

# Reverse Engineering of Temporal Queries Mediated by LTL Ontologies

Marie Fontin<sup>1</sup>, Boris Konev<sup>2</sup>, Vladislav Ryzhikov<sup>3</sup>, Yury Savateev<sup>4</sup>,  
Frank Wolter<sup>2</sup> and Michael Zakharyashev<sup>3</sup>

<sup>1</sup>Université Paris Cité, CNRS, IRIF, France

<sup>2</sup>Department of Computer Science, University of Liverpool, UK

<sup>3</sup>Department of Computer Science and Information Systems, Birkbeck, University of London, UK

<sup>4</sup>School of Electronics and Computer Science, University of Southampton, UK

mfortin@irif.fr, {boris.konev,wolter}@liverpool.ac.uk, {vlad,michael}@dcs.bbk.ac.uk,  
y.savateev@soton.ac.uk

## Abstract

In reverse engineering of database queries, we aim to construct a query from a given set of answers and non-answers; it can then be used to explore the data further or as an explanation of the answers and non-answers. We investigate this query-by-example problem for queries formulated in positive fragments of linear temporal logic *LTL* over timestamped data, focusing on the design of suitable query languages and the combined and data complexity of deciding whether there exists a query in the given language that separates the given answers from non-answers. We consider both plain *LTL* queries and those mediated by *LTL*-ontologies.

## 1 Introduction

Supporting users of databases by constructing a query from examples of answers and non-answers to the query has been a major research area since the 2000s [Martins, 2019]. In the database community, research has focussed on standard query languages such as SQL, graph query languages, and SPARQL [Zhang *et al.*, 2013; Weiss and Cohen, 2017; Kalashnikov *et al.*, 2018; Deutch and Gilad, 2019; Staworko and Wieczorek, 2012; Barceló and Romero, 2017; Cohen and Weiss, 2016; Arenas *et al.*, 2016]. The KR community has been concerned with constructing queries from examples under the open world semantics and with background knowledge given by an ontology [Gutiérrez-Basulto *et al.*, 2018; Ortiz, 2019; Cima *et al.*, 2021; Jung *et al.*, 2021; Jung *et al.*, 2022]. A fundamental problem that has been investigated by both communities is known as *separability* or *query-by-example* (QBE), a term coined by Zloof [1977]:

**Given:** sets  $E^+$  and  $E^-$  of pairs  $(\mathcal{D}, \mathbf{d})$  with a database instance  $\mathcal{D}$  and a tuple  $\mathbf{d}$  in  $\mathcal{D}$ , a (possibly empty) ontology  $\mathcal{O}$ , and a query language  $\mathcal{Q}$ .

**Problem:** decide whether there exists a query  $q \in \mathcal{Q}$  separating  $(E^+, E^-)$  in the sense that  $\mathcal{O}, \mathcal{D} \models q(\mathbf{d})$  for all  $(\mathcal{D}, \mathbf{d}) \in E^+$  and  $\mathcal{O}, \mathcal{D} \not\models q(\mathbf{d})$  for all  $(\mathcal{D}, \mathbf{d}) \in E^-$ .

If such a  $q$  exists, then  $(E^+, E^-)$  is often called *satisfiable* w.r.t.  $\mathcal{Q}$  under  $\mathcal{O}$ , and the construction of  $q$  is called *learning*.

In many applications, the input data is timestamped and queries are naturally formulated in languages with temporal operators. In this paper, we investigate temporal query-by-example by focusing on the basic but very useful case where data  $\mathcal{D}$  is a set of timestamped atomic propositions. Our query languages are positive fragments of *linear temporal logic LTL* with the temporal operators  $\diamond$  (eventually),  $\circ$  (next), and  $\cup$  (until) interpreted under the strict semantics [Demri *et al.*, 2016]. To enforce generalisation, we do not admit  $\vee$ . Our most expressive query language  $\mathcal{Q}[\cup]$  is thus defined as the set of formulas constructed from atoms using  $\wedge$  and  $\cup$  (via which  $\circ$  and  $\diamond$  are expressible); the fragments  $\mathcal{Q}[\diamond]$  and  $\mathcal{Q}[\circ, \diamond]$  are defined analogously. Ontologies can be given in full *LTL* or its fragments  $LTL^{\square, \diamond}$  (known as the *Prior logic* [Prior, 1956]), which only uses the operators  $\square$  (always in the future) and  $\diamond$ , and the Horn fragment  $LTL_{horn}^{\square, \circ}$  containing axioms of the form  $C_1 \wedge \dots \wedge C_k \rightarrow C_{k+1}$ , where the  $C_i$  are atoms possibly prefixed by  $\square$  and  $\circ$  for  $i \leq k + 1$ , and also by  $\diamond$  for  $i \leq k$ . Ontology axioms are supposed to hold at all times. In fact, already this basic ‘one-dimensional’ temporal ontology-mediated querying formalism provides enough expressive power in those real-world situations where the interaction among individuals in the object domain is not important and can be disregarded in data modelling; see [Artale *et al.*, 2021] and also Example 1 and the references before it.

Within this temporal setting, we take a broad view of the potential applications of the QBE problem. On the one hand, there are non-expert users who would like to explore data via queries but are not familiar with temporal logic. They usually are, however, capable of providing data examples illustrating the queries they are after. QBE supports such users in the construction of those queries. On the other hand, the positive and negative data examples might come from an application, and the user is interested in possible explanations of the examples. Such an explanation is then provided by a temporal query separating the positive examples from the negative ones. In this case, our goal is similar to recent work on learning *LTL* formulas in explainable planning and program synthesis [Lemieux *et al.*, 2015; Neider and Gavran, 2018; Camacho and McIlraith, 2019; Fijalkow and Lagarde, 2021; Raha *et al.*, 2022; Fortin *et al.*, 2022].

**Example 1.** Imagine an engineer whose task is to explain the behaviour of the monitored equipment (say, why an engine stops) in terms of qualitative sensor data such as ‘low temperature’ ( $T$ ), ‘strong vibration’ ( $V$ ), etc. Suppose the engine stopped after the runs  $\mathcal{D}_1^+$  and  $\mathcal{D}_2^+$  below but did not stop after the runs  $\mathcal{D}_1^-, \mathcal{D}_2^-, \mathcal{D}_3^-$ , where we assume the runs to start at 0 and measurements to be recorded at moments 0, 1, 2, . . . :

$$\begin{aligned} \mathcal{D}_1^+ &= \{T(2), V(4)\}, \mathcal{D}_2^+ = \{T(1), V(4)\}, \\ \mathcal{D}_1^- &= \{T(1)\}, \mathcal{D}_2^- = \{V(4)\}, \mathcal{D}_3^- = \{V(1), T(2)\}. \end{aligned}$$

The  $\diamond$ -query  $q = \diamond(T \wedge \diamond V)$  is true at 0 in the  $\mathcal{D}_i^+$ , false in  $\mathcal{D}_i^-$ , and so gives a possible explanation of what could cause the engine failure. The example set  $(\{\mathcal{D}_3^+, \mathcal{D}_4^+\}, \{\mathcal{D}_4^-\})$  with

$$\begin{aligned} \mathcal{D}_3^+ &= \{T(1), V(2)\}, \mathcal{D}_4^+ = \{T(1), T(2), V(3)\}, \\ \mathcal{D}_4^- &= \{T(1), V(3)\} \end{aligned}$$

is explained by the U-query  $TUV$ . Using background knowledge, we can compensate for sensor failures resulting in incomplete data. To illustrate, suppose  $\mathcal{E}_1^+ = \{H(3), V(4)\}$ , where  $H$  means ‘heater is on’. If an ontology  $\mathcal{O}$  has the axiom  $\bigcirc H \rightarrow T$  saying that a heater can only be triggered by the low temperature at the previous moment, then the same  $q$  separates  $\{\mathcal{E}_1^+, \mathcal{D}_2^+\}$  from  $\{\mathcal{D}_1^-, \mathcal{D}_2^-, \mathcal{D}_3^-\}$  under  $\mathcal{O}$ .  $\dashv$

Query  $q$  in Example 1 is of a particular ‘linear’ form, in which the order of atoms is fixed and not left open as, for instance, in the ‘branching’  $\diamond T \wedge \diamond V$ . More precisely, *path  $\bigcirc \diamond$ -queries* in the class  $\mathcal{Q}_p[\bigcirc, \diamond]$  take the form

$$q = \rho_0 \wedge \mathbf{o}_1(\rho_1 \wedge \mathbf{o}_2(\rho_2 \wedge \dots \wedge \mathbf{o}_n \rho_n)), \quad (1)$$

where  $\mathbf{o}_i \in \{\bigcirc, \diamond\}$  and  $\rho_i$  is a conjunction of atoms;  $\mathcal{Q}_p[\diamond]$  restricts  $\mathbf{o}_i$  to  $\{\diamond\}$ ; and *path U-queries*  $\mathcal{Q}_p[U]$  look like

$$q = \rho_0 \wedge (\lambda_1 U (\rho_1 \wedge (\lambda_2 U (\dots (\lambda_n U \rho_n) \dots))), \quad (2)$$

where  $\lambda_i$  is a conjunction of atoms or  $\perp$ . Path queries are motivated by two observations. First, if a query language admits conjunctions of queries—unlike our classes of path queries—then, dually to overfitting for  $\vee$ , multiple negative examples become redundant: if  $q_{\mathcal{D}}$  separates  $(E^+, \{\mathcal{D}\})$ , for each  $\mathcal{D} \in E^-$ , then  $\bigwedge_{\mathcal{D} \in E^-} q_{\mathcal{D}}$  separates  $(E^+, E^-)$ . Second, numerous natural query types known from applications can be captured by path queries. For example, the existence of a *common subsequence* of the positive examples (regarded as words) that is not a subsequence of any negative one corresponds to the existence of a separating  $\mathcal{Q}_p[\diamond]$ -query with  $\rho_0 = \top$  and  $\rho_i \neq \top$  for  $i > 0$ , and the existence of a *common subword* of the positive examples that is not a subword of any negative one corresponds to the existence of a separating query of the form  $\diamond(\rho_1 \wedge \bigcirc(\rho_2 \wedge \dots \wedge \bigcirc \rho_n))$ . These and similar queries are the basis of data comparison programs with numerous applications in computational linguistics, bioinformatics, and revision control systems [Bergroth *et al.*, 2000; Chowdhury *et al.*, 2010; Blum *et al.*, 2021].

While path queries express the intended separating pattern of events in many applications, branching queries are needed if the order of events is irrelevant for separation.

**Example 2.** In the setting of Example 1, the positive examples  $\{T(2), V(4)\}$  and  $\{V(1), T(4)\}$  are separated from the negative  $\{T(1)\}$  and  $\{V(4)\}$  by the branching  $\mathcal{Q}[\diamond]$ -query  $\diamond T \wedge \diamond V$  while no path query is capable of doing this.  $\dashv$

Branching  $\mathcal{Q}[\bigcirc, \diamond]$ -queries express transparent existential conditions and can be regarded as *LTL CQs*. However, branching  $\mathcal{Q}[U]$ -queries with nestings of  $U$  on the left-hand side correspond to complex first-order formulas with multiple alternations of quantifiers  $\exists$  and  $\forall$ , which are hard to comprehend. So we also consider the language  $\mathcal{Q}[U_s] \supseteq \mathcal{Q}_p[U]$  of ‘simple’  $\mathcal{Q}[U]$ -queries without such nestings.

In this paper, we take the first steps towards understanding the complexity and especially feasibility of the query-by-example problems  $\text{QBE}(\mathcal{L}, \mathcal{Q})$  with  $\mathcal{L}$  an ontology and  $\mathcal{Q}$  a query language. We are particularly interested in whether there is a difference in complexity between path and branching queries and whether it can be reduced by bounding the number of positive or negative examples. Our results in the ontology-free case are summarised in Table 1, where

QBE for	b+, b−	b+	b− or unbounded
$\mathcal{Q}_p[\diamond]/\mathcal{Q}_p[\bigcirc, \diamond]$	$\leq \text{P}$	$= \text{NP}$	$= \text{NP}$
$\mathcal{Q}[\diamond]/\mathcal{Q}[\bigcirc, \diamond]$	$\leq \text{P}$	$\leq \text{P}$	$= \text{NP}$
$\mathcal{Q}_p[U]$	$= \text{NP}$		
$\mathcal{Q}[U_s]$	$\leq \text{P}$	$\leq \text{P}$	$\geq \text{NP}, \leq \text{PSPACE}$
$\mathcal{Q}[U]$	$\leq \text{PSPACE}$		

Table 1: Complexity in the ontology-free case.

b+ / b− indicate that the number of positive / negative examples is bounded<sup>1</sup>. Note that path queries are indeed harder than branching ones when the number of positive examples is bounded but not in the unbounded case. Our proof techniques range from reductions to common subsequence existence problems [Maier, 1978; Fraser, 1996] and dynamic programming to mimicking separability by path and branching U-queries in terms of containment and simulation of transition systems [Kupferman and Vardi, 1996]. The key to NP upper bounds is the *polynomial separation property* (PSP) of the respective languages: any separable example set is separated by a polynomial-size query. The complexity for  $\mathcal{Q}_p[\diamond]$ ,  $\mathcal{Q}[\diamond]$  can also be obtained from [Fijalkow and Lagarde, 2021] who studied separability by  $\mathcal{Q}[\diamond]$ -queries of bounded size.

In the presence of ontologies, we distinguish between the combined complexity of  $\text{QBE}(\mathcal{L}, \mathcal{Q})$ , when both data and ontology are regarded as input, and the data complexity, when the ontology is deemed fixed or negligibly small compared with the data. We obtain encouraging results:  $\mathcal{Q}_p[\diamond]$ - and  $\mathcal{Q}[\diamond]$ -queries mediated by  $LTL^{\bigcirc \diamond}$ -ontologies and all of our queries mediated by  $LTL_{horn}^{\bigcirc \diamond}$ -ontologies enjoy the same data complexity as in Table 1. The combined complexity results for queries with  $LTL_{horn}^{\bigcirc \diamond}$ -ontologies we have obtained so far are given in Table 2. Interestingly, QBE for query classes with  $\diamond$  and  $\bigcirc$  only is PSPACE-complete— not harder than satisfiability. The upper bound is proved by establishing the *exponential separation property* for all of these classes of queries and using the canonical (aka minimal) model property of

<sup>1</sup>We do not consider queries with  $\bigcirc$  only as separability is trivially in P and does not detect any useful patterns.

$\mathcal{Q}[\diamond] / \mathcal{Q}_p[\diamond]$	= PSPACE
$\mathcal{Q}[\circ, \diamond] / \mathcal{Q}_p[\circ, \diamond]$	
$\mathcal{Q}[U_s]$	
$\mathcal{Q}_p[U]$	
$\mathcal{Q}[U]$	$\geq \text{PSPACE}, \leq \text{EXPTIME}$
	$\geq \text{NEXPTIME}, \leq \text{EXPSpace}$
	$\geq \text{PSPACE}, \leq 2\text{EXPTIME}$

Table 2: Combined complexity of  $\text{QBE}(LTL_{\text{horn}}^{\square\circ}, \mathcal{Q})$  in both bounded and unbounded cases.

Horn *LTL*. The upper bounds for U-queries are by reduction to the simulation and containment problems for exponential-size transition systems. For arbitrary *LTL*-ontologies, this technique only gives a  $2\text{EXPTIME}$  upper bound for  $\mathcal{Q}[U_s]$  and a  $2\text{EXPSpace}$  one for  $\mathcal{Q}_p[U]$ . Separability by (path)  $\diamond$ -queries under  $LTL^{\square\circ}$  ontologies turns out to be  $\Sigma_2^p$ -complete, where the upper bound is shown by establishing the PSP.

Compared with non-temporal QBE, our results are very encouraging: QBE is  $\text{CONEXPTIME}$ -complete for conjunctive queries (CQs) over standard relational databases [Willard, 2010; ten Cate and Dalmau, 2015] and even undecidable for CQs under  $\mathcal{ELI}$  or  $\mathcal{ALC}$  ontologies [Funk *et al.*, 2019; Jung *et al.*, 2020].

## 2 Further Related Work

We now briefly comment on a few other related research areas. One of them is concept learning in description logic (DL), as proposed by [Badea and Nienhuys-Cheng, 2000] who, inspired by inductive logic programming, used refinement operators to construct a concept separating positive and negative examples in a DL ABox. There has been significant interest in this approach [Lehmann and Haase, 2009; Lehmann and Hitzler, 2010; Lisi and Straccia, 2015; Sarker and Hitzler, 2019; Lisi, 2012; Rizzo *et al.*, 2020]. Prominent systems include the DL LEARNER [Bühmann *et al.*, 2016], DL-FOIL [Fanizzi *et al.*, 2018] and its extension DL-FOCL [Rizzo *et al.*, 2018], SPaCEL [Tran *et al.*, 2017], YINYANG [Iannone *et al.*, 2007], PFOIL-DL [Straccia and Mucci, 2015], and EVOLARNER [Heindorf *et al.*, 2022]. However, this work has not considered the complexity of separability. Also closely related is the work on the separability of two formal (e.g., regular) languages using a weaker (e.g., FO-definable) language [Place and Zeitoun, 2016; Hofman and Martens, 2015; Place and Zeitoun, 2022]. When translated into a logical separability problem, the main difference to our results is that one demands  $\mathcal{O}, \mathcal{D} \models \neg q(d)$ —and not just  $\mathcal{O}, \mathcal{D} \not\models q(d)$ —for all  $(\mathcal{D}, d) \in E^-$ .

## 3 Preliminaries

*LTL-formulas* are built from *atoms*  $A_i$ ,  $i < \omega$ , using the Booleans and (future-time) temporal operators  $\circ, \diamond, \square, U$ , which we interpret under the *strict semantics* [Gabbay *et al.*, 2003; Demri *et al.*, 2016]. An *LTL-interpretation*  $\mathcal{I}$  identifies those atoms  $A_i$  that are *true* at each time instant  $n \in \mathbb{N}$ , written  $\mathcal{I}, n \models A_i$ . The truth-relation for atoms is extended inductively to *LTL-formulas* by taking  $\mathcal{I}, n \models \varphi \cup \psi$  iff  $\mathcal{I}, m \models \psi$ , for some  $m > n$ , and  $\mathcal{I}, k \models \varphi$  for all  $k \in (n, m)$ , and using the standard clauses for the Booleans and equiva-

lences  $\circ\varphi \equiv \perp \cup \varphi$ ,  $\diamond\varphi \equiv \top \cup \varphi$  and  $\square\varphi \equiv \neg\diamond\neg\varphi$  with Boolean *constants*  $\perp$  and  $\top$  for ‘false’ and ‘true’.

An *LTL-ontology*,  $\mathcal{O}$ , is any finite set of *LTL-formulas*, called the *axioms* of  $\mathcal{O}$ . An interpretation  $\mathcal{I}$  is a *model* of  $\mathcal{O}$  if all axioms of  $\mathcal{O}$  are true at *all times* in  $\mathcal{I}$ . As mentioned in the introduction, apart from full *LTL* we consider its Prior  $\square\circ$ -fragment  $LTL^{\square\circ}$  and  $LTL_{\text{horn}}^{\square\circ}$  whose axioms take the form

$$C_1 \wedge \dots \wedge C_k \rightarrow C_{k+1} \quad (3)$$

with  $C_i$  given by  $C ::= A_i \mid \perp \mid \square C \mid \circ C$ . In fact, we could allow  $\diamond$  on the left-hand side of (3) as  $\diamond C \rightarrow C'$  can be replaced by  $\circ C \rightarrow A, \circ A \rightarrow A, A \rightarrow C'$  with fresh  $A$ .

A *data instance* is a finite set  $\mathcal{D}$  of atoms  $A_i(\ell)$  with a *timestamp*  $\ell \in \mathbb{N}$ ;  $\max \mathcal{D}$  is the maximal timestamp in  $\mathcal{D}$ . We access data by means of *LTL* analogues of conjunctive queries: our *queries*,  $\varkappa$ , are constructed from atoms,  $\perp$  and  $\top$  using  $\wedge, \circ, \diamond$  and  $U$ . The class of queries that only use operators from  $\Phi \subseteq \{\circ, \diamond, U\}$  is denoted by  $\mathcal{Q}[\Phi]$ ;  $\mathcal{Q}_p[\Phi]$  is its subclass of *path-queries*, which take the form (1) or (2); and  $\mathcal{Q}[U_s]$  comprises *simple queries* in  $\mathcal{Q}[U]$  that do not contain subqueries  $\varkappa_1 U \varkappa_2$  with an occurrence of  $U$  in  $\varkappa_1$ . Note that  $\mathcal{Q}_p[U] \subseteq \mathcal{Q}[U_s]$ . The *temporal depth*  $\text{tdp}(\varkappa)$  of  $\varkappa$  is the maximum number of nested temporal operators in  $\varkappa$ .

An interpretation  $\mathcal{I}$  is a *model* of a data instance  $\mathcal{D}$  if  $\mathcal{I}, \ell \models A_i(\ell)$  for all  $A_i(\ell) \in \mathcal{D}$ .  $\mathcal{O}$  and  $\mathcal{D}$  are *consistent* if they have a model. We call  $k \leq \max \mathcal{D}$  a (*certain*) *answer* to the *ontology-mediated query*  $(\mathcal{O}, \varkappa)$  over  $\mathcal{D}$  and write  $\mathcal{O}, \mathcal{D} \models \varkappa(k)$  if  $\mathcal{I}, k \models \varkappa$  in all models  $\mathcal{I}$  of  $\mathcal{O}$  and  $\mathcal{D}$ .

Let  $\mathcal{L}$  and  $\mathcal{Q}$  be an ontology and query language defined above. The *query-by-example problem*  $\text{QBE}(\mathcal{L}, \mathcal{Q})$  we are concerned with in this paper is formulated as follows:

**given** an  $\mathcal{L}$ -ontology  $\mathcal{O}$  and an *example set*  $E = (E^+, E^-)$  with finite sets  $E^+$  and  $E^-$  of *positive* and, respectively, *negative* data instances,

**decide** whether  $E$  is *Q-separable* under  $\mathcal{O}$  in the sense that there is a  $\mathcal{Q}$ -query  $\varkappa$  with  $\mathcal{O}, \mathcal{D} \models \varkappa(0)$  for all  $\mathcal{D} \in E^+$  and  $\mathcal{O}, \mathcal{D} \not\models \varkappa(0)$  for all  $\mathcal{D} \in E^-$ .

If  $\mathcal{L} = \emptyset$ , we shorten  $\text{QBE}(\emptyset, \mathcal{Q})$  to  $\text{QBE}(\mathcal{Q})$ . We also consider the QBE problems with the input example sets having a *bounded* number of positive and/or negative examples, denoted  $\text{QBE}^{\text{b}^+}(\mathcal{L}, \mathcal{Q})$ ,  $\text{QBE}^{\text{b}}(\mathcal{L}, \mathcal{Q})$ , or  $\text{QBE}^{\text{b}^+}(\mathcal{L}, \mathcal{Q})$ . Notations like  $\text{QBE}_{1-}^{2+}(\mathcal{L}, \mathcal{Q})$  should be self-explanatory. The size of  $\mathcal{O}, E, \varkappa$ , denoted  $|\mathcal{O}|, |E|, |\varkappa|$ , respectively, is the number of symbols in it with the timestamps given in unary.

The next example illustrates the definitions and relative expressive power of queries with different temporal operators.

**Example 3.** (a) Let  $E = (\{\mathcal{D}_1\}, \{\mathcal{D}_2\})$  with  $\mathcal{D}_1 = \{A(1)\}$ ,  $\mathcal{D}_2 = \{A(2)\}$ . Then  $\circ A$  separates  $E$  but no  $\mathcal{Q}[\diamond]$ -query does.  $E$  is not separable under  $\mathcal{O} = \{\circ A \rightarrow A\}$  by any query  $\varkappa$  as  $\mathcal{O}, \mathcal{D}_1 \models \varkappa(0)$  implies  $\mathcal{O}, \mathcal{D}_2 \models \varkappa(0)$ .

(b) Let  $E = (\{\mathcal{D}_1, \mathcal{D}_2\}, \{\mathcal{D}_3\})$  with  $\mathcal{D}_1 = \{A(1), B(2)\}$ ,  $\mathcal{D}_2 = \{A(2), B(3)\}$ ,  $\mathcal{D}_3 = \{A(3), B(5)\}$ . Then the query  $\diamond(A \wedge \circ B)$  separates  $E$  but no query in  $\mathcal{Q}[\diamond]$  does.

(c)  $A \cup B$  separates  $(\{\{B(1)\}, \{A(1), B(2)\}\}, \{\{B(2)\}\})$  but no  $\mathcal{Q}[\circ, \diamond]$ -query does.  $\dashv$

We now establish a few important *polynomial-time reductions*,  $\leq_p$ , among the QBE-problems for various query

classes, including  $\mathcal{Q}_p^\circ[\diamond]$ -queries of the form

$$\varkappa = \rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond\rho_n)), \quad (4)$$

where each  $\rho_i$  is a  $\mathcal{Q}_p[\circ]$ -query (i.e.,  $\diamond$ -free  $\mathcal{Q}_p[\circ, \diamond]$ -query).

**Theorem 4.** *The following polynomial-time reductions hold:*

(i.1)  $\text{QBE}(\mathcal{L}, \mathcal{Q}) \leq_p \text{QBE}_{1-}(\mathcal{L}, \mathcal{Q})$ , for any  $\mathcal{Q}$  closed under  $\wedge$ , and any  $\mathcal{L}$  (including  $\mathcal{L} = \emptyset$ ),

(i.2)  $\text{QBE}(\mathcal{L}, \mathcal{Q}) \leq_p \text{QBE}^{2+}(\mathcal{L}, \mathcal{Q})$ , for  $\mathcal{L} \in \{\text{LTL}, \text{LTL}^{\diamond\diamond}\}$ ,

(i.3)  $\text{QBE}(\mathcal{L}, \mathcal{Q}[\circ, \diamond]) \leq_p \text{QBE}(\mathcal{L}, \mathcal{Q}_p^\circ[\diamond])$  and  $\text{QBE}(\mathcal{L}, \mathcal{Q}[\diamond]) \leq_p \text{QBE}(\mathcal{L}, \mathcal{Q}_p[\diamond])$ , for any  $\mathcal{L}$ ,

(ii.1)  $\text{QBE}(\mathcal{Q}_p[\diamond]) \leq_p \text{QBE}(\mathcal{Q}_p[\circ, \diamond])$  and  $\text{QBE}(\mathcal{Q}_p[\diamond]) \leq_p \text{QBE}(\mathcal{Q}_p[\text{U}]) \leq_p \text{QBE}_{1-}(\mathcal{Q}_p[\text{U}])$ ,

(ii.2)  $\text{QBE}(\mathcal{Q}[\circ, \diamond]) =_p \text{QBE}(\mathcal{Q}[\diamond]) \leq_p \text{QBE}(\mathcal{Q}[\text{U}_s])$ .

*Reductions (i.1)–(i.3) work for combined complexity; (i.1), (i.3) also work for data complexity. The reductions preserve boundedness of the number of positive/negative examples.*

*Proof.* In (i.1),  $(E^+, E^-)$  with  $E^- = \{\mathcal{D}_1, \dots, \mathcal{D}_n\}$  is  $\mathcal{Q}$ -separable under  $\mathcal{O}$  iff each  $(E^+, \{\mathcal{D}_i\})$  is because if  $\varkappa_i$  separates  $(E^+, \{\mathcal{D}_i\})$ , then  $\varkappa_1 \wedge \dots \wedge \varkappa_n$  separates  $(E^+, E^-)$ .

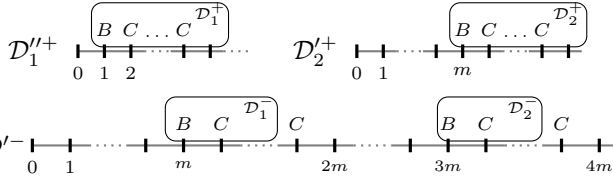
In (i.2),  $(E^+, E^-)$  with  $E^+ = \{\mathcal{D}_1, \dots, \mathcal{D}_n\}$ ,  $n > 1$ , is  $\mathcal{Q}$ -separable under  $\mathcal{O}$  iff  $(E^+, E^-)$  is  $\mathcal{Q}$ -separable under  $\mathcal{O}'$  that extends  $\mathcal{O}$  with the following axioms simulating  $E^+$ :

$$S_1 \rightarrow A_1 \vee \dots \vee A_n, \quad S_2 \rightarrow A_1 \vee \dots \vee A_n, \\ C_i \wedge \diamond A_j \rightarrow X, \quad D_i \wedge \diamond A_j \rightarrow X, \quad \text{for } X(i) \in \mathcal{D}_j,$$

where  $S_1, S_2, A_k, C_l, D_l$ , for  $l \leq n' = \max_i \max \mathcal{D}_i$ , are fresh and  $E'^+$  consists of  $\{C_0(0), \dots, C_k(n'), S_1(n'+1)\}$  and  $\{D_0(0), \dots, D_k(n'), S_2(n'+1)\}$ .

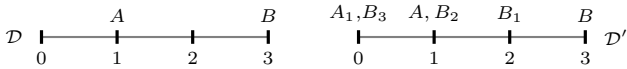
(i.3) Using  $[\rho_0 \wedge \diamond(\rho_1 \wedge \bigwedge_i \diamond \varkappa_i)] \equiv [\rho_0 \wedge \bigwedge_i \diamond(\rho_1 \wedge \diamond \varkappa_i)]$ ,  $\circ \diamond \varkappa \equiv \diamond \circ \varkappa$  and  $\circ(\varkappa \wedge \varkappa') \equiv (\circ \varkappa \wedge \circ \varkappa')$  we convert, in polytime, each  $\mathcal{Q}[\circ, \diamond]$ -query to an equivalent conjunction of  $\mathcal{Q}_p^\circ[\diamond]$ -queries. Thus, there is  $\mathbf{q} \in \mathcal{Q}[\circ, \diamond]$  separating  $(E^+, E^-)$  iff there are polysize  $\mathbf{q}_{\mathcal{D}} \in \mathcal{Q}_p^\circ[\diamond]$  separating  $(E^+, \{\mathcal{D}\})$ , for each  $\mathcal{D} \in E^-$ .

(ii.1) The first two reductions are shown by adding to  $E^+ \ni \mathcal{D}$ , for some  $\mathcal{D}$ , the data instance  $\mathcal{D}' = \{A(mn) \mid A(n) \in \mathcal{D}\}$  with  $m = \max \mathcal{D} + 2$ . Now, if  $\mathcal{D} \models \varkappa(0)$  and  $\mathcal{D}' \models \varkappa(0)$ , for  $\varkappa \in \mathcal{Q}_p[\text{U}]$ , then  $\varkappa$  is equivalent to a  $\mathcal{Q}_p[\diamond]$ -query. The third reduction, illustrated below for  $E^+ = \{\mathcal{D}_1^+, \mathcal{D}_2^+\}$  and  $E^- = \{\mathcal{D}_1^-, \mathcal{D}_2^-\}$ , transforms  $E$  into two positive and one negative example using ‘pads’ of fresh atoms  $B, C$ . We show



that  $E$  is  $\mathcal{Q}_p[\text{U}]$ -separable iff  $(\{\mathcal{D}_1''+, \mathcal{D}_2''+\}, \{\mathcal{D}'-\})$  is.

(ii.2) The first reduction is established by modifying every  $\mathcal{D}$  in the given  $E$  as illustrated below using fresh atoms  $A_i$  and  $B_j$  that encode  $\circ^i A$  and  $\circ^j B$ , respectively:



Then  $E$  is  $\mathcal{Q}[\circ, \diamond]$ -separable iff  $E'$  is  $\mathcal{Q}[\diamond]$ -separable. The converse and the second reduction are similar to (ii.1).  $\dashv$

## 4 QBE without Ontologies

We start investigating the complexity of the QBE problems for LTL by considering queries without mediating ontologies.

**Theorem 5.** *The QBE-problems for the classes of queries defined above (with the empty ontology) belong to the complexity classes shown in Table 1.*

We comment on the proof in the remainder of this section.

**$\circ \diamond$ -queries.** NP-hardness is established by reduction of the consistent subsequence existence problems [Fraser, 1996, Theorems 2.1, 2.2] in tandem with Theorem 4; membership in NP follows from the fact that separating queries, if any, can always be taken of polynomial size.

Tractability is shown using dynamic programming. We explain the idea for  $\text{QBE}_{b-}^+(\mathcal{Q}_p[\circ, \diamond])$ ,  $E^+ = \{\mathcal{D}_1^+, \mathcal{D}_2^+\}$  and  $E^- = \{\mathcal{D}_1^-, \mathcal{D}_2^-\}$ . Suppose  $\varkappa$  takes the form (1) with  $\rho_n \neq \top$ . Then  $\mathcal{D} \models \varkappa(0)$  iff there is a strictly monotone map  $f: [0, n] \rightarrow [0, \max \mathcal{D}]$  with  $f(0) = 0$ ,  $f(i+1) = f(i) + 1$  if  $\circ i = \circ$ , and  $\rho_i \subseteq t_{\mathcal{D}}(f(i)) = \{A \mid A(f(i)) \in \mathcal{D}\}$ . We call such an  $f$  a *satisfying assignment* for  $\varkappa$  in  $\mathcal{D}$ . Let  $S_{i,j}$  be the set of tuples  $(k, \ell_1, \ell_2, n_1, n_2)$  such that  $\ell_1 \leq i \leq \max \mathcal{D}_1^+$ ,  $\ell_2 \leq j \leq \max \mathcal{D}_2^+$ , and there is  $\varkappa = \rho_0 \wedge \circ_1(\rho_1 \wedge \dots \wedge \circ_k \rho_k)$  for which (i) there are satisfying assignments  $f_1, f_2$  in  $\mathcal{D}_1^+$  and  $\mathcal{D}_2^+$  with  $f_1(k) = \ell_1$  and  $f_2(k) = \ell_2$ , respectively, and (ii)  $n_1$  is minimal with a satisfying assignment  $f$  for  $\varkappa$  in  $\mathcal{D}_1^-$  having  $f(k) = n_1$ , and  $n_1 = \infty$  if there is no such  $f$ ; and similarly for  $n_2, \mathcal{D}_2^-$ . It suffices to compute  $S_{\max \mathcal{D}_1^+, \max \mathcal{D}_2^+}$  in polytime. This can be done incrementally by initially observing that  $S_{0,0}$  can only contain  $(0, 0, 0, 0, 0)$ , which is the case if there is  $\rho_0 \subseteq t_{\mathcal{D}_1^+}(0)$ ,  $\rho_0 \subseteq t_{\mathcal{D}_2^+}(0)$  and  $\rho_0 \not\subseteq t_{\mathcal{D}_1^-}(0)$ ,  $\rho_0 \not\subseteq t_{\mathcal{D}_2^-}(0)$  (and similarly for  $S_{i,0}$ ).

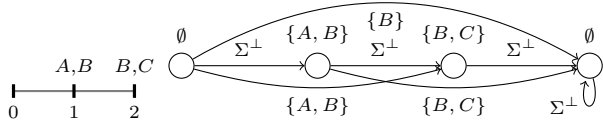
**U-queries.** NP-hardness for  $\mathcal{Q}_p[\text{U}]$ ,  $\mathcal{Q}[\text{U}_s]$  follows from Theorem 4 (ii.1), (ii.2) and NP-hardness for  $\circ \diamond$ -queries.

The upper bounds are shown by reduction of  $\mathcal{Q}_p[\text{U}]$ - and  $\mathcal{Q}[\text{U}_s]$ -separability to the simulation and containment problems for transition systems [Kupferman and Vardi, 1996]. A *transition system*,  $S$ , is a digraph each of whose nodes and edges is labelled by some set of symbols from a node or, respectively, edge alphabet;  $S$  also has a designated set  $S_0$  of *start nodes*. A *run* of  $S$  is a path in digraph  $S$ , starting in  $S_0$ , together with all of its labels. The *computation tree* of  $S$  is the tree unravelling  $\mathfrak{T}_S$  of  $S$ . For systems  $S$  and  $S'$  over the same alphabets, we say that  $S$  is *contained* in  $S'$  if, for every run  $r$  of  $S$ , there is a run  $r'$  of  $S'$  such that  $r$  and  $r'$  have the same length and the labels on the states and edges in  $r$  are subsumed by the corresponding labels in  $r'$ .  $S$  is *simulated* by  $S'$  if  $\mathfrak{T}_S$  is *finitely embeddable* into  $\mathfrak{T}_{S'}$  in the sense that every finite subtree<sup>2</sup> of  $\mathfrak{T}_S$  can be homomorphically mapped into  $\mathfrak{T}_{S'}$ , preserving (subsumption of) node and edge labels.

Now, let  $E = (E^+, E^-)$  with  $E^\sigma = \{\mathcal{D}_i \mid i \in I^\sigma\}$ , for  $\sigma \in \{+, -\}$  and disjoint  $I^+$  and  $I^-$ , and let  $\Sigma$  be the signature of  $E$ . For each  $i \in I^+ \cup I^-$ , we take a transition system  $S^i$  with states  $0^i, \dots, (\max \mathcal{D}_i + 1)^i$ , where  $(\max \mathcal{D}_i + 1)^i$  is labelled with  $\emptyset$  and the remaining  $j^i$  by  $\{A \mid A(j) \in \mathcal{D}_i\}$ . Transitions are  $j^i \rightarrow k^i$ , for  $0 \leq j < k \leq \max \mathcal{D}_i + 1$ , that

<sup>2</sup>A *subtree* is a convex subset of  $\mathfrak{T}_S$ 's nodes with some start node.

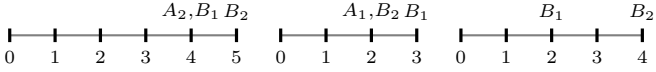
are labelled by  $\{A \in \Sigma \cup \{\perp\} \mid A(n) \in \mathcal{D}_i, n \in (j, k)\}$  and  $(\max \mathcal{D}_i + 1)^i \rightarrow (\max \mathcal{D}_i + 1)^i$  with label  $\Sigma^\perp = \Sigma \cup \{\perp\}$ . Thus,  $\mathcal{D}_i$  shown on the left below gives rise to  $S^i$  on the right:



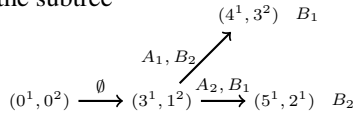
We form the direct product (synchronous composition)  $\mathfrak{P}$  of  $\{S^i \mid i \in I^+\}$ , for  $I^+ = \{1, \dots, l\}$ , whose states are vectors  $(s_1, \dots, s_l)$  of states  $s_i \in S^i$ , which are labelled by the intersection of the labels of  $s_i$  in  $S^i$ , with transitions  $(s_1, \dots, s_l) \rightarrow (p_1, \dots, p_l)$ , if  $s_i \rightarrow p_i$  in  $S^i$  for all  $i$ , also labelled by the intersection of the component transition labels. On the other hand, we take the disjoint union  $\mathfrak{N}$  of  $S^i$ , for  $i \in I^-$ , and establish the following separability criterion:

**Theorem 6.** (i)  $E$  is not  $\mathcal{Q}[U_s]$ -separable iff  $\mathfrak{P}$  is simulated by  $\mathfrak{N}$ . (ii)  $E$  is not  $\mathcal{Q}_p[U]$ -separable iff  $\mathfrak{P}$  is contained in  $\mathfrak{N}$ .

**Example 7.** For the example set depicted below, in which the only negative instance is on the right-hand side,



$\mathfrak{T}_{\mathfrak{P}}$  contains the subtree

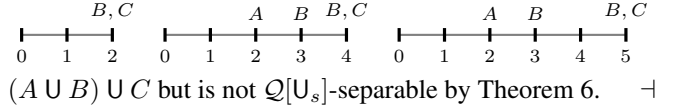


where only the last  $\mathfrak{P}$ -node of a  $\mathfrak{T}_{\mathfrak{P}}$ -node (a sequence) is indicated together with the atoms that are true at nodes and on edges. Intuitively,  $\mathfrak{T}_{\mathfrak{P}}$  ‘represents’ all possible  $\mathcal{Q}[U_s]$ -queries and its paths represent  $\mathcal{Q}_p[U]$ -queries  $\varkappa$  such that  $\mathcal{O}, \mathcal{D} \models \varkappa(0)$  for all  $\mathcal{D} \in E^+$ . The  $\mathcal{Q}[U_s]$ -query given by the subtree above is  $\varkappa = \diamond(((A_1 \wedge B_2) \cup B_1) \wedge ((A_2 \wedge B_1) \cup B_2))$ . The subtree is not embeddable into  $\mathfrak{T}_{\mathfrak{N}}$  (obtained for the negative instance), so  $\varkappa$  separates  $E$ . Observe that every path in  $\mathfrak{T}_{\mathfrak{P}}$  (and in the subtree above) is embeddable into  $\mathfrak{T}_{\mathfrak{N}}$ .

By inspecting the structure of  $\mathfrak{P}$  and  $\mathfrak{N}$  we observe that if  $\mathfrak{P}$  has a run that is not embeddable into any run of  $\mathfrak{N}$ , then we can find such a run of length  $\leq M = \min\{\max \mathcal{D}_i \mid i \in I^+\}$  (any longer run has  $\emptyset$ -labels on its states after the  $M$ th one). Thus, we can guess the required run and check in P if it is correct, establishing the NP upper bound for  $\mathcal{Q}_p[U]$ . To show the PSPACE upper bound for  $\mathcal{Q}[U_s]$ , we notice that if there is a finite subtree of  $\mathfrak{T}_{\mathfrak{P}}$  that is not embeddable into  $\mathfrak{T}_{\mathfrak{N}}$ , then the full subtree  $\mathfrak{T}_{\mathfrak{P}}^M$  of depth  $M$  is not embeddable into  $\mathfrak{T}_{\mathfrak{N}}$ , which can be checked by constructing  $\mathfrak{T}_{\mathfrak{P}}^M$  branch-by-branch while checking all possible embeddings of these branches into  $\mathfrak{T}_{\mathfrak{N}}$ . Finally, we have the P upper bound for  $\mathcal{Q}[U_s]$  with a bounded number of positive examples because  $\mathfrak{P}$  is constructible in polytime and checking simulation between transition systems is P-complete [Kupferman and Vardi, 1996]. Interestingly, the smallest separating query we can construct in this case is of the same size as  $\mathfrak{T}_{\mathfrak{P}}^M$ , i.e., exponential in  $|E^+|$ ; however, we can check its existence in polytime.

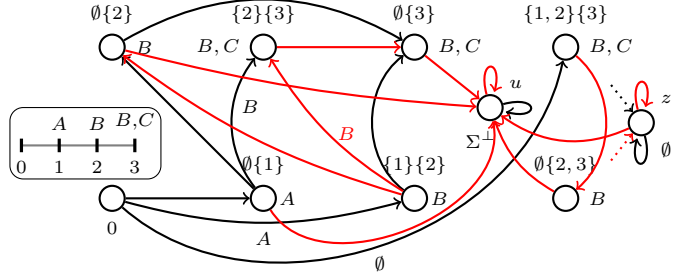
The PSPACE upper bound for  $\mathcal{Q}[U]$  requires a more sophisticated notion of simulation between transition systems.

**Example 8.** The example set below, where only the rightmost instance is negative, is separated by the  $\mathcal{Q}[U]$ -query



$(A \cup B) \cup C$  but is not  $\mathcal{Q}[U_s]$ -separable by Theorem 6.  $\dashv$

We prove a  $\mathcal{Q}[U]$ -inseparability criterion using transition systems whose non-initial/sink states are *pairs* of sets of numbers, and transitions are of *two* types. The picture below shows a data instance and the induced transition system (where  $z$  has incoming arrows labelled by  $\Sigma^\perp$  from all states



but  $u$ , which are all omitted). Each arrow from 0 leads to a state  $\{1, \dots, n-1\}\{n\}$ ; it represents a formula  $\varphi \cup \psi$  that is true at 0, with the arrow label indicating the non-nested atoms of  $\varphi$  and the state label indicating the atoms of  $\psi$ . Each black (resp., red) arrow from  $s_1 s_2$  to  $s'_1 s'_2$  represents a  $U$ -formula  $\alpha_{s_2 \rightarrow s'_1 s'_2}$  (resp.,  $\alpha_{s_1 \rightarrow s'_1 s'_2}$ ) that is true at all points in  $s_2$  (resp.,  $s_1$ ). The black and red transitions are arranged in such a way that a transition from  $s'_1 s'_2$  to  $s_1 s_2$  with an arrow label  $\lambda$  and  $s_1 s_2$ -label  $\mu$  represents the  $U$ -formula  $(\lambda \wedge \bigwedge \alpha_{s_1 \rightarrow s'_1 s'_2}) \cup (\mu \wedge \bigwedge \alpha_{s_2 \rightarrow s'_1 s'_2})$  and similarly for the transitions from 0. A version of Theorem 6 for  $\mathcal{Q}[U]$  and a PSPACE-algorithm are given in the full paper.

## 5 QBE with $LTL_{horn}^{\square\circ}$ -Ontologies

Recall from [Artale *et al.*, 2021] that, for any  $LTL_{horn}^{\square\circ}$ -ontology  $\mathcal{O}$  and data instance  $\mathcal{D}$  consistent with  $\mathcal{O}$ , there is a *canonical model*  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  of  $\mathcal{O}$  and  $\mathcal{D}$  such that, for any query  $\varkappa$  and any  $k \in \mathbb{N}$ , we have  $\mathcal{O}, \mathcal{D} \models \varkappa(k)$  iff  $\mathcal{C}_{\mathcal{O}, \mathcal{D}} \models \varkappa(k)$ .

Let  $sub_{\mathcal{O}}$  be the set of subformulas of the  $C_i$  in the axioms (3) of  $\mathcal{O}$  and their negations. A *type for  $\mathcal{O}$*  is any maximal subset  $tp \subseteq sub_{\mathcal{O}}$  consistent with  $\mathcal{O}$ . Let  $\mathcal{T}$  be the set of all types for  $\mathcal{O}$ . Given an interpretation  $\mathcal{I}$ , we denote by  $tp_{\mathcal{I}}(n)$  the type for  $\mathcal{O}$  that holds at  $n \in \mathbb{N}$  in  $\mathcal{I}$ . For  $\mathcal{O}$  consistent with  $\mathcal{D}$ , we abbreviate  $tp_{\mathcal{C}_{\mathcal{O}, \mathcal{D}}}$  to  $tp_{\mathcal{O}, \mathcal{D}}$ . The canonical models have a periodic structure in the following sense:

**Proposition 9.** For any  $LTL_{horn}^{\square\circ}$  ontology  $\mathcal{O}$  and any data instance  $\mathcal{D}$  consistent with  $\mathcal{O}$ , there are  $s_{\mathcal{O}, \mathcal{D}} \leq 2^{|\mathcal{O}|}$  and  $p_{\mathcal{O}, \mathcal{D}} \leq 2^{2^{|\mathcal{O}|}}$  such that  $tp_{\mathcal{O}, \mathcal{D}}(n) = tp_{\mathcal{O}, \mathcal{D}}(n + p_{\mathcal{O}, \mathcal{D}})$ , for all  $n \geq \max \mathcal{D} + s_{\mathcal{O}, \mathcal{D}}$ . Deciding  $\mathcal{C}_{\mathcal{O}, \mathcal{D}} \models \xi(\ell)$ , for a binary  $\ell$  and a conjunction of atoms  $\xi$ , is in PSPACE/P for combined/data complexity.

We now show that the combined complexity of QBE with  $\diamond$ - and  $\circ$ ,  $\diamond$ -queries is PSPACE-complete in both bounded and unbounded cases, i.e., as complex as  $LTL_{horn}^{\square\circ}$  reasoning.

**Theorem 10.** Let  $\mathcal{Q} \in \{\mathcal{Q}[\circ, \diamond], \mathcal{Q}[\diamond], \mathcal{Q}_p[\circ, \diamond], \mathcal{Q}_p[\diamond]\}$ . Then QBE( $LTL_{horn}^{\square\circ}, \mathcal{Q}$ ) and QBE $_b^+(\mathcal{Q}, LTL_{horn}^{\square\circ})$  are both PSPACE-complete for combined complexity.

*Proof.* PSPACE-hardness is inherited from that of  $LTL_{horn}^{\square\bigcirc}$ . We briefly sketch the proof of the matching upper bound for  $\mathcal{Q}[\bigcirc, \diamond]$  using the reduction of Theorem 4 (i.3). We can assume that  $\mathcal{O}$  and  $\mathcal{D}$  are consistent for any  $\mathcal{D} \in E^+ \cup E^-$ . For if  $\mathcal{O}$  and  $\mathcal{D} \in E^-$  are inconsistent, then  $E$  is not  $\mathcal{Q}$ -separable under  $\mathcal{O}$  as  $\mathcal{O}, \mathcal{D} \models \varkappa(0)$  for all  $\varkappa \in \mathcal{Q}$ ; if  $\mathcal{O}$  and  $\mathcal{D} \in E^+$  are inconsistent, then  $E$  is separable iff  $(E^+ \setminus \{\mathcal{D}\}, E^-)$  is. Checking consistency is known to be PSPACE-complete.

Given an  $LTL_{horn}^{\square\bigcirc}$ -ontology  $\mathcal{O}$  and an example set  $E$ , let

$$k = \max_{\mathcal{D} \in E^+ \cup E^-} (\max \mathcal{D} + s_{\mathcal{O}, \mathcal{D}}), \quad m = \prod_{\mathcal{D} \in E^+ \cup E^-} p_{\mathcal{O}, \mathcal{D}},$$

where  $s_{\mathcal{O}, \mathcal{D}}$  and  $p_{\mathcal{O}, \mathcal{D}}$  in  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  are from Proposition 9. We show that if  $E$  is  $\mathcal{Q}[\bigcirc, \diamond]$ -separable under  $\mathcal{O}$ , then it is separated by a conjunction of  $|E^-|$ -many  $\varkappa \in \mathcal{Q}_p^{\square\bigcirc}[\diamond]$  of  $\diamond$ -depth  $\leq k + 1$  and  $\bigcirc$ -depth  $\leq k + m$  in (4). Indeed, in this case any  $(E^+, \{\mathcal{D}\})$ , for  $\mathcal{D} \in E^-$ , is separated under  $\mathcal{O}$  by some  $\varkappa$  of the form (4) with the  $\rho_l$  of  $\bigcirc$ -depth  $\leq k + m$  because  $\rho_l = \bigwedge_{i=0}^{\ell} \bigcirc^i \lambda_i$  with  $\ell > k + m$  can be replaced by

$$\bigwedge_{i=0}^k \bigcirc^i \lambda_i \wedge \bigwedge_{j=1}^m \bigcirc^{k+j} \bigwedge_{i \leq \ell, j = (i-k) \bmod m} \lambda_i.$$

In addition, if  $n > k$  in (4), then  $(E^+, \{\mathcal{D}^-\})$  is separated by

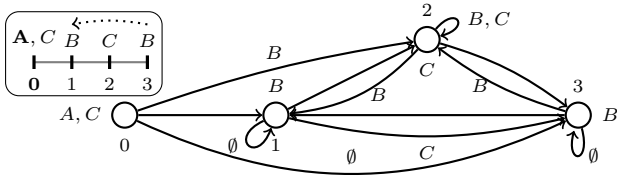
$$\rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond \rho_k)) \wedge \bigwedge_{i=k+1}^n \diamond^{k+1} \rho_i,$$

and so by some  $\rho_0 \wedge \diamond(\rho_1 \wedge \dots \wedge \diamond(\rho_k \wedge \diamond \rho_j))$  with  $k < j \leq n$ . Our nondeterministic PSPACE-algorithm incrementally guesses the  $\rho_l$  and checks if they are satisfiable in the relevant part of the relevant  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  bounded by  $k + m$ .  $\dashv$

The situation is quite different for queries with  $\cup$ :

**Theorem 11.**  $\text{QBE}(LTL_{horn}^{\square\bigcirc}, \mathcal{Q}[U_s])$  is in EXPTIME for combined complexity,  $\text{QBE}(LTL_{horn}^{\square\bigcirc}, \mathcal{Q}_p[U])$  is in EXPSPACE, and  $\text{QBE}_{b^+}^{\square\bigcirc}(LTL_{horn}^{\square\bigcirc}, \mathcal{Q}_p[U])$  is NEXPTIME-hard.

*Proof.* For the upper bounds, we again assume that  $\mathcal{O}$  and  $\mathcal{D}$  are consistent for all  $\mathcal{D} \in E^+ \cup E^-$ . Observe that Theorem 6 continues to hold in the presence of  $LTL_{horn}^{\square\bigcirc}$  ontologies  $\mathcal{O}$  but we need a different construction of transition systems  $S^i$  that represent all  $\mathcal{Q}[U_s]$ -queries mediated by  $\mathcal{O}$  over  $\mathcal{D}_i$ . We illustrate it for  $\mathcal{O} = \{A \rightarrow C \wedge \bigcirc B, B \rightarrow \bigcirc^2 B, B \rightarrow \bigcirc C\}$  and  $\mathcal{D}_i = \{A(0)\}$  below, where the picture on the left shows the canonical model of  $\mathcal{O}, \mathcal{D}_i$  (see Proposition 9) and next to it is  $S^i$  (the omitted labels on transitions are  $\Sigma^{\bigcirc}$ ).

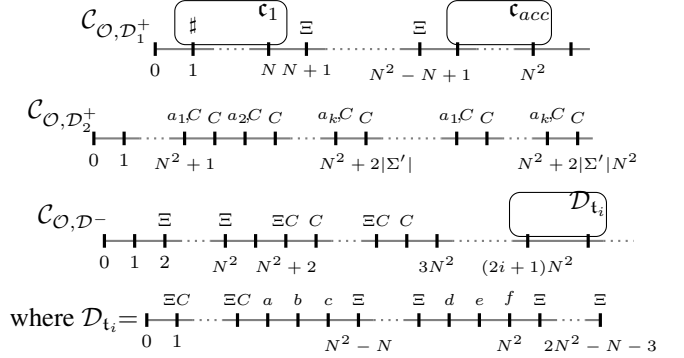


In general, the size of  $S^i$  is  $|\mathcal{D}_i| + O(2^{|\mathcal{O}|})$  and the product of  $S^i$ ,  $\mathcal{D}_i \in E^+$  is of size  $O(2^{|\mathcal{O}| + |E^+|})$ . The upper bounds now follow from P and PSPACE completeness of checking simulation and containment for transition systems.

Now we sketch the proof of the lower bound. Let  $M$  be a non-deterministic Turing machine that accepts  $\Sigma$ -words  $x = x_1 \dots x_n$  in  $N = 2^{poly(|x|)}$  steps and erases the tape after a successful computation. We represent configurations  $c$  of a

computation of  $M$  on  $x$  by an  $N - 1$ -long word (with sufficiently many blanks at the end), in which  $y$  in the active cell is replaced by  $(q, y)$  with the current state  $q \in Q$ . An accepting computation of  $M$  on  $x$  is encoded by the  $N^2$ -long word  $w = \#c_1 \#c_2 \# \dots \#c_{N-1} \#c_N$  over  $\Xi = \Sigma \cup (Q \times \Sigma) \cup \{\#\}$ . Thus, a word  $w$  of length  $N^2$  encodes an accepting computation iff it starts with the initial configuration  $c_1$  preceded by  $\#$ , ends with the accepting configuration  $c_{acc}$ , and every two length 3 subwords at distance  $N$  apart form a *legal tuple* [Sipser, 1997, Theorem 7.37].

We define  $\mathcal{O}$  and  $E = (\{\mathcal{D}_1^+, \mathcal{D}_2^+\}, \{\mathcal{D}^-\})$  so that their canonical models look as follows, for  $\Xi = \{a_1, \dots, a_k\}$ :



where  $\mathcal{D}_{t_i} = \{a, b, c, d, e, f\}$  is the lexicographically  $i$ -th illegal tuple. The parts of the canonical models shown above are of exponential size; however, due to their repetitive nature, they can be described by a polynomial-size  $LTL_{horn}^{\square\bigcirc}$  ontology  $\mathcal{O}$  as in [Ryzhikov *et al.*, 2021]. We show that the  $\mathcal{Q}_p[U]$ -query

$$\varkappa = \diamond(\rho_1 \wedge C U (\rho_2 \wedge C U (\dots (\rho_{N^2-1} \wedge (C U \rho_{N^2})) \dots))),$$

where  $\rho_1 \dots \rho_{N^2}$  encodes an accepting computation of  $M$  on  $x$ , is the only type of query that can separate  $E$  under  $\mathcal{O}$ .  $\dashv$

As for *data* complexity, we show that  $LTL_{horn}^{\square\bigcirc}$ -ontologies come essentially for free:

**Theorem 12.** *The results of Theorem 5 continue to hold for queries mediated by a fixed  $LTL_{horn}^{\square\bigcirc}$ -ontology.*

Intuitively, the reason is that, given a fixed  $LTL_{horn}^{\square\bigcirc}$ -ontology  $\mathcal{O}$ , we can compute the types of the canonical model  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$ , for consistent  $\mathcal{O}$  and  $\mathcal{D}$ , in polynomial time in  $\mathcal{D}$  by Proposition 9, with the length  $M$  from Section 4 being polynomial in  $E$ . Checking consistency of  $\mathcal{D}$  and fixed  $\mathcal{O}$  is known to be in P [Artale *et al.*, 2021].

## 6 QBE with $LTL^{\square\bigcirc}$ -Ontologies

In this section, we investigate separability by  $\diamond$ -queries under  $LTL^{\square\bigcirc}$ -ontologies. Remarkably, we show that, for data complexity,  $LTL^{\square\bigcirc}$ -ontologies also come for free despite admitting arbitrary Boolean operators; cf., [Schaerf, 1993].

**Theorem 13.** *Let  $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}[\diamond]\}$ . If  $E$  is  $\mathcal{Q}$ -separable under an  $LTL^{\square\bigcirc}$ -ontology  $\mathcal{O}$ , then  $E$  can be separated under  $\mathcal{O}$  by a  $\mathcal{Q}$ -query of polysize in  $E$  and  $\mathcal{O}$ .  $\text{QBE}(LTL^{\square\bigcirc}, \mathcal{Q})$  and  $\text{QBE}_{b^+}^{\square\bigcirc}(LTL^{\square\bigcirc}, \mathcal{Q})$  are  $\Sigma_2^p$ -complete for combined complexity. The presence of  $LTL^{\square\bigcirc}$ -ontologies has no effect on the data complexity, which remains the same as in Theorem 5.*

We comment on the proof of this theorem for  $\mathcal{Q}_p[\diamond]$ . Taking into account NP-completeness of checking if  $\mathcal{O}$  is consistent with  $\mathcal{D}$  and tractability of this problem for a fixed  $\mathcal{O}$  [Artale *et al.*, 2021], we can assume, as in Theorem 10, that  $\mathcal{O}$  and  $\mathcal{D}$  are consistent for each  $\mathcal{D} \in E^+ \cup E^-$ . Observe first that if  $E$  is separated by  $\varkappa \in \mathcal{Q}_p[\diamond]$  of the form (1) under an  $LTL^{\square\diamond}$ -ontology  $\mathcal{O}$ , then, as follows from [Ono and Nakamura, 1980], for any  $\mathcal{D} \in E^-$ , there is a model  $\mathcal{J}_{\mathcal{D}} \not\models \varkappa(0)$  of  $\mathcal{O}$  and  $\mathcal{D}$  whose types form a sequence

$$tp_0, \dots, tp_k, tp_{k+1}, \dots, tp_{k+l}, \dots, tp_{k+l}, \dots \quad (5)$$

with  $\max \mathcal{D} \leq k \leq \max \mathcal{D} + |\mathcal{O}|$  and  $l \leq |\mathcal{O}|$ . This allows us to find a separating  $\varkappa$  of polysize in  $E, \mathcal{O}$ . Indeed, let  $K$  be the maximal  $k$  in (5) over all  $\mathcal{D} \in E^-$ . If the depth  $n$  of  $\varkappa$  is  $\leq K$ , we are done. If  $n > K$ , we shorten  $\varkappa$  as follows. Consider the prefix  $\varkappa'$  of  $\varkappa$  formed by  $\rho_0, \dots, \rho_K$ . If  $\mathcal{J}_{\mathcal{D}} \not\models \varkappa'(0)$  for all  $\mathcal{D} \in E^-$ , we are done. Otherwise, for each  $\mathcal{D} \in E^-$ , we pick a  $\rho_i, i > K$ , with  $\rho_i \not\subseteq tp_{k+j}$  for any  $j \leq l$ ; it must exist as  $\mathcal{J}_{\mathcal{D}} \not\models \varkappa(0)$ . Then we construct  $\varkappa''$  by omitting from  $\varkappa$  all  $\rho_i$  that are different from those in  $\varkappa'$  and the chosen  $\rho_i$  with  $i > K$ . Clearly,  $\varkappa''$  is as required.

A  $\Sigma_2^p$ -algorithm guesses  $\varkappa$  and  $\mathcal{J}_{\mathcal{D}}$ , for  $\mathcal{D} \in E^-$ , and checks in polytime that  $\mathcal{J}_{\mathcal{D}} \models \mathcal{O}, \mathcal{D}$  and  $\mathcal{J}_{\mathcal{D}} \not\models \varkappa(0)$  and in cONP [Ono and Nakamura, 1980] that  $\mathcal{O}, \mathcal{D} \models \varkappa(0)$  for all  $\mathcal{D} \in E^+$ . The lower bound is shown by reduction of the validity problem for fully quantified Boolean formulas  $\exists p \forall q \psi$ , where  $p = p_1, \dots, p_k$  and  $q = q_1, \dots, q_m$  are all propositional variables in  $\psi$ . We can assume that  $\psi$  is not a tautology and  $\neg\psi \not\models x$  for  $x \in \{p_i, \neg p_i, q_j, \neg q_j \mid i \leq k, j \leq m\}$ . Let  $E = (E^+, E^-)$  with  $E^+ = \{\mathcal{D}_1, \mathcal{D}_2\}$ ,  $E^- = \{\mathcal{D}_3\}$ , where

$\mathcal{D}_1 = \{B_1(0)\}$ ,  $\mathcal{D}_2 = \{B_2(0)\}$ ,  $\mathcal{D}_3 = \{q_1(0), \dots, q_m(0)\}$ , and let  $\mathcal{O}$  contain the following axioms with fresh atoms  $B_1, B_2, A_i, \bar{A}_i$ , for  $i = 1, \dots, k$ :

$$\begin{aligned} B_1 \vee B_2 &\rightarrow \neg\psi, & p_i &\rightarrow \diamond(\bar{A}_i \wedge \bigwedge_{j \neq i} (A_j \wedge \bar{A}_j)), \\ & & \neg p_i &\rightarrow \diamond(A_i \wedge \bigwedge_{j \neq i} (A_j \wedge \bar{A}_j)). \end{aligned}$$

Then  $\exists p \forall q \psi$  is valid iff  $E$  is  $\mathcal{Q}_p[\diamond]$ -separable under  $\mathcal{O}$ .

We obtain the NP upper bounds in data complexity using the same argument as for the  $\Sigma_2^p$ -upper bound and observing that checking  $\mathcal{O}, \mathcal{D} \models \varkappa(0)$  is in P in data complexity. The NP lower bounds are inherited from the ontology-free case. The proof of the P upper bounds is more involved. We illustrate the idea for  $\mathcal{O}$  with arbitrary Boolean but without temporal operators. In this case, one can show (which is non-trivial) that  $\mathcal{O}, \mathcal{D} \models \varkappa(0)$  iff  $\mathcal{I}_{\mathcal{O}, \mathcal{D}} \models \varkappa(0)$ , where  $\mathcal{I}_{\mathcal{O}, \mathcal{D}}$  is the *completion* of  $\mathcal{D}$ : it contains  $A(\ell)$  iff  $\mathcal{O} \cup \{B \mid B(\ell) \in \mathcal{D}\} \models A$ . For example, if  $\mathcal{O} = \{A \vee B\}$  and  $\mathcal{D} = \{A(1), B(1), A(3), B(3)\}$ , the completion  $\mathcal{I}_{\mathcal{O}, \mathcal{D}}$  is just  $\mathcal{D}$  regarded as an interpretation (so  $\mathcal{I}_{\mathcal{O}, \mathcal{D}}$  does not have to be a model of  $\mathcal{O}$ ). It can be constructed in polytime in  $\mathcal{D}$  and, due to the equivalence above, used to prove the P upper bounds using dynamic programming. That equivalence does not hold for  $LTL^{\square\diamond}$ , but the technique can be extended by applying it to data sets enriched by certain types.

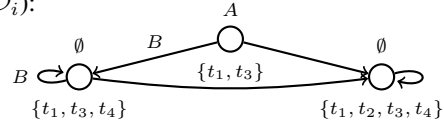
Note that the completion technique does not work for  $\circ, \diamond$ -queries. For example,  $\mathcal{O}, \mathcal{D} \models \diamond(A \wedge \circ B)$  for  $\mathcal{D}$  and  $\mathcal{O}$  defined above, and so the equivalence does not hold. In fact, the complexity of separability by  $\circ, \diamond$ -queries remains open.

## 7 QBE with LTL-Ontologies

For ontologies with arbitrary LTL-axioms, we obtain:

**Theorem 14.** (i)  $\text{QBE}(LTL, \mathcal{Q})$  is in 2EXPTIME, for any  $\mathcal{Q} \in \{\mathcal{Q}[\diamond], \mathcal{Q}[\circ, \diamond], \mathcal{Q}[U_s]\}$ . (ii)  $\text{QBE}(LTL, \mathcal{Q})$  is in 2EXPSpace, for any  $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}_p[\circ, \diamond], \mathcal{Q}_p[U]\}$ .

The proof requires a further modification of the transition systems  $S^i$  in Theorem 6. We illustrate it by an example. Let  $\mathcal{O} = \{A \rightarrow \diamond B, \top \rightarrow A \vee B, A \wedge B \rightarrow \perp\}$  with the set of  $\mathcal{O}$ -types  $\mathbf{T}_{\mathcal{O}} = \{tp_1, tp_2, tp_3\}$ , where  $tp_1 = \{A, \neg B, \diamond B\}$ ,  $tp_2 = \{\neg A, B, \neg \diamond B\}$ ,  $tp_3 = \{A, \neg B, \neg \diamond B\}$ , and  $tp_4 = \{\neg A, B, \diamond B\}$ , from which we omitted subformulas such as  $A \vee B$  that are true or false in all types. For non-empty sets  $\mathbf{T}_1, \mathbf{T}_2 \subseteq \mathbf{T}_{\mathcal{O}}$  and  $\Gamma \subseteq \Sigma^\perp$ , we take the relation  $\mathbf{T}_1 \rightarrow_\Gamma \mathbf{T}_2$ , which, intuitively, says that if there are instants  $n_{\mathcal{I}}$  in all models  $\mathcal{I}$  of  $\mathcal{O}, \mathcal{D}$  such that  $\{tp_{\mathcal{I}}(n_{\mathcal{I}}) \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}\} = \mathbf{T}_1$ , then there exist  $m_{\mathcal{I}} > n_{\mathcal{I}}$  with  $\{tp_{\mathcal{I}}(m_{\mathcal{I}}) \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}\} = \mathbf{T}_2$  and  $\Gamma = \{A \in \Sigma^\perp \mid \mathcal{I}, m \models A \text{ for all } \mathcal{I} \text{ and } n_{\mathcal{I}} < m < m_{\mathcal{I}}\}$ . In our example, we have  $\{tp_1, tp_3\} \rightarrow_{\Sigma^\perp} \{tp_1, tp_2, tp_3, tp_4\}$  and  $\{tp_1, tp_3\} \rightarrow_{\{B\}} \{tp_1, tp_3, tp_4\}$  (among others). Then we construct the following transition system  $S^i$  for, say,  $\mathcal{D}_i = \{A(0)\}$ , which reflects all  $\mathcal{Q}[U_s]$ -queries over  $\mathcal{O}, \mathcal{D}_i$  using  $\mathbf{T}' \subseteq \mathbf{T}_{\mathcal{O}}$  as states (the initial state is  $\{tp_1, tp_3\}$  since  $A(0) \in \mathcal{D}_i$ ):



The  $S^i$  can be constructed in 2EXPTIME in  $|\mathcal{D}_i| + |\mathcal{O}|$  (checking  $\mathbf{T}_1 \rightarrow_\Gamma \mathbf{T}_2$ , for given  $\mathbf{T}_1, \mathbf{T}_2$  and  $\Gamma$ , can be done in EXPSpace). Also, the product of the  $S^i$ , for  $\mathcal{D}_i \in E^+$ , can be constructed in 2EXPTIME in  $|\mathcal{D}_i| + |E^+|$ .

## 8 Conclusions

We have started an investigation of the computational complexity of query-by-example for principal classes of LTL-queries, both with and without mediating ontologies. Our results are encouraging as we exhibit important cases that are tractable for data complexity and not harder than satisfiability for combined complexity. Many interesting and technically challenging problems remain open. Especially intriguing are queries with U. For example, we still need to pinpoint the size of minimal separating  $\mathcal{Q}[U_s]$ - and  $\mathcal{Q}_p[U]$ -queries under a Horn ontology. The tight complexity of QBE for unrestricted U-queries is also open. In general, such queries could be too perplexing for applications; however, they can express useful disjunctive patterns such as ‘in at most  $n$  moments of time’. Note also that sparse data instances with large gaps between timestamps may require binary representations thereof, for which the proofs of some of our results do not go through.

Our results and techniques provide a good starting point for studying QBE with (ontology-mediated) queries over temporal databases with a full relational component [Chomicki *et al.*, 2001; Chomicki and Toman, 2018; Artale *et al.*, 2022] and also for the construction of separating queries satisfying additional conditions such as being a longest/shortest separator [Blum *et al.*, 2021; Fijalkow and Lagarde, 2021] or a most specific/general one [ten Cate *et al.*, 2022].

## Acknowledgements

This work was supported by EPSRC UK grants EP/S032207, EP/S032282, and EP/W025868.

## References

- [Arenas *et al.*, 2016] M. Arenas, G. I. Diaz, and E. Kostylev. Reverse engineering SPARQL queries. In *Proc. of WWW*, pages 239–249, 2016.
- [Artale *et al.*, 2021] A. Artale, R. Kontchakov, A. Kovtunova, V. Ryzhikov, F. Wolter, and M. Zakharyashev. First-order rewritability of ontology-mediated queries in linear temporal logic. *Artif. Intell.*, 299:103536, 2021.
- [Artale *et al.*, 2022] A. Artale, R. Kontchakov, A. Kovtunova, V. Ryzhikov, F. Wolter, and M. Zakharyashev. First-order rewritability and complexity of two-dimensional temporal ontology-mediated queries. *J. Artif. Intell. Res.*, 75:1223–1291, 2022.
- [Badea and Nienhuys-Cheng, 2000] L. Badea and S. Nienhuys-Cheng. A refinement operator for description logics. In *Proc. of ILP*, pages 40–59, 2000.
- [Barceló and Romero, 2017] P. Barceló and M. Romero. The complexity of reverse engineering problems for conjunctive queries. In *Proc. of ICDT*, pages 7:1–7:17, 2017.
- [Bergroth *et al.*, 2000] L. Bergroth, H. Hakonen, and T. Raita. A survey of longest common subsequence algorithms. In *Proc. of SPIRE*, pages 39–48, 2000.
- [Blum *et al.*, 2021] C. Blum, M. Djukanovic, A. Santini, H. Jiang, C. Li, F. Manyà, and G. Raidl. Solving longest common subsequence problems via a transformation to the maximum clique problem. *Comput. Oper. Res.*, 125:105089, 2021.
- [Bühmann *et al.*, 2016] L. Bühmann, J. Lehmann, and P. Westphal. DL-learner - A framework for inductive learning on the semantic web. *J. Web Sem.*, 39:15–24, 2016.
- [Camacho and McIlraith, 2019] A. Camacho and S. McIlraith. Learning interpretable models expressed in linear temporal logic. In *Proc. of ICAPS*, pages 621–630, 2019.
- [Chen and Lin, 1994] C.-C. Chen and I-Peng Lin. The computational complexity of the satisfiability of modal Horn clauses for modal propositional logics. *Theor. Comp. Sci.*, 129(1):95–121, 1994.
- [Chomicki and Toman, 2018] J. Chomicki and D. Toman. Temporal logic in database query languages. In *Encyclopedia of Database Systems*. Springer, 2018.
- [Chomicki *et al.*, 2001] J. Chomicki, D. Toman, and M. Böhlen. Querying ATSQL databases with temporal logic. *ACM Trans. Database Syst.*, 26(2):145–178, 2001.
- [Chowdhury *et al.*, 2010] R. Chowdhury, H.-S. Le, and V. Ramachandran. Cache-oblivious dynamic programming for bioinformatics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 7(3):495–510, 2010.
- [Cima *et al.*, 2021] G. Cima, F. Croce, and M. Lenzerini. Query definability and its approximations in ontology-based data management. In *Proc. of CIKM*, pages 271–280. ACM, 2021.
- [Cohen and Weiss, 2016] S. Cohen and Y. Weiss. The complexity of learning tree patterns from example graphs. *ACM Trans. Database Syst.*, 41(2):14:1–14:44, 2016.
- [Demri *et al.*, 2016] S. Demri, V. Goranko, and M. Lange. *Temporal Logics in Computer Science*. Cambridge University Press, 2016.
- [Deutch and Gilad, 2019] D. Deutch and A. Gilad. Reverse-engineering conjunctive queries from provenance examples. In *Proc. of EDBT*, pages 277–288, 2019.
- [Fanizzi *et al.*, 2018] N. Fanizzi, G. Rizzo, C. d’Amato, and F. Esposito. DLFoil: Class expression learning revisited. In *Proc. of EKAW*, pages 98–113, 2018.
- [Fijalkow and Lagarde, 2021] N. Fijalkow and G. Lagarde. The complexity of learning linear temporal formulas from examples. In *Proc. of 15th Int. Conf. on Grammatical Inference*, volume 153 of PMLR, pages 237–250, 2021.
- [Fortin *et al.*, 2022] M. Fortin, B. Konev, V. Ryzhikov, Y. Savateev, F. Wolter, and M. Zakharyashev. Unique characterisability and learnability of temporal instance queries. In *Proc. of KR*, 2022.
- [Fraser, 1996] C. Fraser. Consistent subsequences and supersequences. *Theor. Comput. Sci.*, 165(2):233–246, 1996.
- [Funk *et al.*, 2019] M. Funk, J. C. Jung, C. Lutz, H. Pulcini, and F. Wolter. Learning description logic concepts: When can positive and negative examples be separated? In *Proc. of IJCAI*, pages 1682–1688, 2019.
- [Gabbay *et al.*, 2003] D. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyashev. *Many-Dimensional Modal Logics: Theory and Applications*, volume 148 of *Studies in Logic*. Elsevier, 2003.
- [Gutiérrez-Basulto *et al.*, 2018] V. Gutiérrez-Basulto, J. Ch. Jung, and L. Sabellek. Reverse engineering queries in ontology-enriched systems: The case of expressive Horn description logic ontologies. In *Proc. of IJCAI*, 2018.
- [Heindorf *et al.*, 2022] S. Heindorf, L. Blübaum, N. Düsterhus, T. Werner, V. N. Golani, C. Demir, and A.-C. Ngonga Ngomo. Evolearner: Learning description logics with evolutionary algorithms. In *Proc. of WWW*, pages 818–828. ACM, 2022.
- [Hofman and Martens, 2015] P. Hofman and W. Martens. Separability by short subsequences and subwords. In *Proc. of ICDT*, volume 31 of *LIPICs*, pages 230–246, 2015.
- [Iannone *et al.*, 2007] L. Iannone, I. Palmisano, and N. Fanizzi. An algorithm based on counterfactuals for concept learning in the Semantic Web. *Appl. Intell.*, 26(2):139–159, 2007.
- [Jung *et al.*, 2020] J. Ch. Jung, C. Lutz, and F. Wolter. Least General Generalizations in Description Logic: Verification and Existence. In *Proc. of AAAI*, 2020.
- [Jung *et al.*, 2021] J. Ch. Jung, C. Lutz, H. Pulcini, and F. Wolter. Separating data examples by description logic concepts with restricted signatures. In *Proc. of KR*, 2021.



- [Jung *et al.*, 2022] J. Ch. Jung, C. Lutz, H. Pulcini, and F. Wolter. Logical separability of labeled data examples under ontologies. *Artif. Intell.*, 313 (2022).
- [Kalashnikov *et al.*, 2018] D. Kalashnikov, L. Lakshmanan, and D. Srivastava. Fastqre: Fast query reverse engineering. In *Proc. of SIGMOD*, pages 337–350, 2018.
- [Kupferman and Vardi, 1996] O. Kupferman and M. Vardi. Verification of fair transisiton systems. In *Proc. of CAV, volume 1102 of LNCS*, pages 372–382. Springer, 1996.
- [Kurtonina and de Rijke, 1997] N. Kurtonina and M. de Rijke. Bisimulations for temporal logic. *J. Log. Lang. Inf.*, 6(4):403–425, 1997.
- [Lehmann and Haase, 2009] J. Lehmann and Ch. Haase. Ideal downward refinement in the  $\mathcal{EL}$  description logic. In *Proc. of ILP*, pages 73–87, 2009.
- [Lehmann and Hitzler, 2010] J. Lehmann and P. Hitzler. Concept learning in description logics using refinement operators. *Machine Learning*, 78:203–250, 2010.
- [Lemieux *et al.*, 2015] C. Lemieux, D. Park, and I. Beschastnikh. General LTL specification mining (T). In *Proc. of ASE*, pages 81–92. IEEE, 2015.
- [Lisi and Straccia, 2015] F. Lisi and U. Straccia. Learning in description logics with fuzzy concrete domains. *Fundamenta Informaticae*, 140(3-4):373–391, 2015.
- [Lisi, 2012] F. Lisi. A formal characterization of concept learning in description logics. In *Proc. of DL*, 2012.
- [Maier, 1978] D. Maier. The complexity of some problems on subsequences and supersequences. *J. ACM*, 25(2):322–336, 1978.
- [Martins, 2019] D. Martins. Reverse engineering database queries from examples: State-of-the-art, challenges, and research opportunities. *Inf. Syst.*, 83:89–100, 2019.
- [Neider and Gavran, 2018] D. Neider and I. Gavran. Learning linear temporal properties. In *Proc. of FMCAD*, pages 1–10. IEEE, 2018.
- [Ono and Nakamura, 1980] H. Ono and A. Nakamura. On the size of refutation Kripke models for some linear modal and tense logics. *Studia Logica*, pages 325–333, 1980.
- [Ortiz, 2019] M. Ortiz. Ontology-mediated queries from examples: a glimpse at the DL-Lite case. In *Proc. of GCAI*, pages 1–14, 2019.
- [Place and Zeitoun, 2016] T. Place and M. Zeitoun. Separating regular languages with first-order logic. *Log. Methods Comput. Sci.*, 12(1), 2016.
- [Place and Zeitoun, 2022] T. Place and M. Zeitoun. A generic polynomial time approach to separation by first-order logic without quantifier alternation. In *Proc. of FSTTCS*, volume 250 of *LIPICs*, pages 43:1–43:22, 2022.
- [Prior, 1956] A. Prior. *Time and Modality*. OUP, 1956.
- [Raha *et al.*, 2022] R. Raha, R. Roy, N. Fijalkow, and D. Neider. Scalable anytime algorithms for learning fragments of linear temporal logic. In *Proc. of TACAS*, volume 13243 of *LNCS*, pages 263–280. Springer, 2022.
- [Rizzo *et al.*, 2018] G. Rizzo, N. Fanizzi, C. d’Amato, and F. Esposito. A framework for tackling myopia in concept learning on the web of data. In *Proc. of EKAW*, pages 338–354. Springer, 2018.
- [Rizzo *et al.*, 2020] G. Rizzo, N. Fanizzi, and C. d’Amato. Class expression induction as concept space exploration: From DL-FOIL to DL-FOCL. *Future Gener. Comput. Syst.*, 108:256–272, 2020.
- [Ryzhikov *et al.*, 2021] V. Ryzhikov, Y. Savateev, and M. Zakharyashev. Deciding FO-rewritability of ontology-mediated queries in linear temporal logic. In *Proc. of TIME*, volume 206 of *LIPICs*, pages 10:1–10:15, 2021.
- [Sarker and Hitzler, 2019] Md. K. Sarker and P. Hitzler. Efficient concept induction for description logics. In *Proc. of AAI*, pages 3036–3043, 2019.
- [Schaerf, 1993] A. Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. *J. Intel. Inf. Sys.*, 2:265–278, 1993.
- [Sipser, 1997] M. Sipser. *Introduction to the theory of computation*. PWS Publishing Company, 1997.
- [Staworko and Wieczorek, 2012] S. Staworko and P. Wieczorek. Learning twig and path queries. In *Proc. of ICDT*, pages 140–154, 2012.
- [Straccia and Mucci, 2015] U. Straccia and M. Mucci. pFOIL-DL: Learning (fuzzy)  $\mathcal{EL}$  concept descriptions from crisp OWL data using a probabilistic ensemble estimation. In *Proc. of SAC*, pages 345–352, 2015.
- [ten Cate and Dalmau, 2015] B. ten Cate and V. Dalmau. The product homomorphism problem and applications. In *Proc. of ICDT*, pages 161–176, 2015.
- [ten Cate *et al.*, 2022] B. ten Cate, V. Dalmau, M. Funk, and C. Lutz. Extremal fitting problems for conjunctive queries. *CoRR*, abs/2206.05080, 2022.
- [Tran *et al.*, 2017] An C. Tran, J. Dietrich, H. W. Guesgen, and S. Marsland. Parallel symmetric class expression learning. *J. Mach. Learn. Res.*, 18:64:1–64:34, 2017.
- [Vardi, 2007] M. Vardi. Automata-theoretic techniques for temporal reasoning. In *Handbook of Modal Logic*, pages 971–989. North-Holland, 2007.
- [Weiss and Cohen, 2017] Y. Weiss and S. Cohen. Reverse engineering SPJ-queries from examples. In *Proc. of PODS*, pages 151–166, 2017.
- [Willard, 2010] R. Willard. Testing expressibility is hard. In *Proc. of CP*, vol. 6308 of *LNCS*, pp. 9–23. Springer, 2010.
- [Zhang *et al.*, 2013] M. Zhang, H. Elmeleegy, C. Procopiuc, and D. Srivastava. Reverse engineering complex join queries. In *Proc. of SIGMOD*, pages 809–820, 2013.
- [Zloof, 1977] M. Zloof. Query-by-example: A data base language. *IBM Syst. J.*, 16(4):324–343, 1977.

## Appendix: Proofs

### A Proofs for Section 3

**Theorem 4.** *The following polynomial-time reductions hold:*

- (i.1)  $\text{QBE}(\mathcal{L}, \mathcal{Q}) \leq_p \text{QBE}_{1-}(\mathcal{L}, \mathcal{Q})$ , for any  $\mathcal{Q}$  closed under  $\wedge$ , and any  $\mathcal{L}$  (including  $\mathcal{L} = \emptyset$ ),
- (i.2)  $\text{QBE}(\mathcal{L}, \mathcal{Q}) \leq_p \text{QBE}^{2+}(\mathcal{L}, \mathcal{Q})$ , for  $\mathcal{L} \in \{\text{LTL}, \text{LTL}^{\square\Diamond}\}$ ,
- (i.3)  $\text{QBE}(\mathcal{L}, \mathcal{Q}[\circ, \diamond]) \leq_p \text{QBE}(\mathcal{L}, \mathcal{Q}_p^\circ[\diamond])$  and  $\text{QBE}(\mathcal{L}, \mathcal{Q}[\diamond]) \leq_p \text{QBE}(\mathcal{L}, \mathcal{Q}_p[\diamond])$ , for any  $\mathcal{L}$ ,
- (ii.1)  $\text{QBE}(\mathcal{Q}_p[\diamond]) \leq_p \text{QBE}(\mathcal{Q}_p[\circ, \diamond])$  and  $\text{QBE}(\mathcal{Q}_p[\diamond]) \leq_p \text{QBE}(\mathcal{Q}_p[\text{U}]) \leq_p \text{QBE}_{1-}(\mathcal{Q}_p[\text{U}])$
- (ii.2)  $\text{QBE}(\mathcal{Q}[\circ, \diamond]) =_p \text{QBE}(\mathcal{Q}[\diamond]) \leq_p \text{QBE}(\mathcal{Q}[\text{U}_s])$ .

*Reductions (i.1)–(i.3) work for combined complexity and (i.1) and (i.3) also work for data complexity. The reductions preserve boundedness of the number of positive/negative example.*

*Proof.* (i.1) Observe that if  $\mathcal{Q}$  is closed under  $\wedge$ ,  $E = (E^+, E^-)$  and  $E^- = \{\mathcal{D}_1^-, \dots, \mathcal{D}_n^-\}$ , then  $E$  is  $\mathcal{Q}$ -separable under  $\mathcal{O}$  iff each  $(E^+, \{\mathcal{D}_i^-\})$ ,  $1 \leq i \leq n$ , is. Indeed, if  $\varkappa_i$  separates  $(E^+, \{\mathcal{D}_i^-\})$ , then  $\varkappa_1 \wedge \dots \wedge \varkappa_n$  separates  $E$ . For such  $\mathcal{Q}$ , we can thus assume that  $E^-$  consists of a single data instance. Note that  $\mathcal{Q}_p[\circ, \diamond]$  and  $\mathcal{Q}_p[\text{U}]$  are not closed under  $\wedge$ .

(i.2) Let  $E^+ = \{\mathcal{D}_1, \dots, \mathcal{D}_n\}$  and let  $k = \max_i \max \mathcal{D}_i$ . We construct an ontology  $\mathcal{O}'$  by taking the fresh atoms  $A_1, \dots, A_n, C_0, \dots, C_k, D_0, \dots, D_k, S_1, S_2$  and adding the following axioms to the given ontology  $\mathcal{O}$ :

$$\begin{aligned} S_1 &\rightarrow A_1 \vee \dots \vee A_n, & S_2 &\rightarrow A_1 \vee \dots \vee A_n, \\ C_i \wedge \diamond A_j &\rightarrow X, & D_i \wedge \diamond A_j &\rightarrow X, \text{ for } X(i) \in \mathcal{D}_j. \end{aligned}$$

Let  $E'^+$  consist of  $\mathcal{D}'_1 = \{C_0(0), \dots, C_k(n'), S_1(n' + 1)\}$  and  $\mathcal{D}'_2 = \{D_0(0), \dots, D_k(n'), S_2(n' + 1)\}$ . Then every model of  $\mathcal{O}'$ ,  $\mathcal{D}'_1$  or  $\mathcal{O}'$ ,  $\mathcal{D}'_2$  contains a model of at least one of the  $\mathcal{O}, \mathcal{D}_i$  and, conversely, every model of any  $\mathcal{O}, \mathcal{D}_i$  can be converted into a model of  $\mathcal{O}'$ ,  $\mathcal{D}'_1$  or  $\mathcal{O}'$ ,  $\mathcal{D}'_2$  by adding only the newly introduced symbols. So, if there is a separating query for  $(E^+, E^-)$  under the ontology  $\mathcal{O}$ , then the same query separates  $(E'^+, E^-)$  under the ontology  $\mathcal{O}'$ . And if there is a separating query for  $(E'^+, E^-)$  under the ontology  $\mathcal{O}'$ , then it cannot contain any symbols from  $\text{sig}(\mathcal{O}') \setminus \text{sig}(\mathcal{O})$ , and so it separates  $(E^+, E^-)$  under the ontology  $\mathcal{O}$ .

(i.3) Recall from the main part of the paper that  $[\rho_0 \wedge \diamond(\rho_1 \wedge \bigwedge_i \diamond \varkappa_i)] \equiv [\rho_0 \wedge \bigwedge_i \diamond(\rho_1 \wedge \diamond \varkappa_i)]$ ,  $\circ \diamond \varkappa \equiv \diamond \circ \varkappa$  and  $\circ(\varkappa \wedge \varkappa') \equiv (\circ \varkappa \wedge \circ \varkappa')$ , and so each  $\mathcal{Q}[\circ, \diamond]$ -query can be equivalently transformed in polynomial time into a conjunction of  $\mathcal{Q}_p^\circ[\diamond]$ -queries. If  $E^-$  is a singleton, then a conjunction of queries in  $\mathcal{Q}_p^\circ[\diamond]$  separates  $(E^+, E^-)$  under an ontology  $\mathcal{O}$  iff a single conjunct separates  $(E^+, E^-)$  under  $\mathcal{O}$ . Thus, there is  $\mathbf{q} \in \mathcal{Q}[\circ, \diamond]$  separating an arbitrary  $(E^+, E^-)$  under  $\mathcal{O}$  iff there are polysize  $\mathbf{q}_{\mathcal{D}} \in \mathcal{Q}_p^\circ[\diamond]$  separating  $(E^+, \{\mathcal{D}\})$  under  $\mathcal{O}$ , for each  $\mathcal{D} \in E^-$ . The second reduction is obtained by dropping  $\circ$  from the argument above.

In (ii.1), the first two reductions are proved by adding to  $E^+ \ni \mathcal{D}$ , for some  $\mathcal{D}$ , the data instance  $\mathcal{D}' = \{A(mn) \mid A(n) \in \mathcal{D}\}$  with  $m = \max \mathcal{D} + 2$ . Now, if  $\mathcal{D} \models \varkappa(0)$  and  $\mathcal{D}' \models \varkappa(0)$ , for  $\varkappa \in \mathcal{Q}_p[\text{U}]$ , then  $\varkappa$  is equivalent to a  $\mathcal{Q}_p[\diamond]$ -query.

For the third reduction, we observe first that without loss of generality one can assume that the positive examples do not contain atoms of the form  $X(0)$ . Indeed, suppose  $E = (E^+, E^-)$ ,  $E^+ = \{\mathcal{D}_1^+, \dots, \mathcal{D}_n^+\}$ , and  $E^- = \{\mathcal{D}_1^-, \dots, \mathcal{D}_k^-\}$ . Let  $\rho = \bigcap_{i=1}^n \{X \mid X(0) \in \mathcal{D}_i^+\}$ ,  $E'^- = \{\mathcal{D} \in E^- \mid \mathcal{D} \not\models \rho\}$ , and let  $D = (D^+, D^-)$ , where  $D^+ = \{\mathcal{D} \setminus \{X(0) \mid X \in \text{sig}(E)\} \mid \mathcal{D} \in E^+\}$  and  $D^- = \{\mathcal{D} \setminus \{X(0) \mid X \in \text{sig}(E)\} \mid \mathcal{D} \in E'^-\}$ . If  $\varphi = \rho_0 \wedge \lambda_1 \cup (\rho_1 \wedge \lambda_2 \cup (\dots (\rho_{l-1} \wedge (\lambda_l \cup \rho_l)) \dots))$  separates  $E$  then  $\varphi'$ , which is  $\varphi$  with  $\rho_0$  replaced by  $\emptyset$ , separates  $D$ . If  $\psi = \rho_0 \wedge \lambda_1 \cup (\rho_1 \wedge \lambda_2 \cup (\dots (\rho_{l-1} \wedge (\lambda_l \cup \rho_l)) \dots))$  separates  $D$ , then  $\rho_0 = \emptyset$  and  $\psi'$ , which is  $\psi$  with  $\rho_0$  replaced by  $\rho$ , separates  $E$ .

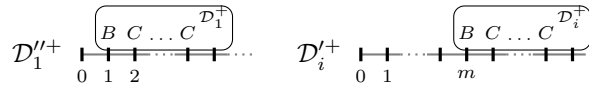
Now, suppose  $E = (E^+, E^-)$ ,  $E^+ = \{\mathcal{D}_1^+, \dots, \mathcal{D}_n^+\}$ ,  $n > 1$ , and  $E^- = \{\mathcal{D}_1^-, \dots, \mathcal{D}_k^-\}$ , and  $X(0) \notin \mathcal{D}_i^+$ , for any  $X$  and  $i$ . Let  $B, C$  be fresh atoms and  $m = \max_{\mathcal{D} \in E^+ \cup E^-} (\max(\mathcal{D})) + 2$ . We set  $E'^+ = \{\mathcal{D}_1'^+, \mathcal{D}_2'^+, \dots, \mathcal{D}_n'^+\}$ , where

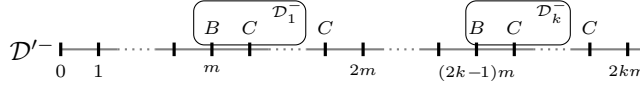
$$\begin{aligned} \mathcal{D}_i'^+ &= \{X(j+m) \mid X(j) \in \mathcal{D}_i^+ \} \cup \{B(m)\} \cup \{C(j) \mid m < j < m + \max(\mathcal{D}_i^+)\}, \\ \mathcal{D}_1'^+ &= \{X(j+1) \mid X(j) \in \mathcal{D}_1^+ \} \cup \{B(1)\} \cup \{C(j) \mid 1 < j < 1 + \max(\mathcal{D}_1^+)\}. \end{aligned}$$

We also set  $E'^- = \{\mathcal{D}'^-\}$  with

$$\mathcal{D}'^- = \{X((2i-1)m+j) \mid X(j) \in \mathcal{D}_i^-, i \in [1, k]\} \cup \{B((2i-1)m) \mid i \in [1, k]\} \cup \{C(i) \mid (2i-1)m < i < 2im, i \in [1, k]\}.$$

See the picture below for an illustration.





Let  $E = (E'^+, E'^-)$ . We prove equiseparability of  $E$  and  $E'$ .

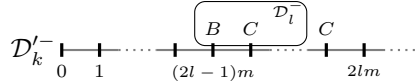
( $\Rightarrow$ ) Suppose  $\varphi = \rho_0 \wedge \lambda_1 \cup (\rho_1 \wedge \lambda_2 \cup (\dots (\rho_{l-1} \wedge (\lambda_l \cup \rho_l)) \dots))$  with  $l < m$  separates  $E$ . Then  $\rho_0 = \top$  (since  $X(0) \notin \mathcal{D}_i^+$ , for any  $X$ ). Consider the query  $\diamond\varphi'$ , where  $\varphi'$  is  $\varphi$  in which  $\rho_0$  is replaced by  $B$  and  $\lambda_i \neq \perp$  are replaced with  $\lambda_i \wedge C$ . Since  $\mathcal{D}_j^+ \models \varphi$  we have  $\mathcal{D} \models \varphi'$  for all  $\mathcal{D} \in E'^+$ . If  $\mathcal{D}'^- \models \diamond\varphi'$ , then  $\mathcal{D}'^- \models \varphi'((2j-1)m)$  for some  $j$ . Since all  $\lambda_i$ 's contain  $C$  and  $l < m$ , we have  $\mathcal{D}_j^- \models \varphi$ , which is impossible. Therefore,  $\diamond\varphi'$  separates  $E'$ .

( $\Leftarrow$ ) Suppose that  $\psi = \rho_0 \wedge \lambda_1 \cup (\rho_1 \wedge \lambda_2 \cup (\dots (\rho_{l-1} \wedge (\lambda_l \cup \rho_l)) \dots))$  with  $\rho_l \neq \top$  separates  $E'$ . Since  $\mathcal{D}_1'^+ \models \psi$ , we have  $l < m$ . Find the smallest  $i$  such that  $\rho_i \neq \top$ . As  $\mathcal{D}_2'^+ \models \psi$ , there is  $i' \leq i$  with  $\lambda_{i'} = \top$ .

If  $B \in \rho_i$  then, since  $\mathcal{D}_1'^+ \models \psi$ , we have  $i = 1$  and  $\lambda_1 = \top$ . Let  $\psi' = \lambda_2' \cup (\rho_2' \wedge \lambda_3' \cup (\dots (\rho_{l-1}' \wedge (\lambda_l \cup \rho_l)) \dots))$  where  $\rho_j' = \rho_j \setminus \{C\}$  and  $\lambda_j' = \lambda_j \setminus \{C\}$ . We see that in this case  $\mathcal{D}_j'^+ \models \psi'(m)$ , and so  $\mathcal{D}_j^+ \models \psi'$  for all  $j$ . Also since  $\mathcal{D}'^- \not\models \psi$ , we have  $\mathcal{D}_j^- \not\models \psi'$ , and so  $\psi'$  separates  $E$ .

If  $B \notin \rho_i$ , let  $\psi' = \diamond(\rho_i \wedge \lambda_{i+1} \cup (\rho_{i+1} \wedge \lambda_{i+2} \cup (\dots (\rho_{l-1} \wedge (\lambda_l \cup \rho_l)) \dots)))$ . Then  $\mathcal{D}^+ \models \psi'$  for all  $\mathcal{D}^+ \in E'^+$ . If  $\mathcal{D}'^- \models \psi'$ , then we have  $\mathcal{D}'^- \models \psi$  as  $\lambda_{i'} = \top$ , and so  $\psi'$  also separates  $E'$ . Consider the instance  $\mathcal{D}_l'^-$  (corresponding to some  $\mathcal{D}_l^-$ ) shown below:

$$\mathcal{D}_l'^- = \{X((2l-1)m + j) \mid X(j) \in \mathcal{D}_l^-\} \cup \{B((2l-1)m)\} \cup \{C(j) \mid (2l-1)m < j < 2lm\}.$$



Since  $\mathcal{D}_l'^- \subseteq \mathcal{D}'^-$ , we have  $\mathcal{D}_l'^- \not\models \psi'$ , and so  $\mathcal{D}_l^- \not\models \psi''$ , where  $\psi''$  is  $\psi'$  with the  $\rho_j$  replaced by  $\rho_j \setminus \{C\}$  and  $\lambda_j$  replaced by  $\lambda_j \setminus \{B, C\}$ , for  $\lambda_j \neq \emptyset$ . Clearly,  $\mathcal{D}^+ \models \psi''$  for all  $\mathcal{D}^+ \in E^+$ , and so  $\psi''$  separates  $E$ .

(ii.2) To show  $\text{QBE}(\mathcal{Q}[\circ, \diamond]) \leq_p \text{QBE}(\mathcal{Q}[\diamond])$ , suppose that  $E = (E^+, E^-)$  is given. Let  $m$  be the maximum over all  $\max \mathcal{D}$  with  $\mathcal{D} \in E^+$ . Introduce, for every  $A$  such that  $A(\ell) \in \mathcal{D}$  for some  $\mathcal{D} \in E^+$ , a fresh atom  $A_k$ ,  $0 < k \leq m$ , and extend any  $\mathcal{D} \in E^+ \cup E^-$  to a data instance  $\mathcal{D}'$  by adding  $A_k(\ell)$  to  $\mathcal{D}$  if  $A(k+\ell) \in \mathcal{D}$ . Let  $F^+ = \{\mathcal{D}' \mid \mathcal{D} \in E^+\}$  and  $F^- = \{\mathcal{D}' \mid \mathcal{D} \in E^-\}$ . Then clearly  $E$  is  $\mathcal{Q}[\diamond, \circ]$ -separable iff  $F$  is  $\mathcal{Q}[\diamond]$ -separable.

The converse reduction and  $\text{QBE}(\mathcal{Q}[\diamond]) \leq_p \text{QBE}(\mathcal{Q}[u_s])$  are proved similarly to (ii.1).  $\dashv$

## B Proofs for Section 4

We show the complexity results in Table 1. To this end, we first introduce some notation for sequence problems. Let  $\Sigma$  be an alphabet of symbols. A *word over*  $\Sigma$  is a finite sequence of symbols from  $\Sigma$ . A word  $\alpha$  is a *subsequence* of a word  $\beta$  if  $\alpha$  can be obtained from  $\beta$  by removing zero or more symbols anywhere in  $\beta$ . For a set  $S$  of words, we call a word  $\alpha$  a *common subsequence of*  $S$  if it is a subsequence of every word in  $S$ . The *consistent subsequence problem (CSSP)* is formulated as follows:

**Given:** sets  $S^+$  and  $S^-$  of words over an alphabet  $\Sigma$ .

**Problem:** decide whether there exists a common subsequence of  $S^+$  that is a not subsequence of any word in  $S^-$ .

The following is shown in [Fraser, 1996]:

**Theorem 15.** (i) *CSSP is NP-hard even if both the alphabet and  $S^+$  have cardinality two.*

(ii) *CSSP is NP-hard even if  $S^-$  is a singleton.*

Another problem of interest for us is the following *common subsequence problem (KsubS)*:

**Given:** a set  $S$  of words over an alphabet  $\Sigma$  and a number  $k$ .

**Problem:** decide whether there exists a common subsequence of  $S$  of length at least  $k$ .

The following is shown in [Maier, 1978]:

**Theorem 16.** *KsubS is NP-hard even if the alphabet has cardinality 2.*

We are now in a position to prove the results for  $\circ\diamond$ -queries in Table 1. We start by proving the NP-lower bounds for  $\mathcal{Q}_p[\diamond]$ . We actually show a slightly stronger result than in the table.

**Lemma 17.**  *$\text{QBE}(\mathcal{Q}_p[\diamond])$  with two positive examples or a single negative example is NP-hard.*

*Proof.* The proof by polynomial-time reduction of CSSP (as formulated in Theorem 15) is trivial. It is also of interest to give a proof of the second claim (a single negative example) via a polynomial-time reduction of KsubS. The proof also works directly for  $\mathcal{Q}_p^\circ[\diamond]$ -queries. Suppose that an instance  $S, k$  of KsubS over alphabet  $\{A, B\}$  is given. We define  $E$  of the form  $(E^+, \{\mathcal{D}^-\})$  such that the following conditions are equivalent:

- there exists a common subsequence of  $S$  of length  $k$ ;
- there exists  $q \in \mathcal{Q}_p[\diamond]$  that separates  $(E^+, \{\mathcal{D}^-\})$ ;
- there exists  $q \in \mathcal{Q}_p^\circ[\diamond]$  that separates  $(E^+, \{\mathcal{D}^-\})$ .

We represent each word  $w \in S$  as a data instance  $\mathcal{D}_w$  starting at time point 1 (for example, the word  $w = ABBA$  is represented as  $\mathcal{D}_w = \{A(1), B(2), B(3), A(4)\}$ ). Now let

$$\mathcal{D}^+ = \{A(i(k+2)), B(i(k+2)) \mid 1 \leq i \leq k\}$$

and

$$\mathcal{D}^- = \mathcal{D}^+ \setminus \{A(k(k+2)), B(k(k+2))\}$$

and let  $E^+ = \{\mathcal{D}_w \mid w \in S\} \cup \{\mathcal{D}^+\}$ . Assume first that there exists a common subsequence  $C_1 \cdots C_k$  of  $S$  of length  $k$ . Then  $\diamond(C_1 \wedge \diamond(C_2 \wedge \cdots \wedge \diamond C_k))$  separates  $(E^+, \{\mathcal{D}^-\})$ . Now assume that a query

$$\varkappa = \rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \cdots \wedge \diamond \rho_n)),$$

where every  $\rho_i$  is a  $\mathcal{Q}[\circ]$ -query, separates  $(E^+, \{\mathcal{D}^-\})$ . As  $\mathcal{D}_w \models \varkappa(0)$  for some  $w \in S$ , we have that  $n \leq k$  and all  $\rho_i$  have depth bounded by  $k$ . Then  $\rho_0 = \top$  and also, as there are ‘gaps’ of length  $k+1$  between any two entries in  $\mathcal{D}^+$  and since  $\mathcal{D}^+ \models \varkappa(0)$  we may assume that each  $\rho_i$ ,  $i > 0$ , is of the form  $\circ^{m_i} \rho'_i$  with  $0 \leq m_i \leq k$  and  $\rho'_i$  a conjunction of atoms. Observe that we can satisfy, in  $\mathcal{D}^+$ ,

- $\rho_1$  in the interval  $\{1, \dots, k+2\}$ ;
- $\rho_2$  in the interval  $\{(k+1)+1, \dots, 2(k+2)\}$ ;
- and so on, with  $\rho_n$  satisfied in the interval  $\{(n-1)(k+2)+1, \dots, n(k+2)\}$ .

In particular, if  $\rho_i$  is a conjunction of atoms, then it can be satisfied in  $i(k+2)$ . If  $n < k$ , then it follows directly that  $\mathcal{D}^- \models \varkappa(0)$ , and we have derived a contradiction. Hence  $n = k$ . Then, as the depth of  $\varkappa$  is bounded by  $k$ ,  $\rho_k$  is a conjunction of atoms. In fact, one can now show by induction starting with  $\rho_{k-1}$  that all  $\rho_i$ ,  $i > 0$ , are nonempty conjunctions of atoms. Otherwise a shift to the left shows that  $\mathcal{D}^- \models \varkappa(0)$  and we have derived a contradiction. Thus  $\varkappa$  takes the form  $\diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \cdots \wedge \diamond \rho_k))$  with all  $\rho_i$  non-empty. It follows from  $\mathcal{D}_w \models \varkappa(0)$  for all  $w \in S$  that  $\varkappa$  defines a common subsequence of  $S$  of length  $k$ , as required.  $\dashv$

The NP-lower bound for  $\text{QBE}(\mathcal{Q}_p[\circ, \diamond])$  with a bounded number of positive examples or a single negative example follows from Lemma 17 and Theorem 4 (ii.1).

We next obtain the NP-lower bound for  $\text{QBE}(\mathcal{Q}[\diamond])$  with a single negative example from Lemma 17 by observing that it follows from the proof of the first part of Theorem 4 (i.3) that any  $(E^+, E^-)$  with  $E^-$  a singleton is  $\mathcal{Q}_p[\diamond]$ -separable if, and only if, it is  $\mathcal{Q}[\diamond]$ -separable. The NP-lower bound for  $\text{QBE}(\mathcal{Q}[\circ, \diamond])$  with a single negative example follows from the NP-lower bound for  $\text{QBE}(\mathcal{Q}[\diamond])$  with a single negative example using Theorem 4 (ii.2).

We come to the NP-upper bounds. Recall that a query language  $\mathcal{Q}$  has the *polynomial separation property (PSP)* under an ontology language  $\mathcal{L}$  if any  $\mathcal{Q}$ -separable example is separated by a query in  $\mathcal{Q}$  of polynomial size. The NP-upper bounds for query languages using  $\diamond$  (and  $\circ$ ) in Table 1 follow trivially from the following result.

**Lemma 18.** *Let  $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}_p[\circ, \diamond], \mathcal{Q}[\diamond], \mathcal{Q}[\circ, \diamond]\}$ . Then  $\mathcal{Q}$  has the PSP under the empty ontology.*

*Proof.* The proof for  $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}_p[\circ, \diamond]\}$  is trivial: if  $\mathcal{D} \models \varkappa(0)$  for some  $\varkappa \in \mathcal{Q}$ , then  $\varkappa$  is clearly equivalent to a query in  $\mathcal{Q}$  whose temporal depth does not exceed the maximal timestamp in  $\mathcal{D}$ , and so is of linear size in  $\mathcal{D}$ .

For  $\mathcal{Q} \in \{\mathcal{Q}[\diamond], \mathcal{Q}[\circ, \diamond]\}$ , the argument is as follows. Assume that  $\varkappa$  separates  $(E^+, E^-)$ . We may assume that  $\varkappa$  is a conjunction of at most  $|E^-|$ -many queries in  $\mathcal{Q}_p^\circ[\diamond]$  of the form

$$\varkappa = \rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \cdots \wedge \diamond \rho_n)),$$

where every  $\rho_i$  is a query in  $\mathcal{Q}_p[\circ]$ . Then the conjuncts of  $\varkappa$  are equivalent to queries in which  $n$  does not exceed the maximal timestamps in data instances in  $E^+$  and each  $\rho_i$  is a query in  $\mathcal{Q}_p[\circ]$  whose temporal depth also does not exceed the maximal timestamps in data instances in  $E^+$ .  $\dashv$

We next complete the description of the polynomial-time algorithm solving  $\text{QBE}_b^+(\mathcal{Q}_p[\circ, \diamond])$  for  $E^+ = \{\mathcal{D}_1^+, \mathcal{D}_2^+\}$  and  $E^- = \{\mathcal{D}_1^-, \mathcal{D}_2^-\}$ . The extension to arbitrary  $E^+, E^-$  is straightforward. Recall that we assume that  $\varkappa$  takes the form (1) with  $\rho_n \neq \top$ . Also recall that  $S_{i,j}$  is the set of tuples  $(k, \ell_1, \ell_2, n_1, n_2)$  such that

1.  $\ell_1 \leq i \leq \max \mathcal{D}_1^+$ ,
2.  $\ell_2 \leq j \leq \max \mathcal{D}_2^+$ ,

and there is  $\varkappa = \rho_0 \wedge \circ_1(\rho_1 \wedge \cdots \wedge \circ_k \rho_k)$  for which

1. there are satisfying assignments  $f_1, f_2$  in  $\mathcal{D}_1^+$  and  $\mathcal{D}_2^+$  with  $f_1(k) = \ell_1$  and  $f_2(k) = \ell_2$ , respectively, and
2.  $n_1$  is minimal with a satisfying assignment  $f$  for  $\varkappa$  in  $\mathcal{D}_1^-$  such that  $f(k) = n_1$ , and  $n_1 = \infty$  if there is no such  $f$ ;  $n_2$  is minimal with a satisfying assignment  $f$  for  $\varkappa$  in  $\mathcal{D}_2^-$  such that  $f(k) = n_2$ , and  $n_2 = \infty$  if there is no such  $f$ .

Then clearly there is a  $\varkappa \in \mathcal{Q}_p[\circ, \diamond]$  separating  $(E^+, E^-)$  if there are  $k, \ell_1, \ell_2$  such that  $(k, \ell_1, \ell_2, \infty, \infty) \in S_{\max \mathcal{D}_1^+, \max \mathcal{D}_2^+}$ .

So it suffices to compute  $S_{\max \mathcal{D}_1^+, \max \mathcal{D}_2^+}$  in polytime incrementally, starting with  $S_{0,0}$ . We have computed  $S_{0,j}$  and  $S_{i,0}$  already. Recall that  $t_{\mathcal{D}}(i) = \{A \mid A(i) \in \mathcal{D}\}$ . To obtain  $S_{i+1,j+1}$ , we add to  $S_{i+1,j} \cup S_{i,j+1}$  any tuple  $(k, \ell_1, \ell_2, n_1, n_2)$  for which there is  $(k', \ell'_1, \ell'_2, n'_1, n'_2) \in S_{i+1,j} \cup S_{i,j+1}$  with  $k' < k, \ell'_1 < \ell_1 \leq i+1, \ell'_2 < \ell_2 \leq j+1$  such that  $\ell_1 = i+1$  or  $\ell_2 = j+1$  and, for  $m = k - k' + 1 \geq 0$ , we have  $\ell_1 - m > \ell'_1, \ell_2 - m > \ell'_2$  and some sets of atoms  $\rho_1, \dots, \rho_m$  with

$$\rho_1 \subseteq t_{\mathcal{D}_1^+}(\ell_1 - m) \cap t_{\mathcal{D}_2^+}(\ell_2 - m), \dots, \rho_m \subseteq t_{\mathcal{D}_1^+}(\ell_1) \cap t_{\mathcal{D}_2^+}(\ell_2)$$

such that

- either  $n_1$  is minimal with  $n_1 - n'_1 > m$  and

$$\rho_1 \subseteq t_{\mathcal{D}_1^-}(n_1 - m), \dots, \rho_m \subseteq t_{\mathcal{D}_1^-}(n_1)$$

or, if no such  $n_1$  exists,  $n_1 = \infty$ , and

- either  $n_2$  is minimal with  $n_2 - n'_2 > m$  and

$$\rho_1 \subseteq t_{\mathcal{D}_2^-}(n_2 - m), \dots, \rho_m \subseteq t_{\mathcal{D}_2^-}(n_2)$$

or, if no such  $n_2$  exists,  $n_2 = \infty$ .

Thus, we obtain  $S_{i+1,j+1}$  from  $S_{i+1,j} \cup S_{i,j+1}$  by adding any tuple that describes a query obtained from a query  $\varkappa$  described by a tuple in  $S_{i+1,j} \cup S_{i,j+1}$  by attaching the query  $\diamond(\rho_1 \wedge \circ(\rho_2 \wedge \dots \wedge \circ\rho_m))$  with  $m \geq 0$  to it. Clearly  $S_{i+1,j+1}$  can be computed in polynomial time from  $S_{i+1,j}$  and  $S_{i,j+1}$ . This finishes the proof for  $\text{QBE}_{\text{b}^-}^{\text{b}^+}(\mathcal{Q}_p[\circ, \diamond])$ .

The proof for  $\text{QBE}_{\text{b}^-}^{\text{b}^+}(\mathcal{Q}_p[\diamond])$  is obtained by dropping  $\circ$  from the proof above. The P-upper bound for  $\text{QBE}(\mathcal{Q}[\diamond])$  with a bounded number of positive examples can be proved in two steps: (1) by Theorem 4 (i.1) it suffices to prove the P-upper bound for  $\text{QBE}_{\text{b}^-}^{\text{b}^+}(\mathcal{Q}[\diamond])$ ; (2) by Theorem 4 (i.3),  $\text{QBE}_{\text{b}^-}^{\text{b}^+}(\mathcal{Q}[\diamond]) \leq_p \text{QBE}_{\text{b}^-}^{\text{b}^+}(\mathcal{Q}_p[\diamond])$ . Finally, the P-upper bound for  $\text{QBE}(\mathcal{Q}[\circ, \diamond])$  with a bounded number of positive examples follows from the P-upper bound for  $\text{QBE}(\mathcal{Q}[\diamond])$  with a bounded number of positive examples by Theorem 4 (ii.2).

We now prove the results for query languages with U in Table 1. We start with the NP-lower bounds. The NP-lower bound for  $\text{QBE}(\mathcal{Q}_p[\text{U}])$  with a bounded number of positive and negative examples follows from the NP-lower bound for  $\text{QBE}(\mathcal{Q}_p[\diamond])$  with a bounded number of positive examples (shown above) and the second part of Theorem 4 (ii.1) which reduces the number of negative examples from unbounded to a singleton. The NP-lower bound for  $\text{QBE}(\mathcal{Q}[\text{U}_s])$  with a bounded number of negative examples follows from the NP-lower bound for  $\text{QBE}(\mathcal{Q}[\diamond])$  with a bounded number of negative examples (shown above) and Theorem 4 (ii.2). This completes the proof of the NP-lower bounds.

The NP-upper bound for  $\text{QBE}(\mathcal{Q}_p[\text{U}])$  follows from its PSP under the empty ontology which is proved in the same way as Lemma 18:

**Lemma 19.**  $\mathcal{Q}_p[\text{U}]$  has the PSP under the empty ontology.

*Proof.* If  $\mathcal{D} \models \varkappa(0)$  for some  $\varkappa \in \mathcal{Q}_p[\text{U}]$ , then  $\varkappa$  is clearly equivalent to a  $\mathcal{Q}_p[\text{U}]$ -query whose temporal depth does not exceed the maximal timestamp in  $\mathcal{D}$ , and so is of linear size in  $\mathcal{D}$ .  $\dashv$

To obtain the P and PSPACE upper bounds for queries with U, we require the machinery and separability criterion that will be developed in the next section.

## C Separability Criteria for $\mathcal{Q}_p[\text{U}]$ , $\mathcal{Q}[\text{U}_s]$ , and $\mathcal{Q}[\text{U}]$

A transition system is a tuple  $S = (\Sigma_1, \Sigma_2, W, L, R, W_0)$ , where  $\Sigma_1$  (respectively,  $\Sigma_2$ ) is a state (respectively, transition) label alphabet,  $W$  is a set of states and  $W_0 \subseteq W$  is a set of initial states. A state labelling,  $L$ , is a map  $W \rightarrow \Sigma_1$ ; a transition labelling,  $R$ , is a partial map  $W \times W \rightarrow \Sigma_2$ . We write  $s \rightarrow_b s'$ , for  $s, s' \in W$ , if  $R(s, s') = b$  and we write  $s \rightarrow s'$  if  $R(s, s')$  is defined. A run or computation on  $S$  is a finite sequence  $\mathfrak{s} = s_0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$ , for  $n \geq 0$ , such that  $s_{i-1} \rightarrow s_i$  for all  $i$  and  $s_0 \in W_0$ . A computation tree  $\mathfrak{T}_S$  of  $S$  is (an infinite) tree—forest, to be more precise—in which the vertices are runs  $\mathfrak{s}$  on  $S$  and the successor relation is  $\mathfrak{s} \rightarrow \mathfrak{s}'$  for all  $\mathfrak{s} = s_0 \rightarrow \dots \rightarrow s_n$  and  $\mathfrak{s}' = s_0 \rightarrow \dots \rightarrow s_n \rightarrow s_{n+1}$ . The vertices  $\mathfrak{s}$  of the tree are labelled with  $L(s_n)$ , while the edges  $\mathfrak{s} \rightarrow \mathfrak{s}'$  are labelled with  $b$  such that  $s_n \rightarrow_b s_{n+1}$ . A tree  $\mathfrak{T}$  is a subtree of  $\mathfrak{T}_S$  if the set of vertices of  $\mathfrak{T}$  is a convex subset of the set of vertices of  $\mathfrak{T}_S$  containing a root (from  $W_0$ ).

Let  $S$  be a transition system such that  $\Sigma_1 = 2^\Sigma$  and  $\Sigma_2 = 2^{\Sigma \cup \{\perp\}}$  for some signature  $\Sigma$ . We then say that  $S$  is a transition system over the signature  $\Sigma$ . For a pair  $S, T$  of transition systems over  $\Sigma$ , we say that  $S$  is simulated by  $T$  if every finite subtree

$\mathfrak{T}'$  of  $\mathfrak{T}_S$  is homomorphically embeddable into  $\mathfrak{T}_T$ , i.e., there is a map  $h$  from the set of vertices of  $\mathfrak{T}'$  to the set of vertices of  $\mathfrak{T}_T$  such that (i)  $\mathfrak{s}$  is labelled by  $a$  implies  $h(\mathfrak{s})$  is labelled by  $a' \supseteq a$ , (ii)  $\mathfrak{s} \rightarrow \mathfrak{s}'$  in  $\mathfrak{T}'$  labelled by  $b$  implies  $h(\mathfrak{s}) \rightarrow h(\mathfrak{s}')$  in  $\mathfrak{T}_T$  and labelled by  $b' \supseteq b$ . We say that  $S$  is *contained* in  $T$  if every finite path in  $\mathfrak{T}_S$ , (i.e., a run in  $S$ ) is homomorphically embeddable into  $\mathfrak{T}_T$ .

Let  $T = (\Sigma_1, \Sigma_2, W', L', R', W'_0)$ . We define the *direct product* (aka synchronous composition) of  $S$  and  $T$  as a transition system  $S \times T = (\Sigma_1, \Sigma_2, W'', L'', R'', W''_0)$  with  $W'' = W \times W'$ ,  $W''_0 = W_0 \times W'_0$ ,  $L''((s, s')) = L(s) \cap L'(s')$  for all  $(s, s') \in W''$ . Then  $R''((s, s'), (t, t'))$  is defined iff both  $R(s, t)$  and  $R'(s', t')$  are defined, in which case  $R''((s, s'), (t, t')) = R(s, t) \cap R'(s', t')$ . The *disjoint union* of  $S$  and  $T$  is a transition system  $S \uplus T = (\Sigma_1, \Sigma_2, W'', L'', R'', W''_0)$  that is obtained by renaming states in  $T$  if necessary to make  $W$  and  $W'$  disjoint, and then taking  $W'' = W \cup W'$ ,  $L'' = L \cup L'$ ,  $R'' = R \cup R'$ , and  $W''_0 = W \cup W'_0$ . The definitions of the product and disjoint union are straightforwardly extended to a collection of transition systems  $S_1, \dots, S_n$ .

### C.1 Representations for $\mathcal{Q}[\mathsf{U}_s]$

Let  $\varkappa$  be a  $\mathcal{Q}[\mathsf{U}_s]$ -query over a signature  $\Sigma$ . We can naturally associate  $\varkappa$  with a tree  $\mathfrak{T}_\varkappa$  as follows. Let  $\varkappa = \varrho'_0 \wedge \bigwedge_i (\varrho_i \cup \psi_i)$ , where  $\varrho'_0, \varrho_i$  is a conjunction of  $\Sigma$ -atoms,  $\psi_i = \varrho'_i \wedge \bigwedge_j (\varrho_j \cup \psi_j)$ , and  $\varrho'_i$  is a conjunction of  $\Sigma$ -atoms. We do not distinguish between a conjunction and a set of atoms. The root of the tree is  $r$  and the tree has  $r \rightarrow (\varrho_i \cup \psi_i)$  for each  $i$ , i.e., there are vertices  $\varrho_i \cup \psi_i$ . The root  $r$  is labelled with  $\varrho'_0$  and each  $\varrho_i \cup \psi_i$  is labelled with  $\varrho'_i$ . Each edge  $r \rightarrow (\varrho_i \cup \psi_i)$  is labelled with  $\varrho_i$ . The tree then contains  $(\varrho_i \cup \psi_i) \rightarrow (\varrho_j \cup \psi_j)$  for each  $j$ , where each such edge is labelled with  $\varrho_j$  and  $(\varrho_j \cup \psi_j)$  is labelled with the set of atoms of  $\psi_j$ , and so on. Thus, we will treat any  $\varkappa \in \mathcal{Q}[\mathsf{U}_s]$  as a tree. The other way round, every finite tree  $\mathfrak{T}$  with vertices labelled with subsets of  $\Sigma$  and edges labelled with subsets of  $\Sigma \cup \{\perp\}$  corresponds to a  $\mathcal{Q}[\mathsf{U}_s]$ -query. Indeed, let  $x$  be any leaf of  $\mathfrak{T}$ . Then we define a  $\mathcal{Q}[\mathsf{U}_s]$ -query  $\varkappa_x = \bigwedge \varrho$ , where  $\varrho$  is a label of  $x$ . Suppose now we have  $x \rightarrow y_i$ , for  $i \in I$  in  $\mathfrak{T}$ , and the label of  $x$  is  $\varrho$  while the label of  $x \rightarrow y_i$  is  $\varrho_i$ . We define  $\varkappa_x = (\bigwedge \varrho) \wedge \bigwedge_{i \in I} ((\bigwedge \varrho_i) \cup \varkappa_{y_i})$ . The query  $\varkappa_r$ , where  $r$  is the root of  $\mathfrak{T}$ , is the required query representing  $\mathfrak{T}$ . We denote it by  $\varkappa_\mathfrak{T}$ .

Let  $\mathcal{D}$  be a data instance and  $\mathcal{O}$  an *LTL*-ontology over a signature  $\Sigma$ . Let  $S$  be a transition system over  $\Sigma$ . We say that  $S$  *represents*  $\mathcal{O}, \mathcal{D}$  if the following conditions hold: (i)  $\mathcal{O}, \mathcal{D} \models \varkappa_\mathfrak{T}'$  for every finite subtree  $\mathfrak{T}'$  of  $\mathfrak{T}_S$ ; (ii)  $\mathfrak{T}_\varkappa$  is homomorphically embeddable into  $\mathfrak{T}_S$  for each  $\varkappa$  with  $\mathcal{O}, \mathcal{D} \models \varkappa$ .

**Lemma 20.** *Let  $E = (E^+, E^-)$ ,  $E^+ = \{\mathcal{D}_i \mid i \in I^+\}$ ,  $E^- = \{\mathcal{D}_i \mid i \in I^-\}$ , and let  $\mathcal{O}$  be an *LTL* ontology. Let  $S^i$  represent  $\mathcal{O}, \mathcal{D}_i$ , for  $i \in I^+ \cup I^-$ . Then (i)  $E$  is not  $\mathcal{Q}[\mathsf{U}_s]$ -separable under  $\mathcal{O}$  iff  $\prod_{i \in I^+} S^i$  is simulated by  $\uplus_{i \in I^-} S^i$ ; (ii)  $E$  is not  $\mathcal{Q}_p[\mathsf{U}]$ -separable under  $\mathcal{O}$  iff  $\prod_{i \in I^+} S^i$  is contained in  $\uplus_{i \in I^-} S^i$ .*

*Proof.* We show (i). For  $(\Rightarrow)$ , suppose  $S^+ = \prod_{i \in I^+} S^i$  is not simulated by  $S^- = \uplus_{i \in I^-} S^i$ . It follows that there exists a finite subtree  $\mathfrak{T}$  of  $\mathfrak{T}_{S^+}$  that is not homomorphically embeddable into  $\mathfrak{T}_{S^-}$ . We claim that  $\varkappa_\mathfrak{T}$  separates  $E$  under  $\mathcal{O}$ . First, we show that  $\mathcal{O}, \mathcal{D}_i \models \varkappa_\mathfrak{T}$  for each  $i \in I^+$ . Indeed, for any such  $i$ , let  $\mathfrak{T}^i$  be a projection of  $\mathfrak{T}$  to the runs of  $S^i$ . Clearly,  $\mathfrak{T}^i$  is a finite subtree of  $\mathfrak{T}_{S^i}$  and  $\varkappa_{\mathfrak{T}^i} \models \varkappa_\mathfrak{T}$  ( $\mathfrak{T}$  is homomorphically embeddable into  $\mathfrak{T}_i$ ). Because  $S^i$  represents  $\mathcal{O}, \mathcal{D}_i$ , we obtain  $\mathcal{O}, \mathcal{D}_i \models \varkappa_\mathfrak{T}$ . Second, we show that  $\mathcal{O}, \mathcal{D}_i \not\models \varkappa_\mathfrak{T}$  for each  $i \in I^-$ . For the sake of contradiction, suppose  $\mathcal{O}, \mathcal{D}_i \models \varkappa_\mathfrak{T}$  for some such  $i$ . Because  $S^i$  represents  $\mathcal{O}, \mathcal{D}_i$ , it follows that  $\mathfrak{T}$  is homomorphically embeddable into  $\mathfrak{T}_{S^i}$ , and so  $\mathfrak{T}$  is homomorphically embeddable into  $\mathfrak{T}_{S^-}$ , which is a contradiction. The proofs of  $(\Leftarrow)$  and (ii) are similar.  $\dashv$

**Constructing representations for queries without an ontology.** Given  $\mathcal{D}$ , we construct a transition system  $S$  with the states  $0, \dots, (\max \mathcal{D} + 1)$ , where  $(\max \mathcal{D} + 1)$  is labelled with  $\emptyset$  and the remaining  $j$  by  $\{A \mid A(j) \in \mathcal{D}\}$ . Transitions are  $j \rightarrow k$ , for  $0 \leq j < k \leq \max \mathcal{D} + 1$ , that are labelled by  $\{A \in \Sigma \cup \{\perp\} \mid A(n) \in \mathcal{D}, n \in (j, k)\}$  and  $(\max \mathcal{D} + 1) \rightarrow (\max \mathcal{D} + 1)$  with label  $\Sigma^\perp = \Sigma \cup \{\perp\}$ .

**Lemma 21.**  *$S$  represents  $\emptyset, \mathcal{D}$ .*

*Proof.* First, we show that  $\mathcal{D} \models \varkappa_\mathfrak{T}'$  for every finite subtree  $\mathfrak{T}'$  of  $\mathfrak{T}_S$ . Let  $\mathcal{I}_\mathcal{D}$  be an *LTL* interpretation such that  $\mathcal{I}_\mathcal{D}, n \models A$  iff  $A(n) \in \mathcal{D}$ , for any atom  $A$ . We observe that any *LTL* interpretation  $\mathcal{I}$  can be viewed as a transition system with the states  $n \in \mathbb{N}$  that are labelled with (sets of) atoms  $A$  holding at  $n$ . The transitions hold between any pair of states  $n < m$  and each such transition is labelled with atoms  $A$  or  $\perp$  that hold at each  $i \in (n, m)$ . It is clear that  $\mathcal{D} \models \varkappa_\mathfrak{T}'$  iff  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}_\mathcal{D}}$ . We define an embedding  $h$  of  $\mathfrak{T}'$  into  $\mathcal{I}_\mathcal{D}$  as follows. We set  $h(0) = 0$ . Suppose  $h(\mathfrak{s})$  for  $\mathfrak{s} = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$  has been defined and let  $\mathfrak{s}' = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_{n+1}$ . If  $s_{n+1} < \max \mathcal{D} + 1$ , then we set  $h(\mathfrak{s}') = s_{n+1}$ . If  $s_{n+1} = \max \mathcal{D} + 1$ , then we set  $h(\mathfrak{s}') = h(\mathfrak{s}') + 1$ . Clearly,  $h$  is a homomorphism. Therefore,  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}_\mathcal{D}}$  and  $\mathcal{D} \models \varkappa_\mathfrak{T}'$ .

Second, we show that  $\mathfrak{T}_\varkappa$  is homomorphically embeddable into  $\mathfrak{T}_S$  for each  $\varkappa$  such that  $\mathcal{D} \models \varkappa$ . Take any  $\varkappa$  such that  $\mathcal{D} \models \varkappa$ . It follows that  $\mathfrak{T}_\varkappa$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}_\mathcal{D}}$ . It remains to observe that  $\mathfrak{T}_{\mathcal{I}_\mathcal{D}}$  is homomorphically embeddable into  $\mathfrak{T}_S$  (the definition of  $h$  is left to the reader).  $\dashv$

The criterion of Theorem 6 now follows immediately from the two previous lemmas. The remaining P and PSPACE upper bounds from Table 1 are explained in the main part of the paper.

**Constructing representations for queries with an  $LTL_{horn}^{\square\circ}$ -ontology.** Let  $\mathcal{D}$  be a data instance and  $\mathcal{O}$  an  $LTL_{horn}^{\square\circ}$ -ontology. Let  $\mathcal{C}_{\mathcal{O},\mathcal{D}}$  be the canonical model of  $\mathcal{O},\mathcal{D}$  and  $s_{\mathcal{O},\mathcal{D}}, p_{\mathcal{O},\mathcal{D}}$  the numbers from Proposition 9. We define  $S$  with the states  $\{0, \dots, \max \mathcal{D} + s_{\mathcal{O},\mathcal{D}} + p_{\mathcal{O},\mathcal{D}} - 1\}$ . The label of each state  $n$  is  $\{A \in \Sigma \mid \mathcal{C}_{\mathcal{O},\mathcal{D}}, n \models A\}$ . There are transitions from  $n$  to  $m$ , for each pair of states  $n < m$  labelled with  $\{A \in \Sigma \cup \{\perp\} \mid \mathcal{C}_{\mathcal{O},\mathcal{D}}, k \models A \text{ for all } k \in (n, m)\}$ . Moreover, there are transitions from  $n$  to  $m$ , for  $n, m \in [\max \mathcal{D} + s_{\mathcal{O},\mathcal{D}}, \max \mathcal{D} + s_{\mathcal{O},\mathcal{D}} + p_{\mathcal{O},\mathcal{D}})$  such that  $n \geq m$ . A label for such a transition is  $\{A \in \Sigma \cup \{\perp\} \mid \mathcal{C}_{\mathcal{O},\mathcal{D}}, k \models A \text{ for all } k \in (n, \max \mathcal{D} + s_{\mathcal{O},\mathcal{D}} + p_{\mathcal{O},\mathcal{D}}) \cup [\max \mathcal{D} + s_{\mathcal{O},\mathcal{D}}, m)\}$ .

**Lemma 22.**  $S$  represents  $\mathcal{O}, \mathcal{D}$ .

*Proof.* First, we show that  $\mathcal{D} \models \varkappa_{\mathfrak{T}'}$  for every finite subtree  $\mathfrak{T}'$  of  $\mathfrak{T}_S$ . It is clear from the properties of  $\mathcal{C}_{\mathcal{O},\mathcal{D}}$  (see Section 5) that  $\mathcal{D} \models \varkappa_{\mathfrak{T}'}$  iff  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$ . We define an embedding  $h$  of  $\mathfrak{T}'$  into  $\mathcal{C}_{\mathcal{O},\mathcal{D}}$  as follows. We set  $h(0) = 0$ . Suppose  $h(\mathfrak{s})$  for  $\mathfrak{s} = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$  has been defined and let  $\mathfrak{s}' = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_{n+1}$ . If  $s_{n+1} > s_n$ , we set  $h(\mathfrak{s}') = h(\mathfrak{s}) + (s_{n+1} - s_n)$ . Otherwise, we set  $h(\mathfrak{s}') = h(\mathfrak{s}) + p_{\mathcal{O},\mathcal{D}} - (s_n - s_{n+1})$ . It is readily verified that  $h$  is a homomorphism. Therefore,  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$  and  $\mathcal{O}, \mathcal{D} \models \varkappa_{\mathfrak{T}'}$ .

Second, we show that  $\mathfrak{T}_{\varkappa}$  is homomorphically embeddable into  $\mathfrak{T}_S$  for each  $\varkappa$  such that  $\mathcal{O}, \mathcal{D} \models \varkappa$ . Take any  $\varkappa$  such that  $\mathcal{O}, \mathcal{D} \models \varkappa$ . It follows that  $\mathfrak{T}_{\varkappa}$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$ . It remains to show that  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$  is homomorphically embeddable into  $S$ . To this end, we define a map  $r: \mathbb{N} \rightarrow [0, \max \mathcal{D} + s_{\mathcal{O},\mathcal{D}} + p_{\mathcal{O},\mathcal{D}})$  by setting  $r(n) = n$  if  $n \in [0, \max \mathcal{D} + s_{\mathcal{O},\mathcal{D}} + p_{\mathcal{O},\mathcal{D}})$  and  $r(n) = ((n - \max \mathcal{D} - s_{\mathcal{O},\mathcal{D}}) \bmod p_{\mathcal{O},\mathcal{D}}) + \max \mathcal{D} + s_{\mathcal{O},\mathcal{D}}$ , otherwise. Now, we define  $h(\mathfrak{s})$  for  $\mathfrak{s} = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$  (note that  $s_i \in \mathbb{N}$  and  $s_{i+1} > s_i$ ) to be equal to  $r(s_n)$ . It is readily verified that  $h$  is a homomorphism from  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$  into  $S$ .  $\dashv$

Now we can explain the data complexity upper bounds from Theorem 12 (the upper bounds from Theorem 11 are explained in the main paper) for  $\text{QBE}(LTL_{horn}^{\square\circ}, \mathcal{Q})$  with  $\mathcal{Q} \in \{\mathcal{Q}[U_s], \mathcal{Q}_p[U]\}$ . The result for  $\text{QBE}^{\text{b}^+}(LTL_{horn}^{\square\circ}, \mathcal{Q}[U_s])$  (and so for  $\text{QBE}^{\text{b}^+}(LTL_{horn}^{\square\circ}, \mathcal{Q}[U_s])$ ) follows from the fact that  $\mathfrak{P} = \prod_{i \in I^+} S^i$  and  $\mathfrak{U} = \uplus_{i \in I^-} S^i$  are constructible in P in the size of  $E$ . This is not the case for  $\text{QBE}(LTL_{horn}^{\square\circ}, \mathcal{Q}[U_s])$ ; here, we use the observation that if there exists a finite subtree  $\mathfrak{T}$  of  $\mathfrak{T}_{\mathfrak{P}}$  that is not homomorphically embeddable into  $\mathfrak{U}$ , then there exists such  $\mathfrak{T}$  satisfying the property that every  $s$  from  $\mathfrak{P}$  occurs on each path of  $\mathfrak{T}$  at most  $|\mathfrak{U}|$ -many times. Let  $N^+ = \prod_{i \in I^+} p_{\mathcal{O},\mathcal{D}_i}$ . We claim that  $\mathfrak{T}$  is a subtree of  $\mathfrak{T}_{\mathfrak{P}}^M$  for  $M = \max_{i \in I^+} \{\max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}\} + N^+|\mathfrak{U}|$ . Indeed, in the required  $\mathfrak{T}$ , if there is a path that is longer than  $M$ , the property above would be violated. By our construction of  $S^i$ , any  $n$ -th element, for  $n \geq \max_{i \in I^+} \{\max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}\}$ , of any path of  $\mathfrak{T}_{\mathfrak{P}}$  is of the form  $(t_1, \dots, t_{|I^+|})$ , where  $t_i \in [\max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}, \max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i} + p_{\mathcal{O},\mathcal{D}_i})$ . Observe that  $(t_1, \dots, t_{|I^+|}) \rightarrow (s_1, \dots, s_{|I^+|})$  in  $\mathfrak{P}$ , for  $(t_1, \dots, t_{|I^+|})$  as above, implies  $s_i \in [\max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}, \max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i} + p_{\mathcal{O},\mathcal{D}_i})$  and any sequence  $(t_1, \dots, t_{|I^+|}) \rightarrow \dots \rightarrow (s_1, \dots, s_{|I^+|})$  as above in  $\mathfrak{P}$  longer than  $N^+|\mathfrak{U}|$  will have some  $(t_1, \dots, t_{|I^+|})$  repeated more than  $|\mathfrak{U}|$  times. Thus, in order to decide  $\text{QBE}(LTL_{horn}^{\square\circ}, \mathcal{Q}[U_s])$ , we need to check if  $\mathfrak{T}_{\mathfrak{P}}^M$  is homomorphically embeddable into  $\mathfrak{U}$ . The latter can be checked by constructing  $\mathfrak{T}_{\mathfrak{P}}^M$  branch-by-branch while checking all possible embeddings of these branches into  $\mathfrak{U}$ . Since  $M$  is polynomial in  $E$ , this algorithm works in PSPACE in the size of  $E$ .

It remains to explain the NP upper bound for  $\text{QBE}(LTL_{horn}^{\square\circ}, \mathcal{Q}_p[U])$ . From Lemma 20 and the argument above, it follows that  $E$  is separable under  $\mathcal{O}$  iff there exists a path in  $\mathfrak{T}_{\mathfrak{P}}^M$  that is not embeddable into  $\mathfrak{U}$ . Such a path (if exists) is of the size polynomial in  $E$ . Embeddability of such a path into  $\mathfrak{U}$  can be checked in P from  $E^-$  and the size of the path.

**Constructing representations for queries with an  $LTL$ -ontology.** Let  $\mathcal{D}$  be a data instance and  $\mathcal{O}$  an  $LTL$ -ontology. We can assume that  $\max \mathcal{D} = 0$ . Indeed, for a given  $\mathcal{O}$  and  $E$ , we can construct in polytime an  $LTL$ -ontology  $\mathcal{O}'$  and  $E'$  such that  $\max \mathcal{D}' = 0$  for each  $\mathcal{D}'$  in  $E'$  and  $E$  is  $\mathcal{Q}[U_s]$ -separable under  $\mathcal{O}$  iff  $E'$  is  $\mathcal{Q}[U_s]$ -separable under  $\mathcal{O}'$ . Let  $\mathcal{T}_{\mathcal{O}}$  be the set of  $\mathcal{O}$ -types. For  $\mathcal{T} \subseteq \mathcal{T}_{\mathcal{O}}$ , we say that  $\mathcal{T}$  is *realisable in  $\mathcal{O}, \mathcal{D}$*  if there are instants  $n_{\mathcal{I}}$  in all models  $\mathcal{I}$  of  $\mathcal{O}, \mathcal{D}$  such that  $\{tp_{\mathcal{I}}(n_{\mathcal{I}}) \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}\} = \mathcal{T}$ . For  $\mathcal{T}_1, \mathcal{T}_2 \subseteq \mathcal{T}_{\mathcal{O}}$  realisable in  $\mathcal{O}, \mathcal{D}$  and  $\Gamma \subseteq \Sigma \cup \{\perp\}$ , we define  $\mathcal{T}_1 \rightarrow_{\Gamma} \mathcal{T}_2$  if there are instants  $n_{\mathcal{I}} < m_{\mathcal{I}}$  in all models  $\mathcal{I}$  of  $\mathcal{O}, \mathcal{D}$  such that (i)  $\{tp_{\mathcal{I}}(n_{\mathcal{I}}) \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}\} = \mathcal{T}_1$ , (ii)  $\{tp_{\mathcal{I}}(m_{\mathcal{I}}) \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}\} = \mathcal{T}_2$ , (iii)  $\{tp_{\mathcal{I}}(k) \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}, n_{\mathcal{I}} < k < m_{\mathcal{I}}\} \cap \mathcal{T}_2 = \emptyset$ , (iv)  $\Gamma = \{A \in \Sigma^{\perp} \mid \mathcal{I}, k \models A \text{ for all } \mathcal{I} \models \mathcal{O}, \mathcal{D}, n_{\mathcal{I}} < k < m_{\mathcal{I}}\}$ . We observe that condition (iii) ensures that there exists at most one  $\Gamma$  for given  $\mathcal{T}_1, \mathcal{T}_2$  such that  $\mathcal{T}_1 \rightarrow_{\Gamma} \mathcal{T}_2$ . We define  $S$  with the states  $\mathcal{T} \subseteq \mathcal{T}_{\mathcal{O}}$  realisable in  $\mathcal{O}, \mathcal{D}$ . A single initial state of  $S$  is  $\mathcal{T}_0 = \{tp_{\mathcal{I}}(0) \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}\}$ . The states  $\mathcal{T}$  are labelled with  $\{A \in \Sigma \mid A \in tp \text{ for all } tp \in \mathcal{T}\}$ . There are transitions from  $\mathcal{T}_1$  to  $\mathcal{T}_2$  where  $\mathcal{T}_1 \rightarrow_{\Gamma} \mathcal{T}_2$  holds for some  $\Gamma$ , labelled with  $\Gamma$ .

**Lemma 23.**  $S$  represents  $\mathcal{O}, \mathcal{D}$ .

*Proof.* Let  $\mathcal{I} = \{\mathcal{I} \mid \mathcal{I} \models \mathcal{O}, \mathcal{D}\}$ . We treat every  $\mathcal{I} \in \mathcal{I}$  as a transition system as we did above. Take the product  $\prod_{\mathcal{I} \in \mathcal{I}} \mathcal{I}$  and denote it (slightly abusing notation) by  $\mathcal{I}$ . Note that the states  $s$  of  $\mathcal{I}$  are maps  $s: \mathcal{I} \rightarrow \mathbb{N}$ .

First, we show that  $\mathcal{O}, \mathcal{D} \models \varkappa_{\mathfrak{T}'}$  for every finite subtree  $\mathfrak{T}'$  of  $\mathfrak{T}_S$ . It should be clear that  $\mathcal{O}, \mathcal{D} \models \varkappa_{\mathfrak{T}'}$  iff  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}}$ . We define an embedding  $h$  of  $\mathfrak{T}'$  into  $\mathcal{I}$  as follows. We set  $h(\mathcal{T}_0) = s_0$ , where  $s_0$  is the initial state of  $\mathcal{I}$  satisfying  $s_0(\mathcal{I}) = 0$  for every  $\mathcal{I} \in \mathcal{I}$ . Suppose  $h(\mathfrak{s})$  for  $\mathfrak{s} = \mathcal{T}_0 \rightarrow \mathcal{T}_1 \rightarrow \dots \rightarrow \mathcal{T}_n$  has been defined equal to  $s$ . Our induction hypothesis will be that  $\mathcal{T}_n = \{tp_{\mathcal{I}}(s(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\}$ . It can be readily verified that it holds for  $\mathfrak{s} = \mathcal{T}_0$ . Let  $\mathfrak{s}' = \mathcal{T}_0 \rightarrow \mathcal{T}_1 \rightarrow \dots \rightarrow \mathcal{T}_{n+1}$  and  $\mathcal{T}_n \rightarrow_{\Gamma} \mathcal{T}_{n+1}$ . For each  $\mathcal{I} \in \mathcal{I}$ , we select  $m_{\mathcal{I}} > s(\mathcal{I})$  such that

$\mathcal{T}_{n+1} = \{tp_{\mathcal{I}}(m_{\mathcal{I}}) \mid \mathcal{I} \in \mathcal{I}\}$  and  $\Gamma = \{A \in \Sigma^{\perp} \mid \mathcal{I}, k \models A \text{ for all } \mathcal{I} \in \mathcal{I}, s(\mathcal{I}) < k < m_{\mathcal{I}}\}$ . That this selection is always possible follows from the IH. We then set  $h(s') = s'$  such that  $s'(\mathcal{I}) = m_{\mathcal{I}}$  for  $\mathcal{I} \in \mathcal{I}$ .

Now, we show that  $\mathfrak{T}_{\mathcal{X}}$  is homomorphically embeddable into  $\mathfrak{T}_S$  for each  $\mathcal{X}$  such that  $\mathcal{O}, \mathcal{D} \models \mathcal{X}$ . Take any  $\mathcal{X}$  such that  $\mathcal{O}, \mathcal{D} \models \mathcal{X}$ . It follows that  $\mathfrak{T}_{\mathcal{X}}$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}}$ . It remains to show that  $\mathfrak{T}_{\mathcal{I}}$  is homomorphically embeddable into  $S$ . We define  $h$  so that  $h(s_0) = T_0$ . Consider now  $s_1 = s_0 \rightarrow s_1$ . Instead of  $s_1$ , we can always select  $s'_1$  such that  $s_0 \rightarrow s'_1$ ,  $\{tp_{\mathcal{I}}(s'_1(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\} \subseteq \{tp_{\mathcal{I}}(s_1(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\}$ ,  $\{tp_{\mathcal{I}}(k) \mid \mathcal{I} \in \mathcal{I}, s_0(\mathcal{I}) < k < s'_1(\mathcal{I})\} \cap \{tp_{\mathcal{I}}(s'_1(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\} = \emptyset$  and, finally,  $\{tp_{\mathcal{I}}(k) \mid \mathcal{I} \in \mathcal{I}, s_0(\mathcal{I}) < k < s'_1(\mathcal{I})\} \subseteq \{tp_{\mathcal{I}}(k) \mid \mathcal{I} \in \mathcal{I}, s_0(\mathcal{I}) < k < s_1(\mathcal{I})\}$ . We define  $h(s_1) = \{tp_{\mathcal{I}}(s'_1(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\}$ . It can be readily verified, given the subsumptions above, that (subsumption of the) label of the edge from  $s_0$  to  $s_1$  is preserved under  $h$  and (subsumption of) the node  $s_1$  is preserved under  $h$ .

Consider now  $s_2 = s_0 \rightarrow s_1 \rightarrow s_2$ . We can always select  $s'_2$  so that  $s'_1 \rightarrow s'_2$ ,  $\{tp_{\mathcal{I}}(s'_2(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\} \subseteq \{tp_{\mathcal{I}}(s_2(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\}$ ,  $\{tp_{\mathcal{I}}(k) \mid \mathcal{I} \in \mathcal{I}, s'_1(\mathcal{I}) < k < s'_2(\mathcal{I})\} \cap \{tp_{\mathcal{I}}(s'_2(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\} = \emptyset$  and, finally,  $\{tp_{\mathcal{I}}(k) \mid \mathcal{I} \in \mathcal{I}, s'_1(\mathcal{I}) < k < s'_2(\mathcal{I})\} \subseteq \{tp_{\mathcal{I}}(k) \mid \mathcal{I} \in \mathcal{I}, s_1(\mathcal{I}) < k < s_2(\mathcal{I})\}$ . We define  $h(s_2) = \{tp_{\mathcal{I}}(s'_2(\mathcal{I})) \mid \mathcal{I} \in \mathcal{I}\}$ .

Clearly, we can extend this argument to arbitrarily-many steps to define  $h$  for any  $s = s_0 \rightarrow \dots \rightarrow s_n$ . This completes the proof of the lemma.  $\dashv$

## C.2 Representations for $\mathcal{Q}[\mathcal{U}]$

Let  $\mathcal{X}$  be a  $\mathcal{Q}[\mathcal{U}]$ -query over a signature  $\Sigma$ . Let all the subformulas of  $\mathcal{X}$ , which are either conjunctions (sets) of atoms  $\gamma, \lambda$ , or  $\varphi \cup \psi$ , or conjunctions thereof, be enumerated. We assume that there are no subformulas of the form  $\varphi \cup \varphi$  (such formulas are equivalent to  $\perp \cup \varphi$ ). We can associate  $\mathcal{X}$  with a tree  $\mathfrak{T}_{\mathcal{X}}$  having edges of two types: black and red. Let  $\mathcal{X} = \gamma_0 \wedge \bigwedge_{i \in J_0} \mathcal{X}_i$  and let  $\mathcal{X}_i = (\lambda_i \wedge \bigwedge_{j \in I_i} \mathcal{X}_j) \cup (\gamma_i \wedge \bigwedge_{j \in J_i} \mathcal{X}_j)$ , where  $\mathcal{X}_j$ , for  $j \in I_i \cup J_i$ , is of the form  $\varphi \cup \psi$ . Then the root  $r$  of the tree is labelled with  $\gamma_0$  and there are black edges  $r \rightarrow \mathcal{X}_i$ , for  $i \in J_0$ . Each such edge is labelled with  $\lambda_i$  and node  $\mathcal{X}_i$ ,  $i \in J_0$ , is labelled with  $\gamma_i$ . Now take any  $i \in J_0$ . There is a black edge  $\mathcal{X}_i \rightarrow \mathcal{X}_j$ , for each  $j \in J_i$ , and there is a red edge  $\mathcal{X}_i \rightarrow \mathcal{X}_j$ , for each  $j \in I_i$ . To define the label of each such black or red  $\mathcal{X}_i \rightarrow \mathcal{X}_j$  edge and the label of the corresponding  $\mathcal{X}_j$ , we look at the form of the  $\mathcal{X}_j$ . Suppose  $\mathcal{X}_j = (\lambda_j \wedge \bigwedge_{k \in I_j} \mathcal{X}_k) \cup (\gamma_j \wedge \bigwedge_{k \in J_j} \mathcal{X}_k)$ . Then the label of  $\mathcal{X}_i \rightarrow \mathcal{X}_j$  is  $\lambda_j$  and the label of  $\mathcal{X}_j$  is  $\gamma_j$ . The construction of the edges  $\mathcal{X}_j \rightarrow \mathcal{X}_k$ , their labels, and the construction of the subsequent tree is done analogously (by treating  $j$  as  $i$  in the previous construction). Thus, we can and will treat any  $\mathcal{X} \in \mathcal{Q}[\mathcal{U}_s]$  as the tree  $\mathfrak{T}_{\mathcal{X}}$ .

Let a *black/red tree* be a tree where each edge has either black or red colour, but not both. Every finite black/red tree  $\mathfrak{T}$ , where vertices are labelled with subsets of  $\Sigma$  and edges are labelled with subsets of  $\Sigma \cup \{\perp\}$ , corresponds to a  $\mathcal{Q}[\mathcal{U}]$ -query. Indeed, let  $x \rightarrow y$  be any edge such that  $y$  is a leaf of  $\mathfrak{T}$ . Then we define a  $\mathcal{Q}[\mathcal{U}]$ -query  $\mathcal{X}_{x \rightarrow y}$  as  $\lambda \cup \gamma$ , where  $\lambda$  is the label of  $x \rightarrow y$  while  $\gamma$  is the label of  $y$ . Suppose now we have black (respectively, red) transitions  $x \rightarrow y_i$ , for  $i \in J$  (respectively,  $i \in I$ ), for an edge  $z \rightarrow x$  in  $\mathfrak{T}$  labelled with  $\lambda$  for  $x$  labelled with  $\gamma$ . We define  $\mathcal{X}_{z \rightarrow x} = (\lambda \wedge \bigwedge_{i \in I} \mathcal{X}_{x \rightarrow y_i}) \cup (\gamma \wedge \bigwedge_{i \in J} \mathcal{X}_{x \rightarrow y_i})$ . Let  $r$  be the root of  $\mathfrak{T}$  labelled with  $\gamma$ . Then the required  $\mathcal{X}_{\mathfrak{T}}$  representing  $\mathfrak{T}$  is  $\gamma \wedge \bigwedge_{r \rightarrow y \text{ in } \mathfrak{T}} \mathcal{X}_{r \rightarrow y}$ .

Further, we define a *black/red transition system*  $S$  by adding either black or red colour, but not both, to each transition  $s \rightarrow s'$  of the transition system  $S$  defined above. The computation tree  $\mathfrak{T}_S$  of  $S$  is defined as before, however, every edge  $s \rightarrow s'$  in  $\mathfrak{T}_S$  has either red or black (but not both) colour that is equal to the colour of  $s_n \rightarrow s_{n+1}$ . In the definition of the direct product  $S \times T$ , we now require that  $R''((s, s'), (t, t'))$  is red (respectively, black) iff both  $R(s, t)$  and  $R'(s', t')$  are red (respectively, black) (the labels are defined as before). Finally, in the definition of a homomorphic embedding of a black/red (labelled) tree  $\mathfrak{T}'$  to another such tree  $\mathfrak{T}''$  we require, additionally, that  $h(s) \rightarrow h(s')$  is black (respectively, red) in  $\mathfrak{T}''$  if  $s \rightarrow s'$  is black (respectively, red) in  $\mathfrak{T}'$ . For a data instance  $\mathcal{D}$ , an LTL ontology  $\mathcal{O}$ , a signature  $\Sigma$ , and a black/red transition system  $S$ , the definition of  $S$  representing  $\mathcal{O}, \mathcal{D}$  continues to hold (with  $\mathcal{Q}[\mathcal{U}_s]$  changed to  $\mathcal{Q}[\mathcal{U}]$  in (ii)). Moreover, the same proof as in Lemma 24 (i) shows that we have:

**Lemma 24.** *Let  $E = (E^+, E^-)$ ,  $E^+ = \{\mathcal{D}_i \mid i \in I^+\}$ ,  $E^- = \{\mathcal{D}_i \mid i \in I^-\}$ , and let  $\mathcal{O}$  be an LTL ontology. Let  $S^i$  represent  $\mathcal{O}, \mathcal{D}_i$ , for  $i \in I^+ \cup I^-$ . Then  $E$  is not  $\mathcal{Q}[\mathcal{U}]$ -separable under  $\mathcal{O}$  iff  $\prod_{i \in I^+} S^i$  is simulated by  $\uplus_{i \in I^-} S^i$ .*

**Constructing representations for queries without an ontology.** Let  $d, e \subseteq \mathbb{N}$  be finite and nonempty. For any  $d \in d$ , let  $\mu(d) = \min\{e \in e \mid d < e\}$ . If  $\mu$  is a surjective  $d \rightarrow e$  function, we write  $d \leq e$  and set

$$\nabla(d, e) = \bigcup_{d \in d} \{d' \in \mathbb{N} \mid d < d' < \mu(d)\}.$$

**Example 25.** Let  $d = \{1, 2, 3\}$  and  $e = \{3, 4\}$ . Then  $d \leq e$  with  $\nabla(d, e) = \{2\}$ . However, for  $d = \{1, 2\}$  and  $e = \{3, 4\}$ , we have neither  $d \leq e$  (because  $\mu$  is not a surjection) nor  $e \leq d$  (because  $\mu$  is not defined).

Given a data instance  $\mathcal{D}$ ,  $e \subseteq \mathbb{N}$  and an atom  $A$  (possibly  $\perp$ ), we write  $\mathcal{D}, e \models A$  if  $A(e) \in \mathcal{D}$  for all  $e \in e$ . We construct a black/red transition system  $S$  with a set of states  $\{0, z, u\} \cup \{de \mid d, e \subseteq \{0, \dots, \max \mathcal{D}\}\}$ . The label of 0 is  $\{A \mid A(0) \in \mathcal{D}\}$ , the label of  $z$  is  $\emptyset$ , the label of  $u$  is  $\Sigma^{\perp}$ , and the label of  $de$  is  $\{A \mid A(e) \in \mathcal{D} \text{ for all } e \in e\}$ . The alphabet of the transition labels is  $2^{\Sigma \cup \{\perp\}}$ . From 0, we have

- (i) a black transitions to every  $ed$  such that  $\{0\} < d$  (this implies that  $|d| = 1$ ) and  $e = \nabla(\{0\}, d)$ , labelled with the set  $\{A \in \Sigma^{\perp} \mid \mathcal{D}, e \models A\}$ .



From each  $ed$ , we have

- (ii) a black transition to every  $fg$  such that  $d < g$  and  $f = \nabla(d, g)$ , labelled with  $\{A \in \Sigma^\perp \mid \mathcal{D}, f \models A\}$ ,
- (iii) a red transition to every  $fg$  such that  $e < g$  and  $f = \nabla(e, g)$ , labelled with  $L = \{A \in \Sigma^\perp \mid \mathcal{D}, f \models A\}$ .

The state  $z$  has a black and a red transition to itself labelled with  $\Sigma^\perp$  and the same holds for  $u$ . We have a black transition to  $z$  from every  $ed$  labelled with  $\{A \in \Sigma \cup \{\perp\} \mid \mathcal{D}, \{\max d, \dots, \max \mathcal{D} - 1\} \models A\}$ , and we have a red transition to  $z$  from every  $ed$  labelled with  $\{A \in \Sigma^\perp \mid \mathcal{D}, \{\max e, \dots, \max \mathcal{D} - 1\} \models A\}$ . Finally, we have a red transition from every  $\emptyset d$  to  $u$  as well as from  $z$  to  $u$  labelled with  $\Sigma^\perp$ .

**Lemma 26.**  $S$  represents  $\emptyset, \mathcal{D}$ .

*Proof.* First, we show that  $\mathcal{D} \models \varkappa_{\mathfrak{T}'}$ , for every finite subtree  $\mathfrak{T}'$  of  $\mathfrak{T}_S$ . Let  $\mathcal{I}_{\mathcal{D}}$  be an *LTL* interpretation such that  $\mathcal{I}_{\mathcal{D}}, n \models A$  iff  $A(n) \in \mathcal{D}$ , for any atom  $A$ . We regard any *LTL* interpretation  $\mathcal{I}$  as a black/red transition system with the states  $\{0, u\} \cup \{\mathbf{de} \mid \mathbf{d}, \mathbf{e} \subseteq \mathbb{N}, \mathbf{e} \neq \emptyset\}$ . The state 0 is labelled with  $\{A \in \Sigma \mid \mathcal{D}, 0 \models A\}$ ,  $u$  is labelled with  $\Sigma^\perp$ , while each state  $\mathbf{de}$  is labelled with  $\{A \in \Sigma \mid \mathcal{D}, \mathbf{e} \models A\}$ . From 0, there are black transitions according to (i). From each  $ed$ , we have black and red transitions according to (ii) and (iii), respectively. The state  $u$  has a black and a red transition to itself and a transition from each state  $\emptyset e$  all labelled with  $\Sigma^\perp$ . It should be clear that  $\mathcal{D} \models \varkappa_{\mathfrak{T}'}$  iff  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}_{\mathcal{D}}}$ . We define an embedding  $h$  of  $\mathfrak{T}'$  into  $\mathcal{I}_{\mathcal{D}}$  as follows. We set  $h(0) = 0$ . Suppose  $h(\mathfrak{s})$  for  $\mathfrak{s} = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$  has been defined and let  $\mathfrak{s}' = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_{n+1}$ . Suppose, first,  $s_{n+1} = \mathbf{de}$  for  $\mathbf{e} \subseteq [0, \max \mathcal{D}]$ . Then we set  $h(\mathfrak{s}') = \mathbf{de}$ . Suppose  $s_{n+1} = z$ . Then  $h(\mathfrak{s}_n) = \mathbf{de}$ , for some  $\mathbf{d}, \mathbf{e} \subseteq \mathbb{N}$ . If  $s_n \rightarrow s_{n+1}$  is black, we set  $h(\mathfrak{s}') = \emptyset\{e + 1 \mid e \in \mathbf{e}\}$  and if it is red, we set  $h(\mathfrak{s}') = \emptyset\{d + 1 \mid d \in \mathbf{d}\}$ . Finally, if  $s_{n+1} = u$ , then we set  $h(\mathfrak{s}') = u$ . It is straightforwardly verified that  $h$  is a homomorphism. Therefore,  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}_{\mathcal{D}}}$  and  $\mathcal{D} \models \varkappa_{\mathfrak{T}'}$ .

Second, we show that  $\mathfrak{T}_\varkappa$  is homomorphically embeddable into  $\mathfrak{T}_S$  for each  $\varkappa$  such that  $\mathcal{D} \models \varkappa$ . Take any  $\varkappa$  such that  $\mathcal{D} \models \varkappa$ . It follows that  $\mathfrak{T}_\varkappa$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{I}_{\mathcal{D}}}$ . It remains to observe that  $\mathfrak{T}_{\mathcal{I}_{\mathcal{D}}}$  is homomorphically embeddable into  $S$ . Indeed, we define  $h(0) = 0$ . Let  $\mathfrak{s} = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$  for  $n \geq 1$ . If  $s_n = \mathbf{de}$  for  $\mathbf{e} \subseteq [0, \max \mathcal{D}]$ , then  $h(\mathfrak{s}) = s_n$ . If  $\max \mathbf{e} > \max \mathcal{D}$ , we set  $h(\mathfrak{s}) = z$ . Finally, if  $s_n = u$ , we set  $h(\mathfrak{s}) = u$ . It is readily verified that  $h$  is a homomorphism from  $\mathfrak{T}_{\mathcal{I}_{\mathcal{D}}}$  into  $S$ .  $\dashv$

Now we explain why  $\text{QBE}(\mathcal{Q}[U])$  is in PSPACE. To this end we observe that every run of  $S$  of length  $> \max \mathcal{D}$  results in either  $s = z$  or  $s = u$ . Moreover, if  $s = u$  then all the subsequent states of the run are also  $u$ . Thus, any run of  $\mathfrak{P} = \prod_{i \in I^+} S^i$  of length  $> \max_{i \in I^+} \{\max \mathcal{D}\}$  is in a state  $s = (t_1, \dots, t_{|I^+|})$  where  $t_i \in \{u, z\}$ . Then  $\mathfrak{T}_{\mathfrak{P}}^M$  for  $M = \max_{i \in I^+ \cup I^-} \{\max \mathcal{D}\} + 1$  is mapped into  $\mathfrak{U} = \uplus_{i \in I^-} S^i$ , if a map  $h$  exists, in such a way that  $h(s_0 \rightarrow \dots \rightarrow s_M)$  is either  $z$  or  $u$  (in the corresponding  $S$  representing  $\mathcal{D}_i$ ,  $i \in I^-$ ). So, we obtain that if  $\mathfrak{T}_{\mathfrak{P}}^M$  is homomorphically embeddable into  $\mathfrak{U}$ , then any finite subtree of  $\mathfrak{T}_{\mathfrak{P}}$  is homomorphically embeddable into  $\mathfrak{U}$ . To decide  $\text{QBE}(\mathcal{Q}[U])$ , we can check the embeddability of  $\mathfrak{T}_{\mathfrak{P}}^M$  in a branch-by-branch fashion similarly to the case of  $\mathcal{Q}[U_s]$ . Note, however, that the existence of a polynomial algorithm for  $\text{QBE}^{\text{b}^+}(\mathcal{Q}[U])$  and  $\text{QBE}_{\text{b}}^{\text{b}^+}(\mathcal{Q}[U])$  remains open as bounding the number of positive examples does not result in  $\mathfrak{P}$  of polynomial size.

**Constructing representations for queries with an  $LTL_{\text{horn}}^{\square\circ}$ -ontology.** Let  $\mathbf{d}, \mathbf{e}$  be finite and nonempty subsets of the interval  $[0, P)$ , for some  $P \in \mathbb{N}$ , and  $M \in \mathbb{N}$ . For any  $d \in \mathbf{d}$ , let  $\text{succs}(d, \mathbf{e}) = \{e \in \mathbf{e} \mid d < e\}$  and let

$$\mu(d) = \begin{cases} \min \text{succs}(d, \mathbf{e}), & \text{if either } d \in [0, M) \text{ or both } d \in [M, P) \text{ and } \text{succs}(d, \mathbf{e}) \neq \emptyset, \\ \min(\mathbf{e}), & \text{if } d \in [M, P) \text{ and } \text{succs}(d, \mathbf{e}) = \emptyset. \end{cases}$$

If  $\mu$  is a surjective  $\mathbf{d} \rightarrow \mathbf{e}$  function, we write  $\mathbf{d} \triangleleft_{M,P} \mathbf{e}$  and set  $\nabla_{M,P}(\mathbf{d}, \mathbf{e}) = \bigcup_{d \in \mathbf{d}} \{d' \in \text{bwn}(d, \mu(d))\}$ , where  $\text{bwn}(d, e) = (d, e)$  if  $d < e$  and  $(e, P) \cup [M, d)$  if  $d \geq e$ .

**Example 27.** Let  $M = 2$ ,  $P = 8$ ,  $\mathbf{d} = \{1, 4, 6, 7\}$  and  $\mathbf{e} = \{3, 5\}$ . Then  $\mathbf{d} \triangleleft_{M,P} \mathbf{e}$  with  $\nabla_{M,P}(\mathbf{d}, \mathbf{e}) = \{2, 7\}$ .

Let  $\mathcal{D}$  be a data instance and  $\mathcal{O}$  an  $LTL_{\text{horn}}^{\square\circ}$ -ontology. Let  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  be the canonical model of  $\mathcal{O}, \mathcal{D}$  and  $s_{\mathcal{O}, \mathcal{D}}, p_{\mathcal{O}, \mathcal{D}}$  be the numbers from Proposition 9. Let  $\max \mathcal{D} + s_{\mathcal{O}, \mathcal{D}} = M$  and  $\max \mathcal{D} + s_{\mathcal{O}, \mathcal{D}} + p_{\mathcal{O}, \mathcal{D}} = P$ . We define  $S$  with the states  $\{0, u\} \cup \{\mathbf{de} \mid \mathbf{d}, \mathbf{e} \subseteq [0, P)\}$ . The label of 0 is  $\{A \in \Sigma \mid \mathcal{C}_{\mathcal{O}, \mathcal{D}}, 0 \models A\}$ , the label of  $\mathbf{de}$  is  $\{A \in \Sigma \mid \mathcal{C}_{\mathcal{O}, \mathcal{D}}, \mathbf{e} \models A\}$  and the label of  $u$  is  $\Sigma^\perp$ . We define the (red and black) transitions between 0 and  $\mathbf{de}$  as specified by (i)–(iii) above but using  $\triangleleft_{M,P}$  instead of  $<$ ,  $\nabla_{M,P}$  instead of  $\nabla$ , and  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}, \mathbf{e}$  instead of  $\mathcal{D}, \mathbf{e}$  (the same applies to  $\mathbf{f}$ ). Finally, we define the transitions between  $\emptyset e$  and  $u$  as defined above.

**Lemma 28.**  $S$  represents  $\mathcal{O}, \mathcal{D}$ .

*Proof.* First, we show that  $\mathcal{D} \models \varkappa_{\mathfrak{T}'}$  for every finite subtree  $\mathfrak{T}'$  of  $\mathfrak{T}_S$ . It is clear from the properties of  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  that  $\mathcal{O}, \mathcal{D} \models \varkappa_{\mathfrak{T}'}$  iff  $\mathfrak{T}'$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O}, \mathcal{D}}}$ . We define an embedding  $h$  of  $\mathfrak{T}'$  into  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  as follows. We set  $h(0) = 0$ . Suppose  $h(\mathfrak{s})$  for  $\mathfrak{s} = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$  has been defined and  $\mathfrak{s}' = 0 \rightarrow s_1 \rightarrow \dots \rightarrow s_{n+1}$ . If  $s_n = u$ , we set  $h(\mathfrak{s}') = u$ . Suppose  $s_n = \mathbf{de}$  and  $s_{n+1} = \mathbf{fg}$ . We will have an IH that  $r(d') = \mathbf{d}$  and  $r(e') = \mathbf{e}$  for the map  $r$  from Lemma 22. First,

we assume  $s_n \rightarrow s_{n+1}$  is a black transition. Then  $e \leq_{M,P} g$  and let  $\mu_{M,P}: e \rightarrow g$  be the corresponding (surjective) map. We construct a map  $\mu': e' \rightarrow \mathbb{N}$  by taking

$$\mu'(e) = \begin{cases} e + \mu_{M,P}(r(e)) - r(e), & \text{if } \mu_{M,P}(r(e)) > r(e); \\ e + p_{\mathcal{O},\mathcal{D}} - r(e) + \mu_{M,P}(r(e)), & \text{otherwise,} \end{cases}$$

for each  $e \in e'$ . We set  $h(s') = f'g'$ , where  $g' = \mu'(e')$  and  $f' = \nabla(e, g')$ . We note that  $r(f') = f$  and  $r(g') = g$ , so IH continues to hold. The case when  $s_n \rightarrow s_{n+1}$  is a red transition is similar and left to the reader. It can be readily verified that  $h$  is a homomorphism from  $\mathfrak{T}'$  to  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$ .

Now, we show that  $\mathfrak{T}_{\mathcal{X}}$  is homomorphically embeddable into  $\mathfrak{T}_S$  for each  $\mathcal{X}$  such that  $\mathcal{O}, \mathcal{D} \models \mathcal{X}$ . Take any  $\mathcal{X}$  such that  $\mathcal{O}, \mathcal{D} \models \mathcal{X}$ . It follows that  $\mathfrak{T}_{\mathcal{X}}$  is homomorphically embeddable into  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$ . It remains to show that  $\mathfrak{T}_{\mathcal{C}_{\mathcal{O},\mathcal{D}}}$  is homomorphically embeddable into  $S$ . We define  $h$  so that  $h(0) = 0$ ,  $h(s) = u$  for  $s = s_0 \rightarrow s_1 \rightarrow \dots \rightarrow u$ . For  $s = s_0 \rightarrow s_1 \rightarrow \dots \rightarrow de$ , we set  $h(s) = r(d)r(e)$ . It is readily verified that  $h$  is homomorphism.  $\dashv$

To justify the 2EXPTIME upper bound for  $\text{QBE}(LTL_{\text{hom}}^{\square\circ}, \mathcal{Q}[U])$ , we observe that  $S$  above representing  $\mathcal{O}, \mathcal{D}$  can be constructed in time  $O(2^{2^{|\mathcal{O}|+|\mathcal{D}|}})$  as  $S$  has such number of states. To justify the PSPACE upper bound, let  $N_i$ , for  $i \in I^+ \cup I^-$ , be the number of states in  $S$  representing  $\mathcal{O}, \mathcal{D}_i$  of the form either  $u, z$  or  $de$ , for  $(d \cup e) \cap [0, \max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}] = \emptyset$ . We set  $N^+ = \prod_{i \in I^+} N_i$ . Similarly to the argument after Lemma 22, we observe that if there exists a finite subtree  $\mathfrak{T}$  of  $\mathfrak{T}_{\mathfrak{P}}$  that is not homomorphically embeddable into  $\mathfrak{U}$ , then there exists such  $\mathfrak{T}$  satisfying the property that every  $s$  from  $\mathfrak{P}$  occurs on each path of  $\mathfrak{T}$  at most  $K^- = \max_{i \in I^-} \{\max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i} + N_i\}$ -many times. It follows that the required  $\mathfrak{T}$ , if exists, is a subtree of  $\mathfrak{T}_{\mathfrak{P}}^M$  for  $M = \max_{i \in I^+} \{\max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}\} + N^+ K^-$ . Indeed, in the required  $\mathfrak{T}$ , if there is a path that is longer than  $M$ , the property above would be violated. By our construction of  $S^i$ , any  $n$ -th element, for  $n \geq \max_{i \in I^+} \{\max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}\}$ , of any path of  $\mathfrak{T}_{\mathfrak{P}}$  is of the form  $(t_1, \dots, t_{|I^+|})$ , where  $t_i$  is either  $u, z$  or  $de$  satisfying  $(d \cup e) \cap [0, \max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}] = \emptyset$ . Observe that  $(t_1, \dots, t_{|I^+|}) \rightarrow (s_1, \dots, s_{|I^+|})$  in  $\mathfrak{P}$ , for  $(t_1, \dots, t_{|I^+|})$  as above, implies  $s_i$  is either  $u, z$  or  $de$  satisfying  $(d \cup e) \cap [0, \max \mathcal{D}_i + s_{\mathcal{O},\mathcal{D}_i}] = \emptyset$ . Any sequence  $(t_1, \dots, t_{|I^+|}) \rightarrow \dots \rightarrow (s_1, \dots, s_{|I^+|})$  as above in  $\mathfrak{P}$  longer than  $N^+ K^-$  will have some  $(t_1, \dots, t_{|I^+|})$  repeated more than  $K^-$  times.

## D Proofs for Section 5

Let  $E^+ = \{\mathcal{D}_1^+, \dots, \mathcal{D}_n^+\}$  and  $E^- = \{\mathcal{D}_1^-, \dots, \mathcal{D}_l^-\}$  and let  $\mathcal{O}$  be a  $LTL_{\text{hom}}^{\square\circ}$  ontology. For every  $\mathcal{C}_{\mathcal{O},\mathcal{D}}$  with  $\mathcal{D} \in E^+ \cup E^-$ , let  $s_{\mathcal{O},\mathcal{D}} \leq 2^{|\mathcal{O}|}$  and  $p_{\mathcal{O},\mathcal{D}} \leq 2^{2^{|\mathcal{O}|}}$  be the length of the ‘handle’ and the length of the ‘period’ in  $\mathcal{C}_{\mathcal{O},\mathcal{D}}$ , respectively (provided by Proposition 9). Set

$$k = \max_{\mathcal{D} \in E^+ \cup E^-} (\max \mathcal{D} + s_{\mathcal{O},\mathcal{D}}), \quad m = \prod_{\mathcal{D} \in E^+ \cup E^-} p_{\mathcal{O},\mathcal{D}}.$$

**Lemma 29.** (i) If  $E$  is  $\mathcal{Q}[\circ, \diamond]$ -separable under  $\mathcal{O}$ , then it is separated by a conjunction of at most  $l$ -many  $\mathcal{X} \in \mathcal{Q}_p^\circ[\diamond]$  of  $\diamond$ -depth  $\leq k + 1$  and  $\circ$ -depth  $\leq k + m$ .

(ii) If  $E$  is  $\mathcal{Q}[\diamond]$ -separable under  $\mathcal{O}$ , then it is separated by a conjunction of at most  $l$ -many  $\mathcal{X} \in \mathcal{Q}_p[\diamond]$  of  $\diamond$ -depth  $\leq k + 1$ .

(iii) If  $E$  is  $\mathcal{Q}_p[\circ, \diamond]$ -separable under  $\mathcal{O}$ , then it is separated by some  $\mathcal{X} \in \mathcal{Q}_p[\circ, \diamond]$  of  $\diamond$ -depth  $\leq k + l$  and  $\circ$ -depth  $\leq k + m$ .

(iv) If  $E$  is  $\mathcal{Q}_p[\diamond]$ -separable under  $\mathcal{O}$ , then it is separated by some  $\mathcal{X} \in \mathcal{Q}_p[\diamond]$  of  $\diamond$ -depth  $\leq k + l$ .

*Proof.* Recall that the types of any  $\mathcal{C}_{\mathcal{O},\mathcal{D}}$  form a sequence

$$tp_0, \dots, tp_k, tp_{k+1}, \dots, tp_{k+m}, \dots, tp_{k+1}, \dots, tp_{k+m}, \dots$$

(i) Recall from the proof of Theorem 4 (i.3) that we may assume that for any  $\mathcal{D}^- \in E^-$  there is a query  $\mathcal{X} \in \mathcal{Q}_p^\circ[\diamond]$  that separates  $(E^+, \{\mathcal{D}^-\})$ . So let  $\mathcal{D}^- \in E^-$  and assume that  $(E^+, \{\mathcal{D}^-\})$  is  $\mathcal{Q}_p^\circ[\diamond]$ -separable under  $\mathcal{O}$ . Then there is a separator

$$\mathcal{X} = \rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond \rho_n))$$

in which each  $\rho_r$  has  $\circ$ -depth  $\leq k + m$ . Indeed, in view of the form of the canonical models, if  $\rho_r = \bigwedge_{i=0}^{\ell} \circ^i \lambda_i$  with  $\ell > k + m$ , then one can replace  $\rho_r$  with

$$\rho'_r = \bigwedge_{i=0}^k \circ^i \lambda_i \wedge \bigwedge_{j=1}^m \circ^{k+j} \bigwedge_{\substack{0 \leq i \leq \ell \\ j = (i-k) \bmod m}} \lambda_i.$$

In addition, if  $n > k$ , then the query

$$\rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond \rho_k)) \wedge \bigwedge_{i=k+1}^n \diamond^{k+1} \rho_i$$

still separates  $(E^+, \{\mathcal{D}^-\})$  under  $\mathcal{O}$ , and so some  $\rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond(\rho_k \wedge \diamond \rho_j)))$  with  $k < j \leq n$  separates  $(E^+, \{\mathcal{D}^-\})$  under  $\mathcal{O}$ .

(ii) is proved by dropping the  $\circ$ -queries from the proof of (i).

(iii) The proof of (i) shows that if  $E$  is  $\mathcal{Q}_p[\circ, \diamond]$ -separable under  $\mathcal{O}$ , then there is a separator

$$\rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond \rho_k)) \wedge \bigwedge_{i=k+1}^n \diamond^{k+1} \rho_i$$

in which each  $\rho_r$  has  $\circ$ -depth  $\leq k + m$ . Now we can select, for each negative example  $\mathcal{D}^-$ , a  $j$  such that

$$\rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond(\rho_k \wedge \diamond \rho_j)))$$

separates  $(E^+, \{\mathcal{D}^-\})$  under  $\mathcal{O}$ . Let  $j_1, \dots, j_l$  be thus selected. Then

$$\rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond(\rho_k \wedge \diamond(\rho_{j_1} \wedge \dots \wedge \diamond \rho_{j_l}))))$$

separates  $E$  under  $\mathcal{O}$ .

(iv) is proved by dropping the  $\circ$ -queries from the proof of (iii). \(\dashv\)

**Theorem 10.** Let  $\mathcal{Q} \in \{\mathcal{Q}[\circ, \diamond], \mathcal{Q}[\diamond], \mathcal{Q}_p[\circ, \diamond], \mathcal{Q}_p[\diamond]\}$ . Then  $\text{QBE}(LTL_{\text{hom}}^{\square \circ}, \mathcal{Q})$  and  $\text{QBE}_{\text{b}}^{\text{b}+}(LTL_{\text{hom}}^{\square \circ}, \mathcal{Q})$  are both PSPACE-complete for combined complexity.

*Proof.* The lower bound follows from [Chen and Lin, 1994]. We first give the upper bound proof for  $\mathcal{Q}_p[\diamond]$ .

Let  $E^+ = \{\mathcal{D}_1^+, \dots, \mathcal{D}_n^+\}$  and  $E^- = \{\mathcal{D}_1^-, \dots, \mathcal{D}_l^-\}$ . We use Lemma 29 (iv). Let  $k$  and  $m$  be as in Lemma 29. The nondeterministic algorithm starts by guessing a conjunction of atoms  $\rho_0$  and checking in PSPACE that  $\mathcal{O}, \mathcal{D}_i^+ \models \rho_0(0)$  for all  $i \in [1, n]$ . We use numbers  $d_i^+, d_j^- \leq k + m$ , for  $i \in [1, n], j \in [1, l]$ , and a set  $N \subseteq [1, l]$  that will keep track of the negative examples yet to be separated. Initially, we set all  $d_i^+, d_j^- = 0$  and  $N = \{j \in [1, l] \mid \mathcal{O}, \mathcal{D}_j^- \models \rho_0(0)\}$ . Then we repeat the following steps until  $N = \emptyset$ , in which case the algorithm terminates accepting the input:

- Guess a conjunction  $\rho$  of atoms in the signature of  $\mathcal{O}$  and  $E$ .
- For every  $i \in [1, n]$ , check in PSPACE that  $\mathcal{O}, \mathcal{D}_i^+ \models \diamond \rho(d_i^+)$  and reject if this is not so.
- Guess  $d_i^{+'}$  such that  $\min(d_i^+, k) < d_i^{+'} \leq k + m$  and  $\mathcal{O}, \mathcal{D}_i^+ \models \rho(d_i^{+'})$ .
- For each,  $j \in N$  check that  $\mathcal{O}, \mathcal{D}_j^- \models \diamond \rho(d_i^-)$ . If no, remove  $j$  from  $N$ . Otherwise, find in PSPACE the smallest  $d_i^{-'}$  such that  $\min(d_i^-, k) < d_i^{-'} \leq k + m$  and  $\mathcal{O}, \mathcal{D}_j^- \models \rho(d_i^{-'})$ .
- Set  $d_i^+ := d_i^{+'}$  and, for all  $j$  still in  $N$ , set  $d_j^- := d_i^{-'}$ .

Let  $\varphi_i = \rho_0 \wedge \diamond(\rho_1 \wedge \diamond(\dots (\rho_{i-1} \wedge \diamond \rho_i) \dots))$ , where  $\rho_i$  is the conjunction of atoms guessed in the  $i$ -th iteration. Let  $N_i$  be the set  $N$  after the  $i$ -th iteration. Then, for all  $j \in [1, n]$ , we have  $\mathcal{O}, \mathcal{D}_j^+ \models \varphi_i(0)$ , for all  $j \in N_i$  we have  $\mathcal{O}, \mathcal{D}_j^- \models \varphi_i(0)$ , and for all  $j \in [1, l] \setminus N_i$  we have  $\mathcal{O}, \mathcal{D}_j^- \not\models \varphi_i(0)$ . So the algorithm accepts after the  $\ell$ -th iteration iff  $\varphi_\ell$  separates  $(E^+, E^-)$ .

By Lemma 29 (ii), this also gives a PSPACE algorithm for  $\mathcal{Q}[\diamond]$ . By Lemma 29 (i), for  $\mathcal{Q}[\circ, \diamond]$  it suffices to give a PSPACE algorithm for  $\mathcal{Q}_p^{\circ}[\diamond]$ , which can be obtained by modifying the algorithm above.

It starts by guessing a conjunction of atoms  $\lambda_0$  and checking that  $\mathcal{O}, \mathcal{D}_i^+ \models \lambda_0(0)$  for all  $i \in [1, n]$ , which can be done in PSPACE. We use numbers  $d_i^+, d_i^- \leq k + m$  (for  $\diamond$ -subformulas) and  $c \leq k + m$  (for  $\circ$ -formulas), and a set  $N \subseteq [1, l]$  that will keep track of the negative examples yet to be separated. Initially, we set all  $d_i^+, d_j^- = 0$ ,  $c = 0$ , and  $N = [1, l]$ . Then we repeat (1) or (2) until  $N = \emptyset$ , in which case the algorithm terminates accepting the input:

- (1) – Set  $c = 0$ .
  - Guess a conjunction  $\lambda$  of atoms in the signature of  $\mathcal{O}$  and  $E$ .
  - For every  $i \in [1, n]$ , check in PSPACE that  $\mathcal{O}, \mathcal{D}_i^+ \models \diamond \lambda(d + i)$  and reject if this is not so.
  - Guess  $d_i^{+'}$  such that  $\min(d +, k) < d^{+'} \leq k + m$  and  $\mathcal{O}, \mathcal{D}_i^+ \models \lambda(d^{+'})$ .
  - For each  $j \in N$ , check that  $\mathcal{O}, \mathcal{D}_j^- \models \diamond \lambda(d_i^-)$ . If no, remove  $j$  from  $N$ . Otherwise, find in PSPACE the smallest  $d_i^{-'}$  such that  $\min(d_i^-, k) < d_i^{-'} \leq k + m$  and  $\mathcal{O}, \mathcal{D}_j^- \models \lambda(d_i^{-'})$ .
  - Set  $d_i^+ = d_i^{+'}$  and, for all  $j$  still in  $N$ , set  $d_j^- = d_i^{-'}$ .
- (2) – Increment  $c$ , provided  $c < m + k$ .

- Guess a conjunction  $\lambda$  of atoms in the signature of  $\mathcal{O}$  and  $E$ .
- For every  $i \in [1, n]$ , check in PSPACE that  $\mathcal{O}, \mathcal{D}_i^+ \models \lambda(d_i^+ + c)$  and reject if this is not so.
- For each  $j \in N$ , check that  $\mathcal{O}, \mathcal{D}_j^- \models \lambda(d_j^- + c)$ . If no, remove  $j$  from  $N$ .

A PSPACE algorithm for  $\mathcal{Q}_p[\circ, \diamond]$  is similar to the one above: it uses Lemma 29 (iii) for guessing the next temporal operator  $\circ$  or  $\diamond$  in the query.  $\dashv$

**Theorem 11.**  $\text{QBE}(LTL_{\text{horn}}^{\square\circ}, \mathcal{Q}[U_s])$  is in EXPTIME for combined complexity,  $\text{QBE}(LTL_{\text{horn}}^{\square\circ}, \mathcal{Q}_p[U])$  is in EXPSPACE, and  $\text{QBE}_{\text{b}}^{\text{b}+}(LTL_{\text{horn}}^{\square\circ}, \mathcal{Q}_p[U])$  is NEXPTIME-hard.

*Proof.* The upper bound follows from Lemmas 20 and 22 above as explained in the main part of the paper.

Now we establish the NEXPTIME lower bound. Let  $M$  be a non-deterministic Turing machine that accepts words  $x$  over its tape alphabet in at most  $N = 2^{p(|x|)}$  steps, for some polynomial  $p$ . Given such an  $M$  and an input  $x$ , our aim is to define an  $LTL_{\text{horn}}^{\square\circ}$  ontology  $\mathcal{O}$  and an example set  $E = (E^+ = \{\mathcal{D}_1^+, \mathcal{D}_2^+\}, E^- = \{\mathcal{D}^-\})$  of size polynomial in  $M$  and  $x$  such that  $E$  is separated by a  $\mathcal{Q}_p[U]$ -query under  $\mathcal{O}$  iff  $M$  accepts  $x$ .

Suppose  $M$  has a set  $Q$  of states, tape alphabet  $\Sigma$  with  $\text{b}$  for blank, initial state  $q_0$ , and accepting state  $q_{\text{acc}}$ . Without loss of generality we assume that  $M$  erases the tape before accepting and its head is at the left-most cell in any accepting configuration.

Given an input word  $x = x_1 \dots x_n$  over  $\Sigma$ , we represent configurations  $c$  of a computation of  $M$  on  $x$  by the  $(N - 1)$ -long word written on the tape (with sufficiently many blanks at the end), in which the symbol  $y$  in the active cell is replaced by the pair  $(q, y)$  with the current state  $q$ . An accepting computation of  $M$  on  $x$  is encoded by the word  $w = \#c_1\#c_2\#\dots\#c_{N-1}\#c_N$  over the alphabet  $\Xi = \Sigma \cup (Q \times \Sigma) \cup \{\#\}$ , where  $c_1, c_2, \dots, c_N$  are the subsequent configurations in the computation. In particular,  $c_1$  is the initial configuration  $(q_0, x_1)x_2 \dots x_n \text{b} \dots \text{b}$ , and  $c_N$  is the accepting configuration  $c_{\text{acc}} = (q_{\text{acc}}, \text{b})\text{b} \dots \text{b}$ . Thus, any accepting computation is encoded by a word of length  $N^2$  in the alphabet  $\Xi$  (we allow  $c_{\text{acc}}$  to follow  $c_{\text{acc}}$ ).

A tuple  $t = (a, b, c, d, e, f) \in (\Xi)^6$  is called *legal* [Sipser, 1997, Theorem 7.37] if there exist two consecutive configurations  $c_1$  and  $c_2$  of  $M$  and a number  $i$  such that

$$abcdef = c_1[i]c_1[i+1]c_1[i+2]c_2[i]c_2[i+1]c_2[i+2],$$

where  $c_j[i]$  is the  $i$ th symbol in  $c_j$ . Let  $\mathfrak{L} \subseteq (\Xi)^6$  be the set of all legal tuples (plus a few additional 6-tuples to take care of  $\#$ ) with the following property: a word  $w$  encodes an accepting computation iff it starts with the initial configuration preceded by  $\#$ , ends with the accepting configuration, and every two length 3 subwords at distance  $N$  apart form a legal tuple. Let  $\mathfrak{L} = (\Xi)^6 \setminus \mathfrak{L}$ .

For any  $k > 0$ , by a *k-counter* we mean a set  $\mathbb{A} = \{A_j^i \mid i = 0, 1, j = 1, \dots, k\}$  of atomic concepts that will be used to store values between 0 and  $2^k - 1$ , which can be different at different time points. The counter  $\mathbb{A}$  is *well-defined* at a time point  $n \in \mathbb{N}$  in an interpretation  $\mathcal{I}$  if  $\mathcal{I}, n \models A_j^0 \wedge A_j^1 \rightarrow \perp$  and  $\mathcal{I}, n \models A_j^0 \vee A_j^1$ , for any  $j = 1, \dots, k$ . In this case, the *value of  $\mathbb{A}$*  at  $n$  in  $\mathcal{I}$  is given by the unique binary number  $b_k \dots b_1$  for which  $\mathcal{I}, n \models A_1^{b_1} \wedge \dots \wedge A_k^{b_k}$ . We require the following formulas, for  $c = b_k \dots b_1$  (provided that  $\mathbb{A}$  is well-defined):

- $[\mathbb{A} = c] = A_1^{b_1} \wedge \dots \wedge A_k^{b_k}$ , for which  $\mathcal{I}, n \models [\mathbb{A} = c]$  iff the value of  $\mathbb{A}$  is  $c$ ;
- $[\mathbb{A} < c] = \bigvee_{\substack{k \geq i \geq 1 \\ b_i = 1}} (A_i^0 \wedge \bigwedge_{j=i+1}^k A_j^{b_j})$  with  $\mathcal{I}, n \models [\mathbb{A} < c]$  iff the value of  $\mathbb{A}$  is  $< c$ ;
- $[\mathbb{A} > c] = \bigvee_{\substack{k \geq i \geq 1 \\ b_i = 0}} (A_i^1 \wedge \bigwedge_{j=i+1}^k A_j^{b_j})$  with  $\mathcal{I}, n \models [\mathbb{A} > c]$  iff the value of  $\mathbb{A}$  is  $> c$ .

We regard the set  $(\circ_F \mathbb{A}) = \{\circ_F A_j^i \mid i = 0, 1, j = 1, \dots, k\}$  as another counter that stores at  $n$  in  $\mathcal{I}$  the value stored by  $\mathbb{A}$  at  $n + 1$  in  $\mathcal{I}$ . Thus, we can use formulas like  $[\mathbb{A} > c_1] \rightarrow [(\circ_F \mathbb{A}) = c_2]$ , which says that if the value of  $\mathbb{A}$  at  $n$  in  $\mathcal{I}$  is greater than  $c_1$ , then the value of  $\mathbb{A}$  at  $n + 1$  in  $\mathcal{I}$  is  $c_2$ . Also, for  $l \leq k$ , we can use formulas like  $[\mathbb{A} = i \pmod{2^l}]$  with self-explaining meaning. Another important formula we need is defined by:

$$[\mathbb{A} = \mathbb{B} + 1] = \bigwedge_{i=1}^k ((B_i^0 \wedge B_{i-1}^1 \wedge \dots \wedge B_1^1 \rightarrow A_i^1 \wedge A_{i-1}^0 \wedge \dots \wedge A_1^0) \wedge \bigwedge_{j < i} ((B_i^0 \wedge B_j^0 \rightarrow A_i^0) \wedge (B_i^1 \wedge B_j^0 \rightarrow A_i^1))).$$

It says that the value of  $\mathbb{A}$  is one greater than the value of  $\mathbb{B}$ .

To define  $\mathcal{O}$  and  $E = (E^+, E^-)$  for given  $M$  and  $x = x_1 \dots x_n$ , we assume that  $\Xi = \{a_1, \dots, a_{2m}\}$  and  $k = 6m + 2 \lceil \log N \rceil + 1$ . We use the following atomic concepts in  $\mathcal{O}$  and  $E$ : the symbols in  $\Xi$ , the atoms  $C, S, N, T$ , and those atoms that are needed in  $k$ -counters  $\mathbb{S}, \mathbb{N}, \mathbb{T}$ .

We set  $\mathcal{D}_1^+ = \{T(0)\}$ ,  $\mathcal{D}_2^+ = \{S(0)\}$ , and  $\mathcal{D}^- = \{N(0)\}$ .

The following axioms initialise the corresponding  $m$ -counters:

$$T \rightarrow [(\circ \mathbb{T}) = 0] \quad S \rightarrow [(\circ \mathbb{S}) = 0] \quad N \rightarrow [(\circ \mathbb{N}) = 0].$$

These and all other axioms of  $\mathcal{O}$  can be easily transformed to equivalent sets of polynomially-many  $LTL_{\text{horn}}^{\square\circ}$  axioms.

The behaviour of each counter is specified by the axioms below whose meaning is illustrated by the structure of the canonical model of the corresponding example restricted to  $\Xi \cup \{C\}$ .

The  $T$ -axioms

$$\begin{aligned} [\mathbb{T} < N^2] &\rightarrow [\circ\mathbb{T} = \mathbb{T} + 1], \\ [\mathbb{T} = 0] &\rightarrow \sharp, \quad [\mathbb{T} = 1] \rightarrow (q_1, x_1), \\ [\mathbb{T} = 2] &\rightarrow x_2, \dots, [\mathbb{T} = n] \rightarrow x_n, \\ [\mathbb{T} > n] \wedge [\mathbb{T} < N] &\rightarrow \mathbf{b}, \\ [\mathbb{T} > N] \wedge [\mathbb{T} < N^2 - N] &\rightarrow \Xi, \\ [\mathbb{T} = N^2 - N] &\rightarrow \sharp, \\ [\mathbb{T} = N^2 - N + 1] &\rightarrow (q_{acc}, \mathbf{b}), \\ [\mathbb{T} > N^2 - N + 1] &\rightarrow \mathbf{b} \end{aligned}$$

together with the data instance  $\mathcal{D}_1^+$  give rise to the canonical model of the form

$$\emptyset, \sharp, (q_1, x_1), x_2, \dots, x_n, \mathbf{b}^{N-n-1}, \sharp, \Xi^{N^2-2N-1}, \sharp, (q_{acc}, \mathbf{b}), \mathbf{b}^{N-2}, \dots$$

The  $S$ -axioms

$$\begin{aligned} [\mathbb{S} < (2|\Xi| + 1)N^2] &\rightarrow [\circ\mathbb{S} = \mathbb{S} + 1], \\ [\mathbb{S} > N^2] &\rightarrow C, \\ [\mathbb{S} > N^2 \wedge \mathbb{S} = 2i \pmod{2^{m+1}}] &\rightarrow a_{i+1}, \text{ for all } i \in [1, 2^m] \end{aligned}$$

and  $\mathcal{D}_2^+$  generate the canonical model

$$\emptyset, \emptyset^{N^2}, \emptyset, (a_1 C, C, \dots, a_{2^k} C, C)^{N^2}, \emptyset, \emptyset, \dots$$

Let  $\bar{\mathcal{L}} = \{t_1 = (a_1, b_1, c_1, d_1, e_1, f_1), \dots, t_l\}$ . Let  $t = N^2 - N - 3$ . The  $N$ -axioms comprise the following, for each  $i \in [1, l]$ :

$$\begin{aligned} [\mathbb{N} < (2l + 3)N^2] &\rightarrow [\circ\mathbb{N} = \mathbb{N} + 1], \\ [0 < \mathbb{N} < N^2] &\rightarrow \Xi, \\ [N^2 < \mathbb{N} < 3N^2] &\rightarrow C, \\ [N^2 < \mathbb{N} < 3N^2] \wedge N_0^1 &\rightarrow \Xi, \\ [(2i + 1)N^2 < \mathbb{N} < (2i + 1)N^2 + t + 1] &\rightarrow \Xi, \\ [\mathbb{N} = (2i + 1)N^2 + t + 1] &\rightarrow a_i, \\ [\mathbb{N} = (2i + 1)N^2 + t + 2] &\rightarrow b_i, \\ [\mathbb{N} = (2i + 1)N^2 + t + 3] &\rightarrow c_i, \\ [(2i + 2)N^2 - N < \mathbb{N} < (2i + 2)N^2 - 2] &\rightarrow \Xi, \\ [\mathbb{N} = (2i + 2)N^2 - 2] &\rightarrow d_i, \\ [\mathbb{N} = (2i + 2)N^2 - 1] &\rightarrow e_i, \\ [\mathbb{N} = (2i + 2)N^2] &\rightarrow f_i, \\ [(2i + 2)N^2 < \mathbb{N} < (2i + 2)N^2 + t + 1] &\rightarrow \Xi. \end{aligned}$$

The data instance  $\mathcal{D}^-$  gives the canonical model

$$\emptyset, \emptyset, \Xi^{N^2-1}, \emptyset, (\Xi C, C)^{N^2-2}, \Xi C, \emptyset, \mathcal{D}_{t_1}, \emptyset^{N+2}, \mathcal{D}_{t_2}, \dots, \mathcal{D}_{t_l}, \emptyset, \dots$$

where  $\mathcal{D}_{t_i} = \emptyset, \Xi^t, a_i, b_i, c_i, \Xi^{N-3}, d_i, e_i, f_i, \Xi^t$ .

We denote the set of the axioms above by  $\mathcal{O}$  and show that  $E$  is separated by a  $\mathcal{Q}_p[U]$ -query  $\varkappa$  under  $\mathcal{O}$  iff  $M$  accepts  $\mathbf{x}$ .

( $\Leftarrow$ ) Suppose  $\rho_1 \dots \rho_{N^2}$  encodes an accepting computation of  $M$  on  $\mathbf{x}$ . Consider the  $\mathcal{Q}_p[U]$ -query

$$\varkappa = \diamond(\rho_1 \wedge C \cup (\rho_2 \wedge C \cup (\dots (\rho_{N^2-1} \wedge (C \cup \rho_{N^2})) \dots))).$$

It is not hard to show by inspecting the respective canonical models described above that

$$\mathcal{O}, \mathcal{D}_1^+ \models \varkappa(0), \quad \mathcal{O}, \mathcal{D}_2^+ \models \varkappa(0), \quad \mathcal{O}, \mathcal{D}_1^- \not\models \varkappa(0).$$

To prove the last one, we first notice that  $\emptyset, \Xi^{N^2-1} \not\models \varkappa(0)$ , and  $\emptyset, (\Xi C, C)^{N^2-2}, \Xi C \not\models \varkappa(0)$ . We have  $\mathcal{D}_t \not\models C(j)$  for all  $t$  and  $j$ . So, if  $\mathcal{D}_t \models \varkappa(0)$  for some  $t$ , then there is  $i < t$  such that  $\emptyset, \mathcal{D}_t \models \rho_j(i+j)$  for all  $j \in [1, N^2]$ . But then  $\rho_{t-i}\rho_{t-i+1}\rho_{t-i+2}\rho_{t+N-i}\rho_{t+N-i+1}\rho_{t+N-i+2} = t \in \bar{\mathcal{L}}$ , which is a contradiction. So we have  $\mathcal{D}_t \not\models \varkappa(0)$  for all  $t \in \bar{\mathcal{L}}$ , and therefore  $\mathcal{O}, \mathcal{D}^- \not\models \varkappa(0)$ .

( $\Rightarrow$ ) Suppose the query

$$\varkappa = \lambda_1 \text{ U } (\rho_1 \wedge \lambda_2 \text{ U } (\rho_2 \dots (\rho_{K-1} \wedge (\lambda_K \text{ U } \rho_K)) \dots))$$

with  $\rho_K \neq \top$  separates  $E$  under  $\mathcal{O}$ . Since  $\mathcal{O}, \mathcal{D}_1^+ \models \varkappa(0)$ , we have  $K \leq N^2$  and  $\rho_i \subseteq \Xi$  for all  $i$ . Since  $\emptyset, \emptyset, \Xi^{N^2-1} \not\models \varkappa(0)$  we have  $\rho_1 \neq \emptyset$ . Since  $\mathcal{O}, \mathcal{D}_2^+ \models \varkappa(0)$  we have  $\lambda_1 = \top$ . Now if  $K < N^2$ , then  $\emptyset, \emptyset, \Xi^{N^2-1} \models \varkappa(0)$ , so  $K = N^2$ .

Since  $\mathcal{O}, \mathcal{D}_2^+ \models \varkappa(0)$ , we have  $|\rho_i| \leq 1$  for all  $i$ . Let  $y_1 < \dots < y_{N^2}$  be such that  $\mathcal{O}, \mathcal{D}_2^+ \models \rho_j(y_j)$  and  $\mathcal{O}, \mathcal{D}_2^+ \models \lambda(i)$  for all  $j \in [1, N^2]$  and  $i \in (y_j, y_{j+1})$ . We see that if  $y_j$  is odd, then  $\rho_j = \emptyset$  and if  $y_j$  is even we can assume that  $\rho_j = a \in \Xi$  where  $\mathcal{O}, \mathcal{D}_2^+ \models a(y_j)$ . Let construct  $z_j$  in the following way:  $z_1 = N^2 + 2$  and if we already have  $z_j$ , then  $z_{j+1}$  is the smallest number bigger than  $z_j$  with the same parity as  $y_{j+1}$ . We can see that, for all  $j < N^2$ , we have  $\mathcal{O}, \mathcal{D}^- \models \rho_j(z_j)$  with  $\mathcal{O}, \mathcal{D}^- \models \lambda_j(y)$  for all  $y \in (z_j, z_{j+1})$  and if there is an odd  $y_j$ , then  $z_{N^2} < 3N^2 - 1$ ,  $\mathcal{O}, \mathcal{D}^- \models \rho_{N^2}(z_{N^2})$ , and therefore  $\mathcal{O}, \mathcal{D}^- \models \varkappa(0)$  which cannot happen. So there are no odd  $y_j$ 's and  $|\rho_i| = 1$  for all  $i$ .

In view of  $\mathcal{O}, \mathcal{D}_1^+ \models \varkappa(0)$ , the word  $\rho_1 \dots \rho_{N^2}$  starts with the starting configuration preceded by  $\sharp$  and ends with the accepting one. Suppose there is some  $i$  such that  $(\rho_i, \rho_{i+1}, \rho_{i+2}, \rho_{N+i}, \rho_{N+i+1}, \rho_{N+i+2}) = t \in \bar{\mathcal{L}}$ . Let  $y_j = t - i + j$  for  $j \in [1, N^2]$ . We have  $\mathcal{D}_t \models \rho_j(y_j)$ , and so  $\mathcal{D}_t \models \varkappa(0)$ , and therefore  $\mathcal{O}, \mathcal{D}^- \models \varkappa(0)$ . So every two length 3 subwords at distance  $N$  apart form a legal tuple and  $\rho_1 \dots \rho_{N^2}$  encodes a successful computation of  $M$  on  $x$ .  $\dashv$

**Theorem 12.** For data complexity, the results of Theorem 5 continue to hold for queries mediated by an  $LTL_{horn}^{\square\circ}$ -ontology.

We first consider  $\circ\Diamond$ -queries, and then come to U-queries.

$\circ\Diamond$ -queries. The NP-lower bounds are inherited from the ontology-free case. For the NP-upper bounds observe that by Lemma 29 and since  $\mathcal{O}$  is fixed, we always have a separating query of polynomial size whenever a separating query exists. The NP-upper bounds follow in the standard way. We now come to the P-upper bounds. We prove the P-upper bound for  $QBE_{\mathcal{D}}^{\text{b}+}(LTL_{horn}^{\square\circ}, \mathcal{Q}_p[\circ, \Diamond])$  by modifying the dynamic programming algorithm we gave in the ontology-free case. The P-upper bound for  $QBE_{\mathcal{D}}^{\text{b}+}(LTL_{horn}^{\square\circ}, \mathcal{Q}_p[\Diamond])$  is obtained by dropping  $\circ$  from the proof. The P-upper bound for  $QBE(LTL_{horn}^{\square\circ}, \mathcal{Q}[\Diamond])$  with a bounded number of positive examples can be again proved in two steps: (1) by Theorem 4 (i.1) it suffices to prove the P-upper bound for  $QBE_{\mathcal{D}}^{\text{b}+}(LTL_{horn}^{\square\circ}, \mathcal{Q}[\Diamond])$ ; (2) by Theorem 4 (i.3),  $QBE_{\mathcal{D}}^{\text{b}+}(LTL_{horn}^{\square\circ}, \mathcal{Q}[\Diamond]) \leq_p QBE_{\mathcal{D}}^{\text{b}+}(LTL_{horn}^{\square\circ}, \mathcal{Q}_p[\Diamond])$ . Finally, the P-upper bound for  $QBE(LTL_{horn}^{\square\circ}, \mathcal{Q}[\circ, \Diamond])$  with a bounded number of positive examples follows from the P-upper bound for  $QBE(LTL_{horn}^{\square\circ}, \mathcal{Q}[\Diamond])$  with a bounded number of positive examples using the same ‘trick’ as in the proof of Theorem 4 (ii.2) in the ontology-free case: we modify the models  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  by adding fresh atoms  $A_i$  encoding  $\circ^i A$  and interpreting them in  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  in the same way as  $\circ^i A$ . By Lemma 29 it suffices to do this for  $i \leq k + m$  (which is polynomial in  $E$  as  $|\mathcal{O}|$  is fixed).

To prove the P-upper bound for  $\mathcal{Q}_p[\circ, \Diamond]$ , we extend the notion of a satisfying assignment for a query  $\varkappa$  in a data instance to a satisfying assignment in the canonical model  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  in the obvious way: suppose  $\varkappa$  takes the form (1) with  $\rho_n \neq \top$ . Then  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}, 0 \models \varkappa$  iff there is a satisfying assignment  $f$  for  $\varkappa$  in  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  in the sense that  $f$  is a strictly monotone map  $f: [0, n] \rightarrow \mathbb{N}$  with  $f(0) = 0$ ,  $f(i+1) = f(i) + 1$  if  $\circ_i = \circ$ , and  $\rho_i \subseteq tp(f(i)) = \{A \mid \mathcal{C}_{\mathcal{O}, \mathcal{D}}, f(i) \models A\}$ , for all  $i \leq n$ . We first observe the following lemma (using the notation and numbers  $k, m$  introduced for Lemma 29):

**Lemma 30.**  $(E^+, E^-)$  is  $\mathcal{Q}_p[\circ, \Diamond]$ -separable under  $\mathcal{O}$  iff there exists  $\varkappa$  of the form (1) with  $\rho_n \neq \top$ , of  $\Diamond$ -depth  $\leq k + l$  and  $\circ$ -depth  $\leq k + m$  such that

1. for any  $\mathcal{D} \in E^+$  there is a satisfying assignment for  $\varkappa$  into  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  with  $f(n) \leq N := k + (k + l + 1)(k + m)$ ;
2. for any  $\mathcal{D} \in E^-$  there is no satisfying assignment for  $\varkappa$  into  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$  with  $f(n) \leq N$ .

*Proof.* First assume that  $(E^+, E^-)$  is  $\mathcal{Q}_p[\circ, \Diamond]$ -separable under  $\mathcal{O}$ . By Lemma 29 there is a query  $\varkappa$  of the form (1) with  $\rho_n \neq \top$ , of  $\Diamond$ -depth  $\leq k + l$  and  $\circ$ -depth  $\leq k + m$  that separates  $(E^+, E^-)$  under  $\mathcal{O}$ . Take any satisfying assignment  $f$  for  $\varkappa$  in  $\mathcal{C}_{\mathcal{O}, \mathcal{D}}$ . Clearly then we can assume that  $f(i+1) - f(i) \leq k + m$  for any  $i$  with  $\circ_i = \Diamond$ . Point 1 follows directly. Point 2 follows from  $\mathcal{O}, \mathcal{D}^- \not\models \varkappa(0)$  for  $\mathcal{D}^- \in E^-$ .

Conversely, assume that there exists  $\varkappa$  of the form (1) with  $\rho_n \neq \top$ , of  $\Diamond$ -depth  $\leq k + l$  and  $\circ$ -depth  $\leq k + m$  such that Points 1 and 2 hold. We show that  $\varkappa$  separates  $(E^+, E^-)$  under  $\mathcal{O}$ . But  $\mathcal{O}, \mathcal{D}^+ \models \varkappa(0)$  for  $\mathcal{D}^+ \in E^+$  follows from Point 1 and  $\mathcal{O}, \mathcal{D}^- \not\models \varkappa(0)$  for  $\mathcal{D}^- \in E^-$  follows from Point 2 using the same argument as in the proof of Point 1 in the converse direction.  $\dashv$

We explain the modifications of the dynamic programming algorithm for  $QBE_{\mathcal{D}}^{\text{b}+}(\mathcal{Q}_p[\circ, \Diamond])$  for  $E^+ = \{\mathcal{D}_1^+, \mathcal{D}_2^+\}$  and  $E^- = \{\mathcal{D}_1^-, \mathcal{D}_2^-\}$ .

We modify the parameters stored in the tuples in the set  $S_{i,j}$  slightly. Instead of the length of the query a tuple describes, we store its  $\Diamond$ -depth  $K$  and its  $\circ$ -depth  $M$ . Thus, let  $S_{i,j}$  be the set of tuples  $(K, M, \ell_1, \ell_2, n_1, n_2)$  such that

1.  $K \leq k + l$ ;
2.  $M \leq k + m$ ;
3.  $\ell_1 \leq i \leq N$ ,
4.  $\ell_2 \leq j \leq N$ ,

and there is  $\varkappa = \rho_0 \wedge \mathbf{o}_1(\rho_1 \wedge \dots \wedge \mathbf{o}_k \rho_k)$  of  $\diamond$ -depth  $K$  and  $\circ$ -depth  $M$  for which

1. there are satisfying assignments  $f_1, f_2$  in  $\mathcal{C}_{\mathcal{O}, \mathcal{D}_1^+}$  and  $\mathcal{C}_{\mathcal{O}, \mathcal{D}_2^+}$  with  $f_1(1 + K + M) = \ell_1$  and  $f_2(1 + K + M) = \ell_2$ , respectively, and
2.  $n_1$  is minimal with a satisfying assignment  $f$  for  $\varkappa$  in  $\mathcal{C}_{\mathcal{O}, \mathcal{D}_1^-}$  such that  $f(1 + K + M) = n_1 \leq N$ , and  $n_1 = \infty$  if there is no such  $f$ ; and  $n_2$  is minimal with a satisfying assignment  $f$  for  $\varkappa$  in  $\mathcal{C}_{\mathcal{O}, \mathcal{D}_2^-}$  such that  $f(k) = n_2 \leq N$ , and  $n_2 = \infty$  if there is no such  $f$ .

It suffices to compute  $S_{N,N}$  in polynomial time because there exists a query in  $\mathcal{Q}_p[\circ, \diamond]$  separating  $(E^+, E^-)$  iff there are  $K \leq k + l$ ,  $M \leq k + m$ ,  $\ell_1 \leq N$ , and  $\ell_2 \leq N$  such that  $(K, M, \ell_1, \ell_2, \infty, \infty) \in S_{N,N}$ .  $S_{i,j}$  with  $i \leq N$  and  $j \leq N$  can be computed in essentially the same way as in the ontology-free case incrementally starting with  $S_{0,0}$ .

The bounds for U-queries were explained in Section C after Lemma 22.

## E Proofs for Section 6

**Theorem 13.** *Let  $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}[\diamond]\}$ . If  $E$  is  $\mathcal{Q}$ -separable under an  $LTL^{\square\diamond}$ -ontology  $\mathcal{O}$ , then  $E$  can be separated under  $\mathcal{O}$  by a  $\mathcal{Q}$ -query of polysize in  $E$  and  $\mathcal{O}$ .  $\text{QBE}(LTL^{\square\diamond}, \mathcal{Q})$  and  $\text{QBE}_b^+(LTL^{\square\diamond}, \mathcal{Q})$  are  $\Sigma_2^p$ -complete for combined complexity. The presence of  $LTL^{\square\diamond}$ -ontologies has no effect on the data complexity, which remains the same as in Theorem 5.*

*Proof.* We start by giving a few more details of the  $\Sigma_2^p$ -lower bound proof. Recall that we reduce the validity problem for fully quantified Boolean formulas of the form

$$\exists \mathbf{p} \forall \mathbf{q} \psi,$$

where  $\psi$  is a propositional formula, and  $\mathbf{p} = p_1, \dots, p_k$  and  $\mathbf{q} = q_1, \dots, q_m$  are lists of propositional variables. We assume w.l.o.g. that  $\psi$  is not a tautology. We also assume that  $\neg\psi \not\models x$  for  $x \in \{p_i, \neg p_i, q_j, \neg q_j \mid 1 \leq i \leq k, 1 \leq j \leq m\}$ . Indeed, if  $\neg\psi \models x$  then  $\psi \equiv \neg x \vee \psi'$ , for some  $\psi'$ , and when  $x \in \{p_i, \neg p_i\}$  the QBF formula  $\exists \mathbf{p} \forall \mathbf{q} \psi$  is vacuously valid whereas when  $x \in \{q_j, \neg q_j\}$  the QBF formula  $\exists \mathbf{p} \forall \mathbf{q} \psi$  is valid iff  $\exists \mathbf{p} \forall \mathbf{q}' \psi'$  is, where  $\mathbf{q}'$  is obtained from  $\mathbf{q}$  by removing  $q_j$ . We regard propositional variables as atoms and also use fresh atoms  $A_1, \dots, A_k, \bar{A}_1, \dots, \bar{A}_k$  and  $B$ .

Let  $E = (E^+, E^-)$  with  $E^+ = \{\mathcal{D}_1, \mathcal{D}_2\}$ ,  $E^- = \{\mathcal{D}_3\}$ , where

$$\mathcal{D}_1 = \{B_1(0)\}, \quad \mathcal{D}_2 = \{B_2(0)\}, \quad \mathcal{D}_3 = \{q_1(0), q_2(0), \dots, q_m(0)\},$$

and let  $\mathcal{O}$  contain (the normal forms of) the following axioms, for all  $i = 1, \dots, k$ :

$$B_1 \rightarrow \neg\psi, \quad B_2 \rightarrow \neg\psi, \tag{6}$$

$$p_i \rightarrow \diamond(\bar{A}_i \wedge \bigwedge_{j \neq i} (A_j \wedge \bar{A}_j)), \quad \neg p_i \rightarrow \diamond(A_i \wedge \bigwedge_{j \neq i} (A_j \wedge \bar{A}_j)), \tag{7}$$

We show that  $\exists \mathbf{p} \forall \mathbf{q} \psi$  is valid iff  $E$  is  $\mathcal{Q}_p[\diamond]$ -separable under  $\mathcal{O}$ .

( $\Rightarrow$ ) Suppose  $\exists \mathbf{p} \forall \mathbf{q} \psi$  is valid. Take an assignment  $\mathbf{a}$  for the variables  $\mathbf{p}$  such that under all assignments  $\mathbf{b}$  for the variables  $\mathbf{q}$  formula  $\psi$  is true. Let  $C$  be the conjunction of all  $A_i$  with  $\mathbf{a}(p_i) = 1$  and all  $\bar{A}_i$  with  $\mathbf{a}(p_i) = 0$ , and let  $\varkappa = \diamond C$ . We show that  $\varkappa$  separates  $E$ . Define an interpretation  $\mathcal{I}$  by taking

- $\mathcal{I}, 0 \models p_i$  iff  $\mathbf{a}(p_i) = 1$ , for  $i = 1, \dots, k$  and  $\mathcal{I}, 0 \models q_j$ , for  $j = 1, \dots, m$ ;
- if  $\mathcal{I}, 0 \models p_i$ , then  $\mathcal{I}, i \models \bar{A}_i \wedge \bigwedge_{j \neq i} (A_j \wedge \bar{A}_j)$ ;
- if  $\mathcal{I}, 0 \not\models p_i$ , then  $\mathcal{I}, i \models A_i \wedge \bigwedge_{j \neq i} (A_j \wedge \bar{A}_j)$ .

By the definition,  $\mathcal{I}$  is a model of  $\mathcal{O}$  and  $\mathcal{D}_3$  with  $\mathcal{I}, 0 \not\models \varkappa$ . On the other hand, let  $\mathcal{I}$  be a model of  $\mathcal{O}$  and some  $\mathcal{D}_l$ ,  $l = 1, 2$ . By (6),  $\mathcal{I}, 0 \not\models \psi$ . Then the truth values of the  $p_i$  in  $\mathcal{I}$  at 0 cannot reflect the truth values of the  $p_i$  under  $\mathbf{a}$  (for otherwise  $\psi$  would be true at 0 in  $\mathcal{I}$ ). Take some  $i_0$  for which these truth values of  $p_{i_0}$  differ, say  $\mathbf{a}(p_{i_0}) = 1$  but  $\mathcal{I}, 0 \not\models p_{i_0}$ . Then  $\mathcal{I}, 0 \models \diamond(A_{i_0} \wedge \bigwedge_{j \neq i_0} (A_j \wedge \bar{A}_j))$ , and so  $\mathcal{I}, 0 \models \varkappa$ .

( $\Leftarrow$ ) Suppose a  $\mathcal{Q}_p[\diamond]$ -query  $\varkappa$  separates  $E$  but  $\exists \mathbf{p} \forall \mathbf{q} \psi$  is not valid. From our conditions on  $\psi$ , it is easy to see by considering possible models of  $\mathcal{O}$  and  $\mathcal{D}_l$ ,  $l = 1, 2, 3$ , that  $\varkappa$  does not contain occurrences of  $B_1, B_2, p_i, q_j$ ,  $1 \leq i \leq k, 1 \leq j \leq m$ . Let  $\mathcal{I}$  be a model of  $\mathcal{O}$  and  $\mathcal{D}_3$  such that  $\mathcal{I}, 0 \not\models \varkappa$ . Let  $\mathbf{a}$  be the assignment for  $\mathbf{p}$  given by  $\mathcal{I}$  at 0. As  $\exists \mathbf{p} \forall \mathbf{q} \psi$  is not valid, there is

an assignment  $\mathbf{b}$  for  $\mathbf{q}$  such that  $\psi$  is false under  $\mathbf{a}$  and  $\mathbf{b}$ . Consider an interpretation  $\mathcal{I}$  such that  $\mathcal{I}, 0 \models B_1$ , the truth values of  $\mathbf{p}$  and  $\mathbf{q}$  at 0 are given by  $\mathbf{a}$  and  $\mathbf{b}$ , and all other atoms are interpreted as in  $\mathcal{J}$ . Then  $\mathcal{I}$  is a model of  $\mathcal{O}$  and  $\mathcal{D}_1$ , and so  $\mathcal{I}, 0 \models \varkappa$ . But then  $\mathcal{J} \models \varkappa$ , as  $\varkappa$  can only contain atoms  $A_i$  and  $\bar{A}_i$ , which is a contradiction showing that  $\exists \mathbf{p} \forall \mathbf{q} \psi$  is valid.

We now prove the results for data complexity. The NP-lower bounds are inherited from the ontology-free case. We show the NP-upper bound for  $\text{QBE}(LTL^{\square\Diamond}, \mathcal{Q}_p[\Diamond])$  and the P-upper bound for  $\text{QBE}_{\mathbf{b}}^{\text{b}^+}(LTL^{\square\Diamond}, \mathcal{Q}_p[\Diamond])$ . The P-upper bound for  $\text{QBE}(LTL^{\square\Diamond}, \mathcal{Q}[\Diamond])$  with a bounded number of positive examples can be again proved in two steps: (1) by Theorem 4 (i.1) it suffices to prove the P-upper bound for  $\text{QBE}_{\mathbf{b}}^{\text{b}^+}(LTL^{\square\Diamond}, \mathcal{Q}[\Diamond])$ ; (2) by Theorem 4 (i.3),  $\text{QBE}_{\mathbf{b}}^{\text{b}^+}(LTL^{\square\Diamond}, \mathcal{Q}[\Diamond]) \leq_p \text{QBE}_{\mathbf{b}}^{\text{b}^+}(LTL^{\square\Diamond}, \mathcal{Q}_p[\Diamond])$ .

Assume an  $LTL^{\square\Diamond}$ -ontology  $\mathcal{O}$  is given. We show that one can construct in *polynomial* time for any data instance  $\mathcal{D}$  a set  $\mathcal{M}_{\mathcal{O}, \mathcal{D}}$  of models of  $\mathcal{D}$  whose types form a sequence

$$tp_0, \dots, tp_{k_0}, tp_{k_0+1}, \dots, tp_{k_0+l}, \dots, tp_{k_0+1}, \dots, tp_{k_0+l}, \dots \quad (8)$$

with  $\max \mathcal{D} \leq k_0 \leq \max \mathcal{D} + |\mathcal{O}|$  and  $l \leq |\mathcal{O}|$  such that for any  $\varkappa \in \mathcal{Q}_p[\Diamond]$ ,  $\mathcal{D}, \mathcal{O} \models \varkappa(0)$  iff  $\mathcal{I}, 0 \models \varkappa$  for all  $\mathcal{I} \in \mathcal{M}_{\mathcal{O}, \mathcal{D}}$ . Note that, in particular, every set  $\mathcal{M}_{\mathcal{O}, \mathcal{D}}$  is of polynomial size in  $\mathcal{D}$ . Interestingly, the models in  $\mathcal{M}_{\mathcal{O}, \mathcal{D}}$  are not necessarily models of  $\mathcal{O}$  (unless  $\mathcal{O}$  is a Horn ontology). Then, to show the NP-upper bound for  $\text{QBE}(LTL^{\square\Diamond}, \mathcal{Q}_p[\Diamond])$  and the P-upper bound for  $\text{QBE}_{\mathbf{b}}^{\text{b}^+}(LTL^{\square\Diamond}, \mathcal{Q}_p[\Diamond])$  one constructs for any  $\mathcal{D} \in E^+ \cup E^-$  the set  $\mathcal{M}_{\mathcal{O}, \mathcal{D}}$  and then decides, using that polysize separating queries exist if separating queries exist at all, whether there exists  $\varkappa \in \mathcal{Q}_p[\Diamond]$  such that

- for all  $\mathcal{D} \in E^+$ :  $\mathcal{I}, 0 \models \varkappa$ , for all  $\mathcal{I} \in \mathcal{M}_{\mathcal{O}, \mathcal{D}}$ ;
- for all  $\mathcal{D} \in E^-$ :  $\mathcal{I}, 0 \not\models \varkappa$ , for some  $\mathcal{I} \in \mathcal{M}_{\mathcal{O}, \mathcal{D}}$

in either NP (by guessing the polysize query and then verifying it in polynomial time) or P (by applying essentially the same dynamic programming algorithm as for  $\text{QBE}_{\mathbf{b}}^{\text{b}^+}(\mathcal{Q}_p[\Diamond])$ ).

We come to the construction of  $\mathcal{M}_{\mathcal{O}, \mathcal{D}}$ . Let  $\mathcal{D}$  be a data instance. A type  $tp$  is *consistent with  $\mathcal{D}$  at  $k$*  if  $A(k) \in \mathcal{D}$  implies  $\neg A \notin tp$ , for any atom  $A$ . We next define the notion of a decoration. Let  $I_0, \dots, I_n$  be a partition of  $\mathbb{N}$  into nonempty intervals  $I_0, \dots, I_n$  with  $I_n$  of the form  $[m, \infty]$  for some  $m$  with  $\max \mathcal{D} < m \leq \max \mathcal{D} + |\mathcal{O}| + 1$  and  $\max I_k + 1 = \min I_{k+1}$  for all  $k < n$ . Let  $f$  be a function that associates with each interval  $k \leq n$  a nonempty set  $f(k)$  of  $\mathcal{O}$ -satisfiable types. Intuitively, the types in  $f(k)$  are types that we aim to satisfy in the interval  $I_k$ . We then call  $D = (I_0, \dots, I_n, f)$  a *pre-decoration* of  $\mathcal{D}$ . We say that a model  $\mathcal{I}$  is *consistent with  $D = (I_0, \dots, I_n, f)$*  if it is defined by a sequence

$$tp_0, tp_1, \dots$$

of types  $tp_i$  such that

1. if  $i \in I_k$ , then  $tp_i \in f(k)$  and  $tp_i$  is consistent with  $\mathcal{D}$  at  $i$ , for all  $i \geq 0$ ;
2. each  $tp \in f(n)$  occurs infinitely often as  $tp_i$  in  $\mathcal{I}$  for  $i \geq m$ .

Then  $D = (I_0, \dots, I_n, f)$  is a *decoration of  $\mathcal{D}$  for  $\mathcal{O}$*  if every model  $tp_0, tp_1, \dots$  that is consistent with  $D$  satisfies  $tp_i$  at timepoint  $i$  (and this ia, in particular, a model of  $\mathcal{O}$ ). Note that models that are consistent with  $D$  are trivially models of  $\mathcal{D}$ . Thus, any  $D$  defines the set  $\mathcal{M}_D$  of models that are consistent with  $D$  and these are also always models of  $\mathcal{O}$  if  $D$  is a decoration of  $\mathcal{D}$  for  $\mathcal{O}$ .  $D = (I_0, \dots, I_n, f)$  also defines a *canonical model  $\mathcal{I}_D$*  as follows: fix any ordering  $tp_0, \dots, tp_{j-1}$  of  $f(n)$  and assume  $I_n = [m_D, \infty]$ . Then let  $\mathcal{I}_D$  be defined by setting

- for  $i \in I_k$  with  $k < n$ ,  $i \in A^{\mathcal{I}_D}$  if  $A \in tp$  for all  $tp \in f(k)$  that are consistent with  $\mathcal{D}$  at  $i$ ;
- for  $i = m_D + j_0 + kj$  with  $j_0 < j$ ,  $i \in A^{\mathcal{I}_D}$  if  $A \in tp_{j_0}$ .

Thus, for  $i < m_D$ ,  $\mathcal{I}_D$  is defined as the intersection of all models that are consistent with  $D$  and for  $i \geq m_D$  we repeat the pattern  $tp_0, \dots, tp_{j-1}$  again and again. Note that  $\mathcal{I}_D$  is of the form defined in (8). We show the following lemma connecting  $\mathcal{M}_D$  and  $\mathcal{I}_D$ .

**Lemma 31.** *For every  $\varkappa \in \mathcal{Q}[\Diamond]$  and every  $i < m_D$ , we have  $\mathcal{J}, i \models \varkappa$  for all  $\mathcal{J} \in \mathcal{M}_D$  iff  $\mathcal{I}_D, i \models \varkappa$ .*

*Proof.* Obtain  $\mathcal{M}$  from  $\mathcal{M}_D$  by replacing for each  $\mathcal{J} \in \mathcal{M}_D$  the final part of  $\mathcal{J}$  based on the interval  $I_n$  by the final part of  $\mathcal{I}_D$  based on  $I_n$ . Then clearly  $\mathcal{J}, 0 \models \varkappa$  for all  $\mathcal{J} \in \mathcal{M}_D$  iff  $\mathcal{J}, 0 \models \varkappa$  for all  $\mathcal{J} \in \mathcal{M}$ . It is therefore sufficient to prove the claim for  $\mathcal{M}$  instead of  $\mathcal{M}_D$ .

The proof is by induction on  $\ell$  for  $\varkappa$  of the form  $\rho_0 \wedge \Diamond(\rho_1 \wedge \Diamond(\rho_2 \wedge \dots \wedge \Diamond\rho_\ell))$ . For  $\ell = 0$  the claim follows from the definition.

Assume that the claim has been proved for  $\ell \geq 0$ ,  $\varkappa = \rho_0 \wedge \Diamond(\rho_1 \wedge \Diamond(\rho_2 \wedge \dots \wedge \Diamond\rho_{\ell+1}))$ , and  $\mathcal{J}, i \models \varkappa$  for some  $i < m_D$  and all  $\mathcal{J} \in \mathcal{M}$ . We have to show that  $\mathcal{I}_D, i \models \varkappa$ .

If there exists  $i \geq m_D$  such that  $\mathcal{J}, i \models \rho_1 \wedge \Diamond(\rho_2 \wedge \dots \wedge \Diamond\rho_\ell)$ , then we are done.

Otherwise, we show that there exists  $i'$  with  $i < i' < m_D$  such that  $\mathcal{J}, i' \models \rho_1 \wedge \Diamond(\rho_2 \wedge \dots \wedge \Diamond\rho_\ell)$  for all  $\mathcal{J} \in \mathcal{M}$ . Then the claim follows by IH.



We first observe that there exists  $i'$  with  $i < i' < m_D$  such that for  $i' \in I_k$  we have  $\rho_1 \subseteq tp$  for all  $tp \in f(k)$  that are consistent with  $\mathcal{D}$  at  $i'$ .

For assume that this is not the case. Then construct a model  $\mathcal{J} \in \mathcal{M}$  by choosing for every  $j$  with  $i < j < m_D$  such that  $j \in I_k$  a  $tp \in f(k)$  that is consistent with  $\mathcal{D}$  at  $j$  such that  $\rho_1 \not\subseteq tp$ . Define  $\mathcal{J}$  using these  $tp_j$ . Then  $\mathcal{J}, i \not\models \varkappa$ , a contradiction.

Let  $i'$  be minimal  $i < i' < m_D$  such that for  $i' \in I_k$  we have  $\rho_1 \subseteq tp$  for all  $tp \in f(k)$  that are consistent with  $\mathcal{D}$  at  $i'$ .

We next show that  $\mathcal{J}, i' \models \diamond(\rho_2 \wedge \dots \wedge \diamond\rho_\ell)$  for all  $\mathcal{J} \in \mathcal{M}$ . Assume that this is not the case. Let  $\mathcal{J}$  be a witness. Then we construct a new model  $\mathcal{J}' \in \mathcal{M}$  by refuting  $\rho_1$  between  $i$  and  $i'$  (possible by minimality of  $i'$ ) and then adding  $\mathcal{J}$  from  $i'$ . Then  $\mathcal{J}', i \not\models \varkappa$ , a contradiction.

It follows that  $\mathcal{J}, i' \models \rho_1 \wedge \diamond(\rho_2 \wedge \dots \wedge \diamond\rho_\ell)$  for all  $\mathcal{J} \in \mathcal{M}$ , as required.  $\dashv$

We next define the decorations we work with. Given any model  $\mathcal{I}$  of  $\mathcal{D}$  and  $\mathcal{O}$  of the form (8), we obtain a decoration  $D_{\mathcal{I}} = (I_0, \dots, I_n, f)$  with  $n \leq 2|\mathcal{O}| + 2$  as follows. Call a node  $i$  maximal in  $\mathcal{I}$  for  $\mathcal{O}$  if there exists  $C$  with  $\Box C \in \text{sub}(\mathcal{O})$  such that  $i \models \Box C \wedge \neg C$ .

Assume  $I_0, \dots, I_\ell$  and  $f(0), \dots, f(\ell)$  have been defined already and  $I_\ell$  is not of the form  $[m, \infty]$  (if  $I_\ell$  is of the form  $[m, \infty]$  we are done). We next define  $I_{\ell+1}$  (and possibly  $I_{\ell+2}$ ).

1. If  $\max I_\ell < k_0$ , then we proceed as follows: if  $\max I_\ell + 1$  is either maximal for  $\mathcal{O}$  in  $\mathcal{I}$  or  $\max I_\ell + 1 = k_0$ , then set  $I_{\ell+1} = \{\max I_\ell + 1\}$  and  $f(\ell + 1) = \{tp_{\mathcal{I}}(\max I_\ell + 1)\}$ .  
Otherwise let

$$k := \min\{k > \max I_\ell \mid k \text{ is maximal for } \mathcal{O} \text{ in } \mathcal{I} \text{ or } k = k_0\}$$

and set  $I_{\ell+1} = [\max I_\ell, k - 1]$ ,  $I_{\ell+2} = \{k\}$ ,  $f(\ell + 1) = \{tp_{\mathcal{I}}(k) \mid k \in I_{\ell+1}\}$ , and  $f(\ell + 2) = \{tp_{\mathcal{I}}(k)\}$ .

2. Otherwise  $\max I_\ell \geq k_0$ . Then let  $I_{\ell+1} = [k_0, \infty]$  and  $f(\ell + 1) = \{tp_{k_0+1}, \dots, tp_{k_0+l}\}$ .

One can easily show that  $D_{\mathcal{I}} = (I_1, \dots, I_n, f)$  is indeed a decoration of  $\mathcal{D}$  for  $\mathcal{O}$  and  $n \leq 2|\mathcal{O}| + 2$ . Note also that  $\mathcal{I}$  itself is consistent with  $D_{\mathcal{I}}$ . The following lemma summarises our findings.

**Lemma 32.** *For any  $\mathcal{D}$  one can construct in polynomial time a set  $\mathcal{F}_{\mathcal{O}, \mathcal{D}}$  of decorations  $D = (I_0, \dots, I_n, f)$  of  $\mathcal{D}$  for  $\mathcal{O}$  such that  $n \leq 2|\mathcal{O}| + 2$  and the following are equivalent for any  $\varkappa \in \mathcal{Q}[\diamond]$ :*

1.  $\mathcal{O}, \mathcal{D} \models \varkappa(0)$ ;
2.  $\mathcal{I}, 0 \models \varkappa$  for every  $\mathcal{I} \in \mathcal{M}_D$  and  $D \in \mathcal{F}_{\mathcal{O}, \mathcal{D}}$ ;
3.  $\mathcal{I}_D, 0 \models \varkappa$  for every  $D \in \mathcal{F}_{\mathcal{O}, \mathcal{D}}$ .

*Proof.* Models of the form (8) satisfying  $\mathcal{O}$  and  $\mathcal{D}$  are complete in the sense that the following conditions are equivalent for all  $\varkappa \in \mathcal{Q}_p[\diamond]$ :

- $\mathcal{O}, \mathcal{D} \models \varkappa(0)$ ;
- $\mathcal{I}, 0 \models \varkappa$  for all models  $\mathcal{I}$  of  $\mathcal{O}$  and  $\mathcal{D}$  of the form (8).

Hence the class of models  $\mathcal{M}_{D_{\mathcal{I}}}$  with  $\mathcal{I}$  a model of  $\mathcal{O}$  and  $\mathcal{D}$  of the form (8) is also complete. Hence the equivalence of Points 1. to 2. holds if we define  $\mathcal{F}_{\mathcal{O}, \mathcal{D}}$  as the class of decorations  $D = (I_0, \dots, I_n, f)$  of  $\mathcal{D}$  for  $\mathcal{O}$  with  $n \leq 2|\mathcal{O}| + 2$ . The equivalence of Points 2. and 3. follows from Lemma 31. It remains to show that  $\mathcal{F}_{\mathcal{O}, \mathcal{D}}$  can be constructed in polynomial time. The set of pre-decorations  $(I_0, \dots, I_n, f)$  of  $\mathcal{D}$  with  $n \leq 2|\mathcal{O}| + 2$  can clearly be constructed in polynomial time in  $|\mathcal{D}|$ . It thus remains to check in polynomial time whether a pre-decoration is a decoration. But such a check is straightforward as a pre-decoration  $D = (I_0, \dots, I_n, f)$  is a decoration if, and only if, the following condition holds: for any subformula  $\Box C$  of  $\mathcal{O}$ , any  $i \leq n$ , any  $tp \in f(i)$ , and any  $k \in f(i)$  with  $tp$  consistent with  $\mathcal{D}$  at  $k$ :  $\Box C \in tp$  iff no  $tp' \in f(i)$  with  $C \notin tp'$  is consistent with  $\mathcal{D}$  at any  $k' \in I_i \cap [k + 1, \infty]$  and no  $tp' \in f(j)$  with  $C \notin tp'$  and  $j > i$  is consistent with  $\mathcal{D}$  at any  $k' \in I_j$ .  $\dashv$

The set  $\mathcal{M}_{\mathcal{O}, \mathcal{D}}$  of models required for the construction of the algorithms is now defined by setting  $\mathcal{M}_{\mathcal{O}, \mathcal{D}} = \bigcup_{D \in \mathcal{F}_{\mathcal{O}, \mathcal{D}}} \mathcal{M}_D$ .  $\dashv$

## F Proofs for Section 7

**Theorem 14.** (i)  $\text{QBE}(LTL, \mathcal{Q})$  is in  $2\text{EXPTIME}$ , for any  $\mathcal{Q} \in \{\mathcal{Q}[\diamond], \mathcal{Q}[\Box, \diamond], \mathcal{Q}[U_s]\}$ . (ii)  $\text{QBE}(LTL, \mathcal{Q})$  is in  $2\text{EXPSpace}$ , for any  $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}_p[\Box, \diamond], \mathcal{Q}_p[U]\}$ .

*Proof.* Let  $E$  be an example set and  $\mathcal{O}$  an LTL ontology. To show the results for  $\text{QBE}(LTL, \mathcal{Q}[U_s])$  and  $\text{QBE}(LTL, \mathcal{Q}_p[U])$ , it is enough, by Lemma 20, to show that the construction of  $S$  in Lemma 23, representing  $\mathcal{D}$  from  $E$  and  $\mathcal{O}$ , can be done in  $2\text{EXPTIME}$ . Consider an unlabelled transition system  $S$  (which can be defined as a transition system of the kind we have with the unary alphabet  $\Sigma_1 = \Sigma_2 = \{\emptyset\}$ ) with the states  $tp$ , where  $tp$  is a type realisable in  $\mathcal{O}, \mathcal{D}$ . Given  $S$  and a set of realisable types  $T$ , we set  $S_T$  to be  $S$  restricted to  $tp$  that are reachable from some  $tp' \in T$  and the initial states  $T$ . For given  $T_1, \Gamma$  and  $T_2$ , we have  $T_1 \rightarrow_\Gamma T_2$  iff, for each path  $s$  in  $\mathcal{T}_{S_{T_1}}$ , there is a position  $p_s > 0$  satisfying the following conditions: (i) the set of

types at all  $p_s$  coincides with  $T_2$ ; (ii)  $A \in tp$  for every  $tp$  at a position  $p \in (0, p_s)$  for every  $s$  iff  $A \in \Gamma$ , for each  $A \in \Sigma^\perp$ ; (iii) the set of all  $tp$  from (ii) does not intersect with  $T_2$ . Next, we observe that if the positions  $p_s$  satisfying (i), (ii), (iii) exist, then there exist such positions  $p_s \leq |\mathcal{S}|$ . (Intuitively, this is because whenever  $p_s > |\mathcal{S}|$ , there exists a type in  $s$  that repeats itself.) Thus, we need to check conditions (i)–(iii) in a tree of depth  $|\mathcal{S}|$ . This can be done in a branch-by-branch fashion using a (non-deterministic) algorithm working in PSPACE in  $|\mathcal{S}|$ . It remains to observe that  $\mathcal{S}$  itself can be constructed in EXPTIME in  $|\mathcal{O}|$  and that to construct  $S$  we need to check  $T_1 \rightarrow_\Gamma T_2$  for  $O(2^{2^{|\mathcal{O}|}})$ -many pairs  $(T_1, T_2)$ .

To obtain the results of the theorem for  $\mathcal{Q}[\circ, \diamond]$ , we construct  $S$  (cf. Lemma 23) as above but using the transition relation  $T_1 \rightarrow'_\Gamma T_2$  for  $\Gamma \in \{\emptyset, \Sigma^\perp\}$  only (i.e.,  $\Sigma_2 = \{\emptyset, \Sigma^\perp\}$  in the definition of a transition system). We set  $T_1 \rightarrow'_{\Sigma^\perp} T_2$  if  $T_1 \rightarrow_{\Sigma^\perp} T_2$  and  $T_1 \rightarrow'_\emptyset T_2$  if  $T_1 \rightarrow_\Gamma T_2$  for  $\Gamma \neq \Sigma^\perp$ . It is easy to verify that  $S$  represents  $\mathcal{O}, \mathcal{D}$  for the class of  $\mathcal{Q}[\circ, \diamond]$  queries. The case  $\mathcal{Q}[\diamond]$  is left to the reader.  $\dashv$