Answering Conjunctive Queries over $\mathcal{EL}$ Knowledge Bases with Transitive and Reflexive Roles

Giorgio Stefanoni and **Boris Motik**
Department of Computer Science, University of Oxford
Wolfson Building, Parks Road,
Oxford, OX1 3QD, UK

Abstract

Answering conjunctive queries (CQs) over $\mathcal{EL}$ knowledge bases (KBs) with complex role inclusions is PSPACE-hard and in PSPACE in certain cases; however, if complex role inclusions are restricted to role transitivity, a tight upper complexity bound has so far been unknown. Furthermore, the existing algorithms cannot handle reflexive roles, and they are not practicable. Finally, the problem is tractable for acyclic CQs and $\mathcal{ELH}$, and NP-complete for unrestricted CQs and $\mathcal{ELHO}$ KBs. In this paper we complete the complexity landscape of CQ answering for several important cases. In particular, we present a practicable NP algorithm for answering CQs over $\mathcal{ELHO}^s$ KBs—a logic containing all of OWL 2 EL, but with complex role inclusions restricted to role transitivity. Our preliminary evaluation suggests that the algorithm can be suitable for practical use. Moreover, we show that, even for a restricted class of so-called *arborescent* acyclic queries, CQ answering over $\mathcal{EL}$ KBs becomes NP-hard in the presence of either transitive or reflexive roles. Finally, we show that answering arborescent CQs over $\mathcal{ELHO}$ KBs is tractable, whereas answering acyclic CQs is NP-hard.

1 Introduction

Description logics (DLs) (Baader et al. 2007) are a family of knowledge representation languages that logically underpin the Web Ontology Language (OWL 2) (Cuenca Grau et al. 2008). DL knowledge bases (KBs) provide modern information systems with a flexible graph-like data model, and answering conjunctive queries (CQs) over such KBs is a core reasoning service in various applications (Calvanese et al. 2011). Thus, the investigation of the computational properties of CQ answering, as well as the development of practicable algorithms, have received a lot of attention lately.

For expressive DLs, CQ answering is at least exponential in *combined complexity* (Glimm et al. 2008; Ortiz, Rudolph, and Simkus 2011)—that is, measured in the combined size of the query and the KB. The problem is easier for the DL-Lite (Calvanese et al. 2007) and the $\mathcal{EL}$ (Baader, Brandt, and Lutz 2005) families of DLs, which logically underpin the QL and the EL profiles of OWL 2, respectively, and worst-case optimal, yet practicable algorithms are known (Kontchakov et al. 2011; Rodriguez-Muro, Kontchakov, and Zakharyaschev 2013; Eiter et al. 2012; Venetis, Stoilos, and Stamou 2014). One can reduce the complexity by restricting the query shape; for example, answering *acyclic* CQs (Yannakakis 1981) is tractable in relational databases. Bienvenu et al. (2013) have shown that answering acyclic CQs in DL-Lite$_{core}$ and $\mathcal{ELH}$ is tractable, whereas Gottlob et al. (2014) have shown it to be NP-hard in DL-Lite$_{\mathcal{R}}$.

In this paper, we consider answering CQs over KBs in the $\mathcal{EL}$ family of languages. No existing practical approach for $\mathcal{EL}$ supports complex role inclusions—a prominent feature of OWL 2 EL that can express complex properties of roles, including role transitivity. The known upper bound for answering CQs over $\mathcal{EL}$ KBs with complex role inclusions (Krötzsch, Rudolph, and Hitzler 2007) runs in PSPACE and uses automata techniques that are not practicable due to extensive don't-know nondeterminism. Moreover, this algorithm does not handle transitive roles specifically, but considers complex role inclusions. Hence, it is not clear whether the PSPACE upper bound is optimal in the presence of transitive roles only; this is interesting because role transitivity suffices to express simple graph properties such as reachability, and it is a known source of complexity of CQ answering (Eiter et al. 2009). Thus, to complete the landscape, we study the combined complexity of answering CQs over various extensions of $\mathcal{EL}$ and different classes of CQs. Our contributions can be summarised as follows.

In Section 3 we present a novel algorithm running in NP for answering CQs over $\mathcal{ELHO}^s$ KBs—a logic containing all of OWL 2 EL, but with complex role inclusions restricted to role transitivity—and thus settle the open question of the complexity for transitive and (locally) reflexive roles. Our procedure generalises the *combined approach with filtering* (Lutz et al. 2013) for $\mathcal{ELHO}$ by Stefanoni, Motik, and Horrocks (2013). We capture certain consequences of an $\mathcal{ELHO}^s$ KB by a datalog program; then, to answer a CQ, we evaluate the query over the datalog program to obtain *candidate answers*, and then we filter out unsound candidate answers. Transitive and reflexive roles, however, increase the complexity of the filtering step: unlike the filtering procedure for $\mathcal{ELHO}$, our filtering procedure runs in nondeterministic polynomial time, and we prove that this is worst-case optimal—that is, checking whether a candidate answer is sound is an NP-hard problem. To obtain a goal-directed filtering procedure, we developed optimisations that reduce the number of nondeterministic choices. Finally, our filtering procedure runs in NP only for candidate answers that depend on both the existential knowledge in the KB, and transitive or reflexive roles—that is, our algorithm exhibits pay-as-you-go behaviour. To evaluate the feasibility of our approach, we implemented a prototypical CQ answering system and we carried out a preliminary evaluation. Our results suggest that, although some queries may be challenging, our algorithm can be practicable in many cases.

In Section 4 we study the complexity of answering acyclic CQs over KBs expressed in various extensions of $\mathcal{EL}$. We introduce a new class of *arborescent* queries—tree-shaped acyclic CQs in which all roles point towards the parent. We prove that answering arborescent queries over $\mathcal{EL}$ KBs with either a single transitive role or a single reflexive role is NP-hard; this is interesting because Bienvenu et al. (2013) show that answering acyclic queries over $\mathcal{ELH}$ KBs is tractable, and it shows that our algorithm from Section 3 is optimal for arborescent (and thus also acyclic) queries. Moreover, we show that answering unrestricted acyclic CQs is NP-hard for $\mathcal{ELHO}$, but it becomes tractable for arborescent queries.

All proofs of our results are provided in the appendix.

2 Preliminaries

We use the standard notions of constants, (ground) terms, atoms, and formulas of first-order logic with the equality predicate $\approx$ (Fitting 1996); we assume that $\top$ and $\bot$ are unary predicates without any predefined meaning; and we often identify a conjunction with the set of its conjuncts. A substitution σ is a partial mapping of variables to terms; $\mathsf{dom}(\sigma)$ and $\mathsf{rng}(\sigma)$ are the domain and the range of σ, respectively; for convenience, we extend each σ to identity on ground terms; $\sigma|_S$ is the restriction of σ to a set of variables S; and, for α a term or a formula, $\sigma(\alpha)$ is the result of simultaneously replacing each free variable x occurring in α with $\sigma(x)$. Finally, $[i, j]$ is the set $\{i, i+1, \dots, j-1, j\}$.

Rules and Conjunctive Queries An *existential rule* is a formula $\forall\vec{x}\forall\vec{y}.\varphi(\vec{x},\vec{y}) \to \exists\vec{z}.\psi(\vec{x},\vec{z})$ where φ and ψ are conjunctions of function-free atoms over variables $\vec{x}\cup\vec{y}$ and $\vec{x}\cup\vec{z}$, respectively. An *equality rule* is a formula of the form $\forall\vec{x}.\varphi(\vec{x}) \to s \approx t$ where φ is a conjunction of function-free atoms over variables $\vec{x}$, and s and t are function-free terms with variables in $\vec{x}$. A *rule base* Σ is a finite set of rules and function-free ground atoms; Σ is a *datalog program* if $\vec{z} = \emptyset$ for each existential rule in Σ. Please note that Σ is always satisfiable, as $\top$ and $\bot$ are ordinary unary predicates. We typically omit universal quantifiers in rules.

A *conjunctive query* (CQ) is a formula $q = \exists\vec{y}.\psi(\vec{x},\vec{y})$ where ψ is a conjunction of function-free atoms over variables $\vec{x}\cup\vec{y}$. Variables $\vec{x}$ are the *answer variables* of q. Let $N_V(q) = \vec{x}\cup\vec{y}$ and let $N_T(q)$ be the set of terms occurring in q. When $\vec{x}$ is empty, we call q a *Boolean CQ*.

For τ a substitution, let $\tau(q) = \exists\vec{z}.\tau(\psi)$, where $\vec{z}$ is obtained from $\vec{y}$ by removing each variable $y \in \vec{y}$ such that $\tau(y)$ is a constant, and by replacing each variable $y \in \vec{y}$ such that $\tau(y)$ is a variable with $\sigma(y)$.

Let Σ be a rule base and let $q = \exists\vec{y}.\psi(\vec{x},\vec{y})$ be a CQ over the predicates in Σ. A substitution π is a *certain answer* to q over Σ, written $\Sigma \models \pi(q)$, if $\mathsf{dom}(\pi) = \vec{x}$, each element of $\mathsf{rng}(\pi)$ is a constant, and $\mathcal{I} \models \pi(q)$ for each model $\mathcal{I}$ of Σ.

Table 1: Translating $\mathcal{ELHO}^s$ Axioms into Rules

Type	Axiom		Rule
1	$A \sqsubseteq B$	$\rightsquigarrow$	$A(x) \to B(x)$
2	$A \sqsubseteq \{a\}$	$\rightsquigarrow$	$A(x) \to x \approx a$
3	$A_1 \sqcap A_2 \sqsubseteq A$	$\rightsquigarrow$	$A_1(x) \land A_2(x) \to A(x)$
4	$\exists R.A_1 \sqsubseteq A$	$\rightsquigarrow$	$R(x,y) \land A_1(y) \to A(x)$
5	$S \sqsubseteq R$	$\rightsquigarrow$	$S(x,y) \to R(x,y)$ $\mathsf{Self}_S(x) \to \mathsf{Self}_R(x)$
6	$\mathsf{range}(R, A)$	$\rightsquigarrow$	$R(x,y) \to A(y)$
7	$A_1 \sqsubseteq \exists R.A$	$\rightsquigarrow$	$A_1(x) \to \exists z.R(x,z) \land A(z)$
8	$\mathsf{trans}(R)$	$\rightsquigarrow$	$R(x,y) \land R(y,z) \to R(x,z)$
9	$\mathsf{refl}(R)$	$\rightsquigarrow$	$\top(x) \to R(x,x) \land \mathsf{Self}_R(x)$
10	$A \sqsubseteq \exists R.\mathsf{Self}$	$\rightsquigarrow$	$A(x) \to R(x,x) \land \mathsf{Self}_R(x)$
11	$\exists R.\mathsf{Self} \sqsubseteq A$	$\rightsquigarrow$	$\mathsf{Self}_R(x) \to A(x)$

The DL $\mathcal{ELHO}^s$ is defined w.r.t. a signature consisting of mutually disjoint and countably infinite sets N_C, N_R, and N_I of *atomic concepts* (i.e., unary predicates), *roles* (i.e., binary predicates), and *individuals* (i.e., constants), respectively. We assume that $\top$ and $\bot$ do not occur in N_C. Each $\mathcal{ELHO}^s$ knowledge base can be *normalised* in polynomial time without affecting CQ answers (Krötzsch 2010), so we consider only normalised KBs. An $\mathcal{ELHO}^s$ *TBox* $\mathcal{T}$ is a finite set of axioms of the form shown in the left-hand side of Table 1, where $A_{(i)} \in N_C \cup \{\top\}$, $B \in N_C \cup \{\top, \bot\}$, $S, R \in N_R$, and $a \in N_I$; furthermore, TBox $\mathcal{T}$ is in $\mathcal{ELHO}$ if it contains only axioms of types 1–7. Relation $\sqsubseteq^*_\mathcal{T}$ is the smallest reflexive and transitive relation on the set of roles occurring in $\mathcal{T}$ such that $S \sqsubseteq^*_\mathcal{T} R$ for each $S \sqsubseteq R \in \mathcal{T}$. A role R is *simple* in $\mathcal{T}$ if $\mathsf{trans}(S) \notin \mathcal{T}$ for each $S \in N_R$ with $S \sqsubseteq^*_\mathcal{T} R$. An *ABox* $\mathcal{A}$ is a finite set of ground atoms constructed using the symbols from the signature. An $\mathcal{ELHO}^s$ *knowledge base* (KB) is a tuple $\mathcal{K} = \langle\mathcal{T}, \mathcal{A}\rangle$, where $\mathcal{T}$ is an $\mathcal{ELHO}^s$ TBox and an $\mathcal{A}$ is an ABox such that each role R occurring in axioms of types 10 or 11 in $\mathcal{T}$ is simple.

Let ind be a fresh atomic concept and, for each role R, let Self_R be a fresh atomic concept uniquely associated with R. Table 1 shows how to translate an $\mathcal{ELHO}^s$ TBox $\mathcal{T}$ into a rule base $\Xi_\mathcal{T}$. Furthermore, rule base $\mathsf{cls}_\mathcal{K}$ contains an atom $\mathsf{ind}(a)$ for each individual a occurring in $\mathcal{K}$, a rule $A(x) \to \top(x)$ for each atomic concept occurring in $\mathcal{K}$, and the following two rules for each role R occurring in $\mathcal{K}$.

$$\mathsf{ind}(x) \land R(x,x) \to \mathsf{Self}_R(x) \tag{1}$$

$$R(x,y) \to \top(x) \land \top(y) \tag{2}$$

For $\mathcal{K}$ a KB, let $\Xi_\mathcal{K} = \Xi_\mathcal{T} \cup \mathsf{cls}_\mathcal{K} \cup \mathcal{A}$; then, $\mathcal{K}$ is *unsatisfiable* iff $\Xi_\mathcal{K} \models \exists y.\bot(y)$. For q a CQ and π a substitution, we write $\mathcal{K} \models \pi(q)$ iff $\mathcal{K}$ is unsatisfiable or $\Xi_\mathcal{K} \models \pi(q)$. Our definition of the semantics of $\mathcal{ELHO}^s$ is unconventional, but equivalent to the usual one (Krötzsch 2010).

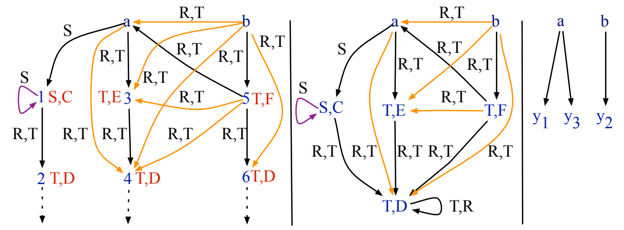

Figure 1: The models of $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$, and the skeleton for q

3 Answering CQs over $\mathcal{ELHO}^s$ KBs

In this section, we present an algorithm for answering CQs over $\mathcal{ELHO}^s$ KBs running in NP. In the rest of this section, we fix $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ to be an arbitrary $\mathcal{ELHO}^s$ KB.

Certain answers to a CQ over $\Xi_\mathcal{K}$ can be computed by evaluating the CQ over a so-called *canonical model* that can be homomorphically embedded into all other models of $\Xi_\mathcal{K}$. It is well known (Krötzsch, Rudolph, and Hitzler 2007) that such models can be seen as a family of directed trees whose roots are the individuals occurring in $\Xi_\mathcal{K}$, and that contain three kinds edges: *direct edges* point from parents to children or to the individuals in $\Xi_\mathcal{K}$; *transitive edges* introduce shortcuts between these trees; and *self-edges* introduce loops on tree nodes. We call non-root elements *auxiliary*. Moreover, each auxiliary element can be uniquely associated with a rule of the form $A_1(x) \rightarrow \exists z.R(x, z) \wedge A(z)$ in $\Xi_\mathcal{K}$ that was used to generate it, and we cal R, A the element's *type*. Example 1 illustrates these observations.

Example 1. *Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be an $\mathcal{ELHO}^s$ KB whose $\mathcal{T}$ contains the following axioms and $\mathcal{A} = \{A(a), B(b)\}$.*

$A \sqsubseteq \exists S.C$	$E \sqsubseteq \exists T.D$	$G \sqsubseteq \{a\}$
$C \sqsubseteq \exists S.\mathsf{Self}$	$B \sqsubseteq \exists T.F$	$D \sqsubseteq \exists T.D$
$C \sqsubseteq \exists T.D$	$F \sqsubseteq \exists T.D$	$T \sqsubseteq R$
$A \sqsubseteq \exists T.E$	$F \sqsubseteq \exists T.G$	$\mathsf{trans}(T)$

The left part of Figure 1 shows a canonical model I of $\Xi_\mathcal{K}$. Each auxiliary element is represented as a number showing the element's type. Direct edges are black, transitive edges are orange, and self-edges are purple. Axiom $D \sqsubseteq \exists T.D$ makes I is infinite; in the figure, we show the infinitely many successors of 2*,* 4*, and* 6 *using a black dotted edge.*

A canonical model I of $\Xi_\mathcal{K}$ can be infinite, so a terminating CQ answering algorithm for $\mathcal{ELHO}^s$ cannot materialise I and evaluate CQs in it. Instead, we first show how to translate $\mathcal{K}$ into a datalog program $\mathsf{D}_\mathcal{K}$ that finitely captures the canonical model I of $\Xi_\mathcal{K}$; next, we present a CQ answering procedure that uses $\mathsf{D}_\mathcal{K}$ to answer CQs over $\Xi_\mathcal{K}$.

Datalog Translation

Krötzsch, Rudolph, and Hitzler (2008) translate $\mathcal{K}$ into datalog for the purposes of ontology classification, and Stefanoni, Motik, and Horrocks (2013) use this translation to answer CQs over $\mathcal{ELHO}$ KBs. Let $o_{R,A}$ be an *auxiliary individual* not occurring in N_I and uniquely associated with each role $R \in N_R$ and each atomic concept $A \in N_C \cup \{\top\}$ occurring in $\mathcal{K}$; intuitively, $o_{R,A}$ represent all auxiliary terms of type R, A in a canonical model I of $\Xi_\mathcal{K}$. We extend this translation by uniquely associating with each role R a *direct predicate* d_R to represent the direct edges in I.

Definition 1. *For each axiom $\alpha \in \mathcal{T}$ not of type* 7*, set $\mathsf{D}_\mathcal{T}$ contains the translation of α into a rule as shown in Table 1; moreover, for each axiom $A_1 \sqsubseteq \exists R.A \in \mathcal{T}$, set $\mathsf{D}_\mathcal{T}$ contains rule $A_1(x) \rightarrow R(x, o_{R,A}) \wedge \mathsf{d}_R(x, o_{R,A}) \wedge A(o_{R,A})$; finally, for each axiom $S \sqsubseteq R \in \mathcal{T}$, set $\mathsf{D}_\mathcal{T}$ contains rule $\mathsf{d}_S(x, y) \rightarrow \mathsf{d}_R(x, y)$. Then, $\mathsf{D}_\mathcal{K} = \mathsf{D}_\mathcal{T} \cup \mathsf{cls}_\mathcal{K} \cup \mathcal{A}$ is the* datalog program for $\mathcal{K}$.

Example 2. *The middle part of Figure 1 shows model J of the datalog program $\mathsf{D}_\mathcal{K}$ for the KB from Example 1. For clarity, auxiliary individuals $o_{R,A}$ are shown as R, A. Note that auxiliary individual $o_{T,G}$ is 'merged' in model J with individual a since $\mathsf{D}_\mathcal{K} \models o_{T,G} \approx a$. We use the notation from Example 1 to distinguish various kinds of edges.*

The following proposition shows how to use $\mathsf{D}_\mathcal{K}$ to test whether $\mathcal{K}$ is unsatisfiable.

Proposition 2. *$\mathcal{K}$ is unsatisfiable iff $\mathsf{D}_\mathcal{K} \models \exists y.\bot(y)$.*

The CQ Answering Algorithm

Program $\mathsf{D}_\mathcal{K}$ can be seen as a strengthening of $\Xi_\mathcal{K}$: all existential rules $A_1(x) \rightarrow \exists z.R(x, z) \wedge A(z)$ in $\Xi_\mathcal{K}$ are satisfied in a model J of $\mathsf{D}_\mathcal{K}$ using a single auxiliary individual $o_{R,A}$. Therefore, evaluating a CQ q in J produces a set of *candidate answers*, which provides us with an upper bound on the set of certain answers to q over $\Xi_\mathcal{K}$.

Definition 3. *A substitution τ is a* candidate answer *to a CQ $q = \exists \vec{y}.\psi(\vec{x}, \vec{y})$ over $\mathsf{D}_\mathcal{K}$ if $\mathsf{dom}(\tau) = N_V(q)$, each element of $\mathsf{rng}(\tau)$ is an individual occurring in $\mathsf{D}_\mathcal{K}$, and $\mathsf{D}_\mathcal{K} \models \tau(q)$. Such a candidate answer τ is* sound *if $\Xi_\mathcal{K} \models \tau|_{\vec{x}}(q)$.*

Stefanoni, Motik, and Horrocks (2013) presented a filtering step that removes unsound candidate answers; however, Example 3 shows that this step can be incomplete when the query contains roles that are not simple.

Example 3. *Let $\mathcal{K}$ be as in Example 1 and let*

$$q = \exists y.A(x_1) \wedge R(x_1, y) \wedge B(x_2) \wedge R(x_2, y) \wedge D(y).$$

Moreover, let π be the substitution such that $\pi(x_1) = a$ and $\pi(x_2) = b$, and let τ be such that $\pi \subseteq \tau$ and $\tau(y) = o_{T,D}$. Using models I and J from Figure 1, one can easily see that $\Xi_\mathcal{K} \models \pi(q)$ and $\mathsf{D}_\mathcal{K} \models \tau(q)$. However, q contains a 'fork' $R(x_1, y) \wedge R(x_2, y)$, and τ maps y to an auxiliary individual, so this answer is wrongly filtered as unsound.

Algorithm 1 specifies a procedure $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ that checks whether a candidate answer is sound. We discuss the intuitions using the KB from Example 1, and the query q and the candidate answer τ from Example 4.

Example 4. *Let q and τ be as follows. Using Figure 1, one can easily see that $\mathsf{D}_\mathcal{K} \models \tau(q)$.*

$$q = \exists \vec{y}.\ S(x, y_1) \wedge S(y_1, y_1) \wedge R(x, y_3) \wedge D(y_3) \wedge R(y_2, y_3) \wedge F(y_2) \wedge T(y_2, x)$$

$$\tau = \{x \mapsto a,\ y_1 \mapsto o_{S,C},\ y_2 \mapsto o_{T,F},\ y_3 \mapsto o_{T,D}\}$$

We next show how $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ decides that τ is sound—that is, that a substitution π mapping the variables in q to terms in I exists such that $\pi(x) = a$ and $\pi(q) \subseteq I$. Substitution τ already provides us with some constraints on π: it must map variable y_1 to 1 and variable y_2 to 5, since these are the only elements of I of types S, C and T, F, respectively. In contrast, substitution π can map variable y_3 to either one of 2, 4, and 6. Each such substitution π is guaranteed to satisfy all unary atoms of q, all binary atoms of q that τ maps to direct edges pointing towards (non-auxiliary) individuals from N_I, and all binary atoms of q that contain a single variable and that τ maps to the self-edge in J. Atoms $T(y_2, x)$ and $S(y_1, y_1)$ in q satisfy these conditions, and so we call them *good* w.r.t. τ. To show that τ is sound, we must demonstrate that all other atoms of q are satisfied.

Step 1 of Algorithm 1 implements the 'fork' and 'aux-acyclicity' checks (Stefanoni, Motik, and Horrocks 2013). To guarantee completeness, we consider only those binary atoms in q that contain simple roles, and that τ maps onto direct edges in J pointing towards auxiliary elements. We call these atoms *aux-simple* as they can be mapped onto the direct edges in I pointing towards auxiliary elements. In step 1 we compute a new query $q_\sim$ by applying all constraints derived by the fork rule, and in the rest of the algorithm we consider $q_\sim$ instead of q. In our example, atom $S(x, y_1)$ is the only aux-simple atom, so q does not contain forks and $q_\sim = q$. When all binary atoms occurring in q are good or aux-simple, step 1 guarantees that τ is sound. Query q from Example 4, however, contains binary atoms that are neither good nor aux-simple, so we proceed to step 3.

Next, in step 3 we guess a renaming σ for the variables in $q_\sim$ to take into account that distinct variables in $q_\sim$ that τ maps to the same auxiliary individual can be mapped to the same auxiliary element of I, and so in the rest of Algorithm 1 we consider $\sigma(q_\sim)$ instead of $q_\sim$. In our example, we guess σ to be identity, so $\sigma(q_\sim) = q_\sim = q$.

In step 4, we guess a *skeleton* for $\sigma(q_\sim)$, which is a finite structure that finitely describes the (possibly infinite) set of all substitutions π mapping the variables in $\sigma(q_\sim)$ to distinct auxiliary elements of I. The right part of Figure 1 shows the skeleton $\mathcal{S}$ for our example query. The vertices of $\mathcal{S}$ are the (non-auxiliary) individuals from $\mathsf{D}_\mathcal{K}$ and the variables from $\sigma(q_\sim)$ that τ maps to auxiliary individuals, and they are arranged into a forest rooted in N_I. Such $\mathcal{S}$ represents those substitutions π that map variables y_1 and y_3 to auxiliary elements of I under individual a, and that map variable y_2 to an auxiliary element of I under individual b.

In steps 5–15, our algorithm labels each edge $\langle v', v\rangle \in \mathcal{S}$ with a set of roles $L(v', v)$; after these steps, $\mathcal{S}$ represents those substitutions π that satisfy the following property (E):

> for each role $P \in L(v', v)$, a path from $\tau(v')$ to $\tau(v)$ in J exists that consists only of direct edges labelled by role P pointing to auxiliary individuals.

We next show how atoms of $\sigma(q_\sim)$ that are not good contribute to the labelling of $\mathcal{S}$. Atom $S(x, y_1)$ is used in step 6 to label edge $\langle a, y_1\rangle$. For atom $R(x, y_3)$, in step 8 we let $P = T$ and we label edge $\langle a, y_3\rangle$ with P. Using Figure 1 and the axioms in Example 1, one can easily check that the conditions in steps 8 and 9 are satisfied. For atom $R(y_2, y_3)$, variables y_2 and y_3 are not reachable in $\mathcal{S}$, so we must split the path from y_2 to y_3. Thus, in step 8 we let $P = T$, and in step 13 we let $a_t = a$; hence, atom $R(y_2, y_3)$ is split into atoms $T(y_2, a)$ and $T(a, y_3)$. The former is used in step 14 to check that a direct path exists in J connecting $o_{T,F}$ with a, and the latter is used to label edge $\langle a, y_2\rangle$.

After the for-loop in steps 7–15, skeleton $\mathcal{S}$ represents all substitutions satisfying (E). In step 17, function exist exploits the direct predicates from $\mathsf{D}_\mathcal{K}$ to find the required direct paths in J, thus checking whether at least one such substitution exists (see Definition 11). Using Figure 1, one can check that substitution π where $\pi(y_3) = 4$ and that maps all other variables as stated above is the only substitution satisfying the constraints imposed by $\mathcal{S}$; hence, $\mathsf{isSound}$ returns t, indicating that candidate answer τ is sound.

We now formalise the intuitions that we have just presented. Towards this goal, in the rest of this section we fix a CQ q' and a candidate answer τ' to q' over $\mathsf{D}_\mathcal{K}$.

Due to equality rules, auxiliary individuals in $\mathsf{D}_\mathcal{K}$ may be equal to individuals from N_I, thus not representing auxiliary elements of I. Hence, set $\mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ in Definition 4 provides us with all auxiliary individuals that are not equal to an individual from N_I. Moreover, to avoid dealing with equal individuals, we replace in query q' all terms that τ' does not map to individuals in $\mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ with a single canonical representative, and we do analogously for τ'; this replacement produces CQ q and substitution τ. Since q and τ are obtained by replacing equals by equals, we have $\mathsf{D}_\mathcal{K} \models \tau(q)$. Our filtering procedure uses q and τ to check whether τ' is sound.

Definition 4. *Let $>$ be a total order on ground terms such that $o_{R,A} > a$ for all individuals $o_{R,A}$ and $a \in N_I$ from $\mathsf{D}_\mathcal{K}$. Set $\mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ contains each individual u from $\mathsf{D}_\mathcal{K}$ for which no individual $a \in N_I$ exists such that $\mathsf{D}_\mathcal{K} \models u \approx a$. For each individual u from $\mathsf{D}_\mathcal{K}$, let $u_\approx = u$ if $u \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$; otherwise, let $u_\approx$ be the smallest individual $a \in N_I$ in the ordering $>$ such that $\mathsf{D}_\mathcal{K} \models u \approx a$. Set $\mathsf{ind}_{\mathsf{D}_\mathcal{K}}$ contains $a_\approx$ for each individual $a \in N_I$ occurring in $\mathsf{D}_\mathcal{K}$. Then query q is obtained from q' by replacing each term $t \in N_T(q')$ such that $\tau'(t) \notin \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ with $\tau'(t)_\approx$; substitution τ is obtained by restricting τ' to only those variables occurring in q.*

Next, we define good and aux-simple atoms w.r.t. τ.

Definition 5. *Let $R(s, t)$ be an atom where $\tau(s)$ and $\tau(t)$ are defined. Then, $R(s, t)$ is* good *if $\tau(t) \in N_I$, or $s = t$ and $\mathsf{D}_\mathcal{K} \models \mathsf{Self}_R(\tau(s))$. Furthermore, $R(s, t)$ is* aux-simple *if $s \neq t$, R is a simple role, $\tau(t) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, and $\tau(s) = \tau(t)$ implies $\mathsf{D}_\mathcal{K} \not\models \mathsf{Self}_R(\tau(s))$.*

Note that, if $R(s, t)$ is not good, then t is a variable and $\tau(t) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. Moreover, by the definition of $\mathsf{D}_\mathcal{K}$, if atom $R(s, t)$ is aux-simple, then $\mathsf{D}_\mathcal{K} \models \mathsf{d}_R(\tau(s), \tau(t))$. The following definition introduces the query $q_\sim$ obtained by applying the fork rule by Stefanoni, Motik, and Horrocks (2013) to only those atoms that are aux-simple.

Definition 6. *Relation $\sim \subseteq N_T(q) \times N_T(q)$ for q and τ is the smallest reflexive, symmetric, and transitive relation closed under the* fork *rule.*

$$(\mathsf{fork})\ \frac{s' \sim t'}{s \sim t} \quad \textit{R(s, s') and P(t, t') are aux-simple atoms in q w.r.t. } \tau$$

Query $q_\sim$ is obtained from query q by replacing each term $t \in N_T(q)$ with an arbitrary, but fixed representative of the equivalence class of $\sim$ that contains t.

To check whether $q_\sim$ is aux-acyclic, we next introduce the *connection graph* cg for q and τ that contains a set E_s of edges $\langle v', v\rangle$ for each aux-simple atom $R(v', v) \in q_\sim$. In addition, cg also contains a set E_t of edges $\langle v', v\rangle$ that we later use to guess a skeleton for $\sigma(q_\sim)$ more efficiently. By the definition of aux-simple atoms, we have $E_s \subseteq E_t$.

Definition 7. *The* connection graph for q and τ *is a triple* $\mathsf{cg} = \langle V, E_s, E_t\rangle$ *where* $E_s, E_t \subseteq V \times V$ *are smallest sets satisfying the following conditions.*

- $V = \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \{z \in N_V(q_\sim) \mid \tau(z) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}\}$.
- *Set E_s contains $\langle v', v\rangle$ for all $v', v \in V$ for which a role R exist such that $R(v', v)$ is an aux-simple atom in $q_\sim$.*
- *Set E_t contains $\langle v', v\rangle$ for all $v', v \in V$ such that individuals $\{u_1, \dots, u_n\} \subseteq \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ and roles $R_1, \dots, R_n$ exist with $n > 0$, $u_n = \tau(v)$, and $\mathsf{D}_\mathcal{K} \models \mathsf{d}_{R_i}(u_{i-1}, u_i)$ for each $i \in [1, n]$ and $u_0 = \tau(v')$.*

Function $\mathsf{isDSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ from Definition 8 ensures that τ satisfies the constraints in $\sim$, and that $q_\sim$ does not contain cycles consisting only of aux-simple atoms.

Definition 8. *Function* $\mathsf{isDSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ *returns* t *if and only if the two following conditions hold.*

1. *For all $s, t \in N_T(q)$, if $s \sim t$, then $\tau(s) = \tau(t)$.*
2. *$\langle V, E_s\rangle$ is a directed acyclic graph.*

We next define the notions of a variable renaming for q and τ, and of a skeleton for q and σ.

Definition 9. *A substitution σ with* $\mathsf{dom}(\sigma) = V \cap N_V(q)$ *and* $\mathsf{rng}(\sigma) \subseteq \mathsf{dom}(\sigma)$ *is a* variable renaming for q and τ *if*

1. *for each $v \in \mathsf{dom}(\sigma)$, we have $\tau(v) = \tau(\sigma(v))$,*
2. *for each $v \in \mathsf{rng}(\sigma)$, we have $\sigma(v) = v$, and*
3. *directed graph $\langle \sigma(V), \sigma(E_s)\rangle$ is a forest.*

Definition 10. *A* skeleton *for q and a variable renaming σ is a directed graph $\mathcal{S} = \langle \mathcal{V}, \mathcal{E}\rangle$ where $\mathcal{V} = \sigma(V)$, and $\mathcal{E}$ satisfies $\sigma(E_s) \subseteq \mathcal{E} \subseteq \sigma(E_t)$ and it is a forest whose roots are the individuals occurring in $\mathcal{V}$.*

Finally, we present function exist that checks whether one can satisfy the constraints imposed by the roles $L(v', v)$ labelling a skeleton edge $\langle v', v\rangle \in \mathcal{E}$.

Definition 11. *Given individuals u' and u, and a set of roles L, function* $\mathsf{exist}(u', u, L)$ *returns* t *if and only if individuals $\{u_1, \dots, u_n\} \subseteq \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ with $n > 0$ and $u_n = u$ exist where*

- *if $S \in L$ exists such that $\mathsf{trans}(S) \notin \mathcal{T}$, then $n = 1$; and*
- *$u_0 = u'$, and $\mathsf{D}_\mathcal{K} \models \mathsf{d}_R(u_{i-1}, u_i)$ for each $R \in L$ and each $i \in [1, n]$.*

Candidate answer τ' for q' over $\mathsf{D}_\mathcal{K}$ is sound, if the nondeterministic procedure $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ from Algorithm 1 returns t, as shown by Theorem 12.

Theorem 12. *Let π' be a substitution. Then $\Xi_\mathcal{K} \models \pi'(q')$ iff $\mathcal{K}$ is unsatisfiable, or a candidate answer τ' to q' over $\mathsf{D}_\mathcal{K}$ exists such that $\tau'|_{\vec{x}} = \pi'$ and the following conditions hold:*

1. *for each $x \in \vec{x}$, we have $\tau'(x) \in N_I$, and*
2. *a nondeterministic computation exists such that function* $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ *returns* t.

Algorithm 1: $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$

```
if isDSound(q, D_K, τ) = f then return f
return t if each R(s,t) ∈ q_~ is good or aux-simple
guess a variable renaming σ for q and τ
guess a skeleton S = ⟨V, E⟩ for q, σ, and τ
for ⟨v', v⟩ ∈ E, let L(v', v) = ∅
for aux-simple atom R(s,t) ∈ σ(q_~), add R to L(s,t)
for neither good nor aux-simple R(s,t) ∈ σ(q_~) do
    guess role P s.t. D_K ⊨ P(τ(s), τ(t)) and P ⊑*_T R
    if ⟨s,t⟩ ∉ E and trans(P) ∉ T then return f
    if s reaches t in E then
        let v_0, ..., v_n be the path from s to t in E
    else
        let a_t be the root reaching t in E via v_0, ..., v_n
        if D_K ⊭ P(τ(s), a_t) then return f
    for i ∈ [1, n], add P to L(v_{i-1}, v_i)
for ⟨v', v⟩ ∈ E do
    if exist(τ(v'), τ(v), L(v', v)) = f then return f
return t
```

The following results show that our function $\mathsf{isSound}$ runs in nondeterministic polynomial time.

Theorem 13. *Function* $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ *can be implemented so that*

1. *it runs in nondeterministic polynomial time,*
2. *if each binary atom in q is either good or aux-simple w.r.t. τ, it runs in polynomial time, and*
3. *if the TBox $\mathcal{T}$ and the query q are fixed, it runs in polynomial time in the size of the ABox $\mathcal{A}$.*

Each rule in $\mathsf{D}_\mathcal{K}$ contains a fixed number of variables, so we can compute all consequences of $\mathsf{D}_\mathcal{K}$ using polynomial time. Thus, we can compute CQ q and substitution τ in polynomial time, and by Proposition 2, we can also check whether $\mathcal{K}$ is unsatisfiable using polynomial time; hence, by Theorem 13, we can check whether a certain answer to q' over $\Xi_\mathcal{K}$ exists using nondeterministic polynomial time in combined complexity (i.e., when the ABox, the TBox, and the query are all part of the input), and in polynomial time in data complexity (i.e., when the TBox and the query are fixed, and only the ABox is part of the input).

The filtering procedure by Stefanoni, Motik, and Horrocks (2013) is polynomial, whereas the one presented in this paper introduces a source of intractability. In Theorem 14 we show that checking whether a candidate answer is sound is an NP-hard problem; hence, this complexity increase is unavoidable. We prove our claim by reducing the NP-hard problem of checking satisfiability of a 3CNF formula φ (Garey and Johnson 1979). Towards this goal, we define an $\mathcal{ELHO}^s$ KB $\mathcal{K}_\varphi$ and a Boolean CQ q_φ such that φ is satisfiable if and only if $\Xi_{\mathcal{K}_\varphi} \models q_\varphi$. Furthermore, we define a substitution τ_φ, and we finally show that τ_φ is a unique candidate answer to q_φ over $\mathsf{D}_{\mathcal{K}_\varphi}$.

Theorem 14. *Checking whether a candidate answer is sound is* NP*-hard.*

Table 2: Evaluation results

		Inds.	Unary atoms	Binary atoms	Total atoms	Ratio
U5	before	100,848	169,079	296,941	466,020	
	after	100,873	511,115	1,343,848	1,854,963	3.98
U10	before	202,387	339,746	598,695	938,441	
	after	202,412	1,026,001	2,714,214	3,740,215	3.98
U20	before	426,144	714,692	1,259,936	1,974,628	
	after	426,169	2,157,172	5,720,670	7,877,842	3.99

	q_3^l				q_1^t				q_2^t				q_3^t				q_4^t			
	C	U	F	N	C	U	F	N	C	U	F	N	C	U	F	N	C	U	F	N
U5	10	0	0.06	0	73K	12	1.71	7.55	3K	0	0.01	0	157K	66	1.07	8.6	30K	63	2.44	10.9
U10	22	0	0.06	0	149K	12	1.68	7.54	6K	0	0.01	0	603K	81	1.20	9.6	61K	63	2.44	10.9
U20	43	0	0.07	0	313K	12	1.66	7.55	12K	0	0.01	0	2.6M	90	1.28	10.3	129K	63	2.44	10.9

(C) total number of candidate answers (U) percentage of unsound answers (F) average filtering time in ms (N) average number of nondeterministic choices required for each candidate answer

Preliminary Evaluation

We implemented our algorithm in a prototypical system, and we conducted a preliminary evaluation with the goal of showing that the number of consequences of $\mathsf{D}_\mathcal{K}$ is reasonably small, and that the nondeterminism of the filtering procedure is manageable. Our prototype uses the RDFox (Motik et al. 2014) system to materialise the consequences of $\mathsf{D}_\mathcal{K}$. We ran our tests on a MacBook Pro with 4GB of RAM and a 2.4Ghz Intel Core 2 Duo processor.

We tested our system using the version of the LSTW benchmark (Lutz et al. 2013) by Stefanoni, Motik, and Horrocks (2013). The TBox of the latter is in $\mathcal{ELHO}$, and we extended it to $\mathcal{ELHO}^s$ by making the role *subOganizationOf* transitive and by adding an axiom of type 5 and an axiom of type 7. We used the data generator provided by LSTW to generate KBs U5, U10, and U20 of 5, 10, and 20 universities, respectively. Finally, only query q_3^l from the LSTW benchmark uses transitive roles, so we have manually created four additional queries. Our system, the test data, and the queries are all available online.[1] We evaluated the practicality of our approach using the following two experiments.

First, we compared the size of the materialised consequences of $\mathsf{D}_\mathcal{K}$ with that of the input data. As the left-hand side of Table 2 shows, the ratio between the two is four, which, we believe, is acceptable in most practical scenarios.

Second, we measured the 'practical hardness' of our filtering step on our test queries. As the right-hand side of Table 2 shows, soundness of a candidate answer can typically be tested in as few as several milliseconds, and the test involves a manageable number of nondeterministic choices. Queries q_3^t and q_4^t were designed to obtain a lot of candidate answers with auxiliary individuals, so they retrieve many unsound answers. However, apart from query q_3^t, the percentage of the candidate answers that turned out to be unsound does not change with the increase in the size of the ABox. Therefore, while some queries may be challenging, we believe that our algorithm can be practicable in many cases.

4 Acyclic and Arborescent Queries

In this section, we prove that answering a simple class of tree-shaped acyclic CQs—which we call *arborescent*—over $\mathcal{ELHO}$ KBs is tractable, whereas answering acyclic queries is NP-hard. In addition, we show that extending $\mathcal{EL}$ with transitive or reflexive roles makes answering arborescent queries NP-hard. This is in contrast with the recent result by Bienvenu et al. (2013), who show that answering acyclic CQs over $\mathcal{ELH}$ KBs is tractable. We start by introducing acyclic and arborescent queries.

[1] http://www.cs.ox.ac.uk/isg/tools/EOLO/

Definition 15. *For q a Boolean CQ, $\mathsf{dg}_q = \langle N_V(q), E\rangle$ is a directed graph where $\langle x, y\rangle \in E$ for each $R(x, y) \in q$. Query q is* acyclic *if the graph obtained from dg_q by removing the orientation of edges is acyclic; q is* arborescent *if q contains no individuals and dg_q is a rooted tree with all edges pointing towards the root.*

Definition 16 and Theorem 17 show how to answer arborescent CQs over $\mathcal{ELHO}$ KBs in polynomial time. Intuitively, we apply the fork rule (cf. Definition 6) bottom-up, starting with the leaves of q and spread constraints upwards.

Definition 16. *Let $\mathcal{K}$ be an $\mathcal{ELHO}$ KB, let $\mathsf{D}_\mathcal{K}$ be the datalog program for $\mathcal{K}$, let $\mathsf{ind}_{\mathsf{D}_\mathcal{K}}$ and $\mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ be as specified in Definition 4, and let q be an arborescent query rooted in $r \in N_V(q)$. For each $y \in N_V(q)$ with $y \neq r$, and each $V \subseteq N_V(q)$, sets r_y and P_V are defined as follows.*

$$\mathsf{r}_y = \{R \in N_R \mid R(y, x) \in q \text{ with } x \text{ the parent of } y \text{ in } \mathsf{dg}_q\}$$
$$\mathsf{P}_V = \{y \in N_V(q) \mid \exists x \in V \text{ with } x \text{ the parent of } y \text{ in } \mathsf{dg}_q\}$$

Set RT is the smallest set satisfying the following conditions.

- *$\{r\} \in \mathsf{RT}$ and the level of $\{r\}$ is 0.*
- *For each set $V \in \mathsf{RT}$ with level n, we have $\mathsf{P}_V \in \mathsf{RT}$ and the level of P_V is $n + 1$.*
- *For each set $V \in \mathsf{RT}$ with level n and each $y \in \mathsf{P}_V$, we have $\{y\} \in \mathsf{RT}$ and the level of $\{y\}$ is $n + 1$.*

For each $V \in \mathsf{RT}$, set c_V contains each $u \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$ such that $\mathsf{D}_\mathcal{K} \models B(u)$ for each unary atom $B(x) \in q$ with $x \in V$. By reverse-induction on the level of the sets in RT, each $V \in \mathsf{RT}$ is associated with a set $\mathsf{A}_V \subseteq \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$.

- *For each set $V \in \mathsf{RT}$ of maximal level, let $\mathsf{A}_V = \mathsf{c}_V$.*
- *For $V \in \mathsf{RT}$ a set of level n where A_V is undefined but A_W has been defined for each $W \in \mathsf{RT}$ of level $n + 1$, let $\mathsf{A}_V = \mathsf{c}_V \cap (\mathsf{i}_V \cup \mathsf{a}_V)$, where i_V and a_V are as follows.*

$$\mathsf{i}_V = \{u \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \mid \forall y \in \mathsf{P}_V \exists u' \in \mathsf{A}_{\{y\}}.\mathsf{D}_\mathcal{K} \models \bigwedge_{R \in \mathsf{r}_y} R(u', u)\}$$
$$\mathsf{a}_V = \{u \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}} \mid \exists u' \in \mathsf{A}_{\mathsf{P}_V} \forall y \in \mathsf{P}_V.\mathsf{D}_\mathcal{K} \models \bigwedge_{R \in \mathsf{r}_y} \mathsf{d}_R(u', u)\}$$

Function $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ returns t if and only if $\mathsf{A}_{\{r\}}$ is nonempty.

Theorem 17. *For $\mathcal{K}$ a satisfiable $\mathcal{ELHO}$ KB and q an arborescent query, function $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ returns t if and only if $\Xi_\mathcal{K} \models q$. Furthermore, function $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ runs in time polynomial in the input size.*

Finally, we show that (unless PTIME = NP), answering arbitrary acyclic queries over $\mathcal{ELHO}$ KBs is harder than answering arborescent queries, and we show that adding transitive or reflexive roles to the DL $\mathcal{EL}$ makes answering arborescent queries intractable.

Theorem 18. *For $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ a KB and q a Boolean CQ, checking $\mathcal{K} \models q$ is* NP*-hard in each of the following cases.*

1. *The query q is acyclic and the TBox $\mathcal{T}$ is in $\mathcal{ELHO}$.*
2. *The query q is arborescent and the TBox $\mathcal{T}$ consists only of axioms of type* 1 *and* 7*, and of one axiom of type* 8*.*
3. *The query q is arborescent and the TBox $\mathcal{T}$ consists only of axioms of type* 1 *and* 7*, and of one axiom of type* 9*.*

5 Outlook

In future, we shall adapt our filtering procedure to detect unsound answers already during query evaluation. Moreover, we shall extend Algorithm 1 to handle complex role inclusions, thus obtaining a practicable approach for OWL 2 EL.

Acknowledgements

This work was supported by Alcatel-Lucent; the EU FP7 project OPTIQUE; and the EPSRC projects MASI[3], Score!, and DBOnto.

References

Abiteboul, S.; Hull, R.; and Vianu, V. 1995. *Foundations of Databases*. Addison-Wesley.

Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and Patel-Schneider, P. F., eds. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ envelope. In Kaelbling, L. P., and Saffiotti, A., eds., *IJCAI 2005*, 364–369.

Bienvenu, M.; Ortiz, M.; Simkus, M.; and Xiao, G. 2013. Tractable queries for lightweight description logics. In *IJCAI 2013*.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. of Automated Reasoning* 39(3):385–429.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; Poggi, A.; Rodriguez-Muro, M.; Rosati, R.; Ruzzi, M.; and Savo, D. F. 2011. The Mastro system for ontology-based data access. *Semantic Web Journal* 2(1):43–53.

Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; and Stein, C. 2009. *Introduction to Algorithms (3. ed.)*. MIT Press.

Cuenca Grau, B.; Horrocks, I.; Motik, B.; Parsia, B.; Patel-Schneider, P.; and Sattler, U. 2008. OWL 2: The next step for OWL. *Journal of Web Semantics* 6(4):309–322.

Dantsin, E.; Eiter, T.; Gottlob, G.; and Voronkov, A. 2001. Complexity and expressive power of logic programming. *ACM Computing Surveys* 33(3):374–425.

Eiter, T.; Lutz, C.; Ortiz, M.; and Simkus, M. 2009. Query answering in description logics with transitive roles. In *IJCAI 2009*, 759–764.

Eiter, T.; Ortiz, M.; Simkus, M.; Tran, T.-K.; and Xiao, G. 2012. Query rewriting for Horn-$\mathcal{SHIQ}$ plus rules. In *AAAI 2012*.

Fitting, M. 1996. *First-order logic and automated theorem proving (2nd ed.)*. Springer-Verlag New York, Inc.

Garey, M. R., and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co.

Glimm, B.; Horrocks, I.; Lutz, C.; and Sattler, U. 2008. Conjunctive query answering for the description logic $\mathcal{SHIQ}$. *Journal of Artif. Intell. Res.* 31:151–198.

Gottlob, G.; Kikot, S.; Kontchakov, R.; Podolskii, V. V.; Schwentick, T.; and Zakharyaschev, M. 2014. The price of query rewriting in ontology-based data access. *Artif. Intell.* 213:42–59.

Kontchakov, R.; Lutz, C.; Toman, D.; Wolter, F.; and Zakharyaschev, M. 2011. The combined approach to ontology-based data access. In Walsh, T., ed., *IJCAI 2011*, 2656–2661.

Krötzsch, M.; Rudolph, S.; and Hitzler, P. 2007. Conjunctive queries for a tractable fragment of OWL 1.1. In Aberer, K., et al., eds., *ISWC 2007*, 310–323.

Krötzsch, M.; Rudolph, S.; and Hitzler, P. 2008. ELP: Tractable rules for OWL 2. In Sheth, A., et al., eds., *ISWC 2008*, 649–664.

Krötzsch, M. 2010. Efficient inferencing for OWL EL. In *JELIA 2010*, volume 6341, 234–246.

Lutz, C.; Seylan, I.; Toman, D.; and Wolter, F. 2013. The combined approach to OBDA: Taming role hierarchies using filters. In *ISWC 2013*, volume 8218, 314–330.

Marnette, B. 2009. Generalized schema-mappings: from termination to tractability. In *PODS 2009*, 13–22.

Motik, B.; Nenov, Y.; Piro, R.; Horrocks, I.; and Olteanu, D. 2014. Parallel materialisation of datalog programs in centralised, main-memory RDF systems. In *AAAI 2014*.

Motik, B.; Shearer, R.; and Horrocks, I. 2009. Hypertableau reasoning for description logics. *Journal of Artif. Intell. Res.* 36:165–228.

Ortiz, M.; Rudolph, S.; and Simkus, M. 2011. Query answering in the Horn fragments of the description logics $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$. In Walsh, T., ed., *IJCAI 2011*, 1039–1044.

Rodriguez-Muro, M.; Kontchakov, R.; and Zakharyaschev, M. 2013. Ontology-based data access: Ontop of databases. In *ISWC 2013*, 558–573. Springer.

Stefanoni, G.; Motik, B.; and Horrocks, I. 2013. Introducing nominals to the combined query answering approaches for EL. In *AAAI 2013*.

Venetis, T.; Stoilos, G.; and Stamou, G. B. 2014. Query extensions and incremental query rewriting for OWL 2 QL ontologies. *J. Data Semantics* 3(1):1–23.

Yannakakis, M. 1981. Algorithms for acyclic database schemes. In *VLDB 1981*, 82–94.

A Skolem Chase and Universal Interpretations

In this section, we present a special variant of *Skolem chase* (Marnette 2009). Our chase uses *merging* (Abiteboul, Hull, and Vianu 1995) and *pruning* (Motik, Shearer, and Horrocks 2009) to deal with equality rules and guarantee that on rule base $\Xi_\mathcal{K}$ the so-called universal interpretation it produces satisfies the structural properties described in Section 3. We next formally present our chase variant.

Let Σ be a rule base and assume, w.l.o.g., that each variable occurring in Σ is quantified over exactly once in Σ; furthermore, let $>$ be an arbitrary total order on ground terms. Next, we first define some auxiliary notions, after which we define our Skolem chase variant and universal interpretations of Σ.

The *skolemisation of an existential rule* $\varphi(\vec{x},\vec{y}) \to \exists\vec{z}.\psi(\vec{x},\vec{z})$ is the formula ψ_{sk} obtained from ψ by substituting $f_z(\vec{x})$ for each $z \in \vec{z}$ with f_z a fresh function symbol of arity $|\vec{x}|$. A *chase instance* is a set $I = I_\Sigma \cup I_{\mathsf{eq}}$ where I_Σ is a set of ground atoms over the predicates in Σ, and I_{eq} is a set of assertions of the form $w \rightsquigarrow w'$ with $\rightsquigarrow$ a fresh predicate. For each term w, let $\|w\|_I = w$, if no term w' exists such that $w \rightsquigarrow w' \in I$; otherwise, let $\|w\|_I = w'$ where w' is the smallest term in the ordering $>$ for which terms $w_0, \ldots, w_n$ with $w_0 = w$ and $w_n = w'$ exist such that $w_{i-1} \rightsquigarrow w_i \in I$ for each $i \in [1, n]$. For a formula α, $\|\alpha\|_I$ is the result of uniformly substituting each term w occurring in α with $\|w\|_I$.

An *existential rule* $\varphi(\vec{x},\vec{y}) \to \exists\vec{z}.\psi(\vec{x},\vec{z})$ *is applicable* to a chase instance $I = I_\Sigma \cup I_{\mathsf{eq}}$, if a substitution σ exists such that $\sigma(\varphi) \subseteq I$ and $\|\sigma(\psi_{\mathsf{sk}})\|_I \not\subseteq I$; the *result* of applying such a rule to I is obtained by adding $\|\sigma(\psi_{\mathsf{sk}})\|_I$ to I_Σ. An *equality rule* $\varphi(\vec{x}) \to s \approx t$ *is applicable* to a chase instance $I = I_\Sigma \cup I_{\mathsf{eq}}$, if a substitution σ exists such that $\sigma(\varphi) \subseteq I$ and $\|\sigma(s)\|_I \neq \|\sigma(t)\|_I$; for $\{w, w'\} = \{\|\sigma(s)\|_I, \|\sigma(t)\|_I\}$ such that $w > w'$, the *result* of applying such a rule to I is the chase instance $I'_\Sigma \cup I'_{\mathsf{eq}}$ where I'_{eq} is the result of adding $w \rightsquigarrow w'$ to I_{eq}, and I'_Σ is obtained by removing all atoms occurring in I_Σ that contain a term w_2 with w a proper subterm of w_2, and by replacing each occurrence of w in the resulting instance with w'.

A *chase for* Σ *w.r.t.* $>$ is a sequence of chase instances $I_0, I_1, \ldots$ where I_0 contains each ground atom in Σ, and, for each $i \geq 1$, chase instance I_{i+1} is the result of an (arbitrarily chosen) rule in Σ applicable to I_i, and $I_{i+1} = I_i$, if no rule is applicable to I_i. This sequence must be *fair*—that is, if a rule $\varphi \to \psi$ in Σ is applicable to some I_i under a specific substitution σ and $\sigma(\varphi) \subseteq I_j$, for each $j \geq i$, then $k \geq i$ exists such that I_{k+1} is the result of $\varphi \to \psi$ on I_k w.r.t. substitution σ. Set $I = \bigcup_{i\in\mathbb{N}} \bigcap_{j\geq i} I_j$ is a *universal interpretation of* Σ *(w.r.t.* $>$*)* and its *domain* is the set Δ^I containing $\|w\|_I$ for each term w that occurs in I.

Please note that for each term w that occurs in I in at least one atom over a predicate from Σ, we have $\|w\|_I = w$ and $w \in \Delta^I$. Also, owing to the fairness of the chase sequence, no rule in Σ is applicable to I. Furthermore, if Σ does not contain equality rules, then I is independent from the order in which rules are applied, and so it is *the universal interpretation* of Σ. Otherwise, if Σ contains equality rules, then I is homomorphically equivalent w.r.t. the predicates occurring in Σ to any other universal interpretation I' of Σ, although I and I' may disagree on the assertions over $\rightsquigarrow$. Finally, it is well known (Marnette 2009) that I can be homomorphically embedded into any model of Σ; so I can be used to answer arbitrary CQs over Σ.

Fact 19. *For each CQ q and each substitution π, $\Sigma \models \pi(q)$ if and only if a substitution π_* with* $\mathsf{dom}(\pi_*) = N_V(q)$ *exists such that $\pi \subseteq \pi_*$ and $\|\pi_*(q)\|_I \subseteq I$.*

A.1 Universal Interpretations of $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$

Let $\mathcal{K}$ be an $\mathcal{ELHO}^s$ knowledge base, and let $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$ be the rule base and the datalog program associated with $\mathcal{K}$, respectively. In the rest of this appendix, we shall make the two following assumptions. First, we associate to each rule $A_1(x) \to \exists z.R(x,z) \wedge A(z)$ in $\Xi_\mathcal{K}$ a fresh unary function symbol $f^{A_1}_{R,A}$, and we assume that the skolemisation of such a rule is given by $R(x, f^{A_1}_{R,A}(x)) \wedge A(f^{A_1}_{R,A}(x))$. Second, we assume that the chase for $\Xi_\mathcal{K}$ and for $\mathsf{D}_\mathcal{K}$, respectively, is w.r.t. the total order $>$ specified in Definition 4, and that $f^{A_1}_{R,A}(w) > a$ for each ground term w and each individual a.

B Proof of Proposition 2

We fix an $\mathcal{ELHO}^s$ knowledge base $\mathcal{K}$. Let $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$ be the rule base and the datalog program associated with $\mathcal{K}$, respectively; moreover, let I and J be universal interpretations of $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$, respectively.

We next define a function δ that maps each term w occurring in I to a term $\delta(w)$ as follows:

$$\delta(w) = \begin{cases} w & \text{if } w \in N_I, \\ o_{R,A} & \text{if } w \text{ is of the form } w = f^{A_1}_{R,A}(w'). \end{cases}$$

The following two Lemmas show that δ establishes a tight connection between I and J. The proofs of these results is given in Appendix C.

Lemma 20. *Mapping δ satisfies the following properties for all terms w_1 and w_2 occurring in I, each individual $a \in N_I$, each role $R \in N_R$, and each concept $C \in N_C \cup \{\top, \bot\}$.*

H1. $C(w_1) \in I$ *implies that* $C(\delta(w_1)) \in J$.

H2. $R(w_1, w_2) \in I$ *implies that* $R(\delta(w_1), \delta(w_2)) \in J$.

H3. $R(w_1, w_2) \in I$ *and* w *is of the form* $f^{A_1}_{P,A}(w_1)$ *imply that* $\mathsf{d}_R(\delta(w_1), \delta(w_2)) \in J$.

H4. $\|w_1\|_I = a$ *implies that* $\|\delta(w_1)\|_J = a$.

Lemma 21. *Mapping* δ *satisfies the following properties for all terms* w_1 *and* w_2 *occurring in* I*, each individual* $a \in N_I$*, each role* $R \in N_R$*, and each concept* $C \in N_C \cup \{\top, \bot\}$*.*

D1. $C(\delta(w_1)) \in J$ *implies that* $C(w_1) \in I$.

D2. $\mathsf{Self}_R(\delta(w_1)) \in J$ *implies that* $R(w_1, w_1) \in I$.

D3. $R(\delta(w_1), \delta(w_2)) \in J$ *and* $\delta(w_2) \in N_I$ *imply that* $R(w_1, w_2) \in I$.

D4. $R(\delta(w_1), \delta(w_2)) \in J$ *and* $\delta(w_2)$ *is of the form* $o_{P,A}$ *imply that*

- *a term* $w'_1 \in \Delta^I$ *exists such that* $R(w'_1, w_2) \in I$*, and*
- *a term* $w'_2 \in \Delta^I$ *exists such that* $\delta(w'_2) = o_{P,A}$ *and* $R(w_1, w'_2) \in I$.

D5. $\mathsf{d}_R(\delta(w_1), \delta(w_2)) \in J$ *and* $\delta(w_2)$ *is of the form* $o_{P,A}$ *imply that*

- *a term* $w'_3 \in \Delta^I$ *of the form* $f^{A_1}_{P,A}(w_1)$ *exists such that* $P(w_1, w'_3) \in I$*, and*
- $P \sqsubseteq^*_{\mathcal{T}} R$.

D6. $\|\delta(w_1)\|_J = \delta(a)$ *implies that* $\|w_1\|_I = a$.

D7. For each individual $u \in \Delta^J$*, term* $w \in \Delta^I$ *exists such that* $\delta(w) = u$.

We are now ready to show that $\mathsf{D}_{\mathcal{K}}$ can be used to check the satisfiability of $\mathcal{K}$.

Proposition 2. $\mathcal{K}$ *is unsatisfiable iff* $\mathsf{D}_{\mathcal{K}} \models \exists y.\bot(y)$.

Proof. By the definition of $\mathcal{ELHO}^s$ semantics, $\mathcal{K}$ is unsatisfiable if and only if $\Xi_{\mathcal{K}} \models \exists x.\bot(x)$. By Lemmas 20 and 21, for each term $w \in \Delta^I$, we have $\bot(w) \in I$ if and only if $\bot(\delta(w)) \in J$. Consequently, $\mathcal{K}$ is unsatisfiable if and only if $\mathsf{D}_{\mathcal{K}} \models \exists x.\bot(x)$. □

C Proofs of Lemmas 20 and 21

We prove Lemmas 20 and 21 in various stages. To start, we next show that mapping δ satisfies a slightly relaxed version of properties H1–H4 from Lemma 20.

Lemma 22. *Mapping* δ *satisfies the following four properties for all terms* w_1 *and* w_2 *occurring in* I*, each individual* $a \in N_I$*, each role* $R \in N_R$*, and each concept* $C \in N_C \cup \{\top, \bot\}$*.*

(i) $C(w_1) \in I$ *implies that* $\|C(\delta(w_1))\|_J \in J$.

(ii) $R(w_1, w_2) \in I$ *implies that* $\|R(\delta(w_1), \delta(w_2))\|_J \in J$.

(iii) $R(w_1, w_2) \in I$ *and* w *is of the form* $f^{A_1}_{P,A}(w_1)$ *imply that* $\|\mathsf{d}_R(\delta(w_1), \delta(w_2))\|_J \in J$.

(iv) $w_1 \rightsquigarrow a \in I$ *implies that* $\|\delta(w_1)\|_J = \|a\|_J$.

Proof. Let $I_0, I_1, \ldots$ be the chase for $\Xi_{\mathcal{K}}$ w.r.t. $>$ used to construct I. We show by induction on this sequence that each I_n satisfies the properties.

Base case. Consider chase instance I_0. By the definition, I_0 contains only ground atoms constructed using the predicates in $N_C \cup \{\top, \bot, \mathsf{ind}\} \cup N_R$ and the individuals in N_I. Therefore, for each term w occurring in I_0, we have $\|w\|_{I_0} = w$ and $w \in N_I$, and so $\delta(w) = w$. Consider an arbitrary atom $\phi \in I_0$. By the definition of $\mathsf{D}_{\mathcal{K}}$, ϕ is an atom in $\mathsf{D}_{\mathcal{K}}$. Since J satisfies all ground atoms in $\mathsf{D}_{\mathcal{K}}$, we have $\|\phi\|_J \in J$, so properties (i)–(iv) hold.

Inductive step. Consider an arbitrary $n \in \mathbb{N}$ and assume that I_n satisfies properties (i)–(iv). By considering each rule in $\Xi_{\mathcal{K}}$, we assume that the rule is applicable to I_n, and we show that the properties hold for all fresh atoms in the resulting instance.

(Datalog Rule) Consider a datalog rule $\varphi(\vec{x}, \vec{y}) \rightarrow \psi(\vec{x})$ in $\Xi_{\mathcal{K}}$, and assume that a substitution σ with $\mathsf{dom}(\sigma) = \vec{x} \cup \vec{y}$ exists such that $\sigma(\varphi) \subseteq I_n$. Let σ' be the substitution such that $\sigma'(x) = \delta(\sigma(x))$ for each variable $x \in \vec{x} \cup \vec{y}$. By the inductive hypothesis, we have $\|\sigma'(\psi)\|_J \subseteq J$. Since $\varphi(\vec{x}, \vec{y}) \rightarrow \psi(\vec{x}) \in \mathsf{D}_{\mathcal{K}}$ and no rule is applicable to J, we have $\|\sigma'(\psi)\|_J \subseteq J$.

(Existential Rule) Consider $A_1(x) \rightarrow \exists z.R(x, z) \wedge A(z)$ in $\Xi_{\mathcal{K}}$, assume that $A_1(w_1) \in I_n$, and let $w_2 = \|f^{A_1}_{R,A}(w_1)\|_{I_n}$. By the inductive hypothesis, $\|A_1(\delta(w_1))\|_J \in J$ holds. Program $\mathsf{D}_{\mathcal{K}}$ contains $A_1(x) \rightarrow R(x, o_{R,A}) \wedge \mathsf{d}_R(x, o_{R,A}) \wedge A(o_{R,A})$ and no rule is applicable to J, so $\|R(\delta(w_1), o_{R,A}) \wedge \mathsf{d}_R(\delta(w_1), o_{R,A}) \wedge A(o_{R,A})\|_J \subseteq J$ holds. We distinguish two cases.

- $w_2 = f^{A_1}_{R,A}(w_1)$. Then, $\delta(w_2) = o_{R,A}$, and so $\|o_{R,A}\|_J = \|\delta(w_2)\|_J$.
- $w_2 \neq f^{A_1}_{R,A}(w_1)$. By the form of equality rules in $\Xi_{\mathcal{K}}$ and due to $w_2 = \|f^{A_1}_{R,A}(w_1)\|_{I_n}$, we have $w_2 \in N_I$. Then, terms $u_0, \ldots, u_n$ with $u_0 = f^{A_1}_{R,A}(w_1)$ and $u_n = w_2$ exist in I_n such that $u_{i-1} \rightsquigarrow u_i \in I_n$ for each $i \in [1, n]$. By the inductive hypothesis, we have $\|\delta(f^{A_1}_{R,A}(w_1))\|_J = \|\delta(w_2)\|_J$, and so $\|o_{R,A}\|_J = \|\delta(w_2)\|_J$.

As stated above, we have $\|R(\delta(w_1), o_{R,A}) \wedge \mathsf{d}_R(\delta(w_1), o_{R,A}) \wedge A(o_{R,A})\|_J \subseteq J$, so properties (i)–(iii) are satisfied.

(Equality Rule) Consider a rule $A(x) \to x \approx a$ in $\Xi_\mathcal{K}$ and assume that $A(w_1) \in I_n$. As $w_1 \in \Delta^{I_n}$, we have $\|w_1\|_{I_n} = w_1$. Then let terms w and w' be such that $\{w, w'\} = \{w_1, \|a\|_{I_n}\}$ and $w > w'$. By the inductive hypothesis, we either have $\|A(\delta(w))\|_J \in J$ and $\|\delta(w')\|_J = \|a\|_J$, or $\|A(\delta(w'))\|_J \in J$ and $\|\delta(w)\|_J = \|a\|_J$. In either cases, since $A(x) \to x \approx a \in \mathsf{D}_\mathcal{K}$ and no rule is applicable to J, we have $\|\delta(w)\|_J = \|\delta(w')\|_J$, and property (iv) is satisfied. We next consider the atoms in I_n that get replaced by the application of this rule.

- $C(w) \in I_n$. By the inductive hypothesis, we have $\|C(\delta(w))\|_J \in J$; thus $\|C(\delta(w'))\|_J \in J$.
- $R(w, w) \in I_n$. By the inductive hypothesis, we have $\|R(\delta(w), \delta(w))\|_J \in J$; thus $\|R(\delta(w'), \delta(w'))\|_J \in J$.
- $R(w, w_2) \in I_n$. By the inductive hypothesis, we have $\|R(\delta(w), \delta(w_2))\|_J \in J$; thus $\|R(\delta(w'), \delta(w_2))\|_J \in J$.
- $R(w_2, w) \in I_n$. By the inductive hypothesis, we have $\|R(\delta(w_2), \delta(w))\|_J \in J$; thus, $\|R(\delta(w_2), \delta(w'))\|_J \in J$. □

We next show that J satisfies a slightly relaxed version of properties D1–D7 from Lemma 21.

Lemma 23. *Mapping δ satisfies the following properties for all terms w_1 and w_2 occurring in I, each individual $a \in N_I$, each role $R \in N_R$, and each concept $C \in N_C \cup \{\top, \bot\}$.*

(a) $C(\delta(w_1)) \in J$ *implies that* $\|C(w_1)\|_I \in I$.
(b) $\mathsf{Self}_R(\delta(w_1)) \in J$ *implies that* $\|R(w_1, w_1)\|_I \in I$.
(c) $R(\delta(w_1), \delta(w_2)) \in J$ *and* $\delta(w_2) \in N_I$ *imply that* $\|R(w_1, w_2)\|_I \in I$.
(d) $R(\delta(w_1), \delta(w_2)) \in J$ *and* $\delta(w_2)$ *is of the form* $o_{P,A}$ *imply that*
  - *a term w_1' from I exists such that $\|R(w_1', w_2)\|_I \in I$, and*
  - *a term w_2' from I exists such that $\delta(w_2') = o_{P,A}$ and $\|R(w_1, w_2')\|_I \in I$.*

(e) $\mathsf{d}_R(\delta(w_1), \delta(w_2)) \in J$ *and* $\delta(w_2)$ *is of the form* $o_{P,A}$ *imply that*
  - *a term w_3' from I of the form $f^{A_1}_{P,A}(\|w_1\|_I)$ exists such that $\|P(w_1, w_3')\|_I \in I$, and*
  - $P \sqsubseteq^*_\mathcal{T} R$.

(f) $\delta(w_1) \rightsquigarrow a \in J$ *implies that* $\|w_1\|_I = \|a\|_I$.
(g) *For each individual u occurring in J, a term w occurring in I exists such that $\delta(w) = u$.*

Proof. Let $J_0, J_1, \ldots$ be the chase for $\mathsf{D}_\mathcal{K}$ used to construct J. We prove by induction on this sequence that each J_n satisfies the properties.

Base case. Consider J_0. By the definition, J_0 does not contain assertions over $\rightsquigarrow$. Moreover, $\mathsf{D}_\mathcal{K}$ and $\Xi_\mathcal{K}$ contain the same ground atoms, all of which are constructed using the individuals from N_I and the predicates in $N_C \cup \{\top, \bot, \mathsf{ind}\} \cup N_R$. Finally, I satisfies all the ground atoms in $\Xi_\mathcal{K}$, so properties (a)–(g) are satisfied.

Inductive step. Consider an arbitrary $n \in \mathbb{N}$ and assume that J_n satisfies properties (a)–(g). By considering each rule in $\mathsf{D}_\mathcal{K}$, we assume that the rule is applicable to J_n, and we show that the properties hold for all fresh atoms in the resulting instance.

Consider a rule of the form $\mathsf{Self}_P(x) \to \mathsf{Self}_R(x)$ in $\mathsf{D}_\mathcal{K}$ and assume that $\mathsf{Self}_P(\delta(w_1)) \in J_n$. By the inductive hypothesis, we have $\|\mathsf{Self}_P(w_1) \wedge P(w_1, w_1)\|_I \subseteq I$. By the definition of program $\mathsf{D}_\mathcal{K}$ and $\Xi_\mathcal{K}$, we have $P \sqsubseteq R \in \mathcal{T}$, and therefore rules $\mathsf{Self}_P(x) \to \mathsf{Self}_R(x)$ and $P(x, y) \to R(x, y)$ are contained in $\Xi_\mathcal{K}$. Since no rule is applicable to I, we have $\|\mathsf{Self}_R(w_1) \wedge R(w_1, w_1)\|_I \subseteq I$.

Consider a rule of the form $\varphi(x) \to B(x)$ in $\mathsf{D}_\mathcal{K}$ with φ a conjunction of unary atoms over variable x and B a concept. Let σ and σ' be substitutions such that $\sigma(x) = \delta(w_1)$ and $\sigma'(x) = w_1$. Assume that $\sigma(\varphi) \subseteq J_n$. By the inductive hypothesis, we have $\|\sigma'(\varphi)\|_I \subseteq I$. Since no rule is applicable to I, we have $\|B(w_1)\|_I \in I$.

Consider a rule of the form $A(x) \to x \approx a$ and assume that $A(\delta(w_1)) \in J_n$. As $\delta(w_1) \in \Delta^{J_n}$, we have $\|\delta(w_1)\|_{J_n} = \delta(w_1)$. Let w and w' be terms such that $\{\delta(w), \delta(w')\} = \{\delta(w_1), \|a\|_{J_n}\}$ and $\delta(w) > \delta(w')$. By the inductive hypothesis, we either have $\|A(w)\|_I \in I$ and $\|w'\|_I = \|a\|_I$, or $\|A(w')\|_I \in I$ and $\|w\|_I = \|a\|_I$. In either cases, since no rule is applicable to I, we have $\|w\|_I = \|w'\|_I$. We next consider the atoms in J_n that get replaced by the application of this rule.

- $\mathsf{Self}_R(\delta(w)) \in J_n$. By the inductive hypothesis, we have $\|\mathsf{Self}_R(w) \wedge R(w, w)\|_I \subseteq I$; so $\|\mathsf{Self}_R(w') \wedge R(w', w')\|_I \subseteq I$.
- $C(\delta(w)) \in J_n$. By the inductive hypothesis, we have $\|C(w)\|_I \in I$, and so $\|C(w')\|_I \in I$.
- $R(\delta(w), \delta(w)) \in J_n$. We distinguish two cases.
  - $\delta(w) \in N_I$. By the inductive hypothesis, we have $\|R(w, w)\|_I \in I$, and so $\|R(w', w')\|_I \in I$.
  - $\delta(w)$ is of the form $o_{P,A}$. By the inductive hypothesis, a term w_2' exists such that $\delta(w_2') = o_{P,A}$ and $\|R(w, w_2')\|_I \in I$. Since $\delta(w_2') = \delta(w)$, we have $\|w_2'\|_I = \|w'\|_I$, and so $\|R(w', w')\|_I \in I$.
- $R(\delta(w_2), \delta(w)) \in J_n$. We distinguish two cases.
  - $\delta(w) \in N_I$. By the inductive hypothesis, we have $\|R(w_2, w)\|_I \in I$, and so $\|R(w_2, w')\|_I \in I$.

- $\delta(w)$ is of the form $o_{P,A}$. By the inductive hypothesis, a term w_2' exists such that $\delta(w_2') = o_{P,A}$ and $\|R(w_2, w_2')\|_I \in I$. Since $\delta(w_2') = \delta(w)$, we have $\|w_2'\|_I = \|w'\|_I$, and so $\|R(w_2, w')\|_I \in I$.

- $R(\delta(w), \delta(w_2)) \in J_n$. We distinguish two cases.
  - $\delta(w_2) \in N_I$. By the inductive hypothesis, $\|R(w, w_2)\|_I \in I$, thus $\|R(w', w_2)\|_I \in I$.
  - $\delta(w_2)$ is of the form $o_{P,A}$. By the inductive hypothesis, a term w_2' with $\delta(w_2') = o_{P,A}$ exists such that $\|R(w, w_2')\|_I \in I$, and so $\|R(w', w_2')\|_I \in I$.
- $\mathsf{d}_R(\delta(w), \delta(w_2)) \in J_n$ and $\delta(w_2)$ is of the form $o_{P,A}$. Property (e) is satisfied by the inductive hypothesis.

Consider a rule of the form $R(x, y) \wedge A_1(y) \rightarrow A(x)$ in $\mathsf{D}_\mathcal{K}$ and assume that $\{R(\delta(w_1), \delta(w_2)), A_1(\delta(w_2))\} \subseteq J_n$. We distinguish two cases.

- $\delta(w_2) \in N_I$. By the inductive hypothesis, we have $\|R(w_1, w_2)\|_I \subseteq I$.
- $\delta(w_2) = o_{P,A}$. By the inductive hypothesis, a term w_2' exists such that $\delta(w_2') = o_{P,A}$ and $\|R(w_1, w_2') \wedge A(w_2')\|_I \in I$.

In either cases, since no rule is applicable to I, we have $\|A(w_1)\|_I \in I$.

Consider a rule $T(x, y) \rightarrow R(x, y)$ in $\mathsf{D}_\mathcal{K}$ and assume that $T(\delta(w_1), \delta(w_2)) \in J_n$. We distinguish two cases.

- $\delta(w_2) \in N_I$. By the inductive hypothesis, $\|T(w_1, w_2)\|_I \in I$. As no rule is applicable to I, we have $\|R(w_1, w_2)\|_I \in I$.
- $\delta(w_2)$ is of the form $o_{P,A}$. By the inductive hypothesis, we have that terms w_1' and w_2' exist such that $\delta(w_2') = o_{P,A}$ and $\|T(w_1, w_2') \wedge T(w_1', w_2)\|_I \subseteq I$. Since no rule is applicable to I, we have $\|R(w_1, w_2') \wedge R(w_1', w_2)\|_I \subseteq I$.

Consider a rule $\mathsf{d}_T(x, y) \rightarrow \mathsf{d}_R(x, y)$ in $\mathsf{D}_\mathcal{K}$ and assume that $\mathsf{d}_T(\delta(w_1), \delta(w_2)) \in J_n$ and $\delta(w_2)$ is of the form $o_{P,A}$. By the inductive hypothesis, a term w_3 of the form $f_{P,A}^{A_1}(\|w_1\|_I)$ exists such that $\|P(w_1, w_3')\|_I \in I$ and $P \sqsubseteq_\mathcal{T}^* T$. By the definition of $\mathsf{D}_\mathcal{K}$, we have $T \sqsubseteq R \in \mathcal{T}$. Consequently, we have $P \sqsubseteq_\mathcal{T}^* R \in \mathcal{T}$ and property (e) holds.

Consider a rule $R(x, y) \rightarrow A(y)$ in $\mathsf{D}_\mathcal{K}$, and assume that $R(\delta(w_1), \delta(w_2)) \in J_n$.

- $\delta(w_2) \in N_I$. By the inductive hypothesis, we have $\|R(w_1, w_2)\|_I \subseteq I$.
- $\delta(w_2)$ is of the form $o_{P,A}$. By the inductive hypothesis, a term w_1' exists such that $\|R(w_1', w_2)\|_I \in I$.

In either cases, since no rule is applicable to I, we have $\|A(w_2)\|_I \in I$.

Consider a rule $A_1(x) \rightarrow R(x, o_{R,A}) \wedge \mathsf{d}_R(x, o_{R,A}) \wedge A(o_{R,A})$ in $\mathsf{D}_\mathcal{K}$, and assume that $A(\delta(w_1)) \in J_n$. Let w_2 be a term such that $\delta(w_2) = \|o_{R,A}\|_{J_n}$. By the inductive hypothesis, we have $\|A(w_1)\|_I \in I$. By the definition of $\mathsf{D}_\mathcal{K}$, rule base $\Xi_\mathcal{K}$ contains $A_1(x) \rightarrow \exists z.R(x, z) \wedge A(z)$. Let $w_3' = f_{R,A}^{A_1}(\|w_1\|_I)$. Since no rule is applicable to I, we have $\|R(w_1, w_3') \wedge A(w_3')\|_I \subseteq I$. We distinguish two cases.

- $\delta(w_2) \neq o_{R,A}$. By the form of equality rules occurring in $\mathsf{D}_\mathcal{K}$, we have $\delta(w_2) \in N_I$. Then individuals $u_0, \ldots, u_n$ with $u_0 = o_{R,A}$ and $u_n = \delta(w_2)$ exist such that $u_{i-1} \rightsquigarrow u_i \in J_n$ for each $i \in [1, n]$. Moreover, given that $\delta(w_3') = o_{R,A}$, by the inductive hypothesis, we have $\|w_3'\|_I = \|w_2\|_I$. Hence, we have $\|R(w_1, w_2) \wedge A(w_2)\|_I \subseteq I$. By the inductive hypothesis, property (g) is satisfied.
- $\delta(w_2) = o_{R,A}$. Then w_2 is of the form $f_{R,A}^{A_2}(w')$. Because such term can only be introduced in I by the application of a rule of the form $A_2(x) \rightarrow \exists z.R(x, z) \wedge A(z)$, a term w_1' must exist such that $\|R(w_1', w_2) \wedge A(w_2)\|_I \subseteq I$. As stated above, we also have $\|R(w_1, w_3')\|_I \in I$. By the reflexivity of $\sqsubseteq_\mathcal{T}^*$, we also have $R \sqsubseteq_\mathcal{T}^* R$, so properties (d) and (e) are satisfied. As $\delta(w_3') = o_{R,A}$, property (g) is also satisfied.

Consider a rule $R(x, y) \wedge R(y, z) \rightarrow R(x, z)$, and assume $\{R(\delta(w_1), \delta(w_2)), R(\delta(w_3), \delta(w_4))\} \subseteq J_n$ and $\delta(w_2) = \delta(w_3)$. It follows that $R(\delta(w_2), \delta(w_4)) \in J_n$. We distinguish four cases.

- $\delta(w_2) \in N_I$ and $\delta(w_4) \in N_I$. By the inductive hypothesis, we have $\|R(w_1, w_2) \wedge R(w_2, w_4)\|_I \in I$. Since no rule is applicable to I, we have $\|R(w_1, w_4)\|_I \in I$.
- $\delta(w_2)$ is of the form $o_{P,A}$ and $\delta(w_4) \in N_I$. By the inductive hypothesis, a term w_2' exists such that $\delta(w_2') = o_{P,A}$ and $\|R(w_1, w_2')\|_I \in I$. Due to $\delta(w_2') = \delta(w_2)$, we have $\|R(\delta(w_2'), \delta(w_4))\|_{J_n} \in J_n$. By the inductive hypothesis, we have $\|R(w_2', w_4)\|_I \in I$. Since no rule is applicable to I, we have $\|R(w_1, w_4)\|_I \in I$.
- $\delta(w_2) \in N_I$ and $\delta(w_4)$ is of the form $o_{R,B}$. By the inductive hypothesis, we have $\|R(w_1, w_2)\|_I \in I$ and terms w_2' and w_4' exist such that $\delta(w_4') = o_{R,B}$ and $\|R(w_2', w_4) \wedge R(w_2, w_4')\|_I \subseteq I$. As no rule is applicable to I, we have $\|R(w_1, w_4')\|_I \in I$.
- $\delta(w_2)$ is of the form $o_{P,A}$ and $\delta(w_4)$ is of the form $o_{R,B}$. By the inductive hypothesis, a term w_2' exists such that $\delta(w_2') = o_{P,A}$ and $\|R(w_1, w_2')\|_I \in I$. Due to $\delta(w_2') = \delta(w_2)$, we have $\|R(\delta(w_2'), \delta(w_4))\|_{J_n} \in J_n$. By the inductive hypothesis, terms u_2' and w_4' exist such that $\delta(w_4') = o_{R,B}$ and $\|R(u_2', w_4) \wedge R(w_2', w_4')\|_I \subseteq I$. As no rule is applicable to I, we have $\|R(w_1, w_4')\|_I \in I$.

Consider a rule $A(x) \to R(x,x) \wedge \mathsf{Self}_R(x)$, assume $A(\delta(w_1)) \in J_n$, and let w_2 be a term such that $\delta(w_2) = \delta(w_1)$. By the inductive hypothesis, we have $\|A(w_1) \wedge A(w_2)\|_I \subseteq I$. Since no rule is applicable to I, we have $\|R(w_1, w_1) \wedge \mathsf{Self}_R(w_1)\|_I \subseteq I$ and $\|R(w_2, w_2) \wedge \mathsf{Self}_R(w_2)\|_I \subseteq I$, so property (a) holds. We distinguish two cases.

- $\delta(w_2) \in N_I$. Then individuals $u_0, \ldots, u_n$ with $u_0 = \delta(w_1)$ and $u_n = \delta(w_2)$ exist such that $u_{i-1} \rightsquigarrow u_i \in J_n$ for each $i \in [1, n]$. By the inductive hypothesis, we have, $\|w_1\|_I = \|w_2\|_I$; thus $\|R(w_1, w_2)\|_I \in I$.
- $\delta(w_2)$ is of the form $o_{P,A}$. Thus, $\delta(w_1) = o_{P,A}$. As stated above, we have $\|R(w_1, w_1) \wedge R(w_2, w_2)\|_I \subseteq I$.

Consider a rule $R(x,x) \wedge \mathsf{ind}(x) \to \mathsf{Self}_R(x)$. Assume that $\{R(\delta(w_1), \delta(w_2)), \mathsf{ind}(\delta(w_2))\} \subseteq J_n$ and $\delta(w_2) = \delta(w_1)$. By the definition of $\mathsf{cls}_\mathcal{K}$ and since no rule in $\mathsf{D}_\mathcal{T}$ derives atoms over ind, we have that $\delta(w_2) \in N_I$. By the inductive hypothesis, we then have $\|R(w_1, w_2) \wedge \mathsf{ind}(w_2)\|_I \subseteq I$. Since δ is the identity on N_I and $\delta(w_1) = \delta(w_2)$, we have $w_1 = w_2$, and so $\|R(w_1, w_1) \wedge R(w_2, w_2)\|_I \subseteq I$. As no rule is applicable to I, we have $\|\mathsf{Self}_R(w_1) \wedge \mathsf{Self}_R(w_2)\|_I \subseteq I$. □

Lemmas 20 and 21 follow immediately from Lemmas 22 and 23, and the following result.

Lemma 24. *For each term w occurring in I and each individual $a \in N_I$, the following two properties hold.*

E1. $\|w\|_I = a$ *if and only if* $\|\delta(w)\|_J = a$.

E2. $\|w\|_I = w$ *if and only if* $\|\delta(w)\|_J = \delta(w)$.

Proof. By the definition of $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$, each equality rule occurring in these rule bases is of the form $A(x) \to x \approx a$ with $a \in N_I$. Consequently, for all terms u and u' such that $u \rightsquigarrow u' \in I \cup J$, we have $u' \in N_I$. Let w be an arbitrary term occurring in I, and let $a \in N_I$ be an arbitrary individual.

We first prove E1. ($\Rightarrow$) Assume that $\|w\|_I = a$. Then terms $w_0, \ldots, w_n$ with $w_0 = w$ and $w_n = a$ exist such that for each $i \in [1, n]$ we have $w_i \in N_I$ and $w_{i-1} \rightsquigarrow w_i \in I$. By Lemma 22, we have $\|\delta(w_0)\|_J = \|\delta(w_n)\|_J$; that is, $\|\delta(w)\|_J = \|a\|_J$. We show that $\|a\|_J = a$. Assume the opposite; hence, an individual b exists such that $a \rightsquigarrow b \in J$ and $a > b$. By Lemma 23, we have $\|a\|_I = \|b\|_I$. Since $a > b$, we have $\|w\|_I = b$, which is a contradiction. ($\Leftarrow$) Assume that $\|\delta(w)\|_I = a$. Then individuals $u_0, \ldots, u_n$ with $u_0 = \delta(w)$ and $u_n = a$ exist such that for each $i \in [1, n]$ we have $u_i \in N_I$ and $u_{i-1} \rightsquigarrow u_i \in I$. By Lemma 23, we have $\|w\|_I = \|a\|_I$. We show that $\|a\|_I = a$. Assume the opposite; hence, an individual b exists such that $a \rightsquigarrow b \in I$ and $a > b$. By Lemma 22, we have $\|a\|_J = \|b\|_J$. Since $a > b$, we have $\|\delta(w)\|_I = b$, which is a contradiction.

Next, we prove property E2 by contraposition. ($\Rightarrow$) Assume that $\|\delta(w)\|_J \neq \delta(w)$; hence, an individual $b \in N_I$ exists such that $\delta(w) \neq b$ and $\|\delta(w)\|_J = b$. By the definition of δ, we have $w \neq b$. By property E1, we have $\|w\|_I = b$, as required. ($\Leftarrow$) Assume that $\|w\|_I \neq w$; hence, an individual $b \in N_I$ exists such that $w \neq b$ and $\|w\|_I = b$. By the definition of δ, we have $\delta(w) \neq b$. By property E1, we have $\|\delta(w)\|_I = b$, as required. □

D Proof of Theorem 12

Let KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a satisfiable $\mathcal{ELHO}^s$ KB, let $\Xi_\mathcal{K}$, and $\mathsf{D}_\mathcal{K}$ be the rule base and the datalog program associated with $\mathcal{K}$, respectively; moreover, let I and J be universal interpretations of $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$, respectively. To prove Theorem 12, we first show that our function is sound, after which we show that it is also complete.

D.1 Soundness

Let $q' = \exists \vec{y}.\, \psi(\vec{x}, \vec{y})$ be a CQ, let τ' be a candidate answer to q' over $\mathsf{D}_\mathcal{K}$, and let $\pi' = \tau'|_{\vec{x}}$. Assume that the two following conditions hold:

1. for each $x \in \vec{x}$, we have $\tau'(x) \in N_I$, and
2. a nondeterministic computation exists such that function $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ returns $\mathtt{t}$.

By the definition of candidate answer, we have $\mathsf{dom}(\tau') = N_V(q')$, each element of $\mathsf{rng}(\tau')$ is an individual occurring in $\mathsf{D}_\mathcal{K}$, and $\mathsf{D}_\mathcal{K} \models \tau'(q')$. Since $\tau'|_{\vec{x}} \subseteq N_I$, we have that $\pi'(x) \in N_I$ for each $x \in \vec{x}$. In the rest of this proof we show that $\Xi_\mathcal{K} \models \pi'(q')$.

Let CQ q and substitution τ be as specified in Definition 4; and let relation $\sim$, CQ $q_\sim$ and the connection graph $\mathsf{cg} = \langle V, E_s, E_t \rangle$ be as determined by isSound. By the construction of q and τ, we have $\mathsf{D}_\mathcal{K} \models \tau(q)$ and $\tau(q) \subseteq J$. We next construct a substitution π with $\mathsf{dom}(\pi) = N_V(q)$ such that $\pi(q) \subseteq I$ and the following property holds.

(1) For each term $t \in N_T(q)$, we have $\delta(\pi(t)) = \tau(t)$.

Later, we will show that property (1) and $\pi(q) \subseteq I$ imply that $\Xi_\mathcal{K} \models \pi'(q')$, thus proving the soundness claim.

To construct substitution π, we proceed in two steps: we first show how to construct π in case our algorithm returns $\mathtt{t}$ in step 2; after which we show how to construct π in case our algorithm returns $\mathtt{t}$ in step 18.

Case 1: $\mathsf{isSound}$ **returns** t **in step** 2 Assume that $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ returns t in step 2. By condition 2 in Definition 8, directed graph $\langle V, E_s\rangle$ is acyclic; we next show that $\langle V, E_s\rangle$ is a forest, after which we will show how to construct substitution π by structural induction on this forest.

Lemma 25. *Directed graph* $\langle V, E_s\rangle$ *is a forest.*

Proof. Since directed graph $\langle V, E_s\rangle$ is acyclic, we are left to show that for each $v \in V$, there exists at most one vertex v' such that $\langle v', v\rangle \in E_s$. Assume the opposite; hence, vertices v_1, v_2, and v exist in V such that $v_1 \neq v_2$ and $\{\langle v_1, v\rangle, \langle v_2, v\rangle\} \subseteq E_s$. Then, roles R and P exist such that $R(v_1, v)$ and $P(v_2, v)$ are aux-simple atoms in $q_\sim$ and $\tau(v) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. By the definition of $\sim$, we have $v_1 \sim v_2$; and, by the construction of $q_\sim$, we have $v_1 = v_2$, which is a contradiction. □

We next construct substitution π with $\mathsf{dom}(\pi) = N_V(q)$ that will satisfy (1) and the two following properties:

(2) for all terms $s, t \in N_T(q)$ such that $s \sim t$, we have $\pi(s) = \pi(t)$, and

(3) for each $\langle v', v\rangle \in E_s$ with $\tau(v)$ of the form $o_{P,A}$, we have $P(\pi(v'), \pi(v)) \in I$.

We define π by structural induction on the forest $\langle V, E_s\rangle$; later we show that $\pi(q) \subseteq I$.

Base case. Consider a root $v \in V$. Fix an arbitrary term $w \in \Delta^I$ such that $\delta(w) = \tau(v)$. For each term $s \in N_T(q)$ with $s \sim v$, let $\pi(s) = w$. By condition 1 in Definition 8, we have $\tau(s) = \tau(v)$. Thus property (1) and (2) hold.

Inductive step. Consider an arbitrary $\langle v', v\rangle \in E_s$ with $\pi(v')$ defined and $\pi(v)$ undefined. By the definition of E_s, a role $R \in N_R$ exists such that $R(v', v)$ is an aux-simple atom in $q_\sim$. Hence, we have $\mathsf{d}_R(\tau(v'), \tau(v)) \in J$ and $\tau(v)$ is of the form $o_{P,A}$. Since by property (1) we have $\delta(\pi(v')) = \tau(v')$, and due to property D5 in Lemma 21, a term $w \in \Delta^I$ exists such that $P(\pi(v'), w) \in I$. Then, for each term $s \in N_T(q)$ with $s \sim v$, let $\pi(s) = w$. Properties (2) and (3) immediately hold. By condition 2 in Definition 8, we have $\tau(s) = \tau(v)$, and so property (1) holds.

Lemma 26. *Substitution* π *satisfies* $\pi(q) \subseteq I$.

Proof. Recall that $\tau(q) \subseteq J$ and $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ returns t in step 2; we next show that $\pi(q) \subseteq I$.

Consider an atom $A(s)$ in q. By assumption, we have $A(\tau(s)) \in J$. By Lemma 21, for each $w \in \Delta^I$ with $\delta(w) = \tau(s)$, we have $A(w) \in I$. By property (1) in the definition of π, we have $A(\pi(s)) \in I$.

Consider an atom $R(s', t')$ in q. By the definition of $q_\sim$, an atom $R(s, t)$ occurs in $q_\sim$ such that $s' \sim s$ and $t' \sim t$. By condition 1 in the definition of $\mathsf{isDSound}$, we have $\tau(s') = \tau(s)$ and $\tau(t') = \tau(t)$. By assumption, we have $R(\tau(s'), \tau(t')) \in J$ and so $R(\tau(s), \tau(t)) \in J$. By property (2) in the definition of π, it suffices to show that $R(\pi(s), \pi(t)) \in I$. Given that our algorithm returns t in step 2, exactly one of the following holds.

$R(s, t)$ is such that $\tau(t) \in N_I$. By Lemma 21, for all terms $w', w \in \Delta^I$ with $\delta(w') = \tau(s)$ and $\delta(w) = \tau(t)$, we have $R(w', w) \in I$. By property (1) in the definition of π, we have $R(\pi(s), \pi(t)) \in I$.

$R(s, t)$ is such that $s = t$, $\tau(t) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, and $\mathsf{Self}_R(\tau(t)) \in J$. By Lemma 21, for each term $w \in \Delta^I$ with $\delta(w) = \tau(t)$, we have $R(w, w) \in I$. By property (1) in the definition of π, we have $R(\pi(t), \pi(t)) \in I$.

$R(s, t)$ is aux-simple. It follows that $\tau(t)$ is of the form $o_{P,A}$ and $\mathsf{d}_R(\tau(s), \tau(t)) \in J$. By property D5 of Lemma 21, we have $P \sqsubseteq^*_\mathcal{T} R$. By the definition of E_s, we have $\langle s, t\rangle \in E_s$. By property (3) in the definition of π, we have $P(\pi(s), \pi(t)) \in I$. Since no rule is applicable to I and $P \sqsubseteq^*_\mathcal{T} R$, we have $R(\pi(s), \pi(t)) \in I$. □

Case 2: $\mathsf{isSound}$ **returns** t **in step** 18 We are left to show that, if our function returns t in step 18, then a substitution π with $\mathsf{dom}(\pi) = N_V(q)$ exists such that $\pi(q) \subseteq I$ and property (1) is satisfied.

Assume that $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ returns t in step 18. Let variable renaming σ, skeleton $\mathcal{S} = \langle \mathcal{V}, \mathcal{E}\rangle$, and function L be as determined by $\mathsf{isSound}$. By the definition of a skeleton for q and σ, graph $\mathcal{S}$ is a forest rooted in $\mathcal{V} \cap \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$. We next define substitution π that will satisfy property (1) as well as the following properties:

(2) for all terms $s, t \in N_T(q)$ such that $s \sim t$ or $\sigma(s) = t$, we have $\pi(s) = \pi(t)$,

(3) for each $\langle v', v\rangle \in \mathcal{E}$ and each role $P \in L(v', v)$, we have $P(\pi(v'), \pi(v)) \in I$.

We define π by structural induction on the forest $\mathcal{S} = \langle \mathcal{V}, \mathcal{E}\rangle$; later we show that $\pi(q) \subseteq I$.

Base case. Consider a root $v \in \mathcal{V} \cap \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$. Given that each element in $\mathsf{rng}(\sigma)$ is a variable, no term $s \in N_T(q)$ exists such that $\sigma(s) = v$. Then, for each term $s \in N_T(q)$ with $s \sim v$, let $\pi(s) = v$. By condition 1 in Definition 8, we have $\tau(s) = \tau(v)$. Thus, properties (1) and (2) are satisfied.

Inductive step. Consider $\langle v', v\rangle \in \mathcal{E}$ such that $\pi(v')$ has been defined, but $\pi(v)$ has not; and let $u_0 = \tau(v')$ and $w_0 = \pi(v')$. By the definition of exist, individuals $\{u_1, \dots, u_n\} \subseteq \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ with $n > 0$ and $u_n = \tau(v)$ exist such that for each $i \in [1, n]$, we have u_i is of the form o_{T_i,A_i}, and $\mathsf{d}_{P_j}(u_{i-1}, u_i) \in J$ for each $P_j \in L(v', v)$. Then, for each $i \in [1, n]$, by property D5 of Lemma 21, a term w_i of the form $f^{B_i}_{T_i,A_i}(w_{i-1})$ exists in Δ^I such that $T_i(w_{i-1}, w_i) \in I$; moreover, $T_i \sqsubseteq^*_\mathcal{T} P_j$ for each role $P_j \in L(v', v)$. Since no rule is applicable to I and $T_i \sqsubseteq^*_\mathcal{T} P_j$, we have $P_j(w_{i-1}, w_i) \in I$. For each term $s \in N_T(q)$ such that $s \sim v$ or $\sigma(s) = v$, let $\pi(s) = w_n$. Property (2) is clearly satisfied. Property (1) is also satisfied, since $\sigma(s) = v$ implies that

$\tau(s) = \tau(v)$, by construction of σ, and $s \sim v$ implies that $\tau(s) = \tau(v)$, by condition 1 in Definition 8. For property (3) we distinguish two cases.

- A role $P \in L(v', v)$ exists such that $\mathsf{trans}(P) \notin \mathcal{T}$. By the definition of function exist, we then have $n = 1$. Consequently, $\pi(v') = w_0$ and $\pi(v) = w_1$; thus, $P_j(\pi(v'), \pi(v)) \in I$ for each $P_j \in L(v', v)$.
- For each $P_j \in L(v', v)$ we have $\mathsf{trans}(P_j) \in \mathcal{T}$. Since no rule is applicable and $P_j(w_{i-1}, w_i) \in I$ for each $i \in [1, n]$, we have $P_j(w_0, w_n) \in I$; that is, $P_j(\pi(v'), \pi(v)) \in I$.

Lemma 27. *Substitution π satisfies $\pi(q) \subseteq I$.*

Proof. We show that $\pi(q) \subseteq I$ by considering the various atoms occurring in q.

Consider an atom $A(s)$ in q. By assumption, we have $A(\tau(s)) \in J$. By Lemma 21, for each $w \in \Delta^I$ with $\delta(w) = \tau(s)$, we have $A(w) \in I$. By property (1) in the definition of π, we have $A(\pi(s)) \in I$.

Consider an atom $R(s', t')$ in q. By assumption, we have $R(\tau(s'), \tau(t')) \in J$. By the definition of $q_\sim$, terms s'' and t'' occur in $q_\sim$ such that $s' \sim s''$, $t' \sim t''$, and $R(s'', t'')$ is an atom in $q_\sim$. By condition 1 in the definition of isDSound, we have $\tau(s') = \tau(s'')$ and $\tau(t') = \tau(t'')$. Therefore, $R(\tau(s''), \tau(t'')) \in J$. By the definition of $\sigma(q_\sim)$, terms s and t occur in $\sigma(q_\sim)$ such that $\sigma(s'') = s$, $\sigma(t'') = t$, and $R(s, t)$ is an atom in $\sigma(q_\sim)$. By the definition of variable renaming, we have $\tau(t'') = \tau(t)$ and $\tau(s'') = \tau(s)$. Thus $R(\tau(s), \tau(t)) \in J$. By property (2) in the definition of π, it suffices to show that $R(\pi(s), \pi(t)) \in I$. Towards this goal, we consider four distinct cases.

$R(s, t)$ is such that $\tau(t) \in N_I$. By Lemmas 21, for all terms $w', w \in \Delta^I$ with $\delta(w') = \tau(s)$ and $\delta(w) = \tau(t)$, we have $R(w', w) \in I$. By property (1) in the definition of π, we have $R(\pi(s), \pi(t)) \in I$.

$R(s, t)$ is such that $s = t$, $\tau(t) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, and $\mathsf{Self}_R(\tau(t)) \in J$. By Lemma 21, for each term $w \in \Delta^I$ with $\delta(w) = \tau(t)$, we have $\|R(w, w)\|_I \in I$. By property (1) in the definition of π, we have $R(\pi(t), \pi(t)) \in I$.

$R(s, t)$ is aux-simple. By the definition of E_s and $\mathcal{E}$, we have $\langle s, t\rangle \in \sigma(E_s) \cap \mathcal{E}$, and $R \in L(s, t)$ by step 6. By property (3) in the definition of π, we have $R(\pi(s), \pi(t)) \in I$.

$R(s, t)$ is neither good nor aux-simple. Let P and $v_0, \ldots, v_n$ be as determined in steps 8–15 when Algorithm 1 considers atom $R(s, t)$. Then by step 8 we have $P \sqsubseteq^*_\mathcal{T} R$. Furthermore, for each $i \in [1, n]$, we have $P \in L(v_{i-1}, v_i)$ by step 15; but then, by property (3) we have $P(\pi(v_{i-1}), \pi(v_i)) \in I$. Next, we distinguish two cases.

- s reaches t in $\mathcal{E}$. If $\langle s, t\rangle \in \mathcal{E}$, then $n = 1$ and, by property (3) in the definition of π, we have $P(\pi(s), \pi(t)) \in I$. Otherwise, we have $\mathsf{trans}(P) \in \mathcal{T}$ and, since no rule is applicable to I, we have $P(\pi(s), \pi(t)) \in I$.
- s does not reach t in $\mathcal{E}$. Then, we have $\mathsf{trans}(P) \in \mathcal{T}$, $v_0 = a_t$, and $P(\tau(s), v_0) \in J$. As $v_0 \in N_I$ and $\pi(v_0) = v_0$, by Lemma 21, we have $P(\pi(s), \pi(v_0)) \in I$. Due to $\mathsf{trans}(P) \in \mathcal{T}$ and no rule is applicable to I, we have $\|P(\pi(s), \pi(t))\|_I \in I$.

Since $P \sqsubseteq^*_\mathcal{T} R$ and no rule is applicable to I we have $R(\pi(s), \pi(t)) \in I$. □

Finally, we prove the soundness claim.

Lemma 28. *Substitution π' satisfies $\Xi_\mathcal{K} \models \pi'(q')$.*

Proof. To prove the lemma, we show that a substitution π'_* with $\mathsf{dom}(\pi'_*) = N_V(q')$ exists such that $\pi' \subseteq \pi'_*$ and $\|\pi'_*(q')\|_I \subseteq I$. By the construction of q and τ, we have $\mathsf{D}_\mathcal{K} \models \tau(q)$ and $\tau(q) \subseteq J$. Let π be the substitution specified just above Lemma 26, if $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ returns t in step 2, otherwise, let π be the substitution specified just above Lemma 27. By Lemmas 26 and 27, we have $\pi(q) \subseteq I$; furthermore, π satisfies property (1). Let γ be the mapping from $N_T(q')$ to $N_T(q)$ such that q is obtained by replacing each $t \in N_T(q')$ with $\gamma(t)$.

Let π'_* be the substitution such that for each $z \in N_V(q')$, we have $\pi'_*(z) = \pi(z)$, if $\gamma(z) = z$; otherwise, we define $\pi'_*(z)$ as an arbitrary term w occurring in I such that $\delta(w) = \tau'(z)$. Since π satisfies property (1) and by construction π'_*, for each term $t \in N_T(q')$, we have $\delta(\pi'_*(t)) = \tau'(t)$. Since $\tau'(x) \in N_I$ for each $x \in \vec{x}$ and given that δ is the identity on N_I, we have $\pi' \subseteq \pi'_*$. To prove that $\|\pi'_*(q')\|_I \subseteq I$, we show that for each term $t \in N_T(q')$, we have $\|\pi'_*(t)\|_I = \pi(\gamma(t))$. The property clearly holds for each term $t \in N_T(q')$ such that $\gamma(t) = t$. Then consider an arbitrary term $t \in N_T(q')$ such that $\gamma(t) \neq t$. By the definition of γ, we have $\gamma(t) = \|\tau'(t)\|_J$ and $\gamma(t) \in N_I$. By Lemma 24, for each term w occurring in I with $\delta(w) = \tau'(t)$, we have $\|w\|_I = \gamma(t)$. By the definition of π'_*, we then have $\|\pi'_*(t)\|_I = \gamma(t) = \pi(\gamma(t))$. □

D.2 Completeness

To prove the completeness claim, we start by establishing two properties of the universal interpretation I. To this end, we start with a couple of definitions.

Let $\prec$ be the smallest irreflexive and transitive relation on the set of terms occurring in I such that $w \prec f^{A_1}_{R,A}(w)$ for each term occurring in I, each role R, and all concepts $A, A_1 \in \{\top\} \cup N_C$. Furthermore, an atom $R(w', w) \in I$ is a *self-loop* if $w' = w$ and $\mathsf{Self}_R(w) \in I$.

Lemma 29. *Interpretation I satisfies the following properties for each role $R \in N_R$ and all terms $w', w \in \Delta^I$ with $w \notin N_I$.*

1. $R(w', w) \in I$ *is not a self-loop and* R *is simple imply that* w *is of the form* $f^{A_1}_{T,A}(w')$.

2. $R(w', w) \in I$ *is not a self-loop implies that a role* $P \in N_R$ *and terms* $w_0, \ldots, w_m$ *from* Δ^I *with* $w_m = w$ *exist where*

2a. if $m > 1$ *or* $w' \not\prec w$*, then* $\mathsf{trans}(P) \in \mathcal{T}$*,*

2b. $P \sqsubseteq^*_{\mathcal{T}} R$*,*

2c. $w_0 = w'$*, if* $w' \prec w$*; otherwise,* w_0 *is the unique individual* $a \in N_I$ *with* $a \prec w$*, and* $P(w', w_0) \in I$*, and*

2d. for each $i \in [1, m]$*,* w_i *is of the form* $f^{B_i}_{S_i,A_i}(w_{i-1})$ *and* $P(w_{i-1}, w_i) \in I$*.*

Proof. Let $I_0, I_1, \ldots$ be the chase sequence used to construct I. Please observe that, by virtue of the pruning step in the application of equality rules, for each term w occurring in I of the form $w = f^{A_1}_{T,A}(w')$ we either have $\|w\|_I = a$ for some individual $a \in N_I$, or $\|w'\|_I = w'$ and $\|w\|_I = w$. Then to prove the lemma, we show by induction on this sequence that each I_n satisfies the following properties for each role $R \in N_R$ and all terms w' and w occurring in I such that $w \notin N_I$.

A. $R(w', w) \in I_n$ is not a self-loop and R is simple imply that w is of the form $f^{A_1}_{T,A}(w')$.

B. $R(w', w) \in I_n$ is not a self-loop implies that a role $P \in N_R$ and terms $w_0, \ldots, w_m$ from I with $w_m = w$ exist where
   (i) if $m > 1$ or $w' \not\prec w$, then $\mathsf{trans}(P) \in \mathcal{T}$,
   (ii) $P \sqsubseteq^*_{\mathcal{T}} R$,
   (iii) $w_0 = w'$, if $w' \prec w$; otherwise, w_0 is the unique individual $a \in N_I$ with $a \prec w$, and $\|P(w', w_0)\|_I \in I$, and
   (iv) for each $i \in [1, m]$, w_i is of the form $f^{B_1}_{S_i,A_i}(w_{i-1})$ and $\|P(w_{i-1}, w_i)\|_I \in I$.

Base case. Consider I_0. All atoms in I_0 are over the individuals in N_I, so properties A and B hold vacuously.

Inductive step. Consider an arbitrary $n \in \mathbb{N}$ and assume that I_n satisfies properties A and B. By considering each rule in $\Xi_{\mathcal{K}}$ that derives binary atoms, we assume that the rule is applicable to I_n, and we show that the properties hold for all fresh atoms in the resulting instance.

Consider a rule $A_1(x) \to \exists z.R(x, z) \wedge A(z)$, and assume that $A_1(w') \in I_n$. Furthermore, let $w = f^{A_1}_{R,A}(w')$, and assume that w' and w occur in I. Clearly, if $\|w\|_{I_n} = a$ for some individual $a \in N_I$, then the properties hold vacuously; hence we consider the case in which $\|w\|_{I_n} = w$. Please note that $w' \prec w$. Since w' occurs in I, we have $\|A_1(w')\|_I \in I$. As no rule is applicable to I, we have $\|R(w', w)\|_I \in I$. Property A holds, and property B is satisfied for role R and terms $w_0 = w'$ and $w_1 = w$.

Consider a rule $A(x) \to x \approx a$, and assume that $A(u) \in I_n$. Let u' and w' be terms occurring in I such that $\{u', w'\} = \{\|u\|_{I_n}, \|a\|_{I_n}\}$ and $u' > w'$. By the definition of $>$, we have $w' \in N_I$. Since u occurs in I, we have $\|A(u)\|_I \in I$. Since no rule is applicable to I, we have $\|u'\|_I = \|w'\|_I$. We next show that the properties are preserved for all atoms in I_n that get replaced by the application of this rule. Please observe that the properties hold vacuously for each atom $R(w, u') \in I_n$ and each atom $R(u', b) \in I_n$ with $b \in N_I$. Then consider an atom $R(u', w) \in I_n$ with $w \neq u'$ and $w \notin N_I$. Then $R(u', w) \in I_n$ is not a self-loop. Since this atom is replaced, not removed, by the application of the rule, we must have that $u' \not\prec w$. Let $c \in N_I$ be the unique individual such that $c \prec w$. By the inductive hypothesis, a role $P \in N_R$ and terms $w_0, \ldots, w_m$ with $w_0 = c$ and $w_m = w$ exist satisfying properties *(i)–(iv)*. Then $\|P(u', w_0)\|_I \in I$, $\mathsf{trans}(P) \in \mathcal{T}$, and $P \sqsubseteq^*_{\mathcal{T}} R$. Role R is not simple, thus property A holds. We next distinguish two cases.

- $w' \prec w$. Since $w' \in N_I$, we have $w' = w_0$. Then role P and terms $w_0, \ldots, w_m$ satisfy properties *(i)–(iv)*.
- $w' \not\prec w$. As stated above, we have $\mathsf{trans}(P) \in \mathcal{T}$ and $\|P(u', w_0)\|_I \in I$. Thus, $\|P(w', w_0)\|_I \in I$, and role P and terms $w_0, \ldots, w_m$ satisfy properties *(i)–(iv)*.

Consider a rule $T(x, y) \to R(x, y)$, and assume that $T(w', w) \in I_n$ is not a self-loop. For property A, assume that R is a simple role. Hence, T is a simple role as well, and property A follows from the inductive hypothesis. For property B, by the inductive hypothesis, a role P and terms $w_0, \ldots, w_m$ with $w_m = w$ exist satisfying *(i)–(iv)*. By the definition of $\Xi_{\mathcal{K}}$, we have $T \sqsubseteq R \in \mathcal{T}$. Since $P \sqsubseteq^*_{\mathcal{T}} T$, we then have $P \sqsubseteq^*_{\mathcal{T}} R$, and properties *(i)–(iv)* are satisfied.

Consider a rule $R(x, y) \wedge R(y, z) \to R(x, z)$, and assume that $\{R(w', w''), R(w'', w)\} \subseteq I_n$. Property A holds vacuously, as R is not simple. For property B, we distinguish four cases.

- $w' \prec w''$ and $w'' \prec w$. By the inductive hypothesis, a role P_1 and terms $w^1_0, \ldots, w^1_{m_1}$ with $w^1_0 = w'$ and $w^1_{m_1} = w''$ exist satisfying properties *(i)–(iv)*; moreover, a role P_2 and terms $w^2_0, \ldots, w^2_{m_2}$ with $w^2_0 = w''$ and $w^2_{m_2} = w$ exist satisfying properties *(i)–(iv)*. Then, for $j = 1, 2$, we have $P_j \sqsubseteq^*_{\mathcal{T}} R$. Furthermore, for each $i \in [1, m_j]$, we have $\|P_j(w^j_{i-1}, w^j_i)\|_I \in I$. As no rule is applicable to I, we have $\|R(w^j_{i-1}, w^j_i)\|_I \in I$. Then, property B is satisfied for role R and terms $w^1_0, \ldots, w^1_{m_1}, w^2_1, \ldots w^2_{m_2}$.

- $w' \prec w''$ and $w'' \not\prec w$. By the inductive hypothesis, a role P_1 and terms $w_0^1, \ldots, w_{m_1}^1$ with $w_0^1 = w'$ and $w_{m_1}^1 = w''$ exist satisfying properties *(i)–(iv)*. Let $b \in N_I$ be the unique individual such that $b \prec w$. Since $w \notin N_I$, by the inductive hypothesis, a role P_2 and terms $w_0^2, \ldots, w_{m_2}^2$ with $w_0^2 = b$ and $w_{m_2}^2 = w$ exist satisfying properties *(i)–(iv)*. By property *(iii)*, we have $\|P_2(w_{m_1}^1, w_0^2)\|_I \in I$. Then, for $j = 1, 2$, we have $P_j \sqsubseteq_{\mathcal{T}}^* R$. Moreover, for each $i \in [1, m_j]$, we have $\|P_j(w_{i-1}^j, w_i^j)\|_I \in I$. As no rule is applicable to I, we have $\|R(w_{i-1}^j, w_i^j)\|_I \in I$ and $\|R(w_{m_1}^1, w_0^2)\|_I \in I$. Since $\mathsf{trans}(R) \in \mathcal{T}$ and no rule is applicable to I, we have $\|R(w_0^1, w_0^2)\|_I \in I$. Property B is satisfied for role R and terms $w_0^2, \ldots, w_{m_2}^2$.
- $w' \not\prec w''$ and $w'' \prec w$. By the inductive hypothesis, a role P_2 and terms $w_0^2, \ldots, w_{m_2}^2$ with $w_0^2 = w''$ and $w_{m_2}^2 = w$ exist satisfying properties *(i)–(iv)*. Then, we have $P_2 \sqsubseteq_{\mathcal{T}}^* R$. Furthermore, for each $i \in [1, m_2]$, we have $\|P_2(w_{i-1}^2, w_i^2)\|_I \in I$. As no rule is applicable to I, we have $\|R(w_{i-1}^2, w_i^2)\|_I \in I$. We distinguish two cases.
  - $w'' \in N_I$. It follows that $w_0^2 = w''$. Then property B is satisfied for role R and terms $w_0^2, \ldots, w_{m_2}^2$.
  - $w'' \notin N_I$. Let $a \in N_I$ be the unique individual such that $a \prec w''$. By the inductive hypothesis, a role P_1 and terms $w_0^1, \ldots, w_{m_1}^1$ with $w_0^1 = a$ and $w_{m_1}^1 = w''$ exist satisfying properties *(i)–(iv)*. Then, we have $P_1 \sqsubseteq_{\mathcal{T}}^* R$ and $\|P_1(w', a)\|_I \in I$. Furthermore, for each $i \in [1, m_1]$, we have $\|P_1(w_{i-1}^1, w_i^1)\|_I \in I$. As no rule is applicable to I, we have $\|R(w', a)\|_I \in I$ and $\|R(w_{i-1}^1, w_i^1)\|_I \in I$. Then the property is satisfied for role R and terms $w_0^2, \ldots, w_{m_2}^2$.
- $w' \not\prec w''$ and $w'' \not\prec w$. By assumption, we have $w \notin N_I$. Let $b \in N_I$ be the unique individual such that $b \prec w$. By the inductive hypothesis, a role P_2 and terms $w_0^2, \ldots, w_{m_2}^2$ with $w_0^2 = b$ and $w_{m_2}^2 = w$ exist satisfying properties *(i)–(iv)*. Then, we have $P_2 \sqsubseteq_{\mathcal{T}}^* R$ and $\|P_2(w'', w_0^2)\|_I \in I$. Furthermore, for each $i \in [1, m_2]$, we have $\|P_2(w_{i-1}^2, w_i^2)\|_I \in I$. As no rule is applicable to I, we have $\|R(w'', w_0^2)\|_I \in I$ and $\|R(w_{i-1}^2, w_i^2)\|_I \in I$. We distinguish two cases.
  - $w'' \in N_I$. As stated above, we have $\|R(w', w'')\|_I \in I$ and $\|R(w'', w_0^2)\|_I \in I$. Since $\mathsf{trans}(R) \in I$ and no rule is applicable to I, we have $\|R(w', w_0^2)\|_I \in I$. Then property B is satisfied for role R and terms $w_0^2, \ldots, w_{m_2}^2$.
  - $w'' \notin N_I$. Let $a \in N_I$ be the unique individual such that $a \prec w''$. By the inductive hypothesis, a role P_1 and terms $w_0^1, \ldots, w_{m_1}^1$ with $w_0^1 = a$ and $w_{m_1}^1 = w''$ exist satisfying properties *(i)–(iv)*. Then, we have $P_1 \sqsubseteq_{\mathcal{T}}^* R$ and $\|P_1(w', w_0^1)\|_I \in I$. Furthermore, for each $i \in [1, m_1]$, we have $\|P_1(w_{i-1}^1, w_i^1)\|_I \in I$. As no rule is applicable to I, we have $\|R(w', w_0^1)\|_I \in I$ and $\|R(w_{i-1}^1, w_i^1)\|_I \in I$. Since $\mathsf{trans}(R) \in \mathcal{T}$ and no rule is applicable to I, we have $\|R(w', w_0^2)\|_I \in I$. Property B is satisfied for role R and terms $w_0^2, \ldots, w_{m_2}^2$. □

Let $q' = \exists \vec{y}.\psi(\vec{x}, \vec{y})$ be a CQ and let π' be a substitution such that $\mathsf{dom}(\pi') = \vec{x}$ and each element in $\mathsf{rng}(\pi')$ is an individual from N_I. Assume that $\Xi_{\mathcal{K}} \models \pi'(q')$; we next show that a substitution τ' with $\mathsf{dom}(\tau') = N_V(q')$ exists such that $\tau'|_{\vec{x}} = \pi'$, each element in $\mathsf{rng}(\tau')$ is an individual occurring in $\mathsf{D}_{\mathcal{K}}$, and all of the following conditions hold:

1. for each $x \in \vec{x}$, we have $\tau'(x) \in N_I$,
2. $\mathsf{D}_{\mathcal{K}} \models \tau'(q')$, and
3. a nondeterministic computation exists such that function $\mathsf{isSound}(q, \mathsf{D}_{\mathcal{K}}, \tau)$ returns t.

Since $\Xi_{\mathcal{K}} \models \pi'(q)$, a substitution π'_* with $\mathsf{dom}(\pi'_*) = N_V(q')$ exists such that $\pi' \subseteq \pi'_*$ and $\|\pi'_*(q')\|_I \subseteq I$. Let substitution τ' be such that, for each variable $z \in N_V(q')$, we have $\tau'(z) = \delta(\pi'_*(z))$. Since δ is the identity on N_I and $\pi'_*(x) \in N_I$ for each $x \in \vec{x}$, we have $\pi' \subseteq \tau'$ and $\tau'(x) \in N_I$. By Lemmas 20 and 24, we have $\|\tau'(q')\|_J \subseteq J$, and so $\mathsf{D}_{\mathcal{K}} \models \tau'(q')$. Hence, in the following, we show that property (3) holds.

Let q and τ be as specified in Definition 4. Then, we have $\mathsf{D}_{\mathcal{K}} \models \tau(q)$ and $\tau(q) \subseteq J$. Let π_* be the restriction of π'_* to only those variables that occur in q. By the definition of τ and π_*, for each variable $z \in N_V(q)$, we have $\delta(\pi_*(z)) = \tau(z)$. We next show that $\pi_*(q) \subseteq I$.

Lemma 30. *Substitution* π_* *satisfies* $\pi_*(q) \subseteq I$

Proof. Let π'_* and π_* be as specified above; furthermore, let γ be the mapping from $N_T(q')$ to $N_T(q)$ such that q is obtained by replacing each term $t \in N_T(q')$ with $\gamma(t)$. To prove that $\pi_*(q) \subseteq I$, we show that for each term $t \in N_T(q')$, we have $\pi_*(\gamma(t)) = \|\pi'_*(t)\|_I$. We distinguish two cases.

- Consider an arbitrary variable $z \in N_V(q')$ such that $\tau'(z) \in \mathsf{aux}_{\mathsf{D}_{\mathcal{K}}}$. Then $\gamma(z) = z$. By the definition of $\mathsf{aux}_{\mathsf{D}_{\mathcal{K}}}$, for each individual u such that $\mathsf{D}_{\mathcal{K}} \models \tau'(z) \approx u$, we have $u \notin N_I$ and $\tau'(z) = u$; therefore, $\|\tau'(z)\|_J = \tau'(z)$. But then, by the construction of τ' and by Lemma 24, we have $\|\pi'_*(z)\|_I = \pi_*(z)$.
- Consider an arbitrary term $t \in N_T(q')$ such that $\tau'(t) \notin \mathsf{aux}_{\mathsf{D}_{\mathcal{K}}}$. By the definition of q, we have $\gamma(t) = \|\tau'(t)\|_J$ and $\gamma(t) \in N_I$. By Lemma 24, for each term w occurring in I with $\delta(w) = \tau'(t)$, we have $\|w\|_I = \gamma(t)$. By the definition of τ', we then have $\pi(\gamma(t)) = \gamma(t) = \|\pi'_*(t)\|_I$. □

Let relation $\sim$ and query $q_\sim$ be as specified in Definition 6; moreover, let connection graph $\mathsf{cg} = \langle V, E_s, E_t \rangle$ be as specified in Definition 7. We next show that function $\mathsf{isDSound}(q, \mathsf{D}_{\mathcal{K}}, \tau)$ returns t.

Lemma 31. *Function* $\mathsf{isDSound}(q, \mathsf{D}_{\mathcal{K}}, \tau)$ *returns* t.

Proof. We next prove that $\mathsf{isDSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ return t by showing that the two conditions of Definition 8 are satisfied.

(Condition 1) We prove that, for each $s \sim t$, we have $\tau(s) = \tau(t)$ and $\pi_*(s) = \pi_*(t)$. We proceed by induction on the number of steps required to derive $s \sim t$. For the base case, the empty relation $\sim$ clearly satisfies the two properties. For the inductive step, consider an arbitrary relation $\sim$ obtained in n steps that satisfies these constraints; we show that the same holds for all constraints derivable from $\sim$. We focus on the (fork) rule, as the derivation of $s \sim t$ due to reflexivity, symmetry, or transitivity clearly preserves the required properties. Let s_1', s_2, s_2', and s_2 be arbitrary terms in $N_T(q)$, and let R_1 and R_2 be arbitrary roles such that $s_1' \sim s_2'$ is obtained in n steps, atoms $R_1(s_1, s_1')$ and $R_2(s_2, s_2')$ occur in q and are aux-simple, and $\tau(s_2') \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. By the definition of τ, we have $\pi(s_2') \notin N_I$. Moreover, by the definition of aux-simple atom, for $i = 1, 2$ we have $s_i \neq s_i'$, role R_i is simple, and $\tau(s_i) = \tau(s_i')$ implies that $\mathsf{Self}_{R_i}(\tau(s_i')) \notin J$. By the inductive hypothesis, we have $\tau(s_1') = \tau(s_2')$ and $\pi_*(s_1') = \pi_*(s_2')$. By Lemma 20, atoms $\{R_1(\pi_*(s_1), \pi_*(s_1')), R_2(\pi_*(s_2), \pi_*(s_2'))\} \subseteq I$ are not self-loops. Moreover, since R_1 and R_2 are simple roles, by property 1 in Lemma 29, we have $\pi_*(s_1) = \pi_*(s_2)$. Therefore, $\pi_*'(s_1) = \pi_*'(s_2)$. By the construction of τ', we have $\tau'(s_1) = \tau'(s_2)$, and so $\tau(s_1) = \tau(s_2)$.

(Condition 2) We show that $\langle V, E_s\rangle$ is a DAG. Assume the opposite; hence, vertices $v_0, \ldots, v_m \in V$ with $v_m = v_0$ exist such that $m > 0$ and $\langle v_{i-1}, v_i\rangle \in E_s$ for each $i \in [1, m]$. By the definition of E_s, we have $E_s \subseteq V \times (V \cap N_V(q))$. Thus, for each $i \in [0, m]$ we have $v_i \notin \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$, and so $\tau(v_i) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. Consider an arbitrary $i \in [1, m]$ and the corresponding edge $\langle v_{i-1}, v_i\rangle \in E_s$. By the definition of E_s, a role R_i exists such that $R_i(v_{i-1}, v_i)$ is an aux-simple atom in $q_\sim$. By the definition of aux-simple atom, we have $v_{i-1} \neq v_i$, role R_i is simple, and $\tau(v_{i-1}) = \tau(v_i)$ implies that $\mathsf{Self}_{R_i}(\tau(v_i)) \notin J$. By Lemma 20 and by the construction of τ, atom $R_i(\pi_*(v_{i-1}), \pi_*(v_i)) \in I$ is not a self-loop. Since $\tau(v_i) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, we have $\pi_*(v_i) \notin N_I$. Then, by property 1 in Lemma 29, we have $\pi_*(v_{i-1}) \prec \pi_*(v_i)$. Thus, $\pi_*(v_0) \prec \pi_*(v_m)$ which contradicts $v_m = v_0$. □

We are now ready to prove the completeness claim.

Lemma 32. *A nondeterministic computation exists such that* $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ *returns* t.

Proof. Recall that $\pi_*(q) \subseteq I$ and that $\tau(q) \subseteq J$. By Lemma 31, the condition in step 1 in Algorithm 1 is not satisfied. If each binary atom $R(s, t)$ occurring in $q_\sim$ is either good or aux-simple, then our algorithm returns t in step 2; hence, in the rest of this proof, we assume that this is not the case.

For the variable renaming σ in step 3, for each variable $z \in V$, let $\sigma(z)$ be an arbitrary, but fixed, variable $z' \in V$ such that $\pi_*(z) = \pi_*(z')$. It is straightforward to see that $\pi_*(\sigma(q_\sim)) \subseteq I$.

For the skeleton $\mathcal{S} = \langle \mathcal{V}, \mathcal{E}\rangle$ in step 3, let $\mathcal{V} = \sigma(V)$ and let $\mathcal{E}$ be the smallest set containing $\langle v', v\rangle \in \mathcal{E}$ for all $v', v \in \mathcal{V}$ such that $\pi_*(v') \prec \pi_*(v)$ and no vertex $v'' \in \mathcal{V}$ exists such that $\pi_*(v') \prec \pi_*(v'') \prec \pi_*(v)$. By the definition of $\prec$, graph $\langle \mathcal{V}, \mathcal{E}\rangle$ is a forest rooted in $\mathcal{V} \cap \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$.

We next show that for each $\langle v', v\rangle \in \sigma(E_s)$, we have $\langle v', v\rangle \in \mathcal{E}$. Consider an arbitrary edge $\langle v', v\rangle \in \sigma(E_s)$. Then vertices u' and u exist in V such that $\langle u', u\rangle \in E_s$, $\pi_*(u') = \pi_*(v')$, and $\pi_*(u) = \pi_*(v)$. By the definition of E_s, role R exist such that $R(u', u)$ is an aux-simple atom in $q_\sim$. By the definition of aux-simple atom, we have $\tau(u) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, $u' \neq u$, role R is simple, and $\tau(u') = \tau(u)$ implies that $\mathsf{Self}_R(\tau(u)) \notin J$. By Lemma 20 and by the construction of τ, atom $R(\pi_*(u), \pi_*(u)) \in I$ is not a self-loop. Since $\tau(u) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, we have $\pi_*(u) \notin N_I$. Then, by property 1 in Lemma 29, we have $\pi_*(u)$ is of the form $f^B_{T,A}(\pi_*(u'))$. Thus, $\pi(u') \prec \pi(u)$ and no vertex $v'' \in \mathcal{V}$ exists such that $\pi(u') \prec \pi(v'') \prec \pi(u)$. Since $\pi_*(u) = \pi_*(v)$ and $\pi_*(u') = \pi_*(v')$, we have $\langle v', v\rangle \in \mathcal{E}$, as required. Therefore, $\sigma(E_s) \subseteq \mathcal{E}$. Please note that since $\langle \mathcal{V}, \mathcal{E}\rangle$ is a forest, so is $\langle \sigma(V), \sigma(E_s)\rangle$, as required by condition 3 in Definition 9.

It remains to show that each edge $\langle v', v\rangle \in \mathcal{E}$ occurs in $\sigma(E_t)$. Consider an arbitrary edge $\langle v', v\rangle \in \mathcal{E}$. By the definition of $\mathcal{E}$, we have $\pi_*(v') \prec \pi_*(v)$, and so $\pi_*(v)$ is a functional term. Then let $w_0, \ldots, w_k$ be terms such that $w_0 = \pi_*(v')$, $w_k = \pi_*(v)$, and w_i is of the form $f^{B_i}_{T_i,A_i}(w_{i-1})$ for each $i \in [1, k]$. Note that these terms are uniquely defined by the edge, and that, by the construction of I, for each $i \in [1, k]$, we have $T_i(w_{i-1}, w_i) \in I$. By Lemma 20, we have $\mathsf{d}_{T_i}(\delta(w_{i-1}), \delta(w_i)) \in J$. By the definition of τ, we have $\delta(w_0) = \tau(v')$ and $\delta(w_k) = \tau(v)$. Thus, $\langle v', v\rangle \in \sigma(E_t)$, as required by Definition 10.

Finally, consider an edge $\langle v', v\rangle \in \mathcal{E}$ and let $w_0, \ldots, w_k$ be terms uniquely associated with this edge. Then a role R is *compatible* with $\langle v, v'\rangle$ if $\mathsf{d}_R(\delta(w_{i-1}), \delta(w_i)) \in J$ for each $i \in [1, k]$, and $\mathsf{trans}(R) \notin \mathcal{T}$ implies that $k = 1$. In the rest of this proof we will show the following property.

(◊) Each role $R \in L(v', v)$ is compatible with the edge $\langle v', v\rangle$.

By the definition of function exist (Definition 11) and the above definition of compatibility, property (◊) implies that the condition in step 17 is not satisfied for edge $\langle v', v\rangle \in \mathcal{E}$.

For the loop in step 6; let $R(s, t)$ be an arbitrary aux-simple atom in $\sigma(q_\sim)$. By the definition of aux-simple atom, we have $\tau(t) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, role R is simple, and $s \neq t$. By the definition of σ, we have $\pi_*(s) \neq \pi_*(t)$. Thus, $R(\pi_*(s), \pi_*(t)) \in I$ is not a self-loop. Since R is simple, by property 1 in Lemma 29, $\pi_*(t)$ is of the form $f^{A_1}_{T,A}(\pi_*(s))$. By Lemma 20, we have $\mathsf{d}_R(\delta(\pi_*(s)), \delta(\pi_*(t))) \in I$. Let $w_0, \ldots, w_k$ be the terms associated with $\langle s, t\rangle \in \mathcal{E}$. By the form of $\pi_*(t)$ and since $w_0 = \pi_*(s)$ and $w_k = \pi_*(t)$, we have $k = 1$. Thus, property (◊) is satisfied.

For the loop in steps 7–15; let $R(s,t)$ be an arbitrary atom in $\sigma(q_\sim)$ that is neither good nor aux-simple. We next determine the nondeterministic choices that preserve $(\Diamond)$ in step 15, and that satisfy the conditions in steps 8, 9, and 14. By assumption, we have $R(\pi_*(s), \pi_*(t)) \in I$. Since $R(s,t)$ is not good, we have $\tau(t) \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, and either $s \neq t$ or $\mathsf{Self}_R(\tau(t)) \notin J$. By Lemma 20 and by the definition of τ, atom $R(\pi_*(s), \pi_*(t)) \in I$ is not a self-loop. Hence, a role P and terms $w_0, \ldots, w_m$ with $w_m = \pi_*(t)$ exist in Δ^I that satisfy property 2 in Lemma 29. Since $P \sqsubseteq^*_\mathcal{T} R$ and by properties *(2c)* and *(2d)*, we have $P(\pi_*(s), \pi_*(t)) \in I$. By Lemma 20 and by the construction of τ, we have $P(\tau(s), \tau(t)) \in J$; consequently, the conditions in step 8 are satisfied. Assume that $\mathsf{trans}(P) \notin \mathcal{T}$. By property 2 in Lemma 29, we have $m = 1$ and $\pi_*(v) \prec \pi_*(t)$. Thus, $\pi_*(t)$ is of the form $f^B_{T,A}(\pi_*(s))$. By the definition of skeleton $\mathcal{S}$, we have $\langle s,t\rangle \in \mathcal{E}$; thus, the condition in step 9 is not satisfied. Next, we consider two cases.

- s reaches t in $\mathcal{E}$. It follows that $\pi_*(s) \prec \pi_*(t)$. Let $v_0, \ldots, v_n$ be the unique path connecting s to t in $\mathcal{S}$. Since $\pi_*(s) \prec \pi_*(t)$, we have $w_0 = \pi_*(v_0)$ and $w_m = \pi_*(v_n)$. Thus, for each $i \in [1,n]$, a unique index ℓ_i exists such that $\pi_*(v_i) = w_{\ell_i}$. By property *(2d)* in Lemma 29 and by Lemma 20, role P is compatible with $\langle v_{i-1}, v_i\rangle \in \mathcal{E}$.
- s does not reach t in $\mathcal{E}$. Hence, $\pi_*(s) \not\prec \pi_*(t)$. Let a_t be the root of t in $\mathcal{E}$, and let $v_0, \ldots, v_n$ be the unique path connecting a_t to t in $\mathcal{E}$. Since $\pi_*(s) \not\prec \pi_*(t)$, we have $w_0 = a_t$, $w_0 \prec w_m$, and $w_m = \pi_*(v_n)$; moreover, $\mathsf{trans}(P) \in \mathcal{T}$, and $P(\pi_*(s), a_t) \in I$. Thus, for each $i \in [1,n]$, a unique index ℓ_i exists such that $\pi_*(v_i) = w_{\ell_i}$. By Lemma 20, we have $P(\tau(s), a_t) \in J$, so condition in step 17 is not satisfied. Then, by property *(2d)* in Lemma 29 and by Lemma 20, role P is compatible with $\langle v_{i-1}, v_i\rangle \in \mathcal{E}$. □

E Proof of Theorem 13

Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A}\rangle$ be a satisfiable $\mathcal{ELHO}^s$ KB and let $\mathsf{D}_\mathcal{K}$ be the datalog program for $\mathcal{K}$; furthermore, let $q = \exists \vec{y}.\psi(\vec{x}, \vec{y})$ be a CQ, and let τ be a candidate answer for q and $\mathsf{D}_\mathcal{K}$.

Theorem 13. *Function* $\mathsf{isSound}(q, \mathsf{D}_\mathcal{K}, \tau)$ *can be implemented so that*

1. *it runs in nondeterministic polynomial time,*
2. *if each binary atom in q is either good or aux-simple w.r.t. τ, it runs in polynomial time, and*
3. *if the TBox $\mathcal{T}$ and the query q are fixed, it runs in polynomial time in the size of the ABox $\mathcal{A}$.*

Proof. Please note that since the number of variables occurring in each rule in $\mathsf{D}_\mathcal{K}$ is fixed, we can compute the set of all consequences of $\mathsf{D}_\mathcal{K}$ in polynomial time (Dantsin et al. 2001). Thus all the following operations can be implemented to run in polynomial time:

- computing sets $\mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ and $\mathsf{ind}_{\mathsf{D}_\mathcal{K}}$,
- computing the connection graph cg for q and τ,
- given two individuals u' and u and a set of role L, checking whether $\mathsf{exist}(u', u, L)$ returns t, and
- checking whether a binary atom is good or aux-simple w.r.t. τ.

Next, we argue that we can compute relation $\sim$ in polynomial time. As stated above, we can evaluate in polynomial time the precondition of the (fork) rule. In addition, the size of relation $\sim$ is bounded by $|N_T(q)|^2$, the rules used to compute it are monotonic, and each inference can be applied in polynomial time, so the claim follows.

Also, we can check in linear time whether a directed graph is a acyclic by searching for a topological ordering of its vertices (Cormen et al. 2009). Therefore, condition 2 in Definition 8 can be checked in polynomial time. Therefore, steps 1 and 2 in Algorithm 1 run in time polynomial in the input size, and property 2 holds.

Since steps 3–15 in Algorithm 1 can all clearly be implemented to run in nondeterministic polynomial time in the input size, property 1 also holds.

For property 3, assume that the TBox $\mathcal{T}$ and the query q are fixed. Then the set of all consequences of $\mathsf{D}_\mathcal{K}$ can be computed in time polynomial in $\mathcal{A}$. Given that the number of variables occurring in q is fixed, the number of guessing steps required in steps 3 and 4 is fixed; also, the number of alternatives for these steps is linear in the size of $\mathcal{A}$. Thus, steps 3 and 4 require polynomial time. Moreover, the maximum number of iterations of the for-loop in steps 7-15 is fixed. The number of alternatives for the guessing steps in line 8 is fixed as well. Therefore, steps 7-15 require time polynomial in the size of $\mathcal{A}$. All other steps can clearly be implemented in time polynomial in the size of $\mathcal{A}$, thus $\mathsf{isSound}$ runs in time polynomial in the size of $\mathcal{A}$. □

F Proof of Theorem 14

Theorem 14. *Checking whether a candidate answer is sound is* NP*-hard.*

Proof. The proof is by reduction from the NP-hard problem of checking the satisfiability of a 3CNF formula (Garey and Johnson 1979). Let $\varphi = \bigwedge_{j=1}^m C_j$ be a 3CNF formula over variables $\{v_1, \ldots, v_n\}$, where each C_j is a set of three literals $C_j = \{l_{j,1}, l_{j,2}, l_{j,3}\}$. A sequence $\nu = l_{j_1,k_1}, \ldots, l_{j_\ell,k_\ell}$ of literals from φ is *consistent* if $\{v_i, \neg v_i\} \not\subseteq \nu$ for each $i \in [1,n]$.

Such a ν is a *truth assignment* for φ if for each $j \in [1, m]$, a literal $l \in \nu$ exists such that $l = l_{j,k}$ for some $k \in [1, 3]$. Then φ is satisfiable if and only if there exists a consistent truth assignment for φ.

Let φ be a 3CNF formula. We next define an $\mathcal{ELHO}^s$ KB $\mathcal{K}_\varphi$ and a Boolean CQ q_φ such that $\mathcal{K}_\varphi$ does not contain axioms of type 2, and $\mathcal{K}_\varphi \models q_\varphi$ if and only if there exists a consistent truth assignment for φ. Let D_φ be the translation of $\mathcal{K}_\varphi$ into datalog. By Theorem 12 and since $\mathcal{K}_\varphi$ does not contain axioms of type 2, we have $\mathcal{K}_\varphi \models q_\varphi$ if and only if a candidate answer τ for q_φ and D_φ exists such that $\mathsf{isSound}(q_\varphi, \mathsf{D}_\varphi, \tau)$ returns t. Then, to prove the theorem, we show that there exists a unique candidate answer τ_φ for q_φ and D_φ. Thus, $\mathcal{K}_\varphi \models q_\varphi$ if and only if $\mathsf{isSound}(q_\varphi, \mathsf{D}_\varphi, \tau_\varphi)$ returns t, hence, $\mathsf{isSound}(q_\varphi, \mathsf{D}_\varphi, \tau_\varphi)$ returns t if and only if φ is satisfiable.

For convenience, in the following we will specify $\mathcal{K}_\varphi$ using its equivalent formalisation as a rule base Ξ_φ. Moreover, rule base Ξ_φ will not contain equality rules; consequently Ξ_φ has a unique universal interpretation I. We will present our construction of Ξ_φ in stages, and for each we will describe how it affects the universal interpretation I.

Our encoding of Ξ_φ uses a fresh individual a, fresh concepts A and G, fresh roles R and T, a fresh concept $L_{j,k}$ and a fresh role $S_{j,k}$ uniquely associated to each literal $l_{j,k}$, a fresh concept C_j uniquely associated to each clause C_j, and fresh roles P_i, N_i, and T_i uniquely associated to each variable v_i.

Before presenting Ξ_φ, we need a couple of definitions. Given two terms w' and w from I, and a word $\rho = T_1 \cdots T_\ell$ over N_R, we write $\rho(w', w) \in I$, if terms $w_0, \ldots, w_\ell$ with $w_0 = w'$ and $w_m = w$ exist such that $T_i(w_{i-1}, w_i) \in I$ for each $i \in [1, \ell]$. In the following, we uniquely associate to each sequence $\nu = l_{j_1,k_1}, \ldots, l_{j_\ell,k_\ell}$ of literals from φ the word $\rho_\nu = R \cdot S_{j_1,k_1} \cdot R \cdot S_{j_2,k_2} \cdots R \cdot S_{j_\ell,k_\ell} \cdot R$.

We next present Ξ_φ which consists of four parts.

The first part of rule base Ξ_φ contains atom (3) and rules (4)–(7). Then for each sequence ν of literals from φ, a term w_ν exists in I such that $\rho_\nu(a, w_\nu) \in I$ and $G(w_\nu) \in I$.

$$A(a) \tag{3}$$

$$A(x) \to \exists z.R(x, z) \wedge C_j(z) \qquad \forall j \in [1, m] \tag{4}$$

$$A(x) \to \exists z.R(x, z) \wedge G(z) \tag{5}$$

$$C_j(x) \to \exists z.S_{j,k}(x, z) \wedge L_{j,k}(z) \qquad \forall j \in [1, m]\ \forall k \in [1, 3] \tag{6}$$

$$L_{j,k}(x) \to A(x) \qquad \forall j \in [1, m]\ \forall k \in [1, 3] \tag{7}$$

The second part of rule base Ξ_φ contains rules (8) and (9). Consider an arbitrary literal $l_{j,k}$ and arbitrary terms w' and w in I such that $S_{j,k}(w', w) \in I$. Then, for each variable v_i, we have $P_i(w', w) \in I$ if and only if sequence $v_i, l_{j,k}$ is consistent; and $N_i(w', w) \in I$ if and only if sequence $\neg v_i, l_{j,k}$ is consistent.

$$S_{j,k}(x, y) \to P_i(x, y) \qquad \forall j \in [1, m], \forall k \in [1, 3], \forall i \in [1, n] \text{ with } l_{j,k} = v_i \text{ or } l_{j,k} \text{ is not over } v_i \tag{8}$$

$$S_{j,k}(x, y) \to N_i(x, y) \qquad \forall j \in [1, m], \forall k \in [1, 3], \forall i \in [1, n] \text{ with } l_{j,k} = \neg v_i \text{ or } l_{j,k} \text{ is not over } v_i \tag{9}$$

The third part of rule base Ξ_φ contains rules (10)–(13). Then for each sequence ν of literals from φ with $\rho_\nu(a, w_\nu) \in I$ and each $i \in [1, n]$, we have $P_i(a, w_\nu) \in I$ if and only if ν, v_i is consistent; and $N_i(a, w_\nu) \in I$ if and only if $\nu, \neg v_i$ is consistent.

$$R(x, y) \to P_i(x, y) \qquad \forall i \in [1, n] \tag{10}$$

$$R(x, y) \to N_i(x, y) \qquad \forall i \in [1, n] \tag{11}$$

$$P_i(x, y) \wedge P_i(y, z) \to P_i(x, z) \qquad \forall i \in [1, n] \tag{12}$$

$$N_i(x, y) \wedge N_i(y, z) \to N_i(x, y) \qquad \forall i \in [1, n] \tag{13}$$

The fourth part of rule base Ξ_φ contains rules (14)–(16). Then for each sequence ν of literals from φ with $\rho_\nu(a, w_\nu) \in I$ and each $i \in [1, n]$, we have $T_i(a, w_\nu) \in I$ if and only if $\{v_i, \neg v_i\} \not\subseteq \nu$; furthermore, for each clause C_j, a term w_j exists such that $T(w_j, w_\nu) \in I$ if and only if an index $k \in [1, 3]$ exists such that $l_{j,k} \in \nu$.

$$P_i(x, y) \to T_i(x, y) \qquad \forall i \in [1, n] \tag{14}$$

$$N_i(x, y) \to T_i(x, y) \qquad \forall i \in [1, n] \tag{15}$$

$$L_i(x, y) \to T(x, y) \qquad \forall i \in [1, n] \tag{16}$$

Query q_φ is given in (17). Then $\mathcal{K}_\varphi \models q_\varphi$ if and only if a sequence ν of literals from φ exists such that $T_i(a, w_\rho) \in I$ for each $i \in [1, n]$, and, for each $j \in [1, m]$, a term w_j exists such that $C_j(w_j) \in I$ and $T(w_j, w_\rho) \in I$; hence, $\mathcal{K}_\varphi \models q_\varphi$ if and only if there exists a consistent truth assignment ν for φ.

$$q_\varphi = \exists y \exists z_1, \ldots \exists z_m.\ G(y) \wedge \bigwedge_{i=1}^{n} T_i(a, y) \wedge \bigwedge_{j=1}^{m} C_j(z_j) \wedge T(z_j, y) \tag{17}$$

Program D_φ contains all the atoms and the rules in Ξ_φ apart from rules (4)–(6) which are replaced by rules (18)–(20).

$$A(x) \to R(x, o_{R,C_j}) \wedge C_j(o_{R,C_j}) \qquad \forall j \in [1, m] \tag{18}$$

$$A(x) \to R(x, o_{R,G}) \wedge G(o_{R,G}) \tag{19}$$
$$C_j(x) \to S_{j,k}(x, o_{S_{j,k},L_{j,k}}) \wedge L_{j,k}(o_{S_{j,k},L_{j,k}}) \qquad \forall j \in [1,m], \forall k \in [1,3] \tag{20}$$

We next define substitution τ_φ, and we show that $\mathsf{D}_\varphi \models \tau_\varphi(q_\varphi)$. Substitution τ_φ is given in (21).

$$\tau(y) = o_{R,G} \text{ and } \tau(z_j) = o_{R,C_j} \qquad \forall j \in [1,m] \tag{21}$$

Consider an arbitrary $i \in [1,n]$; we show that $\mathsf{D}_\varphi \models T_i(a, \tau(y))$. By atom (3), and by rule (19) we have $\mathsf{D}_\varphi \models R(a, \tau(y))$. But then, by rules (10), (11), (14) (15), we immediately have $\mathsf{D}_\varphi \models T_i(a, \tau(y))$, as required. Next, consider an arbitrary $j \in [1,m]$; we show that $\mathsf{D}_\varphi \models T(\tau(z_j), \tau(y))$. Please note that by atom (3) and by rules (18), we have $\mathsf{D}_\varphi \models C_j(\tau(z_j))$. Consider an arbitrary literal $l_{j,k}$ in C_j. By rules (20) and (7), there exists an individual u such that $\mathsf{D}_\varphi \models S_{j,k}(\tau(z_j), u)$ and $\mathsf{D}_\varphi \models A(u)$. Then, by rule (19) we have $\mathsf{D}_\varphi \models R(u, \tau(y))$. By rules (10)–(16), we have $\mathsf{D}_\varphi \models T(\tau(z_j), \tau(y))$, as required.

We are left to show that τ_φ is unique. Consider an arbitrary candidate answer ξ for q_φ and D_φ. Since $G(y)$ is an atom in q and only rule (19) derives assertions over G, we must have $\xi(y) = o_{R,G}$. Consider an arbitrary $j \in [1,m]$. Due to atom $C_j(z_j)$ in q, and only rule (18) derives assertions over concept C_j, we must have $\xi(y) = o_{R,C_j}$. Thus, $\xi = \tau_\varphi$, as required. □

G Proof of Theorem 17

Let $\mathcal{K}$ be a satisfiable $\mathcal{ELHO}$ KB, and let $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$ be the rule base and the datalog program associated with $\mathcal{K}$, respectively; furthermore, let $\mathsf{ind}_{\mathsf{D}_\mathcal{K}}$ and $\mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ be as specified in Definition 4, and let q be an arborescent query. We next show that function $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ returns t if and only if $\Xi_\mathcal{K} \models q$, and that $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ runs in time polynomial in the input size.

To this end, we start with a couple of definitions. Let I and J be universal interpretations for $\Xi_\mathcal{K}$ and $\mathsf{D}_\mathcal{K}$, respectively. Moreover, let δ be the mapping as specified at the beginning of Section B. For each set $V \in \mathsf{RT}$, let q_V be the arborescent query obtained from q by replacing each variable $y \in V$ with a fresh variable y_V, and by removing each variable z, and all the atoms involving z, in the resulting query such that $z \neq y_V$ and z is not a descendant of y_V in dg_q.

Lemma 33. *Function* $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ *returns* t *if and only if* $\Xi_\mathcal{K} \models q$*. Furthermore,* $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ *runs in time polynomial in the input size.*

Proof. Since q does not contain individuals, we have $\Xi_\mathcal{K} \models q$ if and only if a substitution π with $\mathsf{dom}(\pi) = N_V(q)$ exists such that $\pi(q) \subseteq I$. In the following, we show that the latter is the case if and only if function $\mathsf{entails}(\mathsf{D}_\mathcal{K}, q)$ returns t.

Let RT be as specified in Definition 16. Let M be the largest $n \in \mathbb{N}$ for which a set $V \in \mathsf{RT}$ exists whose level is n. By the definition, M is the length of the longest path in dg_q, and so it is linearly bounded by the size of q.

Next, we argue that we can compute set RT in polynomial time. This follows from the fact that the rules used to compute it are monotonic, each rule can be applied at most M times, and each rule application introduces at most $|q|$ sets in RT, and the size of each set is linearly bounded by $|q|$.

By Lemmas 20 and 21, and since each rule in $\mathsf{D}_\mathcal{K}$ contains a fixed number of variables, each set $V \in \mathsf{RT}$ satisfies the two following properties for each $u \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ and each term $w \in \Delta^I$ with $\delta(w) = u$.

A. $u \in \mathsf{c}_V$ if and only if $B(w) \in I$ for each concept B such that $B(y) \in q$ for some $y \in V$.

B. c_V can be computed in time polynomial in the size of $\mathsf{D}_\mathcal{K}$ and q.

To prove the lemma, we next show that each set $V \in \mathsf{RT}$ satisfies the following properties for each term $u \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$.

1. $u \in \mathsf{A}_V$ if and only if a substitution π with $\mathsf{dom}(\pi) = N_V(q_V)$ exists such that $\pi(q_V) \subseteq I$ and $\delta(\pi(y_V)) = u$.
2. A_V can be computed in time polynomial in the size of $\mathsf{D}_\mathcal{K}$ and q.

The proof goes by reverse-induction on the level of sets in RT.

Base case. Consider an arbitrary set $V \in \mathsf{RT}$ of level M. By the definition of A_V, we have $\mathsf{A}_V = \mathsf{c}_V$. Furthermore, an atom $B(y_V)$ is in q_V if and only if a variable $y \in V$ exists such that $B(y)$ in q. Properties 1 and 2 follow immediately from A and B.

Inductive step. Consider an arbitrary $n \in \mathbb{N}$. Let $V \in \mathsf{RT}$ be an arbitrary set of level n, and assume that properties 1 and 2 hold for each set $W \in \mathsf{RT}$ of level $n+1$; we show that the properties are satisfied by V.

For property 1, let u be an arbitrary term in $\mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. By the definition of A_V we have $\mathsf{A}_V = \mathsf{c}_V \cap (\mathsf{i}_V \cup \mathsf{a}_V)$. By the definition of q_V and from property A, it suffices to show that $u \in (\mathsf{i}_V \cup \mathsf{a}_V)$ if and only if a substitution π exists such that $\pi(q_V) \subseteq I$ and $\delta(\pi(y_V)) = u$. We consider the two directions of 1 separately.

($\Rightarrow$) Assume that $u \in (\mathsf{i}_V \cup \mathsf{a}_V)$. We distinguish two cases.

- $u \in \mathsf{i}_V$. Thus, $u \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$ and $u \in N_I$. Consider an arbitrary variable $y \in \mathsf{P}_V$ and an arbitrary role $R \in \mathsf{r}_y$. Then, an individual $u' \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ exists such that $u' \in \mathsf{A}_{\{y\}}$ and $R(u', u) \in J$. By the inductive hypothesis, a substitution $\pi_{\{y\}}$ exists such that $\pi_{\{y\}}(q_{\{y\}}) \subseteq I$ and $\delta(w') = u'$, where $w' = \pi_{\{y\}}(y_{\{y\}})$. By Lemma 21, we have $R(w', u) \in I$. Then let π be the substitution such that $\pi(y_V) = u$ and $\pi_{\{y\}} \subseteq \pi$ for each $y \in \mathsf{P}_V$. Then $\pi(q_V) \subseteq I$, as required.

- $u \in \mathsf{a}_V$. Thus, $u \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ and u is of the form $o_{P,A}$. Consider an arbitrary role R such that $R \in \mathsf{r}_y$ for some $y \in \mathsf{P}_V$. Then, an individual $u' \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$ exists such that $u' \in \mathsf{A}_{\mathsf{P}_V}$ and $\mathsf{d}_R(u', u) \in J$. By the inductive hypothesis, a substitution π_{P_V} exists such that $\pi_{\mathsf{P}_V}(q_{\mathsf{P}_V}) \subseteq I$ and $\delta(w') = u'$, where $w' = \pi_{\mathsf{P}_V}(y_{\mathsf{P}_V})$. By Lemma 21, a term $w \in \Delta^I$ exists such that $\delta(w) = o_{P,A}$ and $P(w', w) \in I$ and $P \sqsubseteq^*_\mathcal{T} R$. Since no rule is applicable to I, we have $R(w', w) \in I$. Then let π be the substitution such that $\pi(y_V) = w$, $\pi_{\mathsf{P}_V} \subseteq \pi$, and $\pi(y) = w'$ for each $y \in \mathsf{P}_V$. Then $\pi(q_V) \subseteq I$, as required.

($\Leftarrow$) Assume that a substitution π exists such that $\pi(q_V) \subseteq I$ and $\delta(\pi(y_V)) = u$. We distinguish two cases.

- $\pi(y_V) \in N_I$. By Lemma 24, we have $u \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$ and $u \in N_I$. Consider an arbitrary variable $y \in \mathsf{P}_V$ and an arbitrary role $R \in \mathsf{r}_y$. By the definition of q_V, atom $R(y, y_V)$ occurs in q_V. Since $R(\pi(y), \pi(y_V)) \in I$, by Lemma 20, we have $R(\delta(\pi(y)), \delta(\pi(y_V))) \in J$ and $\delta(\pi(y)) \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. Due to the construction of q_V, $q_{\{y\}}$ is a subquery of q_V; thus, $\pi(q_{\{y\}}) \subseteq I$. Hence, by the inductive hypothesis, we have $\delta(\pi(y)) \in \mathsf{A}_{\{y\}}$. Due to $\delta(\pi(y_V)) = u$ and $u \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}}$, we have $u \in \mathsf{i}_V$, as required.
- $\pi(y_V) \notin N_I$. By Lemma 24, we have $u \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. Consider an arbitrary variable $y \in \mathsf{P}_V$ and an arbitrary role $R \in \mathsf{r}_y$. By the definition of q_V, atom $R(y, y_V)$ occurs in q_V. By assumption, we have $R(\pi(y), \pi(y_V)) \in I$. Since $\mathcal{K}$ is an $\mathcal{ELHO}$ knowledge base, role R is simple and $\mathsf{Self}_R(\pi(y_V)) \notin I$. By Lemma 29, term $\pi(y_V)$ is of the form $f^B_{P,A}(\pi(y))$. Therefore, for each variable $z \in \mathsf{P}_V$, we have $\pi(z) = \pi(y)$. Furthermore, by Lemma 20, we have $\mathsf{d}_R(\delta(\pi(y)), \delta(\pi(y_V))) \in J$ and $\delta(\pi(y)) \in \mathsf{ind}_{\mathsf{D}_\mathcal{K}} \cup \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$. Let substitution π_{P_V} for q_{P_V} be obtained from π by setting $\pi_{\mathsf{P}_V}(y_{\mathsf{P}_V}) = \pi(y)$. Then, we have $\pi_{\mathsf{P}_V}(q_{\mathsf{P}_V}) \subseteq I$. Hence, by the inductive hypothesis, we have $\delta(\pi(y)) \in \mathsf{A}_{\mathsf{P}_V}$. Due to $\delta(\pi(y_V)) = u$ and $u \in \mathsf{aux}_{\mathsf{D}_\mathcal{K}}$, we have $u \in \mathsf{a}_V$, as required.

For property 2, we show that A_V can be computed in time polynomial in the size of $\mathsf{D}_\mathcal{K}$ and q. By property B, set c_V can be computed in time polynomial in q and $\mathsf{D}_\mathcal{K}$, hence in the rest of this proof we focus on sets i_V and a_V.

- i_V. Please note that the number of variables in P_V is bounded by the size of q. Then consider an arbitrary variable $y \in \mathsf{P}_V$. By the inductive hypothesis, set $\mathsf{A}_{\{y\}}$ can be computed in time polynomial in q and $\mathsf{D}_\mathcal{K}$. Then, since each rule in $\mathsf{D}_\mathcal{K}$ contains a constant number of variables, i_V can also be computed in time polynomial in the size of q and $\mathsf{D}_\mathcal{K}$.
- a_V. By the inductive hypothesis, set $\mathsf{A}_{\mathsf{P}_V}$ can be computed in time polynomial in the size of q and $\mathsf{D}_\mathcal{K}$. Then, since each rule in $\mathsf{D}_\mathcal{K}$ contains a constant number of variables, a_V can also be computed in time polynomial in the size of q and $\mathsf{D}_\mathcal{K}$. □

H Proof of Theorem 18

In this section, we prove the lower bounds established in Theorem 18, all of which are proved by reducing the NP-hard problem of checking the satisfiability of a CNF formula ψ (Garey and Johnson 1979). For the rest of this section, we fix a CNF formula $\psi = \bigwedge_{j=1}^m C_j$ where each C_j is a clause over variables $\{v_1, \ldots, v_n\}$.

For convenience, we will assume that $\mathcal{ELHO}$ TBoxes can contain axioms of the form $A_1 \sqsubseteq \exists R.\{a\}$ with $A_1 \in N_C \cup \{\top\}$, R a role, and $\{a\}$ a nominal. The translation into rules for this type of axiom is given by $A_1(x) \to R(x, a)$. Moreover, in the following we will specify DL knowledge bases using their equivalent formalisation as rule bases. Furthermore, all the rules from this section do not contain equality, and so each rule base has exactly one universal interpretation.

Then to prove the theorem, we will first define a rule base Ξ_0 containing only rules of types 1 and 7 from Table 1, and a Boolean CQ q_0 over Ξ_0, after which we will show that

1. a rule base Ξ_1 and a query q_1 exist such that Ξ_1 is in $\mathcal{ELHO}$, query $q_0 \wedge q_1$ is acyclic, and ψ is satisfiable if and only if $\Xi_0 \cup \Xi_1 \models q_0 \wedge q_1$,
2. a rule base Ξ_2 and a query q_2 exist such that Ξ_2 contains a single rule of type 8, query $q_0 \wedge q_2$ is arborescent, and ψ is satisfiable if and only if $\Xi_0 \cup \Xi_2 \models q_0 \wedge q_2$, and
3. a rule base Ξ_3 and a query q_3 exist such that Ξ_3 contains a single rule of type 9, query $q_0 \wedge q_3$ is arborescent, and ψ is satisfiable if and only if $\Xi_0 \cup \Xi_3 \models q_0 \wedge q_3$.

H.1 Construction of Ξ_0 and q_0

Our encoding uses a fresh role R, a fresh concepts G, fresh concepts T_i, F_i, and A_i uniquely associated with each variable v_i in ψ, and fresh concepts C_j uniquely associated to each clause C_j in ψ. Rule base Ξ_0 contains a ground atom (22), and rules (23)–(29). Let I_0 be the universal interpretation of Ξ_0, we next describe how the rules in Ξ_0 model the structure of I_0. Atom (22) and rules (23)–(27) encode a binary tree of depth $n + 1$ in I_0 rooted in individual a in which edges are labelled by role R, each leaf node satisfies concept G, and each node w at depth $1 \leq i \leq n$ satisfies exactly one of T_i and F_i. Then node w represents a positive truth assignment to v_i, if $T_i(w) \in I_0$; otherwise, it represents a negative truth assignment to v_i. Consequently, a path from a to a leaf node in the tree represents a truth assignment to the variables in ψ. Finally, rules (28) and (29) ensure that, for

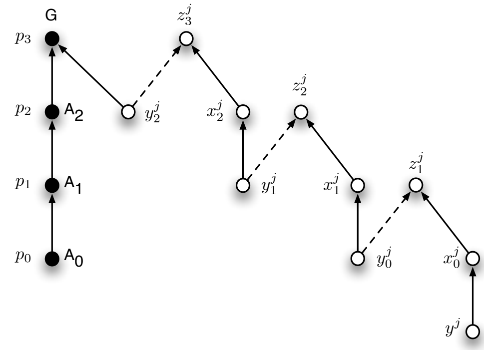

Figure 2: A graphical representation of query $q_0 \wedge q_1$ for $n = 2$ and an arbitrary $j \in [1, m]$

each node w at depth $1 \leq i \leq n$ in the tree, if the truth assignment to v_i represented by w makes clause C_j evaluate to true, then $C_j(w) \in I_0$.

$$A_0(a) \tag{22}$$
$$A_{i-1}(x) \rightarrow \exists z.R(x, z) \wedge T_i(z) \qquad \forall i \in [1, n] \tag{23}$$
$$A_{i-1}(x) \rightarrow \exists z.R(x, z) \wedge F_i(z) \qquad \forall i \in [1, n] \tag{24}$$
$$T_i(x) \rightarrow A_i(x) \qquad \forall i \in [1, n] \tag{25}$$
$$F_i(x) \rightarrow A_i(x) \qquad \forall i \in [1, n] \tag{26}$$
$$A_n(x) \rightarrow \exists z.R(x, z) \wedge G(z) \tag{27}$$
$$T_i(x) \rightarrow C_j(x) \qquad \forall i \in [1, n]\ \forall j \in [1, m] \text{ with } v_i \in C_j \tag{28}$$
$$F_i(x) \rightarrow C_j(x) \qquad \forall i \in [1, n]\ \forall j \in [1, m] \text{ with } \neg v_i \in C_j \tag{29}$$

Query q_0 is given in (30). It should be clear that, for each substitution π with $\mathsf{dom}(\pi) = \vec{p}$ such that $\pi(q_0) \subseteq I_0$, substitution π represents a truth assignment for ψ, $\pi(p_0) = a$, and $\pi(p_{n+1})$ is a leaf.

$$q_0 = \exists p_0 \ldots \exists p_{n+1}. \bigwedge_{i=0}^{n} [A_i(p_i) \wedge R(p_i, p_{i+1})] \wedge G(p_{n+1}) \tag{30}$$

H.2 Construction of Ξ_1 and q_1

Our rule base Ξ_2 uses fresh roles S_j and individuals c_j uniquely associated to each clause C_j of ψ. Then rule base Ξ_1 contains atoms (31), and rules (32) and (33). Let I_1 be the universal interpretation of $\Xi_0 \cup \Xi_1$, then we clearly have $I_0 \subseteq I_1$. We next describe how the rules and atoms in Ξ_1 modify the tree encoded by Ξ_0. Rule (32) ensure that, for each node w in the tree whose truth assignment makes clause C_j evaluate to true, there exists an edge labelled by S_j from w to individual c_j. Atom (31) generates an edge labelled by R connecting c_j to itself. Finally, rule (33) labels each edge in the tree and each looping edge with S_j.

$$R(c_j, c_j) \qquad \forall j \in [1, m] \tag{31}$$
$$C_j(x) \rightarrow S_j(x, c_j) \qquad \forall j \in [1, m] \tag{32}$$
$$R(x, y) \rightarrow S_j(x, y) \qquad \forall j \in [1, m] \tag{33}$$

Our encoding of query q_1 uses variable p_{n+1} from query q_0, as well as fresh variables $x_0^j, \ldots, x_n^j$, $y^j, y_0^j, \ldots, y_n^j$, and $z_1^j, \ldots, z_{n+1}^j$ uniquely associated with each clause C_j. Next, we associate with each C_j a conjunction φ^j as shown in (34). Then query q_1 is the existential closure of $\bigwedge_j \varphi^j$.

$$\varphi^j = R(y_j, x_0^j) \wedge \bigwedge_{i=1}^{n} [\varphi_i^j \wedge R(y_{i-1}^j, x_i^j)] \wedge \varphi_{n+1}^j \wedge R(y_n^j, p_{n+1}) \tag{34}$$
$$\varphi_i^j = R(x_{i-1}^j, z_i^j) \wedge S_j(y_{i-1}^j, z_i^j) \tag{35}$$

Figure 2 shows a graphical representation of query $q_0 \wedge q_1$ for $n = 2$ and an arbitrary $j \in [1, m]$. The black edges denote atoms over role R, and the dashed edges denote atoms over role S_j. It should be clear that query $q_0 \wedge q_1$ is acyclic, as required. Please note that $q_0 \wedge q_1$ is not arborescent: each variable y_i^j in $q_0 \wedge q_1$ has two parent nodes.

Consider an arbitrary substitution π such that $\pi(q_0) \subseteq I_1$. Intuitively, each conjunct φ^j in q_1 ensures that the truth assignment represented by π makes clause C_j true. We next prove that our encoding is correct.

Lemma 34. *ψ is satisfiable if and only if $\Xi_0 \cup \Xi_1 \models q_0 \wedge q_1$.*

Proof. $(\Rightarrow)$ Assume that a truth assignment $\nu : \{v_1, \dots, v_n\} \mapsto \{\mathsf{t}, \mathsf{f}\}$ exists such that $\nu(\psi) = \mathsf{t}$; we next show that a substitution π with $\mathsf{dom}(\pi) = N_V(q_0 \wedge q_1)$ exists such that $\pi(q_0 \wedge q_1) \subseteq I_1$. To this end, let $\pi(p_0) = a$ and, for each $i \in [1, n]$, let $\pi(p_i)$ be the unique successor of $\pi(p_{i-1})$ in I_1 such that $T_i(\pi(p_i)) \in I_1$ if and only if $\nu(v_i) = \mathsf{t}$. Consider an arbitrary clause C_j. Since $\nu(\psi) = \mathsf{t}$, a literal $l \in C_j$ exists such that $\nu(l) = \mathsf{t}$. Let $k \in [1, n]$ be the unique index such that literal l is over variable v_k. Then, formulas Ψ_-^j and Ψ_+^j exist such that φ^j is of the form $\varphi^j = \Psi_-^j \wedge \varphi_{k+1}^j \wedge \Psi_+^j$. Next, we extend substitution π as follows:

- π maps each variable occurring in Ψ_-^j to c_j,
- $\pi(x_k^j) = c_j$, $\pi(z_{k+1}^j) = c_j$, and $\pi(y_k^j) = \pi(p_k)$, and
- for each $\ell \in [k+1, n+1]$, π maps each variable with subscript ℓ occurring in Ψ_+^j to $\pi(p_k)$.

Please note that by rules (32) and (33), and by atoms (31), we have $\pi(\Psi_-^j) \cup \pi(\varphi_{k+1}^j) \cup \pi(\Psi_+^j) \subseteq I_1$.

$(\Leftarrow)$ Assume that a substitution π with $\mathsf{dom}(\pi) = N_V(q_0 \wedge q_1)$ exists such that $\pi(q_0 \wedge q_1) \subseteq I_1$; we next show that a truth assignment $\nu : \{v_1, \dots, v_n\} \mapsto \{\mathsf{t}, \mathsf{f}\}$ exists such that $\nu(\psi) = \mathsf{t}$. Please note that $\pi(p_0) = a$ and, for each $i \in [1, n]$, we have $\pi(p_i)$ is a successor of $\pi(p_{i-1})$ in the binary tree encoded by Ξ_0; moreover, for each $j \in [1, m]$, due to atom $R(y_n^j, p_{n+1})$ in (34), we have $\pi(y_n^j)$ is the unique parent of $\pi(p_{n+1})$ in the tree, and so $\pi(y_n^j) = \pi(p_n)$. Then, let ν be the truth assignment such that, for each $i \in [1, n]$, we have $\nu(v_i) = \mathsf{t}$ if and only if $T_i(\pi(p_i)) \in I_1$. We next show that $\nu(\psi) = \mathsf{t}$; that is, for each clause C_j, we have $\nu(C_j) = \mathsf{t}$. Consider an arbitrary clause C_j. By rules (27) and (28), it suffices to find an index $\ell \in [1, n]$ such that $C_j(\pi(p_\ell)) \in I_1$. Towards this goal, we first show that substitution π satisfies the following property for each $i \in [0, n]$.

$(\diamondsuit)$ If for each $\ell \in [i, n]$ we have $\pi(z_{\ell+1}^j) \notin N_I$, then $\pi(y_i^j) = \pi(p_i)$ and $\pi(x_i^j) = \pi(p_i)$.

We proceed by reverse induction on $i \in [0, n]$.

Base case. Let $i = n$. Assume that $\pi(z_{n+1}^j) \notin N_I$. As stated above, we have $\pi(y_n^j) = \pi(p_n)$. Due to atom $S_j(y_n^j, z_{n+1}^j)$ in φ_{n+1}^j, we have $\pi(z_{n+1}^j)$ is a successor of $\pi(p_n)$. Due to atom $R(x_n^j, z_{n+1}^j)$ in φ_{n+1}^j, we have $\pi(x_n^j)$ is the parent of $\pi(z_{n+1}^j)$, thus $\pi(x_n^j) = \pi(p_n)$, as required.

Inductive step. Consider an arbitrary $i \in [0, n-1]$. Assume that the property holds for $i + 1$, and that $\pi(z_{\ell+1}^j) \notin N_I$ for each $\ell \in [i, n]$. By the inductive hypothesis, we have $\pi(y_{i+1}^j) = \pi(p_{i+1})$ and $\pi(x_{i+1}^j) = \pi(p_{i+1})$. Due to atom $R(y_i^j, x_{i+1}^j)$ in (34), we have $\pi(y_i^j)$ is the unique parent of $\pi(p_{i+1})$ in the tree, and so $\pi(y_i^j) = \pi(p_i)$. Due to atom $S_j(y_i^j, z_{i+1}^j)$ in φ_{i+1}^j, we have $\pi(z_{i+1}^j)$ is a successor of $\pi(p_i)$. Finally, due to atom $R(x_i^j, z_{i+1}^j)$ in φ_{i+1}^j, we have $\pi(x_i^j) = \pi(p_i)$, as required.

We next show that an index $i \in [0, n]$ exists such that $\pi(z_{i+1}^j)$ is mapped to an individual. Assume the opposite, hence, for each $i \in [0, n]$, we have $\pi(z_{i+1}^j) \notin N_I$. By property $(\diamondsuit)$, we then have $\pi(x_0^j) = \pi(p_0)$. Since atom $S(y^j, x_0^j)$ occurs in q_1, $\pi(x_0^j) = \pi(p_0) = a$, and a does not have incoming edges in I_1, this is a contradiction.

Finally, let $\ell \in [0, n]$ be the largest index such that $\pi(z_{\ell+1}^j) \in N_I$. We show that $\pi(y_\ell^j) = \pi(p_\ell)$ by considering two cases.

- $\ell = n$. As stated above, we have $\pi(y_n^j) = \pi(p_n)$.
- $\ell < n$. By property $(\diamondsuit)$, we have $\pi(x_{\ell+1}^j) = \pi(p_{\ell+1})$. Due to atom $R(y_\ell^j, x_{\ell+1}^j)$, we have $\pi(y_\ell^j) = \pi(p_\ell)$.

In either cases, we have $\pi(y_\ell^j) = \pi(p_\ell)$.

We next show that $\ell > 0$. Assume the opposite, hence $\ell = 0$. Since π maps $\pi(z_1^j) \in N_I$ and $\pi(y_0^j) = a$, atom $S_j(y_0^j, z_1^j)$ occurs in q_1, and a is not connected to an individual in I_1, we have $S_j(\pi(y_0^j), \pi(z_1^j)) \notin I_1$, which is a contradiction.

Therefore, we have $\pi(y_\ell^j) = \pi(p_\ell)$ and $\ell > 1$. By rules (32) and (33), and since atom $S_j(y_\ell^j, z_{\ell+1}^j)$ occurs in q_1, we have $\pi(z_{\ell+1}^j) = c_j$ and $C_j(\pi(p_\ell)) \in I_1$, as required. □

H.3 Construction of Ξ_2 and q_2

Rule base Ξ_2 consists only of rule (36). Let I_2 be the universal interpretation of $\Xi_0 \cup \Xi_2$, then we clearly have $I_0 \subseteq I_2$. Note that rule (36) modifies the tree encoded by Ξ_0 by connecting each node w in the tree via an R edge to all nodes that occur on the path connecting w to the root a.

$$R(x, y) \wedge R(y, z) \rightarrow R(x, z) \tag{36}$$

Our encoding of query q_2 uses variable p_{n+1} from query q_0, as well as a fresh variable x_j uniquely associated to each clause C_j. Query q_2 is given in (37).

$$\exists x_1 \ldots \exists x_m \bigwedge_{i=1}^{m} C_j(x_j) \wedge R(x_j, p_{n+1}) \tag{37}$$

It should be clear that $q_0 \wedge q_2$ is arborescent. Now, consider a substitution π such that $\pi(q_0 \wedge q_2) \subseteq I_2$. By the definition of q_2, for each $j \in [1, m]$, a term w_j exists such that w_j occurs on the unique path connecting leaf $\pi(p_{n+1})$ to the root of the tree a and $C_j(w_j) \in I_2$; therefore, an index $\ell_j \in [1, n]$ exists such that $\pi(p_{\ell_j}) = \pi(x_j)$. By the definition of Ξ_0 and q_0, the truth assignment represented by π then makes C_j evaluate to true. Thus, the following holds.

Lemma 35. *ψ is satisfiable if and only if $\Xi_0 \cup \Xi_2 \models q_0 \wedge q_2$.*

H.4 Construction of Ξ_3 and q_3

Rule base Ξ_3 consists of rules (38). Let I_3 be the universal interpretation of $\Xi_0 \cup \Xi_3$, then we clearly have $I_0 \subseteq I_3$. Rules (38) modify the tree encoded by Ξ_0 by generating an edge labelled by R connecting each node to itself.

$$\top(x) \rightarrow \mathsf{Self}_R(x) \wedge R(x, x) \qquad \forall j \in [1, m] \tag{38}$$

Our encoding of query q_3 uses variable p_{n+1} from query q_0, as well as fresh variables $x_0^j, \ldots, x_n^j$ uniquely associated to each clause C_j. Furthermore, we associate to each clause C_j the formula φ^j in (39). Then q_3 is the existential closure of $\bigwedge_j \varphi^j$.

$$\varphi^j = C_j(x_0^j) \wedge R(x^j, x_0^j) \wedge \bigwedge_{i=1}^{n} R(x_{i-1}^j, x_i^j) \wedge R(x_n^j, p_{n+1}) \tag{39}$$

It should be clear that $q_0 \wedge q_3$ is arborescent. Consider an arbitrary substitution π such that $\pi(q_0 \wedge q_3) \subseteq I_3$ and an arbitrary clause C_j. Then, formula φ^j ensures that a path of length $n + 1$ exists connecting leaf node $\pi(p_{n+1})$ to one arbitrary node occurring in the unique path connecting $\pi(p_{n+1})$ to root a. Since the tree has depth $n+1$ and the leaves and the root of the tree do no satisfy concept C_j, variable x_0^j must be mapped to some node at depth $1 \leq i \leq n$. By the definition of Ξ_0 and q_0, the truth assignment represented by π then makes C_j evaluate to true. Thus, the following holds.

Lemma 36. *ψ is satisfiable if and only if $\Xi_0 \cup \Xi_3 \models q_0 \wedge q_3$.*