

Inconsistency-tolerant Conjunctive Query Answering for Simple Ontologies

Meghyn Bienvenu

LRI - CNRS & Université Paris-Sud, France
www.lri.fr/~meghyn/ meghyn@lri.fr

1 Introduction

In recent years, there has been growing interest in using description logic (DL) ontologies to query instance data. An important issue which arises in this setting is how to handle the case in which the data (ABox) is inconsistent with the ontology (TBox). Ideally, one would like to restore consistency by identifying and correcting the errors in the data (using e.g. techniques for debugging or revising DL knowledge bases, cf. [1]). However, such an approach requires the ability to modify the data and the necessary domain knowledge to determine which part of the data is erroneous. When these conditions are not met (e.g. in information integration applications), an alternative is to adopt an inconsistency-tolerant semantics in order to obtain reasonable answers despite the inconsistencies.

The related problem of querying databases which violate integrity constraints has long been studied in the database community (cf. [2] and the survey [3]), under the name of *consistent query answering*. The semantics is based upon the notion of a repair, which is a database which satisfies the integrity constraints and is as similar as possible to the original database. Consistent query answering corresponds to evaluating the query in each of the repairs, and then intersecting the results. This semantics is easily adapted to the setting of ontology-based data access, by defining repairs as the inclusion-maximal subsets of the data which are consistent with the ontology.

Consistent query answering for the *DL-Lite* family of lightweight DLs was investigated in [4, 5]. The obtained complexity results are rather disheartening: the problem was shown in [4] to be co-NP-hard in data complexity, even for instance queries; this contrasts sharply with the very low AC_0 data complexity for (plain) conjunctive query answering in *DL-Lite*. Similarly discouraging results were recently obtained in [6] for another prominent lightweight DL \mathcal{EL}_\perp [7]. In fact, we will see in Example 1 that if we consider conjunctive queries, only a single concept disjointness axiom is required to obtain co-NP-hard data complexity.

In the database community, negative complexity results spurred a line of research [8–10] aimed at identifying cases where consistent query answering is feasible, and in particular, can be done using first-order query rewriting techniques. The idea is to use targeted polynomial-time procedures whenever possible, and to reserve generic methods with worst-case exponential behavior for difficult cases (see [9] for some experimental results supporting such an approach). A similar investigation for *DL-Lite* ontologies was initiated in [11], where general conditions

were identified for proving either first-order expressibility or coNP-hardness of consistent query answering for a given TBox and instance query.

The main objective of the present work is to gain a better understanding of what makes consistent conjunctive query answering in the presence of ontologies so difficult. To this end, we conduct a fine-grained complexity analysis which aims to characterize the complexity of consistent query answering based on the properties of the ontology and the conjunctive query. We focus on simple ontologies, consisting of class subsumption ($A_1 \sqsubseteq A_2$) and class disjointness ($A_1 \sqsubseteq \neg A_2$) axioms, since the problem is already far from trivial for this case. We identify the number of quantified variables in the query as an important factor in determining the complexity of consistent query answering. Specifically, we show that consistent query answering is always first-order expressible for conjunctive queries with at most one quantified variable; the problem has polynomial data complexity (but is not necessarily first-order expressible) when there are two quantified variables; and it may become coNP-hard starting from three quantified variables. For queries having at most two quantified variables, we further identify a necessary and sufficient condition for first-order expressibility.

To obtain positive results for arbitrary conjunctive queries, we propose a novel inconsistency-tolerant semantics which is a sound approximation of the consistent query answering semantics (and a finer approximation than the approximate semantics proposed in [4]). We show that under this semantics, first-order expressibility of consistent query answering is guaranteed for all conjunctive queries. Finally, in order to treat more expressive ontologies, and to demonstrate the applicability of our techniques, we show how our positive results can be extended to handle *DL-Lite_{core}* ontologies without inverse roles.

Full proofs can be found in a long version available on the author's website.

2 Preliminaries

Syntax. All the ontology languages considered in this paper are fragments of *DL-Lite_{core}* [12, 13]. We recall that *DL-Lite_{core}* knowledge bases (KBs) are built up from a set \mathbf{N}_I of *individuals*, a set \mathbf{N}_C of *atomic concepts*, and a set \mathbf{N}_R of *atomic roles*. Complex concept and role expressions are constructed as follows:

$$B \rightarrow A \mid \exists P \quad C \rightarrow B \mid \neg B \quad P \rightarrow R \mid R^-$$

where $A \in \mathbf{N}_C$ and $R \in \mathbf{N}_R$. A *TBox* is a finite set of *inclusions* of the form $B \sqsubseteq C$ (B, C as above). An *ABox* is a finite set of (*ABox*) *assertions* of the form $A(a)$ ($A \in \mathbf{N}_C$) or $R(a, b)$ ($R \in \mathbf{N}_R$), where $a, b \in \mathbf{N}_I$. We use $\text{Ind}(\mathcal{A})$ to denote the set of individuals in \mathcal{A} . A KB consists of a TBox and an ABox.

Semantics An *interpretation* is $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set and $\cdot^{\mathcal{I}}$ maps each $a \in \mathbf{N}_I$ to $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, each $A \in \mathbf{N}_C$ to $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, and each $P \in \mathbf{N}_R$ to $P^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The function $\cdot^{\mathcal{I}}$ is straightforwardly extended to general concepts and roles, e.g. $(\neg A)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus A^{\mathcal{I}}$ and $(\exists S)^{\mathcal{I}} = \{c \mid \exists d : (c, d) \in S^{\mathcal{I}}\}$. \mathcal{I} satisfies $G \sqsubseteq H$ if $G^{\mathcal{I}} \subseteq H^{\mathcal{I}}$; it satisfies $A(a)$ (resp. $P(a, b)$) if $a^{\mathcal{I}} \in A^{\mathcal{I}}$

(resp. $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in P^{\mathcal{I}}$). We write $\mathcal{I} \models \alpha$ if \mathcal{I} satisfies inclusion/assertion α . An interpretation \mathcal{I} is a *model* of $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if \mathcal{I} satisfies all inclusions in \mathcal{T} and assertions in \mathcal{A} . We say a KB \mathcal{K} is *consistent* if it has a model, and that \mathcal{K} *entails* an inclusion/assertion α , written $\mathcal{K} \models \alpha$, if every model of \mathcal{K} is a model of α .

In what follows, it will prove useful to extend the notions of satisfaction and entailment to sets of concepts. We will say that a set of concepts $\{C_1, \dots, C_n\}$ is consistent w.r.t. a TBox \mathcal{T} if there exists a model \mathcal{I} of \mathcal{T} and an element $e \in \Delta^{\mathcal{I}}$ such that $e \in C_i$ for every $1 \leq i \leq n$. Entailment of a concept from a set of concepts is defined in the obvious way: $\mathcal{T} \models S \sqsubseteq D$ if and only if for every model \mathcal{I} of \mathcal{T} , we have $\bigcap_{C \in S} C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

Queries A (*first-order*) *query* is a formula of first-order logic with equality, whose atoms are of the form $A(t)$ ($A \in \mathbf{N}_{\mathbf{C}}$), $R(t, t')$ ($R \in \mathbf{N}_{\mathbf{R}}$), or $t = t'$ with t, t' *terms*, i.e., variables or individuals. *Conjunctive queries* (CQs) have the form $\exists \mathbf{y} \psi$, where \mathbf{y} denotes a tuple of variables, and ψ is a conjunction of atoms of the forms $A(t)$ or $R(t, t')$. *Instance queries* are queries consisting of a single atom with no variables (i.e. ABox assertions). Free variables in queries are called *answer variables*, whereas bound variables are called *quantified variables*. We use $\text{terms}(q)$ to denote the set of terms appearing in a query q .

A *Boolean query* is a query with no answer variables. For a Boolean query q , we write $\mathcal{I} \models q$ when q holds in the interpretation \mathcal{I} , and $\mathcal{K} \models q$ when $\mathcal{I} \models q$ for all models \mathcal{I} of \mathcal{K} . For a non-Boolean query q with answer variables v_1, \dots, v_k , a tuple of individuals (a_1, \dots, a_k) is said to be a certain answer for q w.r.t. \mathcal{K} just in the case that $\mathcal{K} \models q[a_1, \dots, a_k]$, where $q[a_1, \dots, a_k]$ is the Boolean query obtained by replacing each v_i by a_i . Thus, conjunctive query answering is straightforwardly reduced to entailment of Boolean CQs.

First-order rewritability It is shown in [12] that for every *DL-Lite_{core}* TBox \mathcal{T} and CQ q , one can construct a first-order query q' such that for every ABox \mathcal{A} and tuple \mathbf{a} : $\mathcal{T}, \mathcal{A} \models q[\mathbf{a}] \Leftrightarrow \mathcal{I}_{\mathcal{A}} \models q'[\mathbf{a}]$, where $\mathcal{I}_{\mathcal{A}}$ denotes the interpretation with domain $\text{Ind}(\mathcal{A})$ that makes true precisely the assertions in \mathcal{A} .

3 Consistent query answering for description logics

In this section, we formally recall the consistent query answering semantics, present some simple examples which illustrate the difficulty of the problem, and introduce the main problem which will be studied in this paper. For readability, we will formulate our definitions and results in terms of Boolean CQs, but they can be straightforwardly extended to general CQs.

The key notion underlying consistent query answering semantics is that of a repair of an ABox \mathcal{A} , which is an ABox which is consistent with the TBox and as similar as possible to \mathcal{A} . In this paper, we follow common practice and use subset inclusion to compare ABoxes.

Definition 1. A repair of a DL ABox \mathcal{A} w.r.t. a TBox \mathcal{T} is an inclusion-maximal subset \mathcal{B} of \mathcal{A} consistent with \mathcal{T} . We use $\text{Rep}_{\mathcal{T}}(\mathcal{A})$ to denote the set of repairs of \mathcal{A} w.r.t. \mathcal{T} .

Consistent query answering can be seen as performing standard query answering on each of the repairs and intersecting the answers. For Boolean queries, the formal definition is as follows:

Definition 2. A query q is said to be consistently entailed from a DL KB $(\mathcal{T}, \mathcal{A})$, written $\mathcal{T}, \mathcal{A} \models_{\text{cons}} q$, if $\mathcal{T}, \mathcal{B} \models q$ for every repair $\mathcal{B} \in \text{Rep}_{\mathcal{T}}(\mathcal{A})$.

Just as with standard query entailment, we can ask whether consistent query entailment can be tested by rewriting the query and evaluating it over the data.

Definition 3. A first-order query q' is a consistent rewriting of a Boolean query q w.r.t. a TBox \mathcal{T} if for every ABox \mathcal{A} , we have $\mathcal{T}, \mathcal{A} \models_{\text{cons}} q$ iff $\mathcal{I}_{\mathcal{A}} \models q'$.

As mentioned in Section 1, it was shown in [4] that consistent query answering in *DL-Lite_{core}* is co-NP-hard in data complexity, even for instance queries. which means in particular that consistent rewritings need not exist. All known reductions make crucial use of inverse roles, and indeed, we will show in Section 7 that consistent instance checking is first-order expressible for *DL-Lite_{core}* ontologies without inverse. However, in the case of conjunctive queries, the absence of inverses does not guarantee tractability. Indeed, the next example shows that only a single concept disjointness axiom can yield coNP-hardness.

Example 1. We use a variant of UNSAT, called 2+2UNSAT, proved coNP-hard in [14], in which each clause has 2 positive and 2 negative literals, where literals involve either regular variables or the truth constants **true** and **false**. Consider an instance $\varphi = c_1 \wedge \dots \wedge c_m$ of 2+2-UNSAT over v_1, \dots, v_k , **true**, and **false**. Let $\mathcal{T} = \{T \sqsubseteq \neg F\}$, and define \mathcal{A} as follows:

$$\begin{aligned} & \{ P_1(c_i, u), P_2(c_i, x), N_1(c_i, y), N_2(c_i, z) \mid c_i = u \vee x \vee \neg y \vee \neg z, 1 \leq i \leq m \} \\ & \cup \{ T(v_j), F(v_j) \mid 1 \leq j \leq k \} \cup \{ T(\mathbf{true}), F(\mathbf{false}) \} \end{aligned}$$

Then one can show that φ is unsatisfiable just in the case that $(\mathcal{T}, \mathcal{A})$ consistently entails the following query:

$$\exists x y_1 \dots y_4 P_1(x, y_1) \wedge F(y_1) \wedge P_2(x, y_2) \wedge F(y_2) \wedge N_1(x, y_3) \wedge T(y_3) \wedge N_2(x, y_4) \wedge T(y_4)$$

Essentially, $T \sqsubseteq \neg F$ forces the choice of a truth value for each variable, so the repairs of \mathcal{A} correspond exactly to the set of valuations. Importantly, there is only one way to avoid satisfying a 2+2-clause: the first two variables must be assigned false and the last two variables must be assigned true. The existence of such a configuration is checked by q .

We remark that the query in the preceding reduction does not have a particularly complicated structure (in particular, it is tree-shaped). Its only notable property is that it has several quantified variables.

In this paper, we aim to gain a better understanding of what makes consistent conjunctive query answering so difficult (and conversely, what can make it easy). To this end, we will consider the following decision problem:

CONS_{ENT}(q, \mathcal{T}): Is \mathcal{A} such that $\mathcal{T}, \mathcal{A} \models_{\text{cons}} q$?

and we will try to characterize its complexity in terms of the properties of the pair (q, \mathcal{T}) . We will in particular investigate the impact of limiting the number of quantified variables in the query q .

In the next three sections, we focus on *simple ontologies*, consisting of inclusions of the forms $A_1 \sqsubseteq A_2$ and $A_1 \sqsubseteq \neg A_2$ where $A_1, A_2 \in \mathbf{N}_C$. As Example 1 demonstrates, the problem is already non-trivial in this case. All obtained lower bounds transfer to richer ontologies, and we will show in Section 7 that positive results can also be extended to *DL-Lite_{core}* ontologies without inverse roles.

4 Tractability for queries with at most two quantified variables

In this section, we investigate the complexity of consistent query answering in the presence of simple ontologies for CQs having at most two quantified variables. We show this problem has tractable data complexity, and we provide necessary and sufficient conditions for FO-expressibility.

We begin with queries with at most one quantified variable, showing that a consistent rewriting always exists.

Theorem 1. *Let \mathcal{T} be a simple ontology, and let q be a Boolean CQ with at most one quantified variable. Then $\text{CONS}_{\text{ENT}}(q, \mathcal{T})$ is first-order expressible.*

Proof (Sketch). We show how to construct the desired consistent rewriting of q in the case where q has a single quantified variable x . First, for each $t \in \text{terms}(q)$, we set $C_t = \{A \mid A(t) \in q\}$, and we let Σ_t be the set of all $S \subseteq \mathbf{N}_C$ such that every maximal subset $U \subseteq S$ consistent with \mathcal{T} is such that $\mathcal{T} \models U \sqsubseteq C_t$. Intuitively, Σ_t defines the possible circumstances under which the conjunction of concepts in C_t is consistently entailed. We can express this condition with the first-order formula ψ_t :

$$\psi_t = \bigvee_{S \in \Sigma_t} \left(\bigwedge_{A \in S} A(t) \wedge \bigwedge_{A \in \mathbf{N}_C \setminus S} \neg A(t) \right)$$

Now using the ψ_t , we construct q' :

$$q' = \exists x \bigwedge_{R(t, t') \in q} R(t, t') \wedge \bigwedge_{t \in \text{terms}(q)} \psi_t$$

It can be shown that q' is indeed a consistent rewriting of q w.r.t. \mathcal{T} . To see why this is so, it is helpful to remark that the repairs of $(\mathcal{T}, \mathcal{A})$ contain precisely the role assertions in \mathcal{A} , together with a maximal subset of concept assertions consistent with \mathcal{T} for each individual.

The next example shows that Theorem 1 cannot be extended to the class of queries with two quantified variables.

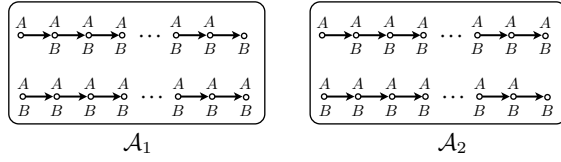


Fig. 1: ABoxes for Example 2. Arrows indicate the role R , and each of the four R -chains has length exceeding 2^k .

Example 2. Consider $q = \exists xy A(x) \wedge R(x, y) \wedge B(y)$ and $\mathcal{T} = \{A \sqsubseteq \neg B\}$. Suppose for a contradiction that q' is a consistent rewriting of q w.r.t. \mathcal{T} , and let k be the quantifier rank of q' . In Figure 1, we give two ABoxes \mathcal{A}_1 and \mathcal{A}_2 , each consisting of two R -chains of length $> 2^k$. It can be verified that q is consistently entailed from $\mathcal{T}, \mathcal{A}_1$. This is because in every repair, the upper chain will have A at one end, B at the other, and either an A or B at all interior points; every such configuration makes q true somewhere along the chain. On the other hand, we can construct a repair for $\mathcal{T}, \mathcal{A}_2$ which does not entail q by always preferring A on the top chain and B on the bottom chain. It follows that the interpretation $\mathcal{I}_{\mathcal{A}_1}$ satisfies q' , whereas $\mathcal{I}_{\mathcal{A}_2}$ does not. However, one can show using standard tools from finite model theory (cf. Ch. 3-4 of [15]) that no formula of quantifier rank k can distinguish $\mathcal{I}_{\mathcal{A}_1}$ and $\mathcal{I}_{\mathcal{A}_2}$, yielding the desired contradiction.

We can generalize the preceding example to obtain sufficient conditions for the inexistence of a consistent rewriting.

Theorem 2. *Let \mathcal{T} be a simple ontology, and let q be a Boolean CQ with two quantified variables x, y . Assume that there do not exist CQs q_1 and q_2 , each with less than two quantified variables, such that $q \equiv q_1 \wedge q_2$. Denote by C_x (resp. C_y) the set of concepts A such that $A(x) \in q$ (resp. $A(y) \in q$). Then $\text{CONSENT}(q, \mathcal{T})$ is not first-order expressible if there exists $S \subseteq \mathbf{N}_C$ such that:*

- for $v \in \{x, y\}$, there is a maximal subset $D_v \subseteq S$ consistent with \mathcal{T} s.t. $\mathcal{T} \not\models D_v \sqsubseteq C_v$
- for every maximal subset $D \subseteq S$ consistent with \mathcal{T} , either $\mathcal{T} \models D \sqsubseteq C_x$ or $\mathcal{T} \models D \sqsubseteq C_y$

Proof (Sketch). The proof generalizes the argument outlined in Example 2. Instead of having a single role connecting successive elements in the chains, we establish the required relational structure for each pair of successive points. We then substitute the set D_y for A , the set D_x for B , and the set S for $\{A, B\}$. The properties of S ensure that if S is asserted at some individual, then we can block the satisfaction of C_x using D_y , and we can block C_y using D_x , but we can never simultaneously block both C_x and C_y . The assumption that q cannot be rewritten as a conjunction of queries with less than two quantified variables is used in the proof of $\mathcal{T}, \mathcal{A}_2 \not\models_{\text{cons}} q$ to show that the only possible matches of q involve successive chain elements (and not constants from the query). To show $\mathcal{I}_{\mathcal{A}_1}$ and $\mathcal{I}_{\mathcal{A}_2}$ cannot be distinguished, we use Ehrenfeucht-Fraïssé games,

rather than Hanf locality, since the latter is inapplicable when there is a role atom containing a constant and a quantified variable.

The following theorem shows that whenever the conditions of Theorem 2 are not met, a consistent rewriting exists.

Theorem 3. *Let \mathcal{T} be a simple ontology, and let q be a Boolean CQ with two quantified variables x, y . Then $\text{CONSENT}(q, \mathcal{T})$ is first-order expressible if q is equivalent to a CQ with at most one quantified variable, or if there is no set S satisfying the conditions of Theorem 2.*

Proof (Sketch). When q is equivalent to a query q' with at most one quantified variable, then Theorem 1 yields a consistent rewriting of q' , and hence of q . Thus, the interesting case is when there is no such equivalent query, nor any set S satisfying the conditions of Theorem 2. Intuitively, the inexistence of such a set S ensures that if at some individual, one can block C_x , and one can block C_y , then it is possible to simultaneously block C_x and C_y (compare this to Example 2 in which blocking A causes B to hold, and vice-versa). This property is key, as it allows different potential query matches to be treated independently.

Together, Theorems 2 and 3 provide a necessary and sufficient condition for the existence of a consistent rewriting. We now reconsider \mathcal{T} and q from Example 2 and outline a polynomial-time method for solving $\text{CONSENT}(q, \mathcal{T})$.

Example 3. Suppose we have an ABox \mathcal{A} , and we wish to decide if $\mathcal{T}, \mathcal{A} \models_{\text{cons}} q$, for $\mathcal{T} = \{A \sqsubseteq \neg B\}$ and $q = \exists xy A(x) \wedge R(x, y) \wedge B(y)$. The basic idea is to try to construct a repair which does not entail q . We start by iteratively applying the following rules until neither rule is applicable: (1) if $R(a, b), A(a), B(a), B(b) \in \mathcal{A}$ but $A(b) \notin \mathcal{A}$, then delete $A(a)$ from \mathcal{A} , and (2) if $R(a, b), A(a), A(b), B(b) \in \mathcal{A}$ but $B(a) \notin \mathcal{A}$, then delete $B(b)$. Note that since the size of \mathcal{A} decreases with every rule application, we will stop after a polynomial number of iterations. Once finished, we check whether there are a, b such that $A(a), R(a, b), B(b) \in \mathcal{A}$, $B(a) \notin \mathcal{A}$, and $A(b) \notin \mathcal{A}$. If so, we return ‘yes’ (to indicate $\mathcal{T}, \mathcal{A} \models_{\text{cons}} q$), and otherwise, we output ‘no’ (for $\mathcal{T}, \mathcal{A} \not\models_{\text{cons}} q$). Note that in the latter case, for all pairs a, b with $A(a), R(a, b), B(b) \in \mathcal{A}$, we have both $B(a)$ and $A(b)$. Thus, we can choose to always keep A , thereby blocking all remaining potential matches.

By carefully generalizing the ideas outlined in Example 3, we obtain a tractability result which covers all queries having at most two quantified variables.

Theorem 4. *Let \mathcal{T} be a simple ontology, and let q be a CQ with at most 2 quantified variables. Then $\text{CONSENT}(q, \mathcal{T})$ is polynomial in data complexity.*

5 An improved coNP lower bound

The objective of this section is to show that the tractability result we obtained for queries with at most two quantified variables cannot be extended further

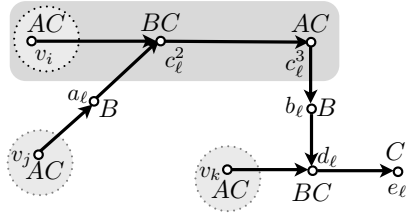


Fig. 2: Abox \mathcal{A}_{c_ℓ} for clause $c_\ell = \neg v_i \vee \neg v_j \vee \neg v_k$.

to the class of conjunctive queries with three quantified variables. We will do this by establishing coNP-hardness for a specific conjunctive query with three quantified variables, thereby improving the lower bound sketched in Example 1. Specifically, we will reduce 3SAT to $\text{CONSENT}(q, \mathcal{T})$ where:

$$\mathcal{T} = \{A \sqsubseteq \neg B, A \sqsubseteq \neg C, B \sqsubseteq \neg C\}$$

$$q = \exists x, y, z A(x) \wedge R(x, y) \wedge B(y) \wedge R(y, z) \wedge C(z).$$

The first component of the reduction is a mechanism for choosing truth values for the variables. For this, we create an ABox $\mathcal{A}_{v_i} = \{A(v_i), C(v_i)\}$ for each variable v_i . It is easy to see that there are two repairs for \mathcal{A}_{v_i} w.r.t. \mathcal{T} : $\{A(v_i)\}$ and $\{C(v_i)\}$. We will interpret the choice of $A(v_i)$ as assigning true to v_i , and the presence of $C(v_i)$ to mean that v_i is false.

Next we need some way of verifying whether a clause is satisfied by the valuation associated with a repair of $\cup_i \mathcal{A}_{v_i}$. To this end, we create an ABox \mathcal{A}_{c_ℓ} for each clause c_ℓ ; the ABox \mathcal{A}_φ encoding φ will then simply be the union of the ABoxes \mathcal{A}_{v_i} and \mathcal{A}_{c_ℓ} . The precise definition of the ABox \mathcal{A}_{c_ℓ} is a bit delicate and depends on the polarity of the literals in c_ℓ . Figure 2 presents a pictorial representation of \mathcal{A}_{c_ℓ} for the case where $c_\ell = \neg v_i \vee \neg v_j \vee \neg v_k$ (the ABoxes \mathcal{A}_{v_i} , \mathcal{A}_{v_j} , and \mathcal{A}_{v_k} are also displayed).

Let us now see how the ABox \mathcal{A}_{c_ℓ} pictured in Figure 2 can be used to test the satisfaction of c_ℓ . First suppose that we have a repair \mathcal{B} of \mathcal{A}_φ which contains $A(v_i)$, $A(v_j)$, and $A(v_k)$, i.e. the valuation associated with the repair does not satisfy c_ℓ . We claim that this implies that q holds. Suppose for a contradiction that q is not entailed from \mathcal{T}, \mathcal{B} . We first note that by maximality of repairs, \mathcal{B} must contain all of the assertions $A(v_j)$, $R(v_j, a_\ell)$, $B(a_\ell)$, and $R(a_\ell, c_\ell^2)$. It follows that including $C(c_\ell^2)$ in \mathcal{B} would cause q to hold, which means we must choose to include $B(c_\ell^2)$ instead. Using similar reasoning, we can see that in order to avoid satisfying q , we must have $C(d_\ell)$ in \mathcal{B} rather than $B(d_\ell)$, which in turn forces us to select $C(c_\ell^3)$ to block $A(c_\ell^3)$. However, this is a contradiction, since we have identified a match for q in \mathcal{B} with $x = v_i$, $y = c_\ell^2$, $z = c_\ell^3$. The above argument (once extended to the other possible forms of \mathcal{A}_{c_ℓ}) is the key to showing that the unsatisfiability of φ implies $\mathcal{T}, \mathcal{A}_\varphi \models q$.

Conversely, it can be proven that if one of c_ℓ 's literals is made true by the valuation, then it is possible to repair \mathcal{A}_{c_ℓ} in such a way that a match for q is avoided. For example, consider again \mathcal{A}_{c_ℓ} from Figure 2, and suppose that

the second literal v_j is satisfied. It follows that $C(v_j) \in \mathcal{B}$, hence $A(v_j) \notin \mathcal{B}$, which means we can keep $C(c_\ell^2)$ rather than $B(c_\ell^2)$, thereby blocking the match at $(v_i, c_\ell^2, c_\ell^3)$. By showing this property holds for the different forms of \mathcal{A}_{c_ℓ} , and by further arguing that we can combine “ q -avoiding” repairs of the \mathcal{A}_{c_ℓ} without inducing a match for q , we can prove that the satisfiability of φ implies $\mathcal{T}, \mathcal{A}_\varphi \not\models q$. We thus have:

Theorem 5. $\text{CONSENT}(q, \mathcal{T})$ is coNP-hard in data complexity for $\mathcal{T} = \{A \sqsubseteq \neg B, A \sqsubseteq \neg C, B \sqsubseteq \neg C\}$ and $q = \exists x, y, z A(x) \wedge R(x, y) \wedge B(y) \wedge R(y, z) \wedge C(z)$.

6 Tractability through approximation

The positive results from Section 4 give us a polynomial algorithm for consistent query answering in the presence of simple ontologies, but only for CQs with at most two quantified variables. In order to be able to handle all queries, we explore in this section alternative inconsistency-tolerant semantics which are sound approximations of the consistent query answering semantics.

One option is to adopt the IAR semantics from [4]. We recall that this semantics (denoted by \models_{IAR}) can be seen as evaluating queries against the ABox corresponding to the *intersection of the repairs*. Conjunctive query answering under IAR semantics was shown in [5] tractable for general CQs in the presence of DL-Lite ontologies (and *a fortiori* simple ontologies) using query rewriting.

To obtain a finer approximation of the consistent query answering semantics, we propose a new inconsistency-tolerant semantics which corresponds to closing repairs with respect to the TBox before intersecting them. In the following definition, we use $cl_{\mathcal{T}}(\mathcal{B})$ to denote the set of assertions entailed from \mathcal{T}, \mathcal{B} .

Definition 4. A Boolean query q is said to be entailed from $(\mathcal{T}, \mathcal{A})$ under ICR semantics (“*intersection of closed repairs*”), written $\mathcal{T}, \mathcal{A} \models_{ICR} q$, if $\mathcal{T}, \mathcal{D} \models q$, where $\mathcal{D} = \bigcap_{\mathcal{B} \in \text{Rep}_{\mathcal{T}}(\mathcal{A})} cl_{\mathcal{T}}(\mathcal{B})$.

The following theorem, which is easy to prove, establishes the relationship among the three semantics.

Theorem 6. For every Boolean CQ q and TBox \mathcal{T} :

$$\mathcal{T}, \mathcal{A} \models_{IAR} q \quad \Rightarrow \quad \mathcal{T}, \mathcal{A} \models_{ICR} q \quad \Rightarrow \quad \mathcal{T}, \mathcal{A} \models_{cons} q$$

The reverse implications do not hold.

The next example illustrates the difference between IAR and ICR semantics:

Example 4. Let $\mathcal{T} = \{A \sqsubseteq C, B \sqsubseteq C, A \sqsubseteq \neg B\}$ and $\mathcal{A} = \{A(a), B(a)\}$. Then $C(a)$ is entailed from $(\mathcal{T}, \mathcal{A})$ under ICR semantics, but not under IAR semantics.

Finally, we show that under ICR semantics, we can answer any conjunctive query in polynomial time using query rewriting.

Theorem 7. *Let \mathcal{T} be a simple ontology and q a Boolean CQ. Then there exists a first-order query q' such that for every ABox \mathcal{A} : $\mathcal{T}, \mathcal{A} \models_{ICR} q$ iff $\mathcal{I}_{\mathcal{A}} \models q'$.*

Proof (Sketch). We first compute, using standard techniques, a union of conjunctive queries φ such that for every \mathcal{A} , we have $\mathcal{T}, \mathcal{A} \models q$ if and only if $\mathcal{I}_{\mathcal{A}} \models \varphi$. Next we use Theorem 1 to find a consistent rewriting $\psi_{A(t)}$ of each concept atom $A(t) \in \varphi$, and we let q' be the first-order query obtained by replacing each occurrence of $A(t)$ in φ by $\psi_{A(t)}$. It can be shown that the query q' is such that $\mathcal{T}, \mathcal{A} \models_{ICR} q$ if and only if $\mathcal{I}_{\mathcal{A}} \models q'$.

7 Extension to inverse-free $DL-Lite_{core}$

In this section, we show how the techniques we developed for simple ontologies can be used to extend our positive results to $DL-Lite_{core}$ ontologies which do not contain inverse roles (we will use $DL-Lite^{no-}$ to refer to this logic).

Our first result shows that the analogues of Theorems 1 and 4 hold for $DL-Lite^{no-}$ ontologies. The main technical difficulty in adapting the proofs of Theorems 1 and 4 is that role assertions may now be contradicted, which means repairs need not have the same set of role assertions as the original ABox.

Theorem 8. *Consider a $DL-Lite^{no-}$ ontology \mathcal{T} , and a Boolean CQ q with at most two quantified variables. Then $\text{CONSENT}(q, \mathcal{T})$ is polynomial in data complexity, and first-order expressible if there is at most one quantified variable.*

We can also extend the general first-order expressibility result for the new ICR semantics (Theorem 7) to the class of $DL-Lite^{no-}$ ontologies.

Theorem 9. *Let \mathcal{T} be a $DL-Lite^{no-}$ ontology, and let q be a Boolean CQ. Then there exists a first-order query q' such that for every ABox \mathcal{A} : $\mathcal{T}, \mathcal{A} \models_{ICR} q$ if and only if $\mathcal{I}_{\mathcal{A}} \models q'$.*

As noted earlier, consistent query answering in (full) $DL-Lite_{core}$ is coNP-hard in data complexity even for instance queries, which means that neither of the preceding theorems can be extended to the class of $DL-Lite_{core}$ ontologies.

8 Conclusion and Future Work

The detailed complexity analysis we conducted for consistent query answering in the presence of simple ontologies provides further insight into the negative complexity results obtained in [4, 6], by making clear how little is needed to obtain first-order inexpressibility or intractability. Our investigation also yielded some positive results, including the identification of novel tractable cases, such as inverse-free $DL-Lite_{core}$ ontologies coupled with CQs with at most two quantified variables (or coupled with arbitrary CQs, under the new ICR semantics).

There are several natural directions for future work. First, it would be interesting to explore how far we can push our positive results. We expect that

adding Horn inclusions and positive role inclusions should be unproblematic, but role disjointness axioms will be more challenging. In order to handle functional roles, we might try to combine our positive results with those which have been obtained for relational databases under functional dependencies [10]. It would also be interesting to try to build upon the results in this paper in order to obtain a criterion for first-order expressibility (or tractability) which applies to all conjunctive queries, regardless of the number of quantified variables.

Finally, we view the present work as a useful starting point in the development of sound but incomplete consistent query answering algorithms for popular lightweight DLs like (full) *DL-Lite_{core}* and \mathcal{EL}_{\perp} . For example, our results could be extended to identify some CQ-TBox pairs in these richer logics for which consistent query answering is tractable. Another idea is to use the new ICR semantics to lift tractability results for IQs (like those in [11]) to classes of CQs.

References

1. Nikitina, N., Rudolph, S., Glimm, B.: Reasoning-supported interactive revision of knowledge bases. In: Proc. of IJCAI. (2011) 1027–1032
2. Arenas, M., Bertossi, L.E., Chomicki, J.: Consistent query answers in inconsistent databases. In: Proc. of PODS, ACM Press (1999) 68–79
3. Chomicki, J.: Consistent query answering: Five easy pieces. In: Proc. of ICDT. (2007) 1–17
4. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant semantics for description logics. In: Proc. of RR. (2010) 103–117
5. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Query rewriting for inconsistent DL-Lite ontologies. In: Proc. of RR. (2011) 155–169
6. Rosati, R.: On the complexity of dealing with inconsistency in description logic ontologies. In: Proc. of IJCAI. (2011) 1057–1062
7. Baader, F., Brandt, S., Lutz, C.: Pushing the \mathcal{EL} envelope. In: Proc. of IJCAI. (2005) 364–369
8. Fuxman, A., Miller, R.J.: First-order query rewriting for inconsistent databases. In: Proc. of ICDT. (2005) 337–351
9. Grieco, L., Lembo, D., Rosati, R., Ruzzi, M.: Consistent query answering under key and exclusion dependencies: algorithms and experiments. In: Proc. of CIKM. (2005) 792–799
10. Wijsen, J.: On the first-order expressibility of computing certain answers to conjunctive queries over uncertain databases. In: Proc. of PODS. (2010) 179–190
11. Bienvenu, M.: First-order expressibility results for queries over inconsistent DL-Lite knowledge bases. In: Proc. of DL. (2011)
12. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning* **39**(3) (2007) 385–429
13. Artale, A., Calvanese, D., Kontchakov, R., Zakharyashev, M.: The DL-Lite family and relations. *Journal of Artificial Intelligence Research* **36** (2009) 1–69
14. Donini, F.M., Lenzerini, M., Nardi, D., Schaerf, A.: Deduction in concept languages: From subsumption to instance checking. *Journal of Logic and Computation* **4**(4) (1994) 423–452
15. Libkin, L.: *Elements of Finite Model Theory*. Springer (2004)

A Omitted Proofs

Theorem 1

Proof. Let \mathcal{T} and q be as stated in the theorem. To simplify the presentation, we assume that q does have a quantified variable, call it x . The proof is similar (but a bit simpler) if there are no quantified variables in q . We first build a formula for each term in q . For $t \in \text{terms}(q)$, we set $C_t = \{A \mid A(t) \in q\}$, and we define Σ_t as the set of all $S \subseteq \mathbf{N}_C$ such that every maximal subset $U \subseteq S$ consistent with \mathcal{T} satisfies $\mathcal{T} \models U \sqsubseteq C_t$. We can then use Σ_t to define the formula ψ_t :

$$\psi_t = \bigvee_{S \in \Sigma_t} \left(\bigwedge_{A \in S} A(t) \wedge \bigwedge_{A \in \mathbf{N}_C \setminus S} \neg A(t) \right)$$

Using the ψ_t , we construct q' :

$$q' = \exists x \bigwedge_{R(t,t') \in q} R(t,t') \wedge \bigwedge_{t \in \text{terms}(q)} \psi_t$$

Our aim is to show that q' is indeed a consistent rewriting of q w.r.t. \mathcal{T} , and hence that $\text{CONSENT}(q, \mathcal{T})$ is first-order expressible.

For the first direction, take some ABox \mathcal{A} with $\mathcal{I}_{\mathcal{A}} \models q'$. Then there must exist an individual a from \mathcal{A} which, when substituted for x in q' , makes all atoms in q' hold in $\mathcal{I}_{\mathcal{A}}$. Let q'_a (resp. q_a) be the ground query obtained by substituting a for x in q' (resp. q). Consider some repair \mathcal{B} of \mathcal{A} w.r.t. \mathcal{T} . Because \mathcal{T} is a simple ontology, there is no way of contradicting a role assertion, and hence \mathcal{B} must contain the same role assertions as \mathcal{A} . As $\mathcal{I}_{\mathcal{A}} \models q'_a$, it follows that all role atoms in q'_a belong to \mathcal{A} , and hence also to \mathcal{B} . As q_a and q'_a have the same role atoms, it follows that all role atoms in q_a appear in \mathcal{B} . Now we wish to show that the concept atoms in q_a are entailed from $(\mathcal{T}, \mathcal{B})$. Since $\mathcal{I}_{\mathcal{A}} \models q'_a$, we know that for every individual name c in q' , the set $\{B \mid B(c) \in \mathcal{A}\}$ belongs to Σ_c , which means that every maximal subset $U \subseteq \{B \mid B(c) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models U \sqsubseteq C_c$. For the same reason, we know the set $\{B \mid B(a) \in \mathcal{A}\}$ belongs to Σ_x , which means every maximal subset $U \subseteq \{B \mid B(a) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models U \sqsubseteq C_x$. To complete the argument, we note that because \mathcal{B} is a repair, and \mathcal{T} does not contain any roles, for each individual b , the set $S_b = \{B \mid B(b) \in \mathcal{B}\}$ must be a maximal subset of $\{B \mid B(b) \in \mathcal{A}\}$ consistent with \mathcal{T} . From this, we can conclude that all concept atoms in q_a are entailed from $(\mathcal{T}, \mathcal{B})$. It follows that q_a , hence q , is consistently entailed from $(\mathcal{T}, \mathcal{A})$.

For the second direction, suppose that $\mathcal{I}_{\mathcal{A}} \not\models q'$. We first consider the case where $\mathcal{I}_{\mathcal{A}} \not\models \psi_c$ for some individual c in q' . If this is the case, then we can find a maximal subset S_c of $\{B \mid B(c) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models S_c \sqsubseteq C_c$. We then let \mathcal{B} be any repair such that $\{B(c) \mid B \in S_c\} \subseteq \mathcal{B}$ (such a repair must exist since S_c is consistent with \mathcal{T}). It is easy to see that $\mathcal{T}, \mathcal{B} \not\models B(c)$ for some $B \in C_c$, which yields $\mathcal{T}, \mathcal{B} \not\models q$. Next suppose instead that $\mathcal{I}_{\mathcal{A}} \models \psi_c$ for every individual c in q' . As $\mathcal{I}_{\mathcal{A}} \not\models q'$, for every individual a , we have $\mathcal{I}_{\mathcal{A}} \not\models q'_a$

(where as above, q'_a denotes q' with a substituted for x). That means that for each individual a , either $\mathcal{I}_A \not\models \alpha_a$ for some role atom $\alpha_a \in q'_a$, or $\mathcal{I}_A \not\models \psi_{x,a}$, where $\psi_{x,a}$ denotes the formula ψ_x with x replaced by a . In the latter case, we can find a maximal subset S_a of $\{B \mid B(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models S_a \subseteq C_x$. We create a repair \mathcal{B} consisting of:

- all role assertions in \mathcal{A}
- for each a such that $\mathcal{I}_A \not\models \varphi_{x,a}$: the assertions $\{B(a) \mid B \in S_a\}$
- for all other individuals a from \mathcal{A} , any maximal subset of $\{B(a) \mid B(a) \in \mathcal{A}\}$ consistent with \mathcal{T}

By construction, the set \mathcal{B} is a maximal subset of \mathcal{A} consistent with \mathcal{T} . We wish to show that $\mathcal{T}, \mathcal{B} \not\models q$. Consider the following interpretation \mathcal{I} defined as follows:

- $\Delta^{\mathcal{I}}$ is the set of individuals from \mathcal{B}
- $a^{\mathcal{I}} = a$, for each individual a from \mathcal{B}
- $A^{\mathcal{I}} = \{d \mid \mathcal{T}, \mathcal{B} \models A(d)\}$
- $R^{\mathcal{I}} = \{(d, e) \mid R(d, e) \in \mathcal{B}\}$

Clearly \mathcal{I} is a model of \mathcal{T}, \mathcal{B} . We need to show that $\mathcal{I} \not\models q_a$ for every individual a . First suppose that a is such that $\mathcal{I}_A \not\models \alpha_a$ for some role atom $\alpha_a \in q'_a$. Then since q_a and q'_a have the same role atoms, and \mathcal{A} and \mathcal{B} agree on role assertions, it follows that $\alpha_a \notin \mathcal{B}$, hence $\mathcal{I} \not\models q_a$. Now consider the other case in which $\mathcal{I}_A \not\models \psi_{x,a}$. Then since S_a is the set of concepts asserted at a in \mathcal{B} , and $\mathcal{T} \not\models S_a \subseteq C_x$, it follows that there is an atom $B(a) \in q_a$ such that $a \notin B^{\mathcal{I}}$, hence $\mathcal{I} \not\models q_a$. We have thus shown that $\mathcal{T}, \mathcal{B} \not\models q$. As a result, we obtain the desired $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. \square

Theorem 2

Proof. Let \mathcal{T} , q , C_x , and C_y be as in the statement of the theorem. Suppose that $S \subseteq \mathbf{N}_C$ is such that:

- for $v \in \{x, y\}$, there is a maximal subset $D_v \subseteq S$ consistent with \mathcal{T} s.t. $\mathcal{T} \not\models D_v \subseteq C_v$
- for every maximal subset $D \subseteq S$ consistent with \mathcal{T} , either $\mathcal{T} \models D \subseteq C_x$ or $\mathcal{T} \models D \subseteq C_y$

Further assume, for a contradiction, that q' is a consistent rewriting of q w.r.t. \mathcal{T} . Let k be the quantifier rank of q' , and fix $m = 2^k + 2$. Let c_1, \dots, c_n be the individuals appearing in q . We wish to create two ABoxes \mathcal{A}_1 and \mathcal{A}_2 such that $\mathcal{T}, \mathcal{A}_1 \models q$ and $\mathcal{T}, \mathcal{A}_2 \models q$. The ABoxes will use fresh individuals a_1, \dots, a_m and b_1, \dots, b_m . They will both have the following role assertions:

- $\{R(c_i, c_j) \mid R(c_i, c_j) \in q\}$
- $\{R(c_i, a_j), R(c_i, b_j) \mid R(c_i, x) \in q, 1 \leq j < m\}$
- $\{R(a_j, c_i), R(b_j, c_i) \mid R(x, c_i) \in q, 1 \leq j < m\}$
- $\{R(c_i, a_j), R(c_i, b_j) \mid R(c_i, y) \in q, 1 < j \leq m\}$
- $\{R(a_j, c_i), R(b_j, c_i) \mid R(y, c_i) \in q, 1 < j \leq m\}$

- $\{R(a_i, a_i)R(b_i, b_i) \mid R(x, x) \in q, 1 \leq j < m\}$
- $\{R(a_i, a_i)R(b_i, b_i) \mid R(y, y) \in q, 1 < j \leq m\}$
- $\{R(a_j, a_{j+1}), R(b_j, b_{j+1}) \mid R(x, y) \in q, 1 \leq j < m\}$
- $\{R(a_{j+1}, a_j), R(b_{j+1}, b_j) \mid R(y, x) \in q, 1 \leq j < m\}$

The ABox \mathcal{A}_1 contains the following concept assertions:

- $\{A(c_i) \mid A(c_i) \in q\}$
- $\{A(a_1) \mid A \in D_y\}$
- $\{A(a_m) \mid A \in D_x\}$
- $\{A(a_i) \mid A \in S, 1 < i < m\}$
- $\{A(b_i) \mid A \in S, 1 \leq i \leq m\}$

The ABox \mathcal{A}_2 contains the concept assertions:

- $\{A(c_i) \mid A(c_i) \in q\}$
- $\{A(a_1) \mid A \in D_y\}$
- $\{A(a_i) \mid A \in S, 1 < i \leq m\}$
- $\{A(b_i) \mid A \in S, 1 \leq i < m\}$
- $\{A(b_m) \mid A \in D_x\}$

We need to show that (a) $\mathcal{I}_{\mathcal{A}_1}$ and $\mathcal{I}_{\mathcal{A}_2}$ are indistinguishable by all first-order formulas of quantifier rank at most k , (b) $\mathcal{I}_{\mathcal{A}_1} \models q'$, and (c) $\mathcal{I}_{\mathcal{A}_2} \models q'$.

We start by showing $\mathcal{I}_{\mathcal{A}_1}$ and $\mathcal{I}_{\mathcal{A}_2}$ to be indistinguishable. For this, we will use the technique of Ehrenfeucht-Fraïssé games (cf. Chapter 3 of [15] for an introduction). We need to show how Duplicator can win the k -round game on the pair of first-order structures $\mathcal{I}_{\mathcal{A}_1}$ and $\mathcal{I}_{\mathcal{A}_2}$. More specifically, we will show that Duplicator can play in such a way as to satisfy the following property (\star) after each round i .

Let $\mathbf{d} = (d_{-n-1}, d_0, d_1, d_2, \dots, d_i)$ be the vector composed of $c_1, \dots, c_n, a_1, a_m$ followed by the sequence of i points played so far in $\mathcal{I}_{\mathcal{A}_1}$, and let $\mathbf{e} = (e_{-n-1}, e_0, e_1, e_2, \dots, e_i)$ be the vector starting by $c_1, \dots, c_n, a_1, b_m$ and finishing with the i points played so far in $\mathcal{I}_{\mathcal{A}_2}$. Then we have the following:

1. if $d_\ell \in \{a_j, b_j\}$, then $e_\ell \in \{a_j, b_j\}$
2. letting $\rho = \max(\{j \mid d_\ell \in \{a_j, b_j\}, d_\ell \neq e_\ell\})$ and $\lambda = \min(\{j \mid d_\ell \in \{a_j, b_j\}, d_\ell = e_\ell\})$, we have $\rho - \lambda > 2^{k-i}$
3. (\mathbf{d}, \mathbf{e}) define a partial isomorphism between $\mathcal{I}_{\mathcal{A}_1}$ and $\mathcal{I}_{\mathcal{A}_2}$

We will proceed by induction on i . The base case is when $i = 0$, i.e. the game has not yet begun. In this case, we have $\mathbf{d} = (c_1, \dots, c_n, a_1, a_m)$ and $\mathbf{e} = (c_1, \dots, c_n, a_1, b_m)$. The first and third conditions clearly hold, and for the second, we remark that $\lambda = 1$ and $\rho = m$, hence $\rho - \lambda = m - 1 = 2^k + 1 > 2^k$.

For the induction step, assume that Duplicator can play so as to satisfy the property for the first i rounds, and we consider what happens in round $i+1$. For simplicity, we will suppose that in round $i+1$ spoiler selects a point d_{i+1} from $\mathcal{I}_{\mathcal{A}_1}$, but the proof is analogous if spoiler picks a point e_{i+1} from $\mathcal{I}_{\mathcal{A}_2}$. There are four cases:

Case 1: $d_{i+1} = d_\ell$ for some $-n - 1 \leq \ell \leq i$.

In this case, Duplicator responds with e_ℓ , and the property trivially holds for round $i + 1$.

Case 2: $d_{i+1} = a_j$ and $d_\ell = b_j$ for some $-1 \leq \ell \leq i$ (but $d_{i+1} \neq d_\ell$ for all $-n - 1 \leq \ell \leq i$).

Then Duplicator plays a_j if $d_\ell = e_\ell$ and otherwise plays b_j . Clearly the first part of the property is satisfied. For the second part, we note that Duplicator's choice ensures that λ and ρ have the same values after rounds i and $i + 1$, and by the induction hypothesis, we had $\rho - \lambda > 2^{k-i}$ following round i . Hence, after round $i + 1$ we have $\rho - \lambda > 2^{k-i} > 2^{k-(i+1)}$. The proof of the third condition is long, so we give only the essential points. We first need to show that $d_{i+1} \in A^{\mathcal{I}\mathcal{A}_1}$ if and only if $e_{i+1} \in A^{\mathcal{I}\mathcal{A}_2}$. If $d_{i+1} = a_1$, then we have $e_{i+1} = b_1$, and we are done, since $\{A \mid A(b_1) \in \mathcal{A}_1\} = \{A \mid A(b_1) \in \mathcal{A}_1\} = S$. If $d_{i+1} = a_m$, then we have $e_{i+1} = b_m$, and $\{A \mid A(a_m) \in \mathcal{A}_1\} = \{A \mid A(b_m) \in \mathcal{A}_1\} = S$. Finally, the other possibility is that d_{i+1} is equal to some a_j with $1 < j < m$. But then both d_{i+1} and e_{i+1} will have the same set of asserted concepts, namely S . For roles, there are several cases to consider, but they are all similar, so we will focus on just one: $(d_{i+1}, d_j) \in R^{\mathcal{I}\mathcal{A}_1}$ for some $-n - 1 \leq j \leq i + 1$. We need to show that $(e_{i+1}, e_j) \in R^{\mathcal{I}\mathcal{A}_2}$. If $d_{i+1} = a_1$, then we know that $e_{i+1} = a_1$, and that d_j is either some constant c_p , a_1 , or a_2 . Now it follows directly from the definition of \mathcal{A}_1 and \mathcal{A}_2 that $R(a_1, c_p) \in \mathcal{A}_1$ if and only if $R(a_1, c_p) \in \mathcal{A}_2$ for every constant c_p . Similarly we have $R(a_1, a_1) \in \mathcal{A}_1$ if and only if $R(a_1, a_1) \in \mathcal{A}_2$. Finally suppose that we have $d_j = a_2$, and so $R(a_1, a_2) \in \mathcal{A}_1$. As currently $\rho - \lambda > 2^{k-i} \geq 1$, it follows that $\rho > 2$, and hence that $e_j = a_2$. It then suffices to remark that the construction of \mathcal{A}_1 and \mathcal{A}_2 ensures $R(a_1, a_2) \in \mathcal{A}_1$ if and only if $R(a_1, a_2) \in \mathcal{A}_2$. The case where $d_{i+1} = a_j$ for some $1 < j \leq m$ can be handled in a similar manner.

Case 3: $d_{i+1} \in \{a_j, b_j\}$ and $d_\ell \notin \{a_j, b_j\}$ for all $-n - 1 \leq \ell \leq i$.

We thus have $j \neq 1$ and $j \neq m$. We only consider the case where $d_{i+1} = a_j$ (the case where $d_{i+1} = b_j$ proceeds analogously). We let λ and ρ be the values computed following round i . If $j \leq \lambda$, then Duplicator plays a_j , and if $j \geq \rho$, Duplicator plays b_j . Note that in both cases the λ and ρ values will not change after round $i + 1$, and so the second part of the property will still be satisfied. The remaining case is when $\lambda < j < \rho$. Since $\rho - \lambda > 2^{k-i}$ from the IH, we know that either $j - \lambda \leq 2^{k-i+1}$ or $\rho - j \leq 2^{k-i+1}$ (but not both). In the former case, Duplicator chooses a_j , and in the latter case, Duplicator picks b_j . In the former case, $j - \lambda \leq 2^{k-i+1}$ together with $\rho - \lambda > 2^{k-i}$ yields $\rho - j > 2^{k-i-1}$. This is exactly what we need since after round $i + 1$ the value of ρ remains unchanged but λ changes to j , and so we get the desired $\rho - \lambda > 2^{k-(i+1)}$. In the second case, $\rho - j \leq 2^{k-i+1}$ and $\rho - \lambda > 2^{k-i}$ yields $j - \lambda > 2^{k-i-1}$. Since after round $i + 1$, λ remains as before but ρ

is replaced by j , we get $\rho - \lambda > 2^{k-(i+1)}$ as required. Thus, the second part of (\star) is satisfied, and clearly, the first part holds as well. To show the third condition, we simply need to use the same ideas as for Case 2. The only interesting situation is role assertions which involve two “chain” individuals, i.e. $R(d_{i+1}, d_\ell)$ or $R(d_\ell, d_{i+1})$ in which $-1 \leq \ell \leq i+1$ with $d_\ell \in \{a_1, \dots, a_m, b_1, \dots, b_m\}$. Given the structure of \mathcal{A}_1 , the only possibilities for d_ℓ are a_{j-1} , a_j , or a_{j+1} . As for Case 2, we can use the fact that $\rho - \lambda > 2$ to infer that $e_\ell \in \{a_{j-1}, a_j, a_{j+1}\}$ if $d_\ell = a_j$, and $e_\ell \in \{b_{j-1}, b_j, b_{j+1}\}$ if $d_\ell = b_j$. This allows us to show that the corresponding role assertion for d_ℓ, e_ℓ holds in $\mathcal{I}_{\mathcal{A}_2}$.

This completes our proof that $\mathcal{I}_{\mathcal{A}_1}$ and $\mathcal{I}_{\mathcal{A}_2}$ are indistinguishable by formulas of quantifier rank at most k . We can thus conclude that $\mathcal{I}_{\mathcal{A}_1} \models q'$ if and only if $\mathcal{I}_{\mathcal{A}_2} \models q'$.

Now we move on to showing that $\mathcal{I}_{\mathcal{A}_1} \models q'$. It is sufficient to show that $\mathcal{T}, \mathcal{A}_1 \models_{cons} q$. Take some repair \mathcal{B} of \mathcal{A}_1 w.r.t. \mathcal{T} . Because \mathcal{T} is a simple ontology, we know that \mathcal{B} contains all role assertions from \mathcal{A} as well as a maximal subset of concept assertions consistent with \mathcal{T} for each individual. We know that for each c_i , the set $\{A \mid A(c_i) \in q\}$ must be consistent with \mathcal{T} , otherwise, we can find a conjunctive query without any quantified variables which is equivalent to q (e.g. the ground query $A(c) \wedge B(c)$ for some A, B such that $\mathcal{T} \models A \sqsubseteq \neg B$). In particular, this means that for every atom $A(c_p) \in q$, we must have $A(c_p) \in \mathcal{B}$, and so all ground concept atoms in q hold in \mathcal{B} . Next we note that since D_y is consistent with \mathcal{T} , \mathcal{B} contains all assertions in $\{A(a_1) \mid A \in D_y\}$. As $\mathcal{T} \models D_y \sqsubseteq C_x$, we obtain $\mathcal{T}, \mathcal{B} \models A(a_1)$ for every $A \in C_x$. Using a similar argument, we get $\mathcal{T}, \mathcal{B} \models A(a_m)$ for every $A \in C_y$. We also know that every maximal subset $D \subseteq S$ consistent with \mathcal{T} is such that either $\mathcal{T} \models D \sqsubseteq C_x$ or $\mathcal{T} \models D \sqsubseteq C_y$. It follows that for each a_i with $1 < i < m$, either $\mathcal{T}, \mathcal{B} \models A(a_i)$ for every $A \in C_x$, or $\mathcal{T}, \mathcal{B} \models A(a_i)$ for every $A \in C_y$. We can thus infer that there must be some $1 \leq i < m$ such that $\mathcal{T}, \mathcal{B} \models A(a_i)$ for every $A \in C_x$ and $\mathcal{T}, \mathcal{B} \models A(a_{i+1})$ for every $A \in C_y$. By the way we have constructed \mathcal{A}_1 , if $R(x, c_p) \in q$, we have $R(a_i, c_p) \in \mathcal{A}$, and if $R(c_p, x) \in q$, we have $R(c_p, a_i) \in \mathcal{A}$. Likewise, if $R(y, c_p) \in q$, we have $R(a_{i+1}, c_p) \in \mathcal{A}$, and if $R(c_p, y) \in q$, we have $R(c_p, a_{i+1}) \in \mathcal{A}$. Additionally, we have that $R(x, y) \in q$ implies $R(a_i, a_{i+1})$, $R(y, x) \in q$ implies $R(a_{i+1}, a_i)$, $R(x, x) \in q$ implies $R(a_i, a_i)$, and $R(y, y) \in q$ implies $R(a_{i+1}, a_{i+1})$. As \mathcal{B} contains all role assertions in \mathcal{A} , it follows that all of these atoms hold in \mathcal{B} . We have thus shown that all atoms in the first-order formula obtained from q by replacing x with a and y with b are entailed from \mathcal{T}, \mathcal{B} . We thus have $\mathcal{T}, \mathcal{B} \models q$, and hence $\mathcal{T}, \mathcal{A} \models_{cons} q$, as desired.

Finally to complete the proof we need to prove $\mathcal{I}_{\mathcal{A}_2} \not\models q'$, which can be shown by proving $\mathcal{T}, \mathcal{A}_2 \not\models_{cons} q$. Consider the ABox $\mathcal{B} \subseteq \mathcal{A}$ consisting of:

- all role assertions from \mathcal{A}
- all concept assertions $A(c_p)$ from \mathcal{A} , for $1 \leq p \leq n$
- all concept assertions $A(a_i)$ with $A \in D_y$, for $1 \leq i \leq m$
- all concept assertions $A(b_i)$ with $A \in D_x$, for $1 \leq i \leq m$

\mathcal{B} is a repair of \mathcal{A} w.r.t. \mathcal{T} since D_y and D_x are both maximal subsets of S consistent with \mathcal{T} , and $\{A \mid A(c_p) \in q\}$ must be satisfiable w.r.t. \mathcal{T} , else one could find a CQ without quantified variables which is equivalent to q . Our objective is to show that $\mathcal{T}, \mathcal{B} \not\models q$. Define \mathcal{I} as follows:

- $\Delta^{\mathcal{I}} = \{a_1, \dots, a_m, b_1, \dots, b_m, c_1, \dots, c_n\}$
- $a_i^{\mathcal{I}} = a_i$, likewise for the b_i and c_i
- $A^{\mathcal{I}} = \{d \mid \mathcal{T}, \mathcal{B} \models A(d)\}$
- $R^{\mathcal{I}} = \{(d, e) \mid R(d, e) \in \mathcal{B}\}$

In other words, the interpretation interprets all individuals as themselves, and makes true precisely the role assertions in \mathcal{B} and the concept assertions entailed from \mathcal{T}, \mathcal{B} . Since $\mathcal{T} \not\models D_y \sqsubseteq C_y$, there is some $A_y \in C_y$ such that $\mathcal{T} \not\models D_y \sqsubseteq A_y$. It follows that for each $1 \leq i \leq m$, we have $a_i \notin A_y^{\mathcal{I}}$. In a similar manner, we can find $A_x \in C_x$ such that for every $1 \leq i \leq m$, $b_i \notin A_x^{\mathcal{I}}$. It follows that no a_i makes the concept atoms for y true, nor can any b_i be used to satisfy the concept atoms for x . As a consequence, we know that for any pair (d, e) such that $d, e \in \{a_1, \dots, a_m, b_1, \dots, b_m\}$ substituting d for x and e for y yields a query which does not hold in \mathcal{I} . Finally, we need to ensure that the remaining individuals c_i cannot be used to make q hold in \mathcal{I} . We know no a_i can be substituted for x , and no b_i can be substituted for y . There are thus three cases to consider: $(d, e) = (a_j, c_i)$, $(d, e) = (c_i, b_j)$, or $(d, e) = (c_i, c_j)$. In the first case, in which $(d, e) = (a_j, c_i)$, it must be the case that for every $R(x, y) \in q$, we have $R(x, c_i) \in q$. Similarly, for every $R(y, x) \in q$, we must have $R(c_i, x) \in q$, and for every $R(y, y) \in q$, we must have $R(c_i, c_i) \in q$. It must also be the case that for every atom $A(y) \in q$, there exists a corresponding atom $B(c_i)$ such that $\mathcal{T} \models B \sqsubseteq A$. Taken together, these facts imply that the query obtained by removing all atoms mentioning y is equivalent to q , and thus we have found an equivalent query with at most one quantified variable. The second case is symmetric and proceeds analogously to the previous one. In the third case, in which $(d, e) = (c_i, c_j)$, we have that: $R(x, y) \in q$ implies $R(c_i, c_j) \in q$, $R(y, x) \in q$ implies $R(c_j, c_i) \in q$, $R(x, x) \in q$ implies $R(c_i, c_i) \in q$, $R(y, y) \in q$ implies $R(c_j, c_j) \in q$, $A(x) \in q$ implies some $B(c_i) \in q$ with $\mathcal{T} \models B \sqsubseteq A$, and $A(y) \in q$ implies some $B(c_j) \in q$ with $\mathcal{T} \models B \sqsubseteq A$. It is not hard to see that by removing all non-ground atoms from q , we obtain a ground CQ which is equivalent to q . Since we know that there does not exist any CQs equivalent to q with less than two quantified variables, it follows that there cannot be a match for q involving one of the c_i . We can thus infer that there is no match for q in the interpretation \mathcal{I} . As \mathcal{I} is a model of \mathcal{T}, \mathcal{B} , it follows that $\mathcal{T}, \mathcal{B} \not\models q$, and hence that $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. \square

Theorem 3

Proof. Let \mathcal{T} and q be as in the statement of the theorem. We start with the case in which q is equivalent to $q_1 \wedge q_2$, where q_1 and q_2 are both CQs with at most one quantified variable. Then we can apply Theorem 1 to q_1 and q_2 to obtain consistent rewritings q'_1 and q'_2 of q_1 and q_2 respectively, given \mathcal{T} . We

then simply remark that $q'_1 \wedge q'_2$ must be a consistent rewriting for $q_1 \wedge q_2$, and hence for q .

Now we consider the more interesting case in which q is not equivalent to any conjunction $q_1 \wedge q_2$ of CQs q_1, q_2 each having less than two quantified variables, and there is no set of concepts S satisfying the conditions of Theorem 2. Note that the former property implies that there is some atom in q which contains both quantified variables x and y . To see why, note that if there is no such atom, we can partition q into q_x which contains all atoms in q which do not mention y , and q_y which contains all atoms which do contain y . It follows from the second property that for every set $S \subseteq \mathbf{N}_C$, if there is a maximal subset $D_x \subseteq S$ consistent with \mathcal{T} s.t. $\mathcal{T} \not\models D_x \sqsubseteq C_x$, and there is a maximal subset $D_y \subseteq S$ consistent with \mathcal{T} s.t. $\mathcal{T} \not\models D_y \sqsubseteq C_y$, then there is a maximal subset $D_{xy} \subseteq S$ consistent with \mathcal{T} s.t. both $\mathcal{T} \not\models D_{xy} \sqsubseteq C_x$ and $\mathcal{T} \not\models D_{xy} \sqsubseteq C_y$. Or in more informal terms, if we can block C_x , and we can block C_y , then we can simultaneously block C_x and C_y .

Interestingly, this property allow us to adopt exactly the same rewriting approach as was used for the single quantified variable case (Theorem 1). As in that proof, for each $t \in \text{terms}(q)$, we set Σ_t equal to the set of all sets $S \subseteq \mathbf{N}_C$ such that every maximal subset $U \subseteq S$ consistent with \mathcal{T} is such that $\mathcal{T} \models U \sqsubseteq C_t$, where $C_t = \{A \mid A(t) \in q\}$. We use the same definition of ψ_t :

$$\psi_t = \bigvee_{S \in \Sigma_t} \left(\bigwedge_{A \in S} A(t) \wedge \bigwedge_{A \in \mathbf{N}_C \setminus S} \neg A(t) \right)$$

We use almost the same definition for q' , just adding the extra quantified variable y :

$$q' = \exists xy \bigwedge_{R(t,t') \in q} R(t,t') \wedge \bigwedge_{t \in \text{terms}(q)} \psi_t$$

It remains to be shown that under our assumptions, q' is indeed a consistent rewriting of q w.r.t. \mathcal{T} . The first direction proceeds almost identically to the single variable case, so we concentrate on the second direction.

For the second direction, suppose that $\mathcal{I}_A \not\models q'$. If $\mathcal{I}_A \not\models \psi_c$ for some individual c in q' , then we can build a repair with $\mathcal{T}, \mathcal{B} \not\models q$ just as in the proof of Theorem 1. So we suppose that $\mathcal{I}_A \models \psi_c$ for every individual c in q' . As $\mathcal{I}_A \not\models q'$, for every pair of individuals (a, b) , we have $\mathcal{I}_A \not\models q'_{a,b}$, where $q'_{a,b}$ denotes q' with x replaced by a and y replaced by b . For each pair (a, b) , it must either be the case that $\mathcal{I}_A \not\models \alpha_{a,b}$ for some role atom $\alpha_{a,b} \in q'_{a,b}$, or that $\mathcal{I}_A \not\models \psi_{x,a}$, or that $\mathcal{I}_A \not\models \psi_{y,b}$ (where $\psi_{x,a}$ is ψ_x with x replaced by a , and similarly for $\psi_{y,b}$). In the second case, we can find a maximal subset $S_{a,x}$ of $\{B \mid B(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models S_{a,x} \sqsubseteq C_x$. In the third case, we can find a maximal subset $S_{b,y}$ of $\{B \mid B(b) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models S_{b,y} \sqsubseteq C_y$. Note that it is possible that both $\mathcal{I}_A \not\models \psi_{x,a}$ and $\mathcal{I}_A \not\models \psi_{y,b}$. If this is the case, then because of our assumption of no sets S satisfying the conditions of Theorem 2, we know that we can also find a maximal subset $S_{a,x,y}$ of $\{B \mid B(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that both $\mathcal{T} \not\models S_{a,x,y} \sqsubseteq C_x$ and $\mathcal{T} \not\models S_{a,x,y} \sqsubseteq C_y$. We can now use these sets to define a repair \mathcal{B} consisting of:

- all role assertions in \mathcal{A}
- for each a such that $\mathcal{I}_{\mathcal{A}} \not\models \psi_{x,a}$ and $\mathcal{I}_{\mathcal{A}} \not\models \psi_{y,a}$: the assertions $\{B(a) \mid B \in S_{a,x,y}\}$
- for each a such that $\mathcal{I}_{\mathcal{A}} \not\models \psi_{x,a}$ and $\mathcal{I}_{\mathcal{A}} \models \psi_{y,a}$: the assertions $\{B(a) \mid B \in S_{a,x}\}$
- for each a such that $\mathcal{I}_{\mathcal{A}} \not\models \psi_{y,a}$ and $\mathcal{I}_{\mathcal{A}} \models \psi_{x,a}$: the assertions $\{B(a) \mid B \in S_{a,y}\}$
- for all other individuals a from \mathcal{A} , any maximal subset of $\{B(a) \mid B(a) \in \mathcal{A}\}$ consistent with \mathcal{T}

By construction, the set \mathcal{B} is a maximal subset of \mathcal{A} consistent with \mathcal{T} , since we kept all role assertions and a maximal subset of concept assertions for each individual. We will now show that $\mathcal{T}, \mathcal{B} \not\models q_{a,b}$ for every pair of individuals (a, b) . The case where (a, b) is such that $\mathcal{I}_{\mathcal{A}} \not\models \alpha_{a,b}$ for some role atom $\alpha_{a,b} \in q'_{a,b}$ is uninteresting, so we consider instead the second case, in which $\mathcal{I}_{\mathcal{A}} \not\models \psi_{x,a}$. Let V be the set of concepts asserted at a in \mathcal{B} . We know from the construction of \mathcal{B} that V is either $S_{a,x}$ or $S_{a,x,y}$, and in both cases, this yields $\mathcal{T} \not\models V \sqsubseteq C_x$. As a consequence, there must be an atom $B(a) \in q_a$ such that $\mathcal{T}, \mathcal{B} \not\models B(a)$, which gives us $\mathcal{T}, \mathcal{B} \not\models q_a$. The final case, in which $\mathcal{I}_{\mathcal{A}} \not\models \psi_{y,b}$, proceeds analogously. It follows that there is no a for which $\mathcal{T}, \mathcal{B} \models q_a$, which means $\mathcal{T}, \mathcal{B} \not\models q$. We can thus conclude that $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. \square

Theorem 4

Proof. Consider the following algorithm which takes as input a simple ontology \mathcal{T} , an ABox \mathcal{A} , and a conjunctive query q with two quantified variables x, y .

1. For each ground atom $\alpha \in q$, output *no* if $\mathcal{T}, \mathcal{A} \not\models_{cons} \alpha$.
2. Set $C_x = \{A \mid A(x) \in q\}$ and $C_y = \{A \mid A(y) \in q\}$.
3. Initialize Σ to the tuples (a, b, ℓ) such that (a, b) satisfy:
 - if $R(x, y) \in q$, then $R(a, b) \in \mathcal{A}$
 - if $R(y, x) \in q$, then $R(b, a) \in \mathcal{A}$
 - if $R(x, x) \in q$, then $R(a, a) \in \mathcal{A}$
 - if $R(y, y) \in q$, then $R(b, b) \in \mathcal{A}$
 - if $R(x, c) \in q$, then $R(a, c) \in \mathcal{A}$
 - if $R(c, y) \in q$, then $R(c, b) \in \mathcal{A}$

and ℓ is defined as follows:

- $\ell = xy$ if there is a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_x \sqsubseteq C_x$, and there is maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_y \sqsubseteq C_y$.
- $\ell = x$ if there is a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_x \sqsubseteq C_x$, but every maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models D_y \sqsubseteq C_y$.
- $\ell = y$ if there is maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_y \sqsubseteq C_y$, but for every maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models D_x \sqsubseteq C_x$.

- $\ell = 0$ if every maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models D_x \sqsubseteq C_x$, and every maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models D_y \sqsubseteq C_y$.
- 4. Apply the following rules until either there is some $(a, b, 0) \in \Sigma$ or there are no applicable rules:
 - R1** If $(a, b, x) \in \Sigma$, then:
 - remove all tuples of the form (a, d, x) or (a, d, xy) from Σ
 - If it is the case that every maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_x \sqsubseteq C_x$ is also such that $\mathcal{T} \models D_x \sqsubseteq C_y$, then:
 - replace each tuple in Σ of the form (d, a, y) by $(d, a, 0)$
 - replace each tuple in Σ of the form (d, a, xy) by (d, a, x)
 - R2** If $(a, b, y) \in \Sigma$, then:
 - remove all tuples of the form (d, b, y) or (d, b, xy) from Σ
 - If it is the case that every maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_y \sqsubseteq C_y$ is also such that $\mathcal{T} \models D_y \sqsubseteq C_x$, then:
 - replace each tuple in Σ of the form (b, d, x) by $(b, d, 0)$
 - replace each tuple in Σ of the form (b, d, xy) by (b, d, y)
- 5. Return *yes* if there is some $(a, b, 0) \in \Sigma$, else return *no*.

Intuitively, at the beginning of Step 4, the set Σ contains all possible substitutions of individuals for x and y which may potentially cause q to hold in a repair of \mathcal{A} . In order to construct a repair which does not entail q , we must block each of these possible matches. The label ℓ indicates which position(s) can be blocked. For instance, the label ‘ x ’ indicates the we can block C_x from holding at the first individual in the tuple but cannot prevent C_y from holding at the second individual; similarly, we use the label 0 to record that we can neither prevent C_x at the first individual nor C_y at the second individual. During Step 4, we block position x whenever that is the only blockable position, and likewise, we block position y whenever it is the only blockable position. We then update Σ by removing any tuples whose possible match has been blocked, and recording which positions are no longer blockable in other tuples.

A simple examination of the definition of Σ and the rules *R1* and *R2* suffices to show that Σ always satisfies the following properties:

- P1 There is at most one tuple (a, b, ℓ) in Σ for each pair (a, b) of individuals from \mathcal{A} .
- P2 If $(a, b, x) \in \Sigma$, then there exists a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_x \sqsubseteq C_x$.
- P3 If $(a, b, y) \in \Sigma$, then there exists a maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_y \sqsubseteq C_y$.

We now proceed to the proof of correctness of the algorithm.

Soundness. Suppose that $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. We wish to show that the algorithm returns *no*. If $\mathcal{T}, \mathcal{A} \not\models_{cons} \alpha$ for some ground atom $\alpha \in q$, then this is immediate. Otherwise, assume all ground atoms are consistently entailed, and let \mathcal{B} be a repair of \mathcal{A} w.r.t. \mathcal{T} such that $\mathcal{T}, \mathcal{B} \not\models q$. Define an interpretation \mathcal{I} as follows:

- $\Delta^{\mathcal{I}}$ is the set of individuals from \mathcal{B}
- $a^{\mathcal{I}} = a$, for each individual a from \mathcal{B}
- $A^{\mathcal{I}} = \{d \mid \mathcal{T}, \mathcal{B} \models A(d)\}$
- $R^{\mathcal{I}} = \{(d, e) \mid R(d, e) \in \mathcal{B}\}$

As \mathcal{I} is the minimal model of \mathcal{T}, \mathcal{B} , and we know $\mathcal{T}, \mathcal{B} \not\models q$, it follows that q does not hold in \mathcal{I} . Now let M be the set of pairs (a, b) such that there is some $(a, b, \ell) \in \Sigma$ at the end of Step 3. We know that for every $(a, b) \in M$, all atoms of the query hold in \mathcal{I} , except possibly atoms of the form $A(x)$ or $A(y)$. It follows that for every $(a, b) \in M$, there is either some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$, or some $A \in C_y$ such that $b \notin A^{\mathcal{I}}$. We define a function $\text{val} : M \rightarrow \{x, y, xy\}$ as follows:

- $\text{val}(a, b) = xy$ if there is both some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$, and some $A \in C_y$ such that $b \notin A^{\mathcal{I}}$.
- $\text{val}(a, b) = x$ if there is some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$, but no $A \in C_y$ such that $b \notin A^{\mathcal{I}}$.
- $\text{val}(a, b) = y$ if there is some $A \in C_y$ such that $b \notin A^{\mathcal{I}}$, but no $A \in C_x$ such that $a \notin A^{\mathcal{I}}$.

We define a partial order \succ of $\{0, x, y, xy\}$ as follows: $xy \succ x$, $xy \succ y$, $x \succ 0$, and $y \succ 0$. Our objective will be to show that the following property always holds for Σ :

$$(\star) \quad \text{if } (a, b, \ell) \in \Sigma, \text{ then } \ell \succeq \text{val}(a, b)$$

This implies that at the beginning of Step 5, there are no tuples $(a, b, 0) \in \Sigma$, and hence the algorithm returns *no*. The proof is by induction on the number of rule applications that have been made. We first note that property (\star) hold for Σ when there have not been any rule applications. This is trivially the case when $\ell = xy$. For (a, b) such that $(a, b, x) \in \Sigma$ before any rule applications, we know from the way Σ was initialized in Step 3 that every maximal subset $D_b \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models D_b \sqsubseteq C_y$. It follows that there is no $A \in C_y$ such that $b \notin A^{\mathcal{I}}$. As $\mathcal{I} \not\models q$, there must be some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$, and so we have $\text{val}(a, b) = x$. The case of (a, b) such that (a, b, y) belongs to the original Σ proceeds analogously. Now suppose that (\star) holds when there have been at most n rule applications, and consider the $n+1$ th rule application. If it is *R1* which is triggered by some tuple $(a, b, x) \in \Sigma$, and the extra condition in *R1* is not met, then we only remove elements from Σ , and so the property (\star) continues to hold. Consider then the more interesting case in which the condition is satisfied, which may cause some tuple (d, a, xy) to be replaced by (d, a, x) , or (d, a, y) to be replaced by $(d, a, 0)$. Because $(a, b, x) \in \Sigma$ at the start of the rule application, we know that $x \succeq \text{val}(a, b)$, and hence $\text{val}(a, b) = x$. It follows that there must be some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$. However, the condition tells us that any maximal subset D_x of $\{A \mid A(a) \in \mathcal{A}\}$ such that $\mathcal{T} \not\models D_x \sqsubseteq C_x$ must also satisfy $\mathcal{T} \models D_x \sqsubseteq C_y$. It follows that for every $A \in C_y$, we have $a \in A^{\mathcal{I}}$. Then since $\mathcal{I} \not\models q$, it follows that for every $(d, a) \in M$, there must be some $A \in C_x$ such that $d \notin A^{\mathcal{I}}$. This means that for every $(d, a) \in M$, we have $\text{val}(d, a) = x$. Thus, from the induction hypothesis,

we know that at the start of the rule application, every tuple $(d, a, \ell') \in \Sigma$ must be such that $\ell' = xy$ or $\ell' = x$. In the former case, we will replace (d, a, xy) by (d, a, x) , which preserves property (\star) . The case where it is R2 which is applied in the $n + 1$ -th rule application proceeds analogously.

Completeness. To show completeness, suppose that the algorithm returns *no*. We need to show that $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. First consider the case where *no* is returned at Step 1. In this case, there is some ground atom $\alpha \in q$ such that $\mathcal{T}, \mathcal{A} \not\models_{cons} \alpha$, which implies also that $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. So let us now consider the more interesting case in which all ground atoms are consistently entailed, and *no* is returned only in Step 5. We will use the rule applications from Step 4 to guide the construction of a repair. First, we note that each application of a rule during Step 4 is concerned with some particular individual. We will say that this individual is the *target* of the rule application. Note that each individual can be the target of at most one rule application, since applying the rule means removing all tuples containing the targeted individual from Σ . Also note that all tuples in Σ at the end of Step 4 must either have label 0 or xy , since any tuple with label x or y makes one of the rules applicable. As we know the algorithm returns *no* in Step 5, there must only be tuples labelled xy . We now proceed to the definition of a repair \mathcal{B} . We include in \mathcal{B} the role assertions in \mathcal{A} together with a maximal subset of concept assertions for each individual a , determined as follows.

- Case 1: a is the target of an application of R1, and the extra condition was not satisfied.
Then property P2 and the non-satisfaction of the extra rule condition ensures that we can find a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that both $\mathcal{T} \not\models D_x \sqsubseteq C_x$ and $\mathcal{T} \not\models D_x \sqsubseteq C_y$. We choose the set of assertions $\{A(a) \mid A \in D_x\}$ for a .
- Case 2: a is the target of an application of R1, and the extra condition was satisfied.
By property P2, we know that we can find a maximal subset D_x of $\{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_x \sqsubseteq C_x$. We let D_x be any such set, and include in \mathcal{B} the set of assertions $\{A(a) \mid A \in D_x\}$.
- Case 3: a is the target of an application of R2, and the extra condition was not satisfied.
Then property P3 and the non-satisfaction of the extra rule condition ensures that we can find a maximal subset $D_y \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that both $\mathcal{T} \not\models D_y \sqsubseteq C_y$ and $\mathcal{T} \not\models D_y \sqsubseteq C_x$. We choose the set of assertions $\{A(a) \mid A \in D_y\}$ for a .
- Case 4: a is the target of an application of R2, and the extra condition was satisfied.
By property P3, there is a maximal subset D_y of $\{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_y \sqsubseteq C_y$. We let D_y be any such set, and include in \mathcal{B} the set of assertions $\{A(a) \mid A \in D_y\}$.
- Case 5: a appears in some tuple of Σ at the end of Step 4, and has not been the target of any rule application.

By above, we know the tuple in Σ containing a at the end of Step 4 must have label ‘ xy ’, which means that this tuple was originally given the label ‘ xy ’ in Step 3. Thus, we can find a maximal subset D_x of $\{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models D_x \sqsubseteq C_x$. We choose the set of assertions $\{A(a) \mid A \in D_x\}$.

- Case 6: a is neither the target of a rule application, nor appears in any tuple at the end of Step 4.

We choose an arbitrary maximal subset of $\{A(a) \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} .

Since \mathcal{B} contains a maximal subset of $\{A(a) \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} for each individual a , it is clear that \mathcal{B} is a repair of \mathcal{A} w.r.t. \mathcal{T} . We let \mathcal{I} be the following interpretation:

- $\Delta^{\mathcal{I}}$ is the set of individuals from \mathcal{B}
- $a^{\mathcal{I}} = a$, for each individual a from \mathcal{B}
- $A^{\mathcal{I}} = \{d \mid \mathcal{T}, \mathcal{B} \models A(d)\}$
- $R^{\mathcal{I}} = \{(d, e) \mid R(d, e) \in \mathcal{B}\}$

We wish to show that q does not hold in \mathcal{I} . Suppose for a contradiction that q does hold when x is replaced by d and y is replaced by e . It follows that the pair (d, e) satisfies the conditions in Step 3, so some tuple (d, e, ℓ) will be added to Σ . The first possibility is that there is a rule application during Step 4 which targets either d or e and causes the removal of a tuple (d, e, ℓ') from Σ (possibly with $\ell \neq \ell'$). First, suppose that (d, e, ℓ') is removed when applying R1 to d . Then d belongs to Case 1 or Case 2, and so $D = \{A \mid A(d) \in \mathcal{B}\}$ is such that $\mathcal{T} \not\models D \sqsubseteq C_x$. If instead the removal is due to applying R2 to e , then e belongs to either Case 3 or Case 4, and $D = \{A \mid A(e) \in \mathcal{B}\}$ satisfies $\mathcal{T} \not\models D \sqsubseteq C_y$. Thus, we either show that d cannot be substituted for x , or that e cannot be substituted for y , contradicting our assumption that (d, e) defines a match of q in \mathcal{I} . Let us now consider the other possibility, which is that Σ contains a tuple (d, e, ℓ') after Step 4. We know from above that $\ell' = xy$. If d was not the target of any rule application, then we fall into Case 5, and we have that $\{A \mid A(d) \in \mathcal{B}\} = D_x$ for some set D_x satisfying $\mathcal{T} \not\models D_x \sqsubseteq C_x$. It is also possible that d was the target of an application of R2, but the condition was not satisfied (Case 3), in which case $\{A \mid A(d) \in \mathcal{B}\} = D_y$ for some set D_y satisfying both $\mathcal{T} \not\models D_y \sqsubseteq C_y$ and $\mathcal{T} \not\models D_y \sqsubseteq C_x$. Either way, we can infer that there is some $A \in C_x$ such that $d \notin A^{\mathcal{I}}$, contradicting our assumption that (d, e) defines a match of q . We have thus found a model \mathcal{I} of \mathcal{T}, \mathcal{B} such that $\mathcal{I} \not\models q$. It follows that $\mathcal{T}, \mathcal{B} \not\models q$, and hence $\mathcal{T}, \mathcal{A} \not\models_{cons} q$.

Complexity. Note that for Step 1, we can use the query rewriting approach from Theorem 1 to decide consistent entailment for the ground atoms in polynomial time in $|\mathcal{A}|$. It is easy to see that the construction of Σ in Step 3 and the application of a rule during Step 4 can both take only polynomial time. Finally, we note that the total number of rule applications in Step 4 cannot exceed $|\mathcal{A}| \times |\mathcal{A}|$ since each rule application decreases the cardinality of the set $\{(a, b) \mid (a, b, \ell) \in \Sigma \text{ for some } \ell\}$. \square

Theorem 5

Proof. We reduce the well-known coNP-complete problem 3SAT to $\text{CONSENT}(q, \mathcal{T})$ where

$$\mathcal{T} = \{A \sqsubseteq \neg B, A \sqsubseteq \neg C, B \sqsubseteq \neg C\}$$

$$q = \exists x, y, z A(x) \wedge R(x, y) \wedge B(y) \wedge R(y, z) \wedge C(z).$$

Consider some instance $\varphi = c_1 \wedge \dots \wedge c_m$ of 3SAT over variables v_1, \dots, v_n . We need to create an ABox \mathcal{A}_φ such that φ is satisfiable if and only if $\mathcal{T}, \mathcal{A}_\varphi \not\models_{\text{cons}} q$. To this end, we show how to create ABoxes which encode each variable and clause. For each v_i , we create the ABox

$$\mathcal{A}_{v_i} = \{A(v_i), C(v_i)\}$$

Intuitively, \mathcal{A}_{v_i} serves to select a truth value for v_i . We remark that every repair of \mathcal{A}_{v_i} must contain precisely one of $A(v_i)$ and $C(v_i)$. When a repair keeps A , we will interpret this as assigning the value T (true) to v_i , and when C is kept, this signifies that v_i receives the value F (false).

The definition of \mathcal{A}_{c_ℓ} for a clause c_ℓ will depend on the polarity of the literals in c_ℓ . In order to cut the number of cases from 8 to 4, we can assume w.l.o.g. that if c_ℓ has one negative literal and two positive literals, then the negative literal is the first literal in c_ℓ . Likewise, we will suppose that if c_ℓ has one positive literal and two negative literals, then the positive literal is the third literal in c_ℓ . Thus there are four possible shapes for c_ℓ : $c_\ell = v_i \vee v_j \vee v_k$, $c_\ell = \neg v_i \vee v_j \vee v_k$, $c_\ell = \neg v_i \neg v_j \vee v_k$, and $c_\ell = \neg v_i \vee \neg v_j \vee \neg v_k$. Figure 3 provides a pictorial representation of the ABox \mathcal{A}_{c_ℓ} for all four cases. The arrows indicate the role R . Note that if the variable v_i appears in one of c_ℓ 's literals, then the ABox \mathcal{A}_{v_i} is a subset of \mathcal{A}_{c_ℓ} . Also note that all individuals appearing in \mathcal{A}_{c_ℓ} which are not of the form v_i are specific to \mathcal{A}_{c_ℓ} , i.e. they do not appear in \mathcal{A}_{c_p} for any $p \neq \ell$. The general ABox \mathcal{A}_φ representing φ will be the union of these ABoxes:

$$\mathcal{A}_\varphi = \bigcup_{1 \leq \ell \leq m} \mathcal{A}_{c_\ell}$$

We need to show that \mathcal{A}_φ satisfies the desired property, that is, $\mathcal{T}, \mathcal{A}_\varphi \not\models_{\text{cons}} q$ if and only if φ is satisfiable.

For the first direction, suppose that φ is satisfiable, and consider some particular satisfying valuation $\text{val} : \{v_1, \dots, v_n\} \rightarrow \{T, F\}$. We will use this valuation to construct a repair \mathcal{B} of \mathcal{A}_φ such that $\mathcal{T}, \mathcal{B} \not\models q$. The set \mathcal{B} will contain all role assertions from \mathcal{A} , as well as all concept assertions $D(a)$ such that D is the only concept asserted at a in \mathcal{B} , since all of these assertions must belong to every repair. Then for each variable v_i , we include $A(v_i)$ in \mathcal{B} if $\text{val}(v_i) = T$, and otherwise, we include $C(v_i)$. Finally, to define the remaining concept assertions, we consider each subABox \mathcal{A}_{c_ℓ} separately. This is unproblematic because of the following property: if we have repairs $\mathcal{B}_{c_1}, \dots, \mathcal{B}_{c_m}$ which all agree with \mathcal{B} on the concept assertions at the individuals v_i , then the set $\bigcup_{\ell=1}^m \mathcal{B}_{c_\ell}$ is a repair of \mathcal{A} .

To see why this is so, we note that the ABoxes interact solely at the v_i , which have concept A or C . Thus it is not possible for a match to “crossover” from one \mathcal{A}_{c_ℓ} to some other $\mathcal{A}_{c_{\ell'}}$. It thus remains to be shown that for every clause c_ℓ we can find a repair \mathcal{B}_{c_ℓ} which contains $A(v_i)$ if $\text{val}(v_i) = T$, and $C(v_i)$ otherwise, and is such that q does not hold. There are many cases to consider, since we need to account for the four different forms for the ABox \mathcal{A}_{c_ℓ} , as well as which literals in c_ℓ are satisfied by val . We give the details just for one form of clause: $c_\ell = \neg v_i \vee \neg v_j \vee \neg v_k$. We consider the seven possible valuations of v_i, v_j, v_k which satisfy c_ℓ , and for each, we list a set of concept assertions which give rise to a repair in which q does not hold:

- $\text{val}(v_i) = F, \text{val}(v_j) = T, \text{val}(v_k) = T$:
fixed: $C(v_i), A(v_j), A(v_k)$
choose: $B(c_\ell^2), C(c_\ell^3), C(g_\ell)$
- $\text{val}(v_i) = T, \text{val}(v_j) = F, \text{val}(v_k) = T$:
fixed: $A(v_i), C(v_j), A(v_k)$
choose: $C(c_\ell^2), C(c_\ell^3), C(g_\ell)$
- $\text{val}(v_i) = T, \text{val}(v_j) = T, \text{val}(v_k) = F$:
fixed: $A(v_i), A(v_j), C(v_k)$
choose: $B(c_\ell^2), A(c_\ell^3), B(g_\ell)$
- $\text{val}(v_i) = T, \text{val}(v_j) = F, \text{val}(v_k) = F$:
fixed: $A(v_i), C(v_j), C(v_k)$
choose: $C(c_\ell^2), C(c_\ell^3), C(g_\ell)$
- $\text{val}(v_i) = F, \text{val}(v_j) = F, \text{val}(v_k) = T$:
fixed: $C(v_i), C(v_j), A(v_k)$
choose: $C(c_\ell^2), C(c_\ell^3), C(g_\ell)$
- $\text{val}(v_i) = F, \text{val}(v_j) = T, \text{val}(v_k) = F$:
fixed: $C(v_i), A(v_j), C(v_k)$
choose: $B(c_\ell^2), C(c_\ell^3), C(g_\ell)$
- $\text{val}(v_i) = F, \text{val}(v_j) = F, \text{val}(v_k) = F$:
fixed: $C(v_i), C(v_j), C(v_k)$
choose: $C(c_\ell^2), C(c_\ell^3), C(g_\ell)$

It can be verified in a similar manner that for each of the other three structures of c_ℓ , we can also construct a repair of \mathcal{A}_{c_ℓ} with the required properties. Then, as explained above, we can take the union of the repairs of the \mathcal{A}_{c_ℓ} to obtain a repair of \mathcal{A}_φ w.r.t. \mathcal{T} which does not make q true. We can thus conclude that $\mathcal{T}, \mathcal{A}_\varphi \not\models_{\text{cons}} q$.

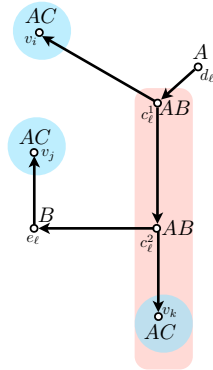
For the second direction, suppose that there is a repair \mathcal{B} of \mathcal{A}_φ such that $\mathcal{T}, \mathcal{B} \models q$. We need to show that φ is satisfiable. Note that because \mathcal{T} contains only concept disjointness axioms, the interpretation $\mathcal{I}_\mathcal{B}$ is a model of \mathcal{T}, \mathcal{B} , and moreover, it is a minimal model for \mathcal{T}, \mathcal{B} . We now that since \mathcal{B} is a repair, it contains either $A(v_i)$ or $C(v_i)$ (but not both). We define a valuation $\text{val} : \{v_1, \dots, v_n\} \rightarrow \{T, F\}$ by setting $\text{val}(v_i) = T$ if $A(v_i) \in \mathcal{B}$, and $\text{val}(v_i) = F$ if $C(v_i) \in \mathcal{B}$. It remains to be shown that val satisfies φ . To do so, we suppose for a contradiction that val does not satisfy c_ℓ , and then we show that if this were the case, then we would have $\mathcal{I}_\mathcal{B} \models q$. There are four cases depending on the

structure of φ . As the ideas are similar throughout, we provide the details for just one case: $c_\ell = \neg v_i \vee \neg v_j \vee \neg v_k$. We first note that since we have assumed val does not satisfy c_ℓ , we have $\text{val}(v_i) = T$, $\text{val}(v_j) = T$, and $\text{val}(v_k) = T$. It follows from the definition of val that \mathcal{B} contains $A(v_i)$, $A(v_j)$, and $A(v_k)$. If $C(c_\ell^2) \in \mathcal{B}$, then we would have a match for q (using $x = v_j$, $y = e_\ell$, and $z = c_\ell^2$). As we know there are no matches for q , it follows that $B(c_\ell^2) \in \mathcal{B}$ (remember that one of $B(c_\ell^2)$ and $C(c_\ell^2)$ must belong to \mathcal{B} , by maximality). Now we consider c_ℓ^3 . We remark that if $B(g_\ell) \in \mathcal{B}$ then there is a match for q (using the substitution $x = v_k$, $y = g_\ell$, $z = f_\ell$). Thus, we must have $C(g_\ell) \in \mathcal{B}$. But this in turn means that $A(c_\ell^3) \in \mathcal{B}$ would cause q to hold, so we must have $C(c_\ell^3)$. However, we know have $A(v_i)$, $B(c_\ell^2)$, and $C(c_\ell^3)$ in \mathcal{B} , and also $R(c_\ell^1, c_\ell^2)$ and $R(c_\ell^2, c_\ell^3)$, since \mathcal{B} must have the same role assertions as \mathcal{A} . This means that $\mathcal{T}, \mathcal{B} \models q$, contradicting our earlier assumption to the contrary. Since assuming that val does not satisfy some c_ℓ leads to a contradiction, we can conclude that val satisfies all clauses, and so φ is satisfiable. \square

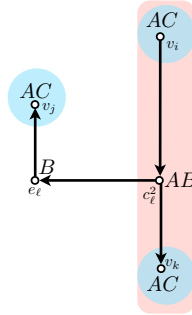
Theorem 7

Proof. Suppose we have a simple ontology \mathcal{T} and a Boolean CQ q . Then because every simple ontology is a *DL-Lite_{core}* ontology, we can use existing algorithms for *DL-Lite_{core}* to find a union of conjunctive queries q' such that for every \mathcal{A} , we have $\mathcal{T}, \mathcal{A} \models q$ if and only if $\mathcal{I}_{\mathcal{A}} \models q'$. Let $A(t)$ be an atom in one of the CQs of q' . If t is an individual, then $A(t)$ is a Boolean CQ with no quantified variables, so using Theorem 1, we can find a consistent rewriting $\varphi_{A(t)}$ of $A(t)$. For atoms $A(t)$ such that t is a variable, we can use Theorem 1 to find a consistent rewriting $\varphi_{A(c)}$ of $A(c)$ (where c is a fresh individual). Then it is easily verified that the formula $\varphi_{A(t)}$ obtained by substituting t for c in $\varphi_{A(c)}$ is a consistent rewriting of the non-Boolean query $A(t)$, i.e. we have that for every individual a , $\mathcal{T}, \mathcal{A} \models_{\text{cons}} A(a)$ if and only if $\mathcal{I}_{\mathcal{A}} \models \varphi_{A(t)}^a$, where $\varphi_{A(t)}^a$ denotes the Boolean query obtained by replacing t with a in $\varphi_{A(t)}$. Now we let ψ be the first-order formula obtained by replacing each atom $A(t)$ by $\varphi_{A(t)}$. We want to show that $\mathcal{T}, \mathcal{A} \models_{\text{ICR}} q$ if and only if $\mathcal{I}_{\mathcal{A}} \models \psi$.

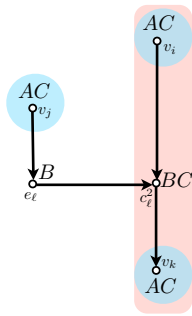
For the first direction, suppose that $\mathcal{T}, \mathcal{A} \models_{\text{ICR}} q$. We want to show that $\mathcal{I}_{\mathcal{A}} \models \psi$. Because $\mathcal{T}, \mathcal{A} \models_{\text{ICR}} q$, we know that $\mathcal{T}, \mathcal{B} \models q$, where $\mathcal{B} = \{\alpha \mid \mathcal{T}, \mathcal{A} \models_{\text{cons}} \alpha\}$ (since the assertions in the intersection of closed repairs are precisely those assertions which are consistently entailed). It follows then that $\mathcal{I}_{\mathcal{B}} \models q'$, and hence there is some CQ σ which is a disjunct of the UCQ q' such that $\mathcal{I}_{\mathcal{B}} \models \sigma$. If σ has variables v_1, \dots, v_k , let (a_1, \dots, a_k) be a substitution for v_1, \dots, v_k which makes σ hold in $\mathcal{I}_{\mathcal{B}}$. Let σ' be the disjunct of ψ which was obtained by replacing each atom $A(t)$ in σ by $\varphi_{A(t)}$. Note that σ' contains exactly the same variables, namely v_1, \dots, v_k . Suppose for a contradiction that $\mathcal{I}_{\mathcal{A}} \not\models \sigma'$. This means in particular that (a_1, \dots, a_k) does not define a match for σ' in $\mathcal{I}_{\mathcal{A}}$. It follows that one of the conjuncts of σ' is not satisfied in $\mathcal{I}_{\mathcal{A}}$ under the substitution (a_1, \dots, a_k) . There are two types of conjuncts: role atoms which were conjuncts of the CQ σ , and formulas of the form $\varphi_{A(t)}$ where $A(t)$ is a conjunct of σ . First suppose that there is a role atom conjunct α which does not hold. Then this means that the role assertion obtained by applying the



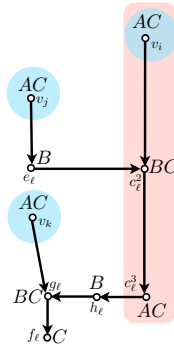
(a) $c_\ell = v_i \vee v_j \vee v_k$



(b) $c_\ell = \neg v_i \vee v_j \vee v_k$



(c) $c_\ell = \neg v_i \vee \neg v_j \vee v_k$



(d) $c_\ell = \neg v_i \vee \neg v_j \vee \neg v_k$

nnn

Fig. 3: ABox A_{c_ℓ}

substitution (a_1, \dots, a_k) to α does not belong to \mathcal{A} . But in this case, this same role assertion will not belong to \mathcal{B} , so this conjunct of σ will not hold in $\mathcal{I}_{\mathcal{B}}$ under substitution (a_1, \dots, a_k) , contradicting our assumption to the contrary. Thus, it must be the case that there is some conjunct $\varphi_{A(t)}$ of σ' which is not satisfied in $\mathcal{I}_{\mathcal{A}}$ by (a_1, \dots, a_k) . This means that the assertion α obtained from $A(t)$ by applying the substitution (a_1, \dots, a_k) is not consistently entailed from \mathcal{T}, \mathcal{A} . It follows that $\alpha \notin \mathcal{B}$, so the atom α , which is a conjunct of σ , does not hold in $\mathcal{I}_{\mathcal{B}}$ under substitution (a_1, \dots, a_k) , again contradicting our earlier assumption. Thus, we can conclude that $\mathcal{I}_{\mathcal{A}} \models \sigma'$, and hence $\mathcal{I}_{\mathcal{A}} \models \psi$.

For the other direction, suppose that $\mathcal{I}_{\mathcal{A}} \models \psi$. We wish to show that $\mathcal{T}, \mathcal{A} \models_{ICR} q$. As $\mathcal{I}_{\mathcal{A}} \models \psi$, there must be some disjunct σ' of ψ and a substitution (a_1, \dots, a_k) for the variables v_1, \dots, v_k appearing in σ' such that the ground CQ τ obtained by applying the substitution (a_1, \dots, a_k) to σ' holds in $\mathcal{I}_{\mathcal{A}}$. It follows that all role assertions which are conjuncts of τ belong to \mathcal{A} . It also follows that all concept assertions which are conjuncts of τ are consistently entailed from \mathcal{T}, \mathcal{A} . Now let $\mathcal{B} = \{\alpha \mid \mathcal{T}, \mathcal{A} \models_{cons} \alpha\}$. We have that all conjuncts of the ground CQ τ belong to \mathcal{B} . Let σ be the CQ in the UCQ q' such that σ' is the formula obtained by replacing each atom $A(t)$ in σ by $\varphi_{A(t)}$. Note that σ' contains exactly the same variables, namely v_1, \dots, v_k . We claim that $\mathcal{I}_{\mathcal{B}} \models \sigma$, and this is witnessed by the substitution (a_1, \dots, a_k) . First take some role atom which is a conjunct of σ . Then we have already seen that the role assertion obtained by applying the substitution (a_1, \dots, a_k) belongs to \mathcal{B} , so this conjunct of σ holds in $\mathcal{I}_{\mathcal{B}}$. Next take some concept atom which is a conjunct of σ . Then the concept assertion obtained by applying the substitution (a_1, \dots, a_k) to this atom is one of the conjuncts of τ . It follows from what we showed above, that this assertion belongs to \mathcal{B} , and hence holds in $\mathcal{I}_{\mathcal{B}}$. We thus have $\mathcal{I}_{\mathcal{B}} \models \sigma$, and hence $\mathcal{I}_{\mathcal{B}} \models q'$. From this we can conclude that $\mathcal{T}, \mathcal{B} \models q$, and as a consequence, $\mathcal{T}, \mathcal{A} \models_{ICR} q$. \square

We prove the two statements in Theorem 8 separately, starting with the first-order expressibility result.

Proposition 1. *Consider a DL-Lite^{no-} ontology \mathcal{T} , and a Boolean CQ q with at most one quantified variable. Then $\text{CONSENT}(q, \mathcal{T})$ is first-order expressible.*

Proof. Let \mathcal{T} and q be as stated in the theorem. We consider only the case where q has one quantified variable, call it x . The proof is easily modified to handle the case where there are no quantified variables in q . Our construction of the rewriting q' will be similar to that given for the case of simple ontologies. We first build a formula for each term in q . For $t \in \text{terms}(q)$, we set $C_t = \{A \mid A(t) \in q\}$, and $R_t = \{\exists R \mid \text{some atom } R(t, t') \in q\}$. We then define Σ_t as the set of all $S \subseteq \mathbf{N}_{\mathcal{C}} \cup \{\exists R \mid R \in \mathbf{N}_{\mathcal{R}}\}$ such that every maximal subset $U \subseteq S$ consistent with \mathcal{T} satisfies (a) $\mathcal{T} \models U \sqsubseteq C_t$, and (b) $R_t \subseteq U$. We next define the formula ψ_t using Σ_t :

$$\psi_t = \bigvee_{S \in \Sigma_t} \left(\bigwedge_{A \in S} A(t) \wedge \bigwedge_{A \in \mathbf{N}_{\mathcal{C}} \setminus S} \neg A(t) \wedge \bigwedge_{\exists R \in S} \exists u R(t, u) \wedge \bigwedge_{\exists R \notin S} \neg \exists u R(t, u) \right)$$

Using the ψ_t , we construct q' :

$$q' = \exists x \bigwedge_{R(t,t') \in q} R(t,t') \wedge \bigwedge_{t \in \text{terms}(q)} \psi_t$$

We need to show that q' is indeed a consistent rewriting of q w.r.t. \mathcal{T} , which gives us the first-order expressibility of $\text{CONSENT}(q, \mathcal{T})$.

For the first direction, take some ABox \mathcal{A} with $\mathcal{I}_{\mathcal{A}} \models q'$. Then there must exist an individual a from \mathcal{A} which, when substituted for x in q' , makes all atoms in q' hold in $\mathcal{I}_{\mathcal{A}}$. Let q'_a (resp. q_a) be the query obtained by substituting a for x in q' (resp. q). Consider some repair \mathcal{B} of \mathcal{A} w.r.t. \mathcal{T} . We need to show that all atoms in q_a are entailed from $(\mathcal{T}, \mathcal{B})$. Since $\mathcal{I}_{\mathcal{A}} \models q'_a$, we know that for every individual name c in q' , the set $\{B \mid B(c) \in \mathcal{A}\} \cup \{\exists R \mid R(c, d) \in \mathcal{A}\}$ belongs to Σ_c , which means that every maximal subset $U \subseteq \{B \mid B(c) \in \mathcal{A}\} \cup \{\exists R \mid R(c, d) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models U \sqsubseteq C_c$ and $R_c \subseteq U$. For the same reason, we know the set $\{B \mid B(a) \in \mathcal{A}\} \cup \{\exists R \mid R(a, d) \in \mathcal{A}\}$ belongs to Σ_x , which means every maximal subset $U \subseteq \{B \mid B(a) \in \mathcal{A}\} \cup \{\exists R \mid R(c, d) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models U \sqsubseteq C_x$ and $R_x \subseteq U$. We then note that because \mathcal{B} is a repair, for each individual b , the set $S_b = \{B \mid B(b) \in \mathcal{B}\} \cup \{\exists R \mid R(c, d) \in \mathcal{B}\}$ must be a maximal subset of $\{B \mid B(b) \in \mathcal{A}\} \cup \{\exists R \mid R(c, d) \in \mathcal{A}\}$ consistent with \mathcal{T} . From this, we can conclude that all concept atoms in q_a are entailed from $(\mathcal{T}, \mathcal{B})$. For a role assertion $R(t, t') \in q_a$, we know from the definition of q' that $R(t, t')$ is a conjunct of q'_a , so $R(t, t')$ must belong to \mathcal{A} . Because \mathcal{T} is a *DL-Lite^{no-}* ontology, a role assertion can only be contradicted via its first argument. So we can use the fact that $\exists R \in S_t$ to infer that there is no assertion in \mathcal{B} which conflicts with $R(t, t')$, and hence $R(t, t')$ belongs to \mathcal{B} . We have thus shown that all atoms in q_a belong to \mathcal{B} . It follows that q_a , hence q , is consistently entailed from $(\mathcal{T}, \mathcal{A})$.

For the second direction, suppose that $\mathcal{I}_{\mathcal{A}} \not\models q'$. We first consider the case where there is some atom $R(c, d) \in q$ with (c, d) individuals such that $\mathcal{I}_{\mathcal{A}} \not\models R(c, d)$. Then $R(c, d) \notin \mathcal{A}$, and since there is no way to entail a such an assertion, we must have $\mathcal{T}, \mathcal{B} \not\models R(c, d)$ for every repair \mathcal{B} , hence $\mathcal{T}, \mathcal{A} \not\models_{\text{cons}} q$. Next we suppose that $\mathcal{I}_{\mathcal{A}} \not\models \psi_c$ for some individual c in q' . This means that we can find a maximal subset S_c of $\{B \mid B(c) \in \mathcal{A}\} \cup \{\exists R \mid R(c, d) \in \mathcal{A}\}$ consistent with \mathcal{T} such that $\mathcal{T} \not\models S_c \sqsubseteq C_c$ or $R_c \not\subseteq S_c$. We then let \mathcal{B} be any repair such that $\{B(c) \mid B \in S_c\} \cup \{R(c, d) \mid R(c, d) \in \mathcal{A}, \exists R \in S_c\} \subseteq \mathcal{B}$ (such a repair must exist since S_c is consistent with \mathcal{T} , and \mathcal{T} has no inverse roles). It is not hard to see that either $\mathcal{T}, \mathcal{B} \not\models B(c)$ for some $B \in C_c$, or $\mathcal{T}, \mathcal{B} \not\models \exists R(c)$ for some $\exists R \in R_c$ (and hence there are no role assertions of the form $R(c, _)$ in \mathcal{B}). In both cases, we obtain $\mathcal{T}, \mathcal{B} \not\models q$. Now consider the remaining case in which $\mathcal{I}_{\mathcal{A}} \models \psi_c$ for every individual c in q' and $\mathcal{I}_{\mathcal{A}} \models R(c, d)$ for every ground atom $R(c, d) \in q$. As $\mathcal{I}_{\mathcal{A}} \not\models q'$, for every individual a , we have $\mathcal{I}_{\mathcal{A}} \not\models q'_a$ (where as above, q'_a denotes q' with a substituted for x). That means that for each individual a , one of the following holds:

- (i) $\mathcal{I}_{\mathcal{A}}$ does not satisfy the formula $\psi_{x,a}$ obtained by replacing x with a in ψ_x

(ii) there is some conjunct $R(t, t')$ of q with $x \in \{t, t'\}$ which is not satisfied after x is replaced by a

If (i) holds, then we can find a maximal subset S_a of $\{B \mid B(a) \in \mathcal{A}\} \cup \{\exists R \mid R(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} such that either $\mathcal{T} \not\models S_a \sqsubseteq C_x$ or $R_x \not\subseteq S_a$. If (ii) holds, then let α_a be the role assertion $R(t, t')$ with x is replaced by a . We have that $\alpha_a \notin \mathcal{A}$. Using the above, we construct a set $\mathcal{B} \subseteq \mathcal{A}$ as follows:

- for each a such that $\mathcal{I}_{\mathcal{A}} \not\models \psi_{x,a}$: the assertions $\{B(a) \mid B \in S_a\} \cup \{R(a, b) \mid R(a, b) \in \mathcal{A}, \exists R \in S_a\}$
- for all other individuals a from \mathcal{A} , any maximal subset of $\{B(a) \mid B(a) \in \mathcal{A}\} \cup \{R(a, b) \mid R(a, b) \in \mathcal{A}\}$ consistent with \mathcal{T}

By construction, the set \mathcal{B} is a maximal subset of \mathcal{A} consistent with \mathcal{T} . We now show that $\mathcal{T}, \mathcal{B} \not\models q$. To do so, we consider the following interpretation \mathcal{J} :

- $\Delta^{\mathcal{J}} = \text{Ind}(\mathcal{B}) \cup \{w_R \mid R \in \mathbf{N}_R\}$
- $a^{\mathcal{J}} = a$, for each individual a from \mathcal{B}
- $A^{\mathcal{J}} = \{d \mid \mathcal{T}, \mathcal{B} \models A(d)\}$
- $R^{\mathcal{J}} = \{(d, e) \mid R(d, e) \in \mathcal{B}\} \cup \{(d, w_R) \mid d \in \text{Ind}(\mathcal{B}) \text{ and } \mathcal{T}, \mathcal{B} \models \exists R(d)\}$

It is easily seen that \mathcal{J} is a model of \mathcal{T}, \mathcal{B} . We intend to show that $\mathcal{J} \not\models q_a$ for every a . First consider the case (i) in which $\mathcal{I}_{\mathcal{A}} \not\models \psi_{x,a}$. Then from our construction of S_a , we know that either $\mathcal{T} \not\models S_a \sqsubseteq C_x$ or $R_x \not\subseteq S_a$. In the former case, there is an atom $B(a) \in q_a$ such that $\mathcal{T}, \mathcal{B} \not\models B(a)$. It follows that $a \notin B^{\mathcal{J}}$, and hence $\mathcal{J} \not\models q_a$. In the latter case, there is some $\exists R \in R_x$ such that there is no assertion of the form $R(a, d) \in \mathcal{B}$. It follows that there is some atom $R(a, e) \in q_a$ such that $(a, e) \notin R^{\mathcal{J}}$, which again yields $\mathcal{J} \not\models q_a$. Finally consider case (ii) in which there is a (ground) role atom α_a of q_a such $\alpha_a \notin \mathcal{A}$. Then we also have $\alpha_a \notin \mathcal{B}$, and so $\mathcal{J} \not\models q_a$. We have thus shown that there is no a for which $\mathcal{J} \models q_a$, which means that $\mathcal{J} \not\models q$. As a result, we have found a repair of \mathcal{A} which does not entail q given \mathcal{T} , which yields the desired $\mathcal{T}, \mathcal{A} \not\models_{\text{cons}} q$. \square

We now prove the remainder of Theorem 8.

Proposition 2. *Consider a DL-Lite^{no-} ontology \mathcal{T} , and a Boolean CQ q with at most two quantified variables. Then $\text{CONSENT}(q, \mathcal{T})$ is polynomial in data complexity, and first-order expressible if there is at most one quantified variable.*

Proof. The proof of Theorem 4, but includes a fair number of modifications in order to account for the existential concepts in the TBox. We will use the following algorithm which takes as input a DL-Lite^{no-} ontology \mathcal{T} , an ABox \mathcal{A} , and a conjunctive query q with two quantified variables x, y .

0. If there is $v \in \{x, y\}$ such that v occurs only once in q , as the second argument of some role assertion $R(t, v)$, then let $q' = q \setminus \{R(t, v)\} \cup \{F(t)\}$ and $\mathcal{T}' = \mathcal{T} \cup \{\exists R \sqsubseteq F\}$, for some fresh concept name F . As q' has at most one quantified variable, we can use the query rewriting approach from Proposition A to check whether $\mathcal{T}', \mathcal{A} \models_{\text{cons}} q'$. We output *yes* if so, and *no* if not.

1. For each ground atom $\alpha \in q$, output *no* if $\mathcal{T}, \mathcal{A} \not\models_{cons} \alpha$ (again we can use the procedure from Proposition A to perform this check). For each atom $R(c, z) \in q$ with $z \in \{x, y\}$ and $c \in \mathbb{N}_1$, output *no* if $\mathcal{T}, \mathcal{A} \not\models_{cons} \exists z R(c, z)$.
2. Set $C_x = \{A \mid A(x) \in q\}$ and $C_y = \{A \mid A(y) \in q\}$.
3. Initialize Σ to the tuples (a, b, ℓ) such that (a, b) satisfy:
 - if $R(x, y) \in q$, then $R(a, b) \in \mathcal{A}$
 - if $R(y, x) \in q$, then $R(b, a) \in \mathcal{A}$
 - if $R(x, x) \in q$, then $R(a, a) \in \mathcal{A}$
 - if $R(y, y) \in q$, then $R(b, b) \in \mathcal{A}$
 - if $R(x, c) \in q$, then $R(a, c) \in \mathcal{A}$
 - if $R(c, y) \in q$, then $R(c, a) \in \mathcal{A}$

and ℓ is defined as follows:

- $\ell = xy$ if there is a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying the condition (X) (see below), and there is a maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(b, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying condition (Y).
- $\ell = x$ if there is a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying the condition (X), and there is no maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(b, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying condition (Y).
- $\ell = y$ if there is a maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(b, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying condition (Y), but no maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying the condition (X).
- $\ell = 0$ if there is no maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying the condition (X), and no maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(b, d) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying condition (Y).

where the conditions (X) and (Y) are as follows:

- (X) $\mathcal{T} \not\models D_x \subseteq C_x$ or $\mathcal{T} \models D_x \subseteq \neg \exists R$ for some R such that there is an atom $R(x, t) \in q$.
- (Y) $\mathcal{T} \not\models D_y \subseteq C_y$ or $\mathcal{T} \models D_y \subseteq \neg \exists R$ for some R such that there is an atom $R(y, t) \in q$.

4. Apply the following rules until either there is some $(a, b, 0) \in \Sigma$ or there are no applicable rules:

R1 If $(a, b, x) \in \Sigma$, then:

- remove all tuples of the form (a, d, x) or (a, d, xy) from Σ

If it is the case that every maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} which satisfies condition (X) is also such that $\mathcal{T} \models D_x \subseteq C_y$ and $\mathcal{T} \not\models D_x \subseteq \neg \exists R$ for every R such that there is an atom $R(y, t) \in q$, then:

- replace each tuple in Σ of the form (d, a, y) by $(d, a, 0)$
- replace each tuple in Σ of the form (d, a, xy) by (d, a, x)

R2 If $(a, b, y) \in \Sigma$, then:

- remove all tuples of the form (d, b, y) or (d, b, xy) from Σ

If it is the case that every maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(b, d) \in \mathcal{A}\}$ consistent with \mathcal{T} which satisfies condition (Y) is also such that $\mathcal{T} \models D_y \subseteq C_x$ and $\mathcal{T} \not\models D_y \subseteq \neg \exists R$ for every R for which there is an atom $R(x, t) \in q$, then:

- replace each tuple in Σ of the form (b, d, x) by $(b, d, 0)$
- replace each tuple in Σ of the form (b, d, xy) by (b, d, y)

5. Return *yes* if there is some $(a, b, 0) \in \Sigma$, else return *no*.

Let us briefly highlight the differences between the above algorithm and the original algorithm from the proof of Theorem 4. We have an additional Step 0, whose purpose is to treat the limit case in which a role atom can be satisfied using an existential concept. We reduce this to the one-variable case, for which Proposition A provides a polynomial-time procedure via query rewriting. In Step 1, in addition for checking consistent entailment of ground atoms, we perform this check also for queries of the form $\exists z R(c, z)$ where c is an individual and z a variable. If the check fails, this implies q is not consistently entailed. If the check succeeds, then we know that there do not exist any assertions in \mathcal{A} which can prevent c from having an R -successor. In Step 3, the definition of the labels ℓ is modified because there are now two ways to spoil a query: either by ensuring that some concept atom does not hold, or by contradicting a role atom. Likewise, the rules R1 and R2 in Step 4 are slightly modified to account for these two ways of spoiling the query.

A simple examination of the definition of Σ and the rules $R1$ and $R2$ suffices to show that Σ always satisfies the following properties:

- P1 There is at most one tuple (a, b, ℓ) in Σ for each pair (a, b) of individuals from \mathcal{A} .
- P2 If $(a, b, x) \in \Sigma$, then there exists a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying (X).
- P3 If $(a, b, y) \in \Sigma$, then there exists a maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\}$ consistent with \mathcal{T} satisfying (Y).

We now proceed to the proof of correctness of the algorithm.

Soundness. Suppose that $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. We wish to show that the algorithm returns *no*, which can happen either in Step 0, Step 1, or Step 5. Consider first the case in which there is $v \in \{x, y\}$ such that v occurs only once in q as the second argument of some role assertion $R(t, v)$. Say that $v = y$ (the case where $v = x$ is similar). Then, we check in Step 0 whether $\mathcal{T}', \mathcal{A} \models_{cons} q'$, for $q' = q \setminus \{R(t, y)\} \cup \{F(t)\}$ and $\mathcal{T}' = \mathcal{T} \cup \{\exists R \sqsubseteq F\}$ (F a fresh concept name). We need to show that $\mathcal{T}', \mathcal{A} \not\models_{cons} q'$ (to get the desired output *no*). Suppose for a contradiction that $\mathcal{T}', \mathcal{A} \models_{cons} q'$. Since $\mathcal{T}, \mathcal{A} \not\models_{cons} q$, we have a repair \mathcal{B} of \mathcal{A} w.r.t. \mathcal{T} such that $\mathcal{T}, \mathcal{B} \not\models_{cons} q$. Note that \mathcal{B} must also be a repair of \mathcal{A} w.r.t. \mathcal{T}' , since \mathcal{T}' does not allow us to infer any new disjointness constraints. Let \mathcal{I} be any model of \mathcal{T}, \mathcal{B} such that $\mathcal{I} \not\models q$, and let \mathcal{I}' be obtained by taking \mathcal{I} and extending it to the augmented vocabulary of \mathcal{T}' by interpreting the new concept F as $F^{\mathcal{I}'} = (\exists R)^{\mathcal{I}'}$. It is easy to see that \mathcal{I}' is a model of $\mathcal{T}', \mathcal{B}$, and so

we must have $\mathcal{I}' \models q'$ since $\mathcal{T}', \mathcal{A} \models_{cons} q'$. Take any element u of $\Delta^{\mathcal{I}'}$ whose substitution for x makes q' hold in \mathcal{I}' . Now if $t = x$, then $F(x) \in q'$, so we also have $u \in (\exists R)^{\mathcal{I}'}$ because $F^{\mathcal{I}'} = (\exists R)^{\mathcal{I}'}$. Similarly if $t = c$ for some individual c , then $F(c) \in q'$, so $c^{\mathcal{I}'} \in F^{\mathcal{I}'} = (\exists R)^{\mathcal{I}'}$. In both case, let w be a witness for membership in $(\exists R)^{\mathcal{I}'}$. Then we can make q hold in \mathcal{I} by substituting u for x and w for y . which contradicts $\mathcal{I} \not\models q$. Thus, $\mathcal{T}', \mathcal{A} \not\models_{cons} q'$, and so the algorithm outputs *no*.

Now let us assume that the conditions in Step 0 do not apply, and so we proceed on to Step 1. In this step, we will output *no* if $\mathcal{T}, \mathcal{B} \not\models_{cons} \alpha$ for some ground atom $\alpha \in q$, or if $\mathcal{T}, \mathcal{B} \not\models_{cons} \exists z R(c, z)$ for some atom $R(c, z) \in q$ with $z \in \{x, y\}$. Thus, let us consider the case where all of these atomic queries are consistently entailed, and we proceed on to Step 3. Let us consider a repair \mathcal{B} of \mathcal{A} w.r.t. \mathcal{T} such that $\mathcal{T}, \mathcal{B} \not\models q$. Define an interpretation \mathcal{I} as follows:

- $\Delta^{\mathcal{I}}$ is the union of the set $\text{Ind}(\mathcal{B})$ of individuals from \mathcal{B} and the set $\{w_R \mid R \in \mathbf{N}_R\}$
- $a^{\mathcal{I}} = a$, for each individual a from \mathcal{B}
- $A^{\mathcal{I}} = \{d \mid \mathcal{T}, \mathcal{B} \models A(d)\}$
- $R^{\mathcal{I}} = \{(d, e) \mid R(d, e) \in \mathcal{B}\} \cup \{(d, w_R) \mid d \in \text{Ind}(\mathcal{B}) \text{ and } \mathcal{T}, \mathcal{B} \models \exists R(d)\}$

By construction, \mathcal{I} is a model of \mathcal{T}, \mathcal{B} (note that because of the lack of inverses, nothing can be deduced concerning the w_R). It is also not hard to see that \mathcal{I} is minimal among models of \mathcal{T}, \mathcal{B} , and so $\mathcal{T}, \mathcal{B} \not\models q$ implies $\mathcal{I} \not\models q$. Now let M be the set of pairs (a, b) such that there is some $(a, b, \ell) \in \Sigma$ at the end of Step 3. We will use q_{ab} to denote the ground CQ obtained by replacing x with a and y with b . Note that because $\mathcal{I} \not\models q$, we also have $\mathcal{I} \not\models q_{ab}$ for every $a, b \in \mathbf{N}_I$. Because of the successful consistent entailment checks in Step 2, and the conditions which must be satisfied to have $(a, b) \in M$, we know that for every $(a, b) \in M$, \mathcal{I} satisfies all atoms in q_{ab} except possibly atoms of the form $A(a)$, $R(a, d)$, $A(b)$, or $R(b, d)$. Since $\mathcal{I} \not\models q$, it follows that some atom in q_{ab} of one of the previous forms must not hold in \mathcal{I} . More precisely, at least one of the following conditions must hold:

- (i) there is some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$
- (ii) there is some $R(a, d) \in q_{ab}$ such that $(a, d) \notin R^{\mathcal{I}}$
- (iii) there is some $A \in C_y$ such that $b \notin A^{\mathcal{I}}$
- (iv) there is some $R(b, d) \in q_{ab}$ such that $(b, d) \notin R^{\mathcal{I}}$

We define a function $\text{val} : M \rightarrow \{x, y, xy\}$ as follows:

- $\text{val}(a, b) = xy$ if at least one of (i) and (ii) holds, and at least one of (iii) and (iv) holds
- $\text{val}(a, b) = x$ if at least one of (i) and (ii) holds, but neither (iii) nor (iv) holds
- $\text{val}(a, b) = y$ if at least one of (iii) and (iv) holds, but neither (i) nor (ii) holds

We define a partial order \succ of $\{0, x, y, xy\}$ as follows: $xy \succ x$, $xy \succ y$, $x \succ 0$, and $y \succ 0$. We want to show that the following property always holds for Σ :

$$(\star) \quad \text{if } (a, b, \ell) \in \Sigma, \text{ then } \ell \succeq \text{val}(a, b)$$

This implies that at the beginning of Step 5, there are no tuples $(a, b, 0) \in \Sigma$, and hence the algorithm returns the desired output *no*.

The proof is by induction on the number of rule applications that have been made. We first note that property (\star) hold for Σ when there have not been any rule applications. This is trivially the case when $\ell = xy$. For (a, b) such that $(a, b, x) \in \Sigma$ before any rule applications, we know from the way Σ was initialized in Step 3 that every maximal subset $D_b \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(b, d) \in \mathcal{A}\}$ consistent with \mathcal{T} is such that $\mathcal{T} \models D_b \sqsubseteq C_y$ and $\mathcal{T} \not\models D_b \sqsubseteq \neg \exists R$. It follows that there is no $A \in C_y$ such that $b \notin A^{\mathcal{I}}$ (so (iii) does not hold), nor any assertion in \mathcal{B} which could contradict a role assertion of the form $R(b, d)$. As \mathcal{T} does not contain any inverse roles, role assertions can only be contradicted via the first argument. Thus, it follows that every assertion in \mathcal{A} of the form $R(b, d)$ must also appear in \mathcal{B} . This means all role atoms in q_{ab} of the form $R(b, d)$ hold in \mathcal{I} (so (iv) does not hold). As $\mathcal{I} \not\models q$, there are only two remaining possibilities: either there is some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$ (condition (i)), or there is some $R(a, d) \in q_{ab}$ such that $(a, d) \notin R^{\mathcal{I}}$ (condition (ii)). Thus, we have shown that (i) or (ii) must hold, and (iii) and (iv) both do not hold, which mean we must have $\text{val}(a, b) = x$. The case of (a, b) such that (a, b, y) belongs to the original Σ proceeds analogously.

Now suppose that (\star) holds when there have been at most n rule applications, and consider the $n + 1$ th rule application. If it is R1 which is triggered by some tuple $(a, b, x) \in \Sigma$, and the extra condition in R1 is not met, then we only remove elements from Σ , and so the property (\star) continues to hold. Consider then the more interesting case in which the condition is satisfied, which may cause some tuple (d, a, xy) to be replaced by (d, a, x) , or (d, a, y) to be replaced by $(d, a, 0)$. Because $(a, b, x) \in \Sigma$ at the start of the rule application, we know that $x \succeq \text{val}(a, b)$, and hence $\text{val}(a, b) = x$. It follows that there must be either some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$, or some $R(a, d) \in q_{ab}$ such that $(a, d) \notin R^{\mathcal{I}}$. We then remark that the set $D_a = \{A \mid A(a) \in \mathcal{B}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{B}\}$ is a maximal subset of $\{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$, since otherwise we could add some additional assertion from \mathcal{A} to \mathcal{B} without causing inconsistency w.r.t. \mathcal{T} (recall again that by the lack of inverses, role assertions can only be contradicted via their first argument). Note that because there is some $A \in C_x$ such that $a \notin A^{\mathcal{I}}$, or some $R(a, d) \in q_{ab}$ such that $(a, d) \notin R^{\mathcal{I}}$, the set D_a must satisfy property (X). Then we can use the fact that the extra condition of R1 is satisfied to infer that $\mathcal{T} \models D_x \sqsubseteq C_y$ and $\mathcal{T} \not\models D_x \sqsubseteq \neg \exists R$ for every R such that there is an atom $R(y, t) \in q$. Then since $\mathcal{I} \not\models q$, it follows that for every $(d, a) \in M$, there must be some $A \in C_x$ such that $d \notin A^{\mathcal{I}}$, or some $R(d, e) \in q_{ab}$ such that $(d, e) \notin R^{\mathcal{I}}$. This means that for every $(d, a) \in M$, we have $\text{val}(d, a) = x$. Thus, from the induction hypothesis, we know that at the start of the rule application, every tuple $(d, a, \ell') \in M$ must be such that $\ell' = xy$ or $\ell' = x$. In the former case, we will replace (d, a, xy) by (d, a, x) , which

preserves property (\star) . The case where it is R2 which is applied in the $n + 1$ -th rule application proceeds analogously.

Completeness. To show completeness, suppose that the algorithm returns *no*. We need to show that $\mathcal{T}, \mathcal{A} \not\models_{cons} q$. The first possibility is that the algorithm returns *no* in Step 0. In this case, there is $v \in \{x, y\}$ such that v occurs only once in q as the second argument of some role assertion $R(t, v)$. We will assume $v = x$ (the case where $v = y$ proceeds analogously). The algorithm will then check whether $\mathcal{T}', \mathcal{A} \models_{cons} q'$, where $q' = q \setminus \{R(t, v)\} \cup \{F(t)\}$ and $\mathcal{T}' = \mathcal{T} \cup \{\exists R \sqsubseteq F\}$, for some fresh concept name F . As the output is *no*, we must have that $\mathcal{T}', \mathcal{A} \not\models_{cons} q'$. Let \mathcal{B} be some repair of \mathcal{A} w.r.t. \mathcal{T}' such that $\mathcal{T}, \mathcal{B} \not\models q'$, and let \mathcal{I}' be a model of $\mathcal{T}', \mathcal{B}$ such that $\mathcal{I}' \not\models q'$. Note that \mathcal{B} must also be a repair of \mathcal{A} w.r.t. \mathcal{T} , since \mathcal{T}' does not allow us to infer any new disjointness constraints. Let \mathcal{I} be the interpretation which is the same as \mathcal{I}' except that it does not interpret F . Suppose for a contradiction that $\mathcal{I} \models q$. Let (u, w) be a substitution which makes q hold in \mathcal{I} . We remark that this means that if $t = x$, then $(u, w) \in R^{\mathcal{I}}$, hence $u \in F^{\mathcal{I}'}$. Similarly, if $t = c$ for some individual c , then $(c^{\mathcal{I}}, w) \in R^{\mathcal{I}}$, hence $c^{\mathcal{I}} \in F^{\mathcal{I}'}$. It follows that we have a match for q' in \mathcal{I}' by setting x to u , contradicting the fact that $\mathcal{I}' \not\models q'$. Thus, it must be the case that $\mathcal{I} \not\models q$, hence $\mathcal{T}, \mathcal{B} \not\models q$, so $\mathcal{T}, \mathcal{A} \not\models_{cons} q$.

Consider next the case where *no* is returned at Step 1. In this case, there is some ground atom $\alpha \in q$ such that $\mathcal{T}, \mathcal{A} \not\models_{cons} \alpha$, or some atom $R(c, z) \in q$ with $z \in \{x, y\}$ and $c \in \mathbf{N}_I$ such that $\mathcal{T}, \mathcal{A} \not\models_{cons} \exists z R(c, z)$. In both cases, we immediately obtain $\mathcal{T}, \mathcal{A} \not\models_{cons} q$.

Let us now consider the last and most interesting case in which all of these atomic queries are consistently entailed, and *no* is returned only in Step 5. We will use the rule applications from Step 4 to guide the construction of a repair. As in the proof of Theorem 4, we will speak of an individual being the *target* of a rule application. Again, we have that each individual can be the target of at most one rule application, since applying the rule means removing all tuples containing the targeted individual from Σ . We also know that all tuples in Σ at the end of Step 4 must either have label 0 or xy , since any tuple with label x or y makes one of the rules applicable. Since the algorithm returns *no* in Step 5, there must only be tuples labelled xy . We now proceed to the definition of a repair \mathcal{B} . We consider each individual a separately and determine which assertions of the form $A(a)$ or $R(a, b)$ to include in \mathcal{B} .

Case 1: a is the target of an application of R1, and the extra condition was not satisfied.

Then property P2 and the non-satisfaction of the extra rule condition ensures that we can find a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} such that

- either $\mathcal{T} \not\models D_x \sqsubseteq C_x$ or $\mathcal{T} \models D_x \sqsubseteq \neg \exists R$ for some R such that there is an atom $R(x, t) \in q$

- either $\mathcal{T} \not\models D_x \sqsubseteq C_y$ or $\mathcal{T} \models D_y \sqsubseteq \neg\exists R$ for some R such that there is an atom $R(y, t) \in q$

We include in \mathcal{B} the concept assertions in $\{A(a) \mid A \in \mathbf{N}_C \cap D_x\}$, and the role assertions in $\{R(a, b) \mid R(a, b) \in \mathcal{A}, \exists R \in D_x\}$.

Case 2: a is the target of an application of R1, and the extra condition was satisfied.

By property P2, we know that we can find a maximal subset $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} such that either $\mathcal{T} \not\models D_x \sqsubseteq C_x$ or $\mathcal{T} \models D_x \sqsubseteq \neg\exists R$ for some R such that there is an atom $R(x, t) \in q$. We let D_x be any such set. We include in \mathcal{B} all concept assertions in $\{A(a) \mid A \in \mathbf{N}_C \cap D_x\}$, and role assertions in $\{R(a, b) \mid R(a, b) \in \mathcal{A}, \exists R \in D_x\}$.

Case 3: a is the target of an application of R2, and the extra condition was not satisfied.

Then property P3 and the non-satisfaction of the extra rule condition ensures that we can find a maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} such that:

- either $\mathcal{T} \not\models D_y \sqsubseteq C_y$ or $\mathcal{T} \models D_y \sqsubseteq \neg\exists R$ for some R such that there is an atom $R(y, t) \in q$
- either $\mathcal{T} \not\models D_y \sqsubseteq C_x$ or $\mathcal{T} \models D_y \sqsubseteq \neg\exists R$ for some R for which there is an atom $R(x, t) \in q$

We include in \mathcal{B} the concept assertions in $\{A(a) \mid A \in \mathbf{N}_C \cap D_y\}$, and the role assertions in $\{R(a, b) \mid R(a, b) \in \mathcal{A}, \exists R \in D_y\}$.

Case 4: a is the target of an application of R2, and the extra condition was satisfied.

By property P3, there is a maximal subset a maximal subset $D_y \subseteq \{A \mid A(b) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} such that either $\mathcal{T} \not\models D_y \sqsubseteq C_y$ or $\mathcal{T} \models D_y \sqsubseteq \neg\exists R$. We let D_y be any such set, and include in \mathcal{B} all concept assertions in $\{A(a) \mid A \in \mathbf{N}_C \cap D_y\}$, and role assertions in $\{R(a, b) \mid R(a, b) \in \mathcal{A}, \exists R \in D_y\}$.

Case 5: a appears in some tuple of Σ at the end of Step 4, and has not been the target of any rule application.

By above, we know the tuple in Σ containing a at the end of Step 4 must have label ' xy ', which means that this tuple was originally given the label ' xy ' in Step 3. Thus, we can find a $D_x \subseteq \{A \mid A(a) \in \mathcal{A}\} \cup \{\exists S \mid \text{some } S(a, d) \in \mathcal{A}\}$ consistent with \mathcal{T} such that either $\mathcal{T} \not\models D_x \sqsubseteq C_x$ or $\mathcal{T} \models D_x \sqsubseteq \neg\exists R$ for some R such that there is an atom $R(x, t) \in q$. We include in \mathcal{B} all concept assertions in $\{A(a) \mid A \in \mathbf{N}_C \cap D_x\}$, and role assertions in $\{R(a, b) \mid R(a, b) \in \mathcal{A}, \exists R \in D_x\}$.

Case 6: a is neither the target of a rule application, nor appears in any tuple at the end of Step 4.

We choose an arbitrary maximal subset of $\{A(a) \mid A(a) \in \mathcal{A}\} \cup \{R(a,b) \in R(a,b) \in \mathcal{A}\}$ consistent with \mathcal{T} .

By the above construction, we know that for each a , we have included in \mathcal{B} a maximal subset of $\{A(a) \mid A(a) \in \mathcal{A}\} \cup \{R(a,b) \in R(a,b) \in \mathcal{A}\}$ consistent with \mathcal{T} . Since there are no inverse roles in \mathcal{T} , it follows that \mathcal{B} is consistent with \mathcal{T} , and so it is a repair of \mathcal{A} w.r.t. \mathcal{T} . We let \mathcal{I} be the following interpretation:

- $\Delta^{\mathcal{I}} = \text{Ind}(\mathcal{B}) \cup \{w_R \mid R \in \mathbf{N}_R\}$
- $a^{\mathcal{I}} = a$, for each individual a from \mathcal{B}
- $A^{\mathcal{I}} = \{d \mid \mathcal{T}, \mathcal{B} \models A(d)\}$
- $R^{\mathcal{I}} = \{(d, e) \mid R(d, e) \in \mathcal{B}\} \cup \{(d, w_R) \mid d \in \text{Ind}(\mathcal{B}) \text{ and } \mathcal{T}, \mathcal{B} \models \exists R(d)\}$

It is easy to see that \mathcal{I} is a model of \mathcal{T}, \mathcal{B} . We wish to show that q does not hold in \mathcal{I} . Suppose for a contradiction that q does hold when x is replaced by d and y is replaced by e . It follows that the pair (d, e) satisfies the conditions in Step 3, so some tuple (d, e, ℓ) will be added to Σ . The first possibility is that there is a rule application during Step 4 which targets either d or e and causes the removal of a tuple (d, e, ℓ') from Σ (possibly with $\ell \neq \ell'$). First, suppose that (d, e, ℓ') is removed when applying R1 to d . Then d belongs to Case 1 or Case 2, and so $D = \{A \mid A(d) \in \mathcal{B}\} \cup \{\exists S \mid S(d, e) \in \mathcal{B}\}$ is such that either $\mathcal{T} \not\models D \sqsubseteq C_x$ or $\mathcal{T} \models D \sqsubseteq \neg \exists R$ for some R such that there is an atom $R(x, t) \in q$. But then q cannot be satisfied by substituting d for x . If instead the removal is due to applying R2 to e , then e belongs to either Case 3 or Case 4, and $D = \{A \mid A(e) \in \mathcal{B}\} \cup \{\exists S \mid S(e, d) \in \mathcal{B}\}$ satisfies $\mathcal{T} \not\models D \sqsubseteq C_y$ or $\mathcal{T} \models D \sqsubseteq \neg \exists R$ for some R such that there is an atom $R(y, t) \in q$. But this means that q cannot be satisfied by substituting e for y . Thus, in either case, we contradict our assumption that (d, e) defines a match of q in \mathcal{I} . Let us now consider the other possibility, which is that Σ contains a tuple (d, e, ℓ') after Step 4. We know from above that $\ell' = xy$. If d was not the target of any rule application, then we fall into Case 5, and we have that $\{A \mid A(d) \in \mathcal{B}\} \cup \{\exists S \mid S(d, f) \in \mathcal{B}\} = D_x$ for a set D_x such that either $\mathcal{T} \not\models D_x \sqsubseteq C_x$ or $\mathcal{T} \models D_x \sqsubseteq \neg \exists R$ for some R such that there is an atom $R(x, t) \in q$. It is also possible that d was the target of an application of R2, but the condition was not satisfied (Case 3), in which case $\{A \mid A(d) \in \mathcal{B}\} \cup \{\exists S \mid S(d, f) \in \mathcal{B}\} = D_y$ for some set D_y which satisfies either $\mathcal{T} \not\models D_y \sqsubseteq C_x$ or $\mathcal{T} \models D_y \sqsubseteq \neg \exists R$ for some R such that there is an atom $R(x, t) \in q$. Either way, we can infer that there is some $A \in C_x$ such that $d \notin A^{\mathcal{I}}$, or some atom $R(x, t) \in q$ such that $d \notin \exists R^{\mathcal{I}}$, contradicting our assumption that (d, e) defines a match of q . We have thus found a model \mathcal{I} of \mathcal{T}, \mathcal{B} such that $\mathcal{I} \not\models q$. It follows that $\mathcal{T}, \mathcal{B} \not\models q$, and hence $\mathcal{T}, \mathcal{A} \not\models_{\text{cons}} q$.

Complexity. Note that for Steps 0 and 1, we can use the query rewriting approach from Proposition A to decide consistent entailment for CQs with at most one quantified variable in polynomial time in $|\mathcal{A}|$. It is easy to see that the construction of Σ in Step 3 and the application of a rule during Step 4 can both take only polynomial time. Finally, we note that the total number of

rule applications in Step 4 cannot exceed $|\mathcal{A}| \times |\mathcal{A}|$ since each rule application decreases the cardinality of the set $\{(a, b) \mid (a, b, \ell) \in \Sigma \text{ for some } \ell\}$. \square

Theorem 9

Proof. The proof is very similar to the proof of Theorem 7, except that in addition to replacing each atom $A(t)$ in q' by its consistent rewriting $\varphi_{A(t)}$ of $A(t)$, we must also do the same for role atoms. Thus, the only interesting part of the proof is to show how to compute the consistent rewritings of atoms in q' . Let $A(t)$ be an atom in one of the CQs of q' . If t is an individual, then $A(t)$ is a Boolean CQ with no quantified variables, so using Proposition , we can find a consistent rewriting $\varphi_{A(t)}$ of $A(t)$. For atoms $A(t)$ such that t is a variable, we can use Proposition to find a consistent rewriting $\varphi_{A(c)}$ of $A(c)$ (where c is a fresh individual). Then the formula $\varphi_{A(t)}$ obtained by substituting t for c in $\varphi_{A(c)}$ is a consistent rewriting of the non-Boolean query $A(t)$, i.e. we have that for every individual a , $\mathcal{T}, \mathcal{A} \models_{cons} A(a)$ if and only if $\mathcal{I}_{\mathcal{A}} \models \varphi_{A(t)}^a$, where $\varphi_{A(t)}^a$ denotes the Boolean query obtained by replacing t with a in $\varphi_{A(t)}$. For a role atom $R(t, t')$, if both t, t' are individuals, then we directly obtain a consistent rewriting $\varphi_{R(t, t')}$ of $R(t, t')$ by applying Proposition . If one or both of t, t' is a variable, then we use the same trick as above. More precisely, we introduce two fresh individuals c, d , and we compute, using Proposition , a consistent rewriting $\varphi_{R(c, d)}$ of $R(c, d)$. Then by substituting t for c and t' for d , we obtain a formula $\varphi_{R(t, t')}$ which is a consistent rewriting of the non-Boolean query $R(t, t')$. \square