

# The Bayesian Ontology Language $\mathcal{BEL}$

İsmail İlkan Ceylan · Rafael Peñaloza

Received: date / Accepted: date

**Abstract** We introduce the new probabilistic description logic (DL)  $\mathcal{BEL}$ , which extends the light-weight DL  $\mathcal{EL}$  with the possibility of expressing uncertainty about the validity of some knowledge. Contrary to other probabilistic DLs,  $\mathcal{BEL}$  is designed to represent classical knowledge that depends on an uncertain context; that is, some of the knowledge may hold or not depending on the current situation. The probability distribution of these contexts is expressed by a Bayesian network (BN).

We study different reasoning problems in  $\mathcal{BEL}$ , providing tight complexity bounds for all of them. One particularly interesting property of our framework is that reasoning can be decoupled between the logical (i.e.,  $\mathcal{EL}$ ), and the probabilistic (i.e., the BN) components. We later generalize all the notions presented to introduce Bayesian extensions of arbitrary ontology languages. Using the decoupling property, we are able to provide tight complexity bounds for reasoning in the Bayesian extensions of many other DLs. We provide a detailed analysis of our formalism w.r.t. the assumptions made and compare it with the existing approaches.

**Keywords** Description Logics · Bayesian Networks · Probabilistic Reasoning · Knowledge Representation and Reasoning

## 1 Introduction

Description Logics (DLs) [7] are a family of knowledge representation formalisms that have been successfully used for representing and reasoning with the knowledge of various application domains. One prominent member of this family is the light-weight DL  $\mathcal{EL}$  [5].  $\mathcal{EL}$  is a very inexpressive DL, incapable of expressing e.g.

---

İ. İ. Ceylan  
Institute for Theoretical Computer Science, Dresden University of Technology,  
E-mail: ceylan@tcs.inf.tu-dresden.de

R. Peñaloza  
KRDB Research Centre Free University of Bozen-Bolzano, Italy  
E-mail: rafael.penaloza@unibz.it

negations or disjunctions of concepts. Despite this limitation, many large knowledge domains have been modelled using slight extensions of  $\mathcal{EL}$ . This logic has been particularly successful in the representation of bio-medical domains. One of the fundamental features of  $\mathcal{EL}$  and some of its extensions is that they allow for polynomial-time standard reasoning. Hence, it remains feasible to reason with huge knowledge bases.

In their classical form, DLs are not suited for handling any kind of uncertainty that may arise in the application domain. This is a large limitation since, in particular in the bio-medical domains, most of the knowledge has some level of uncertainty attached to it. To alleviate this issue, many probabilistic extensions of DLs have been proposed; see [31] for a thorough survey. What differentiates all these formalisms are the assumptions made in the type of probabilities (e.g., subjective vs. statistical), the independence assumptions, and the way in which uncertainty is modeled (e.g., as a concept constructor, or in the axioms).

An important issue when modelling uncertain knowledge is how to represent the probabilistic dependencies between the different elements of the knowledge base. Bayesian networks (BNs) [23] are probabilistic models that use a graphical structure to express conditional independence assumptions between the variables of the network. Over the years, BNs have been used to model probabilistic knowledge in many domains. In particular, they have been used in several biological applications. See [26, 40] for just two of the many instances that can be found in the literature.

We propose the new probabilistic DL  $\mathcal{BEL}$ , which extends  $\mathcal{EL}$  by expressing the uncertainty of the different axioms with the help of a BN. The main assumption in this logic is that the knowledge base contains information that is certain, but dependent on an uncertain context. For example, in a biological application, we have the knowledge that when a cell enters mitosis, then its chromosomes are aligned, and a sequence of other processes is activated. However, our knowledge of whether the cell is in mitosis or not is uncertain, and dependent of other factors. To model this knowledge, we associate each axiom of the knowledge base to a *context* (essentially, a propositional conjunctive clause) that expresses when this axiom is required to hold. The joint probability distribution of these contexts is expressed with the help of a BN. Dually, one can think of  $\mathcal{BEL}$  as a generalization of BNs. Under this view, every valuation of the variables in the BN corresponds to an  $\mathcal{EL}$  KB, rather than just a propositional world. The inference problem in this case extends from asking the probability of a literal to hold, to asking the probability of an implicit consequence (of the different  $\mathcal{EL}$  knowledge bases) to be entailed. It is often useful to consider this view as a compact way to represent several different classical KBs (defined by the contexts), which have different probabilities of being correct.

In this paper, we study the complexity of standard reasoning in  $\mathcal{BEL}$ . We show first that every  $\mathcal{BEL}$  knowledge base is consistent. Thus, we focus later only on the problem of deciding subsumption: whether a concept must be interpreted as a subclass of another one. We start by studying the problem of deciding whether a subsumption relation is guaranteed to hold in a specific context. Afterwards, we take into account the probabilities expressed by the BN and study the problem of finding the probability of a subsumption relation to hold, and two special cases where we are only interested in knowing whether this probability is positive, or exactly 1. We show that, in this case, reasoning can be decoupled between the

logical (i.e.,  $\mathcal{EL}$ ) component and the probabilistic structure (i.e., the BN). We also consider the dual problem of finding, given a subsumption relation, the most likely context in which it is guaranteed to hold.

We obtain tight complexity bounds for all these reasoning problems. Our complexity analysis is supported by the novel *proof structure*, which provides additional insights on the complexity of the reasoning problems. As a rule of thumb, the complexity of each of these problems is usually the same as the maximal complexity of reasoning in the BN and in  $\mathcal{EL}$  separately. This behaviour is consistent with the decoupling of components mentioned before. In particular, we exploit the polynomial-time reasoning methods for  $\mathcal{EL}$  to reduce the  $\mathcal{BEL}$  reasoning problems to standard inferences in an extended BN.

Later in the paper we generalize the notions introduced for  $\mathcal{BEL}$  to obtain Bayesian extensions of any arbitrary ontology language. Using the decoupling between the logical and the probabilistic part, we describe black-box algorithms for reasoning in these languages. These algorithms make repeated calls to a standard logical reasoner and to a BN inference engine. Using this idea, we also obtain tight complexity bounds for many expressive DLs, and good upper bounds for others.

This paper collects, extends, and improves on the results previously published in two international conferences and a workshop [14–16]; see also [13]. In particular, in this paper we show how to handle also conditional probabilistic inferences, as opposed to the material implication used in previous work; we provide a simpler and smaller construction of the proof structure; and provide full proofs for our results.

## 2 Preliminaries

We start by recalling the individual components of our formalism; namely the description logic  $\mathcal{EL}$  and Bayesian networks, followed by a brief overview on some complexity classes relevant to our analysis.

### 2.1 The Description Logic $\mathcal{EL}$

$\mathcal{EL}$  is a light-weight description logic (DL) that has been successfully applied for modelling large application domains, specially in the bio-medical sciences. One of its attracting features is that it allows for polynomial time reasoning [5, 12]. As is the case with all DLs, its main building blocks are *concepts* (corresponding to unary predicates of first-order logic) and *roles* (binary predicates). Starting from two disjoint sets  $\mathbf{N}_C$  and  $\mathbf{N}_R$  of *concept names* and *role names*, respectively,  $\mathcal{EL}$  concepts are built through the syntax rule  $C ::= A \mid \top \mid C \sqcap C \mid \exists r.C$ , where  $A \in \mathbf{N}_C$  and  $r \in \mathbf{N}_R$ .

The semantics of this logic is defined through interpretations. An *interpretation* is a pair  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$  where  $\Delta^{\mathcal{I}}$  is a non-empty set, called the *domain*, and  $\cdot^{\mathcal{I}}$  is the *interpretation function* that maps every  $A \in \mathbf{N}_C$  to a set  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$  and every role name  $r$  to a binary relation  $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ . The interpretation function  $\cdot^{\mathcal{I}}$  is extended to  $\mathcal{EL}$  concepts by defining  $\top^{\mathcal{I}} := \Delta^{\mathcal{I}}$ ,  $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$ , and  $(\exists r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \exists e \in \Delta^{\mathcal{I}} : (d, e) \in r^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}}\}$ .

**Table 1** The  $\mathcal{EL}$  completion rules.

$\mapsto$	Premises (S)	Result ( $\alpha$ )
1	$A \sqsubseteq B_1, B_1 \sqsubseteq B$	$A \sqsubseteq B$
2	$A \sqsubseteq A_1, A \sqsubseteq A_2, A_1 \sqcap A_2 \sqsubseteq B$	$A \sqsubseteq B$
3	$A \sqsubseteq A_1, A_1 \sqsubseteq \exists r.B$	$A \sqsubseteq \exists r.B$
4	$A \sqsubseteq \exists r.A_1, A_1 \sqsubseteq B_1, \exists r.B_1 \sqsubseteq B$	$A \sqsubseteq B$

The knowledge of an application domain is represented through a set of axioms that restrict the possible interpretation of the concepts. A *general concept inclusion (GCI)* is an expression of the form  $C \sqsubseteq D$ , where  $C, D$  are concepts. A *TBox*  $\mathcal{T}$  is a finite set of GCIs. The *signature* of  $\mathcal{T}$  ( $\text{sig}(\mathcal{T})$ ) is the set of concept and role names appearing in  $\mathcal{T}$ . The interpretation  $\mathcal{I}$  satisfies the GCI  $C \sqsubseteq D$  iff  $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ ; it is a *model* of the TBox  $\mathcal{T}$  iff it satisfies all the GCIs in  $\mathcal{T}$ . The main reasoning task in  $\mathcal{EL}$  is deciding the subsumption relations between concepts. A concept  $C$  is *subsumed by*  $D$  w.r.t. the TBox  $\mathcal{T}$  ( $\mathcal{T} \models C \sqsubseteq D$ ) iff  $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$  for all models  $\mathcal{I}$  of  $\mathcal{T}$ . In particular, one can consider without loss of generality only the problem of *atomic subsumption*, where the subsumption relation is to be decided between concept names (see Lemma 16).

It is well known that subsumption in  $\mathcal{EL}$  can be decided in polynomial time through a completion algorithm [5]. This algorithm first transforms the TBox into an equivalent one (w.r.t. atomic subsumption) in *normal form*; i.e., having only axioms of the form

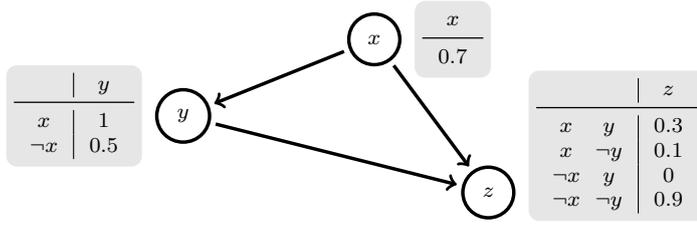
$$A \sqsubseteq B, \quad A_1 \sqcap A_2 \sqsubseteq B, \quad A \sqsubseteq \exists r.B, \quad \text{or} \quad \exists r.A \sqsubseteq B, \quad (1)$$

where  $A, A_1, A_2$ , and  $B$  are concept names or  $\top$ . Every  $\mathcal{EL}$  TBox  $\mathcal{T}$  can be transformed into an equivalent one in normal form, whose size is linear on the size of  $\mathcal{T}$  [5, 12] (see also Table 2).

After this normalization step, the completion algorithm deduces the relevant subsumption relations entailed by the TBox through an exhaustive application of the completion rules described in Table 1. Each rule ( $\mapsto$ ) maps a set of premises  $S$  to its consequence  $\alpha$ . The algorithm uses these rules to extend a set  $\text{comp}(\mathcal{T})$ , initialized as the input TBox  $\mathcal{T}$  together with some tautological information. Let  $\text{ini}(\mathcal{T}) := \mathcal{T} \cup \{A \sqsubseteq A, A \sqsubseteq \top \mid A \in \text{sig}(\mathcal{T}) \cap \mathbf{N}_{\mathcal{C}}\}$ . Starting from  $\text{comp}(\mathcal{T}) := \text{ini}(\mathcal{T})$ , a completion rule is *applicable* to  $\text{comp}(\mathcal{T})$  if  $S \subseteq \text{comp}(\mathcal{T})$  but  $\alpha \notin \text{comp}(\mathcal{T})$ . In that case, its application adds the consequence  $\alpha$  to the TBox  $\text{comp}(\mathcal{T})$ . When no rules are applicable, the resulting TBox contains all the atomic subsumptions that can be derived from the original TBox. More formally, we have that  $\mathcal{T} \models A \sqsubseteq B$  iff  $A \sqsubseteq B \in \text{comp}(\mathcal{T})$ . The completion rules introduce only GCIs in normal form, and do not change the signature. Hence, the algorithm stops after at most  $|\text{sig}(\mathcal{T})|^3$  rule applications. For more details, we refer the interested reader to [5].

## 2.2 Bayesian Networks

We will later extend  $\mathcal{EL}$  to express and handle uncertain knowledge in the form of probabilistic axioms. To encode the conditional probability of the knowledge, we will use Bayesian networks [37]. Formally, a *Bayesian network* (BN) is a pair



**Fig. 1** The BN  $\mathcal{B}_{\text{exa}}$  over  $V_{\text{exa}} = \{x, y, z\}$

$\mathcal{B} = (G, \Phi)$ , where  $G = (V, E)$  is a finite directed acyclic graph (DAG) whose nodes represent Boolean random variables,<sup>1</sup> and  $\Phi$  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 all its non-descendants given its parents. Thus, every BN  $\mathcal{B}$  defines the unique joint probability distribution (JPD) over the set of random variables  $V$  given by the chain rule

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

A very simple BN over the variables  $V_{\text{exa}} = \{x, y, z\}$  is shown in Figure 1. In this network, the parents of  $z$  are  $\pi(z) = \{x, y\}$ . Using the chain rule, we can derive e.g.  $P_{\mathcal{B}_{\text{exa}}}(x, \neg y, z) = P_{\mathcal{B}_{\text{exa}}}(z \mid x, \neg y) \cdot P_{\mathcal{B}_{\text{exa}}}(\neg y \mid x) \cdot P_{\mathcal{B}_{\text{exa}}}(x) = 0.1 \cdot 0 \cdot 0.7 = 0$ .

The main inference problems associated to a BN are to compute the probability of a partial observation (PR), the maximum a posteriori probability given some evidence (MAP) and the most probable explanation for a given observation (MPE). In this paper, we are interested only on their decision variants, which are introduced next.

**Definition 1 (inferences)** 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. Let  $\mathcal{B}$  be a BN over  $V$ . We define the following decision problems.

D-PR Given a  $V$ -context  $\kappa$  and a  $p > 0$ , is  $P_{\mathcal{B}}(\kappa) > p$ ?

D-MPE Given a  $V$ -context  $\kappa$  and a  $p > 0$ , is there a valuation  $\mathcal{W}$  of the variables in  $V$  such that  $\mathcal{W}$  extends  $\kappa$  and  $P_{\mathcal{B}}(\mathcal{W}) > p$ ?<sup>2</sup>

D-MAP Given a  $V$ -context  $\kappa$ ,  $p > 0$  and a set  $W \subseteq V$  is there a valuation  $\mathcal{W}$  of the variables in  $W$  such that  $P_{\mathcal{B}}(\mathcal{W} \cup \kappa) > p$ ?

All these decision problems are NP-hard, and in PSPACE. To provide more precise complexity bounds for these problems, we first introduce some basic probabilistic complexity classes.

<sup>1</sup> In their general form, BNs allow for arbitrary discrete random variables. We restrict w.l.o.g. to Boolean variables for ease of presentation.

<sup>2</sup> We will often see valuations as contexts containing one literal for each Boolean variable.

### 2.3 Probabilistic Complexity

Standard complexity classes do not suffice to obtain a fine-grained complexity analysis of probabilistic decision problems as the ones presented above. This limitation has motivated the study of probabilistic complexity classes, of which the most basic representative is the class PP [27]. Briefly, PP contains all languages that can be recognized by a polynomial-time bounded non-deterministic Turing machine that accepts an input iff more than half of the computation paths are accepting.

It is easily seen that PP contains NP and is contained in PSPACE. From the latter it immediately follows that  $\text{NP}^{\text{PP}} \subseteq \text{NP}^{\text{PSPACE}} = \text{PSPACE}$ . More interestingly, PP is closed under intersection, union, and complement [11]. In addition, PP is usually considered a hard class, since  $\text{P}^{\text{PP}}$  contains the polynomial hierarchy [44].

Using these classes, we can characterise precisely the complexity of the decision problems introduced for BNs. Namely, D-PR is PP-complete [30], D-MPE is NP-complete [43], and D-MAP is  $\text{NP}^{\text{PP}}$ -complete [35].

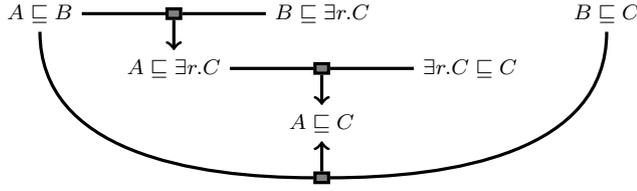
### 3 Proof Structures

As described in Section 2.1, the most typical reasoning problem in  $\mathcal{EL}$  is to decide subsumption between concepts w.r.t. background knowledge expressed via a TBox. For many applications, it is useful not only to decide whether a subsumption relation follows from a TBox  $\mathcal{T}$ , but also to find all the sub-TBoxes of  $\mathcal{T}$  that entail this relation. This task is known as *axiom pinpointing* in the literature [8]. To find these sub-TBoxes, we store all the possible traces of the completion rules using a directed hypergraph. A *directed hypergraph* is a tuple  $H = (V, E)$  where  $V$  is a non-empty set of *vertices* and  $E$  is a set of *directed hyper-edges* of the form  $e = (S, v)$  where  $S \subseteq V$  and  $v \in V$ . Given  $S \subseteq V$  and  $v \in V$ , a *path from  $S$  to  $v$  in  $H$  of length  $n$*  is a sequence  $(S_1, v_1), (S_2, v_2), \dots, (S_n, v_n)$  of hyper-edges where  $v_n = v$  and  $S_i \subseteq S \cup \{v_j \mid 0 < j < i\}$  for all  $i, 1 \leq i \leq n$ .

Given a TBox  $\mathcal{T}$  in normal form, we build the hypergraph  $H_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}})$ , where  $V_{\mathcal{T}} = \text{comp}(\mathcal{T})$  is the set of all GCIs that appear in  $\text{comp}(\mathcal{T})$  after the completion algorithm has terminated and  $E_{\mathcal{T}} = \{(S, \alpha) \mid S \mapsto \alpha, S \subseteq V_{\mathcal{T}}\}$ , with  $\mapsto$  the deduction relation defined in Table 1. We call this hypergraph the *proof structure* of  $\mathcal{T}$ . From the soundness and completeness of the completion algorithm, we get the following lemma.

**Lemma 2** *Let  $\mathcal{T}$  be a TBox in normal form,  $H_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}})$  its proof structure,  $\mathcal{O} \subseteq \mathcal{T}$ , and  $A_0 \sqsubseteq B_0 \in V_{\mathcal{T}}$ . Then, there is a path from  $\text{ini}(\mathcal{O})$  to  $A_0 \sqsubseteq B_0$  in  $H_{\mathcal{T}}$  iff  $\mathcal{O} \models A_0 \sqsubseteq B_0$ .*

The hypergraph  $H_{\mathcal{T}}$  can be seen as a compact representation of all the possible derivations of a consequence from the GCIs in  $\mathcal{T}$  [3, 8]. Traversing this hypergraph backwards from a GCI  $A_0 \sqsubseteq B_0$  entailed by  $\mathcal{T}$ , it is possible to construct all the proofs for  $\alpha$ ; hence the name “proof structure.” It is well known that to decide the existence of a path in a directed hypergraph  $\mathcal{G}$  it is sufficient to consider only paths whose length is at most the same as the number of nodes in  $\mathcal{G}$ . In our case, this means that we can focus on paths of length at most  $|V_{\mathcal{T}}| \leq |\text{sig}(\mathcal{T})|^3$ .



**Fig. 2** Proof structure of  $\mathcal{T}_{\text{exa}}$  from Example 3 (simplified).

*Example 3* Consider the TBox  $\mathcal{T}_{\text{exa}} := \{A \sqsubseteq B, B \sqsubseteq \exists r.C, \exists r.C \sqsubseteq C, B \sqsubseteq C\}$ , which is already in normal form. After the completion rules have been applied, we obtain  $\text{comp}(\mathcal{T}_{\text{exa}}) = \mathcal{T}_{\text{exa}} \cup \{A \sqsubseteq \exists r.C, A \sqsubseteq C\} \cup \{X \sqsubseteq X, X \sqsubseteq \top \mid X \in \{A, B, C\}\}$ . The proof structure  $H_{\mathcal{T}_{\text{exa}}}$  is depicted in Figure 2, where to improve readability all the tautological axioms have been removed. In particular, we can see that the consequence  $A \sqsubseteq C$  follows already from the sub-TBox  $\{A \sqsubseteq B, B \sqsubseteq C\}$ .

Clearly, the proof structure  $H_{\mathcal{T}}$  can be cyclic. To simplify the process of finding the causes of an atomic subsumption relation being entailed, we build an *unfolded* version of this hypergraph by making different copies of each node. In this case, nodes are pairs of axioms and labels, where the latter indicates the level to which the nodes belong in the hypergraph. Given a set of axioms  $S$ , and an index  $i \geq 0$ ,  $S^i := \{(\alpha, i) \mid \alpha \in S\}$  denotes the  $i$ -labeled set of GCIs in  $S$ . Let  $n := |V_{\mathcal{T}}|$ . We define the sets  $W_i, 0 \leq i \leq n$  inductively as follows.  $W_0 := \{(\alpha, 0) \mid \alpha \in \text{ini}(\mathcal{T})\}$ , and for all  $i, 0 \leq i < n$ ,

$$W_{i+1} := \{(\alpha, i+1) \mid S^i \sqsubseteq W_i, S \mapsto \alpha\} \cup \{(\alpha, i+1) \mid \alpha \in \text{ini}(\mathcal{T})\}.$$

In a nutshell, each  $W_i, i, 0 \leq i \leq n$ , contains all the GCIs that can be derived by at most  $i$  applications of the completion rules. The *unfolded proof structure* of  $\mathcal{T}$  is the hypergraph  $H_{\mathcal{T}}^u = (W_{\mathcal{T}}, F_{\mathcal{T}})$ , where  $W_{\mathcal{T}} := \bigcup_{i=0}^n W_i$  and  $F_{\mathcal{T}} := \bigcup_{i=1}^n F_i$ ,

$$F_{i+1} := \{(S^i, (\alpha, i+1)) \mid S^i \sqsubseteq W_i, S \mapsto \alpha\} \cup \{(\{(\alpha, i)\}, (\alpha, i+1)) \mid \alpha \in \text{ini}(\mathcal{T})\}.$$

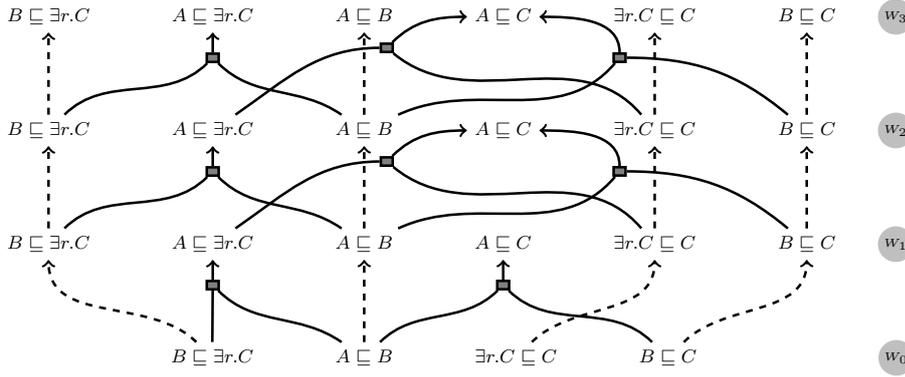
The following theorem is a simple consequence of our constructions and Lemma 2.

**Theorem 4** *Let  $\mathcal{T}$  be a TBox in normal form,  $H_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}})$  its proof structure with  $n = |V_{\mathcal{T}}|$ , and  $H_{\mathcal{T}}^u = (W_{\mathcal{T}}, F_{\mathcal{T}})$  the unfolded proof structure of  $\mathcal{T}$ . Then,*

1.  $\alpha \in V_{\mathcal{T}}$  iff  $(\alpha, n) \in W_{\mathcal{T}}$ ;
2.  $(S, \alpha) \in E_{\mathcal{T}}$  iff  $(S^{n-1}, (\alpha, n)) \in F_{\mathcal{T}}$ ; and
3. for all  $A, B \in \text{sig}(\mathcal{T}) \cap \mathbf{N}_{\mathcal{C}}$  and all  $\mathcal{O} \sqsubseteq \mathcal{T}$ , it holds that  $\mathcal{O} \models A \sqsubseteq B$  iff there is a path from  $\{(\alpha, 0) \mid \alpha \in \text{ini}(\mathcal{O})\}$  to  $(A \sqsubseteq B, n)$  in  $H_{\mathcal{T}}^u$ .

*Proof* The statements 1. and 2. are trivial; thus, we prove only the third claim. If there is a path from  $\{(\alpha, 0) \mid \alpha \in \text{ini}(\mathcal{O})\}$  to  $(A \sqsubseteq B, n)$  in  $H_{\mathcal{T}}^u$ , then by construction and the first points of the theorem, there is a path from  $\text{ini}(\mathcal{O})$  to  $A \sqsubseteq B$  in  $H_{\mathcal{T}}$ . By Lemma 2 it then follows that  $\mathcal{O} \models A \sqsubseteq B$ .

Conversely, if  $\mathcal{O} \models A \sqsubseteq B$ , then by Lemma 2 there is a path from  $\text{ini}(\mathcal{O})$  to  $A \sqsubseteq B$  in  $H_{\mathcal{T}}$ . Without loss of generality, we can assume that this path is of the form  $(S_1, \alpha_1), \dots, (S_k, \alpha_k)$  for  $k \leq n$ . A path from  $\{(\alpha, 0) \mid \alpha \in \text{ini}(\mathcal{O})\}$  to  $(A \sqsubseteq B, n)$  in  $H_{\mathcal{T}}^u$  can then be constructed through the sequence of hyperedges  $(S_i^j, (\alpha_i, j))$  for all  $1 \leq i < j \leq n$ .  $\square$



**Fig. 3** The first four levels ( $W_0$ – $W_3$ ) of  $H_{\mathcal{T}_{\text{exa}}}^u$ . To improve readability, the second component of the nodes is represented visually.

The unfolded proof structure of a TBox  $\mathcal{T}$  is guaranteed to contain the information of all possible causes for a consequence to follow from  $\mathcal{T}$ . Moreover, this hypergraph is acyclic, and has polynomially many nodes, on the size of  $\mathcal{T}$ , by construction. More precisely, the number of nodes of  $H_{\mathcal{T}}^u$  is bounded by  $|V_{\mathcal{T}}|^2$ , where  $|V_{\mathcal{T}}| \leq |\mathcal{T}|^3$ . Yet, this hypergraph may still contain many redundant nodes. Indeed, it can be the case that all the simple paths in  $H_{\mathcal{T}}$  starting from a subset of  $\mathcal{T}$  are of length  $k < n$ . In that case,  $W_i = W_{i+1}$  and  $F_i = F_{i+1}$  hold for all  $i \geq k$ , modulo the second component. For example, the first four levels of the unfolded proof structure for the TBox  $\mathcal{T}_{\text{exa}}$  from Example 3 are shown in Figure 3. As it can be seen, the levels  $W_2$  and  $W_3$  are identical; moreover, if we continued the unfolding of the proof structure, all successive levels will be the same. It can also be seen that in this case, all the axiomatic causes for a consequence can be read by paths of length at most 2. In general, it suffices to unfold the proof structure only up to the length of the longest simple path from  $\text{ini}(\mathcal{T})$  to any element in  $\text{comp}(\mathcal{T})$  in  $H_{\mathcal{T}}$ . For our purposes, we are only interested in knowing that the unfolded proof structure is of polynomial size; hence, we consider the full unfolded proof structure for the rest of this paper. It should be noted, however, that for efficiency reasons, a pruned version can be also used.

In the next section we introduce a probabilistic extension of  $\mathcal{EL}$  based on BNs. The construction of the unfolded proof structure will be helpful in to obtain additional insights and understanding the complexity of reasoning in this logic.

#### 4 The Probabilistic Ontology Language $\mathcal{BEL}$

$\mathcal{BEL}$  is an extension of  $\mathcal{EL}$  capable of expressing uncertain knowledge in the form of probabilistic axioms. As with classical DLs, the main building blocks in  $\mathcal{BEL}$  are *concepts*. Syntactically,  $\mathcal{BEL}$  concepts are constructed exactly as  $\mathcal{EL}$  concepts. The difference arises in the description of axioms, which are now associated to a set of literals from the BN.

**Definition 5 (KB)** 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  $\kappa$  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.

*Example 6* Extending the TBox from Example 3, consider the  $V_{\text{exa}}$ -TBox

$$\mathcal{T}'_{\text{exa}} := \{\langle A \sqsubseteq B : \{x\} \rangle, \langle B \sqsubseteq \exists r.C : \{y, z\} \rangle, \langle \exists r.C \sqsubseteq C : \{x, y\} \rangle, \langle B \sqsubseteq C : \{\neg z\} \rangle\}.$$

Then  $\mathcal{K}_{\text{exa}} = (\mathcal{B}_{\text{exa}}, \mathcal{T}'_{\text{exa}})$  is a  $\mathcal{BEL}$  KB.

The intuition behind the contexts associated to axioms is that a  $V$ -GCI is only required to hold whenever its context is satisfied. To formalize this intuition, we extend the notion of interpretations to evaluate also the Boolean variables appearing in the BN and in the contexts.

**Definition 7 (interpretation)** Let  $V$  be a finite set of Boolean variables. A  $V$ -interpretation is a triple  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$  where  $\Delta^{\mathcal{I}}$  is a non-empty set called the domain,  $\mathcal{V}^{\mathcal{I}} : V \rightarrow \{0, 1\}$  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 omit the prefix  $V$  and speak simply of e.g. a TBox, a KB, or an interpretation. The interpretation function  $\cdot^{\mathcal{I}}$  is extended to arbitrary concepts as done in the classical case. 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  $\kappa$ ,

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

where  $\min_{\ell \in \emptyset} \mathcal{V}^{\mathcal{I}}(\ell) := 1$  as usual. A context  $\kappa$  can be thought as the conjunction of the literals it contains; thus, it is evaluated to 1 iff each of its elements is so and 0 otherwise.

**Definition 8 (model)** 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 behind this semantics is that the  $V$ -GCI  $\langle C \sqsubseteq D : \kappa \rangle$  restricts the interpretations of  $C$  and  $D$ , but only when the context  $\kappa$  is satisfied. Thus, any interpretation that violates the context trivially satisfies the whole axiom. For example, consider the interpretation  $\mathcal{I}_0 = (\{d\}, \cdot^{\mathcal{I}_0}, \mathcal{V}_0)$ , where  $\mathcal{V}_0(\{x, \neg y, z\}) = 1$ ,  $A^{\mathcal{I}_0} = B^{\mathcal{I}_0} = \{d\}$ ,  $C^{\mathcal{I}_0} = \emptyset$ , and  $r^{\mathcal{I}_0} = \emptyset$ . Then  $\mathcal{I}_0$  is a model of  $\mathcal{T}'_{\text{exa}}$ ; in particular, it satisfies the GCI  $\langle B \sqsubseteq C : \{\neg z\} \rangle$  because  $\mathcal{V}_0(\{\neg z\}) = 0$ . However, this same interpretation is not a model of the GCI  $\langle B \sqsubseteq C : \{z\} \rangle$ .

The classical DL  $\mathcal{EL}$  can be seen as a special case of  $\mathcal{BEL}$  in which all GCIs are associated with the empty context; that is, are all 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  $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ , which corresponds to the classical semantics of  $\mathcal{EL}$  [12]. The  $V$ -TBox  $\mathcal{T}$  entails  $\langle C \sqsubseteq D : \emptyset \rangle$  ( $\mathcal{T} \models C \sqsubseteq D$ ) 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  $\mathcal{V}^{\mathcal{I}} = \mathcal{W}$ . For the rest of this paper, we will use the expression  $\langle C \sqsubseteq D \rangle$  to abbreviate the  $V$ -GCI  $\langle C \sqsubseteq D : \emptyset \rangle$ .

When reasoning in  $\mathcal{BEL}$ , it is sometimes useful to focus on the classical  $\mathcal{EL}$  TBox induced by a given valuation of the variables in  $V$ .

**Definition 9 (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

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

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 consider multiple-world semantics. In a nutshell, every  $V$ -interpretation describes a possible world; by assigning a probability distribution over these interpretations, we describe the required probabilities, which should be consistent with the BN.

**Definition 10 (probabilistic model)** A *probabilistic interpretation* is a pair of the form  $\mathcal{P} = (\mathcal{J}, P_{\mathcal{J}})$ , where  $\mathcal{J}$  is a set of  $V$ -interpretations and  $P_{\mathcal{J}}$  is a probability distribution over  $\mathcal{J}$  such that  $P_{\mathcal{J}}(\mathcal{I}) > 0$  only for finitely many interpretations  $\mathcal{I} \in \mathcal{J}$ . This probabilistic interpretation is a *model* of the TBox  $\mathcal{T}$  if every  $\mathcal{I} \in \mathcal{J}$  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

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

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 of the variables in  $V$ . Hence, the probability of a context  $\kappa$  is the sum of the probabilities of all (full) valuations that extend  $\kappa$ .

**Theorem 11** Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be a KB, and  $\kappa$  a context. For every model  $\mathcal{P}$  of  $\mathcal{K}$  it holds that

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

*Proof* By definition, it holds that

$$P_{\mathcal{B}}(\kappa) = \sum_{\mathcal{W}(\kappa) = 1} P_{\mathcal{B}}(\mathcal{W}) = \sum_{\mathcal{W}(\kappa) = 1} \sum_{\mathcal{I} \in \mathcal{J}, \mathcal{V}^{\mathcal{I}} = \mathcal{W}} P_{\mathcal{J}}(\mathcal{I}) = \sum_{\mathcal{I} \in \mathcal{J}, \mathcal{V}^{\mathcal{I}}(\kappa) = 1} P_{\mathcal{J}}(\mathcal{I}). \quad \square$$

In order to reason w.r.t.  $\mathcal{BEL}$  KBs, it is sometimes useful 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 12 (pithy)** The probabilistic interpretation  $\mathcal{P} = (\mathcal{J}, P_{\mathcal{J}})$  is called *pithy* if for every valuation  $\mathcal{W}$  of the variables in  $V$  there exists at most one  $V$ -interpretation  $\mathcal{I} = (\Delta^{\mathcal{I}}, \mathcal{I}, \mathcal{V}^{\mathcal{I}}) \in \mathcal{J}$  such that  $\mathcal{V}^{\mathcal{I}} = \mathcal{W}$ .

In the following section we introduce classical and probabilistic reasoning problems for the DL  $\mathcal{BEL}$ , and analyse their complexity w.r.t. diverse measures.

## 5 Reasoning in $\mathcal{BEL}$

In the previous section we 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 reasoning problem in any DL is to decide whether a knowledge base is consistent; that is, whether it has a model. It turns out that, as for classical  $\mathcal{EL}$ , this problem is trivial in  $\mathcal{BEL}$ .

**Theorem 13** *Every  $\mathcal{BEL}$  KB is consistent.*

*Proof* Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be an arbitrary  $\mathcal{BEL}$  KB. Let  $\Delta^{\mathcal{I}} = \{a\}$  and  $\cdot^{\mathcal{I}}$  be the interpretation function with  $A^{\mathcal{I}} = \{a\}$  and  $r^{\mathcal{I}} = \{(a, a)\}$  for all  $A \in \mathbf{N}_{\mathcal{C}}$  and  $r \in \mathbf{N}_{\mathcal{R}}$ . For every valuation  $\mathcal{W}$ , define the  $V$ -interpretation  $\mathcal{I}_{\mathcal{W}} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{W})$ . It follows that the probabilistic interpretation  $\mathcal{P} = (\mathfrak{J}, P_{\mathfrak{J}})$  where  $\mathfrak{J} = \{\mathcal{I}_{\mathcal{W}} \mid \mathcal{W} \text{ is a valuation}\}$  and  $P_{\mathfrak{J}}(\mathcal{I}_{\mathcal{W}}) = P_{\mathcal{B}}(\mathcal{W})$  is a (pithy) model of  $\mathcal{K}$ .  $\square$

As we have seen in Section 2.1, the main reasoning problem in  $\mathcal{EL}$  is the subsumption between concepts. We generalize this problem to consider also the contexts attached to the GCIs, and probabilities provided by the BN.

**Definition 14 (subsumption)** Let  $C, D$  be two  $\mathcal{BEL}$  concepts,  $\kappa$  a  $V$ -context, and  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  a  $\mathcal{BEL}$  KB.  $C$  is *contextually subsumed by  $D$  in  $\kappa$  w.r.t.  $\mathcal{K}$* , denoted as  $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle$ , if every  $V$ -model of  $\mathcal{T}$  is also a model of the  $V$ -GCI  $\langle C \sqsubseteq D : \kappa \rangle$ .

Given a probabilistic interpretation  $\mathcal{P} = (\mathfrak{J}, P_{\mathfrak{J}})$ , the probability of a consequence is defined by  $P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa \rangle) := \sum_{\mathcal{I} \in \mathfrak{J}, \mathcal{I} \models \langle C \sqsubseteq D : \kappa \rangle} P_{\mathfrak{J}}(\mathcal{I})$ . The *probability of the contextual subsumption relation  $\langle C \sqsubseteq D : \kappa \rangle$  w.r.t.  $\mathcal{K}$*  is

$$P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle) := \inf_{\mathcal{P} \models \mathcal{K}} P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa \rangle). \quad (2)$$

We say that  $C$  is *positively subsumed* by  $D$  in context  $\kappa$  if  $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle) > 0$ , and  $C$  is  *$p$ -subsumed* by  $D$  in context  $\kappa$ , 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.

As before, if the context is empty we will omit it and consider e.g.,  $\langle C \sqsubseteq_{\mathcal{K}} D \rangle$  or  $P(\langle C \sqsubseteq_{\mathcal{K}} D \rangle)$ . We refer to this case as probabilistic subsumption, and to the general case as probabilistic contextual subsumption.

As a simple consequence of the proof of Theorem 13, we have that for every  $\mathcal{BEL}$  KB  $\mathcal{K}$ , every context  $\kappa$ , and concepts  $C, D$ , there is a model  $\mathcal{P}$  of  $\mathcal{K}$  such that  $P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa \rangle) = 1$ . In particular, this means that the reasoning problem obtained by replacing the infimum in Equation (2) with a supremum is trivial in  $\mathcal{BEL}$ . Notice moreover that if  $C$  is subsumed by  $D$  in  $\kappa$  w.r.t. the KB  $\mathcal{K}$ , then for every probabilistic model  $\mathcal{P}$  of  $\mathcal{K}$  we have that  $P(\langle C \sqsubseteq_{\mathcal{P}} D : \kappa \rangle) = 1$ ; and thus  $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle) = 1$ . The converse, however, may not hold: the subsumption relation might be violated in some  $V$ -interpretations that have probability zero.

*Example 15* Consider again the KB  $\mathcal{K}_{\text{exa}}$  from Example 6. For any two concept names  $E, F \in \mathbf{N}_{\mathcal{C}} \setminus \{A, B, C\}$  it holds that  $P(\langle E \sqsubseteq_{\mathcal{K}_{\text{exa}}} F : \{x, \neg y\} \rangle) = 1$  since the GCI  $\langle E \sqsubseteq_{\mathcal{K}_{\text{exa}}} F : \{x, \neg y\} \rangle$  can only be violated in  $V$ -interpretations that have probability 0. However, in general the consequence  $\langle E \sqsubseteq_{\mathcal{K}_{\text{exa}}} F : \{x, \neg y\} \rangle$  does not hold.

**Table 2**  $\mathcal{BEL}$  normalization rules, where  $A \in \mathbf{N}_C \cup \{\top\}$ ,  $C, D \notin \mathbf{N}_C \cup \{\top\}$  and  $X$  is a new concept name.

<b>NF1</b>	$\langle A \sqcap C \sqsubseteq E : \kappa \rangle$	$\longrightarrow$	$\{\langle C \sqsubseteq X : \kappa \rangle, \langle A \sqcap X \sqsubseteq E : \kappa \rangle\}$
<b>NF2</b>	$\langle \exists r.C \sqsubseteq E : \kappa \rangle$	$\longrightarrow$	$\{\langle C \sqsubseteq X : \kappa \rangle, \langle \exists r.X \sqsubseteq E : \kappa \rangle\}$
<b>NF3</b>	$\langle C \sqsubseteq D : \kappa \rangle$	$\longrightarrow$	$\{\langle C \sqsubseteq X : \kappa \rangle, \langle X \sqsubseteq D : \kappa \rangle\}$
<b>NF4</b>	$\langle A \sqsubseteq E \sqcap F : \kappa \rangle$	$\longrightarrow$	$\{\langle A \sqsubseteq E : \kappa \rangle, \langle A \sqsubseteq F : \kappa \rangle\}$
<b>NF5</b>	$\langle A \sqsubseteq \exists r.C : \kappa \rangle$	$\longrightarrow$	$\{\langle A \sqsubseteq \exists r.X : \kappa \rangle, \langle X \sqsubseteq C : \kappa \rangle\}$

For the rest of this paper, we consider only *atomic* subsumption problems; that is, cases where we want to decide, or compute the probability of a contextual subsumption between two concept names. As we show next, this restriction is made without any loss of generality.

**Lemma 16** *Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be a  $\mathcal{BEL}$  KB,  $C, D$  two  $\mathcal{BEL}$  concepts,  $\kappa$  a context, and  $A_0, B_0$  two concept names not appearing in  $\mathcal{T} \cup \{\langle C \sqsubseteq D \rangle\}$ . Consider the KB  $\mathcal{K}' = (\mathcal{B}, \mathcal{T} \cup \{\langle A_0 \sqsubseteq C \rangle, \langle D \sqsubseteq B_0 \rangle\})$ . Then,*

1.  $\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle$  iff  $\langle A_0 \sqsubseteq_{\mathcal{K}'} B_0 : \kappa \rangle$ , and
2.  $P(\langle C \sqsubseteq_{\mathcal{K}} D : \kappa \rangle) = P(\langle A_0 \sqsubseteq_{\mathcal{K}'} B_0 : \kappa \rangle)$ .

*Proof* The result follows from the fact that every model of  $\mathcal{K}'$  is also a model of  $\mathcal{K}$  and, conversely, every model  $\mathcal{I}$  of  $\mathcal{K}$  can be extended to a model of  $\mathcal{K}'$  by setting  $A_0^{\mathcal{I}} = C^{\mathcal{I}}$  and  $B_0^{\mathcal{I}} = D^{\mathcal{I}}$ . The full details of this construction can be developed analogously to [5].  $\square$

Moreover, we can also assume w.l.o.g. that the  $V$ -TBox is in normal form; that is, all the  $V$ -GCIs in  $\mathcal{T}$  are of the form  $\langle C \sqsubseteq D : \kappa \rangle$ , where  $C \sqsubseteq D$  is an  $\mathcal{EL}$  GCI in normal form (see Expression (1)). For every  $V$ -TBox  $\mathcal{T}$ , it is possible to build in linear time a new  $V$ -TBox in normal form that preserves all the subsumption relations between concept names appearing in  $\mathcal{T}$ . More formally, let  $\mathcal{T}$  be a  $V$ -TBox, and  $\mathcal{T}'$  be the TBox obtained after an exhaustive application of the normalization rules from Table 2. Each normalization rule takes a  $V$ -GCI that is not in normal form and replaces it by two simpler  $V$ -GCIs. Notice that the normalization rules never change the context in which the axioms hold. It is easy to see that the resulting TBox  $\mathcal{T}'$  is in normal form. Let now  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  and  $\mathcal{K}' = (\mathcal{B}, \mathcal{T}')$ , where  $\mathcal{B}$  is an arbitrary BN over  $V$ . Then, for every two concept names  $A, B \in \text{sig}(\mathcal{T})$  and every context  $\kappa$ , it holds that  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  iff  $\langle A \sqsubseteq_{\mathcal{K}'} B : \kappa \rangle$ . The full proof of this claim is equivalent to the one presented in [5] for  $\mathcal{EL}$ . Hence, we leave it as an exercise to the interested reader.

We now analyse the reasoning problems defined in this section in detail, starting with contextual subsumption, followed by the case where the probabilistic information is also relevant. Afterwards, we consider other non-standard inferences that can be made over Bayesian KBs.

## 5.1 Contextual Subsumption

In this section we consider the problem of deciding whether a contextual subsumption relation follows from all models of the KB in a classical sense; that

is, whether  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  holds. Contrary to classical  $\mathcal{EL}$ , subsumption in  $\mathcal{BEL}$  is already intractable, even if we consider only the empty context.

**Theorem 17** *Let  $\mathcal{K}$  be a KB and  $A, B \in \mathbf{N}_{\mathcal{C}}$ . Deciding  $\langle A \sqsubseteq_{\mathcal{K}} B \rangle$  is coNP-hard.*

*Proof* We present a reduction from validity of DNF formulas, which is known to be coNP-hard [19]. Let  $\phi = \sigma_1 \vee \dots \vee \sigma_n$  be a DNF formula where each  $\sigma_i$  is a conjunctive clause and let  $V$  be the set of all variables appearing in  $\phi$ . 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  $\sigma = \ell_1 \wedge \dots \wedge \ell_m$  define the TBox  $\mathcal{T}_\sigma := \{\langle B_{\ell_1} \sqcap \dots \sqcap B_{\ell_m} \sqsubseteq C \rangle\}$ . 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  $\phi$  is valid iff  $\langle A \sqsubseteq_{\mathcal{K}} C \rangle$ .  $\square$

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  $\kappa$  not because the axioms explicitly labeled with  $\kappa$  entail the consequence, but rather because any valuation satisfying  $\kappa$  will yield it. That is the main idea in the proof of Theorem 17; the axioms that follow directly from the empty context never entail the subsumption  $A \sqsubseteq C$ , but if  $\phi$  is valid, then this subsumption follows from all valuations. Following this intuition, we can characterize contextual subsumption in terms of classical subsumption.

**Lemma 18** *Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be a KB. Then  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  iff for every valuation  $\mathcal{W}$  with  $\mathcal{W}(\kappa) = 1$ , it holds that  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ .*

*Proof* Suppose that  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$ . Let  $\mathcal{W}$  be a valuation such that  $\mathcal{W}(\kappa) = 1$ , and  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$  be a (classical) model of  $\mathcal{T}_{\mathcal{W}}$ . By definition,  $\mathcal{J} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{W})$  is a  $V$ -model of  $\mathcal{T}$  and hence also of  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$ . In particular,  $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ . Since  $\mathcal{I}$  was an arbitrary model of  $\mathcal{T}_{\mathcal{W}}$ , this implies that  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ .

Conversely, let  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$  be a  $V$ -model of  $\mathcal{T}$ . If  $\mathcal{V}^{\mathcal{I}}(\kappa) = 0$ , then by definition  $\mathcal{I}$  is a model of  $\langle A \sqsubseteq B : \kappa \rangle$ . Otherwise, by assumption we know that  $\mathcal{T}_{\mathcal{V}^{\mathcal{I}}} \models A \sqsubseteq B$ ; i.e.,  $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ . Hence,  $\mathcal{I}$  is also a model of  $\langle A \sqsubseteq B : \kappa \rangle$ .  $\square$

This lemma yields a tight upper bound for the complexity of contextual subsumption. If the subsumption does not hold, then we can guess a valuation  $\mathcal{W}$  and verify in polynomial time that  $\mathcal{W}(\kappa) = 1$  and  $\mathcal{T}_{\mathcal{W}} \not\models A \sqsubseteq B$ .

**Corollary 19** *Contextual subsumption is coNP-complete. The lower bound holds even if  $\kappa = \emptyset$ .*

This result provides a tight complexity bound for the problem of contextual subsumption w.r.t.  $\mathcal{BEL}$  KBs. Notice moreover that Lemma 18 shows that the problem is fixed-parameter tractable over the parameter  $|V|$ .<sup>3</sup> However, the non-deterministic algorithm suggested by this lemma is not practical, as it requires to perform reasoning on all valuations that satisfy the context  $\kappa$ .

We now propose a different approach that is based on techniques developed for axiom-pinpointing [4], access control [9], and context-based reasoning [10]. Our goal is to find, for a given subsumption relation  $A \sqsubseteq B$ , the set of all valuations

<sup>3</sup> Recall that a problem is fixed-parameter tractable if it can be solved in polynomial time, assuming that the parameter has a fixed (i.e., constant) size [25].

**Table 3** The  $\mathcal{B}\mathcal{E}\mathcal{L}$  completion rules.

	Premises (S)	Result ( $\text{lab}(\alpha)$ )
1	$A \sqsubseteq B_1^{\phi_1}, B_1 \sqsubseteq B^{\phi_2}$	$\text{lab}(A \sqsubseteq B) := \text{lab}(A \sqsubseteq B) \vee (\phi_1 \wedge \phi_2)$
2	$A \sqsubseteq A_1^{\phi_1}, A \sqsubseteq A_2^{\phi_2}, A_1 \sqcap A_2 \sqsubseteq B^{\phi_3}$	$\text{lab}(A \sqsubseteq B) := \text{lab}(A \sqsubseteq B) \vee (\phi_1 \wedge \phi_2 \wedge \phi_3)$
3	$A \sqsubseteq A_1^{\phi_1}, A_1 \sqsubseteq \exists r.B^{\phi_2}$	$\text{lab}(A \sqsubseteq \exists r.B) := \text{lab}(A \sqsubseteq \exists r.B) \vee (\phi_1 \wedge \phi_2)$
4	$A \sqsubseteq \exists r.A_1^{\phi_1}, A_1 \sqsubseteq B_1^{\phi_2}, \exists r.B_1 \sqsubseteq B^{\phi_3}$	$\text{lab}(A \sqsubseteq B) := \text{lab}(A \sqsubseteq B) \vee (\phi_1 \wedge \phi_2 \wedge \phi_3)$

$\mathcal{W}$  such that  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ . We will use this set to decide, through a propositional entailment test, whether  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  holds or not. Recall that contextual 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  $\kappa$  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  $\phi$  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 [10], we are interested in finding a *boundary* for a subsumption relation. Given a TBox  $\mathcal{T}$  labeled over the lattice  $\mathbb{B}$  and concept names  $A, B$ , a *boundary* for  $A \sqsubseteq B$  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  $\mathcal{T}_{\psi} \models A \sqsubseteq B$  (see [10] for further details). Notice that the join-prime elements of  $\mathbb{B}$  are exactly the valuations of variables in  $V$ . Using Lemma 18 we obtain the following result.

**Theorem 20** *Let  $\phi$  be a boundary for  $A \sqsubseteq B$  w.r.t.  $\mathcal{T}$  in  $\mathbb{B}$ . Then, for any context  $\kappa$  we have that  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  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  $\mathcal{EL}$  [12], can in fact compute the boundaries for all subsumption relations between concept names that follow from the KB.

Recall that we assume w.l.o.g. that the  $V$ -TBox is in normal form. We extend the completion algorithm from Section 2.1 to include a function  $\text{lab}$  that maps every derived subsumption relation to a Boolean formula over the variables in  $V$ . Intuitively,  $\text{lab}(C \sqsubseteq D) = \phi$  expresses that  $\mathcal{T}_{\mathcal{W}} \models C \sqsubseteq D$  in all valuations  $\mathcal{W}$  that satisfy  $\phi$ . The algorithm is initialized with the labeling of axioms

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

where  $\text{t}$  is a tautology and  $\text{f}$  a contradiction in  $\mathbb{B}$ . Let now  $\mathcal{T}' := \{\alpha \mid \langle \alpha : \kappa \rangle \in \mathcal{T}\}$ . The algorithm initializes the labels of each GCI in  $\mathcal{T}'$  to include all the contexts that are already known to entail them; that is, for every GCI  $\alpha \in \mathcal{T}'$ , we set  $\text{lab}(\alpha) := \bigvee_{\langle \alpha : \kappa \rangle \in \mathcal{T}} \chi_{\kappa}$ . This labeling function is modified by applying the rules from Table 3 where for brevity, we denote  $\text{lab}(\alpha) = \phi$  by  $\alpha^{\phi}$ . Every rule application changes the label of one subsumption for a more general formula; to ensure termination, the rule is only applied if the new label is strictly more general than the previous one. The number of such subsumption relations 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  $\mathcal{T}_{\mathcal{W}} \models C \sqsubseteq D$  for all concept names  $A, B$  and all valuations  $\mathcal{W}$  that satisfy  $\text{lab}(C \sqsubseteq D)$ . It can be shown using techniques from axiom-pinpointing (see e.g. [8, 10]) that after termination the converse also holds; i.e., for every valuation  $\mathcal{W}$ , if  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ , then  $\mathcal{W} \models \text{lab}(A \sqsubseteq B)$ . Thus, we obtain the following result.

**Theorem 21** ([8, 10]) *Let  $\text{lab}$  be the labelling function obtained through the completion algorithm. For every two concept names  $A, B$  appearing in  $\mathcal{T}$ ,  $\text{lab}(A \sqsubseteq B)$  is a boundary for  $A \sqsubseteq B$  w.r.t.  $\mathcal{T}$ .*

Using the boundary  $\phi$  for  $A \sqsubseteq B$  w.r.t.  $\mathcal{T}$ , it is possible to decide whether the contextual subsumption  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  holds; we need only to check if  $\chi_{\kappa} \models \phi$ . This decision is in NP on  $|V|$ .

*Example 22* Consider the  $\mathcal{BEL}$  KB  $\mathcal{K}_{\text{exa}}$  from Example 6. The modified completion algorithm starts with the labeled GCIs

$$A \sqsubseteq B^x, \quad B \sqsubseteq \exists r.C^{y \wedge z}, \quad \exists r.C \sqsubseteq C^{x \wedge y}, \quad B \sqsubseteq C^{\neg z}$$

Applying the first rule with the premises  $A \sqsubseteq B^x, B \sqsubseteq C^{\neg z}$  we obtain  $A \sqsubseteq C^{x \wedge \neg z}$ . From the third rule, we then get  $A \sqsubseteq \exists r.C^{x \wedge y \wedge z}$ . One last application of the fourth rule changes the label of  $A \sqsubseteq C$  to  $(x \wedge \neg z) \vee (x \wedge y \wedge z)$ . From this label, we can deduce that  $\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \{x, \neg y, \neg z\} \rangle$  holds, but  $\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \{x, \neg y, z\} \rangle$  does not. Indeed,  $\mathcal{T}_{\text{exa}}\{x, \neg y, z\} = \{\langle A \sqsubseteq B \rangle\} \not\models A \sqsubseteq C$ .

Clearly, the boundary for the atomic subsumption relation  $A \sqsubseteq B$  provides more information than necessary for deciding whether the subsumption holds in a *given* context  $\kappa$ . It encodes *all* contexts that entail the desired subsumption. We can use this knowledge to deduce other kinds of knowledge from the KB, like the most likely context. Before considering this non-standard inference, we examine the computation of the probability of a subsumption relation.

## 5.2 Probabilistic Subsumption

We consider now the problem of computing the probability of a subsumption and other associated problems; namely, deciding positive,  $p$ -subsumption, and almost-sure subsumption. First, we consider the special case in which the context is empty; i.e., we focus on the problem of finding  $P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle)$ . In other words, we are interested in the probability of a subsumption relation without any knowledge of the context in which it should hold. Afterwards, we generalize our methods to take into account also the contextual information and study first *contextual* positive, almost-sure, and  $p$ -subsumption. At the end of this section, we also introduce the conditional subsumption problems.

We start by proving a fundamental result for this logic. Specifically, that it is possible w.l.o.g. to restrict reasoning to pithy models only (recall Definition 12).

**Lemma 23** *Let  $\mathcal{K}$  be a  $\mathcal{BEL}$  KB, and  $A, B \in \mathbf{N}_{\mathcal{C}}$ . For every probabilistic model  $\mathcal{P}$  of  $\mathcal{K}$  there is a pithy model  $\mathcal{Q}$  of  $\mathcal{K}$  such that  $P(\langle A \sqsubseteq_{\mathcal{Q}} B \rangle) \leq P(\langle A \sqsubseteq_{\mathcal{P}} B \rangle)$ .*

*Proof* Let  $\mathcal{P} = (\mathfrak{J}, P_{\mathfrak{J}})$  be a probabilistic model of  $\mathcal{K}$  and assume w.l.o.g. that  $P_{\mathfrak{J}}(\mathcal{I}) > 0$  for all  $\mathcal{I} \in \mathfrak{J}$ . In particular, this means that  $\mathfrak{J}$  is finite. If  $\mathcal{P}$  is already pithy, then the result holds trivially. Otherwise, there exist two interpretations  $\mathcal{I}, \mathcal{J} \in \mathfrak{J}$  such that  $\mathcal{V}^{\mathcal{I}} = \mathcal{V}^{\mathcal{J}}$ .

If (i)  $\mathcal{I} \models \langle A \sqsubseteq B \rangle$  and  $\mathcal{J} \models \langle A \sqsubseteq B \rangle$ , or (ii)  $\mathcal{I} \not\models \langle A \sqsubseteq B \rangle$  and  $\mathcal{J} \not\models \langle A \sqsubseteq B \rangle$  then set  $\mathfrak{H} := \mathfrak{J} \setminus \{\mathcal{I}\}$ . Otherwise, assume w.l.o.g. that  $\mathcal{I} \models \langle A \sqsubseteq B \rangle$  but  $\mathcal{J} \not\models \langle A \sqsubseteq B \rangle$ ; then, we define  $\mathfrak{H} := \mathfrak{J} \setminus \{\mathcal{I}\}$ . The probabilistic interpretation  $\mathcal{P}' = (\mathfrak{H}, P_{\mathfrak{H}})$  with

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

is still a model of  $\mathcal{K}$  and  $P(\langle A \sqsubseteq_{\mathcal{P}'} B \rangle) \leq P(\langle A \sqsubseteq_{\mathcal{P}} B \rangle)$ . Moreover,  $|\mathfrak{H}| < |\mathfrak{J}|$ ; thus this construction leads to the desired pithy model.  $\square$

Using this lemma, it is possible to show that the probability of a consequence can be computed by a simple algorithm that performs standard (classical) reasoning over the restrictions  $\mathcal{T}_{\mathcal{W}}$  of  $\mathcal{T}$  (recall Definition 9).

**Theorem 24** *Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be a KB, and  $A, B$  two concept names. Then*

$$P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle) = \sum_{\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B} P_{\mathcal{B}}(\mathcal{W}).$$

*Proof* For every valuation  $\mathcal{W}$ , we construct the  $V$ -interpretation  $\mathcal{I}_{\mathcal{W}}$  as follows. If  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ , 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  $\langle A \sqsubseteq B \rangle$ , which must exist by definition. Let now  $\mathcal{P}_{\mathcal{K}} = (\mathfrak{J}, P_{\mathfrak{J}})$  be the probabilistic interpretation where  $\mathfrak{J} = \{\mathcal{I}_{\mathcal{W}} \mid \mathcal{W} \text{ a valuation of } V\}$  and  $P_{\mathfrak{J}}(\mathcal{I}_{\mathcal{W}}) = P_{\mathcal{B}}(\mathcal{W})$  for all  $\mathcal{W}$ . Then  $\mathcal{P}_{\mathcal{K}}$  is a model of  $\mathcal{K}$ . Moreover, it holds that

$$P(\langle A \sqsubseteq_{\mathcal{P}_{\mathcal{K}}} B \rangle) = \sum_{\mathcal{I}_{\mathcal{W}} \models \langle A \sqsubseteq B \rangle} P_{\mathfrak{J}}(\mathcal{I}_{\mathcal{W}}) = \sum_{\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B} P_{\mathcal{B}}(\mathcal{W}). \quad (3)$$

Thus,  $P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle) \leq \sum_{\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B} P_{\mathcal{B}}(\mathcal{W})$ . If this inequality is strict, then there exists a probabilistic model  $\mathcal{P} = (\mathfrak{J}, P_{\mathfrak{J}})$  of  $\mathcal{K}$  with  $P(\langle A \sqsubseteq_{\mathcal{P}} B \rangle) < P(\langle A \sqsubseteq_{\mathcal{P}_{\mathcal{K}}} B \rangle)$ . By Lemma 23, 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

$$\sum_{\mathcal{J}_{\mathcal{W}} \models \langle A \sqsubseteq B \rangle} P_{\mathfrak{J}}(\mathcal{J}_{\mathcal{W}}) < \sum_{\mathcal{I}_{\mathcal{W}} \models \langle A \sqsubseteq B \rangle} P_{\mathfrak{J}}(\mathcal{I}_{\mathcal{W}}).$$

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  $\mathcal{I}_{\mathcal{V}} \models \langle A \sqsubseteq B \rangle$  but  $\mathcal{J}_{\mathcal{V}} \not\models \langle A \sqsubseteq B \rangle$ . As  $\mathcal{J}_{\mathcal{V}}$  is a model of  $\mathcal{T}_{\mathcal{V}}$ , it follows that  $\mathcal{T}_{\mathcal{V}} \not\models A \sqsubseteq B$ . By construction, then we have that  $\mathcal{I}_{\mathcal{V}} \not\models \langle A \sqsubseteq B \rangle$ , which contradicts the conditions made during the construction of  $\mathcal{I}_{\mathcal{V}}$ .  $\square$

*Example 25* Consider again the KB  $\mathcal{K}_{\text{exa}}$  from Example 6. There are three valuations  $\mathcal{W}$  such that  $\mathcal{T}_{\text{exa}\mathcal{W}} \models A \sqsubseteq C$ ; namely,  $\{x, y, \neg z\}$ ,  $\{x, \neg y, \neg z\}$ , and  $\{x, y, z\}$ . Thus,  $P(\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C \rangle) = 0.49 + 0 + 0.21 = 0.7$ .

**Algorithm 1** Probability of Subsumption**Input:** KB  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  over  $V$ ; concept names  $A, B$ **Output:**  $P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle)$ 


---

```

1:  $P \leftarrow 0$ 
2: for all valuations  $\mathcal{W}$  of  $V$  do
3:   if  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$  then
4:      $P \leftarrow P + P_{\mathcal{B}}(\mathcal{W})$ 
5: return  $P$ 

```

---

Based on Theorem 24, 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 subsumption. 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 decide whether  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ , and possibly compute the probability  $P_{\mathcal{B}}(\mathcal{W})$ . The latter can be done in polynomial time on the size of  $\mathcal{B}$ , using the standard chain rule [23], while deciding subsumption w.r.t. an  $\mathcal{EL}$  TBox is polynomial on  $\mathcal{T}$  [12]. Overall, Algorithm 1 runs in time exponential on  $\mathcal{B}$  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.

**Corollary 26** *Deciding  $p$ -subsumption is PP-hard and in PSPACE. Moreover, it is fixed-parameter tractable w.r.t. the parameter  $|V|$ .*

*Proof* The upper bounds follow directly from the correctness of Algorithm 1, which is a consequence of Theorem 24. To prove the lower bound we reduce the D-PR problem for BNs, which is PP-complete. Let  $\mathcal{B}$  be an arbitrary BN over a set  $V$ , and  $\lambda$  be a  $V$ -context. Define the KB  $\mathcal{K} = (\mathcal{B}, \{\langle A \sqsubseteq B : \lambda \rangle\})$ . It then follows that  $P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle) = \sum_{\mathcal{W}(\lambda)=1} P_{\mathcal{B}}(\mathcal{W}) = P_{\mathcal{B}}(\lambda)$ . Hence, we have that  $P_{\mathcal{B}}(\lambda) \geq p$  iff  $P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle) \geq p$ .  $\square$

This corollary provides already an insight on the computational complexity of performing probabilistic reasoning in  $\mathcal{BEL}$ , and a reasoning algorithm that is easy to implement combining state-of-the-art BN inference engines and  $\mathcal{EL}$  reasoners [6, 28, 33]. It is also possible to use the boundary, as described in Section 5.1 to find the valuations  $\mathcal{W}$  such that  $\mathcal{T}_{\mathcal{W}}$  entails the subsumption relation.

To provide a tight complexity bound, we develop a new algorithm that exploits the construction of the (unraveled) proof structure introduced in Section 2.1. We first show that  $p$ -subsumption w.r.t. a  $\mathcal{BEL}$  KB can be reduced in polynomial time to the D-PR problem over a special Bayesian network. Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be an arbitrary but fixed  $\mathcal{BEL}$  KB. From the  $V$ -TBox  $\mathcal{T}$ , we construct the  $\mathcal{EL}$  TBox  $\mathcal{T}' := \{\alpha \mid \langle \alpha : \kappa \rangle \in \mathcal{T}\}$ . That is,  $\mathcal{T}'$  contains the same axioms as  $\mathcal{T}$ , but ignores the contextual information encoded in their labels. Let now  $H_{\mathcal{T}'}^u$  be the unraveled proof structure for  $\mathcal{T}'$ . By construction,  $H_{\mathcal{T}'}^u$  is a directed acyclic *hypergraph*. Our goal is to transform this hypergraph into a DAG and construct a BN, from which all the  $p$ -subsumption relations between concept names can be read through standard BN inferences. The basic idea of the reduction is depicted in Figure 4, using the KB  $\mathcal{K}_{\text{exa}}$  from Example 6. On the upper part of the figure, the unraveled proof structure of  $\mathcal{T}_{\text{exa}}$  has been transformed into a DAG, by adding a new node for each hyper-edge used. Each of the nodes of this DAG is associated with a conditional

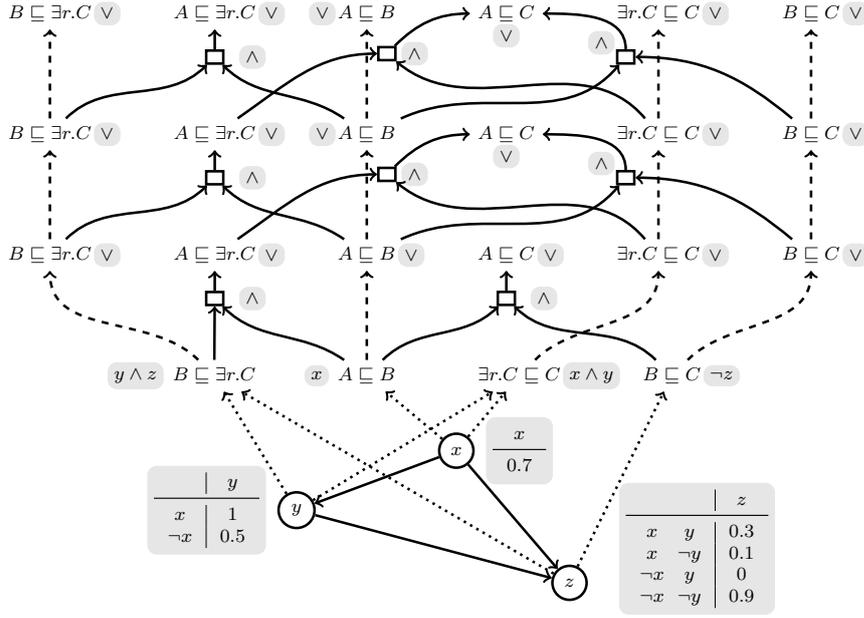


Fig. 4 Reduction of the KB  $\mathcal{K}_{\text{exa}}$  to a BN

probability table expressed by a logical condition. In the lower part, we have the original BN from the KB. The two components are connected at the base through the context associated to each axiom. We explain this construction in detail next.

Recall that hypergraphs generalize graphs by allowing edges to connect many vertices. These hyperedges can be seen as an encoding of a formula in disjunctive normal form. An edge  $(S, v)$  expresses that if all the elements in  $S$  can be reached, then  $v$  is also reachable; we see this as an implication:  $\bigwedge_{w \in S} w \Rightarrow v$ . Several edges sharing the same head  $(S_1, v), (S_2, v), \dots, (S_k, v)$  in the hypergraph can be described through the implication  $\bigvee_{i=1}^k (\bigwedge_{w \in S_i} w) \Rightarrow v$ . We can thus rewrite any directed acyclic hypergraph into a DAG by introducing auxiliary *conjunctive* and *disjunctive* nodes; the proper semantics of these nodes will be guaranteed by the conditional probability distribution defined later. Since the space needed for describing the conditional probability tables in a BN is exponential on the number of parents of the node, we also ensure that all the nodes in this DAG have at most two parent nodes.

Algorithm 2 constructs such a DAG from a directed hypergraph. The algorithm adds a new node  $\wedge_i$  for each hyperedge  $(S, v)$  in the input hypergraph  $H$ , and connects it with all the nodes in  $S$ . If there are  $k$  hyperedges that lead to a single node  $v$ , it creates  $k-1$  nodes  $\vee_i$ . These are used to represent the binary disjunctions among all the hyperedges leading to  $v$ . The algorithm runs in polynomial time on the size of  $H$ , and if  $H$  is acyclic, the resulting graph  $G$  is acyclic too. Moreover, all the nodes  $v \in V$  that existed in the input hypergraph have at most one parent node after the translation; every  $\vee_i$  node has exactly two parents, and the number of parents of a node  $\wedge_i$  is given by the set  $S$  from the hyperedge  $(S, v) \in E$  that generated it. In particular, if the input hypergraph is the unraveled proof structure

**Algorithm 2** Construction of a DAG from a hypergraph**Input:**  $H = (V, E)$  directed acyclic hypergraph**Output:**  $G = (V', E')$  directed acyclic graph

---

```

1:  $V' \leftarrow V, i, j \leftarrow 0$ 
2: for each  $v \in V$  do
3:    $\mathfrak{S} \leftarrow \{S \mid (S, v) \in E\}, j \leftarrow i$ 
4:   for each  $S \in \mathfrak{S}$  do
5:      $V' \leftarrow V' \cup \{\wedge_i\}, E' \leftarrow E' \cup \{(u, \wedge_i) \mid u \in S\}$ 
6:     if  $i > j$  then
7:        $V' \leftarrow V' \cup \{\vee_i\}, E' \leftarrow E' \cup \{(\wedge_i, \vee_i)\}$ 
8:      $i \leftarrow i + 1$ 
9:   if  $i = j + 1$  then ▷ If the GCI has only one explanation
10:     $E' \leftarrow E' \cup \{(\wedge_j, v)\}$ 
11:   else
12:     $E' \leftarrow E' \cup \{(\vee_k, \vee_{k+1}) \mid j < k < i - 1\} \cup \{(\vee_{i-1}, v), (\wedge_j, \vee_{j+1})\}$ 
13: return  $G = (V', E')$ 

```

---

for a TBox  $\mathcal{T}$ , then the size of the generated graph  $G$  is polynomial on the size of  $\mathcal{T}$ , and each node has at most three parent nodes.

The next step is to build a BN that preserves the probabilistic entailments of a  $\mathcal{BEL}$  KB. Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be such a KB, with  $\mathcal{B} = (G, \Phi)$ , and let  $G_{\mathcal{T}} = (V_{\mathcal{T}}, E_{\mathcal{T}})$  be the DAG obtained from the unraveled proof structure of  $\mathcal{T}$  using Algorithm 2. Recall that the nodes of  $G_{\mathcal{T}}$  are either (i) pairs of the form  $(\alpha, i)$ , where  $\alpha$  is a GCI in normal form built from the signature of  $\mathcal{T}$ , or (ii) an auxiliary disjunction  $(\vee_i)$  or conjunction  $(\wedge_i)$  node introduced by Algorithm 2. Moreover,  $(\alpha, 0)$  is a node of  $G_{\mathcal{T}}$  iff there is a context  $\kappa$  with  $\langle \alpha : \kappa \rangle \in \mathcal{T}$ . We assume w.l.o.g. that for node  $(\alpha, 0)$  there is *exactly one* such context. If there were more than one, then we could extend the BN  $\mathcal{B}$  with an additional variable which describes the disjunctions of these contexts, as in the construction from Algorithm 2. Similarly, we assume w.l.o.g. that each context  $\kappa$  appearing in  $\mathcal{T}$  contains at most two literals. For a context  $\kappa$ , let  $\text{var}(\kappa)$  denote the set of all variables appearing in  $\kappa$ . We construct a new BN  $\mathcal{B}_{\mathcal{K}}$  as follows.

Let  $V_{\mathcal{K}} := V \cup V_{\mathcal{T}}$  and  $E_{\mathcal{K}} := E \cup E_{\mathcal{T}} \cup \{(x, (\alpha, 0)) \mid \langle \alpha : \kappa \rangle \in \mathcal{T}, x \in \text{var}(\kappa)\}$ . By construction  $G_{\mathcal{K}} = (V_{\mathcal{K}}, E_{\mathcal{K}})$  is a DAG. We now need only to define the conditional probability tables for the nodes in  $V_{\mathcal{T}}$  given their parents in  $G_{\mathcal{K}}$ ; notice that the structure of the graph  $G$  remains unchanged for the construction of  $G_{\mathcal{K}}$ . For every node  $(\alpha, 0) \in V_{\mathcal{T}}$ , there is a  $\kappa$  such that  $\langle \alpha : \kappa \rangle \in \mathcal{T}$ ; the parents of  $(\alpha, 0)$  in  $G_{\mathcal{K}}$  are then  $\text{var}(\kappa) \subseteq V$ . The conditional probability of  $(\alpha, 0)$  given its parents is defined, for every valuation  $\mathcal{V}$  of  $\text{var}(\kappa)$  as  $P_{\mathcal{B}}((\alpha, 0) = \text{true} \mid \mathcal{V}) = \mathcal{V}(\kappa)$ ; that is, the probability of  $(\alpha, 0)$  being true given a valuation of its parents is 1 if the valuation makes the context  $\kappa$  true; otherwise, it is 0. Recall now that each auxiliary node has at most three parents. The conditional probability of a conjunction node  $\wedge_i$  being true is 1 iff all parents are true, and the conditional probability of a disjunction node  $\vee_i$  being true is 1 iff at least one parent is true. Finally, every  $(\alpha, i)$  with  $i > 0$  has exactly one parent node  $v$ ;  $(\alpha, i)$  is true with probability 1 iff  $v$  is true. From the properties of proof structures and Theorem 4 we have that

$$P_{\mathcal{B}_{\mathcal{K}}}((\alpha, n)) = \sum_{\mathcal{V}} P_{\mathcal{B}_{\mathcal{K}}}((\alpha, n) \mid \mathcal{V}) P_{\mathcal{B}_{\mathcal{K}}}(\mathcal{V}) = \sum_{\mathcal{W} \models \alpha} P_{\mathcal{B}_{\mathcal{K}}}(\mathcal{W}). \quad (4)$$

which yields the following result.

**Theorem 27** *Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be a  $\mathcal{BEL}$  KB,  $A, B \in \mathbf{N}_{\mathcal{C}}$ , and  $n = |\text{sig}(\mathcal{T})|^3$ . Then  $P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle) = P_{\mathcal{B}_{\mathcal{K}}}(\langle A \sqsubseteq B, n \rangle)$ .*

This theorem states that we can reduce  $p$ -subsumption w.r.t. the  $\mathcal{BEL}$  KB  $\mathcal{K}$  to a probabilistic inference in the BN  $\mathcal{B}_{\mathcal{K}}$ . Notice that the size of  $\mathcal{B}_{\mathcal{K}}$  is polynomial on the size of  $\mathcal{K}$ . This means that  $p$ -subsumption is at most as hard as deciding D-PR problems over the BN  $\mathcal{B}_{\mathcal{K}}$ .

**Corollary 28** *Deciding  $p$ -subsumption is PP-complete in the size of the KB.*

*Proof* The upper bound is a direct consequence of Theorem 27 and Equation (5). The lower bound was already shown in Corollary 26.  $\square$

Depending on the intended use of the KB, knowing the precise probability of a consequence is not always fundamental, but it might suffice to know whether it is at least possible (with probability greater than zero), or almost certain. If we consider only these cases as decision problems, then the complexity of probabilistic reasoning decreases to the first level of the polynomial hierarchy.

**Theorem 29** *Positive subsumption is NP-complete, and almost-sure subsumption is coNP-complete.*

*Proof* To decide positive subsumption, we can guess a valuation  $\mathcal{W}$  and verify in polynomial time that (i)  $P_{\mathcal{B}}(\mathcal{W}) > 0$  and (ii)  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ . The correctness of this approach is given by Theorem 24. Thus the problem is in NP.

Recall now that deciding, given a BN  $\mathcal{B}$  and a variable  $x \in V$ , if  $P_{\mathcal{B}}(x) > 0$  is already NP-hard [20]. Consider the KB  $\mathcal{K} = (\mathcal{B}, \{\langle A \sqsubseteq B : \{x\} \rangle\})$ , where  $A, B$  are two arbitrary concept names. It follows from Theorem 24 that  $P_{\mathcal{B}}(x) > 0$  iff  $P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle) > 0$ . Thus positive subsumption is NP-hard. The coNP-completeness of almost-sure subsumption can be shown by analogous (but dual) arguments.  $\square$

Notice 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 as long as the DAG representing the BN 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.

We now look at the general case, where we are interested in  $P(\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle)$  for some potentially non-empty context  $\kappa$ . As for the special case without contexts, the fundamental property for the contextual case is that we can restrict reasoning to pithy models without loss of generality. Formally, for every probabilistic model  $\mathcal{P}$  of  $\mathcal{K}$  there is a pithy model  $\mathcal{Q}$  of  $\mathcal{K}$  such that  $P(\langle A \sqsubseteq_{\mathcal{Q}} B : \kappa \rangle) \leq P(\langle A \sqsubseteq_{\mathcal{P}} B : \kappa \rangle)$ . The proof of this claim follows the same steps as the proof of Lemma 23, and hence we do not present it in detail here. Using this result, we can generalize Theorem 24 to consider the contextual information.

**Theorem 30** *Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be a KB,  $A, B \in \mathbf{N}_{\mathcal{C}}$ , and  $\kappa$  a context. Then*

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

Once again, the proof is very similar to the proof of Theorem 24. The biggest change is in Equation (3), which now becomes

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

*Example 31* Consider again the KB  $\mathcal{K}_{\text{exa}}$  from Example 6. Notice that for every valuation  $\mathcal{W}$  that evaluates  $x$  to false,  $\mathcal{T}_{\mathcal{W}} \not\models A \sqsubseteq C$ . Thus, we have that  $P(\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \{\neg x\} \rangle) = 1 - P_{\mathcal{B}_{\text{exa}}}(\neg x) + 0 = 0.7$ . Moreover, since  $P_{\mathcal{B}_{\text{exa}}}(\neg x, y, z) = 0$ , we get  $P(\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \{\neg x, y, z\} \rangle) = 1$ .

Theorem 30 tells us that, to compute the probability of a contextual subsumption, it suffices to compute the probability of the context in the original BN and the probability of all the valuations that satisfy both, the context and the subsumption relation. To compute the last term in this equation, we can once again use the proof structure as described before. In fact, from the properties of proof structures, it follows that

$$P_{\mathcal{B}_{\kappa}}((\alpha, n), \kappa) = \sum_{\mathcal{V}(\kappa)=1} P_{\mathcal{B}_{\kappa}}((\alpha, n) \mid \mathcal{V}) P_{\mathcal{B}_{\kappa}}(\mathcal{V}) = \sum_{\substack{\mathcal{T}_{\mathcal{W}} \models \alpha \\ \mathcal{W}(\kappa)=1}} P_{\mathcal{B}_{\kappa}}(\mathcal{W}), \quad (5)$$

and hence  $P(\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle) = 1 - P_{\mathcal{B}}(\kappa) + P_{\mathcal{B}_{\kappa}}((A \sqsubseteq B, n), \kappa)$ . Altogether, we have the following result.

**Theorem 32** *Contextual  $p$ -subsumption is PP-complete, contextual positive subsumption is NP-complete, and contextual almost-sure subsumption is coNP-complete.*

The probabilistic semantics of  $\mathcal{BEL}$ , as introduced in the previous section, consider the relationship between a context and the logical subsumption relation that follows the structure of the so-called *material implication*. More precisely, we logically express that *if* the context is satisfied, *then* the subsumption relation must hold too. The  $V$ -GCI is then trivially satisfied whenever the context is violated, following the classical rules of implication. As we have already shown in Example 31, this kind of implication may produce some counter-intuitive results. Indeed, it is easy to see that  $\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \{\neg x, y, z\} \rangle$  does not hold, but  $P(\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \{\neg x, y, z\} \rangle) = 1$ . The reason for this is simply that the context  $\{\neg x, y, z\}$  has probability zero of occurring. To avoid this unexpected behaviour, one possibility is to use a conditional implication and compute the probability of the subsumption to hold, given that the context is known to be satisfied. Intuitively, the conditional implication rescales the probability distribution to only the space where the antecedent of the implication is true. The following definition translates this idea to the case of probabilistic subsumption.

**Definition 33 (conditional subsumption)** Let  $A, B$  be two concept names,  $\kappa$  a  $V$ -context,  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  a  $\mathcal{BEL}$  KB, and  $\mathcal{P} = (\mathcal{J}, P_{\mathcal{J}})$  a probabilistic interpretation. The *conditional probability* of  $A \sqsubseteq B$  given  $\kappa$  is defined by the rule

$$P(\langle A \sqsubseteq_{\mathcal{P}} B \mid \kappa \rangle) = \sum_{\mathcal{I} \in \mathcal{J}, \mathcal{V}^{\mathcal{I}}(\kappa)=1} P_{\mathcal{J}}(\mathcal{I}) = \sum_{\mathcal{I} \in \mathcal{J}, \mathcal{V}^{\mathcal{I}}(\kappa)=1, A^{\mathcal{I}} \subseteq B^{\mathcal{I}}} P_{\mathcal{J}}(\mathcal{I}). \quad (6)$$

The *conditional probability* of  $A \sqsubseteq B$  given  $\kappa$  w.r.t.  $\mathcal{K}$  is

$$P(\langle A \sqsubseteq_{\mathcal{K}} B \mid \kappa \rangle) := \inf_{\mathcal{P} \models \mathcal{K}} P(\langle A \sqsubseteq_{\mathcal{P}} B \mid \kappa \rangle).$$

The notions of  $p$ -subsumption, positive subsumption, and almost-sure subsumption are extended to conditional subsumption in the obvious manner.

Recall from Theorem 11 that for every model  $\mathcal{P}$  of  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  it holds that  $\sum_{\mathcal{I} \in \mathcal{J}, \mathcal{V}^{\mathcal{I}}(\kappa)=1} P_{\mathcal{J}}(\mathcal{I}) = P_{\mathcal{B}}(\kappa)$ . Thus, if  $P_{\mathcal{B}}(\kappa) > 0$ , the rule from Equation (6) is equivalent to

$$P(\langle A \sqsubseteq_{\mathcal{P}} B \mid \kappa \rangle) = \frac{\sum_{\mathcal{I} \in \mathcal{J}, \mathcal{V}^{\mathcal{I}}(\kappa)=1, A^{\mathcal{I}} \subseteq B^{\mathcal{I}}} P_{\mathcal{J}}(\mathcal{I})}{P_{\mathcal{B}}(\kappa)}.$$

Interestingly, the techniques that we have developed for probabilistic contextual subsumption can also be used for deciding the different conditional subsumption problems. Moreover, the complexity of all these decision problems remains the same as for the contextual case.

**Theorem 34** *Conditional  $p$ -subsumption is PP-complete, conditional positive subsumption is NP-complete, and conditional almost-sure subsumption is coNP-complete.*

*Proof* By Theorem 24, we know that

$$P(\langle A \sqsubseteq_{\mathcal{K}} B \rangle) = \sum_{\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B} P_{\mathcal{B}}(\mathcal{W}) = \frac{\sum_{\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B} P_{\mathcal{B}}(\mathcal{W})}{P_{\mathcal{B}}(\emptyset)} = P(\langle A \sqsubseteq_{\mathcal{K}} B \mid \emptyset \rangle).$$

The hardness results follow directly from Corollary 28 and Theorem 29.

To decide conditional  $p$ -subsumption, we can use the proof structure as before. Indeed, from Equation (5), we know that  $P(\langle A \sqsubseteq_{\mathcal{K}} B \mid \kappa \rangle)P_{\mathcal{B}}(\kappa) = P_{\mathcal{B}_{\mathcal{K}}}((\alpha, n), \kappa)$ . This means that  $p$ -subsumption can be decided by a linear number of probabilistic tests on the BN  $\mathcal{B}_{\mathcal{K}}$ , whose size is polynomial on  $\mathcal{K}$ . Thus  $p$ -subsumption is in PP. Conditional positive subsumption can be decided by the non-deterministic algorithm that guesses a valuation  $\mathcal{W}$  that extends  $\kappa$  and verifies in polynomial time that  $P_{\mathcal{B}}(\mathcal{W}) > 0$  and  $\mathcal{T}_{\mathcal{W}} \models A \sqsubseteq B$ . The upper bound for conditional almost-sure subsumption is dual.  $\square$

A simple consequence of the proof of this theorem is that both implications coincide when the context is empty; i.e.,  $P(\langle A \sqsubseteq_{\mathcal{K}} B : \emptyset \rangle) = P(\langle A \sqsubseteq_{\mathcal{K}} B \mid \emptyset \rangle)$ . In terms of practical applications, it is usually this case that is considered: we want to know how likely it is to observe a consequence, without having any knowledge of the context in which we are currently at. In the next section, we consider a dual problem: knowing that an entailment holds, try to deduce the current context.

**Table 4** The contexts  $\kappa$  such that  $\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \kappa \rangle$  holds, and their probability.

Context	Probability
$\{x, y, \neg z\}$	$P(x, y, \neg z) = P(\neg z   x, y)P(y   x)P(x) = 0.7 \cdot 1 \cdot 0.7 = 0.49$
$\{x, \neg y, \neg z\}$	$P(x, \neg y, \neg z) = P(\neg z   x, \neg y)P(\neg y   x)P(x) = 0.9 \cdot 0 \cdot 0.7 = 0$
$\{x, \neg z\}$	$P(x, \neg z) = P(x, y, \neg z) + P(x, \neg y, \neg z) = 0.49 + 0 = 0.49$
$\{x, y, z\}$	$P(x, y, z) = P(z   x, y)P(y   x)P(x) = 0.3 \cdot 1 \cdot 0.7 = 0.21$

### 5.3 Most Likely Context

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; if a consequence holds, we want to find the context with the highest probability in which we are making this observation.

**Definition 35 (most likely context)** Let  $\mathcal{K} = (\mathcal{B}, \mathcal{T})$  be a  $\mathcal{BEL}$  KB, and  $A, B$  two concept names. A  $V$ -context  $\kappa$  is a *most likely context* (mlc) for  $A \sqsubseteq B$  if (i)  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  and (ii) for all contexts  $\kappa'$  with  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa' \rangle$ ,  $P_{\mathcal{B}}(\kappa) \geq P_{\mathcal{B}}(\kappa')$ .

Continuing with our running example, it is easy to see that the mlcs for  $A \sqsubseteq C$  w.r.t.  $\mathcal{K}_{\text{exa}}$  are  $\{x, \neg z\}$  and  $\{x, y, \neg z\}$ . Indeed, from the boundary computed in Example 22, we know that there are only four contexts  $\kappa$  such that  $\langle A \sqsubseteq_{\mathcal{K}_{\text{exa}}} C : \kappa \rangle$  holds. These contexts are shown in the first column of Table 4. In the second column, we compute their probability. Highlighted with a gray background are those contexts that have the highest probability among those that entail the subsumption relation. Since  $\{x, y, \neg z\}$  contains  $\{x, \neg z\}$ , one can think of the latter as a more informative mlc than the former. Usually, one might be interested in finding such mlcs. The techniques that we present next can be easily adapted to find these mlcs, too.

A simple algorithm for computing all most likely contexts for a given subsumption relation is to enumerate all the (exponentially many) contexts  $\kappa$ , test whether  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  holds, and if so compute  $P_{\mathcal{B}}(\kappa)$ . Such an algorithm runs in exponential time on the size of  $\mathcal{K}$ . This complexity bound cannot be further improved, since a single consequence may have exponentially many mlcs. We consider now the problem of finding *one* mlc; or, more precisely, its associated decision problem: decide, given a  $p \in (0, 1]$ , whether there exists a context  $\kappa$  such that  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  and  $P_{\mathcal{B}}(\kappa) > p$ .<sup>4</sup>

This problem is clearly in  $\text{NP}^{\text{PP}}$ : given  $p > 0$  we can guess a  $V$ -context  $\kappa$ , and check with a PP oracle that  $\langle A \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  and  $P_{\mathcal{B}}(\kappa) > p$ , using the construction from Section 5.2. To show that it is also  $\text{NP}^{\text{PP}}$ -hard, we provide a reduction from D-MAP, which corresponds to finding a valuation that maximizes the probability of an event.

**Theorem 36** *Deciding whether  $\kappa$  is an mlc for  $A \sqsubseteq B$  w.r.t.  $\mathcal{K}$  is  $\text{NP}^{\text{PP}}$ -complete.*

*Proof* The upper bound can be shown as described above. Hence we focus here on proving the lower bound by a reduction from D-MAP. Let  $\mathcal{B} = ((V, E), \Phi)$

<sup>4</sup> We use a different formulation than in our previous work [14] to better align the decision problem with existing literature.

be an arbitrary but fixed BN,  $\kappa$  a  $V$ -context,  $Q = \{x_1, \dots, x_k\} \subseteq V$ , and  $p > 0$ . Define  $V' = V \uplus \{x^+, x^- \mid x \in Q\} \uplus \{z\}$ , where  $\uplus$  denotes the disjoint union, and  $E' = E \cup \{(x, x^+), (x, x^-) \mid x \in Q\}$ . We construct  $\mathcal{B}' = ((V', E'), \Phi')$  where  $\Phi'$  contains  $P_{\mathcal{B}'}(v \mid \pi(v)) = P_{\mathcal{B}}(v \mid \pi(x))$  for all  $v \in V$ , and  $P_{\mathcal{B}'}(z) = p$ ,  $P_{\mathcal{B}'}(x^+ \mid x) = 1$ ,  $P_{\mathcal{B}'}(x^+ \mid \neg x) = 0$ ,  $P_{\mathcal{B}'}(x^- \mid x) = 0$ , and  $P_{\mathcal{B}'}(x^- \mid \neg x) = 1$  for all  $x \in Q$ . Let now

$$\mathcal{T} = \{\langle A_{i-1} \sqsubseteq A_i : x_i^+ \rangle, \langle A_{i-1} \sqsubseteq A_i : x_i^- \rangle \mid 1 \leq i \leq k\} \cup \{\langle A_k \sqsubseteq B : \kappa \rangle, \langle A_0 \sqsubseteq B : z \rangle\},$$

and  $\mathcal{K} = (\mathcal{B}', \mathcal{T})$ . For any  $V'$ -context  $\kappa'$  it holds that if  $\langle A_0 \sqsubseteq_{\mathcal{K}} B : \kappa \rangle$  and  $z \notin \kappa'$ , then  $\kappa \subseteq \kappa'$  and for every  $x \in Q$ ,  $\{x^+, x^-\} \cap \kappa' \neq \emptyset$ . Moreover, by construction  $P_{\mathcal{B}}(z) = p$  and  $P_{\mathcal{B}}(x^+, x^-) = 0$  for all  $x \in Q$ .  $\square$

Notice that part of the hardness of this problem arises from the fact that contexts represent only partial valuations of the variables in  $V$ . Computing the probability of these partial valuations is already hard, since it is necessary to consider all the valuations that extend the context itself. If we restrict the search to only the most probable *valuations*, then the complexity of the problem decreases to NP. Formally, we say that a valuation  $\mathcal{W}$  is a *most likely world* (mlw) for  $A \sqsubseteq B$  if  $\langle A \sqsubseteq_{\mathcal{K}} B : \mathcal{W} \rangle$  and for all valuations  $\mathcal{W}'$  such that  $\langle A \sqsubseteq_{\mathcal{K}} B : \mathcal{W}' \rangle$ , it holds that  $P_{\mathcal{B}}(\mathcal{W}) \geq P_{\mathcal{B}}(\mathcal{W}')$ . In our running example, we have only one mlw for  $A \sqsubseteq C$ ; namely  $\{x, y, \neg z\}$  (see Table 4).

**Theorem 37** *Deciding whether  $\mathcal{W}$  is an mlw for  $A \sqsubseteq B$  w.r.t.  $\mathcal{K}$  is NP-complete.*

*Proof* For the upper bound, we can guess a valuation  $\mathcal{W}$  and decide in polynomial time using the chain rule that  $P_{\mathcal{B}}(\mathcal{W}) > p$ . For the lower bound, we reduce the D-MPE problem, following the same approach described in the proof of Theorem 36.  $\square$

This finishes our analysis of the complexity of reasoning in  $\mathcal{BEL}$ . In the next section, we generalize some of these results to arbitrary ontology languages.

## 6 Bayesian Ontology Languages

In Section 4 we defined  $\mathcal{BEL}$  as a Bayesian extension of the classical DL  $\mathcal{EL}$ . Obviously, the main ideas behind this extension are independent of the specific DL used, and can in fact be generalized to any ontology language. An *ontology language*  $\mathcal{L}$  defines possibly infinite classes (i)  $\mathfrak{A}$  of well-formed *axioms*; (ii)  $\mathfrak{C}$  of *consequences*; and (iii)  $\mathfrak{D}$  of *ontologies* such that every element of  $\mathfrak{D}$  is a finite subset of  $\mathfrak{A}$ , and if  $\mathcal{O} \in \mathfrak{D}$ , then every subset of  $\mathcal{O}$  is also an ontology in  $\mathfrak{D}$ . Examples of ontology languages are all usual DLs; e.g.,  $\mathcal{ALC}$  or  $\mathcal{SHOIQ}$ . In this case, *axioms* can be GCIs or assertions, and the class of ontologies may be the set of all TBoxes, or even be restricted to *acyclic* TBoxes only [7]. Consequences in these languages can be, for instance, concept unsatisfiability, concept subsumption, or ontology inconsistency. Notice however that the notion of an ontology language is not restricted to DLs [10].

The semantics of an ontology language is given by a class  $\mathbb{I}$  of *interpretations* and an *entailment relation*  $\models \subseteq \mathbb{I} \times (\mathfrak{A} \cup \mathfrak{C})$  that expresses which interpretations satisfy

which axioms and consequences. An interpretation  $\mathcal{I}$  is a *model* of the ontology  $\mathcal{O}$ , denoted by  $\mathcal{I} \models \mathcal{O}$  if  $\mathcal{I} \models \alpha$  for all  $\alpha \in \mathcal{O}$ .  $\mathcal{O}$  is called *consistent* if it has at least one model.  $\mathcal{O}$  *entails* the consequence  $c$ , denoted by  $\mathcal{O} \models c$  if every model of  $\mathcal{O}$  satisfies  $c$ . Notice that the entailment relation is monotonic; i.e., if  $\mathcal{O} \models c$  and  $\mathcal{O} \subseteq \mathcal{O}' \in \mathfrak{D}$ , then  $\mathcal{O}' \models c$ .

For the rest of this section, we consider  $\mathcal{L}$  to be an arbitrary but fixed ontology language, with axioms  $\mathfrak{A}$ , ontologies  $\mathfrak{O}$ , consequences  $\mathfrak{C}$ , and interpretations  $\mathbb{I}$ . In the Bayesian ontology language  $\mathcal{BEL}$ , ontologies are generalized by annotating every axiom with a context.

**Definition 38 (KB, model)** A *V-axiom* is of the form  $\langle \alpha : \kappa \rangle$  where  $\alpha \in \mathfrak{A}$  and  $\kappa$  is a *V-context*. A *V-ontology* is a finite set  $\mathcal{O}$  of *V-axioms* with  $\{\alpha \mid \langle \alpha : \kappa \rangle \in \mathcal{O}\} \in \mathfrak{D}$ . A *V-consequence* is a pair  $\langle c : \kappa \rangle$ , where  $c \in \mathfrak{C}$  and  $\kappa$  is a *V-context*. A  *$\mathcal{BEL}$  knowledge base (KB)* over  $V$  is a pair  $\mathcal{K} = (\mathcal{B}, \mathcal{O})$  where  $\mathcal{B}$  is a BN over  $V$  and  $\mathcal{O}$  is a *V-ontology*.

A *V-interpretation* is a pair  $\mathbb{I} = (\mathcal{I}, \mathcal{V}^{\mathbb{I}})$  where  $\mathcal{I} \in \mathbb{I}$  and  $\mathcal{V}^{\mathbb{I}} : V \rightarrow \{0, 1\}$  is a *valuation* of the variables in  $V$ . This *V-interpretation* is a *model* of the *V-axiom*  $\langle \alpha : \kappa \rangle$  or the *V-consequence*  $\langle c : \kappa \rangle$ , denoted as  $\mathbb{I} \models \langle \alpha : \kappa \rangle$  (respectively  $\mathbb{I} \models \langle c : \kappa \rangle$ ), iff  $\mathcal{V}^{\mathbb{I}}(\kappa) = 0$  or  $\mathcal{I} \models \alpha$  (resp.,  $\mathcal{I} \models c$ ). It is a *model* of the *V-ontology*  $\mathcal{O}$  iff it is a model of all the *V-axioms* in  $\mathcal{O}$ . The ontology  $\mathcal{O}$  *entails*  $\langle c : \kappa \rangle$  if every model of  $\mathcal{O}$  is a model of  $\langle c : \kappa \rangle$ .

The notion of a probabilistic interpretation is a straightforward adaptation from Definition 10. A *probabilistic interpretation* is a pair  $\mathcal{P} = (\mathfrak{J}, P_{\mathfrak{J}})$ , where  $\mathfrak{J}$  is a set of *V-interpretations* and  $P_{\mathfrak{J}}$  is a probability distribution over  $\mathfrak{J}$  such that  $P_{\mathfrak{J}}(\mathbb{I}) > 0$  only for finitely many interpretations  $\mathbb{I} \in \mathfrak{J}$ . It is a *model* of the *V-ontology*  $\mathcal{O}$  (resp., of the *V-consequence*  $\langle c : \kappa \rangle$ ) if every  $\mathbb{I} \in \mathfrak{J}$  is a model of  $\mathcal{O}$  (resp., of  $\langle c : \kappa \rangle$ ).  $\mathcal{P}$  is *consistent* with the BN  $\mathcal{B}$  if for every valuation  $\mathcal{W}$  of the variables in  $V$  it holds that

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

The probabilistic interpretation  $\mathcal{P}$  is a *model* of the KB  $(\mathcal{B}, \mathcal{O})$  iff it is a (probabilistic) model of  $\mathcal{O}$  and consistent with  $\mathcal{B}$ . This KB is *consistent* if it has at least one model.

**Definition 39 (entailment)** The  $\mathcal{BEL}$  KB  $\mathcal{K}$  *entails* the *V-consequence*  $\langle c : \kappa \rangle$ , denoted as  $\mathcal{K} \models \langle c : \kappa \rangle$ , if every probabilistic model of  $\mathcal{K}$  is a model of  $\langle c : \kappa \rangle$ . The *probability* of  $\langle c : \kappa \rangle$  w.r.t. the probabilistic interpretation  $\mathcal{P} = (\mathfrak{J}, P_{\mathfrak{J}})$  is  $P_{\mathcal{P}}(\langle c : \kappa \rangle) := \sum_{\mathbb{I} \in \mathbb{I}, \mathbb{I} \models \langle c : \kappa \rangle} P_{\mathbb{I}}(\mathbb{I})$ . The *probability* of  $\langle c : \kappa \rangle$  w.r.t.  $\mathcal{K}$  is

$$P_{\mathcal{K}}(\langle c : \kappa \rangle) := \inf_{\mathcal{P} \models \mathcal{K}} P_{\mathcal{P}}(\langle c : \kappa \rangle).$$

Notice that, contrary to the case of  $\mathcal{EL}$ , in an arbitrary ontology language  $\mathcal{L}$  there might be consequences that do not hold in any model of a given ontology. In such cases, it can be meaningful to compute the *supremum* of all the probabilities given by models of the KB, in addition to the infimum, as defined above. Such supremum corresponds to some kind of *brave* reasoning, where one is interested in consequences that follow from at least one model, rather than in all models. However, in general it is not clear how to compute such brave entailments, even in classical ontology languages. For that reason, we focus on entailments as introduced in Definition 39.

Clearly, the complexity of reasoning in  $\mathcal{BL}$  depends on the complexity of deciding entailments in the original ontology language  $\mathcal{L}$ . Hence, it is impossible to provide a general complexity bound for reasoning in arbitrary Bayesian ontology languages. Nonetheless, we can still apply some of the results presented in the previous sections to obtain some understanding of the computational properties of these problems on Bayesian extensions of different ontology languages.

Notice first that if a  $\mathcal{BL}$  KB  $\mathcal{K} = (\mathcal{B}, \mathcal{O})$  is inconsistent, then by definition all entailments hold with probability 1. It is easy to see that  $\mathcal{K}$  is inconsistent iff there is a valuation  $\mathcal{W}$  such that  $P_{\mathcal{B}}(\mathcal{W}) > 0$  and  $\mathcal{O}_{\mathcal{W}}$  is an inconsistent  $\mathcal{L}$ -ontology. Thus,  $\mathcal{BL}$  KB consistency can be decided by guessing such a world  $\mathcal{W}$  and verifying that it satisfies both conditions.

**Theorem 40** *Let  $\mathfrak{C}$  be the complexity of deciding consistency in the ontology language  $\mathcal{L}$ . Then consistency in  $\mathcal{BL}$  is in  $\text{NP}^{\mathfrak{C}}$ .*

Recall that the ontology language  $\mathcal{L}$  can be seen as a special case of its Bayesian extension  $\mathcal{BL}$ , as described in Section 4. In particular, this means that consistency, contextual entailment,  $p$ -entailment, positive entailment, and almost-sure entailment in  $\mathcal{BL}$  are necessarily at least as hard as classical entailment in  $\mathcal{L}$ . To obtain an upper bound for  $p$ -entailment, notice that the proofs of Lemma 23 and Theorem 24 only depend on the fact that the entailment relation in  $\mathcal{EL}$  is monotonic, as is the case for all ontology languages. To decide  $p$ -entailment, it then suffices to make an entailment test and a probability computation for each possible valuation of the variables in  $V$ , as suggested by Algorithm 1. Similarly, for deciding positive and almost-sure entailment it suffices to guess an adequate valuation.

**Theorem 41** *Let  $\mathfrak{C}$  be the complexity of deciding entailment in the ontology language  $\mathcal{L}$ . Then  $p$ -entailment in  $\mathcal{BL}$  is in  $\text{PSPACE}^{\mathfrak{C}}$ , positive entailment is in  $\text{NP}^{\mathfrak{C}}$  and almost-sure entailment is in  $\text{coNP}^{\mathfrak{C}}$ .*

Notice that the same upper bounds hold for the case where the material implication is substituted by the conditional implication. In particular, this theorem implies that if  $\mathfrak{C}$  is a deterministic class containing  $\text{PSPACE}$ , then all these decision problems are  $\mathfrak{C}$ -complete.

When considering contextual entailment, it is easy to see an analogous of Lemma 18 holds for all Bayesian ontology languages. This lemma yields an algorithm for proving that contextual entailment does not hold: guess a valuation  $\mathcal{W}$  and verify using standard reasoning that the entailment does not hold.

**Theorem 42** *Let  $\mathfrak{C}$  be the complexity of deciding entailment in the ontology language  $\mathcal{L}$ . Then contextual entailment is in  $\text{coNP}^{\mathfrak{C}}$ .*

To conclude, we consider the problems of deciding whether a context is a most likely context, or a valuation being a most likely world for a given entailment. As described in Section 5.3, all mlcs and mlws can be computed by an exponential-time algorithm with a standard  $\mathcal{L}$ -entailment oracle. In general, however, we cannot improve this upper bound as done for  $\mathcal{BEL}$ . In this case, the upper bound obtained is slightly more elaborate.

**Theorem 43** *Let  $\mathfrak{C}$  and  $\mathfrak{D}$  be two complexity classes containing  $\text{PP}$  and  $\text{P}$ , respectively. Then (i) if  $\mathcal{L}$ -entailment is decidable in  $\mathfrak{C}$ , then  $\text{mlc}$  is in  $(\Delta_2^{\text{P}})^{\mathfrak{C}}$ ; and (ii) if  $\mathcal{L}$ -entailment is in  $\mathfrak{D}$ , then  $\text{mlw}$  is in  $\text{NP}^{\mathfrak{D}}$ .*

**Table 5** Complexity bounds for reasoning in Bayesian extensions of different DLs.

Logic	classic	$p$	entailment			context	mlc	mlw	consistent
			positive	almost-sure					
$\mathcal{EL}$	P <sub>TIME</sub>	PP	NP	coNP	coNP	NP <sup>PP</sup>	NP	1	
$\mathcal{ALL}_a$	PSP	PSP	PSP	PSP	PSP	PSP	PSP	PSP	
$\mathcal{ALL}$	EXPT	EXPT	EXPT	EXPT	EXPT	EXPT	EXPT	EXPT	
$\mathcal{SHOIQ}$	NE	$\leq$ PSP <sup>NE</sup>	$\leq$ NP <sup>NE</sup>	$\leq$ coNP <sup>NE</sup>	$\leq$ coNP <sup>NE</sup>	$\leq$ ( $\Delta_2^P$ ) <sup>NE</sup>	$\leq$ NP <sup>NE</sup>	$\leq$ NP <sup>NE</sup>	

*Proof* Given a  $p > 0$ , we can guess a context  $\kappa$  and a world  $\mathcal{W}$  with  $\mathcal{W}(\kappa) = 1$ . verify in PP that  $P_{\mathcal{B}}(\kappa) > p$  and in coNP <sup>$\mathcal{C}$</sup>  that  $\mathcal{K} \models \langle c : \kappa \rangle$  (see Theorem 42). For mlw, on the other hand, verifying  $P_{\mathcal{B}}(\mathcal{W}) > p$  can be done in polynomial time and  $\mathcal{K} \models \langle c : \mathcal{W} \rangle$  requires only an  $\mathcal{L}$ -entailment check.  $\square$

Using all these results, we obtain complexity bounds for reasoning in the Bayesian extensions of many DLs. Some of these results, including the tight bounds proven for  $\mathcal{BEL}$  in previous sections, are shown in Table 5. In the table, all the bounds are tight, except for the last row ( $\mathcal{SHOIQ}$ ), where only upper bounds are known (denoted by  $\leq \mathcal{C}$ ). For brevity, we write PSP for PSPACE, EXPT for EXPTIME, and NE for NEXPTIME. In the second row,  $\mathcal{ALL}_a$  stands for  $\mathcal{ALL}$  with acyclic TBoxes. Notice that the results shown in the table refer to only a few of the many possible ontology languages. These are intended only as a sample of the results that one can obtain from the arguments presented in this paper. Moreover, we believe that the upper bounds for  $\mathcal{SHOIQ}$  can be further reduced to NEXPTIME in all cases.

## 7 Related Work

Probabilistic extensions of DLs and other formalisms for representing and reasoning with uncertainty have been studied for several decades. We refer the interested reader to a deep, although slightly outdated survey [31].

One of the first attempts for combining BNs and DLs was P-CLASSIC [29], which extended the CLASSIC system through probability distributions over the interpretation domain. The more recent PR-OWL [21] 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.

The approaches that most resemble ours are perhaps the Bayesian extension of DL-Lite [22] and the extensions of DL with distribution semantics DISPONTE [38, 39]. 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 [38, 39], 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. In practical terms, our Bayesian ontology languages are a strict generalization of the distribution semantics from [39].

Moreover, we provide (tight) complexity bounds for reasoning in these logics. A different approach is followed in [34], where  $\mathcal{EL}$  axioms are attached with weights

that indirectly determine a probability distribution. The main difference is that in [34], these weights are not restricted to be between 0 and 1, and even if they do, they do not correspond directly to the probability of the axiom to hold.

Abstracting from the different logical constructors used, the logic in [22] 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 [22], 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  $\mathcal{B}\mathcal{E}\mathcal{L}$ . For example, using the semantics proposed in [22], the TBox  $\{\langle A \sqsubseteq B : \{x\} \rangle, \langle A \sqsubseteq B : \{\neg x\} \rangle\}$  is inconsistent: every valuation must either satisfy  $x$  or  $\neg x$ . In the first case, the interpretation must satisfy  $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ ; but then, due to the second axiom in the TBox, it must also satisfy  $\neg x$ , which is impossible. The dual argument holds for interpretations that satisfy  $\neg x$ . With our semantics, as given in Definition 8, the TBox given above is equivalent to  $\{\langle A \sqsubseteq B : \emptyset \rangle\}$ ; i.e., to requiring the subsumption relation to hold in every context.

We have proposed two kinds of algorithms for deciding probabilistic subsumption in  $\mathcal{B}\mathcal{E}\mathcal{L}$ . One makes repeated calls to a classical  $\mathcal{E}\mathcal{L}$  reasoner and a BN inference engine to find the probability of all the restricted TBoxes that entail the desired subsumption. The second produces a larger BN from which the probabilities of the subsumption relations can be derived through standard probabilistic inferences. A different approach, implemented in the system BORN [18], is to reduce the reasoning problem to a probabilistic logic program, and use an existing efficient system for these programs, like ProbLog [24].

Although its representation as a hypergraph, and its translation into a DAG are new, the main insights that give rise to our proof structures and their properties have a long history. It is well known that every Horn theory can be represented as a hypergraph, where each Horn clause is a directed hyperedge from the set of all the premises to the conclusion [36]. Under this view, the proof structure corresponds to a simplified version of the formula proposed by Sebastiani and Vescovi for axiom pinpointing in  $\mathcal{E}\mathcal{L}^+$  [41, 42].

## 8 Conclusions

We have introduced  $\mathcal{B}\mathcal{E}\mathcal{L}$ , a Bayesian extension of the light-weight DL  $\mathcal{E}\mathcal{L}$  for expressing uncertainty attached to contextual knowledge of the domain. The main purpose of  $\mathcal{B}\mathcal{E}\mathcal{L}$  is to represent certain knowledge that is dependent on an uncertain context. The uncertainty of the different contexts is expressed by a probability distribution represented via a Bayesian network. One of the main insights obtained studying this logic is that reasoning can be decoupled into the logical part and the probabilistic part. This means that standard  $\mathcal{E}\mathcal{L}$  reasoners and BN inference engines can be used to perform all the studied inferences in  $\mathcal{B}\mathcal{E}\mathcal{L}$ .

The goal of the paper was to provide tight complexity bounds for different reasoning problems in  $\mathcal{B}\mathcal{E}\mathcal{L}$ . All these results are summarized in the first row of Table 5. Interestingly, most of the upper bounds obtained are consequence of a reduction from  $\mathcal{B}\mathcal{E}\mathcal{L}$  into standard inference problems over a Bayesian network. To obtain this translation, we exploit the so-called *proof structure*: a directed hypergraph that preserves the information of all the possible subsumption relations that

can be derived by application of the completion algorithm. Such a proof structure is then modified into a DAG, which serves as the basis for a newly constructed BN. We can thus exploit state-of-the-art BN engines to reason in  $\mathcal{BEL}$  efficiently. We have also extended the notion of a Bayesian logic to arbitrary ontology languages, which include but are not restricted to all standard DLs. Exploiting the decoupling between the logical and the probabilistic components of these logics, we provided upper bounds for the reasoning problems associated to them, which in many cases are tight (see Table 5). These upper bounds can be implemented via a black-box algorithm that makes repeated calls to a standard logical reasoner and a BN inference tool. It is a matter of future work to develop more efficient algorithms that exploit the properties of each ontology language, as we have done for  $\mathcal{EL}$  already.

There exist many directions for future work. The most obvious and urgent of them is the implementation and evaluation of a  $\mathcal{BEL}$  reasoner. Towards this goal, we can further optimize the reasoner BORN [18], or extend the methods for axiom-pinpointing that exploit the proof structure [1, 2, 32] to consider also the probabilistic contexts. In the same direction, we will study further optimizations and special cases where the complexity of reasoning can be further decreased. For instance, it is well known that the complexity of probabilistic inferences in BNs depends strongly on the structure of the underlying DAG [23]. Unfortunately, the construction proposed in Section 5.2 produces a highly interconnected DAG for which these inferences are hard. It would thus be interesting to find restrictions or new constructions that preserve the structure of the original BN.

Here we have considered subsumption as the standard reasoning problem in  $\mathcal{EL}$ . In case there is knowledge about individuals of the application domain, it is also relevant to answer *queries* about these individuals and their properties. We have started studying the problem of probabilistic query answering in  $\mathcal{BEL}$  [17]. However, there still remain many open questions that need to be addressed.

**Acknowledgements** İsmail İlkan Ceylan is supported by DFG within the Research Training Group “RoSI” (GRK 1907) and Rafael Peñaloza is partially supported by DFG within the Cluster of Excellence “cfAED;” part of this work was developed while Rafael Peñaloza was affiliated with TU Dresden, and the Center for Advancing Electronics Dresden, Germany. We would like to thank Gabriele Kern-Isberner for fruitful discussions regarding the material and conditional implications, and Franz Baader for his suggestions for improving the quality and presentation of the paper.

## References

1. Arif F, Mencía C, Ignatiev A, Manthey N, Peñaloza R, Marques-Silva J (2016) BEACON: An efficient SAT-based tool for debugging EL+ ontologies. In: Proceedings of SAT 2016, to appear
2. Arif MF, Mencía C, Marques-Silva J (2015) Efficient MUS enumeration of Horn formulae with applications to axiom pinpointing. CoRR abs/1505.04365, URL <http://arxiv.org/abs/1505.04365>
3. Baader F, Peñaloza R (2010) Automata-based axiom pinpointing. Journal of Automated Reasoning 45(2):91–129
4. Baader F, Peñaloza R (2010) Axiom pinpointing in general tableaux. Journal of Logic and Computation 20(1):5–34

5. Baader F, Brandt S, Lutz C (2005) Pushing the  $\mathcal{EL}$  envelope. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05), Morgan Kaufmann
6. Baader F, Lutz C, Suntisrivaraporn B (2006) CEL—a polynomial-time reasoner for life science ontologies. In: Furbach U, Shankar N (eds) Proceedings of the 3rd International Joint Conference on Automated Reasoning (IJCAR 2006), Springer-Verlag, Lecture Notes in Artificial Intelligence, vol 4130, pp 287–291
7. Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF (eds) (2007) The Description Logic Handbook: Theory, Implementation, and Applications, 2nd edn. Cambridge University Press
8. Baader F, Peñaloza R, Suntisrivaraporn B (2007) Pinpointing in the description logic  $\mathcal{EL}^+$ . In: Proceedings of the 30th German Conference on Artificial Intelligence (KI 2007), Springer-Verlag, Osnabrück, Germany, Lecture Notes in Artificial Intelligence, vol 4667, pp 52–67
9. Baader F, Knechtel M, Peñaloza R (2009) A generic approach for large-scale ontological reasoning in the presence of access restrictions to the ontology’s axioms. In: Proceedings of the 8th International Semantic Web Conference (ISWC 2009), Springer-Verlag, Lecture Notes in Computer Science, vol 5823, pp 49–64
10. Baader F, Knechtel M, Peñaloza R (2012) Context-dependent views to axioms and consequences of semantic web ontologies. *Journal of Web Semantics* 12–13:22–40
11. Beigel R, Reingold N, Spielman DA (1995) PP is closed under intersection. *Journal of Computer and System Sciences* 50(2):191–202
12. Brandt S (2004) Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In: Proceedings of the 16th European Conference on Artificial Intelligence (ECAI-2004), IOS Press, pp 298–302
13. Ceylan II (2013) Context-sensitive Bayesian description logics. Master’s thesis, Dresden University of Technology, Germany
14. Ceylan II, Peñaloza R (2014) The Bayesian description logic BEL. In: Proceedings of the 7th International Joint Conference on Automated Reasoning (IJCAR 2014), Springer-Verlag, Lecture Notes in Computer Science, vol 8562, pp 480–494
15. Ceylan II, Peñaloza R (2014) Bayesian Description Logics. In: Proceedings of the 27th International Workshop on Description Logics (DL’14), CEUR-WS, CEUR Workshop Proceedings, vol 1193, pp 447–458
16. Ceylan II, Peñaloza R (2014) Tight complexity bounds for reasoning in the description logic  $\mathcal{BEL}$ . In: Proceedings of the 14th European Conference on Logics in Artificial Intelligence (JELIA 2014), Springer-Verlag, Lecture Notes in Computer Science, vol 8761, pp 77–91
17. Ceylan II, Peñaloza R (2015) Probabilistic query answering in the Bayesian description logic  $\mathcal{BEL}$ . In: Proceedings of the 9th International Conference on Scalable Uncertainty Management (SUM 2015), Springer-Verlag, Lecture Notes in Artificial Intelligence, vol 9310, pp 1–15
18. Ceylan II, Mendez J, Peñaloza R (2015) The Bayesian ontology reasoner is BORN! In: Proceedings of the 4th International Workshop on OWL Reasoner Evaluation (ORE-2015), CEUR-WS, CEUR Workshop Proceedings, vol 1387, pp 8–14

19. Cook SA (1971) The complexity of theorem-proving procedures. In: Proc. Third Annual ACM Symposium on Theory of Computing, ACM, New York, NY, USA, STOC '71, pp 151–158, DOI 10.1145/800157.805047, URL <http://doi.acm.org/10.1145/800157.805047>
20. Cooper GF (1990) The computational complexity of probabilistic inference using Bayesian belief networks (research note). *Artificial Intelligence* 42(2-3):393–405
21. da Costa PCG, Laskey KB, Laskey KJ (2008) PR-OWL: A Bayesian ontology language for the semantic web. In: *Uncertainty Reasoning for the Semantic Web I, URSW 2005-2007*, Springer, Lecture Notes in Computer Science, vol 5327, pp 88–107
22. d'Amato C, Fanizzi N, Lukasiewicz T (2008) Tractable reasoning with Bayesian description logics. In: *Proc. Second International Conference on Scalable Uncertainty Management (SUM 2008)*, Springer, Lecture Notes in Computer Science, vol 5291, pp 146–159
23. Darwiche A (2009) *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press
24. De Raedt L, Kimmig A, Toivonen H (2007) ProbLog: A probabilistic prolog and its application in link discovery. In: *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07)*, AAAI Press, pp 2462–2467
25. Downey RG, Fellows M (1999) *Parameterized Complexity*. Monographs in Computer Science, Springer-Verlag
26. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. *Journal of Computational Biology* 7:601–620
27. Gill JT (1977) Computational Complexity of Probabilistic Turing Machines. *SIAM Journal on Computing* 6(4):675–695
28. Kazakov Y, Krötzsch M, Simančík F (2014) The incredible ELK: From polynomial procedures to efficient reasoning with  $\mathcal{EL}$  ontologies. *Journal of Automated Reasoning* 53:1–61
29. Koller D, Levy AY, Pfeffer A (1997) P-classic: A tractable probabilistic description logic. In: *Proc. 14th National Conference on Artificial Intelligence (AAAI-97)*, AAAI Press, pp 390–397
30. Littman ML, Majercik SM, Pitassi T (2001) Stochastic Boolean satisfiability. *Journal of Automated Reasoning* 27(3):251–296
31. Lukasiewicz T, Straccia U (2008) Managing uncertainty and vagueness in description logics for the semantic web. *Journal of Web Semantics* 6(4):291–308, DOI 10.1016/j.websem.2008.04.001
32. Manthey N, Peñaloza R (2015) Exploiting SAT technology for axiom pinpointing. Tech. Rep. 15-05, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden
33. Mendez J (2012) jcel: A modular rule-based reasoner. In: *Proceedings of the 1st International Workshop on OWL Reasoner Evaluation (ORE-2012)*, CEUR-WS, CEUR Workshop Proceedings, vol 858
34. Niepert M, Noessner J, Stuckenschmidt H (2011) Log-linear description logics. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, IJCAI/AAAI, pp 2153–2158, URL <http://ijcai.org/papers11/Papers/IJCAI11-359.pdf>

35. Park JD, Darwiche A (2004) Complexity results and approximation strategies for MAP explanations. *Journal of Artificial Intelligence Research* 21:101–133
36. Peñaloza R, Sertkaya B (2010) On the complexity of axiom pinpointing in the EL family of description logics. In: Lin F, Sattler U, Truszczynski M (eds) *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*, AAAI Press
37. Pearl J (1988) *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann Publishers
38. Riguzzi F, Bellodi E, Lamma E, Zese R (2012) Epistemic and statistical probabilistic ontologies. In: *Proc. 8th Int. Workshop on Uncertainty Reasoning for the Semantic Web (URSW-12)*, CEUR-WS, vol 900, pp 3–14
39. Riguzzi F, Bellodi E, Lamma E, Zese R (2015) Probabilistic description logics under the distribution semantics. *Semantic Web Journal* 6(5):477–501, DOI 10.3233/SW-140154, URL <http://dx.doi.org/10.3233/SW-140154>
40. Scutari M, Howell P, Balding DJ, Mackay I (2014) Multiple quantitative trait analysis using bayesian networks. *Genetics* 198:129–137
41. Sebastiani R, Vescovi M (2009) Axiom pinpointing in lightweight description logics via Horn-SAT encoding and conflict analysis. In: *Proceedings of the 22nd International Conference on Automated Deduction*, Springer-Verlag, *Lecture Notes in Computer Science*, vol 5663, pp 84–99
42. Sebastiani R, Vescovi M (2015) Axiom pinpointing in large  $\mathcal{EL}+$  ontologies via SAT and SMT techniques. Tech. Rep. DISI-15-010, University of Trento, Italy, URL [http://disi.unitn.it/%7Erseba/elsat/elsat\\_techrep.pdf](http://disi.unitn.it/%7Erseba/elsat/elsat_techrep.pdf)
43. Shimony SE (1994) Finding maps for belief networks is NP-hard. *Artificial Intelligence* 68(2):399–410
44. Toda S (1989) On the Computational Power of PP and +P. In: *Proc. of the 30th Annual Symposium on Foundations of Computer Science*, IEEE, pp 514–519