The Bayesian Description Logic $\mathcal{BEL}$

İsmail İlkan Ceylan[1*] and Rafael Peñaloza[1,2**]

[1] Theoretical Computer Science, TU Dresden, Germany
[2] Center for Advancing Electronics Dresden
{ceylan,penaloza}@tcs.inf.tu-dresden.de

Abstract. We introduce the probabilistic Description Logic $\mathcal{BEL}$. In $\mathcal{BEL}$, axioms are required to hold only in an associated context. The probabilistic component of the logic is given by a Bayesian network that describes the joint probability distribution of the contexts. We study the main reasoning problems in this logic; in particular, we (i) prove that deciding positive and almost-sure entailments is not harder for $\mathcal{BEL}$ than for the BN, and (ii) show how to compute the probability, and the most likely context for a consequence.

1 Introduction

Description Logics (DLs) [2] are a family of knowledge representation formalisms originally designed for representing the terminological knowledge of a domain in a precise and well-understood manner. They have been successfully employed for creating large knowledge bases, representing real application domains. For instance, they are the logical formalism underlying prominent bio-medical ontologies such as SNOMED CT, GALEN, or the Gene Ontology.

Description logic ontologies are usually composed of axioms that restrict the class of possible interpretations. As these are hard restrictions, DL ontologies can only encode absolute, immutable knowledge. For some application domains, however, knowledge depends on the situation (or context) in which it is considered. For example, the notion of a *luxury hotel* in a small rural center will be different from the one in a large cosmopolitan city. When building an ontology for hotels, it makes sense to contextualize the axioms according to location, and possibly other factors like season, type of weather, etc. Since these contexts refer to notions that are external to the domain of interest, it is not always desirable, or even possible, to encode them directly into the classical DL axioms.

We follow a different approach for handling contextual knowledge. We label every axiom with the context in which it is valid. For example, we could have statements like $\langle \mathsf{LuxuryHotel} \sqsubseteq \exists \mathsf{hasFeature}.\mathsf{MeetingRoom} : \mathsf{city} \rangle$ stating that in the context of a city, every luxury hotel has a meeting room. This axiom imposes no restriction in case the context is not a city: it might still hold, or not, depending on other factors.

* Supported by DFG within the Research Training Group "RoSI" (GRK 1907).
** Partially supported by DFG within the Cluster of Excellence 'cfAED'.

Labeling the axioms in an ontology allows us to give a more detailed description of the knowledge domain. Reasoning in these cases can be used to infer knowledge that is guaranteed to hold in any given context. While the knowledge *within* this context is precise, there might be a level of uncertainty regarding the current context. To model this uncertainty, we attach a probability to each of the possible contexts. Since we cannot assume that the contexts are (probabilistically) independent, we need to describe the joint probability distribution over the space of all contexts. Thus, we consider knowledge bases that are composed of an ontology labeled with contextual information, together with a joint probability distribution over the space of contexts.

To represent the probabilistic component of the knowledge base, we use Bayesian networks (BNs) [14], a well-known probabilistic graphical model that allows for a compact representation of the probability distribution, with the help of conditional independence assumptions. For the logical component, we focus on $\mathcal{EL}$ [1], a light-weight DL that allows for polynomial-time reasoning. These formalisms together yield the Bayesian DL $\mathcal{BEL}$.

We study classical and probabilistic reasoning problems in $\mathcal{BEL}$. Not surprisingly, reasoning in this logic is intractable in general, as is reasoning in BNs already. However, we show that hardness arises exclusively from the probabilistic component: the parameterized complexity of reasoning is polynomial, if the size of the BN is considered as a parameter.

The choice of $\mathcal{EL}$ as underlying logical formalism is meant as a simple prototypical case. It allows us to understand the subtleties of combining BNs with DLs, as a first step towards more expressive formalisms. For a preliminary discussion on more expressive Bayesian DLs, and additional details and examples for $\mathcal{BEL}$, see [9].

2 The Description Logic $\mathcal{BEL}$

The DL $\mathcal{BEL}$ is a probabilistic extension of the light-weight DL $\mathcal{EL}$, where probabilities are encoded using a Bayesian network [14]. Formally, a *Bayesian network* (BN) is a pair $\mathcal{B} = (G, \Phi)$, where $G = (V, E)$ is a finite directed acyclic graph (DAG) whose nodes represent Boolean random variables,[3] and Φ contains, for every node $x \in V$, a conditional probability distribution $P_\mathcal{B}(x \mid \pi(x))$ of x given its parents $\pi(x)$. If V is the set of nodes in G, we say that $\mathcal{B}$ is a BN *over* V.

The idea behind BNs is that $G = (V, E)$ encodes a series of conditional independence assumptions between the random variables. More precisely, every variable $x \in V$ is conditionally independent of its non-descendants given its parents. Thus, every BN $\mathcal{B}$ defines a unique joint probability distribution (JPD) over V given by

$$P_\mathcal{B}(V) = \prod_{x \in V} P_\mathcal{B}(x \mid \pi(x)).$$

[3] In their general form, BNs allow for arbitrary discrete random variables. We restrict w.l.o.g. to Boolean variables for ease of presentation.

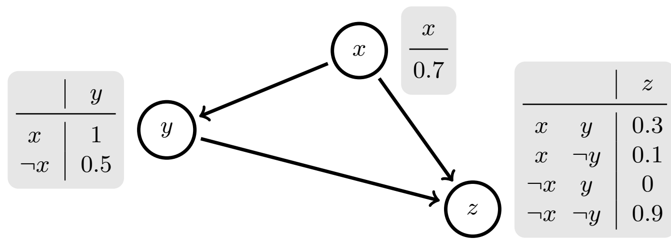

Fig. 1: The BN $\mathcal{B}_0$ over $V_0 = \{x, y, z\}$

A very simple BN is shown in Figure 1. From this network we can derive e.g. $P(x, \neg y, z) = P(z \mid x, \neg y) \cdot P(\neg y \mid x) \cdot P(x) = 0.1 \cdot 0 \cdot 0.7 = 0$.

As with classical DLs, the main building blocks in $\mathcal{BEL}$ are *concepts*, which are syntactically built as $\mathcal{EL}$ concepts. Given two disjoint sets $\mathsf{N_C}$ and $\mathsf{N_R}$ of *concept names* and *role names*, respectively, $\mathcal{BEL}$ concepts are defined through the syntactic rule

$$C ::= A \mid \top \mid C \sqcap C \mid \exists r.C$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. In DLs, the domain knowledge is typically encoded as a finite set of general concept inclusions (GCIs), called a TBox. $\mathcal{BEL}$ generalizes classical TBoxes by annotating the GCIs with a context, defined by a set of literals belonging to a BN.

Definition 1 (KB). *Let V be a finite set of Boolean variables. A V-*literal *is an expression of the form x or $\neg x$, where $x \in V$; a V-*context *is a consistent set of V-literals.*

*A V-*restricted general concept inclusion *(V-GCI) is an expression of the form $\langle C \sqsubseteq D : \kappa \rangle$ where C and D are $\mathcal{BEL}$ concepts and κ is a V-context. A V-*TBox *is a finite set of V-GCIs.*

A $\mathcal{BEL}$ knowledge base *(KB) over V is a pair $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ where $\mathcal{B}$ is a BN over V and $\mathcal{T}$ is a V-TBox.*

Intuitively, a V-GCI is an axiom that is only guaranteed to hold when its context is enforced. The semantics of this logic is defined with the help of interpretations that map concept and role names to unary and binary predicates, respectively; additionally, these interpretations evaluate the random variables from the BN.

Definition 2 (interpretation). *Given a finite set of Boolean variables V, a V-*interpretation *is a tuple $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$ where $\Delta^{\mathcal{I}}$ is a non-empty set called the* domain, *$\mathcal{V}^{\mathcal{I}} : V \to \{0, 1\}$ is a* valuation *of the variables in V, and $\cdot^{\mathcal{I}}$ is an* interpretation function *that maps every concept name A to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and every role name r to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.*

When there is no danger of ambiguity, we will usually ignore the parameter V and speak simply of e.g. a *TBox*, a *KB*, or an *interpretation*.

The interpretation function $\cdot^{\mathcal{I}}$ is extended to arbitrary $\mathcal{BEL}$ concepts by the following rules.

- $\top^{\mathcal{I}} := \Delta^{\mathcal{I}}$
- $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$
- $(\exists r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \exists e \in \Delta^{\mathcal{I}} : (d,e) \in r^{\mathcal{I}} \text{and } e \in C^{\mathcal{I}}\}$

The valuation $\mathcal{V}^{\mathcal{I}}$ is extended to contexts by defining, for every $x \in V$, $\mathcal{V}^{\mathcal{I}}(\neg x) = 1 - \mathcal{V}^{\mathcal{I}}(x)$, and for every context κ,

$$\mathcal{V}^{\mathcal{I}}(\kappa) = \min_{\ell \in \kappa} \mathcal{V}^{\mathcal{I}}(\ell),$$

where $\min_{\ell \in \emptyset} \mathcal{V}^{\mathcal{I}}(\ell) := 1$. Intuitively, a context κ can be thought as a conjunct of literals, which is evaluated to 1 iff each conjunct is so and 0 otherwise. We say that the V-interpretation $\mathcal{I}$ is a *model* of the GCI $\langle C \sqsubseteq D : \kappa \rangle$, denoted as $\mathcal{I} \models \langle C \sqsubseteq D : \kappa \rangle$, iff (i) $\mathcal{V}^{\mathcal{I}}(\kappa) = 0$, or (ii) $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. It is a *model* of the TBox $\mathcal{T}$ iff it is a model of all the GCIs in $\mathcal{T}$. The idea is that the restriction $C \sqsubseteq D$ is only required to hold whenever the context κ is satisfied. Thus, any interpretation that violates the context trivially satisfies the whole axiom.

Example 3. Let $V_0 = \{x, y, z\}$, and consider the V_0-TBox

$$\mathcal{T}_0 := \{\, \langle A \sqsubseteq C : \{x, y\} \rangle \,,\ \langle A \sqsubseteq B : \{\neg x\} \rangle \,,\ \langle B \sqsubseteq C : \{\neg x\} \rangle \}.$$

The interpretation $\mathcal{I}_0 = (\{d\}, \cdot^{\mathcal{I}_0}, \mathcal{V}_0)$ where $\mathcal{V}_0(\{x, \neg y, z\}) = 1$, $A^{\mathcal{I}_0} = \{d\}$, and $B^{\mathcal{I}_0} = C^{\mathcal{I}_0} = \emptyset$ is a model of $\mathcal{T}_0$, but is not a model of the GCI $\langle A \sqsubseteq B : \{x\} \rangle$.

The classical DL $\mathcal{EL}$ can be seen as a special case of $\mathcal{BEL}$ in which all GCIs are associated with an empty context; that is, are of the form $\langle C \sqsubseteq D : \emptyset \rangle$. Notice that every valuation satisfies the empty context $\emptyset$. Thus, a V-interpretation $\mathcal{I}$ satisfies the GCI $\langle C \sqsubseteq D : \emptyset \rangle$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. We say that $\mathcal{T}$ *entails* $\langle C \sqsubseteq D : \emptyset \rangle$, denoted by $\mathcal{T} \models C \sqsubseteq D$, if every model of $\mathcal{T}$ is also a model of $\langle C \sqsubseteq D : \emptyset \rangle$. For a valuation $\mathcal{W}$ of the variables in V, we can define a TBox containing all axioms that must be satisfied in any V-interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$ with $\mathcal{V}^{\mathcal{I}} = \mathcal{W}$.

Definition 4 (restriction). *Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a KB. The* restriction *of $\mathcal{T}$ to a valuation $\mathcal{W}$ of the variables in V is the TBox*

$$\mathcal{T}_{\mathcal{W}} := \{\langle C \sqsubseteq D : \emptyset \rangle \mid \langle C \sqsubseteq D : \kappa \rangle \in \mathcal{T}, \mathcal{W}(\kappa) = 1\}.$$

So far, our semantics have focused on the evaluation of the Boolean variables and the interpretation of concepts, ignoring the probabilistic information provided by the BN. To handle these probabilities, we introduce multiple-world semantics next. Intuitively, a V-interpretation describes a possible world; by assigning a probabilistic distribution over these interpretations, we describe the required probabilities, which should be consistent with the BN.

Definition 5 (probabilistic model). *A* probabilistic interpretation *is a pair $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$, where $\mathfrak{I}$ is a set of V-interpretations and $P_{\mathfrak{I}}$ is a probability distribution over $\mathfrak{I}$ such that $P_{\mathfrak{I}}(\mathcal{I}) > 0$ only for finitely many interpretations $\mathcal{I} \in \mathfrak{I}$.*

This probabilistic interpretation is a model *of the TBox $\mathcal{T}$ if every $\mathcal{I} \in \mathfrak{I}$ is a model of $\mathcal{T}$. $\mathcal{P}$ is* consistent *with the BN $\mathcal{B}$ if for every possible valuation $\mathcal{W}$ of the variables in V it holds that*

$$\sum_{\mathcal{I} \in \mathfrak{I},\, \mathcal{V}^{\mathcal{I}} = \mathcal{W}} P_{\mathfrak{I}}(\mathcal{I}) = P_{\mathcal{B}}(\mathcal{W}).$$

The probabilistic interpretation $\mathcal{P}$ is a model *of the KB $(\mathcal{B}, \mathcal{T})$ iff it is a (probabilistic) model of $\mathcal{T}$ and consistent with $\mathcal{B}$.*

One simple consequence of this semantics is that probabilistic models preserve the probability distribution of $\mathcal{B}$ for subsets of literals; i.e., contexts. The proof follows from the fact that a context corresponds to a partial valuation. Hence, the probability of a context κ is the sum of the probabilities of all valuations that extend κ.

Theorem 6. *Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a KB, and κ a context. For every model $\mathcal{P}$ of $\mathcal{K}$ it holds that*

$$\sum_{\mathcal{I} \in \mathfrak{I},\, \mathcal{V}^{\mathcal{I}}(\kappa) = 1} P_{\mathfrak{I}}(\mathcal{I}) = P_{\mathcal{B}}(\mathcal{V}^{\mathcal{I}}(\kappa)).$$

For the following sections it will be useful for proving our results to consider a special kind of interpretations, which we call *pithy*. These interpretations contain at most one V-interpretation for each valuation of the variables in V. Each of these V-interpretations provides the essential information associated to the corresponding valuation.

Definition 7 (pithy). *The probabilistic interpretation $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$ is called* pithy *if for every valuation $\mathcal{W}$ of the variables in V there exists at most one V-interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}}) \in \mathfrak{I}$ such that $\mathcal{V}^{\mathcal{I}} = \mathcal{W}$.*

We now study classical and probabilistic reasoning problems in $\mathcal{BEL}$, and analyse their complexity.

3 Reasoning in $\mathcal{BEL}$

In the previous section we have described how probabilistic knowledge can be represented using a $\mathcal{BEL}$ KB. We now focus our attention to reasoning with this knowledge. The most basic decision problem in any DL is whether an ontology is consistent. It turns out that, as for classical $\mathcal{EL}$, this problem is trivial in $\mathcal{BEL}$.

Theorem 8. *Every $\mathcal{BEL}$ KB is consistent.*

Proof (Sketch). Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a $\mathcal{BEL}$ KB. Let $\Delta^{\mathcal{I}} = \{a\}$ and $\cdot^{\mathcal{I}}$ be such that $A^{\mathcal{I}} = \{a\}$ and $r^{\mathcal{I}} = \{(a, a)\}$ for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. For every valuation $\mathcal{W}$, define the V-interpretation $\mathcal{I}_{\mathcal{W}} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{W})$. Then, the probabilistic interpretation $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$ where $\mathfrak{I} = \{\mathcal{I}_{\mathcal{W}} \mid \mathcal{W} \text{ is a valuation}\}$ and $P_{\mathfrak{I}}(\mathcal{I}_{\mathcal{W}}) = P_{\mathcal{B}}(\mathcal{W})$ is a model of $\mathcal{K}$.

A more interesting reasoning problem is subsumption: decide whether a concept is interpreted as a subclass of another one. We generalize this problem to consider also the contexts and probabilities provided by the BN.

Definition 9 (subsumption). *Let C, D be two $\mathcal{BEL}$ concepts, κ a context, and $\mathcal{K}$ a KB. C is* contextually subsumed *by D in κ w.r.t. $\mathcal{K}$, denoted as $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle$, if every probabilistic model of $\mathcal{K}$ is also a model of the TBox $\{\langle C \sqsubseteq D : \kappa\rangle\}$. For a probabilistic interpretation $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$, we define the* probability of a consequence *$P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa\rangle) := \sum_{\mathcal{I}\in\mathfrak{I},\mathcal{I}\models\langle C\sqsubseteq D:\kappa\rangle} P_{\mathfrak{I}}(\mathcal{I})$. The* probability of *$\langle C \sqsubseteq D : \kappa\rangle$* w.r.t. *$\mathcal{K}$ is defined as*

$$P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle) := \inf_{\mathcal{P}\models\mathcal{K}} P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa\rangle).$$

We say that C is positively subsumed *by D in κ if $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle) > 0$, and C is p-*subsumed *by D in κ, for $p \in (0,1]$ if $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle) \geq p$. We sometimes refer to 1-subsumption as* almost-sure *subsumption.*

Clearly, if C is subsumed by D in κ w.r.t. a KB $\mathcal{K}$, then $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle) = 1$. The converse, however, may not hold since the subsumption relation might be violated in V-interpretations of probability zero.

Example 10. Consider the KB $\mathcal{K}_0 = (\mathcal{B}_0, \mathcal{T}_0)$, where $\mathcal{B}_0$ is the BN depicted in Figure 1 and $\mathcal{T}_0$ the TBox from Example 3. It follows that $P(\langle A \sqsubseteq_{\mathcal{K}_0} C : \emptyset\rangle) = 1$ and $P(\langle C \sqsubseteq_{\mathcal{K}_0} B : \{x, y\}\rangle) = 0$. Moreover, for any two concepts E, F, it holds that $P(\langle E \sqsubseteq_{\mathcal{K}_0} F : \{x, \neg y\}\rangle) = 1$ since $\langle E \sqsubseteq_{\mathcal{K}_0} F : \{x, \neg y\}\rangle$ can only be violated in V-interpretations that have probability 0. However, in general the consequence $\langle E \sqsubseteq_{\mathcal{K}_0} F : \{x, \neg y\}\rangle$ does not hold.

3.1 Probabilistic Subsumption

We consider first the problem of computing the probability of a subsumption, or deciding positive, p-subsumption, and almost-sure subsumption. As an intermediate step, we show that it is possible w.l.o.g. to restrict reasoning to pithy models.

Lemma 11. *Let $\mathcal{K}$ be a KB. If $\mathcal{P}$ is a probabilistic model of $\mathcal{K}$, then a pithy model $\mathcal{Q}$ of $\mathcal{K}$ can be computed such that for every two concepts C, D and context κ it holds that $P(\langle C \sqsubseteq_{\mathcal{Q}} D : \kappa\rangle) \leq P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa\rangle)$.*

Proof (Sketch). Let $\mathcal{W}$ be a valuation and $\mathcal{I}, \mathcal{I}' \in \mathfrak{I}$ two V-interpretations such that $\mathcal{V}^{\mathcal{I}} = \mathcal{V}^{\mathcal{I}'} = \mathcal{W}$. Construct a new interpretation $\mathcal{J}$ as the disjoint union of $\mathcal{I}$ and $\mathcal{I}'$. The probabilistic interpretation $(\mathfrak{H}, P_{\mathfrak{H}})$ with $\mathfrak{H} = (\mathfrak{I} \cup \{\mathcal{J}\}) \setminus \{\mathcal{I}, \mathcal{I}'\}$ and

$$P_{\mathfrak{H}}(\mathcal{H}) := \begin{cases} P_{\mathfrak{I}}(\mathcal{H}) & \mathcal{H} \neq \mathcal{J} \\ P_{\mathfrak{I}}(\mathcal{I}) + P_{\mathfrak{I}}(\mathcal{I}') & \mathcal{H} = \mathcal{J} \end{cases}$$

is a model of $\mathcal{K}$. Moreover, $\mathcal{J} \models \langle C \sqsubseteq D : \kappa\rangle$ iff both $\mathcal{I} \models \langle C \sqsubseteq D : \kappa\rangle$ and $\mathcal{I}' \models \langle C \sqsubseteq D : \kappa\rangle$. □

As we show next, the probability of a consequence can be computed by reasoning over the restrictions $\mathcal{T}_\mathcal{W}$ of $\mathcal{T}$.

Theorem 12. *Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a KB, C, D two concepts and κ a context.*

$$P(\langle C \sqsubseteq_\mathcal{K} D : \kappa\rangle) = 1 - P_\mathcal{B}(\kappa) + \sum_{\substack{\mathcal{T}_\mathcal{W} \models C \sqsubseteq D \\ \mathcal{W}(\kappa)=1}} P_\mathcal{B}(\mathcal{W}).$$

Proof. For every valuation $\mathcal{W}$ construct the V-interpretation $\mathcal{I}_\mathcal{W}$ as follows. If $\mathcal{T}_\mathcal{W} \models C \sqsubseteq D$, then $\mathcal{I}_\mathcal{W}$ is any model $(\Delta^{\mathcal{I}_\mathcal{W}}, \cdot^{\mathcal{I}_\mathcal{W}}, \mathcal{W})$ of $\mathcal{T}_\mathcal{W}$; otherwise, $\mathcal{I}_\mathcal{W}$ is any model $(\Delta^{\mathcal{I}_\mathcal{W}}, \cdot^{\mathcal{I}_\mathcal{W}}, \mathcal{W})$ of $\mathcal{T}_\mathcal{W}$ that does not satisfy $\langle C \sqsubseteq D : \kappa\rangle$, which must exist by definition. The probabilistic interpretation $\mathcal{P}_\mathcal{K} = (\mathfrak{I}, P_\mathfrak{I})$ such that $\mathfrak{I} = \{\mathcal{I}_\mathcal{W} \mid \mathcal{W} \text{ a valuation of } V\}$ and $P_\mathfrak{I}(\mathcal{I}_\mathcal{W}) = P_\mathcal{B}(\mathcal{W})$ for all $\mathcal{W}$ is a model of $\mathcal{K}$ and

$$\begin{aligned} P(\langle C \sqsubseteq_{\mathcal{P}_\mathcal{K}} D : \kappa\rangle) &= \sum_{\mathcal{I}_\mathcal{W} \models \langle C \sqsubseteq D : \kappa\rangle} P_\mathfrak{I}(\mathcal{I}_\mathcal{W}) \\ &= \sum_{\mathcal{W}(\kappa)=0} P_\mathfrak{I}(\mathcal{I}_\mathcal{W}) + \sum_{\substack{\mathcal{W}(\kappa)=1, \\ \mathcal{I}_\mathcal{W} \models \langle C \sqsubseteq D : \kappa\rangle}} P_\mathfrak{I}(\mathcal{I}_\mathcal{W}) \\ &= 1 - P_\mathcal{B}(\kappa) + \sum_{\substack{\mathcal{T}_\mathcal{W} \models C \sqsubseteq D \\ \mathcal{W}(\kappa)=1}} P_\mathcal{B}(\mathcal{W}). \end{aligned}$$

Thus, $P(\langle C \sqsubseteq_\mathcal{K} D : \kappa\rangle) \leq 1 - P_\mathcal{B}(\kappa) + \sum_{\mathcal{T}_\mathcal{W} \models C \sqsubseteq D, \mathcal{W}(\kappa)=1} P_\mathcal{B}(\mathcal{W})$. Suppose now that the inequality is strict, then there exists a probabilistic model $\mathcal{P} = (\mathfrak{J}, P_\mathfrak{J})$ of $\mathcal{K}$ such that $P(\langle C \sqsubseteq_\mathcal{P} D : \kappa\rangle) < P(\langle C \sqsubseteq_{\mathcal{P}_\mathcal{K}} D : \kappa\rangle)$. By Lemma 11, we can assume w.l.o.g. that $\mathcal{P}$ is pithy, and hence for every valuation $\mathcal{W}$ with $P_\mathcal{B}(\mathcal{W}) > 0$ there exists exactly one $\mathcal{J}_\mathcal{W} \in \mathfrak{J}$ with $\mathcal{V}^{\mathcal{J}_\mathcal{W}} = \mathcal{W}$. We thus have

$$\sum_{\mathcal{J}_\mathcal{W} \models \langle C \sqsubseteq D : \kappa\rangle, \mathcal{W}(\kappa)=1} P_\mathfrak{J}(\mathcal{J}_\mathcal{W}) < \sum_{\mathcal{I}_\mathcal{W} \models \langle C \sqsubseteq D : \kappa\rangle, \mathcal{W}(\kappa)=1} P_\mathfrak{J}(\mathcal{I}_\mathcal{W}).$$

Since $P_\mathfrak{J}(\mathcal{I}_\mathcal{W}) = P_\mathfrak{J}(\mathcal{J}_\mathcal{W})$ for all $\mathcal{W}$, then there must exist a valuation $\mathcal{V}$ such that $\mathcal{I}_\mathcal{V} \models \langle C \sqsubseteq D : \kappa\rangle$ but $\mathcal{J}_\mathcal{V} \not\models \langle C \sqsubseteq D : \kappa\rangle$. Since $\mathcal{J}_\mathcal{V}$ is a model of $\mathcal{T}_\mathcal{V}$ it follows that $\mathcal{T}_\mathcal{V} \not\models C \sqsubseteq D$. By construction, then we have that $\mathcal{I}_\mathcal{V} \not\models \langle C \sqsubseteq D : \kappa\rangle$, which is a contradiction. □

Based on this theorem, we can compute the probability of a subsumption as described in Algorithm 1. The algorithm simply verifies for all possible valuations $\mathcal{W}$, whether $\mathcal{T}_\mathcal{W}$ entails the desired axiom. Clearly, the **for** loop is executed $2^{|V|}$ times; that is, once for each possible valuation of the variables in V. Each of these executions needs to compute the probability $P_\mathcal{B}(\mathcal{W})$ and, possibly, decide whether $\mathcal{T}_\mathcal{W} \models C \sqsubseteq D$. The former can be done in polynomial time on the size of $\mathcal{B}$, using the standard chain rule [14], while deciding entailment from an $\mathcal{EL}$ TBox is polynomial on $\mathcal{T}$ [8]. Overall, Algorithm 1 runs in time exponential on

Algorithm 1 Probability of Subsumption

Input: KB $\mathcal{K} = (\mathcal{B}, \mathcal{T})$, GCI $\langle C \sqsubseteq D : \kappa \rangle$
Output: $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle)$

$P \leftarrow 0, Q \leftarrow 0$
for all valuations $\mathcal{W}$ **do**
  if $\mathcal{W}(\kappa) = 0$ **then**
    $Q \leftarrow Q + P_{\mathcal{B}}(\mathcal{W})$
  else if $\mathcal{T}_{\mathcal{W}} \models C \sqsubseteq D$ **then**
    $P \leftarrow P + P_{\mathcal{B}}(\mathcal{W})$
return $1 - Q + P$

$\mathcal{B}$ but polynomial on $\mathcal{T}$. Moreover, the algorithm requires only polynomial space since the different valuations can be enumerated using only $|V|$ bits. Thus, we obtain the following result.

Theorem 13. *The problem of deciding p-subsumption is in* PSPACE *and fixed-parameter tractable where* $|V|$ *is the parameter.*[4]

As a lower bound, unsurprisingly, p-subsumption is at least as hard as deciding probabilities from the BN. Since this latter problem is hard for the class PP [19], we get the following result.

Theorem 14. *Deciding p-subsumption is* PP*-hard.*

If we are interested only in deciding positive or almost-sure subsumption, then we can further improve these upper bounds to NP and coNP, respectively.

Theorem 15. *Deciding positive subsumption is* NP*-complete. Deciding almost-sure subsumption is co*NP*-complete.*

Proof. To decide positive subsumption, we can simply guess a valuation $\mathcal{W}$ and check in polynomial time that (i) $P_{\mathcal{B}}(\mathcal{W}) > 0$ and (ii) either $\mathcal{W}(\kappa) = 0$ or $\mathcal{T}_{\mathcal{W}} \models C \sqsubseteq D$. The correctness of this algorithm is given by Theorem 12. Thus the problem is in NP.

To show hardness, we recall that deciding, given a BN $\mathcal{B}$ and a variable $x \in V$, whether $P_{\mathcal{B}}(x) > 0$ is NP-hard [11]. Consider the KB $\mathcal{K} = (\mathcal{B}, \emptyset)$ and A, B two arbitrary concept names. It follows from Theorem 12 that $P_{\mathcal{B}}(x) > 0$ iff $P(\langle A \sqsubseteq_{\mathcal{K}} B : \{\neg x\} \rangle) > 0$. Thus positive subsumption is NP-hard. The coNP-completeness of almost-sure subsumption can be shown analogously. □

Notice once again that the non-determinism needed to solve these problems is limited to the number of random variables in $\mathcal{B}$. More precisely, exactly $|V|$ bits need to be non-deterministically guessed, and the rest of the computation runs in polynomial time. In practical terms this means that subsumption is tractable

[4] Recall that a problem is fixed-parameter tractable if it can be solved in polynomial time, assuming that the parameter is fixed [15].

as long as the DAG remains small. On the other hand, Algorithm 1 shows that the probabilistic and the logical components of the KB can be decoupled while reasoning. This is an encouraging result as it means that one can apply the optimized methods developed for BN inference and for DL reasoning directly in $\mathcal{BEL}$ without major modifications.

3.2 Contextual Subsumption

We now turn our attention to deciding whether a contextual subsumption relation follows from all models of the KB in a classical sense; that is, whether $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle$ holds. Contrary to classical $\mathcal{EL}$, subsumption in $\mathcal{BEL}$ is already intractable, even if we consider only the empty context.

Theorem 16. *Let $\mathcal{K}$ be a KB and C, D two concepts. Deciding $\langle C \sqsubseteq_{\mathcal{K}} D : \emptyset\rangle$ is co*NP*-hard.*

Proof. We present a reduction from validity of DNF formulas, which is known to be coNP-hard [10]. Let $\phi = \sigma_1 \vee \ldots \vee \sigma_n$ be a DNF formula where each σ_i is a conjunctive clause and let V be the set of all variables appearing in ϕ. For each variable $x \in V$, we introduce the concept names B_x and $B_{\neg x}$ and define the TBox $\mathcal{T}_x := \{\langle A \sqsubseteq B_x : \{x\}\rangle, \langle A \sqsubseteq B_{\neg x} : \{\neg x\}\rangle\}$. For every conjunctive clause $\sigma = \ell_1 \wedge \ldots \wedge \ell_m$ define the TBox $\mathcal{T}_\sigma := \{\langle B_{\ell_1} \sqcap \ldots \sqcap B_{\ell_m} \sqsubseteq C : \emptyset\rangle\}$. Let now $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ where $\mathcal{B}$ is an arbitrary BN over V and $\mathcal{T} = \bigcup_{x \in V} \mathcal{T}_x \cup \bigcup_{1 \leq i \leq n} \mathcal{T}_{\sigma_i}$. It is easy to see that ϕ is valid iff $\langle A \sqsubseteq_{\mathcal{K}} C : \emptyset\rangle$. □

The main reason for this hardness is that the interaction of contexts might produce consequences that are not obvious at first sight. For instance, a consequence might follow in context κ not because the axioms from κ entail the consequence, but rather because any valuation satisfying κ will yield it. That is the main idea in the proof of Theorem 16; the axioms that follow directly from the empty context never entail the subsumption $A \sqsubseteq C$, but if ϕ is valid, then this subsumption follows from all valuations. We obtain the following result.

Lemma 17. *Let $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ be a KB. Then $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle$ iff for every valuation $\mathcal{W}$ with $\mathcal{W}(\kappa) = 1$, it holds that $\mathcal{T}_{\mathcal{W}} \models C \sqsubseteq D$.*

It thus suffices to identify all valuations that define TBoxes entailing the consequence. To do this, we will take advantage of techniques developed in the area of axiom-pinpointing [6], access control [3], and context-based reasoning [4]. It is worth noticing that subsumption relations depend only on the TBox and not on the BN. For that reason, for the rest of this section we focus only on the terminological part of the KB.

We can think of every context κ as the conjunctive clause $\chi_\kappa := \bigwedge_{\ell \in \kappa} \ell$. In this view, the V-TBox $\mathcal{T}$ is a labeled TBox over the (distributive) lattice $\mathbb{B}$ of all Boolean formulas over the variables V, modulo equivalence. Each formula ϕ in this lattice defines a sub-TBox $\mathcal{T}_\phi$ which contains all axioms $\langle C \sqsubseteq D : \kappa\rangle \in \mathcal{T}$ such that $\chi_\kappa \models \phi$.

Using the terminology from [4], we are interested in finding a boundary for a consequence. Given a TBox $\mathcal{T}$ labeled over the lattice $\mathbb{B}$ and concepts C, D, a *boundary* for $C \sqsubseteq D$ w.r.t. $\mathcal{T}$ is an element $\phi \in \mathbb{B}$ such that for every join-prime element $\psi \in \mathbb{B}$ it holds that $\psi \models \phi$ iff $\mathcal{T}_\psi \models C \sqsubseteq D$ (see [4] for further details). Notice that the join-prime elements of $\mathbb{B}$ are exactly the valuations of variables in V. Using Lemma 17 we obtain the following result.

Theorem 18. *Let ϕ be a boundary for $C \sqsubseteq D$ w.r.t. $\mathcal{T}$ in $\mathbb{B}$. Then, for any context κ we have that $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle$ iff $\chi_\kappa \models \phi$.*

While several methods have been developed for computing the boundary of a consequence, they are based on a *black-box* approach that makes several calls to an external reasoner. We present a *glass-box* approach that computes a compact representation of the boundary directly. This method, based on the standard completion algorithm for $\mathcal{EL}$ [8], can in fact compute the boundaries for all subsumption relations between concept names that follow from the KB.

For our completion algorithm we assume that the TBox is in normal form; i.e., all GCIs are of the form $\langle A_1 \sqcap A_2 \sqsubseteq B : \kappa\rangle$, $\langle A \sqsubseteq \exists r.B : \kappa\rangle$, or $\langle \exists r.A \sqsubseteq B : \kappa\rangle$, where $A, A_1, A_2, B \in \mathsf{N_C} \cup \{\top\}$. It is easy to see that every V-TBox can be transformed into an equivalent one in normal form in linear time.

Given a TBox in normal form, the completion algorithm uses rules to label a set of assertions until no new information can be added. Assertions are tuples of the form (A, B) or (A, r, B) where $A, B \in \mathsf{N_C} \cup \{\top\}$ and $r \in \mathsf{N_R}$ are names appearing in the TBox. The function lab maps every assertion to a Boolean formula ϕ over the variables in V. Intuitively, $\mathsf{lab}(A, B) = \phi$ expresses that $\mathcal{T}_\mathcal{W} \models A \sqsubseteq B$ in all valuations $\mathcal{W}$ that satisfy ϕ; and $\mathsf{lab}(A, r, B) = \phi$ expresses that $\mathcal{T}_\mathcal{W} \models A \sqsubseteq \exists r.B$ in all valuations $\mathcal{W}$ that satisfy ϕ. The algorithm is initialized with the labeling of assertions

$$\mathsf{lab}(\alpha) := \begin{cases} \mathsf{t} & \alpha \text{ is of the form } (A, \top) \text{ or } (A, A) \text{ for } A \in \mathsf{N_C} \cup \{\top\} \\ \mathsf{f} & \text{otherwise,} \end{cases}$$

where t is a tautology and f a contradiction in $\mathbb{B}$. This function is modified by applying the rules from Table 1 where for brevity, we denote $\mathsf{lab}(\alpha) = \phi$ by α^ϕ. Every rule application changes the label of one assertion for a more general formula. The number of assertions is polynomial on $\mathcal{T}$ and the depth of the lattice $\mathbb{B}$ is exponential on $|V|$. Thus, in the worst case, the number of rule applications is bounded exponentially on $|V|$, but polynomially on $\mathcal{T}$.

Clearly, all the rules are sound; that is, at every step of the algorithm it holds that $\mathcal{T}_\mathcal{W} \models A \sqsubseteq B$ for all concept names A, B and all valuations $\mathcal{W}$ that satisfy $\mathsf{lab}(A, B)$, and analogously for (A, r, B). It can be shown using techniques from axiom-pinpointing (see e.g. [7, 4]) that after termination the converse also holds; i.e., for every valuation $\mathcal{W}$, if $\mathcal{T}_\mathcal{W} \models A \sqsubseteq B$, then $\mathcal{W} \models \mathsf{lab}(A, B)$. Thus, we obtain the following result.

Theorem 19. *Let lab be the labelling function obtained through the completion algorithm. For every two concept names A, B appearing in $\mathcal{T}$, $\mathsf{lab}(A, B)$ is a boundary for $A \sqsubseteq B$ w.r.t. $\mathcal{T}$.*

Table 1: Completion rules for subsumption in $\mathcal{BEL}$

If $\langle A_1 \sqcap A_2 \sqsubseteq B : \kappa\rangle \in \mathcal{T}$, $(X, A_1)^{\phi_1}, (X, A_2)^{\phi_2}, (X, B)^{\psi}$, $\chi_\kappa \wedge \phi_1 \wedge \phi_2 \not\models \psi$	**then** $\mathsf{lab}(X, B) := (\chi_\kappa \wedge \phi_1 \wedge \phi_2) \vee \psi$
If $\langle A \sqsubseteq \exists r.B : \kappa\rangle \in \mathcal{T}$, $(X, A)^{\phi}, (X, r, B)^{\psi}$, $\chi_\kappa \wedge \phi \not\models \psi$	**then** $\mathsf{lab}(X, r, B) := (\chi_\kappa \wedge \phi) \vee \psi$
If $\langle \exists r.A \sqsubseteq B : \kappa\rangle \in \mathcal{T}$, $(X, r, Y)^{\phi_1}, (Y, A)^{\phi_2}, (X, B)^{\psi}$, $\chi_\kappa \wedge \phi_1 \wedge \phi_2 \not\models \psi$	**then** $\mathsf{lab}(X, B) := (\chi_\kappa \wedge \phi_1 \wedge \phi_2) \vee \psi$

Once we know a boundary ϕ for $A \sqsubseteq B$ w.r.t. $\mathcal{T}$, we can decide whether $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa\rangle$: we need only to verify whether $\chi_\kappa \models \phi$. This decision is in NP on $|V|$. Although the algorithm is described exclusively for concept names A, B, it can be used to compute a boundary for $C \sqsubseteq D$, for arbitrary $\mathcal{BEL}$ concepts C, D, simply by adding the axioms $\langle A_0 \sqsubseteq C : \emptyset\rangle$ and $\langle D \sqsubseteq B_0 : \emptyset\rangle$, where A_0, B_0 are new concept names, to the TBox, and then computing a boundary for $A_0 \sqsubseteq B_0$ w.r.t. the extended TBox. This yields the following result.

Corollary 20. *Subsumption in $\mathcal{BEL}$ can be decided in exponential time, and is fixed-parameter tractable where $|V|$ is the parameter.*

Clearly, the boundary for $C \sqsubseteq D$ provides more information than necessary for deciding whether the subsumption holds in a *given* context κ. It encodes *all* contexts that entail the desired subsumption. We can use this knowledge to deduce the most likely context.

3.3 Most Likely Context

The problem of finding the most likely context for a consequence can be seen as the dual of computing the probability of this consequence. Intuitively, we are interested in finding the most likely explanation for an event; assuming that a consequence holds, we are interested in finding an explanation for it, in the form of a context, that has the maximal probability of occurring.

Definition 21 (most likely context). *Given a KB $\mathcal{K} = (\mathcal{B}, \mathcal{T})$ and concepts C, D, the context κ is called a* most likely context *for $C \sqsubseteq D$ w.r.t. $\mathcal{K}$ if (i) $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle$, and (ii) for every context κ', if $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa'\rangle$ holds, then $P_{\mathcal{B}}(\kappa') \leq P_{\mathcal{B}}(\kappa)$.*

Notice that we are not interested in maximizing $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle)$ but rather $P_{\mathcal{B}}(\kappa)$. Indeed, these two problems can be seen as dual, since $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa\rangle)$ depends inversely, but not exclusively, on $P_{\mathcal{B}}(\kappa)$ (see Theorem 12).

Algorithm 2 computes the set of all most likely contexts for $C \sqsubseteq D$ w.r.t. $\mathcal{K}$, together with their probability. It maintains a value p of the highest known

Algorithm 2 Compute all most likely contexts

Input: KB $\mathcal{K} = (\mathcal{B}, \mathcal{T})$, concepts C, D
Output: The set Λ of most likely contexts for $C \sqsubseteq D$ w.r.t. $\mathcal{K}$ and probability $p \in [0, 1]$

```
Λ ← ∅, p ← 0
φ ← boundary(C ⊑ D, T)                ▷ compute a boundary for C ⊑ D w.r.t. T
for all contexts κ do
    if χ_κ ⊨ φ then
        if P_B(κ) > p then
            Λ ← {κ}
            p ← P_B(κ)
        else if P_B(κ) = p then
            Λ ← Λ ∪ {κ}
return Λ, p
```

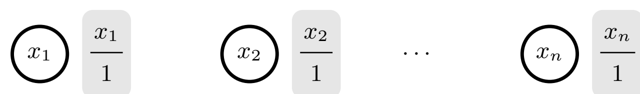

Fig. 2: The BN $\mathcal{B}_n$ over $\{x_1, \ldots, x_n\}$

probability for a context, and a set Λ with all the contexts that have probability p. The algorithm first computes a boundary for the consequence, which is used to test, for every context κ whether $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle$ holds. In that case, it compares $P_{\mathcal{B}}(\kappa)$ with p. If the former is larger, then the highest probability is updated to this value, and the set Λ is restarted to contain only κ. If they are the same, then κ is added to the set of most likely contexts.

Computing a boundary requires exponential time on $\mathcal{T}$. Likewise, the number of contexts is exponential on $\mathcal{B}$, and for each of them we have to test propositional entailment, which is also exponential on $\mathcal{B}$. Overall, we have the following.

Theorem 22. *Algorithm 2 computes all most likely contexts for $C \sqsubseteq D$ w.r.t. $\mathcal{K}$ in exponential time.*

In general, it is not possible to lower this exponential upper bound, since a simple consequence may have exponentially many most likely contexts. For example, given a natural number $n \geq 1$, let $\mathcal{B}_n = (G_n, \Phi_n)$ be the BN where $G = (\{x_1, \ldots, x_n\}, \emptyset)$, i.e., G contains n nodes and no edges connecting them, and for each $i, 1 \leq i \leq n$ Φ_n contains the distribution with $P_{\mathcal{B}}(x_i) = 1$ (see Figure 2). For every context $\kappa \subseteq \{x_1, \ldots, x_n\}$, we have that $P_{\mathcal{B}}(\kappa) = 1$ which means that there are 2^n most likely contexts for $A \sqsubseteq A$ w.r.t. the KB $(\mathcal{B}_n, \emptyset)$.

Algorithm 2 can be adapted to compute *one* most likely context in a more efficient way. The main idea is to order the calls in the **for** loop by decreasing probability. Once one context κ with $\chi_\kappa \models \phi$ has been found, it is guaranteed to be a most likely context and the algorithm may stop. This approach would still require exponential time in the worst case. However, recall that simply *verifying* whether κ is a context for $C \sqsubseteq D$ is already coNP-hard (Theorem 16), and hence deciding whether it is a most likely context is arguably hard for the second level of the polynomial hierarchy. On the other hand, this exponential bound depends exclusively on $|V|$. Hence, as before, we have that deciding whether a context is a most likely context for a consequence is fixed-parameter tractable over $|V|$.

4 Related Work

The amount of work on handling uncertain knowledge with description logics is too vast to cover in detail here. Many probabilistic description logics have been defined, which differ not only in their syntax but also in their use of the probabilities and their application. These logics were recently surveyed in [17]. We discuss here only those logics most closely related to ours.

One of the first attempts for combining BNs and DLs was P-Classic [16], which extended Classic through probability distributions over the interpretation domain. The more recent PR-OWL [12] uses multi-entity BNs to describe the probability distributions of some domain elements. In both cases, the probabilistic component is interpreted providing individuals with a probability distribution; this differs greatly from our multiple-world semantics, in which we consider a probability distribution over a set of classical DL interpretations.

Perhaps the closest to our approach are the Bayesian extension of DL-Lite [13] and DISPONTE [18]. The latter allows for so-called epistemic probabilities that express the uncertainty associated to a given axiom. Their semantics are based, as ours, on a probabilistic distribution over a set of interpretations. The main difference with our approach is that in [18], the authors assume that all probabilities are independent, while we provide a joint probability distribution through the BN. Another minor difference is that in DISPONTE it is impossible to obtain classical consequences, as we do.

Abstracting from the different logical constructors used, the logic in [13] looks almost identical to ours. There is, however, a subtle but important difference. In our approach, an interpretation $\mathcal{I}$ satisfies an axiom $\langle C \sqsubseteq D : \kappa \rangle$ if $\mathcal{V}^{\mathcal{I}}(\kappa) = 1$ implies $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. In [13], the authors employ a closed-world assumption over the contexts, where this implication is substituted for an equivalence; i.e., $\mathcal{V}^{\mathcal{I}}(\kappa) = 0$ also implies $C^{\mathcal{I}} \not\subseteq D^{\mathcal{I}}$. The use of such semantics can easily produce inconsistent KBs, which is impossible in $\mathcal{BEL}$.

5 Conclusions

We have introduced the probabilistic DL $\mathcal{BEL}$, which extends the classical $\mathcal{EL}$ to express uncertainty. Our basic assumption is that we have *certain* knowledge, which depends on an *uncertain* situation, or context. In practical terms, this means that every axiom is associated to a context with the intended meaning that, if the context holds, then the axiom must be true. Uncertainty is represented through a BN that encodes the probability distribution of the contexts. The advantage of using Bayesian networks relies in their capacity of describing conditional independence assumptions in a compact manner.

We have studied the complexity of reasoning in this probabilistic logic. Contrary to classical $\mathcal{EL}$, reasoning in $\mathcal{BEL}$ is in general intractable. More precisely, we have shown that positive subsumption is NP-complete, and almost-sure subsumption is coNP-complete. For the other reasoning problems we have not found tight complexity bounds, but we proved that p-subsumption is NP-hard and in PSpace, while contextual subsumption and deciding most likely contexts are between coNP and ExpTime.

In contrast to these negative complexity results, we have shown that the complexity can be decoupled between the probabilistic and the logical components of the KB. Indeed, all these problems are fixed-parameter tractable over the parameter $|V|$. This means that, if we have a fixed number of contexts, then all these problems can be solved in polynomial time. It is not unreasonable, moreover, to assume that the number of contexts is quite small in comparison to the size of the TBox. Finally notice that reasoning with the BN itself is already intractable. What we have shown is that intractability is a consequence of the contextual and probabilistic components of the KB, and not of the logical one.

There are several directions for future work. First, we would like to tighten our complexity results. Notice that the main bottleneck in our algorithms for deciding contextual subsumption and computing the most likely contexts is the computation of the boundary, which requires exponential time. It has been argued that a compact representation of the pinpointing formula, which is a special case of the boundary, can be computed in polynomial time for $\mathcal{EL}$ using an automata-based approach [5]. If making logical inferences over this compact encoding is not harder than for the formula itself, then we would automatically obtain a Σ_2^P algorithm for deciding contextual subsumption. Likewise, a context could be verified to be a most likely context for a consequence in PSpace.

A different direction will be to extend our semantics to more expressive logics. In particular, we will include assertion and role axioms into our knowledge bases. Since many of our algorithms depend only on the existence of a reasoner for the logic, such extension should not be a problem. Our complexity results, on the other hand, would be affected by these changes. From the probabilistic side, we can also consider other probabilistic graphical models to encode the JPD of the contexts. Finally, we would like to consider problems that tighten the relationship between the probabilistic and the logical components. One of such problems would be to update the BN according to evidence attached to the TBox.

References

1. Baader, F.: Terminological cycles in a description logic with existential restrictions. In: Proc. 18th International Joint Conference on Artificial Intelligence (IJCAI-03). pp. 325–330. Morgan Kaufmann (2003)
2. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2nd edn. (2007)

3. Baader, F., Knechtel, M., Peñaloza, R.: A generic approach for large-scale ontological reasoning in the presence of access restrictions to the ontology's axioms. In: Proc. 8th International Semantic Web Conference (ISWC 2009). LNCS, vol. 5823, pp. 49–64. Springer (2009)
4. Baader, F., Knechtel, M., Peñaloza, R.: Context-dependent views to axioms and consequences of semantic web ontologies. J. of Web Semantics 12–13, 22–40 (2012)
5. Baader, F., Peñaloza, R.: Automata-based axiom pinpointing. J. of Automated Reasoning 45(2), 91–129 (August 2010)
6. Baader, F., Peñaloza, R.: Axiom pinpointing in general tableaux. J. of Logic and Computation 20(1), 5–34 (2010)
7. Baader, F., Peñaloza, R., Suntisrivaraporn, B.: Pinpointing in the description logic $\mathcal{EL}^+$. In: Proc. 30th German Conference on Artificial Intelligence (KI2007). LNAI, vol. 4667, pp. 52–67. Springer, Osnabrück, Germany (2007)
8. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In: Proc. 16th European Conference on Artificial Intelligence (ECAI-2004). pp. 298–302. IOS Press (2004)
9. Ceylan, İ.İ.: Context-Sensitive Bayesian Description Logics. Master's thesis, Dresden University of Technology, Germany (2013)
10. Cook, S.A.: The complexity of theorem-proving procedures. In: Proc. Third Annual ACM Symposium on Theory of Computing. pp. 151–158. STOC '71, ACM, New York, NY, USA (1971), http://doi.acm.org/10.1145/800157.805047
11. Cooper, G.F.: The computational complexity of probabilistic inference using bayesian belief networks (research note). Artif. Intel. 42(2-3), 393–405 (Mar 1990)
12. da Costa, P.C.G., Laskey, K.B., Laskey, K.J.: Pr-owl: A bayesian ontology language for the semantic web. In: Uncertainty Reasoning for the Semantic Web I, URSW 2005-2007. LNCS, vol. 5327, pp. 88–107. Springer (2008)
13. d'Amato, C., Fanizzi, N., Lukasiewicz, T.: Tractable reasoning with bayesian description logics. In: Proc. Second International Conference on Scalable Uncertainty Management (SUM 2008). LNCS, vol. 5291, pp. 146–159. Springer (2008)
14. Darwiche, A.: Modeling and Reasoning with Bayesian Networks. Cambridge University Press (2009)
15. Downey, R.G., Fellows, M.: Parameterized Complexity. Monographs in Computer Science, Springer (1999)
16. Koller, D., Levy, A.Y., Pfeffer, A.: P-classic: A tractable probablistic description logic. In: Proc. 14th National Conference on Artificial Intelligence (AAAI-97). pp. 390–397. AAAI Press (1997)
17. Lukasiewicz, T., Straccia, U.: Managing uncertainty and vagueness in description logics for the semantic web. J. of Web Semantics 6(4), 291–308 (2008)
18. Riguzzi, F., Bellodi, E., Lamma, E., Zese, R.: Epistemic and statistical probabilistic ontologies. In: Proc. 8th Int. Workshop on Uncertainty Reasoning for the Semantic Web (URSW-12). vol. 900, pp. 3–14. CEUR-WS (2012)
19. Roth, D.: On the hardness of approximate reasoning. Artif. Intel. 82(1-2), 273–302 (1996)