



**TECHNISCHE  
UNIVERSITÄT  
DRESDEN**

Technische Universität Dresden  
Institute for Theoretical Computer Science  
Chair for Automata Theory

## **LTCS–Report**

### **Temporal Conjunctive Queries in Expressive DLs with Non-simple Roles**

Franz Baader      Stefan Borgwardt      Marcel Lippmann

LTCS-Report 15-17

This is an extended version of the paper “Temporal Conjunctive Queries in Expressive Description Logics with Transitive Roles”, published in the Proceedings of the 28th Australasian Joint Conference on Artificial Intelligence (AI’15).

Postal Address:  
Lehrstuhl für Automatentheorie  
Institut für Theoretische Informatik  
TU Dresden  
01062 Dresden

<http://lat.inf.tu-dresden.de>

Visiting Address:  
Nöthnitzer Str. 46  
Dresden

# Temporal Conjunctive Queries in Expressive DLs with Non-simple Roles

## **Abstract**

In Ontology-Based Data Access (OBDA), user queries are evaluated over a set of facts under the open world assumption, while taking into account background knowledge given in the form of a Description Logic (DL) ontology. Motivated by situation awareness applications, temporal conjunctive queries (TCQs) have recently been proposed as a useful extension of traditional OBDA to support the processing of temporal information. This paper extends the existing complexity analysis of TCQ entailment to very expressive DLs underlying the OWL 2 standard, and in contrast to previous work also allows for queries containing transitive roles.

# Contents

1	Introduction	3
2	Preliminaries	4
3	Atemporal Queries in <i>SHIQ</i> , <i>SHOQ</i> , <i>SHOI</i>	9
4	Combined Complexity of TCQ Entailment	13
5	Data Complexity without Rigid Roles	14
6	Conclusions	17

# 1 Introduction

Given a (man-made or natural) dynamical system that changes its states over time, it is sometimes useful to monitor the behavior of the system in order to detect and then react to critical situations [2]. To achieve this, one can monitor the running system using sensors (e.g., heart rate and blood pressure sensors for a patient) and store the (possibly aggregated and preprocessed) values in a database. Critical situations (such as “blood pressure too high”) can then be described by database queries, and detecting them can be realized through query answering. However, such a pure database solution is unsatisfactory for several reasons. First, one cannot assume that the sensors provide a complete description of the current state of the system, which clashes with the closed world assumption used by database systems. Second, though one usually does not have a complete specification of the system’s behavior, one may have some background knowledge restricting the possible states of the system, which can help to detect more situations.

These problems are addressed by so-called ontology-based data access (OBDA) [13, 27], where (i) the preprocessed and aggregated data are stored in a Description Logic (DL) ABox, which is interpreted with open world assumption, and (ii) the background knowledge is represented in a TBox (ontology) expressed in an appropriate DL. DLs [4] can be used to formalize knowledge using concepts, which represent sets of elements of an application domain, and roles, which describe binary relations between elements. For example, the concept `Patient` can be used to model the set of all patients in a hospital, while `isTreatedWith` represents a relationship between patients and treatments. Concept constructors can then be used to build complex concepts out of atomic concepts and roles. For example, `Patient  $\sqcap$   $\exists$ isTreatedWith.Antibiotics` describes patients treated with antibiotics. In the TBox, one can state subconcept-superconcept relationships, such as  `$\exists$ isTreatedWith.Antibiotics  $\sqsubseteq$   $\exists$ finding.BacterialInfection`, which says that antibiotics treatment is given only if there is a bacterial infection. In the ABox, one can state specific facts about individuals, such as `isTreatedWith(BOB, PENICILLIN)`.

When monitoring a dynamical system, the situation to be recognized may also depend on states of the system at different points in time (such as “fluctuating heart rate”). For this reason, OBDA was extended to the temporal case in [1, 3]. In [3] the complexity of answering temporal conjunctive queries (TCQs) w.r.t. TBoxes was investigated for TBoxes expressed in DLs between *ALC* and *SHQ*. The results are concerned both with data complexity (which is measured only in the size of the data) and with combined complexity (which additionally takes the size of the query and the TBox into account). In addition, the paper considers rigid concepts and roles, whose interpretations must not change over time.

We extend the results of [3] in two directions. First, while being quite expressive, *SHQ* does not contain the constructors nominals and inverse roles, which are quite useful in many applications. Here, we also consider logics that have these

two constructors. However, the main difference is that, though  $\mathcal{SHQ}$  can express transitivity of roles and sub-role relationships, transitive roles and roles with transitive subroles must not occur in queries in [3]. In the present paper, we dispense with this restriction, which unfortunately leads to a dramatic increase in complexity that reflects the results for standard (atemporal) queries (see [14, 21] and Table 2.11).

As an example that illustrates the benefit of transitive roles in queries, assume that we want to recognize patients who have previously had myocarditis, i.e., an inflammation of the heart muscle. This can be expressed using the TCQ

$$\text{Patient}(x) \wedge \bigcirc^- \diamond^- (\exists y, z. \text{partOf}(y, x) \wedge \text{Heart}(y) \wedge \text{partOf}(z, y) \wedge \text{Muscle}(z) \wedge \text{Inflamed}(z))$$

This query is looking for a patient that, at some past time point, had (as part) a heart that itself had as part a muscle that was inflamed. In this example, we assume that the role `partOf` is transitive and rigid. Transitivity implies that the inflamed muscle was also part of the patient and rigidity ensures that the heart is not part of different patients at different points in time. In addition, we assume that `Heart` and `Muscle` are rigid (hearts and muscles stay hearts and muscles over time), but `Patient` and `Inflamed` are non-rigid (the muscle may, e.g., cease to be inflamed and the patient may be discharged).

In the next section, we introduce the DLs investigated in this paper, as well as TCQs and their semantics. We also give an overview over the already known and the new complexity results (see Table 2.11). Section 3 investigates the complexity of answering certain atemporal queries in a fine-grained way. The reason is that, similar to [3], we split the task of answering TCQs into propositional temporal reasoning on the one hand, and answering atemporal queries on the other hand. In Section 4, we then determine the combined complexity of answering TCQs whereas in Section 5 we deal with the data complexity.

## 2 Preliminaries

In this section, we recall the basic notions of DLs and TCQs. Throughout the paper, let  $\mathbf{N}_C$ ,  $\mathbf{N}_R$ , and  $\mathbf{N}_I$  be non-empty, pairwise disjoint sets of *concept names*, *role names*, and *individual names*, respectively.

**Definition 2.1** (Syntax of DLs). *A role  $r$  is either a role name, i.e.,  $r \in \mathbf{N}_R$ , or an inverse role; see the first part of Table 2.2. The set of concepts is inductively defined starting from concept names  $A \in \mathbf{N}_C$  using the constructors in the second part of Table 2.2, where  $r, s$  are roles,  $a, b \in \mathbf{N}_I$ ,  $n \in \mathbb{N}$ , and  $C, D$  are concepts. The third part of Table 2.2 shows how axioms are defined.*

*Moreover, a TBox is a finite set of general concept inclusions (GCIs), an RBox is a finite set of role inclusions and transitivity axioms, and an ABox is a finite set*

Table 2.2: Syntax and Semantics of DLs

	syntax	semantics
inverse role	$r^-$	$\{(e, d) \mid (d, e) \in r^{\mathcal{I}}\}$
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{d \in \Delta^{\mathcal{I}} \mid \text{there is } e \in C^{\mathcal{I}} \text{ with } (d, e) \in r^{\mathcal{I}}\}$
nominal	$\{a\}$	$\{a^{\mathcal{I}}\}$
at-most restriction	$\leq n r.C$	$\{d \in \Delta^{\mathcal{I}} \mid \#\{e \in C^{\mathcal{I}} \mid (d, e) \in r^{\mathcal{I}}\} \leq n\}$
GCI	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
concept assertion	$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
role assertion	$r(a, b)$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$
role inclusion	$r \sqsubseteq s$	$r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$
transitivity axiom	$\text{trans}(r)$	$r^{\mathcal{I}} = (r^{\mathcal{I}})^+$

of assertions, i.e., concept assertions and role assertions. A knowledge base (KB)  $\mathcal{K} = (\mathcal{A}, \mathcal{T}, \mathcal{R})$  consists of an ABox  $\mathcal{A}$ , a TBox  $\mathcal{T}$ , and an RBox  $\mathcal{R}$ .

Which constructors and types of axioms are available depends on the specific DL. In  $\mathcal{ALC}$ , negation, conjunction, and existential restriction are the only allowed constructors. Also, no inverse roles, role inclusions and transitivity axioms are allowed in  $\mathcal{ALC}$ . If additional concept constructors or types of axioms are allowed, this is denoted by concatenating a corresponding letter:  $\mathcal{I}$  means inverse roles,  $\mathcal{O}$  means nominals,  $\mathcal{Q}$  means at-most restrictions, and  $\mathcal{H}$  means role inclusions. Thus, the DL  $\mathcal{ALCHI}$  is the extension of  $\mathcal{ALC}$  that also allows for nominals and inverse roles. The extension of  $\mathcal{ALC}$  with transitivity axioms is denoted by  $\mathcal{S}$ . Hence, the DL allowing for all the constructors and types of axioms introduced here is called  $\mathcal{SHOIQ}$ . We sometimes write  $\mathcal{L}$ -concept ( $\mathcal{L}$ -KB, ...) for some DL  $\mathcal{L}$  to make clear which DL is used.

The semantics of DLs are defined in a model-theoretic way.

**Definition 2.3** (Semantics of DLs). An interpretation is a pair  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , where  $\Delta^{\mathcal{I}}$  is a non-empty set (called domain), and  $\cdot^{\mathcal{I}}$  is a mapping assigning a set  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$  to every  $A \in \mathbf{N}_{\mathcal{C}}$ , a binary relation  $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  to every  $r \in \mathbf{N}_{\mathcal{R}}$ , and a domain element  $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$  to every  $a \in \mathbf{N}_{\mathcal{I}}$ , such that  $a^{\mathcal{I}} \neq b^{\mathcal{I}}$  for all  $a, b \in \mathbf{N}_{\mathcal{I}}$  with  $a \neq b$  (unique name assumption (UNA)). This function  $\cdot^{\mathcal{I}}$  is extended to roles and concepts inductively as shown in the first two parts of Table 2.2, where  $\#S$  denotes the cardinality of the set  $S$ .

Moreover,  $\mathcal{I}$  is a model of the axiom  $\alpha$  (written  $\mathcal{I} \models \alpha$ ) if the condition in the third part of Table 2.2 is satisfied, where  $\cdot^+$  denotes the transitive closure. Furthermore,  $\mathcal{I}$  is a model of a set of axioms  $\mathcal{X}$  (written  $\mathcal{I} \models \mathcal{X}$ ) if it is a model of all axioms  $\alpha \in \mathcal{X}$ , and  $\mathcal{I}$  is a model of a KB  $\mathcal{K} = (\mathcal{A}, \mathcal{T}, \mathcal{R})$  (written  $\mathcal{I} \models \mathcal{K}$ ) if it is a model

of  $\mathcal{A}$ ,  $\mathcal{T}$ , and  $\mathcal{R}$ . We call  $\mathcal{K}$  consistent if it has a model.

For an RBox  $\mathcal{R}$ , we call a role name  $r \in \mathbf{N}_R$  *transitive* (w.r.t.  $\mathcal{R}$ ) if every model of  $\mathcal{R}$  is a model of  $\text{trans}(r)$ . Moreover,  $r$  is a *subrole* of a role name  $s \in \mathbf{N}_R$  (w.r.t.  $\mathcal{R}$ ) if every model of  $\mathcal{R}$  is a model of  $r \sqsubseteq s$ . Finally,  $r$  is *simple w.r.t.  $\mathcal{R}$*  if it has no transitive subrole. Deciding whether  $r \in \mathbf{N}_R$  is simple can be done in time polynomial in the size of  $\mathcal{R}$  by a simple reachability test. Unfortunately, the problem of deciding whether a given  $\mathcal{SHQ}$ -KB  $\mathcal{K} = (\mathcal{A}, \mathcal{T}, \mathcal{R})$  is consistent is undecidable in general [19]. To regain decidability, we need to make the following syntactic restriction: if  $\leq n r.C$  occurs in  $\mathcal{K}$ , then  $r$  must be simple w.r.t.  $\mathcal{R}$ .

To enable a better separation of the influence the ABoxes have on the complexity of reasoning, we assume in the following that concept assertions use only concept *names* that also occur in the TBox, and that role assertions use only role names that occur in the TBox or the RBox. One can still simulate a complex concept assertion  $C(a)$  using  $A(a)$  and  $A \equiv C$ , where the latter stands for the two GCIs  $A \sqsubseteq C$  and  $C \sqsubseteq A$ .

Before we can define temporal queries, we need to lift the notions of knowledge bases and interpretations to a temporal setting. For this purpose, we assume that there are designated sets  $\mathbf{N}_{RC} \subseteq \mathbf{N}_C$  of *rigid concept names* and  $\mathbf{N}_{RR} \subseteq \mathbf{N}_R$  of *rigid role names*. Intuitively, the interpretation of a rigid name is not allowed to change over time. All individual names are implicitly assumed to be rigid. A concept or role name that is not rigid is called *flexible*.

**Definition 2.4** (TKB). *A tuple  $\mathcal{K} = ((\mathcal{A}_i)_{0 \leq i \leq n}, \mathcal{T}, \mathcal{R})$ , consisting of a finite sequence of ABoxes  $\mathcal{A}_i$ , a TBox  $\mathcal{T}$ , and an RBox  $\mathcal{R}$ , is called a temporal knowledge base (TKB). Let  $\mathfrak{I} = (\mathcal{I}_i)_{i \geq 0}$  be an infinite sequence of interpretations  $\mathcal{I}_i = (\Delta, \cdot^{\mathcal{I}_i})$  over a fixed domain  $\Delta$ . Then  $\mathfrak{I}$  is a model of  $\mathcal{K}$  (written  $\mathfrak{I} \models \mathcal{K}$ ) if*

- $\mathcal{I}_i \models \mathcal{A}_i$  for all  $i, 0 \leq i \leq n$ ,
- $\mathcal{I}_i \models \mathcal{T}$  and  $\mathcal{I}_i \models \mathcal{R}$  for all  $i \geq 0$ , and
- $\mathfrak{I}$  respects *rigid names*, i.e., we have  $x^{\mathcal{I}_i} = x^{\mathcal{I}_j}$  for all  $x \in \mathbf{N}_I \cup \mathbf{N}_{RC} \cup \mathbf{N}_{RR}$  and all time points  $i, j \geq 0$ .

We denote the set of all individual names occurring in a TKB  $\mathcal{K}$  by  $\text{Ind}(\mathcal{K})$ . The syntax of TCQs is defined by combining ordinary conjunctive queries via the operators of LTL [3, 26].

**Definition 2.5** (Syntax of TCQs). *Let  $\mathbf{N}_V$  be a set of variables. A conjunctive query (CQ) is of the form  $\exists y_1, \dots, y_m. \psi$ , where  $y_1, \dots, y_m \in \mathbf{N}_V$  and  $\psi$  is a finite conjunction of atoms of the form  $A(z_1)$  (concept atom),  $r(z_1, z_2)$  (role atom), or  $z_1 \approx z_2$  (equality atom), where  $A \in \mathbf{N}_C$ ,  $r \in \mathbf{N}_R$ , and  $z_1, z_2 \in \mathbf{N}_V \cup \mathbf{N}_I$ .*

Temporal conjunctive queries (TCQs) are built inductively from CQs, using the constructors  $\neg\phi_1$  (negation),  $\phi_1 \wedge \phi_2$  (conjunction),  $\bigcirc\phi_1$  (next),  $\bigcirc^-\phi_1$  (previous),  $\phi_1 \cup \phi_2$  (until), and  $\phi_1 \text{S} \phi_2$  (since), where  $\phi_1$  and  $\phi_2$  are TCQs.

Note that in contrast to [3], we allow also non-simple roles to occur in CQs. A union of conjunctive queries (UCQ) is a disjunction of CQs, which is defined as  $\phi_1 \vee \phi_2 := \neg(\neg\phi_1 \wedge \neg\phi_2)$ . Further, a CQ-literal is either a CQ or a negated CQ. We denote the set of individual names occurring in a TCQ  $\phi$  by  $\text{Ind}(\phi)$ , the set of variables occurring in  $\phi$  by  $\text{Var}(\phi)$ , the set of free variables of  $\phi$  by  $\text{FVar}(\phi)$ . A TCQ  $\phi$  with  $\text{FVar}(\phi) = \emptyset$  is called *Boolean*. As in [3], we assume without loss of generality that all CQs are connected, i.e., all variables and individual names are related (transitively) by roles.

**Definition 2.6** (Semantics of TCQs). *An interpretation  $\mathcal{I} = (\Delta, \cdot^{\mathcal{I}})$  is a model of a Boolean CQ  $\phi$  (written  $\mathcal{I} \models \phi$ ) if there is a homomorphism of  $\phi$  into  $\mathcal{I}$ , which is a mapping  $\pi: \text{Var}(\phi) \cup \text{Ind}(\phi) \rightarrow \Delta$  with*

- $\pi(a) = a^{\mathcal{I}}$  for all  $a \in \text{Ind}(\phi)$ ;
- $\pi(z) \in A^{\mathcal{I}}$  for all concept atoms  $A(z)$  in  $\phi$ ;
- $(\pi(z_1), \pi(z_2)) \in r^{\mathcal{I}}$  for all role atoms  $r(z_1, z_2)$  in  $\phi$ ; and
- $\pi(z_1) = \pi(z_2)$  for all equality atoms  $z_1 \approx z_2$  in  $\phi$ .

An infinite sequence of interpretations  $\mathfrak{I} = (\mathcal{I}_i)_{i \geq 0}$  over a common domain  $\Delta$  is a model of a Boolean TCQ  $\phi$  at time point  $i \geq 0$  iff  $\mathfrak{I}, i \models \phi$  holds, which is defined as follows:

$$\begin{array}{ll}
\mathfrak{I}, i \models \exists y_1, \dots, y_m. \psi & \text{iff } \mathcal{I}_i \models \exists y_1, \dots, y_m. \psi \\
\mathfrak{I}, i \models \neg\phi_1 & \text{iff } \mathfrak{I}, i \not\models \phi_1 \\
\mathfrak{I}, i \models \phi_1 \wedge \phi_2 & \text{iff } \mathfrak{I}, i \models \phi_1 \text{ and } \mathfrak{I}, i \models \phi_2 \\
\mathfrak{I}, i \models \bigcirc\phi_1 & \text{iff } \mathfrak{I}, i + 1 \models \phi_1 \\
\mathfrak{I}, i \models \bigcirc^-\phi_1 & \text{iff } i > 0 \text{ and } \mathfrak{I}, i - 1 \models \phi_1 \\
\mathfrak{I}, i \models \phi_1 \cup \phi_2 & \text{iff } \text{there is } k \geq i \text{ with } \mathfrak{I}, k \models \phi_2 \\
& \text{and } \mathfrak{I}, j \models \phi_1 \text{ for all } j, i \leq j < k \\
\mathfrak{I}, i \models \phi_1 \text{S} \phi_2 & \text{iff } \text{there is } k, 0 \leq k \leq i \text{ with } \mathfrak{I}, k \models \phi_2 \\
& \text{and } \mathfrak{I}, j \models \phi_1 \text{ for all } j, k < j \leq i
\end{array}$$

Given a TKB  $\mathcal{K} = ((\mathcal{A}_i)_{0 \leq i \leq n}, \mathcal{T}, \mathcal{R})$ , we say that  $\mathfrak{I}$  is a model of  $\phi$  w.r.t.  $\mathcal{K}$  if  $\mathfrak{I} \models \mathcal{K}$  and  $\mathfrak{I}, n \models \phi$ . We call  $\phi$  satisfiable w.r.t.  $\mathcal{K}$  if it has a model w.r.t.  $\mathcal{K}$ , and it is entailed by  $\mathcal{K}$  (written  $\mathcal{K} \models \phi$ ) if every model  $\mathfrak{I}$  of  $\mathcal{K}$  satisfies  $\mathfrak{I}, n \models \phi$ .

Given a TCQ  $\phi$ , a mapping  $\mathbf{a}: \text{FVar}(\phi) \rightarrow \text{Ind}(\mathcal{K})$  is a certain answer to  $\phi$  w.r.t.  $\mathcal{K}$  if  $\mathcal{K} \models \mathbf{a}(\phi)$ , where  $\mathbf{a}(\phi)$  is the Boolean TCQ obtained from  $\phi$  by replacing the free variables according to  $\mathbf{a}$ .

As usual [3], in the following we consider only the TCQ entailment problem, which can be used to compute all certain answers. For this purpose, we analyze the satisfiability problem, which has the same complexity as non-entailment. We examine both the *combined complexity* where the whole TKB and the TCQ are considered as the input, and the *data complexity*, where TBox, RBox, and TCQ are fixed, i.e., the complexity is measured only w.r.t. the sequence of ABoxes.

In the remainder of this section, we recall the basic approach from [3] to decide satisfiability by splitting it into two separate satisfiability problems, one for the temporal component and one for the DL component. In the following, let  $\mathcal{K} = ((\mathcal{A}_i)_{0 \leq i \leq n}, \mathcal{T}, \mathcal{R})$  be a TKB and  $\phi$  be a TCQ to be checked for satisfiability. The *propositional abstraction*  $\phi^p$  of  $\phi$  is the propositional LTL-formula obtained from  $\phi$  by replacing all CQs by propositional variables. We assume that  $\alpha_1, \dots, \alpha_m$  are the CQs occurring in  $\phi$ , and that each  $\alpha_i$  is replaced by the propositional variable  $p_i$ ,  $1 \leq i \leq m$ . We now consider a set  $\mathcal{S} \subseteq 2^{\{p_1, \dots, p_m\}}$ , which intuitively specifies the worlds that are allowed to occur in an LTL-structure satisfying  $\phi^p$ . However, guessing such an  $\mathcal{S}$  and then testing whether  $\phi^p$  has a model at time point  $n$  that uses only the worlds from  $\mathcal{S}$  is not sufficient for checking whether  $\phi$  has a model w.r.t.  $\mathcal{K}$ . We must also check whether  $\mathcal{S}$  can indeed be induced by some sequence of interpretations that is a model of  $\mathcal{K}$ . In the following, let  $\mathcal{S} = \{X_1, \dots, X_k\} \subseteq 2^{\{p_1, \dots, p_m\}}$ , and  $\iota: \{0, \dots, n\} \rightarrow \{1, \dots, k\}$  be a mapping that specifies a set  $X_{\iota(i)}$  for each of the ABoxes  $\mathcal{A}_i$ ,  $0 \leq i \leq n$ .

**Definition 2.7** (r-satisfiability). *The set  $\mathcal{S}$  is r-satisfiable w.r.t.  $\iota$  and  $\mathcal{K}$  if there exist interpretations  $\mathcal{J}_1, \dots, \mathcal{J}_k$  and  $\mathcal{I}_0, \dots, \mathcal{I}_n$  that share the same domain, respect rigid names,<sup>1</sup> are models of  $\mathcal{T}$  and  $\mathcal{R}$ , and additionally each  $\mathcal{J}_i$ ,  $1 \leq i \leq k$ , is a model of  $\chi_i := \bigwedge_{p_j \in X_i} \alpha_j \wedge \bigwedge_{p_j \notin X_i} \neg \alpha_j$ , and each  $\mathcal{I}_i$ ,  $0 \leq i \leq n$ , is a model of  $\mathcal{A}_i$  and  $\chi_{\iota(i)}$ .*

This notion formalizes the rigidity constraints that have to be satisfied by  $\mathcal{S}$ . The temporal part of the satisfiability problem is described next.

**Definition 2.8** (t-satisfiability). *The LTL-formula  $\phi^p$  is t-satisfiable w.r.t.  $\mathcal{S}$  and  $\iota$  if there exists an LTL-structure  $\mathfrak{J} = (w_i)_{i \geq 0}$  such that  $\mathfrak{J}, n \models \widehat{\phi}$ ,  $w_i \in \mathcal{S}$  for all  $i \geq 0$ , and  $w_i = X_{\iota(i)}$  for all  $i$ ,  $0 \leq i \leq n$ .*

These two checks together suffice to determine the satisfiability of  $\phi$  w.r.t.  $\mathcal{K}$ .

**Proposition 2.9.** *The TCQ  $\phi$  is satisfiable w.r.t. the TKB  $\mathcal{K}$  iff there are a set  $\mathcal{S} = \{X_1, \dots, X_k\} \subseteq 2^{\{p_1, \dots, p_m\}}$  and a mapping  $\iota: \{0, \dots, n\} \rightarrow \{1, \dots, k\}$  such that  $\mathcal{S}$  is r-satisfiable w.r.t.  $\iota$  and  $\mathcal{K}$ , and  $\phi^p$  is t-satisfiable w.r.t.  $\mathcal{S}$  and  $\iota$ .*

The original proof of this result in [3] considers only  $\mathcal{SHQ}$ , but is actually independent of any specific DL. The complexity of the t-satisfiability problem is obviously also DL-agnostic, and hence we can reuse another result from [3].

<sup>1</sup>This is defined as for infinite sequences of interpretations (see Definition 2.4).

Table 2.11: Summary of known and new complexity results for TCQ entailment, where contributions of this paper are highlighted in boldface. Settings: (i) no rigid names are allowed, (ii) only rigid concept names are allowed, and (iii) arbitrary rigid names are allowed.

	data complexity			combined complexity		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
<i>ALC</i> – <i>ALCHQ</i>	CO-NP	CO-NP	$\leq$ EXP	EXP	CO-NEXP	2-EXP
<i>ALCO</i> – <i>ALCHOQ/ALCHOI</i>	CO-NP	<b>co-NP</b>	<b><math>\leq</math>EXP</b>	$\geq$ CO-NEXP	?	<b>2-EXP</b>
<i>S</i> – <i>SQ</i>	CO-NP	<b>co-NP</b>	<b><math>\leq</math>EXP</b>	$\geq$ CO-NEXP	?	<b>2-EXP</b>
<i>SO</i> – <i>SOQ</i>	$\geq$ CO-NP	?	<b><math>\leq</math>EXP</b>	$\geq$ CO-NEXP	?	<b>2-EXP</b>
<i>SH/ALCI</i> – <i>SHIQ</i>	CO-NP	<b>co-NP</b>	<b><math>\leq</math>EXP</b>	2-EXP	2-EXP	<b>2-EXP</b>
<i>SHO</i> – <i>SHOQ/SHOI</i>	$\geq$ CO-NP	?	<b><math>\leq</math>EXP</b>	2-EXP	2-EXP	<b>2-EXP</b>
<i>ALCOIQ</i> – <i>ALCHOIQ</i>	$\geq$ CO-NP	?	<b>decidable</b>	$\geq$ CO-2-NEXP	?	<b>decidable</b>
<i>SOIQ</i> – <i>SHOIQ</i>	$\geq$ CO-NP	?	?	$\geq$ CO-2-NEXP	?	?

**Proposition 2.10.** *Deciding  $t$ -satisfiability of  $\phi^P$  w.r.t.  $\mathcal{S}$  and  $\iota$  can be done in EXP w.r.t. combined complexity, and in P w.r.t. data complexity.*

Table 2.11 gives an overview over all known complexity results for TCQ entailment. We distinguish the three cases that (i) no rigid names are allowed ( $\mathbf{N}_{RC} = \mathbf{N}_{RR} = \emptyset$ ); (ii) only rigid concept names are allowed, but no rigid role names ( $\mathbf{N}_{RR} = \emptyset$ ); and (iii) arbitrary rigid names are allowed. The first row of the table contains the known results for *ALC/ALCHQ* [3]<sup>2</sup>, and in this paper we derive the upper bounds for cases (ii) and (iii) marked in bold font. Unfortunately, we leave open the precise data complexity for case (iii), as was the case in [3]; we are only able to show an EXP upper bound in most logics, as opposed to the complexity of CO-NP in the atemporal setting.

A question mark indicates that the precise complexity is unknown even for the atemporal CQ entailment problem. For *SHOIQ*, it is not even known whether this problem is decidable, while for *ALCHOIQ* it is only known to be decidable, but no better upper bound has been found so far [25, 28]. The shown lower bounds follow from the complexity of satisfiability of *ALC*-LTL formulae [5] and the complexity of atemporal CQ entailment. More precisely, the latter problem is CO-NP-hard in data complexity already for *ALC* [29]. Under combined complexity, it is CO-NEXP-hard for *ALCO* [23] and *S* [14], 2-EXP-hard for *SH* [14] and *ALCI* [21], and CO-2-NEXP-hard for *ALCOIQ* [18].

### 3 Atemporal Queries in *SHIQ*, *SHOQ*, *SHOI*

Before we can present our complexity results about TCQ entailment, we analyze in more detail the case of a simple conjunction of CQ literals, which is an atemporal

<sup>2</sup>Actually, that paper considers *SHQ*, but restricts the roles in CQs to be simple.

query. In a nutshell, we reduce the satisfiability of such a conjunction to UCQ non-entailment and exploit existing algorithms for this atemporal problem. We consider here only the sublogics  $\mathcal{SHIQ}$ ,  $\mathcal{SHOQ}$ , and  $\mathcal{SHOI}$ , which enjoy (a variant of) the so-called *forest model property* [11], which means that every consistent KB formulated in one of these logics has a model that basically consists of several tree-shaped structures whose roots are arbitrarily interconnected by roles (disregarding role connections due to nominals or transitive roles).

To show the results in the following sections, however, we need to conduct a more fine-grained analysis of the complexity of the atemporal query entailment algorithms. The main insight is that, while UCQ entailment in  $\mathcal{SHIQ}$ ,  $\mathcal{SHOQ}$ , and  $\mathcal{SHOI}$  is in 2-EXP w.r.t. combined complexity, the *number* of CQs in the UCQ only has an exponential influence on the complexity of this decision problem. Likewise, for data complexity, assuming that the number of CQs in the UCQ is linear instead of constant usually has no influence on the complexity. Unfortunately, to the best of our knowledge, the precise data complexity of UCQ entailment is known only for  $\mathcal{SHIQ}$ ,  $\mathcal{ALCHOQ}$ , and  $\mathcal{ALCHOI}$ , while for  $\mathcal{SHOQ}$  and  $\mathcal{SHOI}$  it is still open [25].

In the following, we consider the *size* of a CQ  $\psi$  (written  $|\psi|$ ) to be the number of symbols in  $\psi$ , ignoring constant expressions like ‘(’ and ‘^’, considering each name and variable to be of size 1, and further ignoring the prefix  $\exists y_1, \dots, y_m$  since these variables also occur in the atoms of  $\psi$ . For example,  $\exists x, y. r(x, y) \wedge A(x)$  has size 5. We could also assume that each name or variable is represented by a binary string denoting its name, and hence of size logarithmic in the size of  $\psi$ , but this would not affect our complexity results. Similarly, the size of a knowledge base is computed by ignoring the concept constructors, and hence considers only the number of occurrences of names in the axioms.

**Lemma 3.1.** *Let  $\psi = \rho_1 \wedge \dots \wedge \rho_\ell \wedge \neg\sigma_1 \dots \wedge \neg\sigma_o$  be a Boolean conjunction of CQ-literals,  $\mathcal{K} = (\mathcal{A}, \mathcal{T}, \mathcal{R})$  be a KB formulated in  $\mathcal{SHIQ}$ ,  $\mathcal{SHOQ}$ , or  $\mathcal{SHOI}$ , and  $\|\psi\| := \max\{|\rho_1|, \dots, |\rho_\ell|, |\sigma_1|, \dots, |\sigma_o|\}$ . Then the satisfiability of  $\psi$  w.r.t.  $\mathcal{K}$  can be decided by a deterministic algorithm in time bounded by  $2^{p(\ell, o, |\mathcal{K}|)^{p'(\|\psi\|)}}$ , for two polynomials  $p$  and  $p'$ .*

*In the case of  $\mathcal{SHIQ}$ ,  $\mathcal{ALCHOQ}$ , or  $\mathcal{ALCHOI}$ , if  $\mathcal{T}$ ,  $\mathcal{R}$ , and  $\|\psi\|$  are fixed, then satisfiability of  $\psi$  w.r.t.  $\mathcal{K}$  can be decided by a nondeterministic algorithm in time bounded by  $p(\ell, o, |\mathcal{A}|)$  for some polynomial  $p$ .*

*Proof.* We start as in [3] by reducing the decision whether  $\psi$  has a model w.r.t.  $\mathcal{K}$  to a UCQ non-entailment problem. We instantiate the positive CQs  $\rho_1, \dots, \rho_\ell$  by omitting the existential quantifiers and replacing all variables by fresh individual names. The set  $\mathcal{A}_\rho$  of all resulting assertions can be viewed as an additional ABox. To ensure that the UNA is satisfied, we additionally consider equivalence relations  $\approx$  on  $\text{Ind}(\mathcal{A} \cup \mathcal{A}_\rho)$  with the additional restriction that no two names from  $\text{Ind}(\mathcal{A})$  may be equivalent. We denote by  $\mathcal{A}_\approx$  the ABox resulting from  $\mathcal{A}_\rho$

by replacing each new individual name by a fixed representative of its equivalence class, where this representative is an element of  $\text{Ind}(\mathcal{A})$  whenever possible. It can be shown as in [3] that  $\psi$  is satisfiable w.r.t.  $\mathcal{K}$  iff there is such an equivalence relation  $\approx$  for which

$$(\mathcal{A} \cup \mathcal{A}_{\approx}, \mathcal{T}, \mathcal{R}) \not\models \sigma_1 \vee \dots \vee \sigma_o. \quad (1)$$

Note that the number of equivalence relations  $\approx$  is exponential in the total number of variables in  $\rho_1, \dots, \rho_\ell$ , which is bounded by  $\ell \cdot \|\psi\|$ , but each is of size polynomial in  $\ell \cdot \|\psi\|$ . Hence, one can either enumerate all such equivalences in time exponential in  $\ell \cdot \|\psi\|$ , or guess one of them in time polynomial in  $\ell \cdot \|\psi\|$ .

We now consider the case that  $\mathcal{K}$  is formulated in  $\mathcal{SHIQ}$ . By [16, Lemma 23], for the non-entailment test (1), it suffices to find a so-called *extended knowledge base*  $\mathcal{K}' = (\mathcal{A} \cup \mathcal{A}_{\approx} \cup \mathcal{A}', \mathcal{T} \cup \mathcal{T}', \mathcal{R})$ , where  $\mathcal{A}'$  and  $\mathcal{T}'$  are formulated in  $\mathcal{SHIQ}^\square$ , i.e.,  $\mathcal{SHIQ}$  extended by role conjunctions, such that  $\mathcal{K}'$  is consistent. By [16, Lemma 20 and Definition 21], the size of each  $(\mathcal{A}', \mathcal{T}')$  is bounded by  $p(o \cdot (|\mathcal{K}| + \ell \cdot \|\psi\|))^{p(\|\psi\|)}$  for some polynomial  $p$ , where the term  $\ell \cdot \|\psi\|$  represents the size of the additional ABox  $\mathcal{A}_{\approx}$ . The bound given in [16] is exponential in the total size of the UCQ, i.e.,  $o \cdot \|\psi\|$ , but the exponential blowup comes only from the rewriting of each individual CQ  $\sigma_i$ . Moreover, all pairs  $(\mathcal{A}', \mathcal{T}')$  can be enumerated in time bounded by  $2^{p(o \cdot (|\mathcal{K}| + \ell \cdot \|\psi\|))^{p(\|\psi\|)}}$ . It is important to note that the size of the longest role conjunction occurring in  $(\mathcal{A}', \mathcal{T}')$  is bounded by a polynomial in  $\|\psi\|$ . Hence, by [16, Lemma 28], one can check the consistency of  $\mathcal{K}'$  in time bounded by  $2^{p'(o \cdot (|\mathcal{K}| + \ell \cdot \|\psi\|))^{p'(\|\psi\|)}}$  for some polynomial  $p'$ . This means that we can decide satisfiability of  $\psi$  w.r.t.  $\mathcal{K}$  by enumerating all equivalence relations and extended KBs as above and testing each of them for consistency within the claimed time bound.

If  $\mathcal{T}$ ,  $\mathcal{R}$ , and  $\|\psi\|$  are fixed, then one can guess  $\approx$  in time polynomial in  $\ell$ . Following the proof of [16, Theorem 35], one can also guess  $\mathcal{K}'$  in time  $p(o \cdot (|\mathcal{A}| + \ell))$ , and the following consistency test can be done in (deterministic) polynomial time in the size of the ABox  $\mathcal{A} \cup \mathcal{A}_{\approx} \cup \mathcal{A}'$ , which is polynomial in  $o \cdot (|\mathcal{A}| + \ell)$ . This establishes the second bound for the case of  $\mathcal{SHIQ}$ .

In the case that  $\mathcal{K}$  is formulated in  $\mathcal{SHOQ}$ , it again suffices to find a certain *extended knowledge base*  $\mathcal{K}' = (\mathcal{A} \cup \mathcal{A}_{\approx}, \mathcal{T} \cup \mathcal{T}', \mathcal{R})$  that is consistent in order to decide (1) [17]. The additional TBox  $\mathcal{T}'$  is formulated in  $\mathcal{SHOQ}^\square$ , and the length of the longest role conjunction in  $\mathcal{T}'$  is bounded polynomially in  $\|\psi\|$ . Actually, in [17] the ABox is assumed to be internalized into the TBox, but we keep it separate here to keep track of the size of  $\mathcal{A}_{\approx}$ , which is  $\ell \cdot \|\psi\|$ . The size of each extended knowledge base is again bounded by  $p(o \cdot (|\mathcal{K}| + \ell \cdot \|\psi\|))^{p(\|\psi\|)}$ , while their number is at most  $2^{p(o \cdot (|\mathcal{K}| + \ell \cdot \|\psi\|))^{p(\|\psi\|)}}$ . The bounds implied by [17, Lemma 7] are actually exponential (resp. double-exponential) in  $o \cdot \|\psi\|$ , but again the construction only involves a simple union of sets of queries, which are constructed for each CQ  $\sigma_i$  individually. By [17, Lemma 12], the consistency of  $\mathcal{K}'$  can be reduced to the consistency of a

knowledge base  $\mathcal{K}''$  formulated in  $\mathcal{ALCHOQ}^\square$ , the size of which is still bounded by  $p'(o \cdot (|\mathcal{K}| + \ell \cdot \|\psi\|))^{p'(\|\psi\|)}$  for a polynomial  $p'$ . By [20, Theorem 3.33],<sup>3</sup> the latter test can be done in time bounded by  $2^{p''(o \cdot (|\mathcal{K}| + \ell \cdot \|\psi\|))^{p''(\|\psi\|)}}$ , and hence we obtain the claimed result.<sup>4</sup>

For  $\mathcal{SHOI}$ , we analyze the automata-based construction from [11, 12]. There, the KB  $(\mathcal{A} \cup \mathcal{A}_{\approx}, \mathcal{T}, \mathcal{R})$  is first reduced to a polynomial-size  $\mathcal{ALCOITb}_{reg}$  KB  $\mathcal{K}'$  by replacing all roles by regular role expressions simulating all (transitive) sub-roles. Then, a polynomial-sized *fully enriched automaton* (an extended two-way alternating tree automaton with parity acceptance condition (2ATA) [8])  $\mathbf{A}_{\mathcal{K}'}$  is constructed that accepts all so-called *interpretation trees* that correspond to models of  $\mathcal{K}'$ . This can be transformed into an equivalent (one-way) nondeterministic parity tree automaton (1NTA)  $\mathbf{A}'_{\mathcal{K}'}$  with a number of states bounded by  $2^{p(|\mathcal{K}| + \ell \cdot \|\psi\|)}$  and constant index (size of the parity condition) [11, 30]. Note that the size of the input alphabet of all these automata is also exponential in  $|\mathcal{K}| + \ell \cdot \|\psi\|$ . In the next step, a 1NTA  $\mathbf{A}_{\neg\sigma_i}$  is constructed that accepts an interpretation tree iff it does not contain a match for  $\sigma_i$ . The construction in [11, 12] can be adapted such that the number of states of  $\mathbf{A}_{\neg\sigma_i}$  is double exponential in the size of  $\sigma_i$ , i.e.,  $\|\psi\|$ , but does not depend on  $|\mathcal{K}|$  or  $\ell$ , while its index is exponential in  $\|\psi\|$ .<sup>5</sup> Finally, an automaton  $\mathbf{A}_{\mathcal{K}' \not\models \sigma_1 \vee \dots \vee \sigma_o}$  can be constructed that recognizes the intersection of the languages accepted by  $\mathbf{A}'_{\mathcal{K}'}, \mathbf{A}_{\neg\sigma_1}, \dots, \mathbf{A}_{\neg\sigma_o}$ , and hence accepts all interpretation trees that witness the non-entailment (1). This automaton has at most  $2^{p'(|\mathcal{K}| + \ell \cdot \|\psi\|)^{p'(\|\psi\|)}}$  states and its index is bounded by  $2^{p'(\|\psi\|)}$  for some polynomial  $p'$ . The claim now follows from the fact that (non)emptiness of 1NTAs can be decided in time polynomial in the number of states and input symbols and exponential in the index [15].

Finally, we consider the logics  $\mathcal{ALCHOQ}$  and  $\mathcal{ALCHOI}$  and assume that  $\mathcal{T}, \mathcal{R}$ , and  $\|\psi\|$  are fixed. The paper [24] describes a nondeterministic tableaux algorithm for  $\mathcal{SHIQ}, \mathcal{SHOQ}$ , and  $\mathcal{SHOI}$ , where queries are restricted to simple roles. The main approach to decide non-entailment is to nondeterministically build a so-called *completion graph*, and then check whether it is not possible to map the variables of the input query to the nodes of this graph in order to satisfy the query. By [24, Lemma 4.2, Proposition 4.4, and Footnote 12], the size of such a completion graph is polynomial in (i) the number of individual names occurring in the input KB and triple exponential in (ii) the number of role atoms in the

<sup>3</sup>Note that this Theorem applies to  $\mathcal{SHOQ}^\square$ , but only if all roles in role conjunctions are simple.

<sup>4</sup>We cannot directly use [17, Theorem 20] here since that only applies for unary encoding of the numbers in number restrictions.

<sup>5</sup>This construction depends on  $\mathcal{K}$  and  $\ell$  only because (i) the language of  $\mathbf{A}_{\neg\sigma_i}$  explicitly distinguishes interpretation trees, and (ii) valid match positions are required by  $\mathbf{A}_{\neg\sigma_i}$  to be reachable from the root. But (i) is unnecessary due to the intersection with  $\mathbf{A}'_{\mathcal{K}'}$ , and (ii) is subsumed by the actual condition on matches for all role atoms of  $\sigma_i$ , which requires them to be connected to the root by the universal role.

largest disjunct of the input query. Under our complexity assumptions, (i) is linear in  $|\mathcal{A}|$  and  $\ell$  and (ii) is constant. Since we can assume that the Boolean CQs  $\sigma_1, \dots, \sigma_o$  do not share variables, it then suffices to check for each CQ  $\sigma_i$  individually whether it is not possible to map its variables into the completion graph to satisfy its atoms. This can be checked by a deterministic algorithm in time polynomial in the number  $o$  of CQs and the size of the completion graph [24]. Hence, we obtain the claimed results.  $\square$

## 4 Combined Complexity of TCQ Entailment

Let  $\mathcal{K} = ((\mathcal{A}_i)_{0 \leq i \leq n}, \mathcal{T}, \mathcal{R})$  be a TKB,  $\phi$  be a TCQ, and assume for now that a set  $\mathcal{S} = \{X_1, \dots, X_k\} \subseteq 2^{\{p_1, \dots, p_m\}}$  and a mapping  $\iota: \{0, \dots, n\} \rightarrow \{1, \dots, k\}$  are given. For our complexity results, we employ the *copying* technique from [3, 5] for deciding whether  $\mathcal{S}$  is r-satisfiable w.r.t.  $\iota$  and  $\mathcal{K}$ . The idea is to introduce enough copies of all flexible names in order to combine the separate satisfiability tests of Def. 2.7 into one big atemporal satisfiability test.

Formally, for all  $i$ ,  $1 \leq i \leq k + n + 1$ , and every *flexible* concept name  $A$  (every *flexible* role name  $r$ ) occurring in  $\mathcal{T}$  or  $\mathcal{R}$ , we introduce a copy  $A^{(i)}$  ( $r^{(i)}$ ). We call  $A^{(i)}$  ( $r^{(i)}$ ) the  $i$ -th copy of  $A$  ( $r$ ). The conjunctive query  $\alpha^{(i)}$  (the axiom  $\beta^{(i)}$ ) is obtained from a CQ  $\alpha$  (an axiom  $\beta$ ) by replacing every flexible name by its  $i$ -th copy. Similarly, for  $1 \leq \ell \leq k$ , the conjunction of CQ-literals  $\chi_\ell^{(i)}$  is obtained from  $\chi_\ell$  (see Definition 2.7) by replacing each CQ  $\alpha_j$  by  $\alpha_j^{(i)}$ . Finally, we define

$$\begin{aligned} \chi_{\mathcal{S}, \iota} &:= \bigwedge_{1 \leq i \leq k} \chi_i^{(i)} \wedge \bigwedge_{0 \leq i \leq n} \left( \chi_{\iota(i)}^{(k+i+1)} \wedge \bigwedge_{\alpha \in \mathcal{A}_i} \alpha^{(k+i+1)} \right), \\ \mathcal{T}_{\mathcal{S}, \iota} &:= \{\beta^{(i)} \mid \beta \in \mathcal{T} \text{ and } 1 \leq i \leq k + n + 1\}, \\ \mathcal{R}_{\mathcal{S}, \iota} &:= \{\gamma^{(i)} \mid \gamma \in \mathcal{R} \text{ and } 1 \leq i \leq k + n + 1\}. \end{aligned}$$

The following result, which reduces r-satisfiability to an atemporal satisfiability problem, was shown in [3] for  $\mathcal{SHQ}$  with simple roles in queries, but it remains valid in our setting since it does not depend on the DL under consideration.

**Proposition 4.1.** *The set  $\mathcal{S}$  is r-satisfiable w.r.t.  $\iota$  and  $\mathcal{K}$  iff  $\chi_{\mathcal{S}, \iota}$  is satisfiable w.r.t.  $(\mathcal{T}_{\mathcal{S}, \iota}, \mathcal{R}_{\mathcal{S}, \iota})$ .*

Together with Lemma 3.1, this allows us to show our first complexity results.

**Theorem 4.2.** *Let  $\mathcal{L}$  be a DL that contains  $\mathcal{ALCI}$  or  $\mathcal{SH}$  and is contained in  $\mathcal{SHIQ}$ ,  $\mathcal{SHOQ}$ , or  $\mathcal{SHOI}$ . Then TCQ entailment in  $\mathcal{L}$  is 2-EXP-complete w.r.t. combined complexity, and in EXP w.r.t. data complexity.*

*Proof.* The lower bound directly follows from 2-EXP-hardness of CQ entailment in  $\mathcal{SH}$  [14] and  $\mathcal{ALCI}$  [21]. To check a TCQ  $\phi$  for satisfiability w.r.t. a TKB  $\mathcal{K}$ , we first enumerate all possible sets  $\mathcal{S}$  and mappings  $\iota$ , which can be done in 2-EXP. For each of these double exponentially many pairs  $(\mathcal{S}, \iota)$ , we then check t-satisfiability of  $\phi^p$  w.r.t.  $\mathcal{S}$  and  $\iota$  in exponential time (see Proposition 2.10) and test  $\mathcal{S}$  for r-satisfiability w.r.t.  $\iota$  and  $\mathcal{K}$  (using Proposition 4.1). By Proposition 2.9,  $\phi$  has a model w.r.t.  $\mathcal{K}$  iff at least one pair passes both tests.

For the r-satisfiability test, observe that the conjunction of CQ-literals  $\chi_{\mathcal{S}, \iota}$  contains exponentially many (negated) CQs, each of size polynomial in the size of  $\phi$ , and that  $\mathcal{T}_{\mathcal{S}, \iota}$  and  $\mathcal{R}_{\mathcal{S}, \iota}$  are of exponential size in the size of  $\mathcal{K}$ . By Lemma 3.1, the satisfiability of  $\chi_{\mathcal{S}, \iota}$  w.r.t.  $(\mathcal{T}_{\mathcal{S}, \iota}, \mathcal{R}_{\mathcal{S}, \iota})$  can thus be checked in double exponential time in the size of  $\phi$  and  $\mathcal{K}$ . For the data complexity, observe that the number of CQs in  $\chi_{\mathcal{S}, \iota}$  is linear in the size of the input ABoxes, and their size only depends on  $\phi$  (the size of a single assertion is constant). Moreover,  $\mathcal{T}_{\mathcal{S}, \iota}$  and  $\mathcal{R}_{\mathcal{S}, \iota}$  are of size linear in  $n$ . Lemma 3.1 thus yields the claimed upper bound.  $\square$

By the same arguments, it is easy to see that TCQ entailment in  $\mathcal{ALCHOIQ}$  is decidable since this is the case for UCQ (non-)entailment [28].

## 5 Data Complexity without Rigid Roles

To obtain a tight bound on the data complexity if we disallow rigid role names, we follow a different approach from [3]. Similarly to the previous section, we decide r-satisfiability of  $\mathcal{S}$  w.r.t.  $\iota$  and  $\mathcal{K}$  by constructing conjunctions of CQ-literals of which we want to check satisfiability. However, we do not compile the whole r-satisfiability check into just one conjunction. More precisely, we define the conjunctions of CQ-literals  $\gamma_i \wedge \chi_{\mathcal{S}}$ ,  $0 \leq i \leq n$ , w.r.t.  $(\mathcal{T}_{\mathcal{S}}, \mathcal{R}_{\mathcal{S}})$ , where

$$\begin{aligned} \gamma_i &:= \bigwedge_{\alpha \in \mathcal{A}_i} \alpha^{(\iota(i))}, & \chi_{\mathcal{S}} &:= \bigwedge_{1 \leq i \leq k} \chi_i^{(i)}, \\ \mathcal{T}_{\mathcal{S}} &:= \{\beta^{(i)} \mid \beta \in \mathcal{T} \text{ and } 1 \leq i \leq k\}, \\ \mathcal{R}_{\mathcal{S}} &:= \{\gamma^{(i)} \mid \gamma \in \mathcal{R} \text{ and } 1 \leq i \leq k\}. \end{aligned}$$

This separates the consistency checks for the individual ABoxes  $\mathcal{A}_i$ ,  $1 \leq i \leq n$ , from each other. For r-satisfiability, we additionally have to make sure that rigid consequences of the form  $A(a)$  for a rigid concept name  $A \in \mathbf{N}_{\text{RC}}$  and an individual name  $a \in \mathbf{N}_I$  are shared between all the conjunctions  $\gamma_i \wedge \chi_{\mathcal{S}}$ . It suffices to do this for the set  $\text{RCon}(\mathcal{T})$  of rigid concept names occurring in  $\mathcal{T}$  since those that occur only in ABox assertions cannot affect the entailment of the TCQ  $\phi$ .

For this purpose, we guess a set  $\mathcal{D} \subseteq 2^{\text{RCon}(\mathcal{T})}$  that fixes the combinations of rigid concept names that are allowed to occur in the models of  $\gamma_i \wedge \chi_{\mathcal{S}}$ , and a

function  $\tau: \text{Ind}(\phi) \cup \text{Ind}(\mathcal{K}) \rightarrow \mathcal{D}$  that assigns to each individual name one such combination. To express this formally, we extend the TBox by the axioms in

$$\mathcal{T}_{\mathcal{D}} := \{A_Y \equiv C_Y \mid Y \in \mathcal{D}\},$$

where  $A_Y$  are fresh rigid concept names and, for every  $Y \subseteq \text{RCon}(\mathcal{T})$ ,

$$C_Y := \prod_{A \in Y} A \sqcap \prod_{A \in \text{RCon}(\mathcal{T}) \setminus Y} \neg A.$$

The size of  $\mathcal{T}_{\mathcal{D}}$  is bounded polynomially in the sizes of  $\mathcal{D}$  and  $\text{RCon}(\mathcal{T})$ , which are constant w.r.t. data complexity. We now extend the conjunctions  $\gamma_i \wedge \chi_{\mathcal{S}}$  by

$$\rho_{\tau} := \bigwedge_{a \in \text{Ind}(\phi) \cup \text{Ind}(\mathcal{K})} A_{\tau(a)}(a)$$

in order to fix the behavior of the rigid concept names on the named individuals.

We need one more definition to formulate the main lemma of this section. We say that an interpretation  $\mathcal{I}$  *respects*  $\mathcal{D}$  if

$$\mathcal{D} = \{Y \subseteq \text{RCon}(\mathcal{T}) \mid \text{there is a } d \in \Delta^{\mathcal{I}} \text{ with } d \in (C_Y)^{\mathcal{I}}\},$$

which means that every combination of rigid concept names in  $\mathcal{D}$  is realized by a domain element of  $\mathcal{I}$ , and conversely, the domain elements of  $\mathcal{I}$  may only realize those combinations that occur in  $\mathcal{D}$ .

**Lemma 5.1.** *Let the DL  $\mathcal{L}$  be contained in  $\mathcal{SHIQ}$ ,  $\mathcal{ALCHOQ}$ , or  $\mathcal{ALCHOI}$ . If  $\mathbf{N}_{\text{RR}} = \emptyset$ , then  $\mathcal{S}$  is  $r$ -satisfiable w.r.t.  $\iota$  and  $\mathcal{K}$  iff there exist  $\mathcal{D} \subseteq 2^{\text{RCon}(\mathcal{T})}$  and  $\tau: \text{Ind}(\phi) \cup \text{Ind}(\mathcal{K}) \rightarrow \mathcal{D}$  such that each  $\gamma_i \wedge \chi_{\mathcal{S}} \wedge \rho_{\tau}$ ,  $0 \leq i \leq n$ , has a model w.r.t.  $(\mathcal{T}_{\mathcal{S}} \cup \mathcal{T}_{\mathcal{D}}, \mathcal{R}_{\mathcal{S}})$  that respects  $\mathcal{D}$ .*

*Proof.* For the “if” direction, assume that  $\mathcal{I}_i$ ,  $0 \leq i \leq n$ , are the required models for  $\gamma_i \wedge \chi_{\mathcal{S}} \wedge \rho_{\tau}$  w.r.t.  $(\mathcal{T}_{\mathcal{S}} \cup \mathcal{T}_{\mathcal{D}}, \mathcal{R}_{\mathcal{S}})$ . We can assume w.l.o.g. that their domains  $\Delta_i$  are countably infinite and for each  $Y \in \mathcal{D}$  there are countably infinitely many elements  $d \in (C_Y)^{\mathcal{I}_i}$ . For  $\mathcal{SHIQ}$ , this is a consequence of the Löwenheim-Skolem theorem and the fact that the countably infinite disjoint union of  $\mathcal{I}_i$  with itself is again a model of  $\gamma_i \wedge \chi_{\mathcal{S}} \wedge \rho_{\tau}$  and  $(\mathcal{T}_{\mathcal{S}} \cup \mathcal{T}_{\mathcal{D}}, \mathcal{R}_{\mathcal{S}})$ . The latter follows from the observation that for any CQ there is a homomorphism into  $\mathcal{I}_i$  iff there is a homomorphism into the disjoint union of  $\mathcal{I}_i$  with itself. One direction is trivial, while whenever there is a homomorphism into the disjoint union, we can construct a homomorphism into  $\mathcal{I}_i$  by replacing the elements in the image of this homomorphism by the corresponding elements of  $\Delta_i$ . It is easy to see that the resulting homomorphism still satisfies all atoms of the CQ.

Due to the presence of nominals, this construction does not work in the sublogics  $\mathcal{ALCHOQ}$  and  $\mathcal{ALCHOI}$  since domain elements may need to be connected to

specific named domain elements, of which there cannot be infinitely many copies, but only one. We adapt the construction by requiring that the countably many copies of  $\mathcal{I}_i$  overlap on the set  $\{a^{\mathcal{I}_i} \mid a \in \mathbf{N}_1\}$ . Hence, there may be infinitely many copies of domain elements connected to each named domain element. In *ALCHOI*, this already suffices since the behavior of the infinitely many unnamed copies is the same, which does not affect the satisfaction of any concepts or query atoms on the named elements.

However, in *ALCHOQ* this may cause number restrictions to be violated. For example,  $a$  could be required to have at most 10  $r$ -successors, whereas after our construction it may have infinitely many. In this case, we can without loss of generality make the further assumption that  $\mathcal{I}_i$  is a *quasi-forest model* of  $\gamma_i \wedge \chi_S \wedge \rho_\tau$  and  $(\mathcal{T}_S \cup \mathcal{T}_D, \mathcal{R}_S)$  [11, 20].<sup>6</sup> This means that each named individual acts as the root for a *tree* of domain elements, and there can only be role connections between elements  $x$  and  $y$  if either  $y$  is a child node of  $x$  in such a tree or  $y$  is a named individual, i.e.,  $y = a^{\mathcal{I}_i}$  for some  $a \in \mathbf{N}_1$ . Since *ALCHOQ* does not contain inverse roles, it is not possible for a concept to refer upwards in a tree, but only directly back to a root. We can thus adapt the construction by copying the unnamed part of  $\mathcal{I}_i$  countably often, and removing all role connections of the form  $(a^{\mathcal{I}_i}, x)$ , where  $x$  is an unnamed domain element, except for the connections in the original  $\mathcal{I}_i$ . Note that we still copy the role connections back to the named elements, as these are necessary to satisfy nominal concepts. This construction ensures that all number restrictions remain satisfied. Due to the absence of inverse roles, the KB  $(\mathcal{T}_S \cup \mathcal{T}_D, \mathcal{R}_S)$  is still satisfied. The conjunction  $\gamma_i \wedge \chi_S \wedge \rho_\tau$  also remains satisfied by the same arguments as for *SHIQ* and the fact that all CQs are connected.

Having ensured that the domain of  $\mathcal{I}_i$  is countably infinite and for each  $Y \in \mathcal{D}$  there are countably infinitely many elements  $d \in (C_Y)^{\mathcal{I}_i}$ , we can partition the domains  $\Delta_i$  into the countably infinite sets  $\Delta_i(Y) := \{d \in \Delta_i \mid d \in (C_Y)^{\mathcal{I}_i}\}$  for each  $Y \in \mathcal{D}$ . By the assumptions above and the fact that all  $\mathcal{I}_i$  satisfy  $\rho_\tau$  and  $\mathcal{T}_D$ , there are bijections  $\pi_i: \Delta_0 \rightarrow \Delta_i$ ,  $1 \leq i \leq n$ , such that

- $\pi_i(\Delta_0(Y)) = \Delta_i(Y)$  for all  $Y \in \mathcal{D}$  and
- $\pi_i(a^{\mathcal{I}_0}) = a^{\mathcal{I}_i}$  for all  $a \in \text{Ind}(\phi) \cup \text{Ind}(\mathcal{K})$ .

Thus, we can assume that the models  $\mathcal{I}_i$  actually share the same domain and interpret the rigid names in  $\text{RCon}(\mathcal{T})$  and  $\text{Ind}(\phi) \cup \text{Ind}(\mathcal{K})$  in the same way.

The remainder of the proof can proceed exactly as in [3]. □

---

<sup>6</sup>To apply the result from [11], we can instantiate the positive CQs in  $\gamma_i \wedge \chi_S \wedge \rho_\tau$  and treat the remaining conjunction of negative CQs as a UCQ for which we want to prove non-entailment, as in the proof of Lemma 3.1.

The restriction imposed by  $\mathcal{D}$  can be expressed as the conjunction of CQ-literals

$$\sigma_{\mathcal{D}} := (\neg\exists x.A_{\mathcal{D}}(x)) \wedge \bigwedge_{Y \in \mathcal{D}} \exists x.A_Y(x),$$

where  $A_{\mathcal{D}}$  is a fresh concept name that is defined by the axiom  $A_{\mathcal{D}} \equiv \prod_{Y \in \mathcal{D}} \neg A_Y$  in the TBox. We denote by  $\mathcal{T}'_{\mathcal{S}}$  the resulting extension of  $\mathcal{T}_{\mathcal{S}} \cup \mathcal{T}_{\mathcal{D}}$ , and have now reduced the r-satisfiability of  $\mathcal{S}$  w.r.t.  $\iota$  and  $\mathcal{K}$  to the consistency of  $\gamma_i \wedge \chi_{\mathcal{S}} \wedge \rho_{\tau} \wedge \sigma_{\mathcal{D}}$  w.r.t.  $(\mathcal{T}'_{\mathcal{S}}, \mathcal{R}_{\mathcal{S}})$ .

**Theorem 5.2.** *Let  $\mathcal{L}$  be a DL that contains  $\mathcal{AL}\mathcal{E}$  and is contained in  $\mathcal{SHIQ}$ ,  $\mathcal{ALCHOQ}$ , or  $\mathcal{ALCHOI}$ . Then TCQ entailment in  $\mathcal{L}$  is CO-NP-complete w.r.t. data complexity.*

*Proof.* The lower bound follows from CO-NP-hardness of instance checking in  $\mathcal{AL}\mathcal{E}$  [29]. To test satisfiability of a TCQ  $\phi$  w.r.t. a TKB  $\mathcal{K}$ , we employ the same approach as before, but instead guess  $\mathcal{S}$  and  $\iota$ . Note that  $\mathcal{S}$  is of constant size in the size of the ABoxes and  $\iota$  is of linear size. Hence, this first step can be done in nondeterministic polynomial time. The t-satisfiability test required for Proposition 2.9 can be done in polynomial time by Proposition 2.10, and for the r-satisfiability test, we use Lemma 5.1.

Following the reduction described above, we guess a set  $\mathcal{D} \subseteq 2^{\text{RCon}(\mathcal{T})}$  and a function  $\tau: \text{Ind}(\phi) \cup \text{Ind}(\mathcal{K}) \rightarrow \mathcal{D}$ , which can be done in nondeterministic polynomial time since  $\mathcal{D}$  only depends on  $\mathcal{T}$  and  $\tau$  is of size linear in the size of the input ABoxes. Next, we check the satisfiability of the polynomially many conjunctions  $\gamma_i \wedge \chi_{\mathcal{S}} \wedge \rho_{\tau} \wedge \sigma_{\mathcal{D}}$  w.r.t.  $(\mathcal{T}'_{\mathcal{S}}, \mathcal{R}_{\mathcal{S}})$ . Note that  $\chi_{\mathcal{S}}$ ,  $\sigma_{\mathcal{D}}$ ,  $\mathcal{T}'_{\mathcal{S}}$ , and  $\mathcal{R}_{\mathcal{S}}$  do not depend on the input ABoxes, while  $\gamma_i$  and  $\rho_{\tau}$  are of polynomial size. Furthermore, the size of the CQs in  $\gamma_i$  and  $\rho_{\tau}$  is constant. Hence, Lemma 3.1 yields the desired NP upper bound for these satisfiability tests.  $\square$

## 6 Conclusions

Query answering w.r.t. DL ontologies is currently a very active research area. We have extended complexity results for very expressive DLs underlying the web ontology language OWL2 to the case of temporal queries. Our results show that, w.r.t. worst-case complexity, adding a temporal dimension often comes for free. In fact, in all sublogics of  $\mathcal{SHOIQ}$ , the upper bounds for the combined complexity of TCQ entailment obtained in this paper for the temporal case coincide with the best known upper bounds for atemporal query entailment (even in the presence of rigid roles). From the application point of view, data complexity is more important since the amount of data is often very large, and in comparison the size of the background knowledge and the user query is small. We have shown that, in many cases, the atemporal data complexity of CO-NP does not increase if we consider

TCQs with rigid concepts (specifically, in  $\mathcal{ALCHOQ}$ ,  $\mathcal{ALCHOI}$ ,  $\mathcal{SHIQ}$ , and sublogics). For the remaining logics of Table 2.11, it is an open problem to find a CO-NP algorithm even in the atemporal case.

As part of future work, we will try to obtain CO-NP upper bounds even in the presence of rigid roles, and study extensions of TCQs with concrete domains and inconsistency-tolerant semantics. Since CO-NP is already a rather negative result for data complexity, we could also try to find restricted formalisms with lower data complexity. On the one hand, one could take a less expressive DL to formulate the background ontology, which has already been investigated for  $\mathcal{EL}$  [10] and  $DL-Lite_{horn}^{\mathcal{H}}$  [9], but only the latter choice reduces the data complexity (to ALOGTIME). On the other hand, one could investigate whether the data complexity can be reduced by imposing additional restrictions on the TBox or CQs, as has been done in the atemporal case [6, 7, 22].

## References

- [1] Artale, A., Kontchakov, R., Wolter, F., Zakharyashev, M.: Temporal description logic for ontology-based data access. In: Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI'13). pp. 711–717 (2013)
- [2] Baader, F.: Ontology-based monitoring of dynamic systems. In: Proc. of the 14th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'14). pp. 678–681 (2014)
- [3] Baader, F., Borgwardt, S., Lippmann, M.: Temporal query entailment in the description logic  $\mathcal{SHQ}$ . Journal of Web Semantics 33, 71–93 (2015)
- [4] Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. 2 edn. (2007)
- [5] Baader, F., Ghilardi, S., Lutz, C.: LTL over description logic axioms. ACM Transactions on Computational Logic 13(3), 21:1–21:32 (2012)
- [6] Bienvenu, M., Ortiz, M., Šimkus, M., Xiao, G.: Tractable queries for lightweight description logics. In: Rossi, F. (ed.) Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI'13). pp. 768–774 (2013)
- [7] Bienvenu, M., ten Cate, B., Lutz, C., Wolter, F.: Ontology-based data access: A study through disjunction Datalog, CSP, and MMSNP. ACM Transactions on Database Systems 39(4), 33:1–33:44 (2014)
- [8] Bonatti, P.A., Lutz, C., Murano, A., Vardi, M.Y.: The complexity of enriched  $\mu$ -calculi. Logical Methods in Computer Science 4(3) (2008)

- [9] Borgwardt, S., Thost, V.: Temporal query answering in *DL-Lite* with negation. In: Proc. of the 1st Global Conf. on Artificial Intelligence (GCAI'15) (2015), to appear.
- [10] Borgwardt, S., Thost, V.: Temporal query answering in the description logic  $\mathcal{EL}$ . In: Proc. of the 24th Int. Joint Conf. on Artificial Intelligence (IJCAI'15). pp. 2819–2825 (2015)
- [11] Calvanese, D., Eiter, T., Ortiz, M.: Regular path queries in expressive description logics with nominals. In: Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI'09). pp. 714–720 (2009)
- [12] Calvanese, D., Eiter, T., Ortiz, M.: Answering regular path queries in expressive description logics via alternating tree-automata. *Inf. Comput.* 237, 12–55 (2014)
- [13] Decker, S., Erdmann, M., Fensel, D., Studer, R.: Ontobroker: Ontology based access to distributed and semi-structured information. In: Proc. DS-8. pp. 351–369 (1999)
- [14] Eiter, T., Lutz, C., Ortiz, M., Šimkus, M.: Query answering in description logics with transitive roles. In: Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI'09). pp. 759–764 (2009)
- [15] Emerson, E.A., Jutla, C.S.: The complexity of tree automata and logics of programs. In: Proc. of the 29th Annual Symp. on Foundations of Computer Science (SFCS'88). pp. 328–337 (1988)
- [16] Glimm, B., Horrocks, I., Lutz, C., Sattler, U.: Conjunctive query answering for the description logic *SHIQ*. *Journal of Artificial Intelligence Research* 31(1), 157–204 (2008)
- [17] Glimm, B., Horrocks, I., Sattler, U.: Unions of conjunctive queries in *SHOQ*. In: Proc. of the 11th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'08). pp. 252–262 (2008)
- [18] Glimm, B., Kazakov, Y., Lutz, C.: Status *QIO*: An update. In: Proc. of the 2011 Int. Workshop on Description Logics (DL'11). pp. 136–146 (2011)
- [19] Horrocks, I., Sattler, U., Tobies, S.: Practical reasoning for very expressive description logics. *Logic Journal of the Interest Group in Pure and Applied Logic* 8(3), 239–263 (2000)
- [20] Lippmann, M.: Temporalised Description Logics for Monitoring Partially Observable Events. Ph.D. thesis, TU Dresden, Germany (2014)
- [21] Lutz, C.: The complexity of conjunctive query answering in expressive description logics. In: Proc. of the 4th Int. Joint Conf. on Automated Reasoning (IJCAR'08). pp. 179–193 (2008)

- [22] Lutz, C., Wolter, F.: Non-uniform data complexity of query answering in description logics. In: Brewka, G., Eiter, T., McIlraith, S.A. (eds.) Proc. of the 13th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'12). pp. 297–307 (2012)
- [23] Ngo, N., Ortiz, M., Šimkus, M.: The combined complexity of reasoning with closed predicates in description logics. In: Proc. of the 28th Int. Workshop on Description Logics (DL'15). pp. 249–261 (2015)
- [24] Ortiz, M., Calvanese, D., Eiter, T.: Data complexity of query answering in expressive description logics via tableaux. *Journal of Automated Reasoning* 41(1), 61–98 (2008)
- [25] Ortiz, M., Šimkus, M.: Reasoning and query answering in description logics. In: Reasoning Web. 8th Int. Summer School, chap. 1, pp. 1–53 (2012)
- [26] Pnueli, A.: The temporal logic of programs. In: Proc. of the 18th Annual Symp. on Foundations of Computer Science (SFCS'77). pp. 46–57 (1977)
- [27] Poggi, A., Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Linking data to ontologies. *Journal on Data Semantics X*, 133–173 (2008)
- [28] Rudolph, S., Glimm, B.: Nominals, inverses, counting, and conjunctive queries or: Why infinity is your friend! *Journal of Artificial Intelligence Research* 39(1), 429–481 (2010)
- [29] Schaerf, A.: On the complexity of the instance checking problem in concept languages with existential quantification. *Journal of Intelligent Information Systems* 2(3), 265–278 (1993)
- [30] Vardi, M.Y.: Reasoning about the past with two-way automata. In: Proc. of the 25th Int. Coll. on Automata, Languages and Programming (ICALP'98). pp. 628–641 (1998)