From Conjunctive Queries to Instance Queries in Ontology-Mediated Querying

Cristina Feier¹, Carsten Lutz¹, Frank Wolter²,
¹ University of Bremen, Germany
² University of Liverpool, UK
feier@uni-bremen.de, clu@uni-bremen.de, wolter@liverpool.ac.uk

Abstract

We consider ontology-mediated queries (OMQs) based on expressive description logics of the \textit{ALC} family and (unions) of conjunctive queries, studying the rewritability into OMQs based on instance queries (IQs). Our results include exact characterizations of when such a rewriting is possible and tight complexity bounds for deciding rewritability. We also give a tight complexity bound for the related problem of deciding whether a given MMSNP sentence is equivalent to a CSP.

1 Introduction

An ontology-mediated query (OMQ) is a database-style query enriched with an ontology that contains domain knowledge, aiming to deliver more complete answers. In OMQs, ontologies are often formulated in a description logic (DL) and query languages of interest include conjunctive queries (CQs), unions of conjunctive queries (UCQs), and instance queries (IQs). While CQs and UCQs are widely known query languages that play a fundamental role also in database systems and theory, IQs are more closely related to DLs. In fact, an IQ takes the form \( C(x) \) with \( C \) a concept formulated in the DL that is also used for the ontology, and thus the expressive power of IQs depends on the ontology language. OMQs based on (UC)Qs are more powerful than OMQs based on IQs as the latter only serve to return all objects from the data that are instances of a given class.

It is easy to see that IQs can express tree-shaped CQs with a single answer variable as well as unions thereof. In fact, this observation has been used in many technical constructions in the area, see for example [Calvanese et al., 1998; Glimm et al., 2008; Lutz, 2008; Eiter et al., 2012a]. Intriguingly, though, it was observed by Zolin [2007] that tree-shaped CQs are not the limit of IQ-rewritability when we have an expressive DL such as \textit{ALC} or \textit{ALCII} at our disposal. For example, the CQ \( r(x, x) \), which asks to return all objects from the data that are involved in a reflexive \( r \)-loop, can be rewritten into the equivalent \textit{ALC-IQ} \( P \rightarrow \exists r.P(x) \). Here, \( P \) behaves like a monadic second-order variable due to the open-world assumption made for OMQs: we are free to interpret \( P \) in any possible way and when making \( P \) true at an object we are forced to make also \( \exists r.P \) true if and only if the object is involved in a reflexive \( r \)-loop. It is an interesting question, raised in [Zolin, 2007] [Kikot and Zolin, 2013] [Kikot et al., 2013], to precisely characterize the class of CQs that are rewritable into IQs. An important step into this direction has been made by Kikot and Zolin [2013] who identify a large class of CQs that are rewritable into IQs: a CQ is rewritable into an \textit{ALC-IQ} if it is connected and every cycle passes through the (only) answer variable; for rewritability into an \textit{ALC-IQ}, one additionally requires that all variables are reachable from the answer variable in a directed sense. It remained open whether these classes are depleting, that is, whether they capture all CQs that are IQ-rewritable.

There are two additional motivations to study the stated question. The first one comes from concerns about the practical implementation of OMQs. When the ontology is formulated in a more inexpressive ‘Horn DL’, OMQ evaluation is possible in PTIME data complexity and a host of techniques for practically efficient OMQ evaluation is available, see for example [Pérez-Urbina et al., 2010; Eiter et al., 2012b; Trivela et al., 2015; Lutz et al., 2009]. In the case of expressive DLs such as \textit{ALC} and \textit{ALCII}, OMQ evaluation is \textit{coNP}-complete in data complexity and efficient implementation is much more challenging. In particular, there are hardly any systems that fully support such OMQs when the actual queries are (UC)Qs. In contrast, the evaluation of OMQs based on (expressive DLs and) IQs is supported by several systems such as Pellet, Hermit, and PAGODA [Sirin et al., 2007; Glimm et al., 2014; Zhou et al., 2015]. For this reason, rewriting (UC)Qs into IQs has been advocated in [Zolin, 2007] [Kikot and Zolin, 2013] [Kikot et al., 2013] as an approach towards efficient OMQ evaluation with expressive DLs and (UC)Qs. The experiments and optimizations reported in [Kikot et al., 2013] show the potential (and challenges) of this approach.

The second motivation stems from the connection between OMQs and constraint satisfaction problems (CSPs) [Bienvenu et al., 2014; Lutz and Wolter, 2017]. Let \((\mathcal{L}, \mathcal{Q})\) denote the class of OMQs based on ontologies formulated in the DL \(\mathcal{L}\) and the query language \(\mathcal{Q}\). It was observed in [Bienvenu et al., 2014] that \((\textit{ALCII}, \textit{IQ})\) is closely related to the complement of CSPs while \((\textit{ALC}, \textit{UCQ})\) is closely related to the complement of the logical generalization MMSNP of CSP; we further remark that MMSNP is a notational variant
of the complement of (Boolean) monadic disjunctive Data-
log. Thus, characterizing OMQs from \((\mathcal{ALC}, \mathcal{UCQ})\) that are
rewritable into \((\mathcal{ALC}, \mathcal{IQ})\) is related to characterizing MM-
SNP sentences that are equivalent to a CSP, and we also study
the latter problem. In fact, the main differences to the OMQ
case are that unary queries are replaced with Boolean ones
and that predicates can have unrestricted arity.

The main aim of this paper is to study the rewritability of
OMQs from \((\mathcal{L}, (\mathcal{U})\mathcal{CQ})\) into OMQs from \((\mathcal{L}, \mathcal{IQ})\), consider-
ing \(\mathcal{L}\) the basic expressive DL \(\mathcal{ALC}\) as well as extensions
of \(\mathcal{ALC}\) with inverse roles, role hierarchies, the universal role,
and functional roles. We provide precise characterizations,
tight complexity bounds for deciding whether a given OMQ
is rewritable, and show how to construct the rewritten query
when it exists. In fact, we prove that the classes of CQs from
[Kikot and Zolin, 2013] are depleting, but we go significantly
beyond that: while [Zolin, 2007] [Kikot and Zolin, 2013]
[Kikot et al., 2015] aim to find IQ-rewritings that work for
any ontology, we consider the more fine-grained question
of rewriting into an IQ an OMQ \((T, \Sigma, q(x))\) where \(T\) is a
DL TBox formalizing the ontology, \(\Sigma\) is an ABox signature,
and \(q(x)\) is the actual query. The ‘any ontology’ setup then
corresponds to the special case where \(T\) is empty and \(\Sigma\) is
full. However, giving a non-empty TBox or a non-full ABox
signature results in additional \((\mathcal{U})\mathcal{CQs}\) to become rewritable.
While we admit modification of the TBox during rewriting,
it turns out that this is mostly unnecessary: only in some
rather special cases, a moderate extension of the TBox pays
off. All this requires non-trivial generalizations of the query
interest. We include the letter \(H\) in the name of a TBox
language to indicate that role inclusions \(r \subseteq s\) are also admitted
in the TBox and likewise for the letter \(F\) and functionality
assertions \(\text{func}(r)\) where in both cases \(r, s\) are role names or
inverse roles in case that the concept language used admits
inverse roles. So it should be understood, for example, what
we mean with an \(\mathcal{ALCH}^u\)-TBox and an \(\mathcal{ALCFI}u\)-TBox.
As usual, the semantics is defined in terms of interpretations,
which take the form \(I = (\Delta^I, \mathcal{I})\) with \(\Delta^I\) a non-empty domain
and \(\mathcal{I}\) an interpretation function. An interpretation is a
model of a TBox \(T\) if it satisfies all inclusions and assertions
in \(T\), defined in the usual way. We write \(T \models r \subseteq s\) if every
model of \(T\) also satisfies the role inclusion \(r \subseteq s\).

An ABox is a set of concept assertions \(A(a)\) and role asser-
tions \(r(a, b)\) where \(A\) is a concept name, \(r\) a role name, and
\(a, b\) are individual names. We use \(\text{ind}(A)\) to denote the set of
all individual names that occur in \(A\). An interpretation is a
model of an ABox \(A\) if it satisfies all concept and role assertions
in \(A\), that is, \(a \in \Delta^A\) when \(A(a)\) is in \(A\) and \((a, b) \in r^A\)
when \(r(a, b)\) is in \(A\). An ABox is consistent with a TBox \(T\)
if \(A\) and \(T\) have a common model. A signature \(\Sigma\) is a set of
concept and role names. We use \(\text{sig}(T)\) to denote the set of
concept and role names that occur in the TBox \(T\), and like-
wise for other syntactic objects such as ABoxes. A \(\Sigma\)-ABox
is an ABox \(A\) such that \(\text{sig}(A) \subseteq \Sigma\).

A conjunctive query (CQ) is of the form \(q(x) =\exists y \varphi(x, y)\),
where \(x\) and \(y\) are tuples of variables and \(\varphi(x, y)\) is a
conjunction of atoms of the form \(A(x)\) or \(r(x, y)\) with \(A\)
a concept name, \(r\) a role name, and \(x, y \in x \cup y\). We call
\(x\) the answer variables of \(q(x)\) and \(y\) quantified variables.
For purposes of uniformity, we use \(r^+(x, y)\) as an alterna-
tive notation to denote an atom \(r(y, x)\) in a CQ. In fact, when

2 Preliminaries
We use standard description logic notation and refer to
[Baader et al., 2017] for full details. In contrast to the stan-
dard DL literature, we carefully distinguish between the concept
language and the TBox language. We consider four con-
cept languages. Recall that \(\mathcal{ALC}\)-concepts are formed
according to the syntax rule
\[
C, D ::= A | \neg C | C \sqcap D | C \sqcup D | \exists r.C | \forall r.C
\]
where \(A\) ranges over concept names and \(r\) over role names.
As usual, we use \(C \to D\) as an abbreviation for \(\neg C \sqcap D\).
\(\mathcal{ALC}\)-concepts additionally admit the use of inverse roles
\(\neg^*\) in concept constructors \(\exists^* r.C\) and \(\forall^* r.C\). With a role,
we mean a role name or an inverse role. \(\mathcal{ALC}^u\)-concepts ad-
ditionally admit the use of the universal role \(u\) in concept
constructors \(\exists u.C\) and \(\forall u.C\). In \(\mathcal{ALC}^u\)-concepts, both
inverse roles and the universal role are admitted.

We now introduce several TBox languages. For \(\mathcal{L}\) one of
the four concept languages introduced above, an \(\mathcal{L}\)-TBox
is a finite set of concept inclusions \(C \subseteq D\) where \(C\) and \(D\)
are \(\mathcal{L}\) concepts. So each concept language also serves as a
TBox language, but there are additional TBox languages of
interest. We include the letter \(H\) in the name of a TBox
language to indicate that role inclusions \(r \subseteq s\) are also admitted
in the TBox and likewise for the letter \(F\) and functionality
assertions \(\text{func}(r)\) where in both cases \(r, s\) are role names or
inverse roles in case that the concept language used admits
inverse roles. So it should be understood, for example, what
we mean with an \(\mathcal{ALCH}^u\)-TBox and an \(\mathcal{ALCFI}u\)-TBox.

Regarding IQ-rewritability as a decision problem, we show
NP-completeness for the case of the empty TBox. This
can be viewed as an underapproximation for the case with
non-empty TBox and ABox signature. With non-empty
TBoxes, complexities are higher. When the ABox signature
is full, we obtain 2ExpTime-completeness for DLs with
inverse roles and an ExpTime lower bound and a ConEx-
pTime upper bound for DLs without inverse roles. With
unrestricted ABox signature, the problem is 2NExpTime-
complete for DLs with inverse roles and NExpTime-hard
(and in 2NExpTime) for DLs without inverse roles. All
lower bounds hold for CQs and all upper bounds capture
UCQs. We also prove that it is 2NExpTime-complete to
decide whether a given MMSNP sentence is equivalent to a
CSP. This problem was known to be decidable [Madelaine
and Stewart, 2007], but the complexity was open.

We also consider \(\mathcal{ALCFI}\), the extension of \(\mathcal{ALC}\) with
functional roles, for which IQ-rewritability turns out to be un-
decidable and much harder to characterize. We give a rather
subtle characterization for the case of the empty TBox and
full ABox signature and show that the decision problem is
then decidable and NP-complete. Since it is not clear how to
apply CSP techniques, we use an approach based on ultrafil-
ters, starting from what was done for \(\mathcal{ALC}\) without functional
roles in [Kikot and Zolin, 2013].

Full proofs are in the appendix.
speaking about an atom $r(x,y)$ in a CQ $q(x)$, we generally also include the case that $r = s^-$ and $s(y, x)$ is the actual atom in $q(x)$, unless explicitly noted otherwise. Every CQ $q(x) = \exists y \varphi(x,y)$ gives raise to a directed graph $G_{q}$ whose nodes are the elements of $x \cup y$ and that contains an edge from $x$ to $y$ if $\varphi(x,y)$ contains an atom $r(x,y)$. The corresponding undirected graph is denoted $G_{q}^{u}$ (it might contain self loops). We can thus use standard terminology from graph theory to CSs, saying for example that a CQ is connected.

A homomorphism from $q(x)$ to an interpretation $I$ is a function $h : x \cup y \rightarrow \Delta^{2}$ such that $h(x) \neq h(y)$ for every atom $A$ of $q(x)$ and $(h(x), h(y)) \in A$ for every atom $r$ of $q(x)$. We write $I \models q(a)$ and call $a$ an answer to $q(x)$ on $I$ if there is a homomorphism from $q(x)$ to $I$ with $h(x) = a$.

A union of conjunctive queries (UCQ) $q(x)$ is a disjunction of one or more CSs that all have the same answer variables $x$. We say that a UCQ is connected if every CQ in it is. The arity of a (UC)CQ is the number of answer variables in it. For $L \in \{\text{ALC}, \text{ALC} T, \text{ALC}^{3}, \text{ALC} T^{3}\}$, an L-instance query (L-IQ) takes the form $C(x)$ where $C$ is an L concept and a x variable. We write $I \models C(a)$ if $a \in C^{T}$. All instance queries have arity 1.

An ontomediated query (OMQ) takes the form $Q = (T, \Sigma, q(x))$ with $T$ a TBox, $\Sigma \subseteq \text{sig}(T) \cup \text{sig}(q)$ an ABox signature, and $q(x)$ a query. The arity of $Q$ is the arity of $q(x)$ and $Q$ is Boolean if its arity zero. When $\Sigma = \text{sig}(T) \cup \text{sig}(q)$, then for brevity we denote it with $\Sigma_{\text{full}}$ and speak of the full ABox signature. Let $A$ be a $\Sigma$-ABox. A tuple $a \in \text{ind}(A)$ is an answer to $Q$ on $A$ if $I \models q(a)$ for all models $I$ of $\Sigma$ and $T$. We say that $Q$ is empty if for all $\Sigma$-ABoxes $A$, there is no answer to $Q$ on $A$. Let $Q_{1}, Q_{2}$ be OMQs, $Q_{1} = (T_{1}, \Sigma, q_{1}(x))$ for $i \in \{1, 2\}$. Then $Q_{1}$ is contained in $Q_{2}$, written $Q_{1} \subseteq Q_{2}$, if for all $\Sigma$-ABoxes $A$, every answer to $Q_{1}$ on $A$ is also an answer to $Q_{2}$ on $A$. Further, $Q_{1}$ and $Q_{2}$ are equivalent, written $Q_{1} \equiv Q_{2}$, if $Q_{1} \subseteq Q_{2}$ and $Q_{2} \subseteq Q_{1}$.

We use $(L, Q)$ to refer to the OMQ language in which the TBox is formulated in the language $L$ and where the actual queries are from the language $Q$, such as in $(\text{ALC}X, \text{UCQ})$. For brevity, we generally write $(L, IQ)$ instead of $(L, L^{\prime} \cdot IQ)$ when $L^{\prime}$ is the concept language underlying the TBox language $L$, so for example $(\text{ALC} \text{CH} \text{T}, \text{IQ})$ is short for $(\text{ALC} \text{CH} \text{T}, \text{ALC} T^{3} \cdot \text{IQ})$.

Definition 1. Let $(L, Q)$ be an OMQ language. An OMQ $Q = (T, \Sigma, q(x))$ is $(L, Q)$-rewritable if there is an OMQ $Q^{r}$ from $(L, Q)$ such that the answers to $Q$ and to $Q^{r}$ are identical on any $\Sigma$-ABox that is consistent with $T$. In this case, we say that $Q$ is rewritten into $Q^{r}$ and call $Q^{r}$ a rewriting of $Q$.

Let $(L, Q)$ be an OMQ-language. IQ-rewritability in $(L, Q)$ is the problem to decide whether a given (unary) OMQ $Q = (T, \Sigma, q(x))$ from $(L, Q)$ is $(L, Q)$-rewritable; for brevity, we simply speak of IQ-rewritability of $Q$ when this is the case. The following examples show that IQ-rewritability of $Q$ depends on several factors. All claims made are sanctioned by results established in this paper.

Example 2. (1) IQ-rewritability depends on the topology of the actual query. Let $q_{1}(x) = r(x, x)$. The OMQ $(0, \Sigma_{\text{full}}, q_{1}(x))$ is rewritable into the OMQ $(0, \Sigma_{\text{full}}, C(x))$ from $(\text{ALC} \Sigma), \text{IQ})$ where $C$ is $P \rightarrow \exists r.P$. In contrast, let $q_{2}(x) = \exists y s(x, y) \land r(y, y)$. The OMQ $(0, \Sigma_{\text{full}}, q_{2}(x))$ is not rewritable into an OMQ from $(\text{ALC} T, \Sigma)$. (2) For this we are not allowed to extend the TBox, IQ-rewritability depends on whether or not inverse roles are available. Let $\Sigma = \{r, s\}$ and $q(x) = \exists y r(y, x) \land s(y, x)$. The OMQ $Q = (0, \Sigma, q(x))$ is rewritable into the OMQ $(0, \Sigma, C(x))$ from $(\text{ALC} T, \Sigma)$ where $C$ is $P \rightarrow \exists s.P$. $Q$ is also rewritable into the OMQ $(0, \Sigma, C(x))$ from $(\text{ALC} T, \Sigma)$ where $T = \{\exists s.P \subseteq \forall r.P^\prime\}$, and $C$ is $P \rightarrow P^\prime$, but it is not rewritable into any OMQ $(0, \Sigma, C(x))$ from $(\text{ALC} T, \Sigma)$ with $T = \emptyset$.

(3) IQ-rewritability depends on the TBox. Let $q(x) = \exists x_{1} \exists y_{1} \exists y_{2} \exists z A(x) \land r(x, x_{1}) \land r(x_{1}, y_{1}) \land r(x_{1}, y_{2}) \land r(y_{1}, z) \land r(y_{2}, z) \land B_{1}(y_{1}) \land B_{2}(y_{2})$. The OMQ $(0, \Sigma_{\text{full}}, q(x))$ is not rewritable into an OMQ from $(\text{ALC} X, \text{IQ})$. Let $T = \{A \subseteq \exists r. \exists r. (B_{1} \cap B_{2} \cap \exists r. T^\prime)\}$. The OMQ $(0, \Sigma_{\text{full}}, q(x))$ is rewritable into the OMQ $(0, \Sigma_{\text{full}}, A(x))$ from $(\text{ALC} X, \text{IQ})$.

(4) IQ-rewritability depends on the ABox signature. Let $q(x)$ be the CQ from (3) without the atom $A(x)$ and let $T$ be as in (3). The OMQ $(0, \Sigma_{\text{full}}, q(x))$ is not rewritable into an OMQ from $(\text{ALC} T, \Sigma, \text{IQ})$. Let $\Sigma = \{A\}$. The OMQ $(0, \Sigma_{\text{full}}, q(x))$ is rewritable into the OMQ $(0, \Sigma_{\text{full}}, A(x))$ from $(\text{ALC} T, \Sigma)$.

Note that we are allowed to completely rewrite the TBox when constructing IQ-rewritings, which might seem questionable from a practical perspective. Fortunately, though, it turns out the TBox can always be left untouched or, in some rare cases, only needs to be slightly extended. Also note that an alternative definition of IQ-rewritability obtained by dropping the restriction to ABoxes consistent with $T$ in Definition 1. All results obtained in this paper hold under both definitions. We comment on this throughout the paper and refer to the alternative version as unrestricted IQ-rewritability.

3 Characteristics

We aim to provide characterizations of OMQs that are IQ-rewritable. On the one hand, these characterizations clarify which OMQs are IQ-rewritable and which are not. On the other hand, they form the basis for deciding IQ-rewritability. We first concentrate on the case of DLs (and IQs) with inverse roles and then move on to DLs without inverse roles. In the final part of this section, we consider the case where the TBox is empty, both with and without inverse roles.

3.1 The Case With Inverse Roles

To state the characterization, we need some preliminaries. Let $q(x)$ be a CQ. A cycle in $q(x)$ is a sequence of non-identical atoms $r_{0}(x_{0}, x_{1}), \ldots, r_{n-1}(x_{n-1}, x_{n})$ in $q(x)$, $n \geq 1$, where

1. $r_{0}, \ldots, r_{n-1}$ are (potentially inverse) roles,
2. $x_{i} \neq x_{j}$ for $0 \leq i < j < n$, and $x_{0} = x_{n}$.

We require the atoms be non-identical to prevent $r(x_{0}, x_{1})$, $r^{-1}(x_{1}, x_{0})$ from being a cycle (both atoms are identical).
The length of this cycle is $n$. We say that $q(x)$ is $x$-acyclic if every cycle in it passes through $x$ and use $q^{\text{acyc}}(x)$ to denote the result of restricting $q(x)$ to those atoms that only use variables reachable from $G^a_n$ from $x$. Both notions are lifted to UCQs by applying them to every CQ in the UCQ. A contraction of $q(x)$ is a CQ obtained from $q(x)$ by zero or more variable identifications, where the identification of $x$ with any other variable yields $x$.

Let $T$ be an $\mathcal{ALCHT}^u$-TBox and $q(x)$ a UCQ. We use $q^{\text{acyc}}(x)$ to denote the UCQ that consists of all $x$-acyclic CQs obtained by starting with a contraction of a CQ from $q(x)$ and then replacing zero or more atoms $r(y,z)$ with $s(y,z)$ when $T \models s \subseteq r$. We write $q^{\text{con}}_{\text{acyc}}(x)$ to denote $(q_{\text{acyc}})^{\text{con}}(x)$.

**Theorem 3.** Let $\mathcal{L} \in \{\mathcal{ALC}, \mathcal{ALCHI}\}$ and let $Q = (T, \Sigma, q(x))$ be a unary OMQ from $(\mathcal{L}, \text{UCQ})$ that is non-empty. Then the following are equivalent:

1. $Q$ is IQ-rewritable, that is, it is rewritable into an OMQ $Q'$ from $(\mathcal{L}, \text{IQ})$;
2. $Q$ is rewritable into an OMQ $Q' = (T, \Sigma, C(x))$ from $(\mathcal{L}, \text{IQ})$;
3. $Q \equiv (T, \Sigma, q^{\text{acyc}}(x))$.

When $\mathcal{L}$ is replaced with $\mathcal{L}^u$, then the same equivalences hold except that $q^{\text{acyc}}$ is replaced with $q^{\text{con}}$.

Note that Theorem 3 excludes empty OMQs, but these are trivially IQ-rewritable. It implies that, in the considered cases, it is never necessary to modify the TBox when constructing an IQ-rewriting. Further, it emerges from the proof that it is never necessary to introduce fresh role names in the rewriting (while fresh concept names are crucial). Theorem 3 also applies to unrestricted IQ-rewritability (where also ABoxes are admitted that are inconsistent with the TBox from the OMQ): unrestricted IQ-rewritability trivially implies IQ-rewritability and the converse is an easy consequence of the fact that every OMQ that is IQ-rewritable has an IQ-rewriting based on the same TBox.

We now give some ideas about the proof of Theorem 3.

The most interesting implication is “1 $\Rightarrow$ 3”. A central step is to show that if $Q = (T, \Sigma, q(x))$ is IQ-rewritable into an OMQ $Q'$, then $Q \subseteq Q_{\text{acyc}} := (T, \Sigma, q^{\text{acyc}}(x))$, that is, when $\mathcal{A} \models Q(a)$ for some $\Sigma$-ABox $\mathcal{A}$, then $\mathcal{A} \models Q_{\text{acyc}}(a)$. To this end, we first construct from $\mathcal{A}$ a $\Sigma$-ABox $\mathcal{A}^g$ of high girth (that is, without small cycles) in a way such that (a) $\mathcal{A}^g$ homomorphically maps to $\mathcal{A}$ and (b) from $\mathcal{A} \models Q(a)$ it follows that $\mathcal{A}^g \models Q'(a)$, thus $\mathcal{A}^g \models Q(a)$. Due to the high girth of $\mathcal{A}^g$ and exploiting (a variation of the tree model property for $\mathcal{ALCHT}^u$, we can then show that $\mathcal{A}^g \models Q(a)$ implies $\mathcal{A}^g \models Q_{\text{acyc}}(a)$). Because of (a), it follows that $\mathcal{A} \models Q_{\text{acyc}}(a)$. In the direction “3 $\Rightarrow$ 2”, we construct actual rewritings, based on the following lemma, an extension of a result of Kikot and Zolin [Kikot and Zolin, 2013] with TBoxes and ABox signatures (and UCQs instead of CQs).

**Lemma 4.** Let $Q = (T, \Sigma, q(x))$ be an OMQ from $(\mathcal{ALCHT}^u, \text{UCQ})$. Then

1. if $q(x)$ is $x$-acyclic and connected, then $Q$ is rewritable into an OMQ $(T, \Sigma, C(x))$ with $C(x)$ an $\mathcal{ALCI}$-IQ and
2. if $q(x)$ is $x$-acyclic, then $Q$ is rewritable into an OMQ $(T, \Sigma, C(x))$ with $C(x)$ an $\mathcal{ALCI}$-IQ.

The size of the IQs $C(x)$ is polynomial in the size of $q(x)$.

We give the construction of the $\mathcal{ALCi}$-IQ $q'(x)$ in Point 1 of Lemma 4. Let $Q = (T, \Sigma, q(x))$ be an OM from $(\mathcal{ALCHT}^u, \text{UCQ})$ with $q(x)$ $x$-acyclic and connected. To construct $q'(x)$, we first construct for each CQ $p(x)$ in $q(x)$ an $\mathcal{ELC}$-concept $C_p$, that is, an $\mathcal{ALCI}$-concept that uses only the constructors $\forall, \exists r, C$, and $\exists r \top - C$. In fact, since $p(x)$ is $x$-acyclic and connected, we can repeatedly choose and remove atoms of the form $r(x, y)$ that occur in a cycle in $p(x)$ and will eventually end up with a tree-shaped CQ $p'(x)$. Here, tree-shaped means that the undirected graph $G^a_n$ is a tree and that there are no multi-edges, that is, if $r(y, z)$ is an atom, then there is no atom $s(y, z)$ with $s \neq r$. Next, extend $p'(x)$ to obtain another tree-shaped CQ $p''(x)$ by taking a fresh concept name $P \notin \Sigma$, and adding $r(x', y)$ and $P(x')$ for each removed atom $r(x, y)$, $x'$ a fresh variable. We can now view $p''(x)$ as an $\mathcal{ELC}$-concept $C_p$ in the obvious way. The desired $\mathcal{ALCI}$-IQ $q'(x)$ is $(P \rightarrow p(x)) \bigcup C_P(x)$.

### 3.2 The Case Without Inverse Roles

We consider OMQs whose TBoxes are formulated in a DL $\mathcal{L}$ that does not admit inverse roles. Note that inverse roles are then also not admitted in the IQ used in the rewriting. We first observe that this has less impact than one might expect: inverse roles in the IQ-rewriting can be eliminated and in fact Points 1 and 3 from Theorem 3 are still equivalent. However, there is also a crucial difference: unless the universal role is present, the elimination of inverse roles requires an extension of the TBox and thus the equivalence of Points 1 and 2 of Theorem 3 fails. In fact, this is illustrated by Point (2) of Example 2. We thus additionally characterize IQ-rewritability without modifying the TBox. We also show that, with the universal role, it is not necessary to extend the TBox.

We start with some preliminaries. An extended conjunctive query (eCQ) is a CQ that also admits atoms of the form $C(x)$, $C$ a (potentially compound) concept, and $\exists r \top$ and $\exists r \top - C$ extend ABoxes (eABoxes) are defined analogously. The semantics is defined in the expected way. Every eCQ $q(x)$ gives rise to an eABox $A_q$ by viewing the variables in $q(x)$ as individual names and the atoms as assertions.

Let $q(x)$ be an eCQ. We use $\text{dreach}(q)$ to denote the set of all variables reachable from $x$ in the directed graph $G_q$ and say that $q(x)$ is $x$-accessible if $\text{dreach}(q)$ contains all variables. For $V$ a set of variables from $q(x)$ that includes $x$, $q(x)|_V$ denotes the restriction of $q(x)$ to the atoms that use only variables from $V$.

Let $T$ be an $\mathcal{ALC}$-TBox. An eCQ $p(x)$ is a $T$-decoration of a CQ $q(x)$ if

1. $p(x)$ is obtained from $q(x)$ by adding, for each $y \in \text{dreach}(q)$ and each subconcept $C$ of $T$, the atom $C(y)$ or the atom $\neg C(y)$;
2. the eABox $A_q$ is consistent with $T$.

---

Note that $x$ is the answer variable and recall that we might have $r = s^-$ and thus also choose atoms $s(y, x)$. 

---
For a UCQ $q(x)$, we use $q^{\text{deco}}(x)$ to denote the UCQ that consists of all eCQs $p(x)$, where $p(x)$ is a $T$-decoration of a CQ from $q(x)$. We write $q^{\text{acyc}}(x)$ to denote $(q^{\text{acyc}})^{\text{deco}}(x)$. We now give the results announced above.

**Theorem 5.** Let $L \in \{\text{ALC, ALCI}\}$ and let $Q = (T, \Sigma, q(x))$ be a unary OMQ from $(L, UCQ)$ that is non-empty. Then the following are equivalent:

1. $Q$ is rewritable into an OMQ from $(L, IQ)$;
2. $Q$ is rewritable into an OMQ $(T \cup T', \Sigma, C(x))$ from $(L, IQ)$;
3. $Q$ is rewritable into an OMQ from $(L, IQ)$;
4. $Q$ is rewritable into an OMQ $Q' = (T, \Sigma_{\text{full}}, C(x))$ from $(L, IQ)$;
5. $Q \equiv (T, \Sigma_{\text{full}}, q^{\text{acyc}}(x))$.

If, furthermore, $L$ is replaced with $L'$ and $L' \cap L' = L'$, then Conditions 1 to 3 are further equivalent to:

6. $Q$ is rewritable into an OMQ $Q' = (T, \Sigma, C(x))$ from $(L', IQ)$.

Characterizing IQ-rewritability in the case where $L \in \{\text{ALC, ALCI}\}$, the TBox (is non-empty and) cannot be extended, and $\Sigma \neq \Sigma_{\text{full}}$ remains an open problem.

In the directions “$3 \Rightarrow 2$,” “$5 \Rightarrow 4$,” and “$3 \Rightarrow 6$”, we have to construct IQ-rewritings. This is done by starting with the rewriting from the proof of Lemma 4 and then modifying it appropriately. As in the case of Theorem 5 it is straightforward to see that all results stated in Theorem 5 also apply to unrestricted IQ-rewritability.

### 3.3 The Case of Empty TBoxes

We consider OMQs in which the TBox is empty as an important special case. Since it is then not interesting to have an ABox signature, this corresponds to the rewritability of (U)CQs into $L$-instance queries, for some concept language $L$ (and thus no OMQs are involved). The importance of this case is due to the fact that it provides an ‘underapproximation’ of the IQ-rewritability of OMQs, while also being easier to characterize and computationally simpler.

We say that an UCQ $q(x)$ is $L$-IQ-rewritable if there is an $L$-IQ $q'(x)$ that is equivalent to $q(x)$ in the sense that the OMQs $(\emptyset, \Sigma_{\text{full}}, q(x))$ and $(\emptyset, \Sigma_{\text{full}}, q'(x))$ are equivalent (and in passing, we define the equivalence between two UCQs in exactly the same way). The following proposition makes precise what we mean by underapproximation.

**Proposition 6.** Let $L \in \{\text{ALC, ALCI, ALC}^\Box, \text{ALCI}^\Box\}$. If a UCQ $q(x)$ is $L$-IQ-rewritable, then so is any OMQ $(T, \Sigma, q(x))$ from $(L', UCQ)$.

Proposition 6 is essentially a corollary of Theorem 7 below. As illustrated by Case (3) of Example 2, its converse fails.

We now characterize IQ-rewritability in the case of the empty TBox. A subquery of a CQ $q(x)$ is a CQ $q'(x)$ obtained from $q(x)$ by dropping atoms. A subquery of a UCQ $q(x)$ is a UCQ obtained by including as a CQ at most one subquery of each CQ in $q(x)$.

**Theorem 7.** Let $q(x)$ be a UCQ. Then

1. $q(x)$ is rewritable into an $\text{ALCI-IQ}$ iff there is a subquery $q'(x)$ of $q(x)$ that is $x$-acyclic, connected, and equivalent to $q(x)$;
2. $q(x)$ is rewritable into an $\text{ALCI-IQ}$ iff there is a subquery $q'(x)$ of $q(x)$ that is $x$-acyclic, $x$-accessible, and equivalent to $q(x)$.

When $L$-IQs are replaced with $L^u$-IQs, then the same equivalences hold except that connectedness/$x$-accessibility is dropped.

Note that Theorem 7 also characterizes rewritability of CQs; the query $q'(x)$ is then also a CQ rather than a UCQ. This is in contrast to Theorems 3 and 5 where the queries $q^{\text{acyc}}(x)$ and $q^{\text{deco}}(x)$ are UCQs even when the query $q(x)$ from the OMQ that we start with is a CQ. Another crucial difference is that $q^{\text{acyc}}(x)$ and $q^{\text{deco}}(x)$ can be of size exponential in the size of the original OMQ while the query $q'(x)$ in Theorem 7 is of size polynomial in the size of $q(x)$.

### 4 Complexity

We determine the complexity of deciding IQ-rewritability in various OMQ languages, based on the established characterizations and starting with the case of empty TBoxes.

**Theorem 8.** For every $Q \in \{\text{CQ, UCQ}\}$ and $L \in \{\text{ALC, ALCI, ALC}^\Box, \text{ALCI}^\Box\}$, it is NP-complete to determine whether a given query from $Q$ is $L$-IQ-rewritable.

The upper bound in Theorem 8 is by guessing the query $q'(x)$ from Theorem 7 and verifying that it satisfies the properties stated there. The lower bound is by a reduction from 3-colorability.

We next consider the case where TBoxes can be non-empty, starting with the assumption that the ABox signature is full since this results in (slightly) lower complexity.

**Theorem 9.** Let $Q \in \{\text{CQ, UCQ}\}$. For OMQs based on the full ABox signature, IQ-rewritability is

1. EXPTime-hard in $(\text{ALC}, Q)$ and in coNEXPTime in $(\text{ALCH}, Q)$ and
2. 2EXPTime-complete in $(\text{ALCI}, Q)$ and $(\text{ALCHI}, Q)$.

The lower bounds are by reduction from OMQ evaluation on ABoxes of the form $A(a)$, $A$ a concept name, which is EXPTime-complete in $(\text{ALCH}, \text{CQ})$ and 2EXPTime-complete in $(\text{ALCHI}, \text{CQ})$ [Lutz, 2008]. The upper bounds are derived from the OMQ containment checks suggested by Condition 3 of Theorem 4 and Condition 4 of Theorem 5. Since we work with the full ABox signature, the non-emptyness condition from these theorems is void (there are no empty OMQs) and OMQ containment is closely related to OMQ evaluation, which allows us to derive upper bounds for the former from the latter; in fact, these bounds are exactly the ones stated in Theorem 4. We have to exercise some care, for two reasons: first, we admit UCQs as the actual query and thus the trivial reduction of OMQ containment to OMQ evaluation that is possible for CQs (which can be viewed as an ABox) does not apply. And second, we aim for upper bounds that exactly match the complexity of OMQ containment while
the UCQs \( q_{\text{con}}(x) \) and \( q_{\text{dec}}(x) \) involved in the containment checks are of exponential size. What rescues us is that each of the CQs in these UCQs is only of polynomial size.

We finally consider the case where the ABox signature is unrestricted.

**Theorem 10.** IQ-rewritability is

1. \( \text{NEXP} \text{TIME-hard in } (\mathcal{ALC}, \text{CQ}) \) and

2. \( 2\text{NEXP} \text{TIME-complete in all of } (\mathcal{ALCF}, \text{CQ}), (\mathcal{ALC} \text{I}, \text{CQ}), (\mathcal{ALCH} \text{I}, \text{CQ}), (\mathcal{ALCH} \text{I}, \text{UCQ}). \)

The lower bound in Point 1 is by reduction from OMQ emptiness in \((\mathcal{ALC}, \text{CQ})\), which is \( \text{NEXP} \text{TIME-complete} \) \cite{Baader2016}. For the one in Point 2, we use a reduction from OMQ containment, which is \( 2\text{NEXP} \text{TIME-complete} \) in \((\mathcal{ALC}, \text{CQ})\) \cite{Bourhis2016}. The upper bounds are obtained by appropriate containment checks as suggested by our characterizations, and we again have to deal with UCQs with exponentially many CQs. Note that Theorem 10 leaves open the complexity of IQ-rewritability in \((\mathcal{ALC}, \text{CQ})\), between \( \text{NEXP} \text{TIME} \) and \( 2\text{NEXP} \text{TIME}. \) The same gap exists for OMQ containment \cite{Bourhis2016} as well as in the related problems of FO-rewritability and Datalog-rewritability \cite{Feier2017}.

**5 Functional Roles**

We consider DLs with functional roles. A fundamental observation is that for the basic such DL \( \mathcal{ALCF} \) IQ-rewritability is undecidable. This can be proved by a reduction from OMQ emptiness in \((\mathcal{ALC} \text{I}, \text{CQ})\) \cite{Baader2016}.

**Theorem 11.** In \((\mathcal{ALCF}, \text{CQ})\), IQ-rewritability is undecidable.

In the following, we show that decidability is regained in the case where the TBox is empty (apart from functionality assertions). This is challenging because functionality assertions have a strong and subtle impact on rewritability. As before, the only interesting ABox signature to be combined with ‘empty’ TBoxes is the full ABox signature. We use \( \mathcal{F} \) to denote the TBox language in which TBoxes are sets of functionality assertions and concentrate on rewriting into IQs that may use inverse roles.

**Example 12.** Consider the CQ \( p(x) = \exists y (s(x,y) \land r(y,y)) \) from Point 1 of Example 2. Then \( Q_s = (T_s, \Sigma_{\text{full}}, p(x)) \) and \( Q_r = (T_r, \Sigma_{\text{full}}, p(x)) \) with \( T_s = \{ \text{func}(w) \} \) for \( w \in \{ r, s \} \) are both rewritable into an OMQ \((T_u, \Sigma_{\text{full}}, q_u(x))\) with \( q_u(x) \) an \( \mathcal{ALCF} \)-IQ. The rewritings are neither trivial to find nor entirely easy to understand. In fact, for \( q_r(x) \) we can use \( \forall s. P \rightarrow \exists s. (P \rightarrow \exists r. P) \). For \( q_s(x) \), we introduce three fresh concept names rather than a single one and use them in a way inspired by graph colorings:

\[
q_r(x) = (\forall s. \bigcup_{1 \leq i \leq 3} P_i) \rightarrow (\exists s. (\bigcap_{1 \leq i \leq 3} (P_i \rightarrow \exists r. P_i))).
\]

Before giving a characterization of rewritable queries, we introduce some preliminaries. Let \( q(x) \) be a CQ and \( T \) an \( \mathcal{ALCF} \)-TBox. A sequence \( x_0, \ldots, x_n \) of variables in \( q(x) \) is a functional path in \( q(x) \) from \( x_0 \) to \( x_n \) w.r.t. \( T \) if for all \( i < n \) there is a role \( r \) such that \( \text{func}(r) \in T \) and \( r(x_i, x_{i+1}) \) is in \( q(x) \). We say that \( q(x) \) is \( f \)-acyclic w.r.t. \( T \) if for every cycle \( r_0(x_0, x_1), \ldots, r_{n-1}(x_{n-1}, x_n) \) in \( q(x) \), one of the following holds:

- there is a functional path in \( q(x) \) from some \( x_i \);
- \( \text{func}(r_i) \in T \) or \( \text{func}(r_i^w) \in T \) for all \( i < n \) and there is a functional path \( y_0, \ldots, y_m \) in \( q(x) \) with \( x_0 = y_0 = y_m \) such that \( \{x_0, \ldots, x_{n-1}\} \subseteq \{y_0, \ldots, y_m\} \).

We are now ready to state the characterization.

**Theorem 13.** An OMQ \( Q = (T, \Sigma_{\text{full}}, q(x)) \) from \((\mathcal{F}, \text{UCQ})\) is rewritable into an OMQ from \((\mathcal{F}, \mathcal{ALCI}, \text{IQ})\) iff there is a subquery \( q'(x) \) of \( q(x) \) that is \( f \)-acyclic, connected, and equivalent to \( q(x) \). When \( \mathcal{ALCI} \)-IQ is replaced with \( \mathcal{ALCF} \)-IQ, the same equivalence holds except that connectedness is dropped.

The proof of Theorem 13 extends the ultrafilter construction from \cite{Kikot2013}. We remark that the “if” direction in Theorem 13 even holds for OMQs \( Q = (T, \Sigma, q(x)) \) from \((\mathcal{ALCF}, \text{UCQ})\). Thus, the case of the ‘empty’ TBox can again be seen as an underapproximation of the general case. We further remark that \( T \) remains unchanged in the construction of the IQ-rewritings and that the constructed rewritings are of polynomial size.

**Theorem 14.** For OMQs from \((\mathcal{F}, \text{UCQ})\), rewritability into \((\mathcal{F}, \mathcal{ALCI}-\text{IQ})\) is \( \text{NP-complete} \).

**6 MMSNP and CSP**

Recall from the introduction that the OMQ languages studied in this paper are closely related to CSPs and their logical generalization MMSNP. In fact, the techniques used to establish the results in Sections 3 and 4 can be adapted to determine the complexity of deciding whether a given MMSNP sentence is equivalent to a CSP. In a nutshell, we prove that an MMSNP-sentence is equivalent to a CSP iff it is preserved under disjoint union and equivalent to a generalized CSP (a CSP with multiple templates), and that both properties can be reduced to containment between MMSNP sentences which is \( 2\text{NEXP} \text{TIME-complete} \) \cite{Bourhis2016}. The latter reduction involves constructing an MMSNP sentence \( \varphi_{\text{acyc}} \) that is reminiscent of the query \( q_{\text{acyc}} \) in Theorem 3. Full details are given in the appendix.

**Theorem 15.** It is \( 2\text{NEXP} \text{TIME-complete} \) to decide whether a given MMSNP-sentence is equivalent to a CSP.

**7 Conclusion**

We have made a leap forward in understanding the relation between (U)CQs and IQs in ontology-mediated querying. Interesting open problems include a characterization of IQ-rewritability for DLs with functional roles when the TBox is non-empty and characterizations for DLs with transitive roles. The remarks after Theorem 4 and 10 mention further problems left open. In addition, it would be worthwhile to continue the effort from \cite{Kikot2013} to understand the value of IQ-rewritings for the purposes of efficient practical implementation.

**Acknowledgements.** Cristina Feier and Carsten Lutz were supported by ERC Consolidator Grant 647289 CODA. Frank Wolter was supported by EPSRC UK grant EP/M012646/1.
References


Appendices

A Some Technical Preliminaries

Every interpretation can be viewed as an undirected graph $G^q_2$, analogously to the definition of the undirected graph $G^n_q$ of a CQ $q$. The universal role does not give rise to edges in $G^q_2$.

An interpretation is tree-shaped or a tree interpretation if $G^q_2$ is a tree and there are no multi-edges, that is, $(d,e) \in r^q$ implies $(d,e) \notin s^q$ for all (potentially inverse) roles $s \neq r$. Let $\mathcal{T}$ be an ALCHIT$^+$-TBox and $\mathcal{A}$ an ABox. An interpretation $I$ is a forest model of $\mathcal{A}$ if there are tree interpretations $(I_a)_{a \in \text{ind}(\mathcal{A}) \cup D}$ where $D$ is a (potentially empty) set of individuals, with mutually disjoint domains, and

$$\Delta^I \cap \text{ind}(\mathcal{A}) = \begin{cases} \{a\}, & \text{if } a \in \text{ind}(\mathcal{A}) \\ \emptyset, & \text{if } a \in D, \end{cases}$$

such that $I$ is the (non-disjoint) union of $\mathcal{I}_a$ and $(\mathcal{I}_a)_{a \in \text{ind}(\mathcal{A}) \cup D}$ where $\mathcal{I}_a$ is $\mathcal{A}$ viewed as an interpretation.

An extended forest model $I$ of $\mathcal{A}$ and $\mathcal{T}$ is a model of $\mathcal{A}$ and $\mathcal{T}$ that can be obtained from a forest model $J$ of $\mathcal{A}$ by closing under role inclusions from $\mathcal{T}$, that is, adding $(d,e) \in r^q$ when $(d,e) \in s^q$ and $\mathcal{T} \models s \subseteq r \in \mathcal{T}$. We also say that $J$ underlies $I$.

Lemmas of the following kind have been widely used in the literature on ontology-mediated querying. The proof of the “if” direction uses a standard unraveling argument and is omitted, see for example [Lutz, 2008].

Lemma 16. Let $Q = (\mathcal{T}, \Sigma, q(x))$ be an OMG from (ALCHIT$^+$, UCQ), $\mathcal{A}$ a $\Sigma$-ABox, and $a \subseteq \text{ind}(\mathcal{A})$. Then $\mathcal{A} \models Q(a)$ iff for all extended forest models $I$ of $\mathcal{A}$ and $\mathcal{T}$, $I \models Q(a)$.

We introduce some more helping lemmas. An ABox $\mathcal{A}$ can be seen as a directed graph $G_\mathcal{A}$ and as an undirected graph $G^\mathcal{A}_q$ in the expected way, analogously to the definition of $G_q$ and $G^n_q$ for a CQ $q$. For an ABox $\mathcal{A}$ and $a \in \text{ind}(\mathcal{A})$, we use $\mathcal{A}^\text{con}_a$ to denote the restriction of $\mathcal{A}$ to the individuals reachable in $G^\mathcal{A}_q$ from $a$. We also denote with $\text{CON}_\mathcal{A}$ the set of ABoxes induced by the maximal connected components of $G^\mathcal{A}_q$.

Lemma 17. Let $Q = (\mathcal{T}, \Sigma, q(x))$ be an OMG from (ALCHIT, IQ). Then $\mathcal{A} \models Q(a)$ implies $\mathcal{A}^\text{con}_a \models Q(a)$.

A homomorphism from an ABox $\mathcal{A}$ to an ABox $\mathcal{B}$ is a function $h : \text{ind}(\mathcal{A}) \rightarrow \text{ind}(\mathcal{B})$ such that $A(a) \in \mathcal{A}$ implies $h(A(a)) \in \mathcal{B}$ and $r(a,b) \in \mathcal{A}$ implies $r(h(a), h(b)) \in \mathcal{B}$. We write $\mathcal{A} \rightarrow \mathcal{B}$ to indicate that there is a homomorphism from $\mathcal{A}$ to $\mathcal{B}$. For $a \in \text{ind}(\mathcal{A})$ and $b \in \text{ind}(\mathcal{B})$, we further write $\mathcal{A}(a) \rightarrow (B, b)$ to indicate that there is a homomorphism from $\mathcal{A}$ to $\mathcal{B}$ with $h(a) = b$. The following lemma is well-known, see for example [Bienvenu et al., 2014].

Lemma 18. Let $Q = (\mathcal{T}, \Sigma, q)$ be a unary OMG from (ALCHIT$^+$, UCQ), with $Q \in (UCQ, IQ)$. $\mathcal{A}$ and $\mathcal{B}$ be $\Sigma$-ABoxes, $a \in \text{ind}(\mathcal{A})$, and $b \in \text{ind}(\mathcal{B})$. Then $\mathcal{A}(a) \rightarrow (B, b)$ and $\mathcal{A} \models Q(a)$ implies $\mathcal{B} \models Q(b)$.

B Proofs for Section 3

We start with introducing several lemmas concerned with certain constructions on ABoxes. These lemmas are closely related to the connection between ontology-mediated querying and constraint satisfaction problems (CSPs), see for example [Bienvenu et al., 2014, Lutz and Wolter, 2017].

Note that in assertions $r(x, y)$ in an ABox, $r$ must be a role name but cannot be an inverse role. For purposes of uniformity, we use $r^-(x, y)$ as an alternative notation to denote an assertion $r(y, x)$ in an ABox. A cycle in an ABox is defined exactly like a cycle in a CQ, repeated here for convenience. A cycle in an ABox $\mathcal{A}$ is a sequence of non-identical assertions $r_0(a_0, a_1), \ldots, r_n(a_{n-1}, a_n)$ in $\mathcal{A}$, $n \geq 1$, where

1. $r_0, \ldots, r_{n-1}$ are (potentially inverse) roles,
2. $a_i \neq a_j$ for $0 \leq i < j < n$, and $a_0 = a_n$.

The length of this cycle is $n$. The girth of $\mathcal{A}$ is the length of the shortest cycle in it and $\infty$ if $\mathcal{A}$ has no cycle.

The following is a DL formulation of what is often known as the sparse incompatibility lemma in CSP [Feder and Vardi, 1998].

Lemma 19. For every ABox $\mathcal{A}$ and all $g, s \geq 0$, there is an ABox $\mathcal{A}^g$ of girth exceeding $g$ such that

1. $\mathcal{A}^g \rightarrow \mathcal{A}$ and
2. for every ABox $\mathcal{B}$ with $|\text{ind}(\mathcal{B})| \leq s$, $\mathcal{A} \rightarrow \mathcal{B}$ if $\mathcal{A}^g \rightarrow \mathcal{B}$.

We next establish a ‘pointed’ version of Lemma 19 that is crucial for the subsequent proofs. The $a$-girth of $\mathcal{A}$ is defined exactly like the girth except that we only consider cycles that do not pass through $a$.

Lemma 20. For all ABoxes $\mathcal{A}$, $a \in \text{ind}(\mathcal{A})$, and $g, s \geq 0$, there is an ABox $\mathcal{A}^g$ of $a$-girth exceeding $g$ such that

1. $\mathcal{A}^g \rightarrow (A, a)$
2. for every ABox $\mathcal{B}$ with $|\text{ind}(\mathcal{B})| \leq s$ and every $b \in \text{ind}(\mathcal{B})$, $\mathcal{A}^g, a \rightarrow (B, b)$ if $\mathcal{A}, a \rightarrow (B, b)$.

Proof. Let $\mathcal{A}$ be an ABox, $a \in \text{ind}(\mathcal{A})$, and $g, s \geq 0$. Further, let $\mathcal{A}^g_{a}$ be the ABox obtained from $\mathcal{A}$ by adding the assertion $P(a)$, $P$ a fresh concept name, let $\mathcal{A}^g_{a}$ the ABox obtained from $\mathcal{A}^g_a$, by applying Lemma 19 for $g$ and $s$, and let $h$ be a homomorphism from $\mathcal{A}^g_a$ to $\mathcal{A}^g_{a'}$. Assume w.l.o.g. that the individual name $a$ does not occur in $\mathcal{A}^g_a$. We use $\mathcal{A}^g$ to denote the ABox obtained from $\mathcal{A}^g_a$ by dropping all facts of the form $P(b)$ and identifying all individual names $b$ with $h(b) = a$, replacing them with $a$. We show that $\mathcal{A}^g$ is as required:

(a) $\mathcal{A}^g$ has $a$-girth higher than $g$.

Every cycle in $\mathcal{A}^g$ that does not pass through $a$ is also in $\mathcal{A}^g_{a'}$, thus is of length exceeding $g$.

(b) Point 1 of Lemma 20 is satisfied.

Let $h' : \text{ind}(\mathcal{A}^g_a) \rightarrow \text{ind}(\mathcal{A})$ be such that $h'(a) = a$ and $h'(b) = h(b)$ if $a \neq b$. It can be verified that $h'$ is a homomorphism from $\mathcal{A}^g_a$ to $\mathcal{A}$. It clearly witnesses $(\mathcal{A}^g, a) \rightarrow (\mathcal{A}, a)$, as required.

(c) Point 2 of Lemma 20 is satisfied.

Let $\mathcal{B}$ be an ABox with $|\text{ind}(\mathcal{B})| \leq s$ and $b \in \text{ind}(\mathcal{B})$. We have to show that $\mathcal{A}, a \rightarrow (B, b)$ if $\mathcal{A}^g, a \rightarrow (B, b)$. 


For Point 2, let \( Q = (T, \Sigma, q(x)) \) be an OMQ from \((\text{ALCHIT}^n, \text{UCQ})\) where \( q(x) \) is x-acyclic. Let \( p_0(x), p_1(), \ldots, p_m() \) be the maximal connected components of \( q(x) \). Note that \( p_0(x) \) is x-acyclic and each \( p_i() \) is acyclic in the sense that it contains no cycles at all. We can view \( p_0(x) \) as an \( \text{ELCI} \)-concept and each \( p_i() \) as an \( \text{ELCI}^n \)-concept \( C_p \). For the form \( u_i.C \) with \( u \) the universal role and \( C \) an \( \text{ELCI} \)-concept. Let \( C_p = C_{p,0} \sqcap C_{p,1} \sqcap \ldots \sqcap C_{p,m} \), and \( Q' = (T, \Sigma, (P \rightarrow p(x) \sqcup_{q(x)} C_p(x))) \). One can show that \( Q = Q' \).

Example 22. Let \( Q \) be an OMQ \((T, \Sigma, q(x))\) with \( T = \emptyset \). \( \Sigma = \{ r, s, t, v \} \) and \( q(x) = \exists y_1 \exists y_2 \exists y_3 \ r(x, y_1) \land s(x, y_2) \land t(y_2, y_1) \land r(y_2, y_3) \). It is easy to see that \( q(x) \) is x-acyclic and connected. Towards obtaining an \( \text{ALCHIT} \)-rewriting, we construct a tree-shaped \( \text{CQ} \) \( p''(x) \) from \( q(x) \) by first removing the atom \( s^- (y_2, x) \) and then adding atoms \( s^- (y_2, x') \) and \( P(x') \), with \( x' \) a fresh variable: \( p''(x) = \exists y_1 \exists y_2 \exists y_3 \ P(x') \land r(x', y_1) \land s' (y_2, x') \land t(y_2, y_1) \land r(y_2, y_3) \). The concept \( C_p \) corresponding to \( p''(x) \) is \( \exists r, \exists s' \cdot \forall r, t \cdot \exists s^- \cdot P \cdot \), and thus the desired rewriting is the OMQ \( Q' = (T, \Sigma, q'(x)) \) with \( q'(x) \) the \( \text{ALCHIT} \)-Q:\( (P \rightarrow C_p(x)) \).

Theorem 3. Let \( L \in \{\text{ACI}, \text{ALCHIT} \} \) and let \( Q = (T, \Sigma, q(x)) \) be an OMQ from \((L, \text{UCQ})\) that is non-empty. Then the following are equivalent:

1. \( Q \) is IQ-rewritable, that is, it is rewritable into an OMQ \( Q' = (T', \Sigma, C(x)) \) from \((L, \text{IQ})\).
2. \( Q \) is rewritable into an OMQ \( Q' = (T, \Sigma, C(x)) \) from \((L, \text{IQ})\).
3. \( Q = (T, \Sigma, q_{\text{acyc}}^\text{con}(x)) \).

When \( L \) is replaced with \( L^n \), then the same equivalences hold except that \( q_{\text{acyc}}^\text{con} \) is replaced with \( q_{\text{acyc}} \).

Proof. The implication “\( 2 \Rightarrow 1 \)” is trivial. For “\( 3 \Rightarrow 2 \)” assume that \( Q = (T, \Sigma, q_{\text{acyc}}^\text{con}(x)) \). Since \( q_{\text{acyc}}^\text{con} \) is connected and x-acyclic, we can apply Lemma [6].

For “\( 1 \Rightarrow 3 \)”, we show that whenever an OMQ \( Q \) from \((L, \text{UCQ})\) is IQ-rewritable, then (a) \( Q = Q_{\text{acyc}} \) where \( Q_{\text{acyc}} = (T, \Sigma, q_{\text{acyc}}(x)) \) and (b) \( Q = Q_{\text{acyc}} \) where \( Q_{\text{acyc}} := (T, \Sigma, q_{\text{acyc}}^\text{con}(x)) \). This yields \( Q = (T, \Sigma, q_{\text{acyc}}(x)) \) as desired: if \( Q \) is IQ-rewritable, then (a) yields \( Q = Q_{\text{acyc}} \), thus \( Q_{\text{acyc}} \) is IQ-rewritable and we can apply (b).

Thus, let \( Q \) from \((L, \text{UCQ})\) be IQ-rewritable. Thus there is an OMQ \( Q' = (T', \Sigma, C(x)) \) from \((L, \text{IQ})\) that is equivalent to \( Q \). By Lemma [21] one can find a \( \Sigma \)-ABox \( B \) and \( b_1, \ldots, b_k \in \text{ind}(B) \) such that for every \( \Sigma \)-ABox \( A \) and \( a \in \text{ind}(A) \).

1. \( A \models Q'(a) \iff (A, a) \not\models (B, b_i) \) for \( 1 \leq i \leq k \);
2. \( A \) is consistent with \( T \cup \exists r \cdot \exists s^- \cdot P \).

We show Points (a) and (b) from above.

(a) We have \( Q_{\text{acyc}} \subseteq Q \) by definition of \( Q_{\text{acyc}} \), no matter whether \( Q \) is IQ-rewritable or not, and thus it remains to show that \( Q \subseteq Q_{\text{acyc}} \). If all \( \text{CQs} \) in \( q(x) \) are x-acyclic, the result clearly holds. In the following we assume that at least one \( \text{CQ} \) in \( q(x) \) is not x-acyclic.

The following lemma is a straightforward variation of similar lemmas from [Bienvenu et al., 2014]. The constructed ABoxes are called CSP in [Bienvenu et al., 2014].

**Lemma 4.** Let \( A \) be a \( \Sigma \)-ABox with \( A \models Q(x) \). Then, for every model \( I \) of \( \Sigma \) and every \( a \in \text{ind}(A) \),

1. \( A \models Q(x) \iff (A, a) \not\models (B, b) \) for all \( (B, b) \in I \); and
2. \( A \) is consistent with \( T \) iff \( A \models B \) for some \( (B, b) \in I \).

When \( Q \) is from \((\text{ALCHIT}, \text{IQ})\), then \( \Gamma \) can be chosen so that all ABoxes in it are identical.

An \( \text{ELCI}^n \)-concept is an \( \text{ACI} \)-concept that uses only the constructors \( \land, \exists r.C, \exists r^- C \), and \( \exists u.C \) where \( u \) is the universal role.

**Lemma 21.** Let \( Q = (T, \Sigma, q(x)) \) be an OMQ from \((\text{ALCHIT}^n, \text{UCQ})\). Then

1. if \( q(x) \) is x-acyclic and connected, then \( Q \) is rewritable into an OMQ \((T, \Sigma, C(x))\) with \( C(x) \) an \( \text{ACI} \)-Q:\( and \( \Sigma \) an \( \text{ACI}^n \)-Q:\(.

The size of the \( \text{CQs} \) \( C(x) \) is polynomial in the size of \( q(x) \).

**Proof.** For Point 1, let \( Q = (T, \Sigma, q(x)) \) be an OMQ from \((\text{ALCHIT}^n, \text{UCQ})\) where \( q(x) \) is x-acyclic and connected. Further, let \( Q' = (T, \Sigma, (P \rightarrow p(x) \sqcup_{q(x)} C_p(x))) \) be as constructed in the main part of the paper. We have to show the following:

1. \( "Q \subseteq Q'": \) Let \( A \) be a \( \Sigma \)-ABox with \( A \models Q(x) \). Then, for every model \( I \) of \( \Sigma \) and \( A, I \models p(x) \) for some \( \text{CQ} \) \( p(x) \) in \( \text{q(x)} \), and thus \( I \models \exists r A \triangleright p(x) \) or \( I \models \exists r A \triangleright p(x) \triangleright p(x) \). As \( A \triangleright p(x) \rightarrow A \triangleright p(x) \cup \{p(x)\} \), the latter is the same as \( I \models \triangleright p'(x) \) or \( I \models \triangleright C_p(x) \). Thus \( I \models (P \rightarrow C_p(x)) \), or \( I \models Q'(x) \).

2. \( "Q' \subseteq Q": \) Let \( A \) be a \( \Sigma \)-ABox with \( A \models Q'(x) \). Then, for every model \( I \) of \( \Sigma \) and \( A, I \models \triangleright p(x) \) for some \( \text{CQ} \) \( p(x) \) in \( \text{q(x)} \), from the construction of \( C_p(x) \), \( \triangleright p'(x) \), and from the fact that \( P \) is interpreted as a singleton in \( T \), \( I \models p(x) \), or \( I \models \triangleright p(x) \cup_{q(x)} p(x) \). As \( P \) is fresh, and in particular does not occur in \( q(x) \), \( I \) and \( T \) might differ only w.r.t. the interpretation of \( P: I \models \triangleright p(x) \cup_{q(x)} p(x) \). Thus, \( A \models Q(x) \).
Let $A$ be a $\Sigma$-ABox with $A \models Q(a)$. Thus, $(A, a) \not\models (B, b_i)$ for $1 \leq i \leq k$. We apply Lemma 20 with $g$ the maximum between 2 and the girths of CQs from $q(x)$ which are not $x$-acyclic, obtaining a $\Sigma$-ABox $A'$ of $g$-girth exceeding $g$ such that $(A', a) \rightarrow (A, a)$ and $(A', a) \not\models (B, b_i)$ for $1 \leq i \leq k$. The latter yields $A' \models Q(a)$. We aim to show that $A' \models Q_{acyc}(a)$. Since $(A', a) \rightarrow (A, a)$, it follows by Lemma 18 that $A \models Q_{acyc}(a)$, as desired.

By Lemma 16 it suffices to show that for every extended forest model $I$ of $A'$ and $\mathcal{T}$, we have $I \models q_{acyc}(a)$. Thus let $I$ be such a model. Since $A' \models Q(a)$, we have $I \models q(a)$ and thus there is a CQ $p(x)$ in $q(a)$ such that $I \models p(a)$. Consequently, there is a homomorphism $h$ from $p(x)$ to $I$ with $h(a) = a$. Let $p'(x)$ be the contraction of $p(x)$ obtained by identifying all variables $y_1$ and $y_2$ such that $h(y_1) = h(y_2)$. As witnessed by $h$, $I \models p'(a)$. Note that the $x$-girth of $p'(x)$ is either infinite or it is bounded from above by $g$ since the $x$-girth of $p(x)$ is. Also note that $h$ is an injective homomorphism from $p'(x)$ to $I$. By definition of extended forest models, all cycles in $I$ are either cycles from $A'$, or they are cycles of the form $r(y,z), s(z,y)$. This together with the fact that the girth of $A'$ exceeds $g$ implies that every cycle in $p'(x)$ passes through $x$ or is of the latter kind. In fact, $p'(x)$ is $x$-acyclic when $T$ contains no role inclusions since then $I$ is a forest model of $A'$. Since $p'(x)$ is a CQ in $q_{acyc}(x)$, we are done in that case.

Now for the case when $T$ contains role inclusions. Let $J$ be the forest model of $A'$ underlying $I$. Construct a CQ $p''(x)$ from $p'(x)$ as follows: for all distinct variables $y, z$, with $y \neq x$ and $z \neq x$, whenever $r_1(y,z)$, $r_2(y,z)$, $s_1(z,y)$, $s_2(z,y)$ are all atoms of this form in $p'(x)$, then replace them with $r(x,y)$ if $(h(x), h(y)) \in r''$ and with $r(y,x)$ if $(h(y), h(x)) \in r''$. Note that by definition of extended forest models and due to the fact that $g$, the girth of $A'$, is greater than 2, such an $r$ always exists. As witnessed by $h$, $I \models p''(a)$. Moreover, $p''(x)$ is $x$-acyclic and a CQ in $q_{acyc}(x)$, thus we are again done. 

(b) It is immediate by definition of $Q^\con$ that $Q \subseteq Q^\con$. We thus have to show that $Q^\con \subseteq Q$. Assume the contrary. Then, there is a $\Sigma$-ABox $A$ and an $a \in \text{Ind}(A)$ such that $A \models Q^\con(a)$ and $A \not\models Q(a)$. Note that $A$ must be consistent with $T$. Since $Q$ is non-empty, there is a $\Sigma$-ABox $A_Q$ such that $A_Q \models Q(b)$ for some $b \in \text{Ind}(A_Q)$. Let $A'$ be the disjoint union of $A$ and $A_Q$. We get $A' \models Q(a)$ from Lemma 18 and thus $A' \models Q'(a)$. Since $Q'$ is from $\{\text{ALCI}, \text{IQ}\}$ and $A'^{\con} = A$, the latter and Lemma 17 implies $A \models Q'(a)$, thus $A \models Q(a)$, a contradiction.

When the OMQ language $L$ is replaced by $L^u$, we can show that $Q \equiv Q_{acyc}$ exactly as above. The second part of the proof showing that $Q \equiv Q_{acyc}$ (does not go through and) is no longer needed.

Lemma 23. Let $Q = (T, \Sigma, q(x))$ be an OMQ from $\{\text{ALCH}, \text{UCQ}\}$ such that $q(x)$ is $x$-acyclic and connected. Then $Q$ is rewritable into an OMQ $(T \cup T', \Sigma, q(x))$ from $\{\text{ALCH}, \text{IQ}\}$ whose size is polynomial in the size of $Q$.

Proof. Let $Q = (T, \Sigma, q(x))$ be an OMQ from $\{\text{ALCH}, \text{UCQ}\}$ such that $q(x)$ is $x$-acyclic and connected. From Lemma 3 we know that there is an OMQ $Q' = (T', \Sigma, C(x))$ that is equivalent to $Q$, with $C(x)$ an $\text{ALCI}$-$\text{IQ}$. From the proof of the lemma, we further know that $C$ has the form $P \rightarrow \bigcup p(x) C_p$ where each $C_p$ is an $\text{EL}$-$\text{concept}$. We show how to transform $Q'$ into an equivalent OMQ $(T', \Sigma, C'(x))$ from $\{\text{ALCH}, \text{IQ}\}$. We start with setting $T' := T$ and $C' := C$ and apply the following modification step until no further changes are possible: if $D$ is a subconcept of $C'$ that is of the form $\exists y. E$ with $E$ an $\text{EL}$-concept, then let $P_D$ be a fresh concept name that is not in $\Sigma$ and

- set $T' := T' \cup \{E \subseteq \forall r. P_D\}$ and
- replace $\exists y. E$ in $C'$ with $P_D$.

At the end of the transformation, $C'$ will contain no inverse roles anymore, so the constructed OMQ is from $\{\text{ALCH}, \text{IQ}\}$. Moreover, it is straightforward to show that the described modification step preserves equivalence of the OMQ.

In fact, assume that $Q_2 = (T_2, \Sigma, C_2(x))$ was obtained by a single modification step from $Q_1 = (T_1, \Sigma, C_1(x))$. Let $A$ be a $\Sigma$-ABox and $a \in \text{Ind}(A)$. First assume that $A \not\models Q_1(a)$. Then there is a model $I$ of $A$ and $T_1$ with $a \not\models C_1^\con$. Extend $I$ to the concept name $P_D$ by setting $P_D^I = (\exists y. E)^I$. Clearly, $I$ is then a model of $T_2$. Moreover, by construction of $C_2$ we have $a \not\models C_2^\con$. Conversely, assume that $A \models Q_1(a)$. Then there is a model $I$ of $A$ and $T_2$, we have $(\exists y. E)^I \subseteq P_D^I$. We can modify $I$ by setting $P_D^I = (\exists y. E)^I$ and the resulting $I$ will still be a model of $T_1$ and still satify $a \not\models C_2^\con$ since all occurrences of $P_D$ in $C_2$ are positive. Moreover, by construction of $C_2$ it also satisfies $a \not\models C_2^\con$.

The proof of the following lemma is a much simplified and slightly extended version of a construction from Kikot and Zolin, 2013.

Lemma 24. 1. Every OMQ $Q = (T, \Sigma, q(x))$ from $\{\text{ALCH}, \text{UCQ}\}$ with $q(x)$ $x$-acyclic and $x$-accessible is rewritable into an OMQ $Q = (T, \Sigma, C(x))$ with $C(x)$ an $\text{ACI}$-$\text{IQ}$ and

2. Every OMQ $Q = (T, \Sigma, q(x))$ from $\{\text{ACI}^u, \text{UCQ}\}$ with $q(x)$ $x$-acyclic is rewritable into an OMQ $Q = (T, \Sigma, C(x))$ with $C(x)$ an $\text{ACI}^u$-$\text{IQ}$.

The size of the IQs $C(x)$ is polynomial in the size of $q(x)$.

Proof. We first observe that Lemma 4 extends to the case where the actual query is a UCQ rather than a UCQ. One simply “carries through” atoms $C(x)$ with $C$ a compound concept in the construction of the IQ.

We start with Point 2 since its proof is simpler and prepares for the proof of Point 1. Thus, let $Q = (T, \Sigma, q(x))$ be an OMQ from $\{\text{ALCH}^u, \text{UCQ}\}$ with $q(x)$ $x$-acyclic. By (the extended) Lemma 4, there is an equivalent OMQ $Q' = (T, \Sigma, C(x))$ with $C(x)$ an $\text{ACI}^u$-$\text{IQ}$. In fact, the IQ $C(x)$ constructed in the proof of Lemma 4 is of the form $P \rightarrow \bigcup p(x) C_p$ where each $C_p$ is an $\text{EL}$-$\text{concept}$ decorated with $\text{ACI}^u$-concepts, that is, built according to the syntax rule

$$C, D ::= \top | A | C \sqcap D | \exists r. D | \exists u. D | E$$
where $A$ ranges over all concept names, $r$ over all (potentially inverse) roles, and $E$ over all ALC$^r$-concepts. Note that every $\mathcal{ELT}^r$-concept decorated with ALC$^u$-concepts is an ALCIT$^r$-concept, but that the converse is false.

We construct from $Q'$ an (ACCH, IQ)-rewriting $(T, \Sigma, C'(x))$ of $Q$ where $C'$ has the form $C_{\mathrm{pre}} \rightarrow \con$. To start, let $D_1 = \exists r_1.P, \ldots, D_L = \exists r_L.P$ be all subconcepts of $C'$ that are of this form and let

- $\con$ be obtained from $C$ by replacing each concept $D_i$ with a fresh concept name $P_{D_i} \notin \Sigma$ and
- $C_{\mathrm{pre}} = \forall r_1.P_{D_1} \cap \cdots \cap \forall r_L.P_{D_L}$.

Next, exhaustively apply the following transformation step:

- if $D = \exists r^{-}.E$ is a subconcept of $\con$ where $E$ is an ALC$^u$-concept (that is, does not contain any inverse roles), then
  - replace $D$ in $\con$ with a fresh concept name $P_{D}$ and
  - set $C_{\mathrm{pre}} = C_{\mathrm{pre}} \cap \forall u.(E \rightarrow \forall r.P_{D})$.

We end up with $\con$ being an ALC$^u$-concept because if there is a subconcept $\exists r^{-}.E$ of $\con$ left, then in the innermost such subconcept $E$ must be an ALC$^u$-concept and thus the transformation rule applies. It can be proved that the initial IQ $C_{\mathrm{pre}} \rightarrow \con(x)$ is equivalent to $C(x)$ and that the transformation step is equivalence preserving. We omit details, please see the proof of Lemma 23 for very similar arguments.

We now turn to Point 1. Let $Q = (T, \Sigma, q(x))$ be an OMQ from (ACCH, UeCQ) with $q(x)$ $x$-acyclic and $x$-accessible. Then $q(x)$ is also connected. By (the extended) Lemma 4 there is an equivalent IQ $Q' = (T, \Sigma, C(x))$ with $C(x)$ an ALCIT-IQ. In fact, the IQ $C(x)$ constructed in the proof of Lemma 4 is of the form $P \rightarrow p(x)$ a CQ in $q(x)$, where each $C_p$ is an $\mathcal{ELT}$-concept decorated with ALC-concepts, that is, an $\mathcal{ELT}$-concept decorated with $\mathcal{ALC}$-concepts that does not mention the universal role. However, the syntactic structure of $C$ is even more restricted.

**Claim.** In each subconcept $\exists r^{-}.D$ of $C$, $D = P$ or $D$ has the form $D_0 \cap \exists r^{-}.(D_1 \cap \exists r_2^{-}.(\ldots \cap \exists r_n^{-}.P))$, $n \geq 1$.

**Proof of claim.** Let $p(x)$ be a CQ in $q(x)$. Recall that, when constructing $C(x)$ in the proof of Lemma 4, we first remove atoms of the form $r(x, y)$ from $p(x)$ to obtain a tree-shaped CQ $p'(x)$, then add back $r^{-}(y, u)$ and $P(u)$ for each removed $r(x, y)$ where $u$ is a fresh variable producing a CQ $p''(x)$, and finally view $p''(x)$ as an $\mathcal{ELT}$-concept $C_p$ decorated with ALC-concepts.

Let $\exists r^{-}.D$ be a subconcept of $C_p$. Then there is a variable $y$ in $p''(x)$ and an atom $r^{-}(y, z)$ such that $D$ describes the subtree of $p''(x)$ rooted at $z$ and $z$ is a successor of $z$ in the tree-shaped $p''(x)$, that is, $y$ is on the path from the root $x$ of $p''(x)$ to $z$. First assume that $r^{-}(y, z)$ was one of the atoms added back in the construction of $p''(x)$. Then $D = P$ and we are done. Now assume that $r^{-}(y, z)$ was already in $p'(x)$ and thus in $p(x)$. Since $p(x)$ is $x$-accessible, $z$ is reachable from $x$ in the directed graph $G_p$. Since $z$ is not reachable from $x$ in the directed graph $G_p$, it follows from the construction of $p'(x)$ and $p''(x)$ that $z$ is reachable in $G_p$, from a leaf node labeled with $P$. Consequently, $D$ must have the stated form. This finishes the proof of the claim.

We construct from $Q'$ an (ACCH, IQ)-rewriting $(T, \Sigma, C'(x))$ of $Q$ where $C'$ has the form $C_{\mathrm{pre}} \rightarrow \con$. To start, let $D_1 = \exists r_1.P, \ldots, D_L = \exists r_L.P$ be all subconcepts of $C'$ that are of this form and let

- $\con$ be obtained from $C$ by replacing each concept $D_i$ with a fresh concept name $P_{D_i}$ and
- $C_{\mathrm{pre}} = \forall r_1.P_{D_1} \cap \cdots \cap \forall r_L.P_{D_L}$.

It is easy to see that the following condition is satisfied:

- $\ast$ in every subconcept $D = \exists r^{-}.E$ of $\con$ with $E$ an ALC-concept, $E$ is of the form $F \cap P_{D'}$.

Next, exhaustively apply the following transformation step, which preserves $\ast$: If $D = \exists r^{-}.(F \cap P_{D'})$ is a subconcept of $\con$, where $F$ is an ALC-concept, then

- replace $D$ in $\con$ with a fresh concept name $P_{D}$ and
- replace $P_{D'}$ in $P_{D}$ with $F \rightarrow \forall r.P_{D}$.

It can be verified that, because of the claim, the transformation step indeed preserves $\ast$. It can also be seen that all subconcepts of the form $\exists r^{-}.E$ will eventually be eliminated. Finally, it can be shown that the initial IQ $C_{\mathrm{pre}} \rightarrow \con(x)$ is equivalent to $C(x)$ and that the transformation step is equivalence preserving. We omit details.

**Example 25.** Let $Q$ be the OMQ from Example 22. Towards obtaining an ALC-IQ rewriting, we start with the ALCIT-IQ rewriting $Q'$ described in the same example. The only subconcept of the form $\exists r^{-}.P$ in $C_q$ is $D = \exists s^{-}.P$. We thus introduce a fresh concept name $P_E$ and initialize $C_p$ and $\con$ with $\forall s.P_E$ and $\exists r.\exists s^{-}.(\exists s.T \cap P_{D})$. We consider concepts of the form $\exists r^{-}.(F \cap P_{D})$, with $F$ an ALC concept and $P_{D'}$ previously introduced. The only such concept is $E = \exists t^{-}.(\exists s.T \cap P_{D})$. We replace $E$ in $\con$ with $P_{E}$ and $P_{D'}$ in $P_{E}$ with $\exists s.T \cup \forall r.P_E$. At this point both $C_{\mathrm{pre}} = \forall s.(\forall t.(\neg \exists s.T \cup \forall r.P_E))$ and $\con = \exists r.P_E$ are ALC-concepts, thus no further transformation is possible (and neither needed): $Q$ can be rewritten into an OMQ $Q'' = (T, \Sigma, C'(x))$ with $C'$ the ALC concept $\forall s.(\forall t.(\neg \exists s.T \cup \forall r.P_E)) \rightarrow \exists r.P_E$.

**Theorem 5.** Let $\mathcal{L} \in \{\mathcal{ALC}, \mathcal{ALCH}\}$ and let $Q = (T, \Sigma, q(x))$ be a unary OMQ from $(\mathcal{L}, \mathcal{UCQ})$ that is non-empty. Then the following are equivalent:

1. $Q$ is rewritable into an OMQ from $(\mathcal{L}, \mathcal{IQ})$;
2. $Q$ is rewritable into an OMQ $(T \cup T', \Sigma, C(x))$ from $(\mathcal{L}, \mathcal{IQ})$;
3. $Q$ is rewritable into an OMQ from $(\mathcal{L}, \mathcal{IQ})$;
4. $Q$ is rewritable into an OMQ $Q'' = (T, \Sigma_{\mathrm{full}}, C'(x))$ from $(\mathcal{L}, \mathcal{IQ})$;
5. $Q \equiv (T, \Sigma_{\mathrm{full}}, q_{\mathrm{deco}}(x))$. 

---

\*The first two steps can together be viewed as an unfolding construction.
If, furthermore, $\mathcal{L}$ is replaced with $\mathcal{L}^u$ and $\mathcal{L}^I$ with $\mathcal{L}^I$, then

6. $Q$ is rewritable into an OMQ $Q' = (\mathcal{T}, \Sigma, C(x))$ from $\mathcal{L}^u$.

**Proof.** “$2 \Rightarrow 1$” and “$1 \Rightarrow 3$” are trivial.

“3 $\Rightarrow 2$.” We know from Theorem 3 that $(\mathcal{L}, \mathcal{I}$)-rewritability of $Q$ implies that $q(x)$ is $x$-acyclic and connected. By Lemma 3, $Q$ is rewritable into an OMQ from $(\mathcal{L}, \mathcal{I})$ that is of the desired shape.

“4 $\Rightarrow 5$”. Let $Q = (\mathcal{T}, \Sigma_{\text{full}}, q(x))$ be an OMQ from $(\mathcal{L}^u, \text{UCQ})$ and assume that $Q$ is rewritable into an OMQ $Q' = (\mathcal{T}, \Sigma_{\text{full}}, C(x))$ with $C(x)$ an $\mathcal{ALC}$-IQ. Let $Q_{\text{acyc}} = (\mathcal{T}, \Sigma_{\text{full}}, q_{\text{acyc}}(x))$. It is established in the proof of the “1 $\Rightarrow 3$” direction of Theorem 3 that, since $Q$ is rewritable into $(\mathcal{ALC}(\mathcal{L}, \mathcal{I}), Q) \subseteq Q_{\text{acyc}}$. It thus suffices to show that $Q_{\text{acyc}} = Q_{\text{deco}}$.

Using the definition of $Q_{\text{deco}}$, it can be shown that $Q_{\text{acyc}} = Q_{\text{deco}}$. To establish the converse direction, assume towards a contradiction that there is a $\Sigma_{\text{full}}$-ABox $\mathcal{A}$ such that $\mathcal{A} \models Q_{\text{deco}}(a)$ but $\mathcal{A} \not\models Q_{\text{acyc}}(a)$. Then $\mathcal{A} \not\models Q'(a)$. Take a model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$ such that $\mathcal{I} \not\models C(a)$. We have $\mathcal{I} = q_{\text{deco}}(a)$, thus $\mathcal{I} = p(x)_{\text{dreach}}(p)$ for some $T$-decoration $p(x)$ of a CQ in $q_{\text{acyc}}(x)$. Let $h$ be a homomorphism from $p(x)_{\text{dreach}}(p)$ to $\mathcal{I}$ with $h(x) = a$.

To finish the proof, it suffices to show that we can construct from $\mathcal{I}$ a model $\mathcal{I}'$ of $\mathcal{T}$ such that $\mathcal{I}' \not\models C(a)$ and $\mathcal{I}' = q_{\text{acyc}}(a)$. In fact, we can then take a homomorphic $h'$ from a CQ in $q_{\text{acyc}}(x)$ to $\mathcal{I}'$ with $h'(x) = a$ and let $\mathcal{A}'$ be $\mathcal{I}'$ restricted to the range of $h'$, viewed as an ABox. Clearly, $\mathcal{A}' = q_{\text{acyc}}(a)$ since already $\mathcal{A}' = (\emptyset, \Sigma_{\text{full}}, q_{\text{acyc}}(x))(a)$. Moreover, $\mathcal{I}'$ is a model of $\mathcal{A}'$ and thus $\mathcal{A}' \not\models Q'(a)$, in contradiction to $Q'$ being equivalent to $Q_{\text{acyc}}$.

It thus remains to construct $\mathcal{I}'$. Informally, we do this by adding to $\mathcal{I}$ the part of $p(x)$ that is not reachable from the answer variable along a directed path. By the second condition of $T$-decorations, there is a model $\mathcal{J}$ of $\mathcal{T}$ and a homomorphism $h'$ from $p(x)$ to $\mathcal{J}$. We can assume that $\mathcal{I}$ and $\mathcal{J}$ have disjoint domains.

Let $\mathcal{I}^T$ be the disjoint union of $\mathcal{I}$ and $\mathcal{J}$, extended as follows: for every atom $r(y_1, y_2)$ in $p(x)$ with $y_1 \not\in \text{dreach}(p)$ and $y_2 \in \text{dreach}(p)$, add $(h'(y_1), h(y_2))$ to $\mathcal{I}^T$. It can be verified that the map $h''$ defined by setting $h''(y) = h(y)$ for all $y \in \text{dreach}(p)$ and $h''(y) = h'(y)$ for all variables $y$ in $p$ that are not in $\text{dreach}(p)$ is a homomorphism from $p(x)$ to $\mathcal{I}^T$ with $h''(x) = a$. Thus, $\mathcal{I}' = q_{\text{acyc}}(a)$ as desired. It thus remains to show that $\mathcal{I}'$ is a model of $\mathcal{T}$ and that $\mathcal{I}' \not\models C(a)$. This is a consequence of the following:

(a) for all $\mathcal{ALC}$-concepts $C$ and all $d \in \Delta^T$, $d \in C^T$ iff $d \in C^\mathcal{J}$;
(b) for all subconcepts $C$ of a concept in $\mathcal{T}$ and all $d \in \Delta^\mathcal{J}$, $d \in C^\mathcal{J}$ iff $d \in C^\mathcal{I}$.

Both points are proved by induction on the structure of $C$. This is straightforward for (a) since for every element $d \in \Delta^T$, the subinterpretation of $\mathcal{I}$ induced by the set of elements reachable from $d$ in $\mathcal{I}$ by traveling roles in the forward direction is identical to the corresponding subinterpretation of $\mathcal{I}'$ (and since $\mathcal{ALC}$-concepts do not admit inverse roles). For (b), it is important to observe that if we have added $(h'(y_1), h(y_2))$ to $\mathcal{I}^T$ in the construction of $\mathcal{I}'$, then $h'(y_1)$ has an $r$-successor $d$ in $\mathcal{J}$ such that for all subconcepts $C$ of a concept in $\mathcal{T}$, $d \in C^\mathcal{J}$ iff $h(y_2) \in C^\mathcal{J}$. In fact, this is a consequence of a result that appears in the text with such concepts: when choosing $d = h'(y_1)$, the stated condition must be satisfied.

“5 $\Rightarrow 4$”. We have that $Q = (\mathcal{T}, \Sigma, q(x))$, where each $q_i$ is an $x$-acyclic, accessible eCQ. We can apply Lemma 7 to $\mathcal{I}, \mathcal{L}$-IQ-rewriting.

For the case with the universal role, it is easy to show that $3 \Rightarrow 6$. Again, we can apply Lemma 7.

**Theorem 7.** Let $q(x)$ be a UCQ. Then

1. $q(x)$ is rewritable into an $\mathcal{ALCI}$-IQ iff there is a subquery $q'(x)$ of $q(x)$ that is $x$-acyclic, connected, and equivalent to $q(x)$;

2. $q(x)$ is rewritable into an $\mathcal{ALC}$-IQ iff there is a subquery $q'(x)$ of $q(x)$ that is $x$-acyclic, $x$-accessible, and equivalent to $q(x)$.

When $\mathcal{L}$-IQs are replaced with $\mathcal{L}^u$-IQs, then the same equivalences hold except that connectedness/$x$-accessibility is dropped.

**Proof.** For the statement at Point 1, the “if” direction is a consequence of Lemma 7 while for the statement at Point 2, the same direction is a consequence of Lemma 7 (and similarly for the cases where the universal role is present). We will thus show the “only if” direction in each case. We first show that $\mathcal{L}$-IQ rewritability of a UCQ $q(x)$, for every $\mathcal{L} \in \{\mathcal{ALCI}, \mathcal{ALC}, \mathcal{ALC}^u, \mathcal{ALC}^n\}$, implies the existence of a subquery $q'(x)$ of $q(x)$ that is $x$-acyclic and equivalent to $q(x)$.

A homomorphism minimal CQ (also hom-minimal) is a CQ which does not admit any equivalent strict subquery.

**Claim 1.** Let $q$ and $q'$ be two CQs such that $q \equiv q'$ and $q'$ is hom-minimal. Then

1. $q$ is a subquery of $q$;
2. $q'$ is a contraction of $q$.

**Proof of claim.** Consider any homomorphisms $h_1$ and $h_2$ from $q'$ to $q$ and from $q$ to $q'$, respectively. Then $h_1$ must be injective and $h_2$ must be surjective (otherwise $h_1 \circ h_2$ is a non-injective homomorphism from $q'$ to itself, and thus $q'$ is not hom-minimal). The existence of $h_1$ implies that $q'$ is a subquery of $q$, while the existence of $h_2$ implies that $q'$ is a contraction of $q$.

**Claim 2.** Let $p(x) \in \mathcal{L}^u$ be a CQ in $q_{\text{acyc}}(x)$. Then, there exists a hom-minimal CQ $p'(x)$ in $q_{\text{acyc}}(x)$ such that $p(x) \equiv p'(x)$.

**Proof of claim.** We show that, in fact, every hom-minimal subquery $p'(x)$ of $p(x)$ which is equivalent to $p(x)$ is a CQ in $q_{\text{acyc}}(x)$. From Claim 1, $p'(x) \equiv p(x)$ and $p'(x)$ being hom-minimal, implies that $p'(x)$ is a contraction of $p(x)$. As $p(x)$ is a contraction of a CQ in $q(x)$, it follows that $p'(x)$ is itself a contraction of some CQ in $q(x)$. As $p(x)$ is $x$-acyclic and
$p'(x)$ is a subquery of $p(x)$, it follows that $p'(x)$ is $x$-acyclic. Thus, $p'(x)$ is an $x$-acyclic contraction of some CQ in $q(x)$, or, in other words, $p'(x)$ is a CQ in $\hat{q}_{\text{acyc}}(x)$.

Assume now $\mathcal{L}$-$IQ$ rewritability of $q(x)$. By inspecting Point (a) in the proof of direction “$1 \Rightarrow 3$” of Theorem 3, we observe that $q(x) \equiv \hat{q}_{\text{acyc}}(x)$.

For didactic purposes, we first consider the case where $q(x)$ is a CQ. Then, there must be a CQ $p(x)$ in the UCQ $\hat{q}_{\text{acyc}}(x)$ such that $q(x) \equiv p(x)$. From Claim 2, there must be some CQ $p'(x)$ in $\hat{q}_{\text{acyc}}(x)$ which is hom-minimal and equivalent to $p(x)$, and thus also to $q(x)$. From Claim 1, it follows that $p'(x)$ is a subquery of $q(x)$, and from the fact that $p'(x)$ is a CQ in $\hat{q}_{\text{acyc}}(x)$, it follows that $p'(x)$ is $x$-acyclic.

We now consider the case where $q(x)$ is a UCQ. Let $q_1(x), \ldots, q_k(x)$ be the CQs in $q(x)$ that are minimal in $q(x)$ in the following sense: for all CQs $p(x)$ in $q(x)$, $q_i(x) \subseteq p(x)$ implies $q_i(x) \equiv p(x)$. Take such a minimal CQ $q_i(x)$. Since $q(x) \equiv \hat{q}_{\text{acyc}}(x)$, there must be a CQ $p_i(x)$ in $q_{\text{acyc}}(x)$ such that $q_i(x) \subseteq p_i(x)$. By construction of $\hat{q}_{\text{acyc}}(x)$, $p_i(x)$ must be the contraction of some CQ $\hat{q}_i(x)$ in $q(x)$ and thus $p_i(x) \subseteq \hat{q}_i(x)$. We obtain $q_i(x) \subseteq \hat{q}_i(x)$ and thus $\hat{q}_i(x) \equiv q_i(x)$ and consequently $q_i(x) \equiv p_i(x)$. From Claim 2, there must be some hom-minimal query $p'_i(x) \in q_{\text{acyc}}(x)$ such that $p'_i(x) \equiv p_i(x)$. Then, $q_i(x) \equiv p'_i(x)$ and from Claim 1, $p'_i(x)$ is a sub-query of $q_i(x)$. Let $\hat{q}_{\text{acyc}}(x)$ be the restriction of $q_{\text{acyc}}(x)$ to the chosen CQs $p'_1(x), \ldots, p'_k(x)$. Clearly, $\hat{q}_{\text{acyc}}(x)$ is equivalent to $q(x)$.

Now we concentrate on the “only if” direction for Point 1, i.e. the case where $L$ is $\mathcal{ALC}$, and thus $q(x)$ is $\mathcal{ALC}$-$IQ$ rewritable. We already know that $\hat{q}_{\text{acyc}}(x)$ is equivalent to $q(x)$, thus $\hat{q}_{\text{acyc}}(x)$ is also $\mathcal{ALC}$-$IQ$ rewritable. From the proof of direction “$1 \Rightarrow 3$” Point (b) in Theorem 3, $\mathcal{ALC}$-$IQ$ rewritability implies $q(x) \equiv q^{\text{con}}(x)$, and thus also $q^{\text{con}}(x) \equiv (\hat{q}_{\text{acyc}})^{\text{con}}(x)$ and $(\hat{q}_{\text{acyc}})^{\text{con}}(x)$ is an $x$-acyclic connected subquery of $q(x)$. Point 2 and the cases with universal roles are treated similarly.

\section*{C Proofs for Section 4}

For the proofs in this section, we recall that every CQ $q$ can be viewed in a straightforward way as an ABox $\mathcal{A}_d$ by viewing the atoms as assertions and the variables as individual names.

\textbf{Theorem 8.} For every $Q \in \{\mathcal{CQ}, \mathcal{UCQ}\}$ and $L \in \{\mathcal{ALC}, \mathcal{ALC}_T, \mathcal{ALC}^u, \mathcal{ALC}^T\}$, it is NP-complete to decide whether a given query from $Q$ is $L$-$IQ$-rewritable.

\textbf{Proof.} Let $L \in \{\mathcal{ALC}, \mathcal{ALC}_T, \mathcal{ALC}^u, \mathcal{ALC}^T\}$. We start with the upper bound, that is, given a UCQ $q(x)$, it is in NP to decide whether $q(x)$ is $L$-$IQ$-rewritable.

We guess a subquery $q'(x)$ of the original query $q(x)$ and check whether $q'(x) \equiv q(x)$. This is the case when for every CQ $p(x)$ in $q(x)$ there exists a CQ $p'(x)$ in $q'(x)$ such that $p(x) \subseteq p'(x)$ and vice versa. We guess for every CQ $p(x)$ in $q(x)$ a target CQ $p'(x)$ in $q'(x)$ and a potential homomorphism $h_x : \text{ind}(A_p) \rightarrow \text{ind}(A_p)$. We also guess for every CQ $p'(x)$ in $q'(x)$ a target CQ $p(x)$ in $q(x)$ and a potential homomorphism $h_{x'} : \text{ind}(A_p) \rightarrow \text{ind}(A_p)$. We then check that every $h_x$ and every $h_{x'}$ is an actual homomorphism. If this is the case, $q(x) \equiv q'(x)$ and, provided that $q'(x)$ fulfills the additional conditions in the L-$IQ$ characterisation from Theorem (x-acyclicity, connectedness and/or $x$-accessibility), $q(x)$ is $L$-$IQ$ rewriterable. As the size of our guess is polynomial in the size of $q(x)$ and all checks can be performed in polynomial time, we obtain the desired upper bound.

To show NP-hardness of whether a given CQ is $L$-$IQ$-rewritable, we employ a reduction from the 3-colorability problem ($3\text{COL}$). Let $G = (V, E)$ be an undirected graph, let $q_G$ be $G$ viewed as a conjunctive query where every $\{v_1, v_2\} \in E$ is represented by two atoms $r(v_1, v_2), r(v_2, v_1)$, and choose a $v \in V$. Let

\[ q(x_0) = \exists y q_G \land r(x_0, v) \land r(v, x_0) \land \bigwedge_{i,j \leq 2 \text{ w. } \neq j} \{r(x_i, x_j) \mid i, j \leq 2 \text{ w. } \neq j \} \]

where $y$ contains all elements of $V$ (as variables) as well as the fresh variables $x_1$ and $x_2$.

\textbf{Claim.} $q(x_0)$ is $L$-$IQ$-rewritable iff $G$ is 3-colorable.

\textbf{Proof of claim.} For the “if” direction, assume that $G$ is 3-colorable. Then $G$ admits a homomorphism into the 3-clique (without reflexive loops). Consequently, $q_G(x)$ is homomorphically equivalent to the restriction $q_{3\text{C}}(x_0)$ of $q(x_0)$ to the variables $x_0, x_1, x_2$. In particular, $q(x_0)$ and $q_{3\text{C}}(x_0)$ are then equivalent in the sense of query containment. Since $q_{3\text{C}}(x_0)$ is $x_0$-acyclic and $x_0$-accessible, by Lemma 24, it is rewritable into an $\mathcal{ALC}$-$IQ$.

Conversely, assume that $G$ is not 3-colorable. By Lemma 24, it suffices to show that any subquery $p(x_0)$ of $q(x_0)$ that is equivalent to $q(x_0)$ is not $x_0$-acyclic. Thus let $p(x_0)$ be such a subquery. There is no homomorphism $h$ from $p(x_0)$ to $q_{3\text{C}}(x_0)$ since the equivalence of $p(x_0)$ and $q(x_0)$ implies the existence of a homomorphism $h'$ from $q(x_0)$ to $p(x_0)$ and composing $h'$ with $h$ would establish 3-colorability of $G$. It is easy to verify, though, that when $p(x_0)$ contains no cycle that does not pass through any of $x_0, x_1, x_2$, then there is such a homomorphism $h$. Consequently, $p(x_0)$ is not $x_0$-acyclic.

\textbf{Theorem 9.} Let $Q \in \{\mathcal{CQ}, \mathcal{UCQ}\}$. For OMQs based on the full ABox signature, $IQ$-rewritability is
1. **$\text{ExpTime}$-hard in $\langle \text{ALC}, Q \rangle$ and in $\text{CoNExpTime}$ in $\langle \text{ALCHI}, Q \rangle$**

2. **2$\text{ExpTime}$-complete in $\langle \text{ALCIT}, Q \rangle$ and $\langle \text{ALCHI}, Q \rangle$**

**Proof.** We start with the lower bounds. Points 1 and 2 are treated uniformly. In fact, for $L \in \{\text{ALC}, \text{ALCIT}\}$, we reduce a special case of OMQ evaluation in $(L, \text{CQ})$ to IQ-rewritability in $(L, \text{CQ})$ where $\text{OMQ evaluation}$ in $(L, \text{Q})$ means to decide, given an OMQ $Q = (T, \Sigma_{\text{full}}, q(x))$ from $(L, \text{Q})$, an ABox $A$, and a tuple $a$ whether $A \models Q(a)$. The mentioned special case is that OMQs are Boolean and $A$ takes the form $\{A(a)\}$ and we refer to this as singleton BOMQ evaluation.

Singleton BOMQ evaluation is $2\text{ExpTime}$-hard in $\text{ALCI}$ [Lutz, 2008]. We observe that it is $\text{ExpTime}$-hard in $\text{ALC}$ since concept (un)satisfiability w.r.t. $\text{ALC}$-TBoxes is $\text{ExpTime}$-hard [Schild, 1991] and an $\text{ALC}$-concept $C$ is unsatisfiable w.r.t. an $\text{ALC}$-TBox $T$ iff $\{A(a)\} \models (T \cup \{A \not\subseteq C\}, \Sigma_{\text{full}}, \exists y D(y))$ where $A$ and $D$ are fresh concept names.

Now for the reduction to IQ-rewritability. Let $Q = (T, \Sigma_{\text{full}}, q(x))$ be an OMQ from $(L, \text{CQ})$, $L \in \{\text{ALC}, \text{ALCIT}\}$, and let $A = \{A(a)\}$ be an ABox. Further, let $q'(x)$ is the extension of $q(x)$ with the atom $A(x)$, $x$ a fresh answer variable. It is important to note that $q'(x)$ is a disconnected CQ.

**Claim.** $A \models Q$ iff $Q' = (T, \Sigma_{\text{full}}, q'(x))$ is IQ-rewritable.

**Proof of claim.** If $A \models Q$, then $Q'$ is equivalent to $(T, \Sigma_{\text{full}}, A(x))$ which is from $(L, \text{IQ})$. Conversely, assume that $A \not\models Q$. The query $Q_{\text{con}}$ from Point (b) in the proof of the “$\Rightarrow$” direction of Theorem 3 [Lutz, 2008] applied to $Q'$, is exactly $(T, \Sigma_{\text{full}}, A(x))$. As shown there, IQ-rewritability of $Q'$ implies $Q \models Q_{\text{con}}$, in contradiction to $A \not\models Q$. This finishes the proof of the claim.

For the upper bounds, we use the characterizations from Theorem 3 and Theorem 5 deciding IQ-rewritability in $(\text{ALCHI}, \text{UCQ})$ amounts to checking containment between $Q$ and $(T, \Sigma_{\text{full}}, q_{\text{dec}}, q_{\text{con}}(x))$ while deciding IQ-rewritability in $(\text{ALCH}, \text{UCQ})$ amounts to checking containment between $Q$ and respectively $(T, \Sigma_{\text{full}}, q_{\text{dec}}(x))$. Note that the two involved OMQs share the same TBox and are based on the full ABox signature. There is also an initial emptiness check, which however is just another containment check. We thus have to argue that these containment checks can be carried out in $2\text{ExpTime}$ and $\text{NExpTime}$, respectively.

We start with the case of $(\text{ALCHI}, \text{UCQ})$ and first observe that containment in $(\text{ALCHI}, \text{UCQ})$ is in $2\text{ExpTime}$. In fact, it is shown in [Lutz, 2008] that OMQ evaluation in $(\text{ALCHI}, \text{CQ})$ is in $2\text{ExpTime}$ and the algorithm given there is straightforwardly extended to $(\text{ALCHI}, \text{UCQ})$. It follows that containment between an OMQ $Q_1 = (T, \Sigma_{\text{full}}, q_1(x_1))$ from $(\text{ALCHI}, \text{CQ})$ in an OMQ $Q_2 = (T, \Sigma_{\text{full}}, q_2(x_2))$ from $(\text{ALCHI}, \text{UCQ})$ is in $2\text{ExpTime}$ since $Q_1 \subseteq Q_2$ iff $A_{q_1} \models Q_2(x_1)$. We next observe how this can be lifted to containment in $(\text{ALCHI}, \text{UCQ})$. In fact, it suffices to show that for $Q_1 = (T, \Sigma_{\text{full}}, q_1(x_1))$ from $(\text{ALCHI}, \text{UCQ})$, $x_1 \in \{1, 2\}$, and $q_1 = p_1 \lor \cdots \lor p_k$, we have $Q_1 \subseteq Q_2$ iff $(T, \Sigma_{\text{full}}, p_i) \subseteq Q_2$ for all $i \in \{1, \ldots, k\}$. The “if” direction is trivial. For the “only if” direction, we argue as follows. Assume that $(T, \Sigma_{\text{full}}, p_i) \subseteq Q_2$ for all $i \in \{1, \ldots, k\}$. Let $A \in \text{ABox}$ and let $A_i$ be a model for $A$ such that $A \models Q_1(a)$. It suffices to show that $I \models q_2(a)$ for every finite model $I$ of $A$ and $T$. Let $I$ be such a model and let $A_2$ be $I$ viewed as an ABox. Since $A \models Q_1(a)$, we must have $I \models p_j(a)$ for some $j \in \{1, \ldots, k\}$. Then clearly also $A_2 \models q_2(a)$ since $(T, \Sigma_{\text{full}}, p_j) \subseteq Q_2$. This yields $A_2 \models q_2(a)$. Since $I$ is a model of $A_2$ and $T$, from this we obtain $I \models q_2(a)$, as required.

The argument is not yet complete since the UCQs $q_{\text{con}}(x)$ can be exponentially large. In fact, it may contain exponentially many CQs, but each CQ is only of polynomial size. For checking $(T, \Sigma_{\text{full}}, q_{\text{con}}(x)) \subseteq Q$, using the above argument we can use exponentially many containment checks between an OMQ from $(\text{ALCHI}, \text{CQ})$ and an OMQ from $(\text{ALCHI}, \text{UCQ})$, both of polynomial size. The overall complexity is thus $2\text{ExpTime}$, as required. For checking $Q \subseteq (T, \Sigma_{\text{full}}, q_{\text{con}}(x))$, we observe that, by Lemma 4, $(T, \Sigma_{\text{full}}, q_{\text{con}}(x), C(x))$ is rewrivable into an equivalent OMQ $(T, \Sigma_{\text{full}}, C(x))$ with $C(x)$ an $\text{ALCHI}$-IQ independently of the properties of $Q$ and such that the size of $C(x)$ is polynomial in the size of $q_{\text{con}}(x)$, which in turn is single exponential in the size of $Q$. We can thus replace the check $Q \subseteq (T, \Sigma_{\text{full}}, q_{\text{con}}(x), C(x))$ with $Q \subseteq (T, \Sigma_{\text{full}}, C(x))$. This boils down to deciding OMQ entailment in $(\text{ALCHI}, \text{IQ})$, which is in $\text{ExpTime}$. So despite $C(x)$ being of (single) exponential size, we achieve $2\text{ExpTime}$ overall complexity.

For the case of $(\text{ALCH}, \text{UCQ})$, the argument is essentially the same. However, as also shown in [Lutz, 2008] OMQ evaluation in $(\text{ALCH}, \text{UCQ})$ is in $\text{ExpTime}$ and thus so is our basic containment check between an OMQ from $(\text{ALCH}, \text{CQ})$ and an OMQ from $(\text{ALCH}, \text{UCQ})$. Therefore, the check $(T, \Sigma_{\text{full}}, q_{\text{dec}}(x)) \subseteq Q$ can be implemented in $\text{ExpTime}$ despite the exponential number of CQs in $q_{\text{dec}}(x)$. It is not clear, however, how to implement the containment check $Q \subseteq (T, \Sigma_{\text{full}}, q_{\text{dec}}(x))$ in $\text{ExpTime}$. We give a sketch of how it can be implemented in $\text{CoNExpTime}$. In fact, what we have to implement in $\text{CoNExpTime}$ is the evaluation of an OMQ $Q = (T, \Sigma_{\text{full}}, q(x))$ where $q(x)$ is a UCQ with exponentially many connected CQs, each of polynomial size. Let $A$ be an ABox and $a \in \text{Ind}(A)$. By Lemma 16 $A \not\models Q(a)$ iff there is an extended forest model $I$ of $A$ and $T$ such that $I \not\models q(a)$. It is easy to see that we can further demand that (the tree-shaped parts of) $I$ be of outdegree polynomial in the size of $T$. Our $\text{NExpTime}$ algorithm for the complement of OMQ evaluation is as follows. Let $m$ be the maximum number of variables of a CQ in $q(x)$. We guess an initial piece of the extended forest model $I$ that consists of the ABox part of $A$ together with the tree-shaped parts restricted to depth $m + 1$, along with a type adornment, that is, a function $\mu$ that assigns a $T$-type to every element of the guessed initial part in a way that is consistent with the initial part. Note that we guess an object of single exponential size here. Since the CQ in $q(x)$ are connected and thus are independent of the part of $I$ that lies beyond the guessed initial part, we can verify that $I$ can be extended to a full model by considering the type $\mu(d)$ for every leaf $d$ in the initial part on level $m + 1$ and verifying that $\mu(d)$ is satisfiable with $T$. 
This can be implemented in \textsc{ExpTime}.  

We introduce a preparatory lemma and notation for the proof of Theorem 5. An atomic query (AQ) is an IQ of the form \( A(x) \), with \( A \) a concept name. A Boolean atomic query (BAQ) is a query of the form \( \exists r \ A(x) \), with \( A \) a concept name, and a Boolean conjunctive query (BCQ) is a IQ of arity zero.

**Lemma 26.** Let \( Q \equiv (T, \Sigma, \exists x C(x)) \) be an OMQ from \((\text{ALC}I, \text{BAQ})\). Then, there exists an IQ \( Q' = (T', \Sigma, M(x)) \) from \((\text{ALC}I, \text{AQ})\), such that for all \( \Sigma \)-Boxes \( A \), \( A \models Q \) iff there is an \( a \in \text{ind}(A) \) such that \( A \models Q'(a) \).

**Proof.** We first observe that \( A \models Q \iff A' \models Q \), for some \( A' \in \text{CON}_A \). Thus, it is enough to show the statement of the lemma for connected \( \Sigma \)-Boxes.

Let \( A' \) be a connected \( \Sigma \)-Box and \( M \) a fresh concept name. Let \( T' \) be the TBox obtained from \( T \) by adding \( C \subseteq M \) and \( \exists \! r. \! M \subseteq M \) for every role \( r \) such that \( r \) or the inverse of \( r \) occurs in \( T \). It can be verified that \( Q' = (T', \Sigma, M(x)) \) is as required.

We also introduce a more fine-grained version of a complexity result from [Bourhis and Lutz, 2016] which highlights that the complexity of containment is double exponential only in the maximum size of CQs in the input OMQs, but not in their number. This only requires a careful analysis of the constructions in [Bourhis and Lutz, 2016].

**Theorem 27.** Containment between OMQs from \((\text{ALCH}I, \text{UCQ})\) is in \textsc{2NExptime}. More precisely, for OMQs \( Q_1 = (T_1, \Sigma_1, q_1) \) and \( Q_2 = (T_2, \Sigma_2, q_2) \) with arity \( a \) and \( n_i \) is the number of CQs in \( q_i \) and \( n_i \) the maximum size of a CQ in \( q_i \), \( i \in \{1, 2\} \), it can be decided in time \( 2^{2^{2\log n_1 + 1 + \log n_2 + 1 + \log |T| + \log |\text{log}|} \cdot p \), \( p \) a polynomial.

**Theorem 10.** IQ-rewritability is

1. \textsc{NExptime}-hard in \((\text{ALC}, \text{CQ})\) and
2. \textsc{2NExptime}-complete in all of \((\text{ALCI}, \text{CQ})\), \((\text{ALCI}, \text{UCQ})\), \((\text{ALCH}I, \text{CQ})\), \((\text{ALCH}I, \text{UCQ})\).

**Proof.** We start with the lower bound for Point 1, using a reduction from OMQ emptiness in \((\text{ALC}I, \text{AQ})\) which is known to be \textsc{NExptime}-hard. Let \( Q_0 = (T, \Sigma, A(x)) \) be an OMQ from this language. Also let

\[ q(x) = \exists y. A(x) \land r(x, y) \land r(y, y) \]

where \( r \) is a role name that does not occur in \( T \) and let \( Q = (T, \Sigma \cup \{ r \}, q(x)) \). It suffices to show that \( Q_0 \) is empty iff \( Q \) is IQ-rewritable.

In fact, emptiness of \( Q_0 \) implies emptiness of \( Q \) and thus \((\text{ALC}, \text{AQ})\)-rewritability. Conversely, assume that \( Q_0 \) is non-empty. To show that \( Q \neq Q' \) IQ-rewritability, by Theorem 5 it suffices to show that \( Q \neq Q^{\text{deco}} \) := \((T, \Sigma \cup \{ r \}, q^{\text{deco}}(x)) \) where \( q^{\text{deco}}(x) \) is a UCQ in which every CQ contains the subquery \( r(x, x) \). Since \( Q_0 \) is non-empty, there is a \( \Sigma \)-Box \( A \) and an \( a \in \text{ind}(A) \) such that \( A \models Q_0(a) \). Since \( r \) does not occur in \( T \), we can assume w.l.o.g. that it does not occur in \( A \) as well. Let \( A' = A \cup \{ r(a, b), r(b, b) \} \). By definition of \( Q(a) \), clearly \( A' \models Q(a) \). Moreover, \( A' \neq Q^{\text{deco}}(a) \) because \( A' \) does not contain the assertion \( r(a, a) \).

To establish the lower bound for Point 2, we use a reduction from containment between an OMQ from \((\text{ALCI}, \text{BAQ})\) and an OMQ from \((\text{ALCI}, \text{BCQ})\). This problem has been shown to be \textsc{2NExptime}-hard in [Bourhis and Lutz, 2016]. The reduction presented there uses different TBoxes in the two involved OMQs. However, by Theorem 3 in [Bienvenu et al., 2012], we can assume w.l.o.g. that they both share the same TBox.

Now for the reduction to IQ-rewritability. Let \( Q_1 = (T_1, \Sigma_1, \exists x A(x)) \) be an OMQ from \((\text{ALCI}, \text{BAQ})\) and \( Q_2 = (T_2, \Sigma_2, q()) \) an OMQ from \((\text{ALCI}, \text{BCQ})\). We first show that \( Q_1 \) and \( Q_2 \), which are Boolean, can be replaced with unary OMQs. By Lemma 26, we find an OMQ \( Q'_1 = (T'_1, \Sigma, M(x)) \) from \((\text{ALCI}, \text{AQ})\) such that for all \( \Sigma \)-Boxes \( A, A \models Q_1 \iff A \models Q'_1(a) \) for some \( a \in \text{ind}(A) \). Clearly, the construction from the proof of Lemma 26 is such that \( Q_1 \) is equivalent to \((T'_1, \Sigma, \exists x A(x)) \) and by choosing the fresh concept \( M \) to also not occur in \( q() \) we can further ensure that \( Q_2 \) is equivalent to \((T', \Sigma, q()) \). Thus, we can assume that \( Q_1 \), \( Q_2 \), and \( Q'_1 \) all use the same TBox \( T \) and we can further assume \( T \) contains a CI \( T \subseteq N \) where \( N \) is a concept name not occurring anywhere else, including \( \Sigma \). Set \( Q'_2 = (T, \Sigma, q(x')) \) where \( x' \) is a fresh (answer) variable. It can be verified that \( Q_1 \subseteq Q_2 \iff Q'_1 \subseteq Q'_2 \).

Now let \( q_0(x) \) be \( q() \) extended with the atom \( M(x), x \) a fresh (answer) variable. It is important to note that \( q_0(x) \) is a disconnected.

**Claim.** \( Q'_1 \subseteq Q'_2 \iff Q = (T, \Sigma, q_0(x)) \) is IQ-rewritable.

**Proof of claim.** If \( Q'_1 \subseteq Q'_2 \), then \( Q \) is equivalent to \( Q'_1 \), which is from \((\text{ALCI}, \text{IQ})\). Conversely, assume that \( Q'_1 \not\subseteq Q'_2 \). The query \( Q^{\text{con}} \) from Point (b) in the proof of the “\( 1 \Rightarrow 3 \)” direction of Theorem 3 applied to \( Q \) is exactly \( Q'_1 \). As shown there, IQ-rewritability of \( Q \) implies \( Q \equiv Q^{\text{con}} \), in contradiction to \( Q'_1 \not\subseteq Q'_2 \).

The upper bounds are a consequence of the characterization of IQ-rewritability in \((\text{ALCH}I,\text{UCQ})\) from Theorem 5 in terms of query containment. Containment in \((\text{ALCH}I,\text{UCQ})\) is in \textsc{2NExptime} [Bourhis and Lutz, 2016]. Note that our characterizations use UCQs with exponentially many CQs, each of which is of polynomial size so we cannot apply the containment complexity result as a black box. However, by perusing the refined OMQ containment complexity result from Theorem 27 we observe that containment checking for OMQs is double exponential only in the size of the CQs in the UCQs while it is only exponential in the number of CQs.

Note that we also need an emptiness check beforehand: if the check succeeds and the OMQ is empty, it is also rewritable and so we answer ‘yes’, if not we proceed to perform the containment check. Emptiness is simply a special case of containment, so we end up in the right complexity class.

**D Proofs for Section 5**

We postpone the proof of Theorem 11 as we need the ultrafilter technique introduced in the proof of Theorem 13.
start by discussing the rewritings given in Example 22 in more detail and present an additional example. Recall that 
\( p(x) = \exists y(s(x, y) \land r(x, y)) \) and that we consider the OMQ \( Q = (T, \Sigma_{\text{full}}, p(x)) \) with \( T = \{ \text{func}(r) \} \). We claim that \( Q_r = (T, \Sigma_{\text{full}}, q_r(x)) \) with 
\[
q_r(x) = (\forall s, \bigcup_{i \leq 3} P_i) \rightarrow (\exists s, (\bigcap_{i \leq 3} (P_i \rightarrow \exists r.P_i))
\]
is a rewriting of \( Q \). To prove this claim one can use the following straightforward colorability argument: for every set \( X \) of individual names in an ABox \( A \) which does not contain an atom of the form \( r(c, c) \) with \( c \in X \) and in which \( r \) is functional (in the sense that for any \( a \) there is at most one \( b \) with \( r(a, b) \in A \)) one can color the individual names in \( X \) with three different colors \( p_1, p_2, p_3 \) without having distinct \( c_1, c_2 \) with \( r(c_1, c_2) \in A \) such that \( P_1(c_1) \) and \( P_2(c_2) \) for some \( 1 \leq i \leq 3 \). We give an additional example illustrating this technique which is fundamental for our approach to rewr...
the following holds: for all $0 \leq i < n$, there is a functional path $y_i^0, \ldots, y_i^n$ in $p^i|X$ such that $y_i^0 = y_i^1$, $y_i^k = x_i^k+1$ and $y_i^n = x_i^k+2$. Now take for every $0 \leq i < n$ fresh concept names $P_{X,i}$, $P_{X,i}^2$, $P_{X,i}^3$ and obtain $p_{\text{plain}}^i(x)$ from $p_{\text{plain}}(x)$ by adding the compound atoms

$$P_{X,i}^1 \cup P_{X,i}^2 \cup P_{X,i}^3(y_0^i)$$

Take functional roles $a_{ij}^0, \ldots, a_{ij}^k$ and consider the query

$$s_{ij}^k(y_0^i, y_1^i), \ldots, s_{ij}^k(y_k^i, y_{k+1}^i) \in p^i|X$$

and consider the CQ $p_{X,i}^j(y_0^i)$ defined by taking the conjunction of the atoms $s_{ij}^k(y_0^i, y_1^i), \ldots, s_{ij}^k(y_k^i, y_{k+1}^i)$ and taking $y_0^i$ as the answer variable. Define for $1 \leq j \leq 3$ an eCQ $p_{X,i}^j(y_0^i)$ by adding to $p_{X,i}^j(y_0^i)$ the compound atom

$$\exists r_{i+1}. P_{X,i}^j(y_0^i).$$

Take the $\mathcal{ELI}$ concept $D_{X,i}^j$ corresponding to $p_{X,i}^j(y_0^i)$, for $j = 1, 2, 3$. Obtain the CQ $p_{\text{deco}}^j(x)$ from $p_{\text{deco}}(x)$ by adding the compound atoms

$$\bigwedge_{1 \leq j \leq 3} (P_{X,i}^j \rightarrow D_{X,i}^j(y_0^i))$$

We do this for all non-degenerate clusters $X$ and obtain eCQs $p_{\text{plain}}^j(x)$ and $p_{\text{deco}}^j(x)$ by taking the conjunction of all $p_{\text{plain}}^j(x)$ and $p_{\text{deco}}^j(x)$, respectively. Now define the concepts $C_{\text{plain}}^j$ and $D_{\text{deco}}^j$ in the obvious way. It is not difficult to prove that $(T, \Sigma_{\text{full}}, C_{\text{plain}}^j \rightarrow D_{\text{deco}}^j(x))$ is a rewriting of $(T, \Sigma_{\text{full}}, p(x))$. By constructing $C_{\text{plain}}^j$ and $D_{\text{deco}}^j$ for all CQs $p(x)$ in $q(x)$ and taking the disjunction of $C_{\text{plain}}^j \rightarrow D_{\text{deco}}^j(x)$ we obtain a rewriting of $q(x)$. The extension to $\mathcal{ALCI-FU}$ without the connectedness assumption is straightforward.

For the proof of the other direction of Theorem 13 we require some preparation. We use standard notation and results for ultrafilter extensions of interpretations [Blackburn et al., 2002].

For $U \subseteq \Delta^2$ for an interpretation $\mathcal{I}$ we set $U' = \Delta^2 \setminus U$. The fundamental property of ultrafilter extensions is the following anti-preservation result:

**Lemma 32.** For all interpretations $\mathcal{I}$, $\mathcal{ALCI}$ concept $C$, and roles $r$:

- if $\mathcal{I} \models C(a)$, then $\mathcal{I}^{\text{uf}} \models C(U_a)$.
- if $r$ is functional, then $\mathcal{I}^{\text{uf}}$ is functional.

We are now in the position to prove the second part of Theorem 13.

**Lemma 33.** If an OMQ $Q = (T, \Sigma_{\text{full}}, q(x))$ from $(\mathcal{F}, \mathcal{UCQ})$ is rewritable into an OMQ from $(\mathcal{F}, \mathcal{ALCI-IQ})$, then there is a subquery $q'(x)$ of $q(x)$ that is $f$-acyclic, connected, and equivalent to $q(x)$. When $\mathcal{ALCI-IQ}$ is replaced with $\mathcal{ALCI^{fu}}$-IQ, then the same equivalence holds except that connectedness is dropped.

**Proof.** Assume that $T$ contains functionality assertions only and $q(x)$ is a UCQ such that there does not exist an equivalent subquery $q'(x)$ of $q(x)$ which is $f$-acyclic and connected. We may assume that

- there is no homomorphism from any disjunct $p(x)$ of $q(x)$ to another disjunct $p'(x)$ of $q(x)$;
- every homomorphism from any disjunct $p(x)$ into itself is surjective.

Take a disjunct $p(x)$ of $q(x)$ which is not $f$-acyclic or not connected. We consider the case that $p(x)$ is not $f$-acyclic but connected. The case that $p(x)$ is not connected is straightforward. We have a cycle $r_0(x_0, x_1), \ldots, r_{n-1}(x_{n-1}, x_n)$ in $p(x)$ such that $FC_q(x) \cap \{x_0, \ldots, x_{n-1}\} = \emptyset$ and

1. $r_i$ or $r_i^-$ is not functional w.r.t. $T$ for some $i < n$ or
2. there exists no functional path $y_0, \ldots, y_m$ in $p(x)$ with $x_0 = y_0 = y_m$ such that $\{x_0, \ldots, x_{n-1}\} \subseteq \{y_0, \ldots, y_m\}$.

The basic idea of the proof for both Point 1 and Point 2 is as follows. Assume there exists a rewriting $(T, \Sigma_{\text{full}}, C(x))$ of $(T, \Sigma_{\text{full}}, q(x))$.

(a) Using the CQ $p(x)$ we construct an infinite ABox $A$ such that $T, A \models \lnot q(a)$;
(b) By compactness of FO there exists a forest model $\mathcal{I}$ of $T$ and $A$ with $\mathcal{I} \models \lnot C(A)$;
(c) Then $\mathcal{I}^{\text{uf}} \models \lnot C(U_a)$ by Lemma 32.
(d) Moreover, $I^{\omega}$ is a model of a finite ABox $A'$ with individual name $\Upsilon_0$ such that $\mathcal{T}, A' \models q(\Upsilon_0)$. But this contradicts $I^{\omega} \models -C(\Upsilon_0)$.

For Point 1, the construction of $A$ is a variant of a construction given in [Kikot and Zolin, 2013]. Consider the cycle $r_0(x_0, x_1), \ldots, r_{n-1}(x_{n-1}, x_n)$ in $p(x)$ such that for some $i$ neither $r$ nor $r'$ are functional w.r.t. $T$. We may assume that $i = 0$. Let $V$ be the connected component of $\{x_0, \ldots, x_{n-1}\}$ in $\text{nFC}_p(x)$. Regard the variables of $p(x)$ as individual names. Define an ABox $A$ with individuals $\text{FC}_p(x) \cup (\text{nFC}_p(x) \setminus V) \cup (V \times N)$ by setting:

- for all concept names $A$ and variables $y \in \text{FC}_p(x) \cup (\text{nFC}_p(x) \setminus V): A(y) \in A$ iff $A(y) \in q(x)$;
- for all concept names $A$, variables $y \in V$, and $i \in N$: $A(y, i) \in A$ iff $A(y) \in q(x)$;
- for all roles $r$ and variables $y, z \in \text{FC}_p(x) \cup (\text{nFC}_p(x) \setminus V): r(y, z) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all roles $r$, variables $y, y' \in \text{FC}_p(x) \cup (\text{nFC}_p(x) \setminus V)$ and $z \in V$, and $i \in N$: $r(y, (i, z)) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all roles $r$, variables $y, y' \in V$, and $i \in N$: $r((y, i), (y', i)) \in A$ iff $r(y, y') \in p(x)$ and $r_0(x_0, x_1) \neq r(y, y')$;
- for all roles $r$, variables $y, y' \in V$, and $i \neq j \in N$: $r((y, i), (y', j)) \in A$ iff $i < j$ and $r_0(x_0, x_1) = r(y, y')$.

We now check that $A$ satisfies Points (a) to (d).

Point (a). Clearly all $r$ functional w.r.t. $T$ are functional in $A$. Thus, $\mathcal{T}, A \models q(x)$ iff $\emptyset, A \models p'(x)$ for some disjunct $p'(x)$ of $q(x)$. It therefore suffices to show that there is no homomorphism from any disjunct $p'(x)$ of $q(x)$ to $A$ mapping $x$ to $x$. It has been observed in [Kikot and Zolin, 2013] already that the mapping $\pi$ from $p(x)$ to $p'(x)$ mapping every variable to itself and every $(y, i) \in V \times N$ to $y$ is a homomorphism. Using this observation it has been shown that there is no homomorphism from $p(x)$ to $A$ as the composition of $\pi$ with such a homomorphism would be a non-surjective homomorphism mapping $x$ to $x$ which contradicts our assumptions. It also follows that there is no homomorphism from any disjunct $p'(x)$ of $q(x)$ distinct from $p(x)$ to $A$ as the composition of $\pi$ with such a homomorphism would be a homomorphism from $p'(x)$ to $p(x)$ mapping $x$ to $x$.

Point (b). Assume there exists no model of $\mathcal{T}$ and $A$ with $I \models -C(x)$. Then, by compactness, there exists a finite subset $A'$ of $A$ such that there exists no model of $\mathcal{T}$ and $A'$ with $I \models -C(x)$. But then $\mathcal{T}, A' \models q(x)$ which contradicts Point (a) and the assumption that $(\mathcal{T}, \Sigma_{\text{full}}, C(x))$ is a rewriting of $(\mathcal{T}, \Sigma_{\text{full}}, q(x))$.

Point (c). This is by Lemma 32.

Point (d). It follows directly from [Kikot and Zolin, 2013] that $I^{\omega}$ contains a homomorphic image of $p(x)$ under a homomorphism mapping $x$ to $\Upsilon_0$. Regard this image as an ABox $A'$. Then $\mathcal{T}, A' \models q(\Upsilon_0)$. This finishes the proof if Point 1 holds.

Now suppose that Point 2 does not hold. Thus, there exists no functional path $y_0, \ldots, y_m$ in $p(x)$ with $x_0 = y_0 = y_m$ such that $\{x_0, \ldots, x_{n-1}\} \subseteq \{y_0, \ldots, y_m\}$. Let $p'' := p|_{\text{nFC}_p(x)}$. We observe the following

Claim 1. There exist $V_1 \subseteq \text{nFC}_p(x)$ of the form $V_1 = \text{FC}_p(x')$ for some $x'$ in $p(x)$ such that for $V_2 := \text{FC}_p(x) \setminus V_1$ there are $y_1, y_2 \in V_2$ such that there is a path from $y_1$ to $y_2$ in $V_2$ and $z_1, z_2 \in V_1$ such that there are distinct $s_1(y_1, z_1), s_2(y_2, z_2) \in q(x)$ with $s_1, s_2$ functional w.r.t. $T$.

For the proof of Claim 1 take the cycle $r_0(x_0, x_1), \ldots, r_{n-1}(x_{n-1}, x_n)$ in $p(x)$. There must exist $x_i$ such that some $x_j$ with $x_i \neq x_j$ is not in $\text{FC}_p(x_i)$. Let $V_1 = \text{FC}_p(x_i)$. Then we find $x_j \in \text{FC}_p(x_i)$ such that $x_{j+1} \notin \text{FC}_p(x_i)$ and we find a path (possibly of length 0) from $x_{j+1}$ to some $x_j$ within $V_2$ such that there exist $r'$ and $x_{j''} \in V_1$ with $r'(x_{j'}, x_{j''}) \in p(x)$. Then $s_1(y_1, z_1) := r'(x_{j+1}, x_{j})$ and $s_2(y_2, z_2) := r'(x_{j''}, x_{j''})$ are as required.

We define an ABox $A$ with individual names $\text{FC}_p(x) \cup (V_1 \times N) \cup (V_2 \times I)$ where

$I = \{((\beta, E) \mid E \subseteq N, |E| = |\mathcal{S}|, \beta : E \rightarrow \mathcal{S}$ bijective $\}$

and

$S = \{r(y, z) \in p(x) \mid z \in V_1, y \in V_2\}$
as follows:

- for all concept names $A$ and variables $y \in \text{FC}_p(x)$: $A(y) \in A$ iff $A(y) \in p(x)$;
- for all roles $r$ and variables $y, z \in \text{FC}_p(x)$: $r(y, z) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all concept names $A$, variables $y \in V_1$, and $i \in N$: $A(y, i) \in A$ iff $A(y) \in p(x)$;
- for all concept names $A$, variables $y \in V_2$, and $i \in I$: $A(y, i) \in A$ iff $A(y) \in p(x)$;
- for all roles $r$, variables $y, z \in V_1$ and $i \in N$: $r((y, i), (z, i)) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all roles $r$, variables $y, z \in V_2$ and $i \in I$: $r((y, i), (z, i)) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all roles $r$, variables $y, z \in V_2$ and $i \in I$: $r((y, i), (z, i)) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all roles $r$, variables $y \in \text{FC}_p(x)$ and $z \in V_2$, and $i \in I$: $r((y, i), (z, i)) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all roles $r$, variables $y \in \text{FC}_p(x)$ and $z \in V_2$, and $i \in I$: $r((y, i), (z, i)) \in A$ iff $r(y, z)$ is in $p(x)$;
- for all roles $r$, variables $y \in V_2$ and $i \in I$: $r((y, i), (z, i)) \in A$ iff $r(y, z)$ is in $p(x)$.

This construction of the ABox $A$ achieves the following:

- for all copies $V_2'$ of $V_2$ and any $r(y, z) \in S$, there is a copy $V_1'$ of $V_1$ such that $r(y', z') \in A$ for the copies $y', z'$ of $y$ and $z$ in $V_2'$ and $V_1'$, respectively.
for all copies $V'_1$ of $V_2$ and copies $V'_i$ of $V_1$ there is at most one atom $r(y, z) \in A$ with $y \in V_2$ and $z \in V'_i$;

- Let $V'_1, \ldots, V'_n$ be distinct copies of $V_2$ and $r_1(y_1, z_1), \ldots, r_n(y_n, z_n)$ be distinct atoms in $S$. Then there is a single copy $V'_1$ of $V_1$ such that $r_1(y'_1, z'_1), \ldots, r_n(y'_n, z'_n) \in A$ for the copies $y'_1, \ldots, y'_n$ of $y_1, \ldots, y_n$ in $V'_1, \ldots, V'_n$, respectively, and the copies $z'_1, \ldots, z'_n$ of $z_1, \ldots, z_n$ in $V'_1$.

We first show Point (a) above.

Point (a). $\mathcal{T}, \mathcal{A} \not\models q(a)$.

Proof of Point (a). By construction, all $r$ functional w.r.t. $\mathcal{T}$ are functional in $\mathcal{A}$. Thus, $\mathcal{T}, \mathcal{A} \models q(x)$ iff $\mathcal{A} \models p'(x)$ for some disjunct $p'(x)$ of $q(x)$. It therefore suffices to show that there is no homomorphism from any disjunct $p'(x)$ of $q(x)$ to $A$ mapping $x$ to $x$. Consider the mapping

$$\pi : A \rightarrow p(x)$$

mapping every variable in $FC_p(x)$ to itself and every $(y, i) \in (V_1 \times N) \cup (V_2 \times I)$ to $y$. It is easy to see that $\pi$ is a homomorphism. It also follows that there is no homomorphism from any disjunct $p'(x)$ of $q(x)$ distinct from $p(x)$ to $A$ as the composition of $\pi$ with such a homomorphism would be a homomorphism from $p'(x)$ to $p(x)$ mapping $x$ to $x$. It remains to prove that there is no homomorphism from $p(x)$ to $A$ mapping $x$ to $x$. Assume there is such a homomorphism $h$. Then $\pi \circ h$ is a homomorphism from $p(x)$ to $p(x)$ mapping $x$ to $x$. We obtain a contradiction if we can show that $\pi \circ h$ is surjective. To this end assume that $\pi \circ h$ is surjective. As $p(x)$ is finite, it is an isomorphism. Let $h[p(x)] = \{h(y) \mid y \in p(x)\}$ be the image of $h$ in $A$. Then $h$ is an isomorphism from $p(x)$ onto the restriction $A|_{h[p(x)], p(x)}$ of $p(x)$ to $h[p(x)]$ and the restriction $\pi|_{h[p(x)], p(x)}$ of $\pi$ to $h[p(x)]$ is an isomorphism onto $p(x)$. It follows that $h[p(x)]$ contains for every $y$ in $p(x)$ exactly one individual $a$ with $\pi(a) = y$ and

- $h[p(x)]$ contains $FC_p(x)$;
- as $V_1$ is connected, there exists $i \in N$ such that $h((p(x))) \supseteq V_1 \times \{i\}$ and $h((p(x))) \cap V_1 \times \{j\} = \emptyset$ for all $j \neq i$;
- for every connected component $V$ of $V_2$ of there exists $i \in I$ such that $h((p(x))) \supseteq V \times \{i\}$ and $h((p(x))) \cap V \times \{j\} = \emptyset$ for all $j \neq i$.

Now recall that there are distinct atoms $s_1(y_1, z_1) \in p(x)$ and $s_2(y_2, z_2) \in p(x)$ such that $y_1, y_2$ are in the same connected component in $V_2$ and $z_1, z_2 \in V_1$. Thus, there exists $i \in N$ such that $(y_1, i), (y_2, i) \in h(p(x))$ and for some $j \in I$, $s_1((y_1, i), (z_1, j)) \in A$ and $s_2((y_2, i), (z_2, j)) \in A$. As observed above, no such two atoms exist in $\mathcal{A}$ and we have derived a contradiction. This finishes the proof of Point (a).

**Example 34.** Let $\mathcal{T} = \{\text{func}(s_1), \text{func}(s_2)\}$ and consider the CQ

$$q(x) = \exists y, z(r(x, y) \land s_1(y, z) \land s_2(y, z))$$

Then

$$FC_q(x) = \{x\}, \hspace{1cm} V_1 = \{z\}, \hspace{1cm} V_2 = \{y\}$$

Thus,

$$S = \{s_1(y, z), s_2(y, z)\}$$

and so the individuals of $\mathcal{A}$ are

$$\{x\} \cup \{\{z\} \times N\} \cup \{(y) \times I\}$$

and the essential properties of $\mathcal{A}$ are:

- for every $(y, i) \in I$ there are distinct $(z, i_1), (z, i_2)$ with $s_1((y, i), (z, i_1)), s_2((y, i), (z, i_2)) \in A$;
- for any two $(z, i_1), (z, i_2)$ there exists $(y, i)$ such that $s_1((y, i), (z, i_1)), s_2((y, i), (z, i_2)) \in A$.

Observe that $\mathcal{T}, \mathcal{A} \models q(x)$.

Points (b) and (c) are as before. It remains to show Point (d). Let $\mathcal{I}$ be a model of $\mathcal{T}$ and $\mathcal{A}$ with $\mathcal{I} \models \neg C(a)$ and consider the ultrafilter extension $\mathcal{I}^w$. We define a homomorphism $h$ from $p(x)$ to $\mathcal{I}^w$ mapping $x$ to $\mathcal{U}_x$. For every $y \in FC_p(x)$, we set $h(y) = \mathcal{U}_y$. To define $h$ for the remaining variables, fix a non-principal ultrafilter $\mathcal{U}$ over $\mathcal{N}$. For every variable $z \in V_1$ we obtain an ultrafilter $\mathcal{U}(z)$ over $\mathcal{I}^w$ by setting $U \in \mathcal{U}(z)$ iff $\{i \mid (z, i) \in U \cap (V_1 \times N)\} \in \mathcal{U}$. Observe that for $z_1, z_2 \in V_1$, $r(z_1, z_2) \in p(x)$ implies $(\mathcal{U}(z_1), \mathcal{U}(z_2)) \in r^w$. Observe as well that for $y \in FC_p(x)$ and $z \in V_1$, $r(y, z) \in p(x)$ implies $(\mathcal{U}_y, \mathcal{U}(z)) \in r^w$. We set $h(z) = \mathcal{U}(z)$ for $z \in V_1$. It remains to define $h$ for variables in $V_2$.

Let for $y \in V_2$, and $X \subseteq \mathcal{I}^w$, $\rho_y(X) = \{i \in I \mid (y, i) \in X\}$. By construction, the set

$$\mathcal{X} = \rho_y(\exists r.Z) \cap (r(y, z) \in S \& \mathcal{Z} \in \mathcal{I}(z))$$

has the finite intersection property. Thus, there exists an ultrafilter $\mathcal{J}$ over $\mathcal{I}$ containing $\mathcal{X}$. For every variable $y \in V_2$ we obtain an ultrafilter $\mathcal{J}(y)$ over $\mathcal{I}^w$ by setting $U \in \mathcal{J}(y)$ iff $\{i \mid (y, i) \in U \cap (V(z \times I)\} \in \mathcal{J}$. Observe that for $y_1, y_2 \in V_1$, $r(y_1, y_2) \in p(x)$ implies $(\mathcal{J}(y_1), \mathcal{J}(y_2)) \in r^w$. Observe as well that for $y \in FC_p(x)$ and $y' \in V_2$, $r(y, y') \in p(x)$ implies $(\mathcal{U}_y, \mathcal{J}(y')) \in r^w$. Finally, observe that by construction $r(y, z) \in S$ implies that $(\mathcal{J}(y), \mathcal{I}(z)) \in r^w$. We set $h(y) = \mathcal{J}(y)$ for $y \in V_2$. It follows that $h$ is a homomorphism from $p(x)$ to $\mathcal{I}^w$, as required.

Now let $A'$ be the image of $p(x)$ under $h$, regarded as an ABox. Then $\mathcal{T}, A' \models q(\mathcal{U}_x)$, as required.

The proof when $\mathcal{ALCT}-IQ$ is replaced with $\mathcal{ALCT}^{-}\text{IQ}$ and connectedness is dropped is a straightforward variation of the proof above.

**Theorem 11.** In $\mathcal{ALCT}, \mathcal{CQ}$, IQ-rewritability is undecidable.

**Proof.** We use a reduction from emptiness checking in $\mathcal{ALCT}, \mathcal{AQ}$ which is known to be undecidable [Baader et al., 2016]. Let $Q = (T, \Sigma, A(x))$ be an OMQ from this language and let $q(x) = \exists y.A(x) \land r(x, y) \land r(y, y)$, where $r$ is a role name that does not occur in $T$. We show that $Q$ is empty iff $Q' = (T, \Sigma \cup \{r\}, q(x))$ is IQ-rewritable. Clearly, if $Q$ is empty, then $Q'$ is empty and, therefore, IQ-rewritable. Conversely, if $Q$ is not empty, then we show that
Q’ is not IQ-rewritable. To this end we take a Σ-ABox A₀ and a ∈ ind(A) such that A₀ |= Q(a) and A₀ is consistent with T. Assume for a proof by contradiction that there is a rewriting Q'' = (T', Σ ∪ {r}, C(x)) of Q'. We use a minor modification of the construction in the proof of Theorem [13]

(a) Using the CQ q(x) and A₀ we construct an infinite (Σ ∪ {r})-ABox A ≅ A₀ such that A ∉ Q'(a);
(b) By compactness of FO there exists a forest model I of T' and with |C(a)|; 
(c) Then T'' |= ¬C(∪₀) and I |= T' by Lemma [13];
(d) Moreover, T'' is a model of a finite (∑ ∪ {r})-ABox A' with individual union ₀ such that A' |= Q'(∪₀). But this contradicts T'' |= ¬C(∪₀).

The construction of A is the same as in Point 1 of the proof of Theorem [13] above using the query q(x) and its cycle r(y, y) except that we also attach A₀ to the ABox constructed by identifying x and a. Now the proof of Points (a) to (d) is exactly as before.

E. Proofs for Section 6

We start with definitions of MMSNP and CSP together with some preliminaries. We consider signatures S that consist of predicate symbols with unrestricted arity, known as schemas. An S-fact is an expression of the form S(a₁, ..., aₙ) where S ∈ S is an n-ary predicate symbol, and a₁, ..., aₙ are elements of some fixed, countably infinite set const of constants. An S-instance I is a set of S-facts. The domain of I, denoted dom(I), is the set of constants that occur in some fact in I. The notions of cycles, girth, acyclicity, and connectedness can be lifted from ABoxes to S-instances, for details see [Feier et al. 2017]. We use CONJ to denote the set of S-instances that are the maximal connected components of the S-instance I.

An MMSNP sentence ϕ over schema S has the form

\[ \exists x_1 \cdots \exists x_n \forall x_{1} \cdots \forall x_m \psi, \]

with X₁,...,Xₙ monadic second-order (SO) variables, x₁,…,xₘ first-order (FO) variables, and ψ a conjunction of quantifier-free formulas of the form

\[ \alpha_1 \land \cdots \land \alpha_n \rightarrow \beta_1 \lor \cdots \lor \beta_m \text{ with } n, m \geq 0, \]

where each αᵢ is of the form Xᵢ(xᵢ) or R(x) with R from S, and each βᵢ is of the form Xᵢ(xᵢ). We refer to a formula of the form \( \{ \) as a rule in ϕ, to the conjunction \( \alpha_1 \land \cdots \land \alpha_n \) as its body, and to \( \beta_1 \lor \cdots \lor \beta_m \) as its head. A rule body can be seen as an S ∪ {X₁,...,Xₙ}-instance in the obvious way, which we shall sometimes do implicitly. An MMSNP sentence ϕ is connected if the body of every rule in ϕ is connected. The rule size of ϕ is the maximum size of a rule in ϕ.

Every MMSNP sentence ϕ can be seen as a Boolean query in the obvious way, that is, for an S-instance I, I ||= ϕ whenever ϕ evaluates to true on I. We also consider disjunctions of MMSNP sentences over schema S, that is, sentences of the form \( \bigvee_i \varphi_i \), where each \( \varphi_i \) is an MMSNP sentence over S. For an S-instance I, I ||= \( \bigvee_i \varphi_i \), whenever there exists some i such that \( \varphi_i \) evaluates to true on I. An MMSNP sentence \( \varphi_2 \) over S is contained in an MMSNP sentence \( \varphi_1 \) over S, written \( \varphi_1 \subseteq \varphi_2 \), if for every S-instance I, I ||= \( \varphi_1 \) implies I ||= \( \varphi_2 \). We say that \( \varphi_1 \) and \( \varphi_2 \) are equivalent if \( \varphi_1 \subseteq \varphi_2 \) and \( \varphi_2 \subseteq \varphi_1 \). Containment and equivalence are defined in the same way for disjunctions of MMSNP sentences.

A constraint satisfaction problem (CSP) is defined by an S-instance T that is called the template. The problem associated with T is to decide whether an input S-instance I admits a homomorphism to T, denoted I ||= T. An MMSNP sentence ϕ over schema S is said to be CSP-definable if there exists an S-template T such that for every S-instance I, I ||= ϕ iff I ||= T. A generalized CSP over schema S is defined by a finite set T = \{T₁, ..., Tₙ\} of S-templates. For an S-instance I, we write I ||= T if I ||= Tᵢ for some i. An MMSNP sentence ϕ over schema S is definable by a generalized CSP T if for every S-instance I, I ||= ϕ iff I ||= T.

For two S-instances I₁ and I₂ with disjoint domains, we use I₁ ∪ I₂ to denote the disjoint union of I₁ and I₂. An MMSNP sentence ϕ is preserved under disjoint union if for all S-instances I₁ and I₂ (with disjoint domains), I₁ ||= ϕ and I₂ ||= ϕ implies I₁ ∪ I₂ ||= ϕ.

Lemma 35. An MMSNP sentence ϕ is CSP-definable if it is definable by a generalized CSP and preserved under disjoint union.


“⇐”. Assume that ϕ is definable by a generalized CSP T = \{T₁, ..., Tₙ\} and preserved under disjoint union. For every i, Tᵢ ||= T and thus Tᵢ ||= ϕ. Assume w.l.o.g. that the domains of all templates in T are mutually disjoint. As ϕ is preserved under disjoint union, \( \bigcup_{1 \leq i \leq n} Tᵢ \) ||= ϕ. Thus, \( \bigcup_{1 \leq i \leq n} Tᵢ \rightarrow T \) and consequently \( \bigcup_{1 \leq i \leq n} Tᵢ \rightarrow T_j \) for some j. This implies that Tᵢ ||= Tᵢ for every i and thus the generalized CSP T is equivalent to the CSP Tₙ, which finishes the proof.

We next determine the complexity of deciding preservation under disjoint union for MMSNP sentences. For our final aim, we only need the upper bound, but we also observe a lower bound for the sake of completeness.

Theorem 36. Deciding whether an MMSNP sentence is preserved under disjoint union is 2NEXPTIME-complete.

Proof. For the upper bound, we reduce preservation under disjoint union to a series of (exponentially many) containment checks between MMSNP sentences (of polynomial size) and then invoke the result from [Bourhis and Lutz, 2016] that MMSNP containment can be decided in 2NEXPTIME.

Let \( \varphi \) be an MMSNP sentence over schema S and \( \mathcal{N} \) the set of nullary predicate symbols in \( \varphi \). We assume w.l.o.g. that the number of first-order variables in \( \varphi \) is bounded from below by the largest arity of a predicate in S. For all \( \mathcal{N}_1, \mathcal{N}_2 \subseteq \mathcal{N} \), we construct an MMSNP sentence \( \varphi_{\mathcal{N}_1, \mathcal{N}_2} \) as follows:

- \( \varphi_{\mathcal{N}_1, \mathcal{N}_2} \) has the same quantifiers as \( \varphi \) except for two new existentially quantified second-order variables \( C_1 \) and \( C_2 \);

We do not assume here that I₁ and I₂ contain the same nullary predicate symbols; I₁ ∪ I₂ contains the union of the nullary symbols in I₁ and I₂.
• $\varphi_{N_1,N_2}$ has the following rules:
  - true $\rightarrow C_1(x) \lor C_2(x)$;
  - $C_1(x) \land C_2(x) \rightarrow$ false;
  - $R(y_1,\ldots,y_n) \land C_i(y_j)$ $\rightarrow C_i(y_k)$ whenever $R \in S$ is $n$-ary, $i \in \{0,1\}$, and $j, k \in \{1,\ldots,n\}$ and where $y_1,\ldots,y_n$ are the first $n$ FO variables in $\varphi$;
  - $C_1(x_1) \land \cdots \land C_2(x_n) \land$ body $\rightarrow$ head whenever body $\rightarrow$ head is a rule in $\varphi$ with no predicate symbol from $\mathcal{N} \setminus N_i$ occurring in body and $i \in \{0,1\}$ and where $x_1,\ldots,x_n$ are the FO variables in body.

Intuitively, an $S$-instance $I$ satisfies $\varphi_{N_1,N_2}$ iff there is a coloring of $I$ with the two colors $C_1$ and $C_2$ such that elements from the same maximal connected components receive the same color and each of the resulting two monochromatic subinstances of $I$ satisfies $\varphi$. Note that $I$ is the disjoint union of $I_1$ and $I_2$. The sets $N_1,N_2$ help to disentangle the nullary predicate symbols: $N_1$ contains the predicates true in the monochromatic subinstance colored $C_1$ and likewise for $N_2$ and $C_2$.

Claim. $\varphi$ is preserved under disjoint union iff $\varphi \equiv \bigvee_{N_1,N_2 \subseteq \mathcal{N}} \varphi_{N_1,N_2}$.

``\Rightarrow``. Assume that $\varphi$ is preserved under disjoint union. We have to show the following inclusions:

- $\varphi \subseteq \bigvee_{N_1,N_2 \subseteq \mathcal{N}} \varphi_{N_1,N_2}$. Let $I \models \varphi$. Further let $I_1$ be the extension of $I$ in which every element is colored $C_1$ and let $N_1$ be the set of nullary predicates true in $I$ and $N_2 = \emptyset$. Clearly, $I_1 \models \varphi_{N_1,N_2}$, witnessing $I \models \bigvee_{N_1,N_2 \subseteq \mathcal{N}} \varphi_{N_1,N_2}$. Let $I \models \bigvee_{N_1,N_2 \subseteq \mathcal{N}} \varphi_{N_1,N_2}$. Then $I \models \varphi_{N_1,N_2}$ for some $N_1, N_2 \subseteq \mathcal{N}$. By construction of $\varphi_{N_1,N_2}$, there is thus a partition of $\text{dom}(I)$ into two sets $S_1, S_2$ such that $I_1 \models \varphi$ and $I_2 \models \varphi$ where $I_1$ is the restriction of $I$ to domain $S_1$ and makes exactly the nullary predicates in $N_1$ true, $i \in \{1,2\}$. As $I$ is the disjoint union of $I_1$ and $I_2$, $I \models \varphi$.

``\Leftarrow``. Assume that $\varphi \equiv \bigvee_{N_1,N_2 \subseteq \mathcal{N}} \varphi_{N_1,N_2}$. Let $I_1$ and $I_2$ be $S$-instances with disjoint domain such that $I_1 \models \varphi$ and $I_2 \models \varphi$. Then $I_1 \uplus I_2 \models \varphi_{N_1,N_2}$, where $N_1$ is the set of nullary predicates true in $I_i$, $i \in \{1,2\}$. Thus, $I_1 \uplus I_2 \models \bigvee_{N_1,N_2 \subseteq \mathcal{N}} \varphi_{N_1,N_2}$ and from the original assumption $I_1 \uplus I_2 \models \varphi$. This finishes the proof of the claim.

It remains to note that the inclusion $\varphi \subseteq \bigvee_{N_1,N_2 \subseteq \mathcal{N}} \varphi_{N_1,N_2}$ holds even when $\varphi$ is not preserved under disjoint union (as shown by the proof above) and thus deciding whether $\varphi$ is preserved under disjoint union amounts to checking that $\varphi \supseteq \varphi_{N_1,N_2}$ for all $N_1, N_2 \subseteq \mathcal{N}$. This gives the desired upper bound.

For the lower bound, we consider the (polynomial time) reduction from [Bourhis and Lutz, 2016] to a 2NEexpTIME-hard torus tiling problem to OMQ containment: there, two OMQs $Q_1$ and $Q_2$ are constructed, $Q_1$ from $(\text{ACCCT}, \text{BAQ})$ and $Q_2$ from $(\text{ACCCT}, \text{BCQ})$, such that a double exponentially large torus can be tiled iff $Q_1 \subseteq Q_2$. We sketch a polynomial time reduction from the containment problem for two such OMQs to the preservation under disjoint union of an MMSNP sentence $\varphi$. For the sake of proving the correctness of our reduction, we note that $Q_1$ and $Q_2$ are such that $Q_2 \not\subseteq Q_1$.

It was shown in [Bienvenu et al., 2014] that every OMQ from $(\text{ACCCT}, \text{BAQ})$ such that $Q_1$ is equivalent to the complement of a CSP $T$ in the sense that for every $\Sigma$-ABox $A$, $A \models Q$ iff $A \not\models B$ where $\Sigma$ is the signature of the OMQs $Q_1$ and $Q_2$. Note that this is a variation of Lemma 21 for the case of BAQs. It has also been shown in [Bienvenu et al., 2014] that for every OMQs from $(\text{ACCCT}, \text{BCQ})$, one can construct in polynomial time an MMSNP sentence whose complement is equivalent to the OMQ and thus for $Q_1$ and $Q_2$ we find two such sentences $\varphi_1$ and $\varphi_2$. The size of $\varphi_1$ and $\varphi_2$ is polynomial in that of $Q_1$ and $Q_2$. Summing up, the 2-exp torus can be tiled iff $\varphi_2 \subseteq \varphi_1$, $\varphi_1$ is equivalent to a CSP, and $\varphi_1 \not\subseteq \varphi_2$. We next construct an MMSNP sentence $\varphi$ such that for all $S$-instances $I$.

$I \not\models \varphi$ iff $I \not\models \varphi_1$ and $I \not\models \varphi_2$. (1)

Towards constructing $\varphi$, we start by standardizing apart all FO and SO variables from $\varphi_1$ and $\varphi_2$. For each $i \in \{1,2\}$, let $\psi_i$ be the quantifier-free part of $\varphi_i$ with all rules of the form $\text{body} \rightarrow$ false removed. Then $\varphi$ is the MMSNP sentence which has as SO/FO variables the union of SO/FO variables from $\varphi_1$ and $\varphi_2$ and the following rules:

- all rules from $\psi_1$ and from $\psi_2$,
- all rules of the form $\text{body}_1 \land \text{body}_2 \rightarrow$ false, where $\text{body}_1 \rightarrow$ false is a rule in $\varphi_1$ and $\text{body}_2 \rightarrow$ false a rule in $\varphi_2$.

It can be verified that $\varphi$ satisfies (1).

Claim. $\varphi_2 \subseteq \varphi_1$ iff $\varphi$ is preserved under disjoint union.

``\Rightarrow``. Assume that $\varphi_2 \subseteq \varphi_1$. Then, $\varphi$ is equivalent to $\varphi_1$ and thus to a CSP, consequently it is preserved under disjoint union.

``\Leftarrow``. Assume that $\varphi_2 \not\subseteq \varphi_1$. Thus there exist instances $I_1$ and $I_2$ such that $I_1 \models \varphi_1$, $I_2 \models \varphi_2$, $I_2 \models \varphi_1$. From (1), we obtain $I_1 \models \varphi$ and $I_2 \models \varphi$. We next observe that $I_1 \not\models \varphi_1$ and $I_2 \not\models \varphi_2$ implies $I_1 \uplus I_2 \not\models \varphi_1$ and $I_1 \uplus I_2 \not\models \varphi_2$. Thus $I_1 \uplus I_2 \not\models \varphi$ by (1). Consequently, $I_1$ and $I_2$ witness that $\varphi$ is not preserved under disjoint union.

We next characterize the equivalence of MMSNP sentences to a generalized CSP and analyze the complexity of deciding this property.

For an MMSNP sentence $\varphi$, let $\varphi_{\text{acyc}}$ be the MMSNP sentence with the same quantifier prefix that contains all rules which have an acyclic body and can be obtained from a rule in $\varphi$ by zero or more identifications of variables.

**Theorem 37.** An MMSNP sentence $\varphi$ is definable by a generalized CSP iff $\varphi \equiv \varphi_{\text{acyc}}$.

**Proof.** "only if". Assume that $\varphi$ is definable by a generalized CSP $T = \{T_1,\ldots,T_n\}$. Using the construction of $\varphi_{\text{acyc}}$, it can be verified that $\varphi \subseteq \varphi_{\text{acyc}}$. It thus remains to be shown that $\varphi_{\text{acyc}} \subseteq \varphi$. If the body of each rule in $\varphi$ is acyclic, then this is clearly the case. Otherwise, let $g$ be the maximum girth of a cyclic rule body from $\varphi$. Take an $S$-instance $I$ such that $I \not\models \varphi$. We have to show that $I \not\models \varphi_{\text{acyc}}$. Since $\varphi$ is
equivalent to $T_i$, $I \nrightarrow T_i$ for $1 \leq i \leq n$. From Lemma 19, we obtain an $S$-instance $I^g$ of girth exceeding $g$ such that $I^g \rightarrow I$ and $I^g \nrightarrow T_i$ for $1 \leq i \leq n$. Thus $I^g \nrightarrow \varphi$. As the girth of $I^g$ is higher than the girth of every cyclic rule body in $\varphi$, it follows that $I^g \nrightarrow \varphi_{acyc}$. Since $I^g \rightarrow I$, $I \nrightarrow \varphi_{acyc}$.

“if”. Assume that $\varphi \equiv \varphi_{acyc}$. Since the rule bodies in $\varphi_{acyc}$ are acyclic, it is easy to convert $\varphi_{acyc}$ into an equivalent MMSNP sentence in which each rule body contains at most one atom that uses a predicate symbol from $S$: see [Feder and Vardi, 1998]. It is implicit in that paper (see also [Bienvenu et al., 2014]) that MMSNP sentences of this kind have the same expressive power as generalized CSPs. Thus, $\varphi$ is equivalent to a generalized CSP.

Before showing the main complexity result of this section, we state a slightly refined version of a theorem from [Bourhis and Lutz, 2016] regarding the complexity of MMSNP containment. It emphasizes that the complexity of containment is double exponential only in the size of the rules, but not in their number. This only requires a careful analysis of the constructions in [Bourhis and Lutz, 2016].

**Theorem 38.** Containment between MMSNP sentences is in $2\text{NEEXP}$TIME. More precisely, for MMSNP sentences $\varphi_1$ and $\varphi_2$ where $\varphi_i$ has $n_i$ rules and rule size $r_i$, $i \in \{1, 2\}$, it can be decided in time $2^{2^{\log n_1 + r_1 + \log n_2 + r_2}}$.

**Theorem 15.** It is $2\text{NEEXP}$TIME-complete to decide whether a given MMSNP-sentence is equivalent to a CSP.

**Proof.** We start with the upper bound. Lemma 35 and Theorem 37 suggest an algorithm for deciding CSP-definability of an MMSNP sentence $\varphi$: check whether $\varphi$ is preserved under disjoint union and $\varphi \equiv \varphi_{acyc}$. The first condition can be decided in $2\text{NEEXP}$TIME according to Theorem 36. As for the second check, we note that the size of $\varphi_{acyc}$ might be exponential in the size of $\varphi$ so we cannot apply the MMSNP containment result from [Bourhis and Lutz, 2016] straightaway. However, the rule size of $\varphi_{acyc}$ is polynomial in the size of $\varphi$ and thus by Theorem 38 the second condition can be decided in $2\text{NEEXP}$TIME as well.

For showing that CSP-definability of MMSNP sentences is $2\text{NEEXP}$TIME-hard, we can apply the same reduction as in the proof of Theorem 36: the MMSNP sentences $\varphi_1$, $\varphi_2$, and $\varphi$, constructed in the reduction are such that $\varphi_2 \subseteq \varphi_1$ iff $\varphi$ is equivalent to a CSP.

---

Note that the lemma in its original formulation in [Feder and Vardi, 1998] applies to instances over schemas of any arity, not just to ABoxes.