

Maybe Eventually? Towards Combining Temporal and Probabilistic Description Logics and Queries*

Patrick Koopmann^[0000–0001–5999–2583]

Institute for Theoretical Computer Science, Technische Universität Dresden, Germany
`firstname.lastname@tu-dresden.de`

Abstract. We present some initial results on ontology-based query answering with description logic ontologies that may employ temporal and probabilistic operators on concepts and axioms. Specifically, we consider description logics extended with operators from linear temporal logic (LTL), as well as subjective probability operators, and an extended query language in which conjunctive queries can be combined using these operators. We first show some complexity results for the setting in which either only temporal operators or only probabilistic operators may be used, both in the ontology and in the query, and then show a 2EXP-SPACE lower bound for the setting in which both types of operators can be used together.

Keywords: Description Logics · Temporal Reasoning · Probabilistic Reasoning · Ontology-Based Query Answering.

1 Introduction

Ontology-Based Query Answering (OBQA) received considerable attention in the past, as it allows to query incomplete data in the presence of an ontology providing background knowledge about the data domain. While classically, OBQA considers a setting where the data is both static and certain, there are many applications where this assumption does not hold, which lead to the development of temporal query languages for OBQA [10,33,11,5], and research on OBQA for probabilistic data [21,8,9,7]. Temporal OBQA has been proposed as a technique for querying historical data and to detect situations in streams of data. To describe temporal patterns in a query, *temporal queries* as in [10,11,5] extend conjunctive queries (CQs) with operators from linear temporal logic (LTL). Probabilistic OBQA is motivated by data sets obtained using uncertain methods such as language and image recognition, or uncertain sensor measurements. In this setting, query answers hold true with a certain probability, which may be part of the query result. As historical data can be obtained using language recognition, and situation recognition is often applied in applications that involve temporal data based on uncertain sensor measurements, there exist applications in which

* Supported by the DFG in the CRC 912 (HAEC) and in the TRR 248 (CPEC).

we want to query data that is both temporal and probabilistic in nature. Motivated by this, recently, temporal probabilistic OBQA has been investigated [23], where the temporal query language from [11] is extended with probability operators, and data are considered sequences of probabilistic ABoxes as in [21]. As an example for a probabilistic temporal query, consider a health supervision app on a smartphone which operates on a sequence of data obtained using motion and blood pressure sensors. The following query then detects situations in which the patient was, during the last 10 time units, with a low probability exercising, until with a high probability he had a high blood pressure, in which case the app might issue a warning:

$$q(x) \leftarrow \bigcirc^{-10} (P_{<0.2} \text{Exercising}(x) \cup P_{>0.7} \text{HighBloodPressure}(x)).$$

While the mentioned works allow for an extended expressivity in the query language, they only consider ontologies that are formulated using a classical (atemporal and non-probabilistic) DL. Since the role of the ontology in OBQA is to provide additional background knowledge, temporal and/or probabilistic OBQA would benefit from ontology languages that provide both temporal and probabilistic language constructs. To stay with the current example, this could for instance be used to express that if a patient starts exercising, his blood pressure is likely to remain increased until the patient takes a break:

$$\text{StartsExercising} \sqsubseteq (P_{>0.7} \text{IncreasedBloodPressure}) \cup \text{StopsExercising},$$

where `StartsExercising` and `StopsExercising` are defined in further axioms using temporal concept operators.

Temporal DLs have been well investigated in the literature, and may extend classical DLs with LTL-operators on axioms and concepts [29,6], with MTL-operators [3,36,20], Halpern and Shoham’s interval logic [1,34], or temporal attributes [32]. Similarly, several probabilistic extensions to DLs have been suggested, such as the non-monotonic DL $P\text{-SHIF}(\mathcal{D})/P\text{-SHOIN}(\mathcal{D})$ [26], the DLs $\text{Prob-}\mathcal{ALC}/\text{Prob-}\mathcal{EL}$ for expressing subjective probabilities [19], DLs using log-linear probabilities [31] and the Bayesian DLs \mathcal{BEL} and \mathcal{BALC} [14,12]. There is also research on ontology languages that combine temporal and probabilistic aspects: these consider temporal probabilistic Datalog programs [15], dynamic Bayesian DL networks [13], and temporal extensions of DL-Lite [25], but do not consider expressive query languages, or the full expressivity of temporal DLs such as $\text{LTL-}\mathcal{ALC}$ and $\text{Prob-}\mathcal{ALC}$. There is some research on answering unions of conjunctive queries in temporal DL-Lite [2], and instance retrieval in temporal extensions of \mathcal{EL} [18], but not on answering temporal queries, and to the best of our knowledge, there is no research on OBQA with ontology languages that employ probabilistic concept operators.

The aim of this paper is to theoretically investigate a setting where temporal operators, as well as operators expressing subjective probability, can be used both as part of the ontology language and as part of the query language. While some complexity bounds are still open at this point, we present initial results towards understanding the complexity in such a setting. Specifically, our contributions are the following.

1. In Section 2, we combine the languages studied in [11,29,19] to define the syntax and semantics of *temporal probabilistic DL formulae* (TPDFs), which generalise temporal probabilistic knowledge bases and queries.
2. In Section 3, we give tight complexity bounds for TPDFs with only temporal operators.
3. In Section 4, we give upper bounds for TPDFs with only probability operators.
4. In Section 5, we show that for TPDFs that use both temporal and probability operators, satisfiability is 2EXPSpace-hard.

Details of proofs and definitions can be found in the extended version of the paper [22].

2 Temporal Probabilistic Description Logic Formulae

2.1 Preliminaries

We assume basic knowledge about expressive DLs. Our results concern DLs ranging from \mathcal{ALC} to \mathcal{ALCOQ} and \mathcal{ALCOI} . Details about the DLs relevant for this paper, as well as on query answering, can be found in the extended version of the paper. We assume DL concepts to be composed using the operators of the respective DL based on the pair-wise disjoint, countably infinite sets \mathbf{N}_C , \mathbf{N}_R and \mathbf{N}_I of respectively *concept names*, *role names* and *individual names*. We assume DL axioms to be either *general class inclusions* (GCIs) of the form $C \sqsubseteq D$, or *assertions* of the forms $C(a)$, $r(a, b)$ with C and D being concepts in the respective DL, r , s role names, and a , b individual names. We use $C \equiv D$ as abbreviation for the two GCIs $C \sqsubseteq D$ and $D \sqsubseteq C$. Satisfiability of sets \mathcal{K} of axioms is defined in terms of *interpretations* $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, where $\Delta^{\mathcal{I}}$ is a set of domain elements and $\cdot^{\mathcal{I}}$ is a function mapping individual names to domain elements, concepts to subsets of $\Delta^{\mathcal{I}}$ and roles to subsets of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. Conjunctive queries (CQs) and entailment of Boolean CQs are defined as usual (eg., see [28]): specifically, CQs can contain free variables called *answer variables*, and a Boolean CQ is a CQ without free variables. A *query answer* to a CQ ϕ in a DL KB \mathcal{K} is an assignment of individual names to the free variables in ϕ such that the resulting Boolean CQ is entailed by \mathcal{K} .

To distinguish between different intervals relevant in this paper, we use the notation $[i, j]$ to denote closed intervals over the reals, and the notation $\llbracket i, j \rrbracket$ to denote closed intervals over the integers. A *probability measure* over a (possibly infinite) set W is a function $P : \mathcal{W} \rightarrow [0, 1]$, where $\mathcal{W} \subseteq 2^W$ is a σ -algebra (it contains W and is closed under complement and countable union), s.t. $P(\emptyset) = 0$, $P(W) = 1$, and for any countable set $\mathcal{W}' \subseteq \mathcal{W}$ of pairwise disjoint sets $W' \subseteq W$, we have $P(\bigcup_{W' \in \mathcal{W}'} W') = \sum_{W' \in \mathcal{W}'} P(W')$.

2.2 Syntax

We consider extensions of classical DLs which additionally allow temporal concepts of the form $\bigcirc C$ (next) and CUD (until), and probabilistic concepts of

the form $P_{\otimes p}C$, where $\otimes \in \{<, =, >\}$, $p \in [0, 1]$ and C, D are concepts. These concepts may be used at any place within a concept, and we call the resulting concepts *temporal probabilistic concepts*. Here, we do not fix a particular DL as basis, but may refer to the *underlying DL* which is extended by these operators. While classically, a DL knowledge base is build using DL axioms, against which queries are evaluated, it will be convenient to study queries and DL axioms not in separation, but to allow for an integrated language in which DL axioms and CQs can be arbitrarily mixed within a formula. This further expressivity can for instance be used to specify that a certain DL axiom holds until a Boolean CQ becomes satisfied. For this reason, we collectively call DL axioms and CQs *generalised axioms*. *Temporal probabilistic DL formulae* (TPDFs) α are then built according to the following syntax rule, where X is a generalised axiom that may use temporal probabilistic concepts, $\otimes \in \{<, =, >\}$ and $p \in [0, 1]$:

$$\alpha ::= X \mid \neg\alpha \mid \alpha \wedge \alpha \mid \bigcirc\alpha \mid \alpha \mathbf{U} \alpha \mid P_{\otimes p}\alpha.$$

The operators \neg , \wedge , \bigcirc and \mathbf{U} are called *temporal operators*, while the operators $P_{\otimes p}$ are called *probability operators*. We define further operators as the usual abbreviations, that is, for TPDFs ϕ and ψ , we denote $\mathbf{true} := \psi \vee \neg\phi$ (for some ϕ), $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$, $\diamond\phi := \mathbf{true}\mathbf{U}\phi$ and $\square\phi := \neg\diamond\neg\phi$, $\phi \rightarrow \psi := \neg\phi \vee \psi$ and $\phi \leftrightarrow \psi := (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$, and similar for concepts. A TPDF is called *Boolean* if every CQ in it is Boolean.

As typical for temporal reasoning with DLs, we assume a set $N_{rig} \subseteq N_C \cup N_R$ of *rigid names*, composed of a set $N_{Crig} = N_{rig} \cap N_C$ of *rigid concepts* and a set $N_{Rrig} = N_{rig} \cap N_R$ of *rigid roles*, which denote concept and role names whose interpretation does not change over time.

2.3 Semantics

To define the semantics of TPDFs, we have to take into consideration two dimensions: the temporal dimension and the probabilistic dimension. A *temporal interpretation* is a sequence $(\mathcal{I}_i)_{i \geq 0}$ of interpretations $\mathcal{I}_i = \langle \Delta, \cdot^{\mathcal{I}_i} \rangle$ sharing the same domain $\Delta^{\mathcal{J}}$, such that for any rigid name $X \in N_{rig}$ and $i, j \geq 0$, $X^{\mathcal{I}_i} = X^{\mathcal{I}_j}$. A *probabilistic temporal interpretation* is then a probability measure $\iota : \mathfrak{J} \rightarrow [0, 1]$, over a set \mathcal{J} of temporal interpretations $(\mathcal{I}_i)_{i \geq 0}$ sharing the same set Δ^t of domain elements ($\mathfrak{J} \subseteq 2^{\mathcal{J}}$ is then a sigma algebra). We call \mathcal{J} the *possible worlds* of ι .

To define the semantics of temporal and probabilistic operators, we define the function $\cdot^{\mathcal{I}_i, \iota}$ on concepts, where $(\mathcal{I}_j)_{j \geq 0} \in \mathcal{J}$ and $i \geq 0$. $\cdot^{\mathcal{I}_i, \iota}$ is defined as $\cdot^{\mathcal{I}_i}$ for the concept operators of the underlying DL, and for the remaining operators by

$$\begin{aligned} (\bigcirc C)^{\mathcal{I}_i, \iota} &= C^{\mathcal{I}_{i+1}, \iota} \\ (\mathbf{CUD})^{\mathcal{I}_i, \iota} &= \{d \in \Delta^t \mid \exists j \geq i : d \in D^{\mathcal{I}_j, \iota}, \forall k \in \llbracket i, j-1 \rrbracket : d \in C^{\mathcal{I}_k, \iota}\} \\ (P_{\otimes p}C)^{\mathcal{I}_i, \iota} &= \{d \in \Delta^t \mid \iota(\{(\mathcal{I}'_j)_{j \geq 0} \in \mathcal{J} \mid d \in C^{\mathcal{I}'_i, \iota}\}) \otimes p\}. \end{aligned}$$

Satisfaction of Boolean TPDFs is defined by:

1. $\mathcal{I}_i, \iota \models \alpha$ iff $\mathcal{I}_i \models \alpha$, where α is a Boolean CQ or a role assertion,
2. $\mathcal{I}_i, \iota \models C \sqsubseteq D$ iff $C^{\mathcal{I}_i, \iota} \subseteq D^{\mathcal{I}_i, \iota}$,
3. $\mathcal{I}_i, \iota \models C(a)$ iff $a^{\mathcal{I}_i} \in C^{\mathcal{I}_i, \iota}$,
4. $\mathcal{I}_i, \iota \models \bigcirc \phi$ iff $\mathcal{I}_{i+1}, \iota \models \phi$,
5. $\mathcal{I}_i, \iota \models \phi \mathbf{U} \psi$ iff there exists $j \geq i$ s.t. $\mathcal{I}_j, \iota \models \psi$ and for all $k \in \llbracket j, i-1 \rrbracket$, $\mathcal{I}_k, \iota \models \phi$, and
6. $\mathcal{I}_i, \iota \models \mathbf{P}_{\otimes p} \phi$ iff $\iota(\{(\mathcal{I}'_j)_{j \geq 0} \in \mathcal{J} \mid \mathcal{I}'_i, \iota \models \phi\}) \otimes p$.¹

We say that ι satisfies a Boolean TPDF ϕ , in symbols $\iota \models \phi$, if for all $(\mathcal{I}_i)_{i \geq 0} \in \mathcal{J}$, $\mathcal{I}_0, \iota \models \phi$, in which case ι is a *model of ϕ* . A Boolean TPDF is satisfiable iff it has a model.

The paper focusses on showing complexity bounds for Boolean TPDF satisfiability. Note that other reasoning tasks that are more related to classical OBQA can be easily reduced to TPDF satisfiability. For instance, for the problem of *temporal probabilistic query answering*, we are given a Boolean TPDF ϕ , and a non-Boolean TPDF ψ that contains only CQs and no DL axioms (a temporal probabilistic query), and we want to find an assignment of individual names to the answer variables in ψ so that the resulting TPDF is logically entailed by ϕ . This problem can be reduced to deciding the unsatisfiability of Boolean TPDFs of the form $\phi \wedge \neg \psi'$, where ψ' is obtained from ψ by replacing answer variables by individual names. As from now on, we focus on Boolean TPDFs only, we will omit the ‘‘Boolean’’ and just call them TPDFs in the following.

Remark. There is a subtle difference between our semantics and that of Prob- \mathcal{ALC} /Prob- \mathcal{EL} as introduced in [19], in that we do not require the set of possible worlds to be countable. We believe that, especially if we add a temporal dimension, considering only countable sets of possible worlds is too restrictive. For instance, if we allow a domain element to arbitrarily switch between satisfying a concept A and not satisfying it there are uncountably many possible sequences for this, each corresponding to a real number in between 0 and 1. There is no real reason why some of these sequences should be excluded. As we show in the extended version of the paper, there are TPDFs even without temporal operators that are only satisfiable in interpretations with an uncountable set of possible worlds, which means that our results do not directly transfer to the setting considered in [19].

3 Only Temporal Operators

We first consider the purely temporal case of TPDFs without probability operators. This problem has so far only been studied for *temporal queries* and *temporal DLs*, but not for the combination of both. Our first result concerns TPDFs without temporal concepts, that is, temporal operators can be used on CQs and on axioms, but not within concepts. Here, complexity upper bounds

¹ Note that we implicitly require that \mathfrak{J} contains all subsets of $2^{\mathcal{J}}$ relevant to these definitions.

directly follow from the complexity of temporal query entailment with classical ontologies, as studied in [11,5,4].

Theorem 1. *Satisfiability of TPDFs without probability operators and temporal concepts, and with underlying DL \mathcal{L} , is*

- PSPACE-complete for $\mathcal{L} = \mathcal{EL}$ and $\mathbf{N}_{Crig} = \mathbf{N}_{Rrig} = \emptyset$,
- EXPTIME-complete for $\mathcal{L} \in \{\mathcal{ALC}, \mathcal{ALCQ}\}$ and $\mathbf{N}_{Crig} = \mathbf{N}_{Rrig} = \emptyset$,
- NEXPTIME-complete for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ALC}, \mathcal{ALCQ}\}$ and $\mathbf{N}_{Rrig} = \emptyset$,
- NEXPTIME-complete $\mathcal{L} = \mathcal{EL}$ and $\mathbf{N}_{Rrig} \neq \emptyset$,
- 2EXPTIME-complete for $\mathcal{L} \in \{\mathcal{ALCI}, \mathcal{ALCIQ}, \mathcal{ALCOQ}, \mathcal{ALCOI}\}$, and
- decidable for \mathcal{ALCOIQ} .

If we also allow for temporal concept operators, we have to do a bit more. We first note that with rigid roles, using temporal operators on the level of concepts leads to undecidability already if the underlying DL is \mathcal{EL} [29]. We thus only have to consider the case where $\mathbf{N}_{Rrig} = \emptyset$. To show upper bounds for this case, we extend the method from [37] for temporal DLs based on quasimodels to also incorporate CQs. Namely, we abstract temporal interpretations using sequences of *quasistates*, which each contain a set of CQs and GCIs that hold or do not hold at the corresponding time point, together with a set of *concept types*, which represent the current states of domain elements.

Given a TPDF ϕ , let $\text{con}(\phi)$ denote the set of (sub-)concepts occurring in ϕ , $\text{form}(\phi)$ denote the set of sub-formulae of ϕ , and $\text{ind}(\phi)$ denote the set of individual names occurring in ϕ . Furthermore, define $\mathbf{t}_c(\phi) = \{C, \neg C \mid C \in \text{con}(\phi)\} \cup \{\{a\} \mid a \in \text{ind}(\phi)\}$ and $\mathbf{t}_f(\phi) = \{\psi, \neg\psi \mid \psi \in \text{form}(\phi)\}$. A *concept type* is then a subset $t \subseteq \mathbf{t}_c(\phi)$ s.t.

- C1** for every $\neg C \in \mathbf{t}_c(\phi)$, $\neg C \in t$ iff $C \notin t$, and
- C2** for every $C \sqcap D \in \mathbf{t}_c(\phi)$, $C \sqcap D \in t$ iff $C, D \in t$.

If a concept type t contains a concept of the form $\{a\}$, we call t a *nominal type*. A *formula type* is a subset $t \subseteq \mathbf{t}_f(\phi)$ s.t.

- F1** for every $\neg\psi \in \mathbf{t}_c(\phi)$, $\neg\psi \in t$ iff $\psi \notin t$, and
- F2** for every $\psi_1 \wedge \psi_2 \in \mathbf{t}_c(\phi)$, $\psi_1 \wedge \psi_2 \in t$ iff $\psi_1, \psi_2 \in t$.

A *quasistate* is a set S of formula and concept types s.t. S contains exactly one formula type t_S .

If the formula type only contains GCIs and their negation, there are easy syntactic conditions for when a quasistate can correspond to an element of a temporal interpretation. This becomes however more difficult when it can also contain CQs, which is why we instead formulate a semantic admissibility condition for quasistates. We first introduce the notion of a conceptual abstraction. Since quasistates will also be used in Section 4, we define them here more general for quasistates that may also contain probability operators. Given a concept or TPDF X , its *conceptual abstraction* X^{ca} is obtained by replacing every outermost concept C of the forms $\bigcirc D$, D_1UD_2 , and $P_{\otimes p}D$ by the fresh concept

name A_C , and every outermost TPDF ψ of the forms $\bigcirc\psi_1$, $\psi_1\bigcup\psi_2$ and $\mathsf{P}_{\otimes\mathsf{p}}\psi_1$ by $A_\psi(a)$, where A_ψ is fresh. A quasistate S is then *admissible* iff there exists a (classical) interpretation \mathcal{I} s.t.

- S1** for every TPDF $\alpha \in \text{form}(\phi)$, $\mathcal{I} \models \alpha^{\text{ca}}$ iff $\alpha \in t_S$, and
- S2** for every concept type $t \subseteq \mathbf{t}_c(\phi)$, $\bigcap_{C \in t} (C^{\text{ca}})^{\mathcal{I}} \neq \emptyset$ iff $t \in S$.

While a quasistate can contain up to exponentially many concept types, we can show that for \mathcal{ALCCOQ} and \mathcal{ALCCOI} , it can still be decided in 2EXPTIME wrt. to the input formula whether a given quasistate is admissible, while this can be done in EXPTIME for \mathcal{ALCCQ} .

It remains to represent the temporal dimension, which we do in terms of *runs* and *temporal quasimodels*.

A *concept/formula run* is a sequence $\sigma : \mathbb{N} \rightarrow \mathbf{t}_c(\phi)/\mathbf{t}_f(\phi)$ of concept/formula types s.t. for all $i \geq 0$,

- R1** for every $\bigcirc\alpha \in \mathbf{t}_c(\phi)/\mathbf{t}_f(\phi)$, $\bigcirc\alpha \in \sigma(i)$ iff $\alpha \in \sigma(i+1)$,
- R2** for every $\alpha\bigcup\beta \in \mathbf{t}_c(\phi)/\mathbf{t}_f(\phi)$, $\alpha\bigcup\beta \in \sigma(i)$ iff there exists $j \geq i$ s.t. $\beta \in \sigma(i)$ and for all $k \in \llbracket i, j-1 \rrbracket$, $\alpha \in \sigma(k)$,
- R3** for every $j \in \mathbb{N}$, $\sigma(i) \cap \mathbf{N}_{\text{Crig}} = \sigma(j) \cap \mathbf{N}_{\text{Crig}}$, and
- R4** for every $j \in \mathbb{N}$ and $a \in \mathbf{N}_I$, $\{a\} \in \sigma(i)$ iff $\{a\} \in \sigma(j)$.

A *temporal quasimodel* for ϕ is a tuple $\langle Q, \mathfrak{R} \rangle$, where Q is a sequence mapping each natural number to an admissible quasistate $Q(i)$, and \mathfrak{R} is a set of runs s.t.

- Q1** $\phi \in t_{Q(0)}$,
- Q2** for each $i \geq 0$ and $t \in Q(i)$, there exists a run $\sigma \in \mathfrak{R}$ s.t. $\sigma(i) = t$, and
- Q3** for each run $\sigma \in \mathfrak{R}$, and $i \geq 0$, $\sigma(i) \in Q(i)$.

Temporal quasimodels witness the satisfiability of TPDFs without probability operators. Furthermore, we can use a regularity argument as in [35] to limit the shape of these quasimodels. This is summarized in the following lemma.

Lemma 1. *If the underlying DL is \mathcal{ALCCOQ} or \mathcal{ALCCOI} , then ϕ is satisfiable iff there exists a quasimodel $\langle Q, \mathfrak{R} \rangle$ for ϕ where Q is of the form*

$$Q(0) \dots Q(n)(Q(n+1) \dots Q(n+m))^\omega,$$

with n and m double exponentially bounded in the size of ϕ .

The proof of the lemma makes use of the fact that, in a classical DL interpretation, if the underlying DL is \mathcal{ALCCOQ} or \mathcal{ALCCOI} , we can arbitrarily extend the set of domain elements that belong to a given concept type without affecting entailment of CQs or the extension of other types. This is not so easily possible for DLs that support both inverse roles and counting quantifiers, which is why we do not have results for \mathcal{ALCCIQ} . Using lower bounds for CQ entailment in \mathcal{ALCCI} [27] and \mathcal{ALCCO} [30], and for TPDFs with temporal operators only on concepts and GCIs [37], we obtain the following completeness results.

Theorem 2. *Satisfiability of TPDFs without probability operators is undecidable if $\mathbf{N}_{\text{Crig}} \neq \emptyset$. Otherwise, it is 2EXPTIME -complete if the underlying DL is \mathcal{ALCCO} , \mathcal{ALCCI} or \mathcal{ALCCOI} , and EXPSpace -complete if the underlying DL is \mathcal{ALC} or \mathcal{ALCCQ} .*

4 Only Probability Operators

We next consider the purely probabilistic case, that is, we allow probability operators on the level of concepts, axioms and queries, but no temporal operators. While [19] consider extensions of \mathcal{ALC} and \mathcal{EL} with probability operators on concepts and assertions, they do not consider these operators on GCIs. We extend this setting by allowing probability operators also on GCIs, and additionally allowing CQs.

Our method for deciding entailment of those TPDFs is again based on quasistates and types, over which we this time define probability measures.

A *probabilistic quasistate* is a probability measure $PS : 2^{\mathfrak{S}} \rightarrow [0, 1]$ over a set \mathfrak{S} of quasistates. It is *admissible* iff for every quasistate $S \in \mathfrak{S}$:

PS1 S is admissible, and

PS2 for every $P_{\otimes p} \psi \in \mathfrak{t}_f(\phi)$, $P_{\otimes p} \psi \in t_S$ iff $PS(\{S \in \mathfrak{S} \mid \psi \in t_S\}) \otimes p$.

While every quasistate contains a set of concept types, we might need a more fine-grained probability measure for each concept type to verify the probabilistic concepts in them. For this, we define *probabilistic concept types*. A probabilistic concept type $pt : 2^T \rightarrow [0, 1]$ is a probability measure over a set T of concept types s.t

PT for every $P_{\otimes p} C \in \mathfrak{t}_c(\phi)$ and $t \in T$, $P_{\otimes p} C \in t$ iff $pt(\{t \in T \mid C \in t\}) \otimes p$.

It is *compatible to a probabilistic quasistate* $PS : 2^{\mathfrak{S}} \rightarrow [0, 1]$ iff there exists a probability measure $P_{PS,pt} : 2^{W_{PS,pt}} \rightarrow [0, 1]$ over some set $W_{PS,pt} \subseteq \mathfrak{S} \times T$ s.t.

PC1 $\langle S, t \rangle \in W_{PS,pt}$ implies $t \in S$,

PC2 for every $S \in \mathfrak{S}$, $P_{PS,pt}(\{\langle S', t \rangle \in W_{PS,pt} \mid S' = S\}) = PS(\{S\})$, and

PC3 for every $t \in T$, $P_{PS,pt}(\{\langle S, t' \rangle \in W_{PS,pt} \mid t' = t\}) = pt(\{t\})$.

We call $P_{PS,pt}$ a *joined probability measure for PS and pt* . A probabilistic quasimodel for ϕ is now a tuple $\langle PS, \mathfrak{PT} \rangle$ of a probabilistic quasistate $PS : 2^{\mathfrak{S}} \rightarrow [0, 1]$ and a set \mathfrak{PT} of probabilistic concept types s.t.

PQ1 for every $S \in \mathfrak{S}$, $\phi \in t_S$,

PQ2 every probabilistic concept type $pt \in \mathfrak{PT}$ is compatible to PS , and

PQ3 for every quasistate $S \in \mathfrak{S}$ and concept type $t \in S$, there exists a joined probability measure for PS and some $pt \in \mathfrak{PT}$ s.t. $\langle S, t \rangle \in W_{PS,pt}$.

Note that in Condition **PQ3**, we only require $\langle S, t \rangle \in W_{PS,pt}$, but not $P_{PS,pt}(\{\langle S, t \rangle\}) > 0$. This is still sufficient to ensure that the type t can be instantiated in every possible world corresponding to S , and in fact necessary to ensure completeness, because we allow for uncountable sets of possible worlds in our semantics.

Lemma 2. *A TPDF ϕ without temporal operators, with underlying DL \mathcal{ALCCOQ} or \mathcal{ALCCOL} , is satisfiable iff there exists a probabilistic quasimodel $\langle PT, \mathfrak{PT} \rangle$ for ϕ .*

Probabilistic quasimodels are similar to temporal quasimodels, where instead of sequences, we use probability measures. For some DLs, this difference in structure can be exploited to gain better complexity bounds. While there can be in general double exponentially many quasistates and probabilistic concept types, if the underlying DL is \mathcal{ALCQ} , only exponentially many of each are needed. In contrast to the temporal quasimodels in Section 3, which indeed may always require a double exponential number of quasistates, probabilistic quasimodels benefit from a lack of *order*: this allows us to merge quasistates that agree on their formula type and nominal types, which is the reason why we can bound the size of probabilistic quasimodels for \mathcal{ALCQ} .

Our decision procedure for TPDF satisfiability consists of guessing and verifying a probabilistic quasimodel of the appropriate size. Here, we make use of a result from [17], which is also used in [19] to provide the complexity bounds of Prob- \mathcal{ALC} , to limit the required precision used in the probability measures.

Theorem 3. *Satisfiability of TPDFs without temporal operators is in NEXP-TIME if the underlying DL is \mathcal{ALCQ} , and in N2EXPTIME if the underlying DL is \mathcal{ALCOQ} or \mathcal{ALCOI} .*

Since satisfiability of Prob- \mathcal{ALC} is still in EXPTIME [19], the only known complexity lower bounds stem from the complexity of Boolean query entailment. We leave it as future work to investigate whether our complexity bounds can be optimised.

5 Temporal and Probability Operators

If we allow both temporal and probability operators, satisfiability of TPDFs becomes 2EXPSPACE-hard, even if we disallow rigid names. We show this by a reduction of the double-exponential corridor tiling problem. This problem is formalised as follows. We are given a set T of tiles containing an *initial tile type* $t_0 \in T$ and a *final tile type* $t_f \in T$, two sets $H \subseteq T \times T$ and $V \subseteq T \times T$ of respectively *horizontal* and *vertical tiling conditions*, and a natural number n . The problem is then to decide whether there exists a natural number m and a *tiling* $t : \llbracket 1, 2^{2^n} \rrbracket \times \llbracket 1, m \rrbracket \rightarrow T$ s.t. $t(1, 1) = t_0$, $t(1, m) = t_f$, for every $i \in \llbracket 1, 2^{2^n} - 1 \rrbracket$ and $j \in \llbracket 1, m \rrbracket$, we have $\langle t(i, j), t(i + 1, j) \rangle \in H$, and for every $i \in \llbracket 1, 2^{2^n} \rrbracket$ and $j \in \llbracket 1, m - 1 \rrbracket$, we have $\langle t(i, j), t(i, j + 1) \rangle \in V$. It follows from the relationship between corridor-tilings and space-bounded Turing machines shown in [16] that the double-exponential tiling problem is 2EXPSPACE-complete.

While the full reduction is shown in the extended version of the paper, we sketch the main ideas here. We use 2^{2^n} domain elements to represent the vertical dimension of the tiling, and the time line to represent the horizontal dimension. The probabilistic dimension is used to implement a double exponential counter on each domain element, which is used to identify which row of the tiling it represents. Here, we use temporal and probabilistic concepts to force the existence of exponentially many possible worlds per domain element, which at each time point store the different bit values of the double exponential counter using a

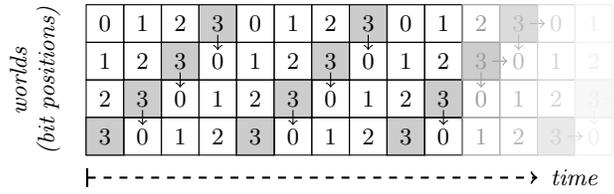


Fig. 1. Illustration of how counters are used to identify neighbouring possible worlds.

concept **Bit**. Specifically, the individual satisfies **Bit** in the i th possible world iff the i th bit of the double exponential counter has the value 1.

A main challenge in the construction is the lack of order in temporal probabilistic models. The set of possible worlds in an interpretation is unordered, which means we cannot directly refer to the “ i th” or “next” possible world. This is however necessary to implement a double exponential counter, since we have to transfer information about carrier bits from one possible world to another. Furthermore, since we do not allow rigid roles, we cannot keep the relationship between the different domain elements stable throughout the time line. As a result, we cannot directly refer to the domain element that refers to the next row in order to test the vertical tiling conditions. For both challenges, we use a similar trick.

For the double-exponential counter, we need to be able to identify possible worlds for the respective domain element that correspond to neighbouring bit positions. To do this, we implement a single-exponential counter in each possible world, which is incremented along the time line, so that in each world, the counter has a different value. This is visualised in Figure 1. To implement these counters, we use concept names A_1, \dots, A_n representing the bit value at the positions 1 to n of this counter. At each time point, two neighbouring possible worlds can be identified easily: the one with a counter value of $2^n - 1$ satisfies $\bigwedge_{i \in [1, n]} A_i$, and unless it corresponds to the last bit position, the next bit position corresponds to the world with a counter value of 0, which satisfies $\bigwedge_{i \in [1, n]} \neg A_i$. Using this mechanism, we can for instance transfer the information on whether the current bit has to be flipped using the following GCIs:

$$\begin{aligned} \bigwedge_{i \in [1, n]} A_i \sqcap \text{Flip} \sqcap \text{Bit} &\sqsubseteq P_{=1}((\bigwedge_{i \in [1, n]} \neg A_i) \rightarrow \text{Flip}) \\ (\bigwedge_{i \in [1, n]} A_i) \sqcap (\neg \text{Flip} \sqcup \neg \text{Bit}) &\sqsubseteq P_{=1}((\bigwedge_{i \in [1, n]} \neg A_i) \rightarrow (\text{FirstBit} \sqcup \neg \text{Flip})). \end{aligned}$$

Using further axioms, this allows us to implement a double exponential counter on each domain element, which is incremented every 2^n time points.

The same technique is used on a different level to identify which domain elements correspond to neighbouring rows in the ceiling. We make sure that eventually, we have at each time point a different double exponential counter value represented by some domain element. At each time point, we can then

identify two neighbouring domain elements easily: the one with a counter value of 0 satisfies $P_{=1}\neg\text{Bit}$, and the one with a counter value of $2^{2^n} - 1$ satisfies $P_{=1}\text{Bit}$. We can thus test the vertical tiling conditions with the following axiom:

$$\Box \bigwedge_{t \in T} \left(\neg(t \sqcap P_{=1}\text{Bit} \sqsubseteq \perp) \rightarrow \bigvee_{\langle t, t' \rangle \in V} (P_{=1}\neg\text{Bit} \sqsubseteq t') \right).$$

The reduction allows us to establish the following theorem.

Theorem 4. *Satisfiability of TPDFs is 2EXPSpace-hard. This already holds if*

- no CQs are used,
- the underlying DL is \mathcal{ALC} ,
- probabilistic operators are only used on the level of concepts,
- $N_{Crig} = N_{Rrig} = \emptyset$, and
- on the axiom level, we only use Boolean connectives and the operator \Box , which does not occur under a negation operator.

6 Conclusion

In the context of description logics, temporal and probabilistic extensions have mostly been investigated in isolation, and similarly, such extensions on DL languages and query languages have not been investigated in combination. In this paper, we presented several results towards filling these gaps. First, we showed tight complexity bounds for a setting where temporal operators are used on the level of axioms and queries, as well as on queries, axioms and concepts in combination, showing that the overall complexity does not increase by such a combination for any DL between \mathcal{ALC} and \mathcal{ALCCQ} or \mathcal{ALCOI} . Second, we considered the setting where probability operators may be used on the level of concepts, axioms and queries, obtaining an NEXPTIME upper bound if the underlying DL is \mathcal{ALCCQ} , and an N2EXPTIME upper bound if the underlying DL is \mathcal{ALCCQ} or \mathcal{ALCOI} . Finally, we showed that the combination of both temporal and probabilistic operators on concepts and axioms results in 2EXPSpace-hardness. We believe that it might be possible to obtain matching upper bounds by a combination of the structures we used in this paper to show our upper bound.

While temporal ABoxes can be easily encoded into TPDFs, our results do not generalise the settings with probabilistic ABoxes studied in [21], or in the work in [23] on temporal probabilistic query answering, since these works assume the probability measure on the possible worlds to be fixed, which is not the case with our semantics. We believe however that extending to such settings does not have an impact on the complexity, as our languages are all already EXPTIME-hard. Another possible direction is investigating special operators that are both temporal and probabilistic in nature, such as the probabilistic diamond-operator introduced in [24,25].

References

1. Artale, A., Kontchakov, R., Ryzhikov, V., Zakharyashev, M.: Tractable interval temporal propositional and description logics. In: Bonet, B., Koenig, S. (eds.) Proc. of the 29th AAAI Conf. on Artificial Intelligence (AAAI'15). pp. 1417–1423. AAAI Press (2015), <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9638>
2. Artale, A., Kontchakov, R., Wolter, F., Zakharyashev, M.: Temporal description logic for ontology-based data access. In: Rossi, F. (ed.) IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013. pp. 711–717. IJCAI/AAAI (2013), <http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6824>
3. Baader, F., Borgwardt, S., Koopmann, P., Ozaki, A., Thost, V.: Metric temporal description logics with interval-rigid names. In: Proc. FroCoS 2017. pp. 60–76. Springer International (2017)
4. Baader, F., Borgwardt, S., Lippmann, M.: Temporal conjunctive queries in expressive description logics with transitive roles. In: Proceedings of the 28th Australasian Joint Conference on Artificial Intelligence (AI'15). LNAI, vol. 9457, pp. 21–33. Springer-Verlag (2015)
5. Baader, F., Borgwardt, S., Lippmann, M.: Temporal query entailment in the description logic \mathcal{SHQ} . J. Web Sem. **33**, 71–93 (2015)
6. Baader, F., Ghilardi, S., Lutz, C.: LTL over description logic axioms. ACM Trans. Comput. Log. **13**(3), 21:1–21:32 (2012). <https://doi.org/10.1145/2287718.2287721>, <https://doi.org/10.1145/2287718.2287721>
7. Baader, F., Koopmann, P., Turhan, A.: Using ontologies to query probabilistic numerical data. In: Proc. FroCoS 2017. pp. 77–94. Springer International (2017)
8. Borgwardt, S., Ceylan, I.I., Lukasiewicz, T.: Ontology-mediated queries for probabilistic databases. In: Singh, S., Markovitch, S. (eds.) Proceedings of the 31st AAAI Conf. on Artificial Intelligence (AAAI'17). pp. 1063–1069. AAAI Press, San Francisco, USA (2017)
9. Borgwardt, S., Ceylan, I.I., Lukasiewicz, T.: Ontology-mediated query answering over log-linear probabilistic data. In: Hentenryck, P.V., Zhou, Z.H. (eds.) Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI'19). AAAI Press, Honolulu, USA (2019), to appear
10. Borgwardt, S., Lippmann, M., Thost, V.: Temporal query answering in the description logic DL-Lite. In: Fontaine, P., Ringeissen, C., Schmidt, R.A. (eds.) Frontiers of Combining Systems - 9th International Symposium, FroCoS 2013, Nancy, France, September 18-20, 2013. Proceedings. Lecture Notes in Computer Science, vol. 8152, pp. 165–180. Springer (2013). https://doi.org/10.1007/978-3-642-40885-4_11, https://doi.org/10.1007/978-3-642-40885-4_11
11. Borgwardt, S., Thost, V.: Temporal query answering in the description logic \mathcal{EL} . In: Proc. IJCAI 2015. pp. 2819–2825. AAAI Press (2015)
12. Botha, L., Meyer, T., Peñaloza, R.: The Bayesian description logic \mathcal{BALC} . In: Ortiz, M., Schneider, T. (eds.) Proceedings of the 31st International Workshop on Description Logics co-located with 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018), Tempe, Arizona, US, October 27th - to - 29th, 2018. CEUR Workshop Proceedings, vol. 2211. CEUR-WS.org (2018), <http://ceur-ws.org/Vol-2211/paper-09.pdf>
13. Ceylan, I.I., Peñaloza, R.: Dynamic Bayesian description logics. In: Proc. DL 2015. CEUR-WS (2015)

14. Ceylan, İ.İ., Peñaloza, R.: The Bayesian ontology language \mathcal{BEL} . *J. Autom. Reasoning* **58**(1), 67–95 (2017). <https://doi.org/10.1007/s10817-016-9386-0>, <https://doi.org/10.1007/s10817-016-9386-0>
15. Dylla, M., Miliaraki, I., Theobald, M.: A temporal-probabilistic database model for information extraction. *Proceedings of the VLDB Endowment* **6**(14), 1810–1821 (2013)
16. van Emde Boas, P.: The convenience of tilings. *Lecture Notes in Pure and Applied Mathematics* pp. 331–363 (1997)
17. Fagin, R., Halpern, J.Y., Megiddo, N.: A logic for reasoning about probabilities. *Information and computation* **87**(1-2), 78–128 (1990)
18. Gutiérrez-Basulto, V., Jung, J.C., Kontchakov, R.: Temporalized \mathcal{EL} ontologies for accessing temporal data: Complexity of atomic queries. In: Kambhampati, S. (ed.) *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*. pp. 1102–1108. IJCAI/AAAI Press (2016), <http://www.ijcai.org/Abstract/16/160>
19. Gutiérrez-Basulto, V., Jung, J.C., Lutz, C., Schröder, L.: Probabilistic description logics for subjective uncertainty. *J. Artif. Intell. Res.* **58**, 1–66 (2017)
20. Gutiérrez-Basulto, V., Jung, J.C., Ozaki, A.: On metric temporal description logics. In: Kaminka, G.A., Fox, M., Bouquet, P., Hüllermeier, E., Dignum, V., Dignum, F., van Harmelen, F. (eds.) *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*. *Frontiers in Artificial Intelligence and Applications*, vol. 285, pp. 837–845. IOS Press (2016). <https://doi.org/10.3233/978-1-61499-672-9-837>, <https://doi.org/10.3233/978-1-61499-672-9-837>
21. Jung, J.C., Lutz, C.: Ontology-based access to probabilistic data with OWL QL. In: Cudré-Mauroux, P., Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Parreira, J.X., Hendler, J., Schreiber, G., Bernstein, A., Blomqvist, E. (eds.) *The Semantic Web - ISWC 2012 - 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I*. *Lecture Notes in Computer Science*, vol. 7649, pp. 182–197. Springer (2012). https://doi.org/10.1007/978-3-642-35176-1_12, https://doi.org/10.1007/978-3-642-35176-1_12
22. Koopmann, P.: Maybe eventually? Towards combining temporal and probabilistic description logics and queries (extended version). *LTCS-Report 19-03*, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2019), <https://tu-dresden.de/inf/lat/reports#Kr-LTCS-19-03>
23. Koopmann, P.: Ontology-based query answering for probabilistic temporal data. In: Hentenryck, P.V., Zhou, Z.H. (eds.) *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI'19)*. AAAI Press, Honolulu, USA (2019), to appear.
24. Kovtunova, A., Peñaloza, R.: Cutting diamonds: A temporal logic with probabilistic distributions. In: Thielscher, M., Toni, F., Wolter, F. (eds.) *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*. pp. 561–570. AAAI Press (2018), <https://aaai.org/ocs/index.php/KR/KR18/paper/view/18037>
25. Kovtunova, A., Peñaloza, R.: Cutting diamonds: Temporal DLs with probabilistic distributions over data. In: Ortiz, M., Schneider, T. (eds.) *Proceedings of the 31st International Workshop on Description Logics co-located with 16th International*

- Conference on Principles of Knowledge Representation and Reasoning (KR 2018), Tempe, Arizona, US, October 27th - to - 29th, 2018. CEUR Workshop Proceedings, vol. 2211. CEUR-WS.org (2018), <http://ceur-ws.org/Vol-2211/paper-23.pdf>
26. Lukasiewicz, T.: Expressive probabilistic description logics. *Artif. Intell.* **172**(6-7), 852–883 (2008). <https://doi.org/10.1016/j.artint.2007.10.017>, <https://doi.org/10.1016/j.artint.2007.10.017>
 27. Lutz, C.: Inverse roles make conjunctive queries hard. In: Proc. DL 2007. CEUR Workshop Proceedings, vol. 250. CEUR-WS.org (2007), http://ceur-ws.org/Vol-250/paper_3.pdf
 28. Lutz, C.: Two upper bounds for conjunctive query answering in *SHIQ*. In: Baader, F., Lutz, C., Motik, B. (eds.) Proceedings of the 21st International Workshop on Description Logics (DL2008), Dresden, Germany, May 13-16, 2008. CEUR Workshop Proceedings, vol. 353. CEUR-WS.org (2008), <http://ceur-ws.org/Vol-353/Lutz.pdf>
 29. Lutz, C., Wolter, F., Zakharyashev, M.: Temporal description logics: A survey. In: Demri, S., Jensen, C.S. (eds.) 15th International Symposium on Temporal Representation and Reasoning, TIME 2008, Université du Québec à Montréal, Canada, 16-18 June 2008. pp. 3–14. IEEE Computer Society (2008). <https://doi.org/10.1109/TIME.2008.14>, <https://doi.org/10.1109/TIME.2008.14>
 30. Ngo, N., Ortiz, M., Simkus, M.: Closed predicates in description logics: Results on combined complexity. In: Proc. KR 2016. pp. 237–246. AAAI Press (2016), <http://www.aaai.org/ocs/index.php/KR/KR16/paper/view/12906>
 31. Niepert, M., Noessner, J., Stuckenschmidt, H.: Log-linear description logics. In: Walsh, T. (ed.) IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011. pp. 2153–2158. IJCAI/AAAI (2011). <https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-359>, <https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-359>
 32. Ozaki, A., Krötzsch, M., Rudolph, S.: Happy ever after: Temporally attributed description logics. In: Ortiz, M., Schneider, T. (eds.) Proceedings of the 31st International Workshop on Description Logics co-located with 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018), Tempe, Arizona, US, October 27th - to - 29th, 2018. CEUR Workshop Proceedings, vol. 2211. CEUR-WS.org (2018), <http://ceur-ws.org/Vol-2211/paper-27.pdf>
 33. Özçep, Ö.L., Möller, R., Neuenstadt, C.: A stream-temporal query language for ontology based data access. In: Lutz, C., Thielscher, M. (eds.) KI 2014: Advances in Artificial Intelligence - 37th Annual German Conference on AI, Stuttgart, Germany, September 22-26, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8736, pp. 183–194. Springer (2014). https://doi.org/10.1007/978-3-319-11206-0_18, https://doi.org/10.1007/978-3-319-11206-0_18
 34. Sanati, M.Y.: A Metric Interval-Based Temporal Description Logic. PhD thesis, McMaster University, Hamilton, Canada (2015), <http://hdl.handle.net/11375/16783>
 35. Sistla, A.P., Clarke, E.M.: The complexity of propositional linear temporal logics. *Journal of the ACM (JACM)* **32**(3), 733–749 (1985)
 36. Thost, V.: Metric temporal extensions of *DL-Lite* and interval-rigid names. In: Wolter, F., Thielscher, M., Toni, F. (eds.) Proc. of the 16th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'18). pp. 665–666. AAAI Press (2018), <https://aaai.org/ocs/index.php/KR/KR18/paper/view/18035>, extended abstract.

37. Wolter, F., Zakharyashev, M.: Temporalizing description logics. In: *Frontiers of Combining Systems II*. LNCS, vol. 1794, pp. 379–402. Springer (1999)