\mathcal{J} &= \{(d,e) \in \Delta^\mathcal{J} \times \Delta^\mathcal{J} \mid t_d \rightsquigarrow_r t_e\}. \end{aligned}$$

First note that $\mathcal{J}$ is clearly a model of $\mathcal{A}$ that respects closed predicates Σ_C. Specifically, if $A(a) \in \mathcal{A}$, then $h(a) \in A^{\mathcal{I}_T}$, thus $A \in h(a) = t_a$ by construction of $\mathcal{I}_T$ which yields $a \in A^\mathcal{J}$ by construction of $\mathcal{J}$; if $r(a,b) \in \mathcal{A}$, then $(h(a), h(b)) \in r^{\mathcal{I}_T}$, thus $t_a \rightsquigarrow_r t_b$ implying $(a,b) \in r^\mathcal{J}$; finally if $A \in \Sigma_C$ and $d \in A^\mathcal{J}$, then we must have $d = a$ for some $a \in \mathsf{Ind}(\mathcal{A})$ by definition of $\mathcal{J}$ and since $d \notin A^\mathcal{I}$ for all $d \in \Delta^\mathcal{I} \setminus \mathsf{Ind}(\mathcal{B})$. Thus, $A \in t_a = h(a)$ by construction of $\mathcal{J}$. This implies $A(a) \in \mathcal{A}$ since otherwise $\overline{A}(a) \in \mathcal{A}'$, which would imply $\overline{A} \in h(a)$, in contradiction to $A \in h(a)$.

It thus remains to show that $\mathcal{J}$ is a model of $\mathcal{T}$ and $\mathcal{J} \not\models q$. By definition, $\mathcal{J}$ satisfies all role inclusions in $\mathcal{T}$. Satisfaction of $\top \sqsubseteq C_\mathcal{T}$ and $\mathcal{J} \not\models q$ follows from the subsequent claim together with the fact that each type in $\mathcal{I}_T$ contains $C_\mathcal{T}$ but not the concept name A_0 from q, and that $\mathcal{I}$ is a model of $\mathcal{T}$ with $\mathcal{I} \not\models q$.

Claim. For all $d \in \Delta^\mathcal{J}$ and $C \in \mathsf{cl}(\mathcal{T})$, we have $d \in C^\mathcal{J}$ iff $C \in t_d$.

The proof is by induction on the structure of C, with the induction start and the cases $C = \neg D$ and $C = D_1 \sqcap D_2$ being trivial. Thus let $C = \exists r.D$ and first assume $d \in C^\mathcal{J}$. Then there is an $e \in D^\mathcal{J}$ with $(d,e) \in r^\mathcal{J}$. Thus $t_d \rightsquigarrow_r t_e$ by definition of $\mathcal{J}$, and IH yields $D \in t_e$. By definition of '$\rightsquigarrow_r$', we must thus have $C \in t_d$ as required. Now let $C \in t_d$. We distinguish two cases:

- $d = a \in \mathsf{Ind}(\mathcal{A})$.
  Let $a' \in \mathsf{Ind}(\mathcal{B})$ be such that $h(a) = \mathsf{tp}_\mathcal{I}(a')$. Since $t_a = h(a)$, we must have $a' \in C^\mathcal{I}$ and thus there is some $e \in D^\mathcal{I}$ with $(a', e) \in r^\mathcal{I}$, which yields $\mathsf{tp}_\mathcal{I}(a') \rightsquigarrow_r \mathsf{tp}_\mathcal{I}(e)$ and $D \in \mathsf{tp}_\mathcal{I}(e)$. If $e = b' \in \mathsf{Ind}(\mathcal{B})$, then since h is surjective there is some $b \in \mathsf{Ind}(\mathcal{A})$ with $h(b) = \mathsf{tp}_\mathcal{I}(b')$. We have $t_a = \mathsf{tp}_\mathcal{I}(a')$ and $t_b = \mathsf{tp}_\mathcal{I}(b')$, thus $t_a \rightsquigarrow_r t_b$ which yields $(a,b) \in r^\mathcal{J}$ by definition of $\mathcal{J}$. We also have $D \in t_b$, which by IH yields $b \in D^\mathcal{J}$.
- $d \notin \mathsf{Ind}(\mathcal{A})$.
  Then $d \in \Delta^\mathcal{I} \setminus \mathsf{Ind}(\mathcal{B})$. Since $C \in t_d$, we thus have $C \in \mathsf{tp}_\mathcal{I}(d)$. Thus, there is an $e \in D^\mathcal{I}$ with $(d,e) \in r^\mathcal{I}$, which implies $\mathsf{tp}_\mathcal{I}(d) \rightsquigarrow_r \mathsf{tp}_\mathcal{I}(e)$ and $D \in \mathsf{tp}_\mathcal{I}(e)$. If $e \notin \mathsf{Ind}(\mathcal{B})$, then the definition of $\mathcal{J}$ and IH yields $d \in C^\mathcal{J}$. Thus assume $e = b' \in \mathsf{Ind}(\mathcal{B})$. Since h is surjective, there is some $b \in \mathsf{Ind}(\mathcal{A})$ with $h(b) = \mathsf{tp}_\mathcal{I}(b')$. Since $t_d = \mathsf{tp}_\mathcal{I}(d)$ and $t_b = h(b)$, we have $t_d \rightsquigarrow_r t_b$, thus $(d,b) \in r^\mathcal{J}$. By IH, $D \in \mathsf{tp}_\mathcal{I}(b') = h(b)$ yields $b \in D^\mathcal{J}$.

($\Rightarrow$). Assume that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_C)} q$. Then there is a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C and such that $\mathcal{I} \not\models q$. Let $\mathcal{I}_T \in \Gamma$ be the corresponding template, that is, $T = \{\mathsf{tp}_\mathcal{I}(a) \mid a \in \mathsf{Ind}(\mathcal{A})\}$. For each $a \in \mathsf{Ind}(\mathcal{A})$, set $h(a) = \mathsf{tp}_\mathcal{I}(a)$; for each $a_B \in \mathsf{Ind}(\mathcal{A}') \setminus \mathsf{Ind}(\mathcal{A})$, set $h(a_B) = d_B$. It is readily checked that h is a surjective homomorphism from $\mathcal{A}'$ to $\mathcal{I}_T$. In particular, $\overline{A}(a) \in \mathcal{A}'$ implies $A(a) \notin \mathcal{A}'$, thus $A \notin \mathsf{tp}_\mathcal{I}(a)$ (since A is closed), which yields $h(a) = \mathsf{tp}_\mathcal{I}(a) \in \overline{A}^{\mathcal{I}_\mathcal{T}}$ by definition of $\mathcal{I}_\mathcal{T}$. ❑

Lemma 28. $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$ *polynomially reduces to the complement of* $(\mathcal{T}, \Sigma_A, \Sigma_C, q)$.

Proof. Let $\mathcal{A}'$ be an input for $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$. An element of $\mathcal{A}'$ is *special for* $A \in \Sigma_C$ if $A(a) \notin \mathcal{A}'$ and $\overline{A}(a) \notin \mathcal{A}'$; it is *special* if it is special for some $A \in \Sigma_C$. First perform the following checks:

1. if there is a non-special element a of $\mathcal{A}'$ such that $A(a) \in \mathcal{A}'$ and $\overline{A}(a) \in \mathcal{A}'$ for some $A \in \Sigma_C$, then return 'no' (there is no template in Γ that has any element to which a can be mapped by a homomorphism);
2. if $\mathcal{A}'$ does not contain a family of distinct elements $(a_A)_{A \in \Sigma_C}$, such that each a_A is special for A, then return 'no' (we cannot map surjectively to the elements d_A of the templates in Γ).

Note that, to check Condition 2, we can go through all candidate families in polytime since the size of Σ_C is constant. If none of the above checks succeeds, then let $\mathcal{A}$ be the ABox obtained from $\mathcal{A}'$ by

- deleting all assertions of the form $\overline{A}(a)$ and
- deleting all special elements.

We have to show that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_C)} q$ iff there exists an $\mathcal{I}_T \in \Gamma$ such that there is a surjective homomorphism from $\mathcal{A}'$ to $\mathcal{I}_T$.

($\Leftarrow$). Let $\mathcal{I}_T \in \Gamma$ and let h be a surjective homomorphism from $\mathcal{A}'$ to $\mathcal{I}_T$. Note that each element a of $\mathcal{A}$ is mapped by h to some element $t \in T$ of $\mathcal{I}_T$ because $A(a) \in \mathcal{A}'$ or $\overline{A}(a) \in \mathcal{A}'$ for every $A \in \Sigma_C$ (which is non-empty). Since $\mathcal{I}_T \in \Gamma$, there is a Σ_A-ABox $\mathcal{B}$ and model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{B}$ that respects closed predicates Σ_C and such that $\mathcal{I} \not\models q$ and $T = \{\mathsf{tp}_\mathcal{I}(a) \mid a \in \mathsf{Ind}(\mathcal{B})\}$. We can now proceed as in the proof of Lemma 27 to build a model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C and such that $\mathcal{J} \not\models q$.

($\Rightarrow$). Assume that $\mathcal{T}, \mathcal{A} \not\models_{c(\Sigma_C)} q$. Then there is a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C and such that $\mathcal{I} \not\models q$. Let $\mathcal{I}_T \in \Gamma$ be the corresponding template, that is, $T = \{\mathsf{tp}_\mathcal{I}(a) \mid a \in \mathsf{Ind}(\mathcal{A})\}$. For each $a \in \mathsf{Ind}(\mathcal{A})$, set $h(a) = \mathsf{tp}_\mathcal{I}(a)$; for each element $a \in \mathsf{Ind}(\mathcal{A}') \setminus \mathsf{Ind}(\mathcal{A})$, we can choose some $A \in \Sigma_C$ such that $A(a) \notin \mathcal{A}'$ and $\overline{A} \notin \mathcal{A}'$, and set $h(a) = d_A$; by Check 2 above, these choices can be made such that the resulting map h is surjective. Moreover, it is readily checked that h is a homomorphism from $\mathcal{A}'$ to $\mathcal{I}_T$. In particular, $\overline{A}(a) \in \mathcal{A}'$ implies $A(a) \notin \mathcal{A}'$ by Check 1, thus $A \notin \mathsf{tp}_\mathcal{I}(a)$ (since A is closed), which yields $h(a) = \mathsf{tp}_\mathcal{I}(a) \in \overline{A}^{\mathcal{I}_\mathcal{T}}$ by definition of $\mathcal{I}_\mathcal{T}$. ❏

E Proofs for Section 6

We refrain from introducing the formal details of monadic disjunctive datalog and instead refer the reader to [Bienvenu *et al.*, 2014]. We allow only unary and binary relation symbols, identifying unary EDB relations with concept names and binary EDB relations with role names; the only exception is the goal predicate, which has arity zero. We say that a monadic disjunctive datalog program Π is *simple* if there is a set $\mathcal{F}$ of binary (EDB) relations (which we call *functional*) such that Π consists of the following rules (where $Y_\mathcal{F}$ abbreviates $\bigwedge_{r\in\mathcal{F}} r(x_r, y_r)$):

(a) a single disjunctive rule $P_1(x) \vee \cdots \vee P_n(x) \leftarrow \top$;

(b) for each $r \in \mathcal{F}$, the rule

$$\mathsf{goal}() \leftarrow r(x,y) \wedge r(x,z) \wedge \neg(y=z) \wedge Y_\mathcal{F};$$

(c) any number of rules of the form $\mathsf{goal}() \leftarrow A(x) \wedge r(x,y) \wedge B(y) \wedge Y_\mathcal{F}$ and $\mathsf{goal}() \leftarrow A(x) \wedge B(x) \wedge Y_\mathcal{F}$.

We use $\Sigma_{\mathsf{EDB}}(\Pi)$ to denote the set of EDB relations used in the monadic disjunctive datalog program Π, that is, $\Sigma_{\mathsf{EDB}}(\Pi)$ contains all relations in Π except $P_1, \dots, P_n$ and the goal relation. For a $\Sigma_{\mathsf{EDB}}(\Pi)$-ABox $\mathcal{A}$, we write $\mathcal{A} \models \Pi$ if every model $\mathcal{I}$ of $\mathcal{A}$ that respects closed predicates $\Sigma_{\mathsf{EDB}}(\Pi)$ satisfies $\mathcal{I} \models \exists\vec{x}\, q$ for some rule $\mathsf{goal}() \leftarrow q$ in Π and where $\vec{x}$ are the variables that occur in q (note that $\exists\vec{x}\, q$ is a BCQ).

Theorem 13. For every simple disjunctive datalog program Π, there exists an OMQC in $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BAQ})$ that is polynomially equivalent to Π. The same is true for $(\text{DL-Lite}_\mathcal{R}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BtUCQ})$.

Proof. Assume a simple disjunctive datalog program Π is given. Assume its set of functional roles is $\mathcal{F}$ and its disjunctive rule is $P_1(x) \vee \cdots \vee P_n(x) \leftarrow \bot$. We define a polynomially equivalent query $Q_\Pi = (\mathcal{T}_\Pi, \Sigma_\Pi, q_\Pi)$ in $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BtUCQ})$ and then obtain the required queries in $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BAQ})$ and $(\text{DL-Lite}_\mathcal{R}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BtUCQ})$ by slight modifications of Q_Π.[2] Define the set of closed predicates of Q_Π as

$$\Sigma_\Pi = \mathcal{F} \cup \{\mathsf{True}, \mathsf{False}, \mathsf{Val}\}$$

where $\mathsf{True}, \mathsf{False}, \mathsf{Val}$ are fresh concept names. We use auxiliary role names $s_1, \dots, s_n$ and, for each $r \in \mathcal{F}$, a role name s_r and concept names A_r and B_r (all fresh symbols). $\mathcal{T}_\Pi$ contains the following inclusions:

$$\begin{array}{rcl} \mathsf{True} & \sqsubseteq & \mathsf{Val} \\ \mathsf{False} & \sqsubseteq & \mathsf{Val} \\ \top & \sqsubseteq & \exists s_i.\mathsf{Val} \text{ for } 1 \leq i \leq n \\ \top & \sqsubseteq & \exists s_r.(\exists r.A_r \sqcap \exists r.B_r) \text{ for all } r \in \mathcal{F}. \end{array}$$

[2] Thus, Σ_Π serves both as the set of closed predicates and as the ABox signature

Intuitively, we simulate the IDB predicate $P_i(x)$ and its complement using

$$\begin{array}{rcl} P_i^t(x) & := & \exists y_{P_i(x)}\, (s_i(x, y_{P_i(x)}) \wedge \mathsf{True}(y_{P_i(x)})) \\ P_i^f(x) & := & \exists y_{P_i(x)}\, (s_i(x, y_{P_i(x)}) \wedge \mathsf{False}(y_{P_i(x)})). \end{array}$$

Define q_Π as the union of the following BCQs, where for brevity we omit the existential quantifiers:

1. $q_\mathcal{F} \wedge X^t$, for all rules $\mathsf{goal}() \leftarrow X$ in Π and
2. $q_\mathcal{F} \wedge \bigwedge_{1\leq i\leq n} s_i(x, y_{P_i(x)}) \wedge \mathsf{False}(y_{P_i(x)})$,

where

$$q_\mathcal{F} = \bigwedge_{r\in\mathcal{F}} (s_r(x, x_r) \wedge r(x_r, y_r) \wedge A_r(y_r) \wedge B_r(y_r))$$

and X^t results from X by replacing every occurence of IDB predicate $P_i(z)$ with $P_i^t(z)$ (without the existential quantifier $\exists y_{P_i(x)}$) and identifying the variable of the leftmost atom in X with the variable x from $q_\mathcal{F}$. This renaming serves the purpose of obtaining a tree-shaped query (with root x) as required rather than a forest-shaped one.

The purpose of the (sub)queries $q_\mathcal{F}$ is to ensure that all relations in $\mathcal{F}$ are partial functions; in fact, one can show that for any Σ_Π-ABox $\mathcal{A}$ such that $\mathcal{T}_\Pi$ and $\mathcal{A}$ have a common model which respects closed predicates $\mathcal{F}$, we have $\mathcal{T}_\Pi, \mathcal{A} \models_{c(\Sigma_\Pi)} \exists\vec{x}\, q_\mathcal{F}$ (where $\vec{x}$ are the variables in $q_\mathcal{F}$) iff all $r \in \mathcal{F}$ are non-empty partial functions in $\mathcal{A}$, that is, $r(a, b_1), r(a, b_2) \in \mathcal{A}$ implies $b_1 = b_2$. Consequently, the answer to Q_Π is trivially 'no' on ABoxes where some $r \in \mathcal{F}$ is not a non-empty partial function. The query under Point 2 above ensures that $\forall x(P_1(x) \vee \cdots \vee P_n(x))$ holds and thus that the disjunctive rule of Π is satisfied. Now one can prove:

Claim. Π and $(\mathcal{T}_\Pi, \Sigma_\Pi, q_\Pi)$ are polynomially equivalent.

($\Rightarrow$) Assume a $\Sigma_{\mathsf{EDB}}(\Pi)$-ABox $\mathcal{A}$ is given as an input to Π. If some $r \in \mathcal{F}$ is empty in $\mathcal{A}$, then output '$\mathcal{A} \not\models \Pi$'. This is correct since all goal rules of Π contain $Y_\mathcal{F}$ in their body. If no $r \in \mathcal{F}$ is empty in $\mathcal{A}$ and some $r \in \mathcal{F}$ is not functional, then output '$\mathcal{A} \models \Pi$'. This follows from the rules of type (b) in Π. Now assume no $r \in \mathcal{F}$ is empty in $\mathcal{A}$ and all $r \in \mathcal{F}$ are functional in $\mathcal{A}$. Let

$$\mathcal{A}' = \{\mathsf{True}(a), \mathsf{False}(b), \mathsf{Val}(a), \mathsf{Val}(b)\},$$

where we asume w.l.o.g. that a, b occur in $\mathcal{A}$. We show that $\mathcal{A} \models \Pi$ iff $\mathcal{T}_\Pi, \mathcal{A}' \models_{c(\Sigma_\Pi)} q_\Pi$.

Assume first that $\mathcal{A} \not\models \Pi$. Let $\mathcal{I}$ be a model of $\mathcal{A}$ that respects closed predicates $\Sigma_{\mathsf{EDB}}(\Pi)$ and satisfies no body of a goal rule in Π. Define $\mathcal{I}'$ in the same way as $\mathcal{I}$ except that $\mathsf{True}^{\mathcal{I}'} = \{a\}$, $\mathsf{False}^{\mathcal{I}'} = \{b\}$, $\mathsf{Val}^{\mathcal{I}'} = \{a, b\}$, and the extension of the fresh auxiliary predicates in $\mathcal{T}_\Pi$ and q_Π is defined as follows:

- For all $r \in \mathcal{F}$ let
$$s_r^{\mathcal{I}'} = \Delta^\mathcal{I} \times \mathsf{dom}(r^\mathcal{I}), \quad A_r^{\mathcal{I}'} = B_r^{\mathcal{I}'} = \Delta^\mathcal{I},$$
where $\mathsf{dom}(r^\mathcal{I})$ denotes the domain of $r^\mathcal{I}$.
- For all $1 \leq i \leq n$, let
$$s_i^{\mathcal{I}'} = (P_i^\mathcal{I} \times \{a\}) \cup ((\Delta^\mathcal{I} \setminus P_i^\mathcal{I}) \times \{b\}).$$

It is straightforward to see that $\mathcal{I}'$ is a model of $\mathcal{T}_\Pi$ and $\mathcal{A}'$ that respects closed predicates Σ_Π and so it remains to show that $\mathcal{I}' \not\models q_\Pi$. To this end it is sufficient to show that

1. No X^t with $\mathsf{goal}() \leftarrow X \in \Pi$ of type (c) is satisfiable in $\mathcal{I}'$;
2. $\mathcal{I}' \not\models \exists x \bigwedge_{1 \leq i \leq n} P_i^f(x)$.

(1.) holds since $P_i^{\mathcal{I}} = \{d \mid \mathcal{I}' \models P_i^t(d)\}$ by definition of $\mathcal{I}'$ and since X is not satisfied in $\mathcal{I}$ for any $\mathsf{goal}() \leftarrow X \in \Pi$ of type (c). (2.) holds since $P_i^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus \{d \mid \mathcal{I}' \models P_i^f(d)\}$ for all $1 \leq i \leq n$ and $\Delta^{\mathcal{I}} = P_1^{\mathcal{I}} \cup \cdots \cup P_n^{\mathcal{I}}$.

Assume now that $\mathcal{T}_\Pi, \mathcal{A}' \not\models_{c(\Sigma_\Pi)} q_\Pi$. Take a model $\mathcal{I}$ of $\mathcal{A}'$ that respects closed predicates Σ_Π and such that $\mathcal{I} \not\models q_\Pi$. Define a model $\mathcal{I}'$ by setting

$$P_i^{\mathcal{I}'} = \{d \mid \mathcal{I} \models P_i^t(d)\}$$

for $1 \leq i \leq n$. Since all $r \in \mathcal{F}$ are non-empty and functional in $\mathcal{A}'$ we have $\mathcal{I} \models \forall x \exists x_r \cdots \exists y_r q_{\mathcal{F}}$. To prove this, observe that

$$\mathcal{I} \models \top \sqsubseteq \exists s_r.(\exists r.A_r \sqcap \exists r.B_r),$$

for all $r \in \mathcal{F}$. By functionality of all $r \in \mathcal{F}$ in $\mathcal{A}'$, we obtain

$$\mathcal{I} \models \top \sqsubseteq \exists s_r.\exists r.(A_r \sqcap B_r),$$

for all $r \in \mathcal{F}$, as required. Thus (i) no X^t is satisfiable in $\mathcal{I}$ for any $\mathsf{goal}() \leftarrow X \in \Pi$ of type (c) and (ii) $\mathcal{I} \not\models \exists x \bigwedge_{1 \leq i \leq n} P_i^f(x)$. We show that $\mathcal{I}'$ satisfies the rule $P_1 \vee \cdots \vee P_n \leftarrow \top$ and no body of a goal rule in Π. The latter condition follows from (i) and the definition of $\mathcal{I}'$. For the first condition observe that since $\mathcal{I} \models \top \sqsubseteq \exists s_i.\mathsf{Val}$ for $1 \leq i \leq n$ and since $\mathsf{True}^{\mathcal{I}}, \mathsf{False}^{\mathcal{I}}$ is a partition of $\mathsf{Val}^{\mathcal{I}}$ we have $\mathcal{I} \models P_i^t(d)$ or $\mathcal{I} \models P_i^f(d)$ for all $d \in \Delta^{\mathcal{I}}$ and all $1 \leq i \leq n$. Thus $P_1^{\mathcal{I}'} \cup \cdots \cup P_n^{\mathcal{I}'} = \Delta^{\mathcal{I}}$ follows from (ii).

($\Leftarrow$) Assume a Σ_Π-ABox $\mathcal{A}$ is given as an input to Q_Π. There exists a model of $\mathcal{T}_\Pi$ and $\mathcal{A}$ that respects closed predicates Σ_Π iff (i) Val is nonempty in $\mathcal{A}$, (ii) $\mathsf{True}, \mathsf{False}$ are both contained in Val in $\mathcal{A}$, and (iii) all $r \in \mathcal{F}$ are nonempty in $\mathcal{A}$. Thus, output $() \in Q_\Pi(\mathcal{A})$ whenever (i), (ii), or (iii) is violated. Now assume (i), (ii), and (iii) hold. If some $r \in \mathcal{F}$ is non-functional, then $\mathcal{T}_\Pi, \mathcal{A} \not\models_{c(\Sigma_\Pi)} q_{\mathcal{F}}$ and so we output $() \notin Q_\Pi(\mathcal{A})$. Thus, assume in addition to (i), (ii) and (iii) that all $r \in \mathcal{F}$ are functional in $\mathcal{A}$.

If $\mathsf{True}, \mathsf{False}$ are not a partition of Val, then output $() \notin Q_\Pi(\mathcal{A})$ if True and False do not cover Val and output $() \in Q_\Pi(\mathcal{A})$ otherwise. It remains to consider the case in which (i), (ii), and (iii) hold, all $r \in \mathcal{F}$ are functional, and $\mathsf{True}, \mathsf{False}$ are a partition of Val. In this case let $\mathcal{A}'$ be the $\Sigma_{\mathsf{EDB}}(\Pi)$-reduct of $\mathcal{A}$. Similarly to the proof of ($\Rightarrow$) one can show that $\mathcal{A}' \models \Pi$ iff $\mathcal{T}_\Pi, \mathcal{A} \models_{c(\Sigma_\Pi)} q_\Pi$.

The modification of Q_Π to obtain an OMQC from $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BAQ})$ is straightforward, c.f. the proof of Theorem 10. We now show how to modify Q_Π to an equivalent query $Q_\Pi^{\text{DL-Lite}}$ from $(\text{DL-Lite}_{\mathcal{R}}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BtUCQ})$. First, to eliminate $\top$ on the right-hand-side of TBox inclusions in $\mathcal{T}_\Pi$, we replace any inclusion $\top \sqsubseteq C$ by the inclusions $A \sqsubseteq C$, $\exists r \sqsubseteq C$, and $\exists r^- \sqsubseteq C$ for any concept name $A \in \Sigma_\Pi$ and role name $r \in \Sigma_\Pi$. Secondly, we employ the standard encoding of qualified existential restrictions in DL-Lite$_{\mathcal{R}}$ by replacing exhaustively any $B \sqsubseteq \exists r.D$ by $B \sqsubseteq \exists s$, $\exists s^- \sqsubseteq A_D$, $A_D \sqsubseteq D$, and $s \sqsubseteq r$, where A_D is a fresh concept name and s is a fresh role name. Let $\mathcal{T}_\Pi^{\text{DL-Lite}}$ be the resulting TBox. Then $(\mathcal{T}_\Pi^{\text{DL-Lite}}, \Sigma_\Pi, q_\Pi)$ is as required. ❏

Theorem 14. For every OMQC in $(\text{DL-Lite}_{\mathsf{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BtCQ})$ there exists a polynomially equivalent OMCQ in $(\text{DL-Lite}_{\mathsf{core}}, \mathsf{N_C}, \mathrm{BtCQ})$.

Proof. Assume an OMQC $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ in $(\text{DL-Lite}_{\mathsf{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \mathrm{BtCQ})$ is given. The starting point to constructing the desired OMQC $(\mathcal{T}', \Sigma'_{\mathsf{A}}, \Sigma'_{\mathsf{C}}, q')$ from $(\text{DL-Lite}_{\mathsf{core}}, \mathsf{N_C}, \mathrm{BtCQ})$ is the observation that, when a role name is not mentioned in the TBox, then declaring it closed or open does not make a difference regarding query entailment. We can thus eliminate a closed role name r by making it open and removing it from the TBox. The latter is achieved by replacing $\exists r$ in the TBox with a fresh concept name A_r and $\exists r^-$ with a fresh concept name A_{r^-}. It remains to ensure that A_r and A_{r^-} represent the domain and range of r, that is, we find $A_r(a)$ in the ABox iff the ABox contains some assertion $r(a, b)$ and we find $A_{r^-}(a)$ in the ABox iff the ABox contains some assertion $r(b, a)$. This requires closing A_r and A_{r^-} and modifying both the TBox and the query.

Formally, denote by R the set of all r, r^- with r a role name in Σ_{C} and take for every $r \in R$ two fresh concept names A_r and A_{r^-}. Let

$$\Sigma'_{\mathsf{C}} = (\Sigma_{\mathsf{C}} \setminus \mathsf{N_R}) \cup \{A_r \mid r \in R\}, \quad \Sigma'_{\mathsf{A}} = \Sigma_{\mathsf{A}} \cup \{A_r \mid r \in R\}.$$

To define $\mathcal{T}'$, take fresh (open) role names s_r and (open) concept names E_r for every $r \in R$ and include in $\mathcal{T}'$ all inclusions $C \sqsubseteq D \in \mathcal{T}$ in which each $\exists r$ with $r \in R$ is replaced by A_r together with the following inclusions for every $r \in R$:

$$\exists r \sqsubseteq A_r, \quad A_r \sqsubseteq \exists s_r, \quad \exists s_r^- \sqsubseteq A_{r^-}, \quad \exists s_r^- \sqsubseteq E_{r^-}.$$

The first inclusion says that the domain and range of r is included in A_r and A_{r^-}, respectively. This achieves one half of our aim. For the second half, we need the query, that is, we will ensure that the domain and range of r is *exactly* A_r and A_{r^-}, respectively, unless our ABox $\mathcal{A}$ is such that $\mathcal{T}', \mathcal{A} \not\models q'$. A first step is done by inclusions two and three in the TBox, which axiomatize some basic behaviour of domains and ranges, namely that A_r is empty iff A_{r^-} is empty. Set

$$q' = q \wedge \bigwedge_{r \in R \cap \mathsf{sig}(q)} E_{r^-}(x_r^1) \wedge r(y_r^1, x_r^1) \wedge E_r(x_r^2) \wedge r(x_r^2, y_r^2).$$

To see how inclusions two, three, and four of $\mathcal{T}'$ and the additional conjuncts in q' play together, consider an ABox $\mathcal{A}$ that contains $A_{r^-}(a)$, but no assertion of the form $r(b, a)$. Then we find a model of $\mathcal{T}'$ and $\mathcal{A}$ where the BCQ $\exists x \exists y\, r(y, x) \wedge E_{r^-}(x)$ has no match, thus $\mathcal{T}', \mathcal{A} \not\models q'$.

Claim. $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ and $(\mathcal{T}', \Sigma'_{\mathsf{A}}, \Sigma'_{\mathsf{C}}, q')$ are polynomially equivalent.

($\Rightarrow$) Assume a Σ_{A}-ABox $\mathcal{A}$ is given as an input to $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$. Define a Σ'_{A}-ABox $\mathcal{A}'$ by adding

$$\{A_r(a) \mid r(a, b) \in \mathcal{A}\} \cup \{A_{r^-}(a) \mid r(b, a) \in \mathcal{A}\}$$

to $\mathcal{A}$. It is readily checked that $\mathcal{T}, \mathcal{A} \models_{c(\Sigma_\mathsf{C})} q$ iff $\mathcal{T}', \mathcal{A}' \models_{c(\Sigma_\mathsf{C})} q'$.

($\Leftarrow$) Assume a Σ'_A-ABox $\mathcal{A}'$ is given as an input to $(\mathcal{T}', \Sigma'_\mathsf{A}, \Sigma'_\mathsf{C}, q')$. Answer $\mathcal{T}', \mathcal{A}' \models_{c(\Sigma_\mathsf{C})} q'$ if $\mathcal{A}'$ is not consistent w.r.t. $\mathcal{T}'$ and closed Σ'_C (this problem is in AC^0 and, therefore, in PTIME, by Theorem 5). Otherwise answer $\mathcal{T}', \mathcal{A}' \not\models_{c(\Sigma_\mathsf{C})'} q'$ if there exists a $r \in R \cap \mathsf{sig}(q)$ and individual a such that $A_r(a) \in \mathcal{A}$ but there does not exist b with $r(a,b) \in \mathcal{A}$ or $A_{r^-}(a) \in \mathcal{A}$ but there does not exist b with $r(b,a) \in \mathcal{A}$. Otherwise, let $\mathcal{A}$ be obtained from $\mathcal{A}'$ by removing all assertions involving any A_r with $r \in R$ from $\mathcal{A}'$ and adding

$$\{r(a,b) \mid r \in R \setminus \mathsf{sig}(q), A_r(a) \in \mathcal{A}, A_{r^-}(b) \in \mathcal{A}\}$$

to $\mathcal{A}'$. It can be shown that $\mathcal{T}', \mathcal{A}' \models_{c(\Sigma'_\mathsf{C})} q'$ iff $\mathcal{T}, \mathcal{A} \models_{c(\Sigma_\mathsf{C})} q$.

Note that the query q' is not yet a tree UCQ. Let x be the root of q' and $B(x)$ a conjunct of q, where B is a basic concept. By adding $B \sqsubseteq \exists s_r$ to $\mathcal{T}'$ for every $r \in R$ such that $r \in \mathsf{sig}(q)$ or $r^- \in \mathsf{sig}(q)$ and adding $s_r(x, x_r^1)$ as well as $s_{r^-}(x, x_r^2)$ to q' we obtain an OMQC of the required form. ❑