Logical Methods in Computer Science
Volume 15, Issue 3, 2019, pp. 23:1–23:54
https://lmcs.episciences.org/

Submitted Sep. 05, 2018
Published Aug. 28, 2019

THE DATA COMPLEXITY OF ONTOLOGY-MEDIATED QUERIES WITH CLOSED PREDICATES

CARSTEN LUTZ, İNANÇ SEYLAN, AND FRANK WOLTER

Fachbereich Informatik, University of Bremen, Germany
e-mail address: clu@uni-bremen.de

ZF Friedrichshafen AG, Germany

Department of Computer Science, University of Liverpool, United Kingdom
e-mail address: wolter@liverpool.ac.uk

Abstract. We study the data complexity of ontology-mediated queries in which selected predicates can be closed (OMQCs), carrying out a non-uniform analysis of OMQCs in which the ontology is formulated in one of the lightweight description logics DL-Lite and $\mathcal{EL}$ or in the expressive description logic $\mathcal{ALCHI}$. We focus on separating tractable from non-tractable OMQCs. On the level of ontologies, we prove a dichotomy between FO-rewritable and coNP-complete for DL-Lite and between PTime and coNP-complete for $\mathcal{EL}$. We also show that in both cases, the meta problem to decide tractability is in PTime. On the level of OMQCs, we show that there is no dichotomy (unless NP equals PTime) if both concept and role names can be closed. For the case where only concept names can be closed, we tightly link the complexity of OMQC evaluation to the complexity of generalized surjective CSPs. We also identify a useful syntactic class of OMQCs based on DL-Lite$_\mathcal{R}$ that are guaranteed to be FO-rewritable.

1. Introduction

The aim of ontology-mediated querying (OMQ) is to facilitate querying incomplete and heterogeneous data by adding an ontology that provides domain knowledge [53, 13, 58]. To account for the incompleteness, OMQ typically adopts the open world assumption (OWA). In some applications, though, there are parts of the data for which the closed world assumption (CWA) is more appropriate. For example, in a data integration application some data may have been extracted from the web and thus be significantly incomplete, suggesting the OWA, while other data may come from curated relational database systems that are known to be complete, thus suggesting the CWA. As an extreme case, one may even use an ontology on top of complete data and thus treat all predicates in the data under the CWA whereas additional predicates that are provided by the ontology for more convenient querying are treated under the OWA [36]. It is argued in [10] that a similar situation emerges when only a subset of the predicates from a complete database is published for privacy reasons, with an ontology linking the 'visible' and 'invisible' predicates. When admitting both types of predicates in queries (e.g. to analyze which parts of the private data can be recovered), the

DOI:10.23638/LMCS-15(3:23)2019

CWA is appropriate for the visible predicates while OWA is required for the invisible ones. A concrete example of mixed OWA and CWA is given in [45], namely querying geo-databases such as OpenStreetMap in which the geo data is complete, thus suggesting the CWA, while annotations are incomplete and suggest the OWA.

In this article, we are interested in ontologies formulated in a description logic (DL). In the area of DLs, quite a number of proposals have been brought forward on how to implement a partial CWA, some of them fairly complex [22, 24, 31, 50, 56]. In OMQ, a particularly straightforward and natural aproach is to simply distinguish between OWA predicates and CWA predicates, as suggested also by the motivating examples given above. The interpretation of CWA predicates is then fixed to what is explicitly stated in the data while OWA predicates can be interpreted as any extension thereof [45].

Making the CWA for some predicates, from now on referred to as *closing* the predicates, has a strong effect on the complexity of query evaluation. We generally concentrate on data complexity where only the data is considered an input while the actual query and ontology are assumed to be fixed; see [51] for an analysis of combined complexity in the presence of closed predicates. The (data) complexity of evaluating (rather restricted forms of) conjunctive queries (CQs) becomes coNP-hard already when ontologies are formulated in inexpressive DLs such as DL-Lite$_{\mathsf{core}}$ and $\mathcal{EL}$ [29] whereas CQ evaluation without closed predicates is FO-rewritable and thus in AC^0 for the former and in PTime for the latter [20, 5, 34]. Here, FO-rewritability is meant in the usual sense of ontology-mediated querying [20, 38, 12, 11], that is, we can find a first-order (FO) query that is equivalent to the original OMQ evaluated w.r.t. the ontology. Since intractability comes so quickly, it is not very informative to analyze complexity on the level of logics, as in the complexity statements just made; instead, one would like to know whether closing a *concrete set of predicates* results in intractability for the *concrete ontology used in an application* or for the *concrete combination of ontology and query that is used*. If it does not, then one should indeed close the predicates since this may result in additional (that is, more complete) answers to queries and additionally enables the use of more expressive query languages for the closed part of the vocabulary. Otherwise, one can resort to full OWA as an approximation semantics for querying or live with the fact that evaluating the concrete query at hand is costly.

Such a *non-uniform analysis* has been carried out in two different ways in [47, 33] and in [14] for classical OMQ (that is, without closed predicates) and expressive DLs such as $\mathcal{ALC}$ which give rise to coNP data complexity even when all predicates are open. The former references aim to classify the complexity of ontologies, *quantifying over the actual query*: evaluating queries formulated in a query language $\mathcal{Q}$ is in PTime for an ontology $\mathcal{O}$ if every query from $\mathcal{Q}$ can be evaluated in PTime w.r.t. $\mathcal{O}$ and it is coNP-hard if there is at least one Boolean query from $\mathcal{Q}$ that is coNP-hard to evaluate w.r.t. $\mathcal{O}$. In the latter reference, an even more fine-grained approach is taken where the query is not quantified away and thus the aim is 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 (where $\cdot_{\mathsf{A}}$ stands for 'ABox'), and q an actual query. In both cases, a close connection to the complexity of constraint satisfaction problems (CSPs) with fixed template is identified. Given a relational structure $\mathcal{I}$, called a *template*, the problem to decide for another relational structure $\mathcal{J}$ whether there is a homomorphism from $\mathcal{J}$ to $\mathcal{I}$ is called the *constraint satisfaction problem defined by* $\mathcal{I}$, and denoted $\mathrm{CSP}(\mathcal{I})$. Investigating the computational complexity of $\mathrm{CSP}(\mathcal{I})$ is an active field of research that brings together algebra, graph theory, and logic [26, 19, 40, 18, 59]. The connection between the complexity of OMQs and CSPs has proved to be very fruitful as it enables the transfer of deep results available for CSPs to OMQ. In fact, it has been used to obtain complexity dichotomies and results on the rewritability of OMQs into more conventional database languages [14, 47, 33, 27].

The aim of this acticle is to carry out both types of analyses, the *quantified query case* and the *fixed query case*, for OMQs with closed predicates and for DLs ranging from the simple Horn DLs DL-Lite and $\mathcal{EL}$ to the expressive DL $\mathcal{ALCHI}$. As the actual queries, we use CQs, unions thereof (UCQs), and several relevant restrictions of CQs and UCQs such as unary tree-shaped CQs, both in the directed and in the undirected sense. Recall that DL-Lite and $\mathcal{EL}$ are underpinning the profiles OWL 2 QL and OWL 2 EL of the prominent OWL 2 ontology language while $\mathcal{ALCHI}$ is related to OWL 2 DL [20, 5, 7]. As a starting point and general backdrop of our investigations, we prove that query evaluation is in coNP when the ontology is formulated in $\mathcal{ALCHI}$, the actual query is a UCQ, and predicates can be closed. Note that this bound is not a consequence of results on ontology-mediated querying in description logics with nominals [52] because nominals are part of the ontology and thus their number is bounded by a constant while closing a predicate corresponds to considering a disjunction of nominals whose number is only bounded by the size of the data (that is, the input size).

In the quantified query case, we aim to classify all *TBoxes with closed predicates*, that is, all pairs $(\mathcal{T}, \Sigma_\mathsf{C})$ where $\mathcal{T}$ is a TBox formulated in the DL under consideration, representing the ontology, and Σ_C is the set of predicates (concept and role names) that are closed; all other predicates are interpreted under the OWA. For the DL-Lite$_\mathcal{R}$ dialect of DL-Lite and for $\mathcal{EL}$, we obtain characterizations that separate the tractable cases from the intractable ones and map out the frontier of tractability in a transparent way (and also cover the fragment DL-Lite$_{\text{core}}$ of DL-Lite$_\mathcal{R}$). They essentially state that evaluating tree-shaped CQs is coNP-hard w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ if $\mathcal{T}$ entails certain concept inclusions that mix open and closed predicates in a problematic way while otherwise UCQ evaluation w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ is tractable, that is, FO-rewritable and in PTime, respectively. Notably, this yields a dichotomy between AC^0 and coNP for DL-Lite$_\mathcal{R}$ TBoxes with closed predicates and between PTime and coNP for $\mathcal{EL}$ TBoxes with closed predicates. It is remarkable that such a dichotomy can be obtained by a rather direct analysis, especially when contrasted with the case of expressive DLs such as $\mathcal{ALC}$ without closed predicates for which a dichotomy between PTime and coNP is equivalent to the dichotomy between PTime and NP for CSPs, a long-standing open problem that was known as the Feder-Vardi conjecture and has been settled only very recently [18, 59]. The proofs are a bit simpler in the case of DL-Lite$_\mathcal{R}$ while they involve the careful use of a certain version of the Craig interpolation property in the $\mathcal{EL}$ case. The characterizations also allow us to prove that it can be decided in PTime whether a given TBox with closed predicates is tractable or coNP-complete (assuming PTime $\neq$ NP), which we from now on call the *meta problem*. It turns out that the tractable cases are precisely those in which closing the predicates in Σ_C does not have an effect on the answers to any query (unless the data is inconsistent with the TBox). This can be interpreted as showing that, in the quantified query case, OMQ with closed predicates is inherently intractable.[1]

Fortunately, this is not true in the fixed query case where we aim to classify all *ontology-mediated queries with closed predicates (OMQCs)* which take the form $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ where $\mathcal{T}$, Σ_A, and q are as in classical OMQs and $\Sigma_\mathsf{C} \subseteq \Sigma_\mathsf{A}$ is a set of closed predicates. Interestingly, switching to fixed queries results in CSPs reentering the picture. While classifying the

[1]It is observed in [45] that this is not the case for the extension $\mathcal{ELI}$ of $\mathcal{EL}$ with inverse roles.

complexity of classical OMQs based on expressive DLs corresponds to classifying standard CSPs, we show that classifying OMQCs is tightly linked to the classification of *generalized surjective CSPs*. Surjective CSPs are defined exactly like standard CSPs except that homomorphisms into the template 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 defined by a template with only six elements whose complexity is not understood [15] while there are no such open cases for standard CSPs. The complexity of surjective CSPs is subject to significant research activities [15, 23] and it appears to be a widely open question whether a dichotomy between PTime and NP holds for the complexity of surjective CSPs. A *generalized* surjective CSP is defined by a finite set Γ of templates rather than by a single template and the problem is to decide whether there is a surjective homomorphism from the input structure to some interpretation in Γ. In the non-surjective case, every generalized CSP can be translated into an equivalent non-generalized CSP [28]. In the surjective case, such a translation is not known. In this part, we consider 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 Boolean UCQs in which all CQs are tree-shaped (BtUCQs). Our result then is that there is a dichotomy between PTime and coNP for such OMQs if and only if there is a dichotomy between PTime and NP for generalized surjective CSPs, a question that is wide open. We find it remarkable that, consequently, 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 be formulated in DL-Lite or in $\mathcal{EL}$ (and queries are BtUCQs). By Ladner's theorem, this precludes the existence of a 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 also show that the meta problem is undecidable.

Our results show that there are many natural tractable OMQs without closed predicates that become intractable when predicates are closed. As a final contribution, we identify a family of OMQC where tractability, and in fact FO-rewritability, is always guaranteed. We obtain this class by using DL-Lite$_{\mathcal{R}}$ as the ontology language, unions of quantifier-free CQs as the query language, and imposing the additional restriction that the ontology contains no role inclusion which states that an open role is contained in a closed one. We believe that this class of OMQCs is relevant for practical applications. We also prove that the restriction on RIs is needed for tractability by showing that dropping it gives rise to OMQCs that are coNP-hard.

This article is structured as follows. In Section 3, we introduce description logics, relevant query languages, and ontology-mediated querying with and without closed predicates. We also observe that one can assume w.l.o.g. that all predicates that occur in the data are closed and that UCQs using open predicates can be combined with FO queries using closed predicates without an impact on the complexity of query evaluation. In Section 4, we prove that UCQ evaluation mediated by $\mathcal{ALCHI}$ TBoxes with closed predicates is always in coNP. In Section 5, we establish the characterizations for the quantified query case and prove the announced complexity dichotomies. In Section 6, we show that it is decidable in PTime whether a given TBox with closed predicates is tractable. We then switch to the case of fixed queries. In Section 7, we establish the link between OMQCs with closed concept names to surjective CSPs and in Section 8 we link the general case where also role names can be closed to the complexity of NP Turing machines and prove that the meta problem is undecidable. In Section 9, we show that evaluating UCQs without quantified variables is FO-rewritable for DL-Lite$_\mathcal{R}$ TBoxes in which no open role is included in a closed role.

2. Related Work

The present article combines and extends the conference publications [44, 45]. Classifications of the complexity of OMQs without closed predicates based on expressive DLs have been studied in [47, 33] in the quantified query case and in [14] in the fixed query case. The combined complexity of ontology-mediated querying with closed predicates has been investigated in [51]. Among other things, it is shown there that the combined complexity of evaluating OMQCs is 2ExpTime-complete when ontologies are formulated in DL-Lite$_\mathcal{R}$ or in $\mathcal{EL}$ and the actual queries are UCQs. The rewritability of OMQCs into disjunctive datalog with negation as failure is considered in [2] and it is shown that a polynomial rewriting is always possible when the ontology is formulated in $\mathcal{ALCHIO}$ and the actual query is of the form $A(x)$, A a concept name.

The subject of [10] is database querying when only a subset of the relations in the schema is visible and the data is subject to constraints, which in its 'instance-level version' is essentially identical to evaluating OMQCs. Among other results, it is proved (stated in our terminology) that when the ontology is formulated in the guarded negation fragment of first-order logic (GNFO) and the actual query is a UCQ, then the combined complexity of evaluating OMQCs is 2ExpTime-complete. The lower bound already applies when the ontology is a set of inclusion dependencies or a set of linear existential rules (which subsume inclusion dependencies). Moreover, there are OMQCs based on inclusion dependencies and UCQs that are ExpTime-hard in data complexity. These results are completemented by the observation from [9] that there are PSpace-hard OMQCs where the ontology is a set of linear existential rules and the actual query Boolean and atomic. It is interesting to contrast the latter two results with our coNP upper bound for $\mathcal{ALCHI}$ and UCQs.

Another related area is the study of combinations of the open and closed world assumption in data exchange [43]. In data exchange one usually assumes an open-world semantics according to which it is possible to extend instances of target schemas in an arbitrary way [4]. In an alternative closed-world semantics approach one only allows to add as much data as needed to the target to satisfy constraints of the schema mapping [32]. In [43], a mixed approach is proposed: one can designate different attributes of target schemas as open or closed. Although similar in spirit to ontology-based data access with closed predicates, the techniques required to analyze the mixed approach to data exchange appear to be very different from those developed in this paper.

More vaguely related to our setup are so-called 'nominal schemas' and 'closed variables' in ontologies that are sets of existential rules, see [41, 42] and [3], respectively. In both cases, the idea is that certain object identifiers (nominals or variables) can only be bound to individuals from the ABox, but not to elements of a model that are introduced by existential quantifiers. When disjunction is not present in the ontology language under consideration, which is the main focus of the present article, then the expressive power of these formalisms is orthogonal to ours. In the presence of disjunction, nominal schemes and closed variables can simulate closed predicates.

3. Preliminaries

We introduce description logics, relevant query languages, and ontology-mediated querying with and without closed predicates. We also observe that one can combine UCQs on open and closed predicates with full first-order queries on closed predicates without adverse effects on the decidability or complexity of query evaluation.

3.1. **Description Logics.**

For a fully detailed introduction to DLs, we refer the reader to [6, 8]. Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be countably infinite sets of *concept names*, *role names*, and *individual names*. An *inverse role* has the form r^- with r a role name. A *role* is a role name or an inverse role. We set $(r^-)^- = r$, for any role name r. We use three *concept languages* in this article. $\mathcal{ALCI}$ *concepts* are defined by the rule

$$C, D := A \mid \top \mid \neg C \mid C \sqcap D \mid \exists r.C \mid \exists r^-.C$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. The constructor $\exists r.C$ is called a *qualified existential restriction*. We use standard abbreviations and write, for example, $C \sqcup D$ for $\neg(\neg C \sqcap \neg D)$ and $\forall r.C$ for $\neg\exists r.\neg C$. *DL-Lite$_{core}$* (or *basic*) *concepts* are defined by the rule

$$B := A \mid \exists r.\top \mid \exists r^-.\top$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. We often use $\exists r$ as shorthand for the concept $\exists r.\top$. $\mathcal{EL}$ *concepts* C are defined by the rule

$$C := A \mid \top \mid \exists r.C$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. Thus, DL-Lite$_{\text{core}}$ and $\mathcal{EL}$ are both fragments of $\mathcal{ALCI}$. Note that DL-Lite$_{\text{core}}$ admits inverse roles but no qualified existential restrictions and $\mathcal{EL}$ admits qualified existential restructions but no inverse roles.

In description logic, ontologies are constructed using concept inclusions and potentially also role inclusions. An $\mathcal{ALCI}$ *concept inclusion (CI)* takes the form $C \sqsubseteq D$ with C, D $\mathcal{ALCI}$ concepts and $\mathcal{EL}$ CIs are defined accordingly. A *DL-Lite$_{core}$ CI* takes the form $B_1 \sqsubseteq B_2$ or $B_1 \sqsubseteq \neg B_2$ with B_1, B_2 basic concepts. For any of these three concept languages $\mathcal{L}$, an $\mathcal{L}$ *TBox* is a finite set of $\mathcal{L}$ CIs. A *role inclusion (RI)* takes the form $r \sqsubseteq s$, where r, s are roles. A *DL-Lite$_\mathcal{R}$ TBox* is a finite set of DL-Lite$_{\text{core}}$ CIs and RIs and an $\mathcal{ALCHI}$ *TBox* is a finite set of $\mathcal{ALCI}$ CIs and RIs.

In description logic, data are stored in *ABoxes* $\mathcal{A}$ which are finite sets of *concept assertions* $A(a)$ and *role assertions* $r(a, b)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. For a role name r, we sometimes write $r^-(a, b) \in \mathcal{A}$ for $r(b, a) \in \mathcal{A}$. We use $\mathsf{Ind}(\mathcal{A})$ to denote the set of individual names used in the ABox $\mathcal{A}$.

DLs are interpreted in standard first-order interpretations $\mathcal{I}$ presented as a pair $(\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set called the *domain* of $\mathcal{I}$ and $\cdot^\mathcal{I}$ is a function that maps each concept name A to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$ and each role name r to a binary relation $r^\mathcal{I}$ on $\Delta^\mathcal{I}$. The extension of $\cdot^\mathcal{I}$ to roles and $\mathcal{ALCI}$ concepts is defined in Table 1. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, a RI $r \sqsubseteq s$ 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 name assumption (SNA) which implies the unique name assumption. An interpretation is a *model of a TBox* $\mathcal{T}$ if it satisfies all inclusions in $\mathcal{T}$ and a *model of an ABox* $\mathcal{A}$ if it satisfies all assertions in $\mathcal{A}$. A concept C is *satisfiable w.r.t. a TBox* $\mathcal{T}$ if there exists a model $\mathcal{I}$ of $\mathcal{T}$ with $C^\mathcal{I} \neq \emptyset$. As usual, we write $\mathcal{T} \models C \sqsubseteq D$ ($\mathcal{T} \models r \sqsubseteq s$) if every model of $\mathcal{T}$ satisfies the CI $C \sqsubseteq D$ (resp. RI $r \sqsubseteq s$).

$$\begin{array}{rcl}
(r^-)^\mathcal{I} & = & \{(e,d) \mid (d,e) \in r^\mathcal{I}\} \\
\top^\mathcal{I} & = & \Delta^\mathcal{I} \\
(\neg C)^\mathcal{I} & = & \Delta^\mathcal{I} \setminus C^\mathcal{I} \\
(C \sqcap D)^\mathcal{I} & = & C^\mathcal{I} \cap D^\mathcal{I} \\
(\exists r.C)^\mathcal{I} & = & \{d \in \Delta^\mathcal{I} \mid \text{there exists } e \in \Delta^\mathcal{I} \text{ such that } (d,e) \in r^\mathcal{I} \text{ and } e \in C^\mathcal{I}\}
\end{array}$$

Table 1: Semantics of roles and $\mathcal{ALCI}$ concepts

A *predicate* is a concept or role name. A *signature* Σ is a finite set of predicates. We use $\mathsf{sig}(C)$ to denote the set of predicates that occur in the concept C and likewise for other syntactic objects such as TBoxes and ABoxes. An ABox is a Σ-*ABox* if it uses only predicates from Σ. We denote by $\mathsf{sub}(C)$ the set of subconcepts of the concept C and by $\mathsf{sub}(\mathcal{T})$ the set of subconcepts of concepts that occur in the TBox $\mathcal{T}$. The *size* of any syntactic object O, denoted $|O|$, is the number of symbols needed to write it with concept, role, and individual names viewed as a single symbol.

It will sometimes be convenient to regard interpretations as ABoxes and vice versa. For an ABox $\mathcal{A}$, the *interpretation $\mathcal{I}_\mathcal{A}$ corresponding to $\mathcal{A}$* is defined as follows:

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

Conversely, every interpretation $\mathcal{I}$ defines the (possibly infinite) ABox $\mathcal{A}_\mathcal{I}$ in which we regard the elements of the domain $\Delta^\mathcal{I}$ of $\mathcal{I}$ as individual names and let $A(d) \in \mathcal{A}_\mathcal{I}$ if $d \in A^\mathcal{I}$ and $r(d,d') \in \mathcal{A}_\mathcal{I}$ if $(d,d') \in r^\mathcal{I}$.

A *homomorphism* h from an interpretation $\mathcal{I}_1$ to an interpretation $\mathcal{I}_2$ is a mapping h from $\Delta^{\mathcal{I}_1}$ to $\Delta^{\mathcal{I}_2}$ such that $d \in A^{\mathcal{I}_1}$ implies $h(d) \in A^{\mathcal{I}_2}$ for all $A \in \mathsf{N_C}$ and $d \in \Delta^{\mathcal{I}_1}$, and $(d,d') \in r^{\mathcal{I}_1}$ implies $(h(d),h(d')) \in r^{\mathcal{I}_2}$ for all $r \in \mathsf{N_R}$ and $d,d' \in \Delta^{\mathcal{I}_1}$. We say that h *preserves* a set $N \subseteq \mathsf{N_I}$ of individual names if $h(a) = a$ for all $a \in N$. The *restriction* $\mathcal{I}|_D$ of an interpretation $\mathcal{I}$ to a non-empty subset D of $\Delta^\mathcal{I}$ is defined by setting $\Delta^{\mathcal{I}|_D} = D$, $A^{\mathcal{I}|_D} = A^\mathcal{I} \cap D$, for all $A \in \mathsf{N_C}$, and $r^{\mathcal{I}|_D} = r^\mathcal{I} \cap (D \times D)$ for all $r \in \mathsf{N_R}$. The Σ-*reduct* $\mathcal{J}$ of an interpretation $\mathcal{I}$ is obtained from $\mathcal{I}$ by setting $P^\mathcal{J} = P^\mathcal{I}$ for all predicates $P \in \Sigma$ and $P^\mathcal{J} = \emptyset$ for all predicates $P \notin \Sigma$.

3.2. **Query Languages.** The query languages used in this article are fragments of first-order logic using predicates of arity one and two only. Fix a countably infinite set $\mathsf{N_V}$ of *variables*. A *first-order query (FOQ)* $q(\vec{x})$ is a first-order formula whose free variables are contained in $\vec{x}$ and that is constructed from atoms $A(x)$ and $r(x,y)$ using conjunction, negation, disjunction, and existential quantification, where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. The variables in $\vec{x}$ are the *answer variables* of $q(\vec{x})$. The *arity* of $q(\vec{x})$ is defined as the length of $\vec{x}$ and a FOQ of arity 0 is called *Boolean*. If the answer variables $\vec{x}$ of a query $q(\vec{x})$ are not relevant, we simply write q for $q(\vec{x})$. An *assignment π in an interpretation $\mathcal{I}$* is a mapping from $\mathsf{N_V}$ into $\Delta^\mathcal{I}$. A tuple $\vec{a} = a_1,\ldots,a_n$ of individual names in $\Delta^\mathcal{I}$ is an *answer to $q(\vec{x})$ in $\mathcal{I}$* if there exists an assignment π in $\mathcal{I}$ such that $\mathcal{I} \models_\pi q$ (in the standard first-order sense) and $\pi(x_i) = a_i$ for $1 \leq i \leq n$. In this case, we write $\mathcal{I} \models q(\vec{a})$.

A *conjunctive query (CQ)* is a FOQ in prenex normal form that uses no operators except conjunction and existential quantification. A *union of CQs (UCQ)* is a disjunction of CQs with the same answer variables. Every CQ q can be viewed as an ABox $\mathcal{A}_q$ by regarding the variables of q as individual names.

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

(1) the directed graph $G_q = (V_q, E_q)$ is a tree with root x, where V_q is the set of variables used in q and E_q contains an edge (x_1, x_2) whenever there is an atom $r(x_1, x_2)$ in q;
(2) if $r(x, y), s(x, y)$ are conjuncts of $q(x)$ then $r = s$.

We sometimes regard a dtCQ q as a $\mathcal{EL}$ concept C_q in the natural way such that for every interpretation $\mathcal{I}$ and $a \in \Delta^{\mathcal{I}}$, $\mathcal{I} \models q(a)$ iff $a \in C_q^{\mathcal{I}}$. Conversely, we denote by q_C the natural dtCQ corresponding to the $\mathcal{EL}$ concept C such that $\mathcal{I} \models q_C(a)$ iff $a \in C^{\mathcal{I}}$ holds for all interpretations $\mathcal{I}$ and $a \in \Delta^{\mathcal{I}}$. It will be convenient to not always strictly distinguish between C and q_C and denote the query q_C by C.

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

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

Similarly to dtCQs, tCQs can be regarded as concepts in the extension $\mathcal{ELI}$ of $\mathcal{EL}$ with inverse roles, see [8]. We use the same notation as for dtCQs.

3.3. **TBoxes and Ontology-Mediated Queries with Closed Predicates.** As explained in the introduction, our central objects of study are TBoxes with closed predicates in the quantified query case and ontology-mediated queries with closed predicates in the fixed query case.

A *TBox with closed predicates* is a pair $(\mathcal{T}, \Sigma_{\mathsf{C}})$ with $\mathcal{T}$ a TBox and Σ_{C} a set of *closed predicates*. An *ontology-mediated query with closed predicates (OMQC)* takes the form $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ where $\mathcal{T}$ is a TBox, Σ_{A} an *ABox signature* which gives the set of predicates that can be used in ABoxes, $\Sigma_{\mathsf{C}} \subseteq \Sigma_{\mathsf{A}}$ a set of *closed predicates*, and q a query (such as a UCQ). The *arity* of Q is defined as the arity of q. If $\Sigma_{\mathsf{A}} = \mathsf{N_C} \cup \mathsf{N_R}$, then we omit Σ_{A} and write $(\mathcal{T}, \Sigma_{\mathsf{C}}, q)$ for $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$. Note that when $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ is an OMQC, then $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is a TBox with closed predicates. When studying TBoxes with closed predicates (in the quantified query case), we generally do not restrict the ABox signature.

The semantics of OMQCs is as follows. We say that 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,

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

Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ be an OMQC and $\mathcal{A}$ a Σ_{A}-ABox. A tuple $\vec{a}$ of elements from $\mathsf{Ind}(\mathcal{A})$, denoted by $\vec{a} \in \mathsf{Ind}(\mathcal{A})$ for convenience, is a *certain answer to Q on $\mathcal{A}$*, written $\mathcal{A} \models Q(\vec{a})$, if $\mathcal{I} \models q(\vec{a})$ for all models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respect Σ_{C}. The *evaluation problem* for Q is the problem to decide, given a Σ_{A}-ABox $\mathcal{A}$ and a tuple $\vec{a} \in \mathsf{Ind}(\mathcal{A})$, whether $\mathcal{A} \models Q(\vec{a})$. Note that this problem parallels the evaluation problem for CQs and other standard query language, but with CQs replaced by OMQCs.

An OMQC $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ with answer variables $\vec{x}$ is *FO-rewritable* if there is a first-order formula $p(\vec{x})$, called an *FO-rewriting of Q*, such that for all Σ_{A}-ABoxes $\mathcal{A}$ and all $\vec{a} \in \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{I}_\mathcal{A} \models p(\vec{a})$ iff $\mathcal{A} \models Q(\vec{a})$. We remind the reader that the query evaluation problem for Q is in AC^0 when Q is FO-rewritable.

Example 3.1. Consider $\mathcal{T} = \{A \sqsubseteq \exists r.B\}$ and $q(x) = \exists y\, r(y,x)$. Let $Q_0 = (\mathcal{T}, \emptyset, q(x))$ be an OMQC without closed predicates and let $Q_1 = (\mathcal{T}, \Sigma_\mathsf{C}, q(x))$ be the corresponding OMQC with closed predicates $\Sigma_\mathsf{C} = \{B\}$. Let $\mathcal{A} = \{A(a), B(b)\}$. Then $\mathcal{A} \not\models Q_0(b)$ since one can define a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ in which $(a,d) \in r^\mathcal{I}$ and $d \in B^\mathcal{I}$ for a fresh element d. However, $\mathcal{A} \models Q_1(b)$ since $B \in \Sigma_\mathsf{C}$. Note that $q(x)$ is an FO-rewriting of Q_0. The FO-rewriting of Q_1 is more complicated and given by

$$q(x) \vee (\exists y\, A(y) \wedge B(x) \wedge \forall y\, (B(y) \rightarrow y = x)) \vee (\exists y\, A(y) \wedge \neg\exists y\, B(y))$$

The second disjunct captures answers for ABoxes in which one has to make x an r-successor of some y because only x satisfies B and the third disjunct captures answers for ABoxes in which there is no common model of $\mathcal{T}$ and the ABox that respects Σ_C.

An *OMQC language* is a triple $(\mathcal{L}, \Sigma, \mathcal{Q})$ with $\mathcal{L}$ a TBox language (such as DL-Lite$_\mathcal{R}$, $\mathcal{EL}$, or $\mathcal{ALCHI}$), Σ a set of predicates (such as $\mathsf{N_C} \cup \mathsf{N_R}$, $\mathsf{N_C}$, or the empty set) from which the closed predicated in OMQCs must be taken, and $\mathcal{Q}$ a query language (such as UCQ or CQ). Then $(\mathcal{L}, \Sigma, \mathcal{Q})$ 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 $q \in \mathcal{Q}$. Note that for $\Sigma = \emptyset$ we obtain the standard languages of ontology-mediated queries without closed predicates [14].

In the quantified query case, we aim to classify the complexity of all TBoxes with closed predicates $(\mathcal{T}, \Sigma_\mathsf{C})$ where $\mathcal{T}$ is formulated in a DL of interest. More precisely, for a query language $\mathcal{Q}$ we say that

- $\mathcal{Q}$ *evaluation w.r.t.* $(\mathcal{T}, \Sigma_\mathsf{C})$ *is in* PTime if for every $q \in \mathcal{Q}$, the evaluation problem for $(\mathcal{T}, \Sigma_\mathsf{C}, q)$ is in PTime;
- $\mathcal{Q}$ *evaluation w.r.t.* $(\mathcal{T}, \Sigma_\mathsf{C})$ *is* coNP*-hard* if there exists $q \in \mathcal{Q}$ such that the evaluation problem for $(\mathcal{T}, \Sigma_\mathsf{C}, q)$ is coNP-hard;
- $\mathcal{Q}$ *evaluation w.r.t.* $(\mathcal{T}, \Sigma_\mathsf{C})$ *is FO-rewritable* if for every $q \in \mathcal{Q}$, the OMQC $(\mathcal{T}, \Sigma_\mathsf{C}, q)$ is FO-rewritable.

In the fixed query case, we aim to classify the complexity of all OMQCs from some OMQC language, in the standard sense. We remind the reader that without closed predicates the complexity of query evaluation is well understood. In fact,

- every OMQC in $(\text{DL-Lite}_\mathcal{R}, \emptyset, \text{UCQ})$ is FO-rewritable [20];
- the evaluation problem for every OMQC in $(\mathcal{EL}, \emptyset, \text{UCQ})$ is in PTime (and there are PTime-hard OMQCs in $(\mathcal{EL}, \emptyset, \text{dtCQ})$) [21, 39]; and
- the evaluation problem for every OMQC in $(\mathcal{ALCHI}, \emptyset, \text{UCQ})$ is in coNP (and there are coNP-hard OMQCs in $(\mathcal{ALCI}, \emptyset, \text{dtCQ})$) [34, 52, 55, 21].

We will often have to deal with ABoxes that contradict the TBox given that certain predicates are closed. We say that an ABox $\mathcal{A}$ is *consistent w.r.t.* $(\mathcal{T}, \Sigma_\mathsf{C})$ if there is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects Σ_C. We further say that *ABox consistency is FO-rewritable* for $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C})$ if there is a Boolean FOQ q such that for all Σ_A-ABoxes $\mathcal{A}$, $\mathcal{I}_\mathcal{A} \models q$ iff $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$. Note that if an ABox is consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$, then it is consistent w.r.t. $(\mathcal{T}, \emptyset)$. The converse does not hold. For example, if $\mathcal{T} = \{A \sqsubseteq B\}$ and $\Sigma_\mathsf{C} = \{B\}$, then $\mathcal{A} = \{A(a)\}$ is not consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ but $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \emptyset)$.

Note that a CI $C \sqsubseteq D$ that uses *only* closed predicates acts as an integrity constraint in the standard database sense [1]. As an example, consider $\mathcal{T} = \{A \sqsubseteq B\}$ and $\Sigma_\mathsf{C} = \{A, B\}$. Then $(\mathcal{T}, \Sigma_\mathsf{C})$ imposes the integrity constraint that if $A(a)$ is contained in an ABox, then so must be $B(a)$. In particular, an ABox $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ iff $\mathcal{A}$ satisfies this integrity constraint. For ABoxes $\mathcal{A}$ that are consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$, $(\mathcal{T}, \Sigma_{\mathsf{C}})$ has no further effect on query answers. In a DL context, integrity constraints are discussed in [22, 24, 48, 49, 50].

3.4. **Basic Observations on OMQCs.** We first show that for DLs that support role inclusions, any OMQC is equivalent to an OMQC in which the ABox signature and the set of closed predicates coincide. This setup was called *DBoxes* in [36, 29]. Assume OMQCs Q_1 and Q_2 have the same arity and ABox signature Σ_{A}. Then Q_1 and Q_2 are *equivalent* if for all Σ_{A}-ABoxes $\mathcal{A}$ and all tuples $\vec{a}$ in $\mathsf{Ind}(\mathcal{A})$, $\mathcal{A} \models Q_1(\vec{a})$ iff $\mathcal{A} \models Q_2(\vec{a})$. 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 article are canonical.

Theorem 3.2. *Let $\mathcal{L} \in \{$DL-Lite$_{\mathcal{R}}, \mathcal{ALCHI}\}$ and $\mathcal{Q}$ be a canonical class of UCQs. 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 OMQC $Q' = (\mathcal{T}', \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{A}}, q')$ with $\mathcal{T}' \in \mathcal{L}$ and $q' \in \mathcal{Q}$.*

Proof. Let $\mathcal{L} \in \{\text{DL-Lite}_{\mathcal{R}}, \mathcal{ALCHI}\}$ and 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 predicate $P \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$, we take a fresh predicate P' of the same arity (if P is a concept name, then P' is a concept name, and if P is a role name, then P' is a role name). Let $\mathcal{T}'$ be the resulting TBox when all $P \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ are replaced by P' and the inclusion $P \sqsubseteq P'$ is added, for each $P \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$. Denote by q' the resulting query when every $P \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$ in q is replaced by P'. We show that $Q' = (\mathcal{T}', \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{A}}, q')$ is equivalent to Q.

First let $\mathcal{A}$ be a Σ_{A}-ABox with $\mathcal{A} \not\models Q(\vec{a})$. Then there is a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ 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 predicates unchanged. It can be verified that $\mathcal{I}'$ is a model of $\mathcal{T}'$ and $\mathcal{A}$ that respects closed predicates Σ_{A} such that $\mathcal{I}' \not\models q'(\vec{a})$. Thus, $\mathcal{A} \not\models Q'(\vec{a})$.

Conversely, let $\mathcal{A}$ be a Σ_{A}-ABox such that $\mathcal{A} \not\models Q'(\vec{a})$. 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

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

and leaving the interpretation of the remaining predicates 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 $\mathcal{I} \not\models q(\vec{a})$. Thus, $\mathcal{A} \not\models Q(\vec{a})$.

$\square$

As observed in [54, 22], a partial CWA enables the use of more expressive query languages without increasing the complexity of query evaluation. This is particularly useful when many predicates are closed—recall that it can even be useful to close all predicates that can occur in the data. We next make this more precise for our particular framework by introducing a concrete class of OMQCs that combine FOQs for closed predicates with UCQs for open predicates. As in the relational database setting, we admit only FOQs that are *domain-independent* and thus correspond to expressions of relational algebra (and SQL queries), see [1] for a formal definition.

Theorem 3.3. *Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q(\vec{x}))$ be an OMQC from $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, CQ)$ and $q'(\vec{x})$ a domain-independent FOQ with $\mathsf{sig}(q') \subseteq \Sigma_{\mathsf{C}}$. If Q is FO-rewritable (evaluating Q is in* PTime*) and ABox-consistency is FO-rewritable (in* PTime*, respectively) for $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}})$, then the OMQC $Q' = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q \wedge q')$ is FO-rewritable (evaluating Q' is in* PTime*, respectively).*

Proof. Assume that p is an FO-rewriting of Q and that p' is a Boolean FOQ such that for all Σ_{A}-ABoxes $\mathcal{A}$, $\mathcal{I}_{\mathcal{A}} \models p'$ iff $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$. Then $\neg p' \vee (p \wedge q')$ is an FO-rewriting of Q'. Next assume that evaluating Q is in PTime and that ABox consistency w.r.t $(\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}})$ is in PTime. To show that evaluating Q' is in PTime, let $\mathcal{A}$ be a Σ_{A}-ABox and $\vec{a}$ a tuple in $\mathcal{A}$. Then $\mathcal{A} \models Q'(\vec{a})$ iff $\mathcal{A}$ is not consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ or $\mathcal{A} \models Q(\vec{a})$ and $\mathcal{I}_{\mathcal{A}} \models q'(\vec{a})$. As both can be checked in polynomial time, one can decide $\mathcal{A} \models Q'(\vec{a})$ in PTime. □

4. A coNP-Upper Bound for Query Evaluation

We show that for our most expressive DL, $\mathcal{ALCHI}$, UCQ evaluation for OMQCs is in coNP. Recall from the introduction that this bound is not a consequence of results on ontology-mediated querying in description logics with nominals because nominals are part of the TBox and thus their number is a constant. 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 4.1. *The evaluation problem for OMQCs in $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, UCQ)$ is in* coNP.

The proof is given by a sequence of lemmas. We first show that it suffices to consider interpretations that are (essentially) forest-shaped when evaluating UCQs and then introduce mosaics as small forest-shaped interpretations. A *forest over an alphabet S* is a prefix-closed set of words over $S^* \setminus \{\varepsilon\}$, where ε denotes the empty word. Let F be a forest over S. A *root of* F is a word in F of length one. 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. The *depth* of $w \in F$ is $|w| - 1$, where $|w|$ is the length of w. The *depth* of a finite forest F is the maximum of the depths of all $w \in F$. 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 4.2. 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 $\Delta^{\mathcal{I}}$, or
- e is a successor of d or d is a successor of e.

$\mathcal{I}$ is of arity k if the forest $\Delta^{\mathcal{I}}$ is of arity k. △

Note that a forest-shaped interpretation is forest-shaped only in a loose sense since it admits edges from any node to the root. We remind the reader of the following easily proved fact.

Lemma 4.3. *Let h be a homomorphism from $\mathcal{I}$ to $\mathcal{J}$ preserving $\mathsf{N_I}$ and let $q(\vec{x})$ be a UCQ and $\vec{a}$ a tuple of individual names. Then $\mathcal{J} \models q(\vec{a})$ if $\mathcal{I} \models q(\vec{a})$.*

As announced, the next lemma shows that it suffices to consider forest-shaped interpretations when evaluating UCQs. We use $\mathsf{cl}(\mathcal{T})$ to denote the closure of $\mathsf{sub}(\mathcal{T})$ under single negation.

Lemma 4.4. *Let $\mathcal{A}$ be a Σ_A-ABox, $\vec{a}$ a tuple in $\mathsf{Ind}(\mathcal{A})$, and $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ a OMQC from $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, UCQ)$. Then the following are equivalent:*

(1) *$\mathcal{A} \models Q(\vec{a})$;*
(2) *$\mathcal{I} \models q(\vec{a})$ for all forest-shaped models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respect Σ_C and such that*
   - *the arity of $\Delta^\mathcal{I}$ is $|\mathcal{T}|$,*
   - *$\mathsf{Ind}(\mathcal{A})$ is the set of roots of $\Delta^\mathcal{I}$,*
   - *for every $d \in \Delta^\mathcal{I} \setminus \mathsf{Ind}(\mathcal{A})$ and $\exists r.C \in \mathsf{cl}(\mathcal{T})$ with $d \in (\exists r.C)^\mathcal{I}$, there exists $a \in \mathsf{Ind}(\mathcal{A})$ with $(d, a) \in r^\mathcal{I}$ and $a \in C^\mathcal{I}$ or there exists a successor d' of d in $\Delta^\mathcal{I}$ such that $(d, d') \in r^\mathcal{I}$ and $d' \in C^\mathcal{I}$.*

The proof is given in the appendix. (1) $\Rightarrow$ (2) is trivial and the proof of (2) $\Rightarrow$ (1) is by unravelling a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ with $\mathcal{I} \not\models q(\vec{a})$ into a forest-shaped model of $\mathcal{T}$ and $\mathcal{A}$ from which there is a homomorphism preserving $\mathsf{N_I}$ to the original model $\mathcal{I}$ and then applying Lemma 4.3.

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

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

In general, 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 $t = \mathsf{tp}_\mathcal{I}(d)$. We use $\mathsf{TP}(\mathcal{T})$ to denote the set of all $\mathcal{T}$-types. For $\mathcal{T}$-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 $t' = \mathsf{tp}_\mathcal{I}(e)$.

We now define the notion of a *mosaic* for an ABox $\mathcal{A}$ and an OMQC $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$. Mosaics are abstract representations of interpretations which add to the ABox $\mathcal{A}$ a tree-shaped interpretation of outdegree bounded by $|\mathcal{T}|$ and depth at most $|q|$. The tree-shaped part is linked to the ABox via roles, where the number of ABox individuals linked to an element of the tree-shaped interpretation is bounded by $|\mathcal{T}|$. We ensure that a mosaic can be extended to a proper model of $\mathcal{T}$ and $\mathcal{A}$ by hooking fresh interpretations to its ABox individuals and the leaves of its tree-shaped interpretation. *Coherent* sets of mosaics will correspond to forest-shaped models of $\mathcal{T}$ and $\mathcal{A}$. We ensure that is can be checked in polynomial time in $|\mathcal{A}|$ whether a set of mosaics is coherent and whether q is satisfied in the interpretation to which is corresponds. A standard guess and check algorithm (which guesses a set of mosaics and checks its coherence and satisfaction of q) then shows that it is NP to decide $\mathcal{A} \not\models Q$.

Definition 4.5. Let $\mathcal{A}$ be a Σ_A-ABox and $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ from $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, \text{UCQ})$. A *mosaic for 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 depth 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 $\mathcal{T}$, $r^{\mathcal{I}} = \{(a,b) \mid r(a,b) \in \mathcal{A}\}$. △

Let $(\mathcal{I},\tau)$ and $(\mathcal{I}',\tau')$ be mosaics. A bijective function $f : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{I}'}$ is an *isomorphism* between $(\mathcal{I},\tau)$ and $(\mathcal{I}',\tau')$ if both f and its inverse f^{-1} are homomorphisms preserving $\mathsf{N_I}$ and $\tau(d) = \tau'(f(d))$, for all $d \in \Delta^{\mathcal{I}}$. We call $(\mathcal{I},\tau)$ and $(\mathcal{I}',\tau')$ *isomorphic* if there is an isomorphism between $(\mathcal{I},\tau)$ and $(\mathcal{I}',\tau')$.

For a forest F, $w \in F$, 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 4.6. 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 exists $(\mathcal{I}',\tau') \in M$ 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 all $(\mathcal{I},\tau) \in M$ and all successors $d \in \Delta^{\mathcal{I}}$ of the root of $\Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})$, there exist $(\mathcal{I}',\tau') \in M$ and an isomorphism f from $(\mathcal{I},\tau)|_{\Delta^{\mathcal{I}}_{d,|q|-1} \cup \mathsf{Ind}(\mathcal{A})}$ to $(\mathcal{I}',\tau')|_{\Delta^{\mathcal{I}'}_{e,|q|-1} \cup \mathsf{Ind}(\mathcal{A})}$ such that $f(d) = e$, where $e \in \Delta^{\mathcal{I}'}$ is the root of $\Delta^{\mathcal{I}'} \setminus \mathsf{Ind}(\mathcal{A})$.

We write $M \vdash q(\vec{a})$ if $\biguplus_{(\mathcal{I},\tau)\in M} \mathcal{I} \models q(\vec{a})$, 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 4.7. *Let $\mathcal{A}$ be a Σ_{A}-ABox, $\vec{a}$ a tuple in $\mathsf{Ind}(\mathcal{A})$, and $Q = (\mathcal{T},\Sigma_{\mathsf{A}},\Sigma_{\mathsf{C}},q)$ a OMQC from $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}$, UCQ$)$. Then the following are equivalent:*
(1) $\mathcal{A} \models Q(\vec{a})$;
(2) $M \vdash q(\vec{a})$, for all coherent sets M of mosaics for Q and $\mathcal{A}$.

Proof. (2) ⇒ (1). Suppose $\mathcal{A} \not\models Q(\vec{a})$. Let $\mathcal{I}$ be a forest-shaped model with $\mathcal{I} \not\models q(\vec{a})$ and satisfying the conditions of Lemma 4.4 (2). 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 Q and $\mathcal{A}$ (to satisfy Condition 9 for mosaics for concept names A and role names r that do not occur in $\mathcal{T}$, 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 $M \not\vdash q(\vec{a})$. But this follows from Lemma 4.3 and the fact that the function h from $\mathcal{I}' = \biguplus_{(\mathcal{J},\tau)\in M} \mathcal{J}$ to $\mathcal{I}$ mapping every $a \in \mathsf{N_I}$ to itself and every copy $d' \in \Delta^{\mathcal{I}'}$ of some $d \in \Delta^{\mathcal{I}}$ to d is a homomorphism from $\mathcal{I}'$ to $\mathcal{I}$ preserving $\mathsf{N_I}$.

(1) $\Rightarrow$ (2). Suppose there is a coherent set M of mosaics for Q and $\mathcal{A}$ with $M \not\models q$. 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 $d \in \Delta^{\mathcal{I}_i} \setminus \mathsf{Ind}(\mathcal{A})$ 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 element of $\Delta^{\mathcal{I}_i} \setminus \mathsf{Ind}(\mathcal{A})$ 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 elements 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 the limit of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$ (cf. proof of Lemma 4.4). It is shown in the appendix that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C} such that $\mathcal{I} \not\models q(\vec{a})$. $\square$

Lemma 4.8. *Let $\mathcal{A}$ be a Σ_{A}-ABox and $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ in $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, UCQ)$. Then, up to isomorphisms, the size of any coherent set M of mosaics for Q and $\mathcal{A}$ is bounded by $(2|\mathcal{A}|)^{|\mathcal{T}|^{f(|q|)}}$, for a linear polynomial f.*

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 $(\mathcal{I}, \tau)$. $\square$

We are now in the position to prove Theorem 4.1. Fix an OMQC $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ in $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, \text{UCQ})$. We show that given a Σ_{A}-ABox $\mathcal{A}$ and tuple $\vec{a}$ in $\mathsf{Ind}(\mathcal{A})$, deciding $\mathcal{A} \not\models Q(\vec{a})$ is in NP. Assume $\mathcal{A}$ and $\vec{a}$ are given. By Lemmas 4.7 and 4.8, $\mathcal{A} \not\models Q(\vec{a})$ iff there exists a coherent set M of mosaics for Q and $\mathcal{A}$ such that $|M| \leq (2|\mathcal{A}|)^{|\mathcal{T}|^{f(|q|)}}$ (f a linear polynomial) and $M \not\models q(\vec{a})$. Thus, it is sufficient to show that it can be decided in time polynomial in the size $|\mathcal{A}|$ of $\mathcal{A}$ whether M is a coherent set of mosaics for Q and $\mathcal{A}$ and whether $M \not\models q(\vec{a})$. 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 time polynomial in $|\mathcal{A}|$ and that checking if $\mathcal{J} \models q(\vec{a})$ is again possible in time polynomial in $|\mathcal{A}|$.

5. Quantified Query Case: Dichotomies for DL-Lite$_\mathcal{R}$ and $\mathcal{EL}$

We consider the quantified query case and show two dichotomy results: for every DL-Lite$_\mathcal{R}$ TBox with closed predicates $(\mathcal{T}, \Sigma_\mathsf{C})$, UCQ evaluation is FO-rewritable or coNP-complete. In the latter case, there is even a tCQ q such that evaluating the OMQ $(\mathcal{T}, \Sigma_\mathsf{C}, q)$ is coNP-hard. It thus follows that a TBox with closed predicates is FO-rewritable for tCQs iff it is FO-rewritable for CQs iff it is FO-rewritable for UCQs, and likewise for coNP-completeness. It also follows that FO-rewritability coincides with tractability, that is, query evaluation in PTime. We obtain the same results for $\mathcal{EL}$ TBoxes with closed predicates except that tCQs are replaced with dtCQs and FO-rewritability is replaced with PTime. In both the DL-Lite$_\mathcal{R}$ case and the $\mathcal{EL}$ case, tractability also implies that query evaluation with closed predicates coincides with query evaluation without closed predicates, unless the data is inconsistent with the TBox. The proof strategy is similar in both cases, but the details are more involved for $\mathcal{EL}$. We first consider the notion of convexity which formalizes the absence of implicit disjunctions in answering tree-shaped queries and show that for $\mathcal{ALCHI}$ TBoxes with closed predicates, non-convexity implies coNP-hardness. We then introduce a syntactic condition for DL-Lite$_\mathcal{R}$ TBoxes (and later also for $\mathcal{EL}$ TBoxes) with closed predicates called safeness and show that non-safeness implies non-convexity while safeness implies tractability.

5.1. Non-Convexity Implies coNP-hardness.

It is well-known that the notion of convexity is closely related to the complexity of query evaluation, see for example [39, 47]. Recall that we omit Σ_A from the OMQC $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ and write $(\mathcal{T}, \Sigma_\mathsf{C}, q)$ if $\Sigma_\mathsf{A} = \mathsf{N_C} \cup \mathsf{N_R}$.

Definition 5.1. Let $\mathcal{Q} \in \{\text{tCQ}, \text{dtCQ}\}$. A TBox with closed predicates $(\mathcal{T}, \Sigma_\mathsf{C})$ is *convex* for $\mathcal{Q}$ if for all ABoxes $\mathcal{A}$, $a \in \mathsf{Ind}(\mathcal{A})$, and $q_1(x), q_2(x) \in \mathcal{Q}$ the following holds: if $\mathcal{A} \models (\mathcal{T}, \Sigma_\mathsf{C}, q_1 \vee q_2)(a)$, then $\mathcal{A} \models (\mathcal{T}, \Sigma_\mathsf{C}, q_i)(a)$ for some $i \in \{1, 2\}$. △

Without closed predicates, every DL-Lite$_\mathcal{R}$ and $\mathcal{EL}$ TBox is convex for tCQs. In fact, it is shown in [47, 33] that for TBoxes in $\mathcal{ALCHI}$ (and even more expressive languages) without closed predicates, convexity for tCQs is a necessary condition for UCQ evaluation to be in PTime (unless PTime = coNP). The following is an example of a DL-Lite$_\mathcal{R}$ TBox with closed predicates that is not convex for tCQs.

Example 5.2. Let $\mathcal{T} = \{A \sqsubseteq \exists r.\top, \exists r^-.\top \sqsubseteq B\}$ and $\Sigma_\mathsf{C} = \{B\}$. We show that $(\mathcal{T}, \Sigma_\mathsf{C})$ is not convex for tCQs. To this end, let

$$\begin{array}{rcl} \mathcal{A} & = & \{A(a), B(b_1), A_1(b_1), B(b_2), A_2(b_2)\} \\ q_i & = & \exists y\, r(x,y) \wedge A_i(y) \text{ for } i \in \{1,2\}. \end{array}$$

Then $\mathcal{A} \models (\mathcal{T}, \Sigma_\mathsf{C}, q_1 \vee q_2)(a)$, whereas $\mathcal{A} \not\models (\mathcal{T}, \Sigma_\mathsf{C}, q_i)(a)$ for any $i \in \{1, 2\}$.

We next show that non-convexity implies that query evaluation is coNP-hard. The result is formulated for $\mathcal{ALCHI}$ as our maximal description logics and comes in a directed version (used lated for $\mathcal{EL}$) and a non-directed one (used for DL-Lite$_\mathcal{R}$).

Lemma 5.3. *Let $(\mathcal{T}, \Sigma_\mathsf{C})$ be an $\mathcal{ALCHI}$ TBox with closed predicates that is not convex for tCQs (resp. dtCQs). Then there exists a tCQ q (resp. dtCQ q) such that the evaluation problem for $(\mathcal{T}, \Sigma_\mathsf{C}, q)$ is* coNP*-hard.*

Proof. The proof is by a reduction of 2+2-SAT inspired by [55]. 2+2-SAT is a variant of propositional satisfiability where each clause contains precisely two positive literals and two negative literals. The queries q_1 and q_2 that witness non-convexity from Definition 5.1 are used as subqueries of the query constructed in the reduction, where they serve the purpose of distinguishing truth values of propositional variables. We give a sketch of the reduction only as it is very similar to a corresponding reduction for TBoxes without closed predicates [47]. Let $(\mathcal{T}, \Sigma_\mathsf{C})$ be not convex for tCQs. Then there are an ABox $\mathcal{A}$ with $a \in \mathsf{Ind}(\mathcal{A})$ and tCQs $q_1(x), q_2(x)$ such that $\mathcal{A} \models (\mathcal{T}, \Sigma_\mathsf{C}, q_1 \vee q_2)(a)$ and $\mathcal{A} \not\models (\mathcal{T}, \Sigma_\mathsf{C}, q_i)(a)$ for all $i \in \{1, 2\}$.

We define 2+2-SAT. A *2+2 clause* is of the form $(p_1 \vee p_2 \vee \neg n_1 \vee \neg n_2)$, where each of p_1, p_2, n_1, n_2 is a propositional letter or a truth constant 0, 1. A *2+2 formula* is a finite conjunction of 2+2 clauses. Now, 2+2-SAT is the problem of deciding whether a given 2+2 formula is satisfiable. It is shown in [55] that 2+2-SAT is NP-complete.

Let $\varphi = c_0 \wedge \cdots \wedge c_n$ be a 2+2 formula in propositional letters $w_0, \ldots, w_m$, and let $c_i = p_{i,1} \vee p_{i,2} \vee \neg n_{i,1} \vee \neg n_{i,2}$ for all $i \leq n$. Our aim is to define an ABox $\mathcal{A}_\varphi$ and a tCQ q_0 such that φ is unsatisfiable iff $\mathcal{A}_\varphi \models (\mathcal{T}, \Sigma_\mathsf{C}, q_0)(f)$, for an individual name f we define shortly. To start, we represent the formula φ in the ABox $\mathcal{A}_\varphi$ as follows:

- the individual name f represents the formula φ;
- the individual names $c_0, \ldots, c_n$ represent the clauses of φ;
- the assertions $c(f, c_0), \ldots, c(f, c_n)$, associate f with its clauses, where c is a role name that does not occur in $\mathcal{T}$;
- the individual names $w_0, \ldots, w_m$ represent propositional letters, and the individual names $0, 1$ represent truth constants;
- the assertions
$$\bigcup_{i \leq n} \{p_1(c_i, p_{i,1}), p_2(c_i, p_{i,2}), n_1(c_i, n_{i,1}), n_2(c_i, n_{i,2})\}$$
associate each clause with the four variables/truth constants that occur in it, where p_1, p_2, n_1, and n_2 are role names that do not occur in $\mathcal{T}$.

We further extend $\mathcal{A}_\varphi$ to enforce a truth value for each of the variables w_i and the truth-constants $0, 1$. To this end, add to $\mathcal{A}_\varphi$ copies $\mathcal{A}_0, \ldots, \mathcal{A}_m$ of the ABox $\mathcal{A}$ obtained by renaming individual names such that $\mathsf{Ind}(\mathcal{A}_i) \cap \mathsf{Ind}(\mathcal{A}_j) = \emptyset$ whenever $i \neq j$. Moreover, assume that a_i coincides with the ith copy of a. Intuitively, the copy $\mathcal{A}_i$ of $\mathcal{A}$ is used to generate a truth value for the variable w_i, where we want to interpret w_i as true in an interpretation $\mathcal{I}$ if $\mathcal{I} \models q_1(a_i)$ and as false if $\mathcal{I} \models q_2(a_i)$. To actually relate each individual name w_i to the associated ABox $\mathcal{A}_i$, we use the role name r that does not occur in $\mathcal{T}$. More specifically, we extend $\mathcal{A}_\varphi$ as follows:

(1) link variable w_i to the ABox $\mathcal{A}_i$ by adding the assertion $r(w_i, a_i)$, for all $i \leq m$; thus, the truth of w_i means that $\mathsf{tt}(x) := \exists y\, (r(x, y) \wedge q_1(y))$ is satisfied and falsity means that $\mathsf{ff}(x) := \exists y\, (r(x, y) \wedge q_2(y))$ is satisfied;
(2) to ensure that 0 and 1 have the expected truth values, add a copy of q_1 viewed as an ABox $\mathcal{A}_{q_1}$ with root $1'$ and a copy of q_2 viewed as an ABox $\mathcal{A}_{q_2}$ with root $0'$; add $r(0, 0')$ and $r(1, 1')$.

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

$$\begin{aligned} q_0(x) \quad = \quad & \exists y, y_1, y_2, y_3, y_4\, \big(c(x, y) \wedge p_1(y, y_1) \wedge \mathsf{ff}(y_1) \wedge p_1(y, y_2) \wedge \mathsf{ff}(y_2) \wedge \\ & n_1(y, y_3) \wedge \mathsf{tt}(y_3) \wedge n_2(y, y_4) \wedge \mathsf{tt}(y_4)\big) \end{aligned}$$

which describes the existence of a clause with only false literals and thus captures falsity of φ. It is straightforward to show that φ is unsatisfiable iff $\mathcal{B} \models (\mathcal{T}, \Sigma_\mathsf{C}, q_0)(f)$. Finally observe that q_0 is a dtCQ if q_1 and q_2 are dtCQs. □

5.2. **Dichotomy for DL-Lite$_\mathcal{R}$.** The next definition gives a syntactic safety condition for DL-Lite$_\mathcal{R}$ TBoxes with closed predicates that turns out to characterize tractability.

Definition 5.4 (Safe DL-Lite$_\mathcal{R}$ TBox)**.** Let $(\mathcal{T}, \Sigma_\mathsf{C})$ be a DL-Lite$_\mathcal{R}$ TBox with closed predicates. Then $(\mathcal{T}, \Sigma_\mathsf{C})$ is *safe* if there are no basic concepts B_1, B_2 and role r such that the following conditions are satisfied:
(1) B_1 is satisfiable w.r.t. $\mathcal{T}$;
(2) $\mathcal{T} \models B_1 \sqsubseteq \exists r$ and $\mathcal{T} \models \exists r^- \sqsubseteq B_2$;
(3) $B_1 \neq \exists r'$, for every role r' with $\mathcal{T} \models r' \sqsubseteq r$;
(4) $\mathsf{sig}(B_2) \subseteq \Sigma_\mathsf{C}$ and $\mathsf{sig}(r') \cap \Sigma_\mathsf{C} = \emptyset$ for every role r' with $\mathcal{T} \models B_1 \sqsubseteq \exists r'$ and $\mathcal{T} \models r' \sqsubseteq r$.
△

The following example illustrates safeness.

Example 5.5. It is easy to see that the TBox with closed predicates from Example 5.2 is not safe. As an additional example, consider

$$\mathcal{T} = \{A \sqsubseteq \exists r.\top, r \sqsubseteq s\} \quad \text{and} \quad \Sigma_\mathsf{C} = \{s\}$$

Then $(\mathcal{T}, \Sigma_\mathsf{C})$ is not safe, which is witnessed by the concepts $B_1 = A$, $B_2 = \exists s^-.\top$, and the role r. Indeed, $(\mathcal{T}, \Sigma_\mathsf{C})$ is not convex for tCQs. This can be proved using the ABox

$$\{A(a), s(a, b_1), A_1(b_1), s(a, b_2), A_2(b_2)\}$$

and the tCQs $q_i = \exists y\, (r(x, y) \wedge A_i(y))$, for $i \in \{1, 2\}$.

We now establish the dichotomy result for DL-Lite$_\mathcal{R}$ TBoxes with closed predicates. Let $(\mathcal{T}_1, \Sigma_1)$ and $(\mathcal{T}_2, \Sigma_2)$ be TBoxes with closed predicates. Then we say that $(\mathcal{T}_1, \Sigma_1)$ and $(\mathcal{T}_2, \Sigma_2)$ are *UCQ-inseparable on consistent ABoxes* [16, 17] if

$$\mathcal{A} \models (\mathcal{T}_1, \Sigma_1, q)(\vec{a}) \quad \text{iff} \quad \mathcal{A} \models (\mathcal{T}_2, \Sigma_2, q)(\vec{a})$$

holds for all UCQs q, all ABoxes $\mathcal{A}$ consistent w.r.t. both $(\mathcal{T}_1, \Sigma_1)$ and $(\mathcal{T}_2, \Sigma_2)$, and all tuples $\vec{a}$ in $\mathsf{Ind}(\mathcal{A})$. The notion of UCQ-inseparability is used in Condition 2(a) of the following dichotomy theorem. Informally, it says that tractable query evaluation implies that query evaluation with closed predicates coincides with query evaluation without closed predicates.

Theorem 5.6. *Let $(\mathcal{T}, \Sigma_\mathsf{C})$ be a DL-Lite$_\mathcal{R}$ TBox with closed predicates. Then*
(1) *If $(\mathcal{T}, \Sigma_\mathsf{C})$ is not safe, then $(\mathcal{T}, \Sigma_\mathsf{C})$ is not convex for tCQs and tCQ evaluation w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ is* coNP*-hard.*
(2) *If $(\mathcal{T}, \Sigma_\mathsf{C})$ is safe, then $(\mathcal{T}, \Sigma_\mathsf{C})$ is convex for tCQs and*
  (a) *$(\mathcal{T}, \Sigma_\mathsf{C})$ and $(\mathcal{T}, \emptyset)$ are UCQ-inseparable on consistent ABoxes.*
  (b) *UCQ evaluation w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ is FO-rewritable.*

Proof. We start with the proof of Point (1). Assume that $(\mathcal{T}, \Sigma_\mathsf{C})$ is not safe. Consider basic concepts B_1, B_2 and a role r satisfying Points (1) to (4) of Definition 5.4. By Points (1) and (4) of Definition 5.4, B_1 is satisfiable w.r.t. $\mathcal{T}$ and $\mathcal{T} \not\models B_1 \sqsubseteq \exists r'$ for any role r' with $\mathsf{sig}(r') \subseteq \Sigma_\mathsf{C}$ and $\mathcal{T} \models r' \sqsubseteq r$. We obtain

$$\mathcal{T} \not\models B_1 \sqsubseteq \bigsqcup_{\mathcal{T} \models r' \sqsubseteq r, \mathsf{sig}(r') \subseteq \Sigma_\mathsf{C}} \exists r'$$

since $(\mathcal{T}, \emptyset)$ is convex. Observe that the CI to the right is a $\mathcal{ALCI}$ CI and $\mathcal{T}$ is an $\mathcal{ALCHI}$ TBox. It is well known that $\mathcal{ALCHI}$ has the finite model property in the sense that any CI that does not follow from an $\mathcal{ALCHI}$ TBox is refuted in a finite model of the TBox [6]. Thus, we can take a finite model $\mathcal{I}$ of $\mathcal{T}$ and some $a_0 \in B_1^\mathcal{I}$ such that $a_0 \notin (\exists r'.\top)^\mathcal{I}$ for any role r' with $\mathsf{sig}(r') \subseteq \Sigma_\mathsf{C}$ and $\mathcal{T} \models r' \sqsubseteq r$. Let $\mathcal{I}_r$ be the interpretation obtained from $\mathcal{I}$ by removing all pairs (a_0, b) from any $r'^\mathcal{I}$ with $\mathcal{T} \models r' \sqsubseteq r$. Take the ABox $\mathcal{A}_r$ corresponding to $\mathcal{I}_r$ and let $\mathcal{A}$ be the disjoint union of two copies of $\mathcal{A}_r$. We denote the individual names of the first copy by $(b, 1)$, $b \in \Delta^{\mathcal{I}_r}$, and the individual names of the second copy by $(b, 2)$, $b \in \Delta^{\mathcal{I}_r}$. Let $\mathcal{A}'$ be defined as

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

where A_1 and A_2 are fresh concept names. Define, for $i \in \{1, 2\}$, the tCQs

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

if B_2 is a concept name. If $B_2 = \exists s$ (or $B_2 = \exists s^-$), for a role name s, then set $q_i(x) = \exists y, z\, (r(x,y) \wedge A_i(y) \wedge s(y,z))$ (or $q_i(x) = \exists y, z\, (r(x,y) \wedge A_i(y) \wedge s(z,y))$, respectively). We use $\mathcal{A}'$ and $q_i(x)$ to prove that $(\mathcal{T}, \Sigma_\mathsf{C})$ is not convex for tCQs.

Claim 1. $\mathcal{A}' \models (\mathcal{T}, \Sigma_\mathsf{C}, q_1 \vee q_2)(a_0, 1)$.

Proof of claim. Let $\mathcal{J}$ be a model of $\mathcal{T}$ and $\mathcal{A}'$ that respects Σ_C. We have $(a_0, 1) \in B_1^\mathcal{J}$ (since, by Point (3) of Definition 5.4, $B_1 \neq \exists r'$ for every r' with $\mathcal{T} \models r' \sqsubseteq r$). It follows from the conditions that $\mathcal{J}$ is a model of $\mathcal{T}$, $\mathcal{T} \models B_1 \sqsubseteq \exists r$, and $\mathcal{T} \models \exists r^- \sqsubseteq B_2$, that there exists $e \in \Delta^\mathcal{J}$ with $((a_0, 1), e) \in r^\mathcal{J}$ and $e \in B_2^\mathcal{J}$. Using the condition that $\mathsf{sig}(B_2) \subseteq \Sigma_\mathsf{C}$ it follows from the definition of $\mathcal{A}'$ that e is of the form (e', i) with $e' \in B_2^\mathcal{I}$ and $i \in \{1, 2\}$. If $i = 1$, we have $A_1(e', 1) \in \mathcal{A}'$ and so $(a_0, 1) \in \exists r.(A_1 \sqcap B_2)^\mathcal{J}$, as required. If $i = 2$, we have $A_2(e', 2) \in \mathcal{A}'$ and so $(a_0, 1) \in \exists r.(A_2 \sqcap B_2)^\mathcal{J}$, as required. ⊣

Claim 2. $\mathcal{A}' \not\models (\mathcal{T}, \Sigma_\mathsf{C}, q_i)(a_0, 1)$, for $i \in \{1, 2\}$.

Proof of claim. Let $i = 1$ (the case $i = 2$ is similar and omitted). We construct a model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}'$ that respects Σ_C such that $(a_0, 1) \notin (\exists r.(A_1 \sqcap B_2))^\mathcal{J}$. $\mathcal{J}$ is defined as the interpretation corresponding to the ABox $\mathcal{A}'$ extended by

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

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

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

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

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

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

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

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

It follows from Claims 1 and 2 that $(\mathcal{T}, \Sigma_\mathsf{C})$ is not convex for tCQs. The coNP-hardness of tCQ evaluation follows from Lemma 5.3. This finishes the proof of Point (1).

We come to the proof of Point (2). The proof relies on the canonical model associated with an ABox and a DL-Lite$_\mathcal{R}$ TBox [20, 37]. Specifically, for every ABox $\mathcal{A}$ that is consistent w.r.t. a DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$ without closed predicates, there is a model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ which is minimal in the sense that for all UCQs q and tuples $\vec{a}$ in $\mathsf{Ind}(\mathcal{A})$:

$$\mathcal{A} \models (\mathcal{T}, \emptyset, q)(\vec{a}) \quad \text{iff} \quad \mathcal{I} \models q(\vec{a}).$$

We show that if $\mathcal{I}$ is constructed in a careful way and $(\mathcal{T}, \Sigma_\mathsf{C})$ is safe, then $\mathcal{I}$ respects Σ_C. This means that a tuple $\vec{a} \in \mathsf{Ind}(\mathcal{A})$ is a certain answer to $(\mathcal{T}, \emptyset, q)$ on $\mathcal{A}$ iff it is a certain answer to $(\mathcal{T}, \Sigma_\mathsf{C}, q)$ on $\mathcal{A}$ since closed predicates can only result in additional answers, but not in invalidating answers. It follows that $(\mathcal{T}, \Sigma_\mathsf{C})$ is convex for tCQs and that $(\mathcal{T}, \Sigma_\mathsf{C})$ and $(\mathcal{T}, \emptyset)$ are UCQ-inseparable on consistent ABoxes (Point (a)). Additionally, we show that ABox consistency w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ is FO-rewritable when $(\mathcal{T}, \Sigma_\mathsf{C})$ is safe, which together with the first observation implies that UCQ evaluation w.r.t. a safe $(\mathcal{T}, \Sigma_\mathsf{C})$ is FO-rewritable: if p_c is an FO-rewriting of ABox consistency w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ and $q'(\vec{x})$ is an FO-rewriting of $(\mathcal{T}, \emptyset, q(\vec{x}))$, then $\neg p_c \vee q'(\vec{x})$ is an FO-rewriting $(\mathcal{T}, \Sigma_\mathsf{C}, q(\vec{x}))$. Thus, Point (b) follows. It thus remains to prove Claims 3 and 4 below.

Claim 3. Let $(\mathcal{T}, \Sigma_\mathsf{C})$ be a safe DL-Lite$_\mathcal{R}$ TBox with closed predicates. Then for every UCQ q, we have

$$\mathcal{A} \models (\mathcal{T}, \Sigma_\mathsf{C}, q)(\vec{a}) \quad \text{iff} \quad \mathcal{A} \models (\mathcal{T}, \emptyset, q)(\vec{a})$$

for all ABoxes $\mathcal{A}$ that are consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ and all $\vec{a} \in \mathsf{Ind}(\mathcal{A})$.

Proof of claim. Let $(\mathcal{T}, \Sigma_\mathsf{C})$ be safe and assume that $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$. We construct a canonical model of $\mathcal{A}$ and $\mathcal{T}$ as the interpretation corresponding to a (possibly infinite) ABox $\mathcal{A}_c$ that is the limit of a sequence of ABoxes $\mathcal{A}_0, \mathcal{A}_1, \ldots$. Let $\mathcal{A}_0 = \mathcal{A}$ and assume $\mathcal{A}_j$ has been defined already. Then $\mathcal{A}_{j+1}$ is obtained from $\mathcal{A}_j$ by applying the following two rules:

(R1) if there exist roles r, s and $a, b \in \mathsf{N_I}$ with $\mathcal{T} \models r \sqsubseteq s$, $r(a, b) \in \mathcal{A}_j$, and $s(a, b) \notin \mathcal{A}_j$, then add $s(a, b)$ to $\mathcal{A}_j$;

(R2) if (R1) does not apply and there are basic concepts B_1, B_2 and $a \in \mathsf{N_I}$ such that $\mathcal{T} \models B_1 \sqsubseteq B_2$, $a \in B_1^{\mathcal{I}_{\mathcal{A}_j}}$, and $a \notin B_2^{\mathcal{I}_{\mathcal{A}_j}}$, then add $B_2(a)$ to $\mathcal{A}_j$ if B_2 is a concept name and add $r(a, b)$ for some fresh $b \in \mathsf{N_I}$ to $\mathcal{A}_j$ if $B_2 = \exists r$ for some role r.

We assume that (R1) and (R2) are applied in a fair way. Now let $\mathcal{I}_{\mathcal{T},\mathcal{A}} = \mathcal{I}_{\mathcal{A}_c}$, where $\mathcal{A}_c = \bigcup_{i\geq 0} \mathcal{A}_i$. It is known [37] (and easy to prove) that $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ with the following properties:

(p1) For all UCQs $q(\vec{x})$ and $\vec{a} \in \mathsf{Ind}(\mathcal{A})$: $\mathcal{A} \models (\mathcal{T}, \emptyset, q)(\vec{a})$ iff $\mathcal{I}_{\mathcal{T},\mathcal{A}} \models q(\vec{a})$.

(p2) For any individual name $b \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A})$ introduced as a witness for Rule (R2) for some CI of the form $B_1 \sqsubseteq \exists s$ and every basic concept B : $b \in B^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$ iff $\mathcal{T} \models \exists s^- \sqsubseteq B$.

To show that $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects Σ_{C} it is sufficient to prove that every assertion using predicates from Σ_{C} in $\mathcal{A}_c$ is contained in $\mathcal{A}$. We first show that for all $a, b \in \mathsf{Ind}(\mathcal{A})$,

- if $A(a) \in \mathcal{A}_c$ and $A \in \Sigma_{\mathsf{C}}$, then $A(a) \in \mathcal{A}$; and
- if $r(a, b) \in \mathcal{A}_c$ and $r \in \Sigma_{\mathsf{C}}$, then $r(a, b) \in \mathcal{A}$.

For a proof by contradiction assume that $A(a) \in \mathcal{A}_c$ but $A(a) \notin \mathcal{A}$ for some concept name $A \in \Sigma_{\mathsf{C}}$. By Point (p1), the former implies $\mathcal{A} \models (\mathcal{T}, \emptyset, A(x))(a)$. Thus, $\mathcal{A} \models (\mathcal{T}, \Sigma_{\mathsf{C}}, A(x))(a)$ which contradicts the assumption that $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$. The argument for role assertions $r(a, b)$ is similar and omitted.

It remains to show there are no $a \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A}_0)$ and basic concept B with $\mathsf{sig}(B) \subseteq \Sigma_{\mathsf{C}}$ such that $a \in B^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$. For a proof by contradiction, assume that there exist an $a \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A}_0)$ and basic concept B with $\mathsf{sig}(B) \subseteq \Sigma_{\mathsf{C}}$ such that $a \in B^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$. Let a be the first such individual name introduced using Rule (R2) in the construction of $\mathcal{A}_c$. By Point (p2) and the construction of $\mathcal{A}_c$ there exist B_1, r, a_0 and $j \geq 0$ such that $\mathcal{T} \models B_1 \sqsubseteq \exists r$, $a_0 \in B_1^{\mathcal{I}_{\mathcal{A}_j}}$, $a_0 \notin (\exists r)^{\mathcal{I}_{\mathcal{A}_j}}$, $(a_0, a) \in r^{\mathcal{I}_{\mathcal{A}_{j+1}}}$, $\mathcal{T} \models \exists r^- \sqsubseteq B$. We show that B_1, B_2, and r satisfy Conditions 1 to 4 from Definition 5.4 for $B_2 := B$ and thus derive a contradiction to the assumption that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is safe. Conditions 1 and 2 are clear. For Condition 3, assume that $B_1 = \exists r'$ for some r' such that $\mathcal{T} \models r' \sqsubseteq r$. Then $(a_0, e) \in (r')^{\mathcal{I}_{\mathcal{A}_j}}$ for some e. But then, since Rule (R1) is exhaustively applied before Rule (R2) is applied, we have $(a_0, e) \in r^{\mathcal{I}_{\mathcal{A}_j}}$ which contradicts $a_0 \notin (\exists r)^{\mathcal{I}_{\mathcal{A}_j}}$. For Condition 4 assume that $\mathcal{T} \models B_1 \sqsubseteq \exists r'$ for some role r' such that $\mathsf{sig}(r') \subseteq \Sigma_{\mathsf{C}}$ and $\mathcal{T} \models r' \sqsubseteq r$. Then $a_0 \in \mathsf{Ind}(\mathcal{A})$ because otherwise a_0 is an individual name introduced before a such that $a_0 \in (\exists r')^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$ and $\mathsf{sig}(\exists r') \subseteq \Sigma_{\mathsf{C}}$, which contradicts our assumption about a. By Point (p1) and the consistency of $\mathcal{A}$ w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$, there is some $b \in \mathsf{Ind}(\mathcal{A})$ such that $(a_0, b) \in (r')^{\mathcal{I}_{\mathcal{A}}}$. But then, again since Rule (R1) is exhaustively applied before Rule (R2) is applied, $(a_0, b) \in r^{\mathcal{I}_{\mathcal{A}_j}}$ which contradicts $a_0 \notin (\exists r)^{\mathcal{I}_{\mathcal{A}_j}}$.

Observe that we have also proved that if $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \emptyset)$, then there do not exist $a \in \mathsf{Ind}(\mathcal{A}_c) \setminus \mathsf{Ind}(\mathcal{A}_0)$ and a basic concept B with $\mathsf{sig}(B) \subseteq \Sigma_{\mathsf{C}}$ such that $a \in B^{\mathcal{I}_{\mathcal{T},\mathcal{A}}}$. ⊣

As the final step in the proof of Point (2), we show that ABox consistency w.r.t. a safe DL-Lite$_{\mathcal{R}}$ TBox with closed predicates is FO-rewritable.

Claim 4. Let $(\mathcal{T}, \Sigma_{\mathsf{C}})$ be a safe DL-Lite$_{\mathcal{R}}$ TBox with closed predicates. Then ABox consistency w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is FO-rewritable.

Proof of claim. It follows immediately from the final remark in the proof of Claim 3 above that an ABox $\mathcal{A}$ is consistent w.r.t. a safe $(\mathcal{T}, \Sigma_{\mathsf{C}})$ if, and only if, (i) $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \emptyset)$, (ii) $\mathcal{A} \models (\mathcal{T}, \emptyset, A(x))(a)$ implies $a \in A^{\mathcal{I}_{\mathcal{A}}}$ for all concept names $A \in \Sigma_{\mathsf{C}}$, (iii) $\mathcal{A} \models (\mathcal{T}, \emptyset, \exists y\, r(x, y))(a)$ implies $a \in (\exists r)^{\mathcal{I}_{\mathcal{A}}}$ for all roles r with $\mathsf{sig}(r) \subseteq \Sigma_{\mathsf{C}}$, and (iv) $\mathcal{A} \models (\mathcal{T}, \emptyset, r(x, y))(a, b)$ implies $(a, b) \in r^{\mathcal{I}_{\mathcal{A}}}$ for all role names $r \in \Sigma_{\mathsf{C}}$.

To obtain an FO-rewriting of ABox consistency w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$, let p_c be an FO-rewriting of ABox consistency w.r.t. $(\mathcal{T}, \emptyset)$, let $q_A(x)$ be an FO-rewriting of $(\mathcal{T}, \emptyset, A(x))$, for $A \in \Sigma_{\mathsf{C}}$, let $q_{\exists r}(x)$ be an FO-rewriting of $(\mathcal{T}, \emptyset, \exists y\, r(x,y))$ for $r \in \Sigma_{\mathsf{C}}$, let $q_{\exists r^-}(x)$ be an FO-rewriting of $(\mathcal{T}, \emptyset, \exists y\, r(y,x))$ for $r \in \Sigma_{\mathsf{C}}$ and let $q_r(x,y) = \bigvee_{\mathcal{T} \models s \sqsubseteq r} s(x,y)$, for $r \in \Sigma_{\mathsf{C}}$. Then $p_c \wedge q_1 \wedge q_2 \wedge q_3 \wedge q_4$ with

$$\begin{aligned}
q_1 &= \forall x \bigwedge_{A \in \Sigma_{\mathsf{C}}} (q_A(x) \rightarrow A(x)) \\
q_2 &= \forall x \bigwedge_{r \in \Sigma_{\mathsf{C}}} (q_{\exists r}(x) \rightarrow \exists y\, r(x,y)) \\
q_3 &= \forall x \bigwedge_{r \in \Sigma_{\mathsf{C}}} (q_{\exists r^-}(x) \rightarrow \exists y\, r(y,x)) \\
q_4 &= \forall x \forall y \bigwedge_{r \in \Sigma_{\mathsf{C}}} (q_r(x,y) \rightarrow r(x,y))
\end{aligned}$$

is an FO-rewriting of ABox consistency w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$. ⊣ □

5.3. **Dichotomy for $\mathcal{EL}$.** We show the announced dichotomy for $\mathcal{EL}$ TBoxes with closed predicates. While we follow the same strategy as in the DL-Lite case, there are some interesting new aspects. In particular, we identify an additional reason for coNP-hardness that we treat by using a variant of the Craig interpolation property for $\mathcal{EL}$. We call a concept E a *top-level conjunct (tlc) of* an $\mathcal{EL}$ concept C if C is of the form $D_1 \sqcap \cdots \sqcap D_n$ and $E = D_i$ for some i. We use the following version of safeness.

Definition 5.7 (Safe $\mathcal{EL}$ TBox)**.** An $\mathcal{EL}$ TBox with closed predicates $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is *safe* if there exists no $\mathcal{EL}$ concept inclusion $C \sqsubseteq \exists r.D$ such that
(1) $\mathcal{T} \models C \sqsubseteq \exists r.D$;
(2) there is no tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$;
(3) one of the following is true:
  (s1) $r \notin \Sigma_{\mathsf{C}}$ and $\mathsf{sig}(D) \cap \Sigma_{\mathsf{C}} \neq \emptyset$;
  (s2) $r \in \Sigma_{\mathsf{C}}$, $\mathsf{sig}(D) \not\subseteq \Sigma_{\mathsf{C}}$, and there is no $\mathcal{EL}$ concept E with $\mathsf{sig}(E) \subseteq \Sigma_{\mathsf{C}}$, $\mathcal{T} \models C \sqsubseteq \exists r.E$, and $\mathcal{T} \models E \sqsubseteq D$. △

Condition 3(s1) captures a reason for non-convexity that is similar to the DL-Lite case. For example, we can recast Example 5.2 using $\mathcal{T} = \{A \sqsubseteq \exists r.B\}$ and $\Sigma_{\mathsf{C}} = \{B\}$. Then the inclusion $A \sqsubseteq \exists r.B$ shows that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe as $r \notin \Sigma_{\mathsf{C}}$ and $B \in \Sigma_{\mathsf{C}}$. However, in $\mathcal{EL}$ there is an additional reason for non-convexity that is captured by Condition 3(s2).

Example 5.8. Let $\mathcal{T} = \{A \sqsubseteq \exists r.B\}$ and $\Sigma_{\mathsf{C}} = \{r\}$. Clearly, by Condition 3(s2), $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe. We show that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not convex for dtCQs. Let

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

Then $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not convex because $\mathcal{A} \models (\mathcal{T}, \Sigma_{\mathsf{C}}, q_1 \vee q_2)(a)$, whereas $\mathcal{A} \not\models (\mathcal{T}, \Sigma_{\mathsf{C}}, q_i)(a)$ for any $i \in \{1, 2\}$. Observe that one cannot reproduce this example in DL-Lite: for example, for the TBox $\mathcal{T}' = \{A \sqsubseteq \exists r, \exists r^- \sqsubseteq B\}$ with $\Sigma'_{\mathsf{C}} = \{r\}$, we have $\mathcal{A} \models (\mathcal{T}', \Sigma'_{\mathsf{C}}, B(x))(b_i)$ for $i = 1, 2$ and thus convexity for dtCQs is not violated.

Note that Condition 3(s2) additionally requires the non-existence of a certain concept E which can be viewed as an interpolant between C and $\exists r.D$ that uses only closed predicates. The following example illustrates why this condition is needed.

Example 5.9. Let $\mathcal{T} = \{A \sqsubseteq \exists r.E, E \sqsubseteq B\}$ and first assume that $\Sigma_{\mathsf{C}} = \{r\}$. Then the CI $A \sqsubseteq \exists r.B$ satisfies Condition 3(s2) and thus $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe. Now let $\Sigma'_{\mathsf{C}} = \{r, E\}$. In this case, the CI $A \sqsubseteq \exists r.B$ does not violate safeness because E can be used as a 'closed interpolant'. Indeed, it is not difficult to show that $(\mathcal{T}, \Sigma'_{\mathsf{C}})$ is both safe and convex for dtCQs.

We now formulate our dichotomy result for $\mathcal{EL}$.

Theorem 5.10. *Let $(\mathcal{T}, \Sigma_{\mathsf{C}})$ be an $\mathcal{EL}$ TBox with closed predicates. Then*
(1) *If $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe, then $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not convex for dtCQs and evaluating dtCQs w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is* coNP*-hard.*
(2) *If $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is safe, then $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is convex for tCQs and*
  (a) *$(\mathcal{T}, \Sigma_{\mathsf{C}})$ and $(\mathcal{T}, \emptyset)$ are UCQ inseparable on consistent ABoxes.*
  (b) *UCQ evaluation w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is in* PTIME.

Towards a proof of Theorem 5.10, we start with introducing canonical models, prove some fundamental lemmas regarding such models, and establish a variant of the Craig interpolation property for $\mathcal{EL}$ that we use to address Condition 3(s2) of Definition 5.7. In fact, we introduce several versions of canonical models. For the proof of Point (1) of Theorem 5.10, we use *finite* canonical models for $\mathcal{EL}$ TBoxes and $\mathcal{EL}$ concepts. For Point (2) and for establishing Craig interpolation, we use (essentially) *tree-shaped* canonical models of $\mathcal{EL}$ TBoxes and possibly infinite ABoxes and, as a special case, the same kind of canonical models of $\mathcal{EL}$ TBoxes and $\mathcal{EL}$ concepts. The constructions of all these canonical models does not involve closed predicates. However, to deal with closed predicates in the proofs, it turns out that we need a more careful definition of (tree-shaped) canonical models than usual.

We start with the definition of finite canonical models for $\mathcal{EL}$ TBoxes $\mathcal{T}$ and $\mathcal{EL}$ concepts C. Take for every $D \in \mathsf{sub}(\mathcal{T}, C)$ an individual name a_D and define the *canonical model* $\mathcal{I}_{\mathcal{T},C} = (\Delta^{\mathcal{I}_{\mathcal{T},C}}, \cdot^{\mathcal{I}_{\mathcal{T},C}})$ of $\mathcal{T}$ and C as follows:

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

Deciding whether $\mathcal{T} \models C \sqsubseteq D$ is in PTIME [7], and thus $\mathcal{I}_{\mathcal{T},C}$ can be constructed in time polynomial in the size of $\mathcal{T}$ and C. The following lemma, shown in [46] as Lemma 12, is the reason for why $\mathcal{I}_{\mathcal{T},C}$ is called a canonical model.

Lemma 5.11. *Let C be an $\mathcal{EL}$ concept and $\mathcal{T}$ an $\mathcal{EL}$ TBox. Then*

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

The next lemma, shown in [46] as Lemma 16, is concerned with the implication of existential restrictions in $\mathcal{EL}$. We will use it in proofs below. It is proved using Lemma 5.11 and the construction of canonical models.

Lemma 5.12. *Suppose $\mathcal{T} \models C \sqsubseteq \exists r.D$, where C, D are $\mathcal{EL}$ concepts and $\mathcal{T}$ is an $\mathcal{EL}$ TBox. Then one of the following holds:*

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

We next construct tree-shaped canonical models. We start with canonical models $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ of an $\mathcal{EL}$ TBox $\mathcal{T}$ and a (possibly infinite) ABox $\mathcal{A}$. In the construction of $\mathcal{J}_{\mathcal{T},\mathcal{A}}$, we use *extended ABoxes* that additionally admit assertions of the form $C(a)$ with C an arbitrary $\mathcal{EL}$ concept. We construct a sequence of extended ABoxes $\mathcal{A}_0, \mathcal{A}_1, \dots$, starting with $\mathcal{A}_0 = \mathcal{A}$. In what follows, we use additional individual names of the form $a \cdot r_1 \cdot C_1 \cdots r_k \cdot C_k$ with $a \in \mathsf{Ind}(\mathcal{A}_0)$, $r_1, \dots, r_k$ role names that occur in $\mathcal{T}$, and $C_1, \dots, C_k \in \mathsf{sub}(\mathcal{T})$. We set $\mathsf{tail}(a \cdot r_1 \cdot C_1 \cdots r_k \cdot C_k) = C_k$. Each extended ABox $\mathcal{A}_{i+1}$ is obtained from $\mathcal{A}_i$ by applying the following rules (the interpretation $\mathcal{I}_{\mathcal{A}_i}$ corresponding to the extended ABox $\mathcal{A}_i$ ignores assertions $C(a)$ with C not a concept name):

(R1) if $C \sqcap D(a) \in \mathcal{A}_i$, then add $C(a)$ and $D(a)$ to $\mathcal{A}_i$;
(R2) if $a \in C^{\mathcal{I}_{\mathcal{A}_i}}$ and $C \sqsubseteq D \in \mathcal{T}$, then add $D(a)$ to $\mathcal{A}_i$;
(R3) if $\exists r.C(a) \in \mathcal{A}_i$ and there exist $b \in \mathsf{Ind}(\mathcal{A}_i)$ with $r(a,b) \in \mathcal{A}_i$ and $\mathcal{A}_i \models (\mathcal{T}, \emptyset, q_C)(b)$, then add $C(b)$ to $\mathcal{A}_i$; otherwise add $r(a, a \cdot r \cdot C)$ and $C(a \cdot r \cdot C)$ to $\mathcal{A}_i$. (Recall that q_C denotes the directed tree CQ corresponding to the concept C.)

Let $\mathcal{A}_c = \bigcup_{i \geq 0} \mathcal{A}_i$. Note that $\mathcal{A}_c$ may be infinite even if $\mathcal{A}$ is finite. Also note that rule (R3) carefully avoids to introduce fresh successors as witnesses for existential restrictions when this is not strictly necessary. This will be useful when closing predicates which might preclude the introduction of fresh successors. Let $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ be the interpretation that corresponds to $\mathcal{A}_c$. Points 1 and 2 of the following lemma show that $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ is canonical, essentially in the sense of Lemma 5.11, and Points 3 and 4 show that, in addition, it is *universal* for UCQs: answers given by $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ coincide with the certain answers.

Lemma 5.13. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $\mathcal{A}$ a possibly infinite ABox. Then*

(1) *$\mathcal{J}_{\mathcal{T},\mathcal{A}}$ is a model of $\mathcal{T}$ and $\mathcal{A}$;*
(2) *for all $p \in \Delta^{\mathcal{J}_{\mathcal{T},\mathcal{A}}} \setminus \mathsf{Ind}(\mathcal{A})$ and all $\mathcal{EL}$ concepts D: $p \in D^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$ iff $\mathcal{T} \models \mathsf{tail}(p) \sqsubseteq D$;*
(3) *for every model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$, there is a homomorphism h from $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ to $\mathcal{I}$ that preserves $\mathsf{Ind}(\mathcal{A})$;*
(4) *for all UCQs $q(\vec{x})$ and tuples $\vec{a}$ in $\mathsf{Ind}(\mathcal{A})$: $\mathcal{A} \models (\mathcal{T}, \emptyset, q)(\vec{a})$ iff $\mathcal{J}_{\mathcal{T},\mathcal{A}} \models q(\vec{a})$.*

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

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

Then $\mathcal{J}_{\mathcal{T},C} := \mathcal{J}_{\mathcal{T},\mathcal{A}_C}$ is the *tree-shaped canonical model* of $\mathcal{T}$ and C. The following is an easy consequence of Lemma 5.13.

Lemma 5.14. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and C an $\mathcal{EL}$ concept. Then*

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

We next give a lemma that connects an answers a to a dtCQs q_C on an ABox $\mathcal{A}$ under a TBox $\mathcal{T}$ with the entailment by $\mathcal{T}$ of concept inclusions of the form $C_a^m \sqsubseteq C$ where C_a^m is obtained by unfolding $\mathcal{A}$ at a up to depth m. More precisely, for every $m \geq 0$ define

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

The following is shown in [46] as Lemma 22.

Lemma 5.15. *For all $\mathcal{EL}$ TBoxes $\mathcal{T}$, $\mathcal{EL}$ concepts C, ABoxes $\mathcal{A}$, and $a \in \mathsf{Ind}(\mathcal{A})$:*

$$\mathcal{A} \models (\mathcal{T}, \emptyset, q_C)(a) \quad \textit{iff} \quad \exists m \geq 0: \quad \mathcal{T} \models C_a^m \sqsubseteq C$$

We now establish a variant of the Craig interpolation property that is suitable for addressing Condition 3(s2) of Definition 5.7. It has been studied before for $\mathcal{ALC}$ and several of its extensions in the context of query rewriting for DBoxes and of Beth definability [36, 57]. Note that it is different from the interpolation property investigated in [46] for $\mathcal{EL}$, which requires the interpolant to be a TBox instead of a concept. For brevity, we set $\mathsf{sig}(\mathcal{T}, C) = \mathsf{sig}(\mathcal{T}) \cup \mathsf{sig}(C)$ for any TBox $\mathcal{T}$ and concept C.

Lemma 5.16 ($\mathcal{EL}$ Interpolation)**.** *Let $\mathcal{T}_1, \mathcal{T}_2$ be $\mathcal{EL}$ TBoxes and let D_1, D_2 be $\mathcal{EL}$ concepts with $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_1 \sqsubseteq D_2$ and $\mathsf{sig}(\mathcal{T}_1, D_1) \cap \mathsf{sig}(\mathcal{T}_2, D_2) = \Sigma$. Then there exists an $\mathcal{EL}$ concept F such that $\mathsf{sig}(F) \subseteq \Sigma$, $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_1 \sqsubseteq F$, and $\mathcal{T}_1 \cup \mathcal{T}_2 \models F \sqsubseteq D_2$.*

Proof. Let $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_1 \sqsubseteq D_2$ with $\mathsf{sig}(\mathcal{T}_1, D_1) \cap \mathsf{sig}(\mathcal{T}_2, D_2) = \Sigma$. Assume that the required $\mathcal{EL}$ concept F does not exist. Consider the tree-shaped canonical model $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}$. Denote by $\mathcal{A}_\Sigma$ the ABox corresponding to the Σ-reduct of $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}$, thus

$$\mathcal{A}_\Sigma = \bigcup_{A\in\Sigma} \{A(a) \mid a \in A^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}}\} \cup \bigcup_{r\in\Sigma} \{r(a,b) \mid r(a,b) \in r^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}}\}$$

We may assume w.l.o.g. that $\mathsf{Ind}(\mathcal{A}_\Sigma) = \Delta^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}}$. Recall that the individual names in $\mathcal{A}_\Sigma$ are paths. For the sake of readability, we denote them by a_p rather than p. Also recall that q_D denotes the dtCQ corresponding to the $\mathcal{EL}$ concept D.

Claim. $\mathcal{A}_\Sigma \not\models (\mathcal{T}_1 \cup \mathcal{T}_2, \emptyset, q_{D_2})(a_{D_1})$.

Proof of claim. Assume for a proof by contradiction that $\mathcal{A}_\Sigma \models (\mathcal{T}_1 \cup \mathcal{T}_2, \emptyset, q_{D_2})(a_{D_1})$. By Lemma 5.15, there is an $\mathcal{EL}$ concept F such that $\mathsf{sig}(F) \subseteq \Sigma$, $\mathcal{T}_1 \cup \mathcal{T}_2 \models F \sqsubseteq D_2$, and $a_{D_1} \in F^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}}$. Then, using Lemma 5.14, we obtain $\mathcal{T}_1 \cup \mathcal{T}_2 \models D_1 \sqsubseteq F$. This contradicts our assumption that no such concept F exists. $\dashv$

Obviously, $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}$ is a model of $\mathcal{T}_1 \cup \mathcal{T}_2$ and $\mathcal{A}_\Sigma$. Then, by Lemma 5.13, there is a homomorphism h from $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}$ to $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}$ with $h(a) = a$ for all $a \in \mathsf{Ind}(\mathcal{A}_\Sigma)$. Conversely, $h' = \{a \mapsto a \mid a \in \Delta^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}}\}$ is a homomorphism from the Σ-reduct of $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}$ to $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}$. Define the interpretation $\mathcal{I}$ as follows:

$$\begin{array}{rcl} \Delta^{\mathcal{I}} & = & \Delta^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}} \\ P^{\mathcal{I}} & = & P^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}} \cup P^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}}, \text{ for all } P \in \mathsf{sig}(\mathcal{T}_1, D_1) \setminus \Sigma \\ P^{\mathcal{I}} & = & P^{\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,\mathcal{A}_\Sigma}}, \text{ for all } P \notin \mathsf{sig}(\mathcal{T}_1, D_1) \setminus \Sigma \end{array}$$

Observe that the mapping h defined above is a homomorphism from $\mathcal{I}$ to $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}$ with $h(a) = a$ for all $a \in \mathsf{Ind}(\mathcal{A}_\Sigma)$. Conversely, the mapping h' defined above is a homomorphism from the $\mathsf{sig}(\mathcal{T}_1, D_1)$-reduct of $\mathcal{J}_{\mathcal{T}_1\cup\mathcal{T}_2,D_1}$ to $\mathcal{I}$. Now it is readily checked that $\mathcal{EL}$ concepts C

are preserved under homomorphisms (if $d \in C^{\mathcal{I}_1}$, then $h(d) \in C^{\mathcal{I}_2}$ if h is a homomorphism from $\mathcal{I}_1$ to $\mathcal{I}_2$). Thus, $\mathcal{I}$ is a model of $\mathcal{T}_1$ since $\mathcal{J}_{\mathcal{T}_1 \cup \mathcal{T}_2, D_1}$ is a model of $\mathcal{T}_1$ and $a_{D_1} \in D_1^{\mathcal{I}}$ since $a_{D_1} \in D_1^{\mathcal{J}_{\mathcal{T}_1 \cup \mathcal{T}_2, D_1}}$. Moreover, by construction, $\mathcal{I}$ is a model of $\mathcal{T}_2$ and $a_{D_1} \notin D_2^{\mathcal{I}}$, by the claim proved above. We have shown that $\mathcal{T}_1 \cup \mathcal{T}_2 \not\models D_1 \sqsubseteq D_2$ and thus derived a contradiction. $\square$

We are now in the position to prove Theorem 5.10. We first prove Point (1). The proof requires two separate constructions that both show non-convexity for dtCQs and address Cases 3(s1) and 3(s2) from Definition 5.7. It then follows from Lemma 5.3 that dtCQ evaluation w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is coNP-hard.

We begin by considering Case 3(s1).

Lemma 5.17. *Let $(\mathcal{T}, \Sigma_{\mathsf{C}})$ be an $\mathcal{EL}$ TBox with closed predicates such that safeness is violated by an inclusion $C \sqsubseteq \exists r.D$ because Condition 3(s1) from Definition 5.7 holds. Then $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not convex for dtCQs.*

Proof. Assume $C \sqsubseteq \exists r.D$ satisfies $\mathcal{T} \models C \sqsubseteq \exists r.D$, there is no tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$, $r \notin \Sigma_{\mathsf{C}}$, and $\mathsf{sig}(D) \cap \Sigma_{\mathsf{C}} \neq \emptyset$. Consider the finite canonical model $\mathcal{I}_{\mathcal{T},C}$ of $\mathcal{T}$ and C. Assume w.l.o.g. that C does not occur in $\mathcal{T}$ (if it does, replace C by $A \sqcap C$ for a fresh concept name A). Note that it follows that there is no $a \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$ with $(a, a_C) \in s^{\mathcal{I}_{\mathcal{T},C}}$ for any role name s.

Let $\mathcal{I}_r$ be the interpretation obtained from $\mathcal{I}_{\mathcal{T},C}$ by removing all pairs (a_C, a_E) from $r^{\mathcal{I}_{\mathcal{T},C}}$ such that $\exists r.E$ is not a tlc of C. Let $\mathcal{A}_r$ be the ABox corresponding to $\mathcal{I}_r$ and let $\mathcal{A}$ be the disjoint union of two copies of $\mathcal{A}_r$. We denote the individual names of the first copy by $(a, 1)$, $a \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$, and the individual names of the second copy by $(a, 2)$, $a \in \Delta^{\mathcal{I}_{\mathcal{T},C}}$. Let A_1 and A_2 be fresh concept names and set

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

Some predicate $P \in \Sigma_{\mathsf{C}}$ occurs in D. If a concept name $E \in \Sigma_{\mathsf{C}}$ occurs in D, then fix one such E and denote, for $i \in \{1, 2\}$, by D_i the resulting concept after one occurrence of E is replaced by $A_i \sqcap E$. For example, if $D = A \sqcap \exists s_1.E \sqcap \exists s_2.E$, $E \in \Sigma_{\mathsf{C}}$ and $A \notin \Sigma_{\mathsf{C}}$, then either $D_i = A \sqcap \exists s_1.(A_i \sqcap E) \sqcap \exists s_2.E$ or $D_i = A \sqcap \exists s_1.E \sqcap \exists s_2.(A_i \sqcap E)$. Similarly, if no concept name from Σ_{C} occurs in D, then let $s \in \Sigma_{\mathsf{C}}$ be a role name such that a concept of the form $\exists s.G$ occurs in D. Denote by D_i the resulting concept after one occurrence of $\exists s.G$ is replaced by $A_i \sqcap \exists s.G$.

We now use $\mathcal{A}'$ and the dtCQs $q_{\exists r.D_i}$ to prove that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not convex for dtCQs. Using the condition $\mathcal{T} \models C \sqsubseteq \exists r.D$ and the construction of D_1 and D_2, it is straightforward to show that $\mathcal{A}' \models (\mathcal{T}, \Sigma_{\mathsf{C}}, q_{\exists r.D_1} \vee q_{\exists r.D_2})(a_C, 1)$. We show that $\mathcal{A}' \not\models (\mathcal{T}, \Sigma_{\mathsf{C}}, q_{\exists r.D_i})(a_C, 1)$ for $i = 1, 2$. Let $i = 1$ (the case $i = 2$ is similar and omitted). We construct a model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}'$ that respects Σ_{C} with $(a_C, 1) \notin (\exists r.D_1)^{\mathcal{J}}$. $\mathcal{J}$ is defined as the interpretation corresponding to the ABox $\mathcal{A}'$ extended by

$$\{r((a_C,1),(e_E,2)), r((a_C,2),(a_E,1)) \mid (a_C,a_E) \in r^{\mathcal{I}_{\mathcal{T},C}} \setminus r^{\mathcal{I}_r}\}$$

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

- there exists a tlc $\exists r.E$ of C such that $(a_E, 1) \in D_1^{\mathcal{J}}$;
- there exists a_E with $(a_C, a_E) \in r^{\mathcal{I}_{\mathcal{T},C}}$ such that $(a_E, 2) \in D_1^{\mathcal{J}}$.

If the first condition holds, then $a_E \in D^{\mathcal{I}_{\mathcal{T},C}}$. Then, by Lemma 5.11, $\mathcal{T} \models E \sqsubseteq D$ for a tlc $\exists r.E$ of C which contradicts Point (2) of the definition of safeness. The second condition does not hold since $(a_E, 2) \in G^{\mathcal{J}}$ iff $(a_E, 2) \in G^{\mathcal{J}|_{\{(a,2)|a\in\Delta^{\mathcal{I}}\}}}$, for every $\mathcal{EL}$ concept G, and $A_1^{\mathcal{J}} \cap \{(a,2) \mid a \in \Delta^{\mathcal{I}}\} = \emptyset$, but D_1 contains A_1. □

We now consider Case 3(s2) from Definition 5.7.

Lemma 5.18. *Let $(\mathcal{T}, \Sigma_{\mathsf{C}})$ be an $\mathcal{EL}$ TBox with closed predicates such that safeness is violated by an inclusion $C \sqsubseteq \exists r.D$ because Condition 3(s2) from Definition 5.7 holds. Then $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not convex for dtCQs.*

Proof. Assume $C \sqsubseteq \exists r.D$ satisfies $\mathcal{T} \models C \sqsubseteq \exists r.D$, there is no tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$, and Condition 3(s2) holds. Let

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

Observe that since there is no tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$, by Lemma 5.12, there exists $G \in K$ with $\mathcal{T} \models G \sqsubseteq D$. We now apply the interpolation lemma. Obtain $\mathcal{T}^i$ from $\mathcal{T}$ by replacing every predicate $P \notin \Sigma_{\mathsf{C}}$ by a fresh predicate P_i of the same arity, $i \in \{1, 2\}$. Similarly, for any $\mathcal{EL}$ concept F we denote by F^i the resulting concept when every predicate $P \notin \Sigma_{\mathsf{C}}$ is replaced by a fresh predicate P_i of the same arity, $i \in \{1, 2\}$. We show the following using the interpolation lemma.

Claim 1. For all $G \in K$: $\mathcal{T}^1 \cup \mathcal{T}^2 \not\models G^1 \sqsubseteq D^2$.

Proof of claim. The proof is indirect. Assume there exists $G \in K$ such that $\mathcal{T}^1 \cup \mathcal{T}^2 \models G^1 \sqsubseteq D^2$. By Lemma 5.16, there exists an $\mathcal{EL}$ concept F with $\mathsf{sig}(F) \subseteq \Sigma_{\mathsf{C}}$ such that $\mathcal{T}^1 \cup \mathcal{T}^2 \models G^1 \sqsubseteq F$ and $\mathcal{T}^1 \cup \mathcal{T}^2 \models F \sqsubseteq D^2$. Then $\mathcal{T} \models G \sqsubseteq F$ and $\mathcal{T} \models F \sqsubseteq D$. But then we obtain from $\mathcal{T} \models C \sqsubseteq \exists r.G$ that $\mathcal{T} \models C \sqsubseteq \exists r.F$ which contradicts Condition 3(s2). ⊣

By Claim 1 we can take the finite canonical models $\mathcal{J}_G := \mathcal{I}_{\mathcal{T}^1\cup\mathcal{T}^2,G^1}$, $G \in K$, and obtain for $a_G := a_{G^0}$ that $a_G \notin (D^2)^{\mathcal{J}_G}$. Let $\mathcal{A}_{G,\Sigma_{\mathsf{C}}}$ be the ABox corresponding to the Σ_{C}-reduct of $\mathcal{J}_G$. We may assume that the sets of individual names $\mathsf{Ind}(\mathcal{A}_{G,\Sigma_{\mathsf{C}}})$ are mutually disjoint, for $G \in K$, and that $a_G \in \mathsf{Ind}(\mathcal{A}_{G,\Sigma_{\mathsf{C}}})$, for all $G \in K$.

Claim 2. For every $G \in K$, there exist

- a model $\mathcal{I}_G^1$ of $\mathcal{T}$ and $\mathcal{A}_{G,\Sigma_{\mathsf{C}}}$ that respects the closed predicates Σ_{C} such that $\Delta^{\mathcal{I}_G^1} = \mathsf{Ind}(\mathcal{A}_{G,\Sigma_{\mathsf{C}}})$, $a_G \in G^{\mathcal{I}_G^1}$, and $a_G \in H^{\mathcal{I}_G^1}$ only if $\mathcal{T} \models G \sqsubseteq H$, for all $\mathcal{EL}$ concepts H;
- a model $\mathcal{I}_G^2$ of $\mathcal{T}$ and $\mathcal{A}_{G,\Sigma_{\mathsf{C}}}$ that respects the closed predicates Σ_{C} such that $\Delta^{\mathcal{I}_G^2} = \mathsf{Ind}(\mathcal{A}_{G,\Sigma_{\mathsf{C}}})$, $a_G \notin D^{\mathcal{I}_G^2}$, and $a_G \in H^{\mathcal{I}_G^2}$ only if $\mathcal{T} \models G \sqsubseteq H$, for all $\mathcal{EL}$ concepts H.

Proof of claim. The interpretation $\mathcal{I}_G^1$ is obtained from $\mathcal{J}_G$ by setting $P^{\mathcal{I}_G^1} := (P^1)^{\mathcal{J}_G}$ for all predicates $P \in \mathsf{sig}(\mathcal{T}, C, D) \setminus \Sigma_{\mathsf{C}}$ and $P^{\mathcal{I}_G^1} := \emptyset$ for all predicates P not in $\mathsf{sig}(\mathcal{T}, C, D) \cup \Sigma_{\mathsf{C}}$. The properties stated follow from the properties of the finite canonical model $\mathcal{I}_{\mathcal{T}^1\cup\mathcal{T}^2,G^1}$. In particular, $a_G \in H^{\mathcal{I}_G^1}$ only if $\mathcal{T} \models G \sqsubseteq H$ follows from Lemma 5.12, Point 2. The interpretation $\mathcal{I}_G^2$ is obtained from $\mathcal{J}_G$ by setting $P^{\mathcal{I}_G^2} := (P^2)^{\mathcal{J}_G}$ for all predicates $P \in \mathsf{sig}(\mathcal{T}, C, D) \setminus \Sigma_{\mathsf{C}}$ and $P^{\mathcal{I}_G^2} := \emptyset$ for all predicates P not in $\mathsf{sig}(\mathcal{T}, C, D) \cup \Sigma_{\mathsf{C}}$. The properties stated follow again from the properties of the finite canonical model $\mathcal{I}_{\mathcal{T}^1\cup\mathcal{T}^2,G^1}$. ⊣

Introduce two copies $\mathcal{A}^1_{G,\Sigma_{\mathsf{C}}}$ and $\mathcal{A}^2_{G,\Sigma_{\mathsf{C}}}$ of $\mathcal{A}_{G,\Sigma_{\mathsf{C}}}$, for $G \in K$. We denote the individual names of the first copy by $(a, 1)$, for $a \in \mathsf{Ind}(\mathcal{A}_{G,\Sigma_{\mathsf{C}}})$, and the individual names of the second copy by $(a,2)$, for $a \in \mathsf{Ind}(\mathcal{A}_{G,\Sigma_\mathsf{C}})$. Let $\mathcal{A}_r$ be the ABox defined in the beginning of the proof of Lemma 5.17. Define the ABox $\mathcal{A}$ by taking two fresh concept names A_1 and A_2 and adding to

$$\mathcal{A}_r \cup \bigcup_{G \in K} \mathcal{A}^1_{G,\Sigma_\mathsf{C}} \cup \mathcal{A}^2_{G,\Sigma_\mathsf{C}}$$

the assertions

- $r(a_C,(a_G,1)), r(a_C,(a_G,2))$, for every $G \in K$;
- $A_1(a_G,1)$, for every $G \in K$;
- $A_1(a_E)$, for every tlc $\exists r.E$ of C;
- $A_2(a_G,2)$, for every $G \in K$.

We use $\mathcal{A}$ and the dtCQs $q_{\exists r.(A_i \sqcap D)}$ to show that $(\mathcal{T},\Sigma_\mathsf{C})$ is not convex for dtCQs. The proof that $\mathcal{A} \models (\mathcal{T},\Sigma_\mathsf{C}, q_{\exists r.(A_1 \sqcap D)} \vee q_{\exists r.(A_2 \sqcap D)})(a_C)$ is straightforward using the condition that $\mathcal{T} \models C \sqsubseteq \exists r.D$ and the construction of $\mathcal{A}$ ($r \in \Sigma_\mathsf{C}$ and all r-successors of a_C in $\mathcal{A}$ are either in A_1 or in A_2). It remains to show $\mathcal{A} \not\models (\mathcal{T},\Sigma_\mathsf{C}, q_{\exists r.(A_i \sqcap D)})(a_C)$, for $i = 1,2$. For $i=2$, construct a witness interpretation $\mathcal{J}$ showing this by expanding all $\mathcal{A}^2_{G,\Sigma_\mathsf{C}}$, $G \in K$, to (isomorphic copies of) $\mathcal{I}^2_G$, all $\mathcal{A}^1_{G,\Sigma_\mathsf{C}}$, $G \in K$, to (isomorphic copies of) $\mathcal{I}^1_G$, and $\mathcal{A}_r$ to $\mathcal{I}_r$. Using the properties of $\mathcal{I}^1_G$ and $\mathcal{I}^2_G$ established in the claim above, it is readily checked that $\mathcal{J}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects Σ_C. Moreover, $a_C \notin (\exists r.(A_2 \sqcap D))^\mathcal{J}$ since $(a_G,2) \notin D^\mathcal{J}$ for any $G \in K$ (by Claim 2).

For $i=1$, construct a witness interpretation $\mathcal{J}$ showing this by expanding all $\mathcal{A}^2_{G,\Sigma_\mathsf{C}}$, $G \in K$, to (isomorphic copies of) $\mathcal{I}^1_G$, all $\mathcal{A}^1_{G,\Sigma_\mathsf{C}}$, $G \in K$, to (isomorphic copies of) $\mathcal{I}^2_G$, and $\mathcal{A}_r$ to $\mathcal{I}_r$. Using again the properties of $\mathcal{I}^1_G$ and $\mathcal{I}^2_G$ established above, it can be checked that $\mathcal{J}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects Σ_C. $a_C \notin (\exists r.(A_1 \sqcap D))^\mathcal{J}$ since $a_E \notin D^\mathcal{J}$ for any tlc $\exists r.E$ of C and since $(a_G,1) \notin D^\mathcal{J}$ for any $G \in K$ (by Claim 2). This finishes the proof. □

This finishes the proof of Point (1) of Theorem 5.10. We now prove Part (a) of Point (2). The proof strategy is exactly the same as in the proof for DL-Lite$_\mathcal{R}$.

Lemma 5.19. *Let $(\mathcal{T},\Sigma_\mathsf{C})$ be a safe $\mathcal{EL}$ TBox with closed predicates. Then for every UCQ q, we have*

$$\mathcal{A} \models (\mathcal{T},\Sigma_\mathsf{C},q)(\vec{a}) \quad \textit{iff} \quad \mathcal{A} \models (\mathcal{T},\emptyset,q)(\vec{a})$$

for all ABoxes $\mathcal{A}$ that are consistent w.r.t. $(\mathcal{T},\Sigma_\mathsf{C})$ and all $\vec{a} \in \mathsf{Ind}(\mathcal{A})$.

Proof. Let $(\mathcal{T},\Sigma_\mathsf{C})$ be safe and assume that $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T},\Sigma_\mathsf{C})$. We consider the tree-shaped canonical model $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ introduced above. It suffices to show that $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ respects Σ_C. To this end it suffices to prove for all $A, r \in \Sigma_\mathsf{C}$:

(1) for all $a \in \mathsf{Ind}(\mathcal{A})$, if $a \in A^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$, then $A(a) \in \mathcal{A}$;
(2) for all $a,b \in \mathsf{Ind}(\mathcal{A})$, if $r(a,b) \in r^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$, then $r(a,b) \in \mathcal{A}$;
(3) for all $a \in \mathsf{Ind}(\mathcal{A})$ and $C \in \mathsf{sub}(\mathcal{T})$, $a \cdot r \cdot C \notin \Delta^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$;
(4) for all $d \in \Delta^{\mathcal{J}_{\mathcal{T},\mathcal{A}}} \setminus \mathsf{Ind}(\mathcal{A})$, there is no $\mathcal{EL}$ concept D with $d \in D^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$ and $\mathsf{sig}(D) \cap \Sigma_\mathsf{C} \neq \emptyset$.

For Item (1), assume $a \in A^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$. By Lemma 5.13, $\mathcal{A} \models (\mathcal{T},\emptyset,A(x))(a)$, and so we obtain $\mathcal{A} \models (\mathcal{T},\Sigma_\mathsf{C},A(x))(a)$. By consistency of $\mathcal{A}$ w.r.t. $(\mathcal{T},\Sigma_\mathsf{C})$, we then have $A(a) \in \mathcal{A}$. Item (2) follows directly from the construction of $\mathcal{J}_{\mathcal{T},\mathcal{A}}$. For Item (3), assume for a proof by contradiction that there are $a \in \mathsf{Ind}(\mathcal{A})$, $r \in \Sigma_\mathsf{C}$, and a concept C such that $a \cdot r \cdot C \in \Delta^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$. By Lemma 5.13, we have $\mathcal{A} \models (\mathcal{T},\emptyset,q_{\exists r.C})(a)$. By Lemma 5.15, this implies that there is some $m \geq 0$ with $\mathcal{T} \models C^m_a \sqsubseteq \exists r.C$, where C^m_a is the unfolding of $\mathcal{A}$ at a of depth m. We show that this contradicts the assumption that $(\mathcal{T},\Sigma_\mathsf{C})$ is safe. There does not exist a tlc $\exists r.C'$

of C_a^m with $\mathcal{T} \models C' \sqsubseteq C$ because otherwise there is some $b \in \mathsf{Ind}(\mathcal{A})$ with $r(a,b) \in \mathcal{A}$ and $\mathcal{A} \models (\mathcal{T}, \emptyset, q_C)(b)$ and thus, $a \cdot r \cdot C$ would have never been introduced by Rule (R3) in the construction of $\mathcal{J}_{\mathcal{T},\mathcal{A}}$. Moreover, there is no $\mathcal{EL}$ concepts E with $\mathsf{sig}(E) \subseteq \Sigma_{\mathsf{C}}$ and $\mathcal{T} \models C_a^m \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq C$ because otherwise there is a $b \in \mathsf{Ind}(\mathcal{A})$ with $r(a,b) \in \mathcal{A}$ and $\mathcal{I}_\mathcal{A} \models E(b)$ since $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$. But then $\mathcal{A} \models (\mathcal{T}, \emptyset, q_C)(b)$ by the fact that $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ is a model of $\mathcal{T}$ and $\mathcal{A}$. Again, in this case, $a \cdot r \cdot C$ would have never been introduced by Rule (R3) in the construction of $\mathcal{J}_{\mathcal{T},\mathcal{A}}$. Hence $C_a^m \sqsubseteq \exists r.C$ witnesses that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe.

For Item (4), assume for a proof by contradiction that there is a $d \in \Delta^{\mathcal{J}_{\mathcal{T},\mathcal{A}}} \setminus \mathsf{Ind}(\mathcal{A})$ and an $\mathcal{EL}$ concept D such that $d \in D^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$ and $\mathsf{sig}(D) \cap \Sigma_{\mathsf{C}} \neq \emptyset$. By definition, $d = a \cdot r_0 \cdot C_0 \cdots r_n \cdot C_n$ for some $a \in \mathsf{Ind}(\mathcal{A})$. Let $G = C_0 \sqcap \exists r_1.\exists r_2 \ldots \exists r_n.D$. Obviously, $\mathsf{sig}(G) \cap \Sigma_{\mathsf{C}} \neq \emptyset$ and $a \cdot r_0 \cdot C_0 \in G^{\mathcal{J}_{\mathcal{T},\mathcal{A}}}$. By the latter and Lemma 5.13, we have $\mathcal{A} \models (\mathcal{T}, \emptyset, q_{\exists r_0.G})(a)$. By Lemma 5.15, this implies that there is some $m \geq 0$ with $\mathcal{T} \models C_a^m \sqsubseteq \exists r_0.G$. We show that it follows that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe. We have $\mathsf{sig}(G) \cap \Sigma_{\mathsf{C}} \neq \emptyset$ and, by Item (3), $r_0 \notin \Sigma_{\mathsf{C}}$. To show that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe it remains to show that there is no tlc $\exists r_0.C'$ of C_a^m with $\mathcal{T} \models C' \sqsubseteq G$. This is indeed the case because otherwise there is some $b \in \mathsf{Ind}(\mathcal{A})$ with $r_0(a,b) \in \mathcal{A}$ and $\mathcal{A} \models (\mathcal{T}, \emptyset, q_{C_0})(b)$ and thus, $a \cdot r_0 \cdot C_0$ would have never been introduced by Rule (R3) in the construction of $\mathcal{J}_{\mathcal{T},\mathcal{A}}$. □

This finishes the proof of Part (a) of Point (2) of Theorem 5.10. Before we prove Part (b) of Point (2) we show the following observation of independent interest.

Lemma 5.20. *Let $(\mathcal{T}, \Sigma_{\mathsf{C}})$ be a safe $\mathcal{EL}$ TBox with closed predicates. Then there exists an $\mathcal{EL}$ TBox $\mathcal{T}'$ equivalent to $\mathcal{T}$ such that for any $C \sqsubseteq D \in \mathcal{T}'$, $\mathsf{sig}(D) \subseteq \Sigma_{\mathsf{C}}$ or $\mathsf{sig}(D) \cap \Sigma_{\mathsf{C}} = \emptyset$.*

Proof. We apply the following three rules exhaustively (and recursively) to $\mathcal{T}$:

- replace any $C \sqsubseteq D_1 \sqcap D_2$ by $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$;
- replace any $C \sqsubseteq \exists r.D$ such that there exists a tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$ by $C' \sqsubseteq D$;
- replace any $C \sqsubseteq \exists r.D$ with $r \in \Sigma_{\mathsf{C}}$ and $\mathsf{sig}(D) \not\subseteq \Sigma_{\mathsf{C}}$ by $C \sqsubseteq \exists r.F$ and $F \sqsubseteq D$, where F is an $\mathcal{EL}$ concepts with $\mathsf{sig}(F) \subseteq \Sigma_{\mathsf{C}}$ such that $\mathcal{T} \models C \sqsubseteq \exists r.F$ and $\mathcal{T} \models F \sqsubseteq D$. (Note that such a concept F always exists by Condition 3(s2).)

It is straightforward to show that the resulting TBox $\mathcal{T}'$ is as required. □

Recall that UCQ evaluation for $\mathcal{EL}$ TBoxes without closed predicates is in PTIME. Thus, the following lemma and Part (a) directly imply Part (b) of Theorem 5.10.

Lemma 5.21. *Let $(\mathcal{T}, \Sigma_{\mathsf{C}})$ be a safe $\mathcal{EL}$ TBox with closed predicates. Then consistency of ABoxes w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is in* PTIME.

Proof. We may assume that $\mathcal{T}$ is in the form of the claim above: for all $C \sqsubseteq D \in \mathcal{T}$, $\mathsf{sig}(D) \subseteq \Sigma_{\mathsf{C}}$ or $\mathsf{sig}(D) \cap \Sigma_{\mathsf{C}} = \emptyset$. We show that the following conditions are equivalent, for every ABox $\mathcal{A}$:

(1) $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$;
(2) for all $C \sqsubseteq F \in \mathcal{T}$ with $\mathsf{sig}(F) \subseteq \Sigma_{\mathsf{C}}$ and all $a \in \mathsf{Ind}(\mathcal{A})$, if $\mathcal{A} \models (\mathcal{T}, \emptyset, q_F)(a)$, then $\mathcal{I}_\mathcal{A} \models F(a)$.

The implication from Condition (1) to Condition (2) is obvious. Conversely, assume that Condition (2) holds. It suffices to show that the tree-shaped canonical model $\mathcal{J}_{\mathcal{T},\mathcal{A}}$ respects closed predicates Σ_{C}. One can readily check that the proofs of Points (2) and (4) of Lemma 5.19 do not use the condition that $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$. Thus, it suffices to prove that Points (1) and (3) of Lemma 5.19 hold for $\mathcal{J}_{\mathcal{T},\mathcal{A}}$. But they follow directly from Condition (2) and the construction of $\mathcal{J}_{\mathcal{T},\mathcal{A}}$. The result now follows from the fact that Condition (2) can be checked in polynomial time in the size of $\mathcal{A}$. □

This finishes the proof of Theorem 5.10.

6. Quantified Query Case: Deciding Tractability of PTime Query Evaluation

We consider the meta problem to decide whether query evaluation w.r.t. a TBox with closed predicates is tractable. We show that the following problems are in PTime:

(1) decide whether UCQ evaluation w.r.t. DL-Lite$_\mathcal{R}$ TBoxes with closed predicates is FO-rewritable (equivalently, in PTime); and
(2) decide whether UCQ evaluation w.r.t. $\mathcal{EL}$ TBoxes with closed predicates is in PTime.

In both cases, we use the characterization via safeness given in the previous section and show that safeness can be decided in PTime. For DL-Lite$_\mathcal{R}$, the proof is actually straightforward: to check safeness of a DL-Lite$_\mathcal{R}$ TBox with closed predicates $(\mathcal{T}, \Sigma_\mathsf{C})$ it suffices to consider all basic concepts B_1, B_2 and roles r from $\mathsf{sig}(\mathcal{T})$ (of which there are only polynomially many) and make satisfiability checks for basic concepts w.r.t. DL-Lite$_\mathcal{R}$ TBoxes and entailment checks of DL-Lite$_\mathrm{core}$ CIs and RIs by DL-Lite$_\mathcal{R}$ TBoxes according to the definition of safeness. Both can be done in polynomial time [20].

Theorem 6.1. *It is in* PTime *to decide whether a DL-Lite$_\mathcal{R}$ TBox with closed predicates is safe.*

Such a straightforward argument does not work for $\mathcal{EL}$ TBoxes since Definition 5.7 quantifies over all $\mathcal{EL}$ concepts C, D, and E, of which there are infinitely many. In the following, we show that, nevertheless, safeness of an $\mathcal{EL}$ TBox with closed predicates $(\mathcal{T}, \Sigma_\mathsf{C})$ can be decided in PTime. The first step of the proof is to convert $\mathcal{T}$ into a *reduced* $\mathcal{EL}$ TBox $\mathcal{T}^*$ with the following properties:

(red1) $\mathcal{T}^*$ contains no CI of the form $C \sqsubseteq D_1 \sqcap D_2$;
(red2) if $C \sqsubseteq \exists r.D \in \mathcal{T}^*$, then there is no tlc $\exists r.C'$ of C with $\mathcal{T}^* \models C' \sqsubseteq D$.

Lemma 6.2. *For every $\mathcal{EL}$ TBox, one can compute in polynomial time an equivalent reduced $\mathcal{EL}$ TBox.*

Proof. Assume that $\mathcal{T}$ is an $\mathcal{EL}$ TBox. Compute $\mathcal{T}^*$ by applying the following two rules exhaustively to $\mathcal{T}$:

- replace any CI $C \sqsubseteq D_1 \sqcap D_2$ with the CIs $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$;
- replace any CI $C \sqsubseteq \exists r.D$ for which there exists a tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D$ by $C' \sqsubseteq D$.

It is straightforward to prove that $\mathcal{T}^*$ is reduced, equivalent to $\mathcal{T}$, and is constructed in polynomial time (using the fact that checking $\mathcal{T} \models C \sqsubseteq D$ is in PTime [7]). □

We now formulate a stronger version of safeness. While Definition 5.7 quantifies over all CIs $C \sqsubseteq \exists r.D$ that are *entailed* by the TBox $\mathcal{T}$, the stronger version only considers CIs of this form that are *contained* in $\mathcal{T}$. For deciding tractability based on safeness, this is clearly a drastic improvement since only the concept E from Definition 5.7 remains universally quantified.

Definition 6.3. An $\mathcal{EL}$ TBox with closed predicates $(\mathcal{T}, \Sigma_\mathsf{C})$ is *strongly safe* if there exists no $\mathcal{EL}$ CI $C \sqsubseteq \exists r.D \in \mathcal{T}$ such that one of the following holds:

(st1) $r \notin \Sigma_{\mathsf{C}}$ and there is some $\mathcal{EL}$ concept E such that $\mathcal{T} \models D \sqsubseteq E$ and $\mathsf{sig}(E) \cap \Sigma_{\mathsf{C}} \neq \emptyset$;
(st2) $r \in \Sigma_{\mathsf{C}}$, $\mathsf{sig}(D) \not\subseteq \Sigma_{\mathsf{C}}$, and there is no $\mathcal{EL}$ concept E with $\mathsf{sig}(E) \subseteq \Sigma_{\mathsf{C}}$ such that $\mathcal{T} \models C \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq D$. △

The crucial observation is that, for $\mathcal{EL}$ TBoxes in reduced form, the original notion of safeness can be replaced by strong safeness.

Lemma 6.4. *If $\mathcal{T}$ is a reduced $\mathcal{EL}$ TBox and Σ_{C} a signature, then $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is safe iff it is strongly safe.*

Proof. Suppose that $\mathcal{T}$ satisfies Conditions (red1) and (red2) for reduced $\mathcal{EL}$ TBoxes.
($\Rightarrow$) Suppose that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not strongly safe, that is, there is some $C \sqsubseteq \exists r.D \in \mathcal{T}$ satisfying (st1) or (st2). If $C \sqsubseteq \exists r.D$ satisfies Condition (st1), then $r \notin \Sigma_{\mathsf{C}}$ and there is some concept E such that $\mathcal{T} \models D \sqsubseteq E$ and $\mathsf{sig}(E) \cap \Sigma_{\mathsf{C}} \neq \emptyset$. We show that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe because the CI $C \sqsubseteq \exists r.(D \sqcap E)$ violates safeness:
(1) $\mathcal{T} \models C \sqsubseteq \exists r.(D \sqcap E)$ since $C \sqsubseteq \exists r.D \in \mathcal{T}$ and $\mathcal{T} \models D \sqsubseteq E$.
(2) there is no tlc $\exists r.C'$ of C with $\mathcal{T} \models C' \sqsubseteq D \sqcap E$; this follows from Condition (red2).
(3) Condition 3(s1) is satisfied because $r \notin \Sigma_{\mathsf{C}}$ and $\mathsf{sig}(D \sqcap E) \cap \Sigma_{\mathsf{C}} \neq \emptyset$ since $\mathsf{sig}(E) \cap \Sigma_{\mathsf{C}} \neq \emptyset$.
If $C \sqsubseteq \exists r.D$ satisfies Condition (st2), then it follows directly that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe because $C \sqsubseteq \exists r.D$ satisfies Condition 3(s2).

($\Leftarrow$) Suppose that $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe. Take any $\mathcal{EL}$ CI $C \sqsubseteq \exists r.D$ violating safeness. In the following, we use the tree-shaped canonical model $\mathcal{J}_{\mathcal{T},C}$ defined above. For the sake of readability denote the individual name p of $\mathcal{J}_{\mathcal{T},C}$ by a_p (in particular, a_C denotes C). Note that Lemma 5.14 yields $a_C \in (\exists r.D)^{\mathcal{J}_{\mathcal{T},C}}$ since $\mathcal{T} \models C \sqsubseteq \exists r.D$. Thus there is some $d \in \Delta^{\mathcal{J}_{\mathcal{T},C}}$ such that $(a_C, d) \in r^{\mathcal{J}_{\mathcal{T},C}}$ and $d \in D^{\mathcal{J}_{\mathcal{T},C}}$. By definition of $\mathcal{J}_{\mathcal{T},C}$, $d = a_{C \cdot r \cdot E}$ for some $\mathcal{EL}$ concept E. By Lemma 5.14 and $d \in D^{\mathcal{J}_{\mathcal{T},C}}$, we have $\mathcal{T} \models E \sqsubseteq D$.

Let $\mathcal{A}_C = \mathcal{A}_0, \mathcal{A}_1, \ldots$ be the ABoxes used in the construction of $\mathcal{J}_{\mathcal{T},C}$. By definition of $\mathcal{A}_C$ and Condition (red2), we have $C \cdot r \cdot E \notin \mathsf{paths}(C)$, that is, $d = a_{C \cdot r \cdot E}$ must have been generated by (R3). Consequently, there is an $i \in \mathbb{N}$ such that $\exists r.E(a_C) \in \mathcal{A}_i$, and $\mathcal{A}_{i+1} = \mathcal{A}_i \cup \{r(a_C, a_{C \cdot r \cdot E}), E(a_{C \cdot r \cdot E})\}$. Using Condition (red1) one can now easily prove that $\exists r.E(a_C)$ can only have been added due to an application of (R2).

Thus, there is some $C' \sqsubseteq \exists r.E \in \mathcal{T}$ with $\mathcal{I}_{\mathcal{A}_j} \models C'(a_C)$. We obtain $a_C \in (C')^{\mathcal{J}_{\mathcal{T},C}}$ and this implies, by Lemma 5.14, that $\mathcal{T} \models C \sqsubseteq C'$. As $(\mathcal{T}, \Sigma_{\mathsf{C}})$ is not safe due to the CI $C \sqsubseteq \exists r.D$, we obtain one of the following cases:

- $C \sqsubseteq \exists r.D$ satisfies Condition 3(s1). Then $r \notin \Sigma_{\mathsf{C}}$ and $\mathsf{sig}(D) \cap \Sigma_{\mathsf{C}} \neq \emptyset$. Since $\mathcal{T} \models E \sqsubseteq D$, we thus have that $C' \sqsubseteq \exists r.E \in \mathcal{T}$ satisfies Condition (st1). We have shown that $\mathcal{T}$ is not strongly safe.
- $C \sqsubseteq \exists r.D$ satisfies Condition 3(s2). Then $r \in \Sigma_{\mathsf{C}}$, $\mathsf{sig}(D) \not\subseteq \Sigma_{\mathsf{C}}$, and there is no $\mathcal{EL}$ concept F with $\mathsf{sig}(F) \subseteq \Sigma_{\mathsf{C}}$ and $\mathcal{T} \models C \sqsubseteq \exists r.F$ and $\mathcal{T} \models F \sqsubseteq D$. We aim at showing that $C' \sqsubseteq \exists r.E \in \mathcal{T}$ satisfies Condition (st2). We already know that $r \in \Sigma_{\mathsf{C}}$. From $\mathcal{T} \models C \sqsubseteq C' \sqsubseteq \exists r.E$ and $\mathcal{T} \models E \sqsubseteq D$, we obtain $\mathsf{sig}(E) \not\subseteq \Sigma_{\mathsf{C}}$ (otherwise set $F := E$ above to derive a contradiction). We also obtain that there is no $\mathcal{EL}$ concept F with $\mathsf{sig}(F) \subseteq \Sigma_{\mathsf{C}}$ and with $\mathcal{T} \models C' \sqsubseteq \exists r.F$ and $\mathcal{T} \models F \sqsubseteq E$. Again it follows that $\mathcal{T}$ is not strongly safe. □

Lemma 6.5. *It is in* PTIME *to decide whether a reduced $\mathcal{EL}$ TBox with closed predicates is strongly safe.*

Proof. Assume $(\mathcal{T},\Sigma_\mathsf{C})$ is a reduced $\mathcal{EL}$ TBox with closed predicates. Assume $C \sqsubseteq \exists r.D \in \mathcal{T}$ is given. It suffices to show that Conditions (st1) and (st2) can be checked in polynomial time. For Condition (st1), it suffices to show that one can check in polynomial time whether there exists an $\mathcal{EL}$ concept E with $\mathcal{T} \models D \sqsubseteq E$ and $\mathsf{sig}(E) \cap \Sigma_\mathsf{C} \neq \emptyset$. We reduce this to a reachability problem in the directed graph induced by the finite canonical model $\mathcal{I}_{\mathcal{T},D}$. In detail, let $G=(V,R)$ be the directed graph with $V=\Delta^{\mathcal{I}_{\mathcal{T},D}}$ and $R=\bigcup_{r\in\mathsf{N_R}} r^{\mathcal{I}_{\mathcal{T},D}}$. Let $T=\bigcup_{A\in\Sigma_\mathsf{C}} A^{\mathcal{I}_{\mathcal{T},D}} \cup \bigcup_{r\in\Sigma_\mathsf{C}} (\exists r)^{\mathcal{I}_{\mathcal{T},D}}$. Using Lemma 5.11, it is readily checked that there exists an $\mathcal{EL}$ concept E with $\mathcal{T} \models D \sqsubseteq E$ and $\mathsf{sig}(E)\cap\Sigma_\mathsf{C}\neq\emptyset$ iff there exists a path from a_D to a node in T in G. The latter reachability problem can be checked in polynomial time.

For Condition (st2), assume $r \in \Sigma_\mathsf{C}$ and let $\mathcal{A}$ denote the ABox corresponding to the Σ-reduct of $\mathcal{I}_{\mathcal{T},C}$. We show that the following conditions are equivalent:

(1) there exists an $\mathcal{EL}$ concept E such that $\mathsf{sig}(E)\subseteq\Sigma_\mathsf{C}$, $\mathcal{T}\models C\sqsubseteq\exists r.E$ and $\mathcal{T}\models E\sqsubseteq D$;
(2) there exists $a\in\mathsf{Ind}(\mathcal{A})$ such that $(a_C,a)\in r^{\mathcal{I}_{\mathcal{T},C}}$ and $\mathcal{A}\models(\mathcal{T},\emptyset,q_D)(a)$.

For the proof of the implication from (1) to (2) take an $\mathcal{EL}$ concept E satisfying (1). Then $a_C\in(\exists r.E)^{\mathcal{I}_{\mathcal{T},C}}$, by Lemma 5.11. Then there exists $a\in\mathsf{Ind}(\mathcal{A})$ such that $(a_C,a)\in r^{\mathcal{I}_{\mathcal{T},C}}$ and $a\in E^{\mathcal{I}_{\mathcal{T},C}}$. Hence $a\in E^{\mathcal{I}_\mathcal{A}}$ as $\mathsf{sig}(E)\subseteq\Sigma_\mathsf{C}$. There is an unfolding C_a^m of $\mathcal{A}$ at a of depth m such that $\emptyset\models C_a^m\sqsubseteq E$. From $\mathcal{T}\models E\sqsubseteq D$ we obtain $\mathcal{T}\models C_a^m\sqsubseteq D$. But then, by Lemma 5.15, $\mathcal{A}\models(\mathcal{T},\emptyset,q_D)(a)$, as required.

Conversely, let $a\in\mathsf{Ind}(\mathcal{A})$ such that $(a_C,a)\in r^{\mathcal{I}_{\mathcal{T},C}}$ and $\mathcal{A}\models(\mathcal{T},\emptyset,q_D)(a)$. By Lemma 5.15, there exists an unfolding C_a^m of $\mathcal{A}$ at a such that $\mathcal{T}\models C_a^m\sqsubseteq D$. It is readily checked that $E=C_d^m$ is as required for Condition (1).

Condition (2) can be checked in PTime since query evaluaton for OMQCs in $(\mathcal{EL},\emptyset,\text{dtCQ})$ is in PTime (in combined complexity). □

Theorem 6.6. *It is in* PTime *to decide whether an $\mathcal{EL}$ TBox with closed predicates is safe.*

Proof. Assume $(\mathcal{T},\Sigma_\mathsf{C})$ is given. By Lemma 6.2, we can construct, in polynomial time, a reduced $\mathcal{EL}$ TBox $\mathcal{T}'$ equivalent to $\mathcal{T}$. By Lemma 6.5, we can check in PTime whether $(\mathcal{T}',\Sigma_\mathsf{C})$ is strongly safe. By Lemma 6.4, strong safeness is equivalent to safeness for $(\mathcal{T}',\Sigma_\mathsf{C})$. □

7. Closing Concept Names in the Fixed Query Case and Surjective CSPs

We now switch from the quantified query case to the fixed query case. In this section, we consider OMQC languages that only admit closing concept names while the case of closing role names is deferred to the subsequent section. Regarding the former, our main aim is to establish a close connection between UCQ evaluation for such OMQC languages and generalized surjective constraint satisfaction problems (CSPs). Let BUtCQ denote the class of Boolean queries that can be obtained from a union of tCQs by existentially quantifying the answer variable and let BAQ denote the class of *Boolean atomic queries* which take the form $\exists x A(x)$, A a concept name. We consider OMQC languages between $(\text{DL-Lite}_\text{core},\mathsf{N_C},\text{BUtCQ})$ and $(\mathcal{ALCHI},\mathsf{N_C},\text{BUtCQ})$ as well as between $(\mathcal{EL},\mathsf{N_C},\text{BUtCQ})$ and $(\mathcal{ALCHI},\mathsf{N_C},\text{BUtCQ})$ and show that for all these, a PTime/coNP dichotomy is equivalent to a PTime/NP dichotomy for generalized surjective CSPs, a problem that is wide open. In fact, understanding the complexity of surjective CSPs, generalized or not, is a very difficult, ongoing research effort. As pointed out in the introduction, there are even concrete surjective CSPs with very few elements whose complexity is unknown and, via the connection established in this section, these problems can be used to derive concrete OMQCs from the mentioned languages whose computational properties are currently not understood.

We next introduce CSPs and then give a more detailed overview of the results obtained in this section. An interpretation $\mathcal{I}$ is a Σ-*interpretation* if it only interprets predicates in Σ, that is, all other predicates are interpreted as empty. For every finite Σ-interpretation $\mathcal{I}$ we denote by $\mathsf{CSP}(\mathcal{I})$ the following *constraint satisfaction problem (in signature* Σ*)*: given a finite Σ-interpretation $\mathcal{J}$, decide whether there is a homomorphism h from $\mathcal{J}$ to $\mathcal{I}$. The *surjective constraint satisfaction problem*, $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$, is the variant of $\mathsf{CSP}(\mathcal{I})$ where we require h to be surjective. $\mathcal{I}$ is then called the *template* of $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$. In this article we only consider CSPs with predicates of arity at most two. 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 $\mathcal{J}$, whether there is a surjective homomorphism from $\mathcal{J}$ to some interpretation in Γ. The interpretations $\mathcal{I}$ in Γ are called the *templates* of $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$.

We first show that for every constraint satisfaction problem $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$, there is an OMQC Q from (DL-Lite$_{\text{core}}$, $\mathsf{N_C}$, BUtCQ) such that the evaluation problem for Q has the same complexity as the complement of $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$, up to polynomial time reductions; we then observe that the same holds for $(\mathcal{EL}, \mathsf{N_C}, \text{BAQ})$. To achieve a cleaner presentation, we first present the construction for non-generalized surjective CSPs and then sketch the modifications required to lift it it to generalized surjective CSPs. Consider $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ in signature Σ. Let A, V, and V_d, $d \in \Delta^{\mathcal{I}}$, be concept names not in Σ, and val, aux_d, and force_d, $d \in \Delta^{\mathcal{I}}$, be role names not in Σ. Define the OMQC $Q_{\mathcal{I}} = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ from (DL-Lite$_{\text{core}}$, $\mathsf{N_C}$, BUtCQ) as follows:

$$
\begin{aligned}
\mathcal{T} &= \{A \sqsubseteq \exists \mathsf{val},\ \exists \mathsf{val}^- \sqsubseteq V\} \cup \\
&\quad\ \{A \sqsubseteq \exists \mathsf{aux}_d,\ \exists \mathsf{aux}_d^- \sqsubseteq V \sqcap V_d \mid d \in \Delta^{\mathcal{I}}\} \cup \\
&\quad\ \{A \sqsubseteq \exists \mathsf{force}_d,\ \exists \mathsf{force}_d^- \sqsubseteq A \mid d \in \Delta^{\mathcal{I}}\} \\
\Sigma_{\mathsf{C}} &= \{A, V\} \cup \{V_d \mid d \in \Delta^{\mathcal{I}}\} \\
\Sigma_{\mathsf{A}} &= \Sigma \cup \Sigma_{\mathsf{C}} \\
q &= q_1 \vee q_2 \vee q_3 \vee q_4
\end{aligned}
$$

where

$$
\begin{aligned}
q_1 &= \bigvee_{d,d' \in \Delta^{\mathcal{I}} \mid d \neq d'} \exists x \exists y_1 \exists y_2\, A(x) \wedge \mathsf{val}(x, y_1) \wedge \\
&\qquad\qquad \mathsf{val}(x, y_2) \wedge V_d(y_1) \wedge V_{d'}(y_2) \\
q_2 &= \bigvee_{d \in \Delta^{\mathcal{I}}, E \in \Sigma \mid d \notin E^{\mathcal{I}}} \exists x \exists y\, A(x) \wedge E(x) \wedge \\
&\qquad\qquad \mathsf{val}(x, y) \wedge V_d(y) \\
q_3 &= \bigvee_{d,d' \in \Delta^{\mathcal{I}}, r \in \Sigma \mid (d,d') \notin r^{\mathcal{I}}} \exists x \exists y \exists x_1 \exists y_1\, A(x) \wedge A(y) \wedge r(x, y) \wedge \\
&\qquad\qquad \mathsf{val}(x, x_1) \wedge \mathsf{val}(y, y_1) \wedge \\
&\qquad\qquad V_d(x_1) \wedge V_{d'}(y_1) \\
q_4 &= \bigvee_{d,d' \in \Delta^{\mathcal{I}} \mid d \neq d'} \exists x \exists y \exists z\, A(x) \wedge \mathsf{force}_d(z, x) \wedge \\
&\qquad\qquad \mathsf{val}(x, y) \wedge V_{d'}(y).
\end{aligned}
$$

The following lemma links $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ to the constructed OMQC $Q_{\mathcal{I}}$.

Lemma 7.1. *The complement of* $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ *and the evaluation problem for* $Q_{\mathcal{I}}$ *are polynomially reducible to each other.*

Proof. Assume that $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ is given. For the polynomial reduction of $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ to the evaluation problem for $Q_{\mathcal{I}}$, 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 corresponding to $\mathcal{J}$. Introduce, for every $d \in \Delta^{\mathcal{I}}$, a fresh individual name a_d and let the ABox $\mathcal{A}$ be defined as

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

Obviously, $\mathcal{A}$ can be constructed in polynomial time. We claim that $\mathcal{J} \in \mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ iff $\mathcal{A} \not\models Q_{\mathcal{I}}$.

($\Rightarrow$) Suppose that there is a surjective homomorphism h from $\mathcal{J}$ to $\mathcal{I}$. Define the interpretation $\mathcal{I}'$ as follows:

$$\begin{array}{rcl}
\Delta^{\mathcal{I}'} & = & \mathsf{Ind}(\mathcal{A}) \\
A^{\mathcal{I}'} & = & \mathsf{Ind}(\mathcal{A}_{\mathcal{J}}) \\
V^{\mathcal{I}'} & = & \Delta^{\mathcal{I}} \\
V_d^{\mathcal{I}'} & = & \{a_d\}, \text{ for all } d \in \Delta^{\mathcal{I}} \\
\mathsf{val}^{\mathcal{I}'} & = & \{(a, a_{h(a)}) \mid a \in \mathsf{Ind}(\mathcal{A}_{\mathcal{J}})\} \\
\mathsf{aux}_d^{\mathcal{I}'} & = & \{(a, a_d) \mid a \in \mathsf{Ind}(\mathcal{A}_{\mathcal{J}})\}, \text{ for all } d \in \Delta^{\mathcal{I}} \\
\mathsf{force}_d^{\mathcal{I}'} & = & \{(a, a') \in \mathsf{Ind}(\mathcal{A}_{\mathcal{J}}) \times \mathsf{Ind}(\mathcal{A}_{\mathcal{J}}) \mid h(a') = d\}, \text{ for all } d \in \Delta^{\mathcal{I}} \\
P^{\mathcal{I}'} & = & P^{\mathcal{J}}, \text{ for all predicates } P \notin (\{A, V, \mathsf{val}\} \cup \{V_d, \mathsf{aux}_d, \mathsf{force}_d \mid d \in \Delta^{\mathcal{I}}\})
\end{array}$$

One can now verify that $\mathcal{I}'$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C}, and that $\mathcal{I}' \not\models q$. Thus, $\mathcal{A} \not\models Q_{\mathcal{I}}$, as required.

($\Leftarrow$) Suppose $\mathcal{A} \not\models Q_{\mathcal{I}}$. 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 $h(a_e) = f$.

We show that h is a homomorphism. We show, using q_3, that h preserves role names. Using q_2, one can show in the same way that h preserves concept names. 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_3:

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

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 $\mathcal{I}' \models q_3$, which contradicts 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_4:

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

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, $\mathcal{I}' \models q_4$ which contradicts our choice of $\mathcal{I}'$. This finishes the proof of the reduction from $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ to evaluating $Q_{\mathcal{I}}$.

We now give the polynomial reduction of the evaluation problem for $Q_{\mathcal{I}}$ to $\mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$. Assume a Σ_{A}-ABox $\mathcal{A}$ is given. To decide whether $\mathcal{A} \models Q_{\mathcal{I}}$, we start with the following:

(1) If $\mathcal{A}$ does not contain any assertion of the form $A(a)$, then $\mathcal{A} \not\models Q_{\mathcal{I}}$. In fact, let $\mathcal{I}_{\mathcal{A}}$ be $\mathcal{A}$ viewed as an interpretation. 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 '$\mathcal{A} \not\models Q_{\mathcal{I}}$'.
(2) Otherwise, if $\mathcal{A}$ does not contain for each $d \in \Delta^{\mathcal{I}}$ an individual name a with $V(a), V_d(a) \in \mathcal{A}$, then $\mathcal{A}$ is not consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$. Thus answer '$\mathcal{A} \models Q_{\mathcal{I}}$'.
(3) Otherwise, if $\mathcal{A}$ contains an individual name a with $V(a) \in \mathcal{A}$ and $V_d(a) \notin \mathcal{A}$ for each $d \in \Delta^{\mathcal{I}}$, then $\mathcal{A} \not\models Q_{\mathcal{I}}$. In fact, we can build a model of $\mathcal{A}$ and $\mathcal{T}$ that makes q false in the following way: 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 trivially be satisfied. All remaining choices can be taken in an arbitrary way.

If none of the above applies, let $\mathcal{A}_{|A}$ be the restriction of $\mathcal{A}$ to $\{a \in \mathsf{Ind}(\mathcal{A}) \mid A(a) \in \mathcal{A}\}$. Since Case (1) above does not apply, $\mathcal{A}_{|A}$ is non-empty. Let $\mathcal{J}_A$ be the Σ-reduct of the interpretation corresponding to $\mathcal{A}_{|A}$. We show that $\mathcal{J}_A \in \mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$ iff $\mathcal{A} \not\models Q_{\mathcal{I}}$.

($\Leftarrow$). Assume that $\mathcal{A} \not\models Q_{\mathcal{I}}$. 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 q_1, for each $a \in \mathsf{Ind}(\mathcal{A}_{|A})$ there is exactly one $d \in \Delta^{\mathcal{I}}$ such that $a \in (\exists \mathsf{val}.V_d)^{\mathcal{J}}$. Define a homomorphism $h : \mathcal{J}_A \to \mathcal{I}$ by mapping each a in $\mathcal{A}_{|A}$ to the value $d \in \Delta^{\mathcal{I}}$ thus determined. By q_2 and q_3, h is indeed a homomorphism. By the third line of $\mathcal{T}$ and q_4 and since A is closed, h must be surjective.

($\Rightarrow$). Assume that $\mathcal{J}_A \in \mathsf{CSP}(\mathcal{I})^{\mathsf{sur}}$, and let h be a surjective homomorphism from $\mathcal{J}_A$ 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 individual name a_d of $\mathcal{A}$ such that $V(a_d)$ and $V_d(a_d)$ are in $\mathcal{A}$. For each individual name a in $\mathcal{A}_{|A}$, 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 individual name a'_d of $\mathcal{A}_{|A}$ with $h(a'_d) = d$. Further extend $\mathcal{J}$ by adding (a, a'_d) to $\mathsf{force}_d^{\mathcal{J}}$ for all $a \in \mathsf{Ind}(\mathcal{A}_{|A})$ 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 $\mathcal{J} \not\models q$. Thus, $\mathcal{A} \not\models Q_{\mathcal{I}}$, as required. □

Note that the same reduction works when DL-Lite$_{\text{core}}$ is replaced with $\mathcal{EL}$. One simply has to replace the TBox $\mathcal{T}$ by the $\mathcal{EL}$ TBox

$$\begin{aligned}\mathcal{T}' \quad = \quad & \{A \sqsubseteq \exists \mathsf{val}.V\} \cup \\ & \{A \sqsubseteq \exists \mathsf{aux}_d.(V \sqcap V_d) \mid d \in \Delta^{\mathcal{I}}\} \cup \\ & \{A \sqsubseteq \exists \mathsf{force}_d.A \mid d \in \Delta^{\mathcal{I}}\}\end{aligned}$$

and observe that all CQs in q have the form $\exists x q'(x)$ with $q'(x)$ a dtCQ which enables the following modification: introduce a fresh concept name B, then for each CQ $\exists x q'(x)$ in q, take the $\mathcal{EL}$ concepts $C_{q'}$ that corresponds to $q'(x)$ and extend $\mathcal{T}'$ with $C_{q'} \sqsubseteq B$, and finally replace q with the BAQ $\exists x\, B(x)$.

We now describe how to extend the reduction from surjective CSPs to generalized surjective CSPs. Let $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$ be such a CSP. Let $\Gamma = \{\mathcal{I}_1, \ldots, \mathcal{I}_n\}$. The main idea is to use n copies of each non-Σ symbol in the above reduction, one for each template in Γ. Let the i-th copy of A be A_i, of val be val_i, and so on. This gives us n copies of the TBox $\mathcal{T}$ and the UCQ q in the above reduction, which we call $\mathcal{T}_1, \ldots, \mathcal{T}_n$ and $q_1, \ldots, q_n$. Note that the $\mathcal{T}_i$ do not share any symbols and that the q_i share only the symbols from Σ. We define $Q_\Gamma = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ where $\mathcal{T} = \mathcal{T}_1 \cup \cdots \cup \mathcal{T}_n$, q is the BUtCQ obtained from $q_1 \wedge \cdots \wedge q_n$ by pulling disjunction outside, and Σ_{A} and Σ_{C} are defined as expected. It is then possible to prove an analogue of Lemma 7.1, we only sketch the required modifications. In the reduction of $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$ to the evaluation problem for Q_Γ, one builds on ABox $\mathcal{A}$ for each $\mathcal{I} \in \Gamma$, each as in the corresponding of the proof of Lemma 7.1, and then takes their union. In the reduction of the evaluation problem for Q_Γ to $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$, one first checks whether for some i, the given Σ_{A}-ABox $\mathcal{A}$ contains an assertion $A_i(a)$, but no assertion $V_i(a)$ and answers '$\mathcal{A} \models Q_\Gamma$' if this is the case (this corresponds to Point (2) in the original proof). One then checks whether for some i there is no assertion of the form $A_i(a)$ and answers '$\mathcal{A} \not\models Q_\Gamma$' if this is the case (corresponding to Point (1) in the original proof). Point (3) and the remainder of the reduction need no major adaptations.

In summary, we have obtained the following result.

Theorem 7.2. *For every $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$, there is an OMQC Q_Γ in ($DL\text{-}Lite_{core}, \mathsf{N_C}, BUtCQ$) such that the complement of $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$ has the same complexity as the evaluation problem for Q_Γ, up to polynomial time reductions. The same holds for $(\mathcal{EL}, \mathsf{N_C}, BAQ)$.*

We note that, as can easily be verified by checking the constructions in the proof of Lemma 7.1, the complement of $\mathsf{CSP}(\Gamma)^{\mathsf{sur}}$ and the evaluation problem for Q_Γ actually have the same complexity up to *FO reductions* [35]. 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 a rather general converse of Theorem 7.2 by showing that for every OMQC Q from $(\mathcal{ALCHI}, \mathsf{N_C}, \mathrm{BUtCQ})$, there is a generalized surjective CSP that has the same complexity as the complement of the evaluation problem for Q, up to polynomial time reductions.

Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ be an OMQC from $(\mathcal{ALCHI}, \mathsf{N_C}, \mathrm{BUtCQ})$. We can assume w.l.o.g. that q is a BAQ, essentially because every tCQ can be rewritten into an $\mathcal{ALCI}$ concept; see the remark on $\mathcal{EL}$ and BAQs made after the proof of Lemma 7.1. Thus, let $q = \exists x\, A_0(x)$ with A_0 a concept name in $\mathcal{T}$. We use the notation for types introduced in Section 5. A subset T of the set $\mathsf{TP}(\mathcal{T})$ of $\mathcal{T}$-types is *realizable in a countermodel of* Q 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 $T = \{\mathsf{tp}_{\mathcal{I}}(a) \mid a \in \mathsf{Ind}(\mathcal{A})\}$. The desired surjective generalized CSP is defined by taking one template for each $T \subseteq \mathsf{TP}(\mathcal{T})$ that is realizable in a countermodel of Q. The signature Σ of the CSP comprises the predicates in Σ_{A} and one concept name $\overline{A}$ for each concept name in Σ_{C}. We assume w.l.o.g. that there is at least one concept name in Σ_{C} and at least one concept name $A_{\mathsf{open}} \in \Sigma_{\mathsf{A}} \setminus \Sigma_{\mathsf{C}}$.

Pick for every $A \in \Sigma_{\mathsf{C}}$ an element d_A. Then for each $T \subseteq \mathsf{TP}(\mathcal{T})$ realizable in a countermodel of Q we define the template $\mathcal{I}_T$ as follows:

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

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_Q be the set of all interpretations $\mathcal{I}_T$ obtained in the described way.

Lemma 7.3. *Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ be an OMQC from $(\mathcal{ALCHI}, \mathsf{N_C}, BUtCQ)$. Then the evaluation problem for Q reduces in polynomial time to the complement of $\mathsf{CSP}(\Gamma_Q)^{\mathsf{sur}}$.*

Proof. Let $\mathcal{A}$ be a Σ_{A}-ABox that is an input for Q and let $\mathcal{A}'$ be its extension with

(1) all assertions $\overline{A}(a)$ such that $A \in \Sigma_{\mathsf{C}}$, $a \in \mathsf{Ind}(\mathcal{A})$, and $A(a) \notin \mathcal{A}$;
(2) assertions $A_{\mathsf{open}}(a_B)$, where a_B is a fresh individual name for each $B \in \Sigma_{\mathsf{C}}$.

We claim that $\mathcal{A} \not\models Q$ iff there is an interpretation $\mathcal{I}_T \in \Gamma_Q$ such that there exists a surjective homomorphism from $\mathcal{I}_{\mathcal{A}'}$ to $\mathcal{I}_T$. The assertions of type (2) are needed to obtain a homomorphism that is surjective in the '$\Rightarrow$' direction, despite the presence of the elements d_B in the templates in Γ_Q.

($\Leftarrow$). Let $\mathcal{I}_T \in \Gamma_Q$ and let h be a surjective homomorphism from $\mathcal{I}_{\mathcal{A}'}$ to $\mathcal{I}_T$. Note that each element a of $\mathsf{Ind}(\mathcal{A})$ 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 $A \in \Sigma_{\mathsf{C}}$ (which is non-empty). Since $\mathcal{I}_T \in \Gamma_Q$, there are a Σ_{A}-ABox $\mathcal{B}$ and a 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:

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

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 $A \in \Sigma_{\mathsf{C}}$ 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 RIs in $\mathcal{T}$. Satisfaction of the CIs in $\mathcal{T}$ and $\mathcal{J} \not\models q$ follow from the subsequent claim together with the condition that no type in T contains A_0 and each type in $\mathcal{I}_T$ is satisfied in a model of $\mathcal{T}$.

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

Proof of claim. 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 $b \in D^{\mathcal{J}}$.

$\dashv$

($\Rightarrow$). Assume that $\mathcal{A} \not\models Q$. 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 $\mathcal{I}_T \in \Gamma_Q$ 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 $h(a_B) = d_B$ (recall that such a_B have been added to $\mathsf{Ind}(\mathcal{A})$ for every $B \in \Sigma_{\mathsf{C}}$). It is readily checked that h is a surjective homomorphism from $\mathcal{I}_{\mathcal{A}'}$ 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 $h(a) = \mathsf{tp}_{\mathcal{I}}(a) \in \overline{A}^{\mathcal{I}_T}$ by definition of $\mathcal{I}_T$. $\square$

Lemma 7.4. *Let $Q = (\mathcal{T}, \Sigma_{\mathsf{A}}, \Sigma_{\mathsf{C}}, q)$ be an OMQC from $(\mathcal{ALCHI}, \mathsf{N_C}, BUtCQ)$. Then $\mathsf{CSP}(\Gamma_Q)^{\mathsf{sur}}$ reduces in polynomial time to the complement of the evaluation problem for Q.*

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

(1) if there is a non-special element a of $\mathsf{Ind}(\mathcal{A}')$ such that $A(a) \in \mathcal{A}'$ and $\overline{A}(a) \in \mathcal{A}'$ for some $A \in \Sigma_{\mathsf{C}}$, then return 'no' (there is no template in Γ_Q 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 $(a_A)_{A \in \Sigma_{\mathsf{C}}}$, 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_Q).

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 $\mathcal{A} \not\models Q$ iff there exists an $\mathcal{I}_T \in \Gamma_Q$ such that there is a surjective homomorphism from $\mathcal{J}$ to $\mathcal{I}_T$.

($\Leftarrow$). Let $\mathcal{I}_T \in \Gamma_Q$ and let h be a surjective homomorphism from $\mathcal{J}$ to $\mathcal{I}_T$. Note that each element a of $\mathsf{Ind}(\mathcal{A})$ 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 $A \in \Sigma_\mathsf{C}$ (which is non-empty). Since $\mathcal{I}_T \in \Gamma_Q$, 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 7.3 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 $\mathcal{A} \not\models Q$. 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 $\mathcal{I}_T \in \Gamma_Q$ 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 $A \in \Sigma_\mathsf{C}$ such that $A(a) \notin \mathcal{A}'$ and $\overline{A}(a) \notin \mathcal{A}'$, 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 $\mathcal{J}$ 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 $\mathcal{I}_\mathcal{T}$. □

We have thus established the following result.

Theorem 7.5. *For every OMQC Q from $(\mathcal{ALCHI}, \mathsf{N_C}, BUtCQ)$, there is a generalized $\mathsf{CSP}(\Gamma_Q)^{\mathsf{sur}}$ such that the evaluation problem for Q has the same complexity as the complement of $\mathsf{CSP}(\Gamma_Q)^{\mathsf{sur}}$, up to polynomial time reductions.*

Again, the theorem can easily be strengthened to state the same complexity up to FO reductions. Note that the DL $\mathcal{ALCHI}$ used in Theorem 7.5 is a significant extension of the DLs referred to in Theorem 7.2 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 $\mathcal{ALCHI}$.

8. Closing Role Names in the Fixed Query Case: Turing Machine Equivalence

We generalize the setup from the previous section by allowing also role names to be closed. Our main results are that for every non-determinstic polynomial time Turing machine M, there is an OMQC Q in (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BUtCQ) such that evaluating Q and the complement of M's word problem are polynomial time reducible to each other, and that it is undecidable whether evaluating OMQCs in (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BUtCQ) is in PTime (unless PTime = NP). By Ladner's theorem, it follows that there are coNP-intermediate OMQCs (unless PTime = NP) and that a full complexity classification of the OMQCs in this language is beyond reach of the techniques available today. As in the previous section, the same results hold for ($\mathcal{EL}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BAQ).

To establish these results, we utilize two related results from [47, 14]: (1) for every NP Turing machine M, there is an ontology-mediated query Q from ($\mathcal{ALCF}$, $\emptyset$, BAQ) such that evaluating Q is reducible in polynomial time to the complement of M's word problem and vice versa, where $\mathcal{ALCF}$ is the extension of $\mathcal{ALC}$ with functional roles; and (2) it is undecidable whether an OMQC from ($\mathcal{ALCF}$, $\emptyset$, BAQ) is in PTime. For using these results in our context, however, it is more convenient to phrase them in terms of (a certain kind of) monadic disjunctive datalog programs with inequality rather than in terms of OMQCs from

$(\mathcal{ALCF}, \emptyset, \text{BAQ})$. This is what we do in the following, starting with the introduction of a suitable version of monadic disjunctive datalog. For a more thorough introduction, see [25].

A *monadic disjunctive datalog rule (MDD rule)* ρ takes the form

$$P_1(x) \vee \cdots \vee P_m(x) \leftarrow R_1(\vec{x}_1) \wedge \cdots \wedge R_n(\vec{x}_n) \quad \text{ or } \quad \mathsf{goal} \leftarrow R_1(\vec{x}_1) \wedge \cdots \wedge R_n(\vec{x}_n)$$

with $m, n > 0$ and where all P_i are unary predicates, goal is the *goal predicate* of arity 0, and all R_i are predicates of arity one or two, including possibly the non-equality predicate $\neq$. We refer to $P_1(x) \vee \cdots \vee P_m(x)$ and, respectively, goal as the *head* of ρ, and to $R_1(\vec{x}_1) \wedge \cdots \wedge R_n(\vec{x}_n)$ as the *body*. A *monadic disjunctive datalog (MDD) program* Π is a finite set of MDD rules containing at least one rule with the goal predicate in its head and no rule with the goal predicate in its body. Predicates that occur in the head of at least one rule of Π are *intensional (IDB) predicates*, denoted $\mathsf{IDB}(\Pi)$, and all remaining predicates in Π are *extensional (EDB) predicates*, denoted $\mathsf{EDB}(\Pi)$. An interpretation $\mathcal{I}$ is a *model* of Π if it satisfies all rules in Π (viewed as universally quantified first-order sentences). Π is *entailed on* a $\mathsf{EDB}(\Pi)$-ABox $\mathcal{A}$, in symbols $\mathcal{A} \models \Pi$, iff goal is true in every model of Π and $\mathcal{A}$. Note that it suffices to consider models that respect closed predicates $\mathsf{EDB}(\Pi)$. The *evaluation problem for* Π is the problem to decide whether Π is entailed by an $\mathsf{EDB}(\Pi)$-ABox $\mathcal{A}$.

For our reduction, we use the following kind of MDD programs that we call basic. A binary predicate r is *functional in an ABox* $\mathcal{A}$ if $r(a, b_1), r(a, b_2) \in \mathcal{A}$ implies $b_1 = b_2$ and r is *empty in* $\mathcal{A}$ if r does not occur in $\mathcal{A}$. Then an MDD program Π is *basic* if

- Π uses exactly two binary predicates, r_1, r_2, and contains exactly the following functionality rules, for $i = 1, 2$:
$$\mathsf{goal} \leftarrow r_i(x, y) \wedge r_i(x, z) \wedge (y \neq z)$$
- all remaining rules of Π are of the form
$$P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \quad \text{ or } \quad \mathsf{goal} \leftarrow q$$
where $n \geq 1$ and q is a dtCQ with root x (with the quantifier prefix removed).
- if r_1, r_2 are functional and at least one r_i is empty in an $\mathsf{EDB}(\Pi)$-ABox $\mathcal{A}$, then $\mathcal{A} \not\models \Pi$.

The following result can be obtained by starting from the results for $(\mathcal{ALCF}, \emptyset, \text{BAQ})$ from [47, 14] mentioned above and translating the involved OMQs into a basic MDD program. Such a translation is given in [14] for the case of $\mathcal{ALC}$ TBoxes and MDD programs without inequality, but the extension to functional roles and inequality is trivial.

Theorem 8.1.

(1) *For every non-deterministic polynomial time Turing machine M, there exists a basic MDD program Π such that the evaluation problem for Π and the complement of M's word problem are polynomial time reducible to each other.*

(2) *It is undecidable whether the evaluation problem for a basic MDD program is in* PTime *(unless* PTime =NP*).*

We next prove the following central theorem.

Theorem 8.2. *For every basic MDD program Π, one can construct an OMQC Q in the language $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, BAQ)$ such that the evaluation problem for Q and Π are polynomial time reducible to each other. The same is true for (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BUtCQ).*

Proof. Assume a basic MDD program Π of the form defined above is given. We first construct an OMQC $Q_\Pi = (\mathcal{T}_\Pi, \Sigma_\Pi, \Sigma_\Pi, q_\Pi)$ in $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BUtCQ})$ and then obtain the required OMQCs in $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BAQ})$ and (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BUtCQ) by rather straightforward modifications of Q_Π. Note that we construct a Q_Π in which the ABox signature and set of closed predicates coincide. We set $\Sigma_\Pi = \mathsf{EDB}(\Pi) \cup \{T, F, V\}$, where T, F, V are fresh concept names. We also use auxiliary predicates which are not in the ABox signature of Q_Π: role names val_P for every unary $P \in \mathsf{IDB}(\Pi)$ and role names s_i and concept names A_i, B_i, for $i = 1, 2$. $\mathcal{T}_\Pi$ contains the following CIs:

$$\begin{array}{lcl} T & \sqsubseteq & V \\ F & \sqsubseteq & V \\ \top & \sqsubseteq & \exists \mathsf{val}_P.V, \text{ for all unary } P \in \mathsf{IDB}(\Pi) \\ \top & \sqsubseteq & \exists s_i.(\exists r_i.A_i \sqcap \exists r_i.B_i), \text{ for } i = 1, 2\ . \end{array}$$

Using $\mathcal{T}_\Pi$, we encode the truth value of IDB predicates P using the CQs

$$P^T(x, y) := (\mathsf{val}_P(x, y) \wedge T(y)), \quad P^F(x, y) := (\mathsf{val}_P(x, y) \wedge F(y)).$$

For any tCQ q, we denote by q^T the result of replacing every occurrence of an IDB $P(x)$ in q by $P^T(x, y_0)$, where the variable y_0 is fresh for every occurrence of $P(x)$, and existentially quantified. Thus, q^T is again a tCQ (and a dtCQ if q is already a dtCQ). The final CI is used to encode functionality of the roles r_1, r_2. We define CQs $q^1_\mathcal{F}$ and $q^2_\mathcal{F}$ by setting

$$q^i_\mathcal{F} = (s_i(x, y) \wedge r_i(y, z) \wedge A_i(z) \wedge B_i(z)),$$

for $i = 1, 2$. Then, for the OMQC $Q_i = (\{\top \sqsubseteq \exists s_i.(\exists r_i.A_i \sqcap \exists r_i.B_i)\}, \Sigma_\Pi, \Sigma_\Pi, \exists y\, \exists z\, q^i_\mathcal{F})$ and any Σ_Π-ABox $\mathcal{A}$:

- if r_i is empty in $\mathcal{A}$, then $\mathcal{A}$ is not consistent w.r.t. $(\{\top \sqsubseteq \exists s_i.(\exists r_i.A_i \sqcap \exists r_i.B_i)\}, \Sigma_\Pi)$, and
- if r_i is not empty in $\mathcal{A}$, then r_i is functional in $\mathcal{A}$ iff $\mathcal{A} \models Q_i(a)$, for some (equivalently, all) $a \in \mathsf{Ind}(\mathcal{A})$.

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

- $q^1_\mathcal{F} \wedge q^1_\mathcal{F} \wedge q^T$, for every rule $\mathsf{goal} \leftarrow q \in \Pi$, where we assume that the only variable shared by any two of the conjuncts $q^1_\mathcal{F}$, $q^2_\mathcal{F}$ and q^T is x.
- $q^1_\mathcal{F} \wedge q^2_\mathcal{F} \wedge q^T \wedge \bigwedge_{1 \leq i \leq n} P^F_i$, for every $P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \in \Pi$, where we assume again that the only variable shared by any two of the conjuncts $q^1_\mathcal{F}$, $q^2_\mathcal{F}$, q^T, P^F_i, $1 \leq i \leq n$, is x.

We prove the following

Claim. The problem of evaluating Π and the problem of evaluating Q_Π are polynomial time reducible to each other.

Proof of claim. ($\Rightarrow$) Assume an $\mathsf{EDB}(\Pi)$-ABox $\mathcal{A}$ is given as an input to Π. If r_1 or r_2 is not functional in $\mathcal{A}$, then output '$\mathcal{A} \models \Pi$'. Otherwise, if r_1 or r_2 is empty, then output '$\mathcal{A} \not\models \Pi$'. Now assume that r_1 and r_2 are not empty and both are functional in $\mathcal{A}$. Let

$$\mathcal{A}' = \mathcal{A} \cup \{T(a), F(b), V(a), V(b)\},$$

where we asume w.l.o.g. that a, b occur in $\mathsf{Ind}(\mathcal{A})$. We show that $\mathcal{A} \models \Pi$ iff $\mathcal{A}' \models Q_\Pi$.

Assume first that $\mathcal{A} \not\models \Pi$. Let $\mathcal{I}$ be a model of $\mathcal{A}$ and Π that respects closed predicates $\mathsf{EDB}(\Pi)$ and satisfies no body of any rule $\mathsf{goal} \leftarrow q \in \Pi$. Define $\mathcal{I}'$ in the same way as $\mathcal{I}$ except that

- $T^{\mathcal{I}'} = \{a\}$, $F^{\mathcal{I}'} = \{b\}$, and $V^{\mathcal{I}'} = \{a, b\}$;
- $s_i^{\mathcal{I}'} = \Delta^\mathcal{I} \times \mathsf{dom}(r_i^\mathcal{I})$ and $A_i^{\mathcal{I}'} = B_i^{\mathcal{I}'} = \Delta^\mathcal{I}$, for $i = 1, 2$, where $\mathsf{dom}(r^\mathcal{I})$ denotes the domain of $r^\mathcal{I}$;

- $\mathsf{val}_P^{\mathcal{I}'} = (P^{\mathcal{I}} \times \{a\}) \cup ((\Delta^{\mathcal{I}} \setminus P^{\mathcal{I}}) \times \{b\})$, for all unary $P \in \mathsf{IDB}(\Pi)$.

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

(1) No q^T with $\mathsf{goal} \leftarrow q \in \Pi$ is satisfied in $\mathcal{I}'$;
(2) No $q^T \wedge \bigwedge_{1\leq i\leq n} P_i^F$ with $P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \in \Pi$ is satisfied in $\mathcal{I}'$.

Point (1) holds since $P^{\mathcal{I}} = \{d \mid \mathcal{I}' \models \exists y\, P^T(d,y))\}$ for all unary $P \in \mathsf{IDB}(\Pi)$, by definition of $\mathcal{I}'$ and since q is not satisfied in $\mathcal{I}$ for any rule $\mathsf{goal}() \leftarrow q \in \Pi$. Point (2) holds since all rules $P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \in \Pi$ are satisfied in $\mathcal{I}$ and $P^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus \{d \mid \mathcal{I}' \models \exists y\, P^F(d,y)\}$ for all unary $P \in \mathsf{IDB}(\Pi)$.

Assume now that $\mathcal{A}' \not\models Q_\Pi$. 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 modifying $\mathcal{I}$ by setting $P^{\mathcal{I}'} = \{d \mid \mathcal{I} \models \exists y\, P^T(d,y)\}$, for all unary $P \in \mathsf{IDB}(\Pi)$. It follows from the condition that r_1, r_2 are non-empty and functional in $\mathcal{A}'$ that $\mathcal{I} \models \forall x(\exists y \exists z q_{\mathcal{F}}^1 \wedge \exists y \exists z q_{\mathcal{F}}^2)$. From $\mathcal{I} \not\models q_\Pi$ we obtain that no q with $\mathsf{goal}() \leftarrow q \in \Pi$ is satisfied in $\mathcal{I}'$ and that all rules $P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \in \Pi$ are satisfied in $\mathcal{I}'$. Thus, $\mathcal{I}'$ is a model of $\mathcal{A}$ and Π witnessing that $\mathcal{A} \not\models \Pi$.

($\Leftarrow$) 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) V is non-empty in $\mathcal{A}$, (ii) T, F are both contained in V in $\mathcal{A}$, and (iii) r_1, r_2 are non-empty in $\mathcal{A}$. Thus, output '$\mathcal{A} \models Q_\Pi$' whenever (i), (ii), or (iii) is violated. Now assume (i), (ii), and (iii) hold. If r_1 or r_2 are not functional in $\mathcal{A}$, then we can construct a model of $\mathcal{T}_\Pi$ and $\mathcal{A}$ that respects closed predicates Σ_Π and such that $\exists x(\exists y \exists z q_{\mathcal{F}}^1 \wedge \exists y \exists z q_{\mathcal{F}}^2)$ is not satisfied in $\mathcal{I}$. Hence, we output '$\mathcal{A} \not\models Q_\Pi$'. Thus, assume in addition to (i), (ii) and (iii) that r_1 and r_2 are functional in $\mathcal{A}$. We distinguish five cases. We only consider the first case in detail, the remaining cases are proved similarly.

(1) If $F^{\mathcal{I}_\mathcal{A}} \cup T^{\mathcal{I}_\mathcal{A}} \neq V^{\mathcal{I}_\mathcal{A}}$, then output '$\mathcal{A} \models Q_\Pi$' if there exists a rule $\mathsf{goal} \leftarrow q \in \Pi$ such that q contains not IDBs and q (which then equals q^T) is satisfied in $\mathcal{I}_\mathcal{A}$. This is clearly correct since $\mathcal{A} \models (\mathcal{T}_\Pi, \Sigma_\Pi, \Sigma_\Pi, q_{\mathcal{F}}^1 \wedge q_{\mathcal{F}}^2 \wedge q^T)$ follows. Otherwise output '$\mathcal{A} \not\models Q_\Pi$'. To prove correctness, let $a \in V^{\mathcal{I}_\mathcal{A}} \setminus (F^{\mathcal{I}_\mathcal{A}} \cup T^{\mathcal{I}_\mathcal{A}})$. Construct a model $\mathcal{I}$ of $\mathcal{T}_\Pi$ and $\mathcal{A}$ that respects Σ_Π by extending $\mathcal{I}_\mathcal{A}$ by setting $\mathsf{val}_P^{\mathcal{I}} = \Delta^{\mathcal{I}} \times \{a\}$ for all unary IDB predicates P and defining $s_i^{\mathcal{I}}, A_i^{\mathcal{I}}, B_i^{\mathcal{I}}$, $i = 1, 2$, arbitrarily so that $\mathcal{T}_\Pi$ is satisfied. Then no q^T with $\mathsf{goal} \leftarrow q \in \Pi$ and no $\bigwedge_{1\leq i\leq n} P_i^F$ with $P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \in \Pi$ is satisfied in $\mathcal{I}$. Thus $\mathcal{I} \not\models q_\Pi$.
(2) If $T^{\mathcal{I}_\mathcal{A}} = F^{\mathcal{I}_\mathcal{A}} = V^{\mathcal{I}_\mathcal{A}}$, then output '$\mathcal{A} \models Q_\Pi$' if there exists a rule $\mathsf{goal} \leftarrow q \in \Pi$ or $P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \in \Pi$ such that q' is satisfied in $\mathcal{I}_\mathcal{A}$ for the query q' obtained from q by removing every atom $P(y)$ from q with P a unary IDB. Otherwise output '$\mathcal{A} \not\models Q_\Pi$'.
(3) If $T^{\mathcal{I}_\mathcal{A}} = V^{\mathcal{I}_\mathcal{A}}$ and $F^{\mathcal{I}_\mathcal{A}} \neq V^{\mathcal{I}_\mathcal{A}}$, then output '$\mathcal{A} \models Q_\Pi$' if there exists a rule $\mathsf{goal} \leftarrow q \in \Pi$ such that q' is satisfied in $\mathcal{I}_\mathcal{A}$ for the query q' obtained from q by removing every atom $P(y)$ from q with P a unary IDB. Otherwise output '$\mathcal{A} \not\models Q_\Pi$'.
(4) If $F^{\mathcal{I}_\mathcal{A}} = V^{\mathcal{I}_\mathcal{A}}$ and $T^{\mathcal{I}_\mathcal{A}} \neq V^{\mathcal{I}_\mathcal{A}}$, then output '$\mathcal{A} \models Q_\Pi$' if there exists a rule $\mathsf{goal} \leftarrow q \in \Pi$ or $P_1(x) \vee \cdots \vee P_n(x) \leftarrow q \in \Pi$ such that q does not contain any IDB and q is satisfied in $\mathcal{I}_\mathcal{A}$. Otherwise output '$\mathcal{A} \not\models Q_\Pi$'.
(5) If none of the four cases above apply, obtain $\mathcal{A}'$ from $\mathcal{A}$ by removing all assertions using T, F, or V. Then $\mathcal{A}' \models \Pi$ iff $\mathcal{A} \models Q$, and we have established the polynomial time reduction. $\dashv$

The modification of Q_Π needed to obtain an OMQC from $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BAQ})$ is the same as in the proof of Theorem 7.2: the query q_Π is a BUdtCQ and so we can replace it with a query of the form $\exists x\, A(x)$: as the disjuncts of q_Π are of the form $\exists x\, q'(x)$ with $q'(x)$ a dtCQ, we can take the $\mathcal{EL}$ concepts $C_{q'}$ corresponding to $q'(x)$ and extend $\mathcal{T}_\Pi$ with $C_{q'} \sqsubseteq A$ for every such disjunct $\exists x\, q'(x)$ of q.

It remains to show how one can modify Q_Π to obtain an equivalent OMQC Q'_Π from the language $(\text{DL-Lite}_\mathcal{R}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BUtCQ})$. First, to eliminate $\top$ on the left-hand-side of CIs in $\mathcal{T}_\Pi$, we replace each CI $\top \sqsubseteq C$ by the CIs $A \sqsubseteq C$, $\exists r \sqsubseteq C$, and $\exists r^- \sqsubseteq C$ for any concept name $A \in \Sigma_\Pi$ and role name $r \in \Sigma_\Pi$. Second, 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 $\mathcal{T}'_\Pi$ be the resulting TBox. Then $Q'_\Pi = (\mathcal{T}'_\Pi, \Sigma_\Pi, \Sigma_\Pi, q_\Pi)$ is as required. □

From Theorems 8.1 and 8.2, we obtain the main result of this section.

Theorem 8.3.

(1) *For every non-deterministic polynomial time Turing machine M one can construct a OMQC Q in the languages $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, BAQ)$ and (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BUtCQ) such that the evaluation problem for Q and M's word problem are polynomial time reducible to each other.*

(2) *It is undecidable whether the evaluation problem for OMQCs in $(\mathcal{EL}, \mathsf{N_C} \cup \mathsf{N_R}, BAQ)$ and (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BUtCQ) is in* PTIME *(unless* PTIME =NP*).*

Note that Theorem 8.3 does not cover DL-Lite$_{\text{core}}$. In fact, the computational status of the language $(\text{DL-Lite}_{\text{core}}, \mathsf{N_C} \cup \mathsf{N_R}, \text{BUtCQ})$ remains open, and in particular it remains open whether Theorem 8.3 can be strengthened to this case.

9. Quantifier-Free UCQs and FO-Rewritability

The results in the previous sections have shown that intractability comes quickly when predicates are closed. The aim of this section is to identify a useful OMQC language whose UCQs are guaranteed to be FO-rewritable. It turns out that one can obtain such a language by combining DL-Lite$_\mathcal{R}$ with quantifier-free UCQs, that is, unions of quantifier-free CQs; we denote this class of queries with UqfCQ. Our main result is that all OMQCs from the language $(\text{DL-Lite}_\mathcal{R}, \mathsf{N_C} \cup \mathsf{N_R}, \text{UqfCQ})$ are FO-rewritable under the mild restriction that there is no RI which requires an open role to be contained in a closed one. We believe that this class of OMQCs is potentially relevant for practical applications. Note that the query language SPARQL, which is used in many web applications, is closely related to UqfCQs and, in fact, does not admit existential quantification under its standard entailment regimes [30]. We also prove that the restriction on RIs is needed for tractability, by constructing a coNP-hard OMQC in $(\text{DL-Lite}_\mathcal{R}, \mathsf{N_C} \cup \mathsf{N_R}, \text{UqfCQ})$.

Theorem 9.1. *Every OMQC $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ from (DL-Lite$_\mathcal{R}$, $\mathsf{N_C} \cup \mathsf{N_R}$, UqfCQ) such that $\mathcal{T}$ contains no RI of the form $s \sqsubseteq r$ with $\mathsf{sig}(s) \not\subseteq \Sigma_\mathsf{C}$ and $\mathsf{sig}(r) \subseteq \Sigma_\mathsf{C}$ is FO-rewritable.*

We first show that ABox consistency w.r.t. $(\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C})$ is FO-rewritable, for every DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$ not containing any RI of the form $s \sqsubseteq r$ with $\mathsf{sig}(s) \not\subseteq \Sigma_\mathsf{C}$ and $\mathsf{sig}(r) \subseteq \Sigma_\mathsf{C}$. We make use of Theorem 3.2 and assume w.l.o.g. that $\Sigma_\mathsf{C} = \Sigma_\mathsf{A}$. Let $\mathsf{con}(\mathcal{T})$ be the set of all concept names in $\mathcal{T}$, 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 $t \subseteq \mathsf{con}(\mathcal{T})$ such that for all $B_1, B_2 \in \mathsf{con}(\mathcal{T})$:

- 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 $t, r_1, \ldots, r_n$ where $t \in T$, $\exists r_1, \ldots, \exists r_n \in \mathsf{con}(\mathcal{T})$ use no predicates from Σ_C, $\exists r_1 \in t$ and for $i \in \{1, \ldots, n-1\}$, $\mathcal{T} \models \exists r_i^- \sqsubseteq \exists r_{i+1}$ and $r_i^- \neq r_{i+1}$. The path is *Σ_C-participating* if for all $i \in \{1, \ldots, n-1\}$, there is no $B \in \mathsf{con}(\mathcal{T})$ with $\mathsf{sig}(B) \subseteq \Sigma_\mathsf{C}$ 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_C-realizable* if for every Σ_C-participating path $t, r_1, \ldots, r_n$ in T, there is some $u \in T$ such that $\{B \in \mathsf{con}(\mathcal{T}) \mid \mathcal{T} \models \exists r_n^- \sqsubseteq B\} \subseteq u$.

A $\mathcal{T}$-typing T provides partial information about a model $\mathcal{I}$ of $\mathcal{T}$ and a Σ_C-ABox $\mathcal{A}$ by taking T to contain the types that are realized in $\mathcal{I}$ by ABox elements. Σ_C-realizability then ensures that we can build from T a model that respects the closed predicates in Σ_C. To make this more precise, define a *$\mathcal{T}$-decoration of* a Σ_C-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 $f(a)|_{\Sigma_\mathsf{C}} = t^a_\mathcal{A}|_{\Sigma_\mathsf{C}}$ where $t^a_\mathcal{A} = \{B \in \mathsf{con}(\mathcal{T}) \mid a \in B^{\mathcal{I}_\mathcal{A}}\}$ and $S|_{\Sigma_\mathsf{C}}$ denotes the restriction of the set S of concepts to those members that only use predicates from Σ_C. The following lemma is proved in the appendix.

Lemma 9.2. *A Σ_C-ABox $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ iff*

(1) *$\mathcal{A}$ has a $\mathcal{T}$-decoration f whose image is a Σ_C-realizable $\mathcal{T}$-typing and*
(2) *if $s(a,b) \in \mathcal{A}$, $\mathcal{T} \models s \sqsubseteq r$, and $\mathsf{sig}(s \sqsubseteq r) \subseteq \Sigma_\mathsf{C}$, then $r(a,b) \in \mathcal{A}$.*

We now construct the required FOQ. 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

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

and for each $\mathcal{T}$-typing $T = \{t_1, \ldots, t_n\}$, 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_C-realizable typings and set

$$\Psi_{\mathcal{T},\Sigma_\mathsf{C}} = \bigvee_{T \in \mathcal{R}} \psi_T \wedge \bigwedge_{\mathcal{T} \models s \sqsubseteq r, \mathsf{sig}(s \sqsubseteq r) \subseteq \Sigma_\mathsf{C}} \forall x \forall y (\psi_s(x,y) \rightarrow \psi_r(x,y)).$$

Note that the two conjuncts of $\Psi_{\mathcal{T},\Sigma_\mathsf{C}}$ express exactly Points (1) and (2) of Lemma 9.2. We have thus shown the following.

Proposition 9.3. *A Σ_C-ABox $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ iff $\mathcal{I}_\mathcal{A} \models \Psi_{\mathcal{T},\Sigma_\mathsf{C}}$.*

The next step is to construct an FO-rewriting of $Q = (\mathcal{T}, \Sigma_\mathsf{C}, \Sigma_\mathsf{C}, q)$ over Σ_C-ABoxes that are consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$. Whereas the FO-rewriting $\Psi_{\mathcal{T},\Sigma_\mathsf{C}}$ above is Boolean and identifies ABoxes that have a common model with $\mathcal{T}$ respecting closed predicates Σ_C, we now aim to construct a FOQ $\Phi_Q(\vec{x})$ such that for all Σ_C-ABoxes $\mathcal{A}$ consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ and $\vec{a} \in \mathsf{Ind}(\mathcal{A})$, we have $\mathcal{I}_\mathcal{A} \models \Phi_Q(\vec{a})$ iff $\mathcal{A} \models Q(\vec{a})$. The desired FO-rewriting of Q is then constructed as $\neg \Psi_{\mathcal{T},\Sigma_\mathsf{C}} \vee \Phi_Q(\vec{x})$. The construction of $\Phi_Q(\vec{x})$ is based on an extended notion of $\mathcal{T}$-typing called *$(\mathcal{T}, q)$-typing* that provides partial information about a model $\mathcal{I}$ of $\mathcal{T}$ and a Σ_C-ABox $\mathcal{A}$ respecting Σ_C which avoids an assignment from $\vec{x}$ to certain individual names $\vec{a}$.

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 $s(x_i, x_j)$, $s \in \Sigma_\mathsf{C}$, 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 an assignment π for q in an ABox $\mathcal{A}$, $f_0(x_i)$ describes the $\mathcal{T}$-type of the ABox individual name $\pi(x_i)$, Γ describes the $\mathcal{T}$-types of ABox individual names that are not in the range of π, and Δ fixes role relationships that do *not* hold between the $\pi(x_i)$. Let X be a set of atoms. Then T *avoids* X if the following conditions hold:

1. for all x_i, $1 \leq i \leq n$, if $A \in f_0(x_i)$, then $A(x_i) \notin X$;
2. for all x_i, $1 \leq i \leq n$, if $\exists s \in f_0(x_i)$, then for $S = \{B \in \mathsf{con}(\mathcal{T}) \mid \mathcal{T} \models \exists s^- \sqsubseteq B\}$ the following holds: (i) S contains no predicate from Σ_C 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 $x' \sim x_i$ and $y' \sim y$ such that $r(x', y') \in X$ and $\mathcal{T} \models s \sqsubseteq r$, or $r(y', x') \in X$ and $\mathcal{T} \models s \sqsubseteq r^-$;
3. if $r(x, y) \in X$, then Δ contains all $s(x, y)$ with $s \in \Sigma_\mathsf{C}$ and $\mathcal{T} \models s \sqsubseteq r$ and all $s(y, x)$ with $s \in \Sigma_\mathsf{C}$ and $\mathcal{T} \models s^- \sqsubseteq r$.

T *avoids* q if it avoids some set X of atoms containing an atom α_i in q_i for any $i \in I$. 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_C-ABox and let π assign individual names $\pi(x_i)$ 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 $r \in \Sigma_\mathsf{C}$. $\mathcal{A}$ *realizes* T using π if there exists a $\mathcal{T}$-decoration f that realizes T using π.

Lemma 9.4. *Let $\mathcal{A}$ be a Σ_C-ABox consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$. Then $\mathcal{A} \not\models Q(\pi(x_1), \ldots, \pi(x_n))$ iff $\mathcal{A}$ realizes some $(\mathcal{T}, q)$-typing T using π that avoids q and such that $\mathsf{tp}(T)$ is Σ_C-realizable.*

The proof is a modification of the proof of Lemma 9.2 and given in the appendix.

We now construct the actual rewriting $\Phi_Q(\vec{x})$. For every $(\mathcal{T}, q)$-typing $T = (\sim, f_0, \Gamma, \Delta)$ with $\Gamma = \{t_1, \ldots, t_k\}$ 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_C-realizable.

Proposition 9.5. *Let $\mathcal{A}$ be a Σ_C-ABox that is consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$. Then $\mathcal{A} \models Q(\vec{a})$ iff $\mathcal{I}_\mathcal{A} \models \Phi_Q(\vec{a})$, for all $\vec{a}$ in $\mathsf{Ind}(\mathcal{A})$.*

Proof. Let $\vec{a} = (a_1, \ldots, a_n)$. Assume $\mathcal{A} \not\models Q(a_1, \ldots, a_n)$. Let $\pi(x_i) = a_i$ for $1 \leq i \leq n$. By Lemma 9.4, $\mathcal{A}$ realizes some $(\mathcal{T}, q)$-typing T using π that avoids q such that $\mathsf{tp}(T)$

is Σ_{C}-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 $\mathcal{I}_{\mathcal{A}} \not\models \Phi_Q(a_1, \dots, a_n)$. Take a $(\mathcal{T}, q)$-typing T that avoids q such that $\mathsf{tp}(T)$ is Σ_{C}-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 $\mathcal{A} \not\models Q(a_1, \dots, a_n)$, by Lemma 9.4. ☐

This finishes the proof of Theorem 9.1.

We now show that without the restriction on RIs adopted in Theorem 9.1, OMQCs from (DL-Lite$_{\mathcal{R}}$, $\mathsf{N_C} \cup \mathsf{N_R}$, UqfCQ) are no longer FO-rewritable. In fact, we prove the following, slightly stronger result by reduction from propositional satisfiability.

Theorem 9.6. *There is a DL-Lite$_{\mathcal{R}}$ TBox with closed predicates $(\mathcal{T}, \Sigma_{\mathsf{C}})$ such that ABox consistency w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ 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 $x_i^\top$, $x_i^\bot$, x_i^{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 $1 \leq i \leq n$;
- $r(c_i, x_j^\bot)$, for all $\neg x_j \in c_i$ and $1 \leq i \leq n$;
- $r'(x_j^\top, x_j^\bot)$, $r'(x_j^\bot, x_j^{\mathsf{aux}})$, 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

$$\mathcal{T} = \{s \sqsubseteq r, A \sqsubseteq \exists s, \exists s^- \sqsubseteq \exists s', s' \sqsubseteq r', \exists s'^- \sqcap \exists s^- \sqsubseteq \bot\}.$$

Let $\Sigma_{\mathsf{C}} = \{A, r, r'\}$. We show that $\mathcal{A}_\varphi$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ iff φ is satisfiable. Assume first that $\mathcal{A}_\varphi$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$. 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 $(c_i, x_j^\top) \in s^{\mathcal{I}}$ and set $v(x_j) = 0$ if there exists i such that $(c_i, x_j^\bot) \in s^{\mathcal{I}}$. Observe that v is well-defined since if $(c_i, x_j^\top) \in s^{\mathcal{I}}, (c_k, x_j^\bot) \in s^{\mathcal{I}}$, then $(x_j^\top, x_j^\bot) \in s'^{\mathcal{I}}$ and so $x_j^\bot \in (\exists s'^- \sqcap \exists s^-)^{\mathcal{I}}$ which contradicts the assumption that $\mathcal{I}$ satisfies $\exists s'^- \sqcap \exists s^- \sqsubseteq \bot$. Next observe that for every c_i there exists a disjunct $\ell \in c_i$ such that $(c_i, x_j^\top) \in s^{\mathcal{I}}$ if $\ell = x_j$ and $(c_i, x_j^\bot) \in s^{\mathcal{I}}$ 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 $\mathcal{I}$ by expanding $\mathcal{I}_{\mathcal{A}_\varphi}$ as follows:

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

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

We close this section with noting that, for the case of $\mathcal{EL}$, quantifier-free queries are computationally no more well-behaved than unrestricted queries. In fact, we have seen that OMQCs in $\mathcal{EL}$ using dtUCQs can be equivalently expressed using atomic database queries $A(x)$ by adding CIs of the form $C_q \sqsubseteq A$ to the TBox.

10. Conclusion

We have investigated the data complexity of ontology-mediated query evaluation with closed predicates, focussing on a non-uniform analysis. At the TBox level we have obtained PTime/coNP dichotomy results for the lightweight DLs $\mathcal{EL}$ and DL-Lite$_\mathcal{R}$. At the query level, the situation is drastically different: there is provably no PTime/coNP dichotomy for neither DL-Lite$_\mathcal{R}$ nor $\mathcal{EL}$ (unless PTime =coNP) and even without closing role names, understanding the complexity of queries is as hard as understanding the complexity of the generalized surjective constraint satisfaction problems. We have also shown that by combining DL-Lite$_\mathcal{R}$ with quantifier-free database queries one obtains FO-rewritable queries and that even for expressive DLs query evaluation is always in coNP. Many challenging open questions remain.

Regarding the data complexity classification at TBox level, it is shown in [45] that the dichotomy proof given for DL-Lite$_\mathcal{R}$ and $\mathcal{EL}$ does not go through for the extension $\mathcal{ELI}$ of $\mathcal{EL}$ with inverse roles. In fact, in contrast to DL-Lite$_\mathcal{R}$ and $\mathcal{EL}$, there are $\mathcal{ELI}$ TBoxes with closed predicates $(\mathcal{T}, \Sigma_\mathsf{C})$ such that CQ evaluation w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ is in PTime, but $(\mathcal{T}, \Sigma_\mathsf{C})$ and $(\mathcal{T}, \emptyset)$ are not CQ-inseparable on consistent ABoxes. In particular, it remains open whether there is a PTime/coNP dichotomy for TBoxes with closed predicates in $\mathcal{ELI}$. The same question remains open for $\mathcal{ALCHI}$ TBoxes (recall that there is a PTime/coNP dichotomy for for $\mathcal{ALCHI}$ TBoxes without closed predicates [47, 33]) and for expressive Horn languages such as Horn-$\mathcal{SHIQ}$. Also of interest are ontologies consisting of tuple-generating dependencies (tgds) which generalizes both DL-Lite$_\mathcal{R}$ and $\mathcal{EL}$. In this case, however, the coNP upper bound established here for $\mathcal{ALCHI}$ does not hold, even for the moderate extension consisting of linear tgds [10, 9].

Regarding the data complexity classification at the OMQC level, it would be of interest to consider DL-Lite$_\text{core}$: it remains open whether there is a PTime/coNP dichotomy for the language (DL-Lite$_\text{core}$, $\mathsf{N_C} \cup \mathsf{N_R}$, BUtCQ) and whether Theorem 8.3 can be strengthened to this case.

Acknowledgments. Frank Wolter was supported by EPSRC grant EP/M012646/1. Carsten Lutz was supported by the ERC Consolidator Grant 647289 CODA. We thank the anonymous reviewers for their suggestions and comments.

References

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

[2] Shqiponja Ahmetaj, Magdalena Ortiz, and Mantas Simkus. Polynomial datalog rewritings for expressive description logics with closed predicates. In *Proc. of IJCAI*, pages 878–885, 2016.

[3] Giovanni Amendola, Nicola Leone, Marco Manna, and Pierfrancesco Veltri. Enhancing existential rules by closed-world variables. In *Proc. of IJCAI*, pages 1676–1682, 2018.

[4] Marcelo Arenas, Pablo Barceló, Leonid Libkin, and Filip Murlak. *Foundations of Data Exchange*. Cambridge University Press, 2014.

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

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

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

[8] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.

[9] Michael Benedikt and Pierre Bourhis. Pspace hardness of mixed world query answering for atomic queries under guarded tgds. Technical report, University of Oxford, 2018.
[10] Michael Benedikt, Pierre Bourhis, Balder ten Cate, and Gabriele Puppis. Querying visible and invisible information. In *Proc. of LICS*, pages 297–306, 2016.
[11] Meghyn Bienvenu, Peter Hansen, Carsten Lutz, and Frank Wolter. First order-rewritability and containment of conjunctive queries in horn description logics. In *Proc. of IJCAI*, pages 965–971, 2016.
[12] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. First-order rewritability of atomic queries in horn description logics. In *Proc. of IJCAI*, pages 754–760, 2013.
[13] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *Proc. of Reasoning Web*, pages 218–307, 2015.
[14] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive datalog, CSP, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33, 2014.
[15] Manuel Bodirsky, Jan Kára, and Barnaby Martin. The complexity of surjective homomorphism problems–a survey. *Discrete Applied Mathematics*, 160(12):1680–1690, 2012.
[16] Elena Botoeva, Boris Konev, Carsten Lutz, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Inseparability and conservative extensions of description logic ontologies: A survey. In *Proc. of Reasoning Web*, pages 27–89, 2016.
[17] Elena Botoeva, Roman Kontchakov, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Games for query inseparability of description logic knowledge bases. *Artif. Intell.*, 234:78–119, 2016.
[18] Andrei A. Bulatov. A dichotomy theorem for nonuniform csps. In *Proc. of FOCS*, pages 319–330, 2017.
[19] Andrei A. Bulatov, Peter Jeavons, and Andrei A. Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34(3):720–742, 2005.
[20] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *Journal of Automated Reasoning*, 39(3):385–429, 2007.
[21] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Data complexity of query answering in description logics. In *KR*, pages 260–270, 2006.
[22] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. EQL-Lite: Effective first-order query processing in description logics. In *IJCAI*, 2007.
[23] Hubie Chen. An algebraic hardness criterion for surjective constraint satisfaction. *Algebra Universalis*, 72(4):393–401, 2014.
[24] Francesco M. Donini, Daniele Nardi, and Riccardo Rosati. Description logics of minimal knowledge and negation as failure. *ACM Transactions on Computational Logic*, 3(2):177–225, 2002.
[25] Thomas Eiter, Georg Gottlob, and Heikki Mannila. Disjunctive datalog. *ACM Trans. Database Syst.*, 22(3):364–418, 1997.
[26] Tomás Feder and Moshe Y. Vardi. Monotone monadic SNP and constraint satisfaction. In *STOC*, pages 612–622, 1993.
[27] Cristina Feier, Antti Kuusisto, and Carsten Lutz. Rewritability in monadic disjunctive datalog, mmsnp, and expressive description logics (invited talk). In *Proc. ICDT*, pages 1:1–1:17, 2017.
[28] Jan Foniok, Jaroslav Nesetril, and Claude Tardif. Generalised dualities and maximal finite antichains in the homomorphism order of relational structures. *Eur. J. Comb.*, 29(4):881–899, 2008.
[29] Enrico Franconi, Yazmin Angélica Ibáñez-García, and İnanç Seylan. Query answering with DBoxes is hard. *Electronic Notes in Theoretical Computer Science*, 278:71–84, 2011.
[30] Birte Glimm and Markus Krötzsch. SPARQL beyond subgraph matching. In *ISWC*, volume 6496 of *LNCS*, pages 241–256. Springer, 2010.
[31] Stephan Grimm and Boris Motik. Closed world reasoning in the semantic web through epistemic operators. In *OWLED*, 2005.
[32] André Hernich, Leonid Libkin, and Nicole Schweikardt. Closed world data exchange. *ACM Trans. Database Syst.*, 36(2):14:1–14:40, 2011.
[33] André Hernich, Carsten Lutz, Fabio Papacchini, and Frank Wolter. Dichotomies in ontology-mediated querying with the guarded fragment. In *Proc. of PODS*, pages 185–199, 2017.
[34] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Data complexity of reasoning in very expressive description logics. In *Proc. of IJCAI*, pages 466–471, 2005.
[35] Neil Immerman. *Descriptive Complexity.* Springer, 1999.

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

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

[38] Roman Kontchakov and Michael Zakharyaschev. An introduction to description logics and query rewriting. In *Proc. of Reasoning Web*, pages 195–244, 2014.

[39] Adila Krisnadhi and Carsten Lutz. Data complexity in the *EL* family of description logics. In *Proc. LPAR*, pages 333–347, 2007.

[40] Andrei A. Krokhin. Tree dualities for constraint satisfaction. In *CSL*, pages 32–33, 2010.

[41] Markus Krötzsch, Frederick Maier, Adila Krisnadhi, and Pascal Hitzler. A better uncle for OWL: nominal schemas for integrating rules and ontologies. In *Proc. of WWW*, pages 645–654. ACM, 2011.

[42] Markus Krötzsch and Sebastian Rudolph. Nominal schemas in description logics: Complexities clarified. In *Proc. of KR.* AAAI Press, 2014.

[43] Leonid Libkin and Cristina Sirangelo. Data exchange and schema mappings in open and closed worlds. *J. Comput. Syst. Sci.*, 77(3):542–571, 2011.

[44] Carsten Lutz, Inanç Seylan, and Frank Wolter. Ontology-mediated queries with closed predicates. In *Proc. of IJCAI*, pages 3120–3126. AAAI Press, 2015.

[45] Carsten Lutz, İnanç Seylan, and Frank Wolter. Ontology-based data access with closed predicates is inherently intractable (sometimes). In *Proc. of IJCAI*, pages 1024–1030, 2013.

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

[47] Carsten Lutz and Frank Wolter. The data complexity of description logic ontologies. *Logical Methods in Computer Science*, 13(4), 2017.

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

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

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

[51] Nhung Ngo, Magdalena Ortiz, and Mantas Simkus. Closed predicates in description logics: Results on combined complexity. In *Proc. of KR*, pages 237–246, 2016.

[52] Magdalena Ortiz, Diego Calvanese, and Thomas Eiter. Data complexity of query answering in expressive description logics via tableaux. *Journal of Automated Reasoning*, 41(1):61–98, 2008.

[53] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *J. Data Semantics*, 10:133–173, 2008.

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

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

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

[57] Balder ten Cate, Enrico Franconi, and İnanç Seylan. Beth definability in expressive description logics. *Journal of Artificial Intelligence Research*, 48:347–414, 2013.

[58] Guohui Xiao, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev. Ontology-based data access: A survey. In *Proc. of IJCAI*, pages 5511–5519, 2018.

[59] Dmitriy Zhuk. A proof of CSP dichotomy conjecture. In *Proc. of FOCS*, pages 331–342, 2017.

Appendix A. Missing Proofs for Section 4

Lemma 4.4 *Let $\mathcal{A}$ be a Σ_A-ABox, $\vec{a}$ a tuple in $\mathsf{Ind}(\mathcal{A})$, and $Q = (\mathcal{T}, \Sigma_\mathsf{A}, \Sigma_\mathsf{C}, q)$ a OMQC from $(\mathcal{ALCHI}, \mathsf{N_C} \cup \mathsf{N_R}, UCQ)$. Then the following are equivalent:*

(1) $\mathcal{A} \models Q(\vec{a})$;
(2) $\mathcal{I} \models q(\vec{a})$ *for all forest-shaped models $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respect Σ_C and such that*
- *the arity of $\Delta^\mathcal{I}$ is $|\mathcal{T}|$,*
- *$\mathsf{Ind}(\mathcal{A})$ is the set of roots of $\Delta^\mathcal{I}$,*
- *for every $d \in \Delta^\mathcal{I} \setminus \mathsf{Ind}(\mathcal{A})$ and $\exists r.C \in \mathsf{cl}(\mathcal{T})$ with $d \in (\exists r.C)^\mathcal{I}$, there exists $a \in \mathsf{Ind}(\mathcal{A})$ with $(d,a) \in r^\mathcal{I}$ and $a \in C^\mathcal{I}$ or there exists a successor d' of d in $\Delta^\mathcal{I}$ such that $(d,d') \in r^\mathcal{I}$ and $d' \in C^\mathcal{I}$.*

Proof. The implication from (1) to (2) is trivial. For the converse direction, suppose $\mathcal{A} \not\models Q(\vec{a})$. Then there is some model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C such that $\mathcal{J} \not\models q(\vec{a})$. We construct, by induction, a sequence of interpretations $\mathcal{I}_0, \mathcal{I}_1, \ldots$. The domain of each $\mathcal{I}_i$ consists of sequences of the form $d_0 \cdot d_1 \cdots d_n$, where $d_j \in \Delta^\mathcal{J}$ for all $j \in \{0, \ldots, n\}$. We call such sequences *paths* and denote the last element in a path p by $\mathsf{tail}(p)$, e.g., $\mathsf{tail}(d_0 \cdots d_n) = d_n$.

We define $\mathcal{I}_0$ as the restriction of $\mathcal{J}$ to $\mathsf{Ind}(\mathcal{A})$.

Assume now that $\mathcal{I}_i$ is given. Let $p \in \Delta^{\mathcal{I}_i}$ such that for some $e \in \Delta^\mathcal{J}$ and $\exists r.C \in \mathsf{cl}(\mathcal{T})$, we have $(\mathsf{tail}(p), e) \in r^\mathcal{J}$ and $e \in C^\mathcal{J}$ and there is no $p' \in \Delta^{\mathcal{I}_i}$ with $\mathsf{tail}(p') = e'$ and $(p,p') \in r^{\mathcal{I}_i}$ 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 setting

$$\begin{aligned}
\Delta^{\mathcal{I}_{i+1}} &= \Delta^{\mathcal{I}_i} \cup \{p \cdot e\} \\
s^{\mathcal{I}_{i+1}} &= s^{\mathcal{I}_i} \cup \{(p, p \cdot e) \mid (\mathsf{tail}(p), e) \in s^\mathcal{J}\} \cup \{(p \cdot e, p) \mid (e, \mathsf{tail}(p)) \in s^\mathcal{J}\} \\
A^{\mathcal{I}_{i+1}} &= A^{\mathcal{I}_i} \cup \{p \cdot e \mid e \in A^\mathcal{J}\}
\end{aligned}$$

for all role names s and concept names A. 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 (p,e) to $s^{\mathcal{I}_i}$, for every role s such that $(\mathsf{tail}(p), e) \in s^\mathcal{J}$.

We assume that the above construction is *fair* in the sense that if the conditions of the inductive step are satisfied for some $p \in \Delta^{\mathcal{I}_i}$, $e \in \Delta^\mathcal{J}$, and $\exists r.C \in \mathsf{cl}(\mathcal{T})$, with $i \geq 0$, then there is some $j > i$ such that the inductive step is applied to p, e, and $\exists r.C$.

Now we define the interpretation $\mathcal{I}$ as the limit of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$:
- $\Delta^\mathcal{I} = \bigcup_{i \geq 0} \Delta^{\mathcal{I}_i}$;
- $P^\mathcal{I} = \bigcup_{i \geq 0} P^{\mathcal{I}_i}$, for all $P \in \mathsf{N_C} \cup \mathsf{N_R}$.

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. That $\mathcal{I}$ is a model of $\mathcal{A}$ is an easy consequence of the facts that $\mathcal{J}$ is a model of $\mathcal{A}$, $\mathcal{I}_0$ is the restriction of $\mathcal{J}$ to $\mathsf{Ind}(\mathcal{A})$, and $\mathcal{I}$ is an extension of $\mathcal{I}_0$. That $\mathcal{I}$ respects closed predicates Σ_C is by definition. We now show that $\mathcal{I}$ is a model of $\mathcal{T}$. The following is easily proved by structural induction.

Claim. For all $p \in \Delta^\mathcal{I}$ and $C \in \mathsf{cl}(\mathcal{T})$, $p \in C^\mathcal{I}$ iff $\mathsf{tail}(p) \in C^\mathcal{J}$.

The fact that $\mathcal{J}$ is a model of $\mathcal{T}$ now implies that $\mathcal{I}$ is a model of every CI in $\mathcal{T}$. That $\mathcal{I}$ is a model of every RI in $\mathcal{T}$ follows by construction. Hence we conclude that $\mathcal{I}$ is a model of $\mathcal{T}$.

Finally, to show that $\mathcal{I} \not\models q(\vec{a})$, observe that $h = \{p \mapsto \mathsf{tail}(p) \mid p \in \Delta^{\mathcal{I}}\}$ is a homomorphism from $\mathcal{I}$ to $\mathcal{J}$ preserving $\mathsf{N_I}$. Thus, $\mathcal{I} \not\models q(\vec{a})$ follows from Lemma 4.3 and $\mathcal{J} \not\models q(\vec{a})$. □

Lemma A.1. *The interpretation $\mathcal{I}$ defined in the proof of Lemma 4.7 is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C} such that $\mathcal{I} \not\models q(\vec{a})$.*

Proof. 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 RI in $\mathcal{T}$;
- $P^{\mathcal{I}} = \{\vec{a} \mid P(\vec{a}) \in \mathcal{A}\}$, for all predicates $P \in \Sigma_{\mathsf{C}}$.

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 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 element of $\{d, e\}$ 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 element of $(\Delta^{\mathcal{I}} \setminus \mathsf{Ind}(\mathcal{A})) \cap \{d, e\}$. 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.

It remains to show that $\mathcal{I} \not\models q(\vec{a})$. Assume $\vec{a} = (a_1, \ldots, a_n)$. For a proof by contradiction, suppose that $\mathcal{I} \models q(\vec{a})$. Then there is a disjunct $\exists \vec{y} \varphi(\vec{x}, \vec{y})$ of q with $\vec{x} = (x_1, \ldots, x_n)$ and φ a conjunction of atoms such that there is an assignment π mapping the variables $\vec{x} \cup \vec{y}$ of φ to $\Delta^{\mathcal{I}}$ with $\pi(x_i) = a_i$ for $1 \leq i \leq n$ and $\mathcal{I} \models_\pi \varphi$. Let $F = \{\pi(x) \mid \pi(x) \notin \mathsf{Ind}(\mathcal{A})\}$. As $\mathcal{I}$ is forest-shaped there are $T_1, \ldots, T_m$ with $F = T_1 \cup \cdots \cup T_m$ such that $T_1, \ldots, T_m$ are maximal and pairwise disjoint trees in F. Fix an $i \in \{1, \ldots, m\}$. Let d be the root of T_i. By the construction of $\mathcal{I}$, there is an isomorphism f_i trom $(\mathcal{I}_d, \tau_d)$ to some $(\mathcal{J}, \tau) \in M$. 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

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

π' is an assignment in $\biguplus_{(\mathcal{J},\tau)\in M} \mathcal{J}$ with $\pi'(x_i) = a_i$ for $1 \leq i \leq n$ such that $\biguplus_{(\mathcal{J},\tau)\in M} \mathcal{J} \models_{\pi'} \varphi$, and so we have derived a contradiction. □

Appendix B. Missing Proofs for Section 9

Lemma 9.2 *A Σ_{C}-ABox $\mathcal{A}$ is consistent w.r.t. $(\mathcal{T}, \Sigma_{\mathsf{C}})$ iff*

(1) *$\mathcal{A}$ has a $\mathcal{T}$-decoration f whose image is a Σ_{C}-realizable $\mathcal{T}$-typing and*
(2) *if $s(a,b) \in \mathcal{A}$, $\mathcal{T} \models s \sqsubseteq r$, and $\mathsf{sig}(s \sqsubseteq r) \subseteq \Sigma_{\mathsf{C}}$, then $r(a,b) \in \mathcal{A}$.*

Proof. ($\Rightarrow$) Let $\mathcal{I}$ be a model of $\mathcal{A}$ and $\mathcal{T}$ that respects closed predicates Σ_{C}. For each $d \in \Delta^{\mathcal{I}}$, let $t^d_{\mathcal{I}} = \{B \in \mathsf{con}(\mathcal{T}) \mid d \in B^{\mathcal{I}}\}$ and let $T_{\mathcal{I}} = \{t^a_{\mathcal{I}} \mid a \in \mathsf{Ind}(\mathcal{A})\}$. We next show that the $\mathcal{T}$-typing $T_{\mathcal{I}}$ is Σ_{C}-realizable. Let $t^a_{\mathcal{I}}, r_1, \ldots, r_n$ be a Σ_{C}-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

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

By definition of Σ_{C}-participating paths, there is some $B^\star \in \mathsf{con}(\mathcal{T})$ with $\mathsf{sig}(B^\star) \subseteq \Sigma_{\mathsf{C}}$ such that $\mathcal{T} \models \exists r_n^- \sqsubseteq B^\star$. By Point (b), 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_{C}, we have $g(n) = b$ for some $b \in \mathsf{Ind}(\mathcal{A})$. By Point (b), $\mathcal{T} \models \exists r_n^- \sqsubseteq B$ implies $B \in t^b_{\mathcal{I}}$ for any $B \in \mathsf{con}(\mathcal{T})$. Thus, $T_{\mathcal{I}}$ is Σ_{C}-realizable. Let $f(a) = t^a_{\mathcal{I}}$ 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_{C}-realizable $\mathcal{T}$-typing. Hence we conclude that $\mathcal{A}$ satisfies Point (1). Point (2) holds by the fact that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C}.

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

(a) for all $a \in \mathsf{Ind}(\mathcal{A})$, we have $t^a_{\mathcal{I}_i} \subseteq f(a)$;
(b) for all $p \in \Delta^{\mathcal{I}_i}$, if $p = t, r_1 \ldots, r_n$, then we have $t^p_{\mathcal{I}_i} \subseteq \{B \in \mathsf{con}(\mathcal{T}) \mid \mathcal{T} \models \exists r_n^- \sqsubseteq B\}$.

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

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

To construct $\mathcal{I}_{i+1}$ from $\mathcal{I}_i$, choose $d \in \Delta^{\mathcal{I}_i}$ and $\exists s \in \mathsf{con}(\mathcal{T})$ such that $\mathsf{sig}(s) \cap \Sigma_{\mathsf{C}} = \emptyset$, $\mathcal{T} \models \bigsqcap t^d_{\mathcal{I}_i} \sqsubseteq \exists s$ 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 (a) and (b), it is easy to verify that q is a path in T. If q is not Σ_{C}-participating, then define $\mathcal{I}_{i+1}$ as follows:

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

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

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

Assume that the choice of $d \in \Delta^{\mathcal{I}_i}$ and $\exists s \in \mathsf{con}(\mathcal{T})$ is fair so that every possible combination of d and $\exists s$ is eventually chosen. Let $\mathcal{I}$ be the limit of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$ (cf. the proof of Lemma 4.4). We claim that $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_{C}. By definition of $\mathcal{I}_0$ and of $\mathcal{T}$-decorations, it is straightforward to see that $\mathcal{I} \models \mathcal{A}$. Moreover, the RIs in $\mathcal{T}$ are clearly satisfied. To show that the CIs are satisfied as well, it is straightforward to first establish the following strengthenings of Conditions (a) and (b) above (details omitted):

(a′) for all $a \in \mathsf{Ind}(\mathcal{A})$, we have $t^a_{\mathcal{I}} = f(a)$;

(b′) for all $p \in \Delta^{\mathcal{I}}$, if $p = t, r_1 \ldots, r_n$, then $t^p_{\mathcal{I}_i} = \{B \in \mathsf{con}(\mathcal{T}) \mid \mathcal{T} \models \exists r_n^- \sqsubseteq B\}$.

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 ($a'$) 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 (b′), 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 (b′) 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 that $f(a)$ 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 $A \in \Sigma_{\mathsf{C}}$, $A^{\mathcal{I}} = \{a \mid A(a) \in \mathcal{A}\}$;
- for all $r \in \Sigma_{\mathsf{C}}$, $r^{\mathcal{I}} = \{(a,b) \mid r(a,b) \in \mathcal{A}\}$.

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 $A \in \Sigma_{\mathsf{C}}$. 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 $A \in \Sigma_{\mathsf{C}}$. Then $A \in f(a)$. The definition of $\mathcal{T}$-decorations yields $A \in t^a_{\mathcal{A}}$, and thus $A(a) \in \mathcal{A}$. Now let $r(a,b) \in \mathcal{A}$ with $r \in \Sigma_{\mathsf{C}}$. 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 $r \in \Sigma_{\mathsf{C}}$.

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 RIs, it follows that $\mathsf{sig}(s) \subseteq \Sigma_\mathsf{C}$. This yields $r(a,b) \in \mathcal{A}$ since $\mathcal{A}$ satisfies Point (2) of Lemma 9.2.

For $i > 0$, we show that the extension of Σ_C-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 $\exists s \in \mathsf{con}(\mathcal{T})$ and let $q = f(a), s$ if $d = a \in \mathsf{Ind}(\mathcal{A})$ and $q = d, s$ otherwise. Then $\mathsf{sig}(s) \cap \Sigma_\mathsf{C} = \emptyset$ and by the restriction on RIs, $\mathsf{sig}(r) \cap \Sigma_\mathsf{C} = \emptyset$ 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_C (no matter whether q is Σ_C-participating or not). In the case where q is Σ_C-participating, there is nothing else to show. If q is not Σ_C-participating, then each concept name A with $\mathcal{T} \models \exists s^- \sqsubseteq A$ is not from Σ_C. Thus also none of the concept names modified in the construction of $\mathcal{I}_{i+1}$ is from Σ_C. $\square$

Lemma 9.4 *Let $\mathcal{A}$ be a Σ_C-ABox consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$. Then $\mathcal{A} \not\models Q(\pi(x_1), \ldots, \pi(x_n))$ iff $\mathcal{A}$ realizes some $(\mathcal{T}, q)$-typing T using π that avoids q and such that $\mathsf{tp}(T)$ is Σ_C-realizable.*

Proof. The proof is a modification of the proof of Lemma 9.2. We only sketch the differences.

($\Rightarrow$) Let $\mathcal{A} \not\models Q(\pi(x_1), \ldots, \pi(x_n))$. We start with a model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{A}$ that respects closed predicates Σ_C 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 $1 \leq i \leq n$;
- $\Gamma = \{t_\mathcal{I}^a \mid a \in \mathsf{Ind}(\mathcal{A})\} \setminus \{\pi(x_1), \ldots, \pi(x_n)\}$;
- $\Delta = \{r(x_i, x_j) \mid r \in \Sigma_\mathsf{C}, r(\pi(x_i), \pi(x_j)) \notin \mathcal{A}\}$.

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 α_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 $\mathsf{sig}(B) \cap \Sigma_\mathsf{C} = \emptyset$ 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 $s \in \Sigma_\mathsf{C}$ 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 $s \in \Sigma_\mathsf{C}$ with $\mathcal{T} \models s^- \sqsubseteq r$. Thus $s(y,x) \in \Delta$ for any such s.

($\Leftarrow$) Assume that a Σ_C-Abox $\mathcal{A}$ that is consistent w.r.t. $(\mathcal{T}, \Sigma_\mathsf{C})$ 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 α_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_C 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 9.2 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 $\exists s \in \mathsf{con}(\mathcal{T})$ such that $\mathsf{sig}(s) \cap \Sigma_\mathsf{C} = \emptyset$, $\mathcal{T} \models \bigsqcap t_{\mathcal{I}_i}^d \sqsubseteq \exists s$ and there is no $(d,e) \in s^{\mathcal{I}_i}$. If $d \notin \{\pi(x_1), \ldots, \pi(x_n)\}$ or $\{B \in \mathsf{con}(\mathcal{T}) \mid \mathcal{T} \models \exists s^- \sqsubseteq B\}$ does not contain a B with $\mathsf{sig}(B) \subseteq \Sigma_\mathsf{C}$ proceed as in the proof of Lemma 9.2. Now assume that $d = \pi(x)$. In the proof of Lemma 9.2 we chose an *arbitrary* $b \in \mathsf{Ind}(\mathcal{A})$ with $\{B \in \mathsf{con}(\mathcal{T}) \mid \exists s^- \sqsubseteq B\} \subseteq t$ 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 $\{B \in \mathsf{con}(\mathcal{T}) \mid \exists s^- \sqsubseteq B\} \subseteq t$ and $t = f(b)$, then we choose such a b and proceed as in Lemma 9.2. 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 $\{B \in \mathsf{con}(\mathcal{T}) \mid \mathcal{T} \models \exists s^- \sqsubseteq B\} \subseteq f_0(y)$ 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 9.2.

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