MASTER'S THESIS

Context-Sensitive Bayesian Description Logics

by

İsmail İlkan Ceylan

in partial fulfilment of the requirements

for the degree of Master of Science

in

International Center for Computational Logic

Faculty of Computer Science

TECHNISCHE UNIVERSITÄT DRESDEN

Supervisor:
Prof. Dr.-Ing. Franz Baader

Advisor:
Dr. rer. nat. Rafael Peñaloza

Dresden, November 2013

Declaration of Authorship

Hereby, I certify that this dissertation titled "*Context-Sensitive Bayesian Description Logics*" is the result of my own work and is written by me. Any help that I have received in my research work has been acknowledged. I certify that I have not used any auxiliary sources and literature except those cited in the thesis.

Author	:	İsmail İlkan Ceylan
Matriculation number	:	3833232
Date of submission	:	19.11.2013
Signature	:	

“Pür eda pür cefa, pek küçük pek güzel... ”

Dede Efendi

Abstract

Research embracing context-sensitivity in the domain of knowledge representation (KR) has been scarce, especially when the context is uncertain. The current study deals with this problem from a logic perspective and provides a framework combining Description Logics (DLs) with Bayesian Networks (BNs). In this framework, we use BNs to describe contexts that are semantically linked to classical DL axioms.

As an application scenario, we consider the Bayesian extension $\mathcal{BEL}$ of the lightweight DL $\mathcal{EL}$. We define four reasoning problems; namely, precise subsumption, positive subsumption, certain subsumption and finding the most likely context for a subsumption. We provide an algorithm that solves the precise subsumption in PSPACE. Positive subsumption is shown to be NP-complete and certain subsumption coNP-complete. We present a completion-like algorithm, which is in EXPTIME, to find the most likely context for a subsumption.

The scenario is then generalised to Bayesian extensions of classic-valued, monotonic DLs, where precise entailment, positive entailment, certain entailment and finding the most likely context for an entailment are defined as lifted reasoning problems. It is shown that precise entailment, positive entailment and certain entailment can be solved by generalising the algorithms developed for the corresponding reasoning problems in $\mathcal{BEL}$. Lastly, the complexities of these problems are shown to be bound with the complexity of entailment checking in the underlying $\mathcal{DL}$, provided this is PSPACE-hard.

Keywords: Description Logics, Bayesian Networks, Context-based Reasoning, Uncertainty, Semantic Web

Acknowledgements

I would like to express my deepest gratitude to Dr. Rafael Peñaloza, my research advisor, for guiding me so patiently and for providing useful critiques at each and every phase of this study. His enthusiasm always encouraged me to go on. No doubt that without the challenging scientific discussions we held, this thesis would not be possible. I would like to extend my thanks to Prof. Franz Baader, the pioneer of the research group on Description Logics, for providing me the opportunity to work in this highly-qualified research environment. I also deeply appreciate all the helpful advices, comments and guidance that Dr. Anni Yasmin Turhan has provided.

Doing my Master's in the International Center for Computational Logic has been very gratifying in that it not only provided funding to me for two years, but also equipped me with solid knowledge on logics and widened my horizon. On behalf of the International Center for Computational Logic, I would like to thank its director, Prof. Steffen Hölldobler. I cannot forget to thank those people who have made life significantly easier with their efforts. Our ambitious secretaries, Sylvia, Julia and Kerstin, were always there when needed, helping us with every tiny detail. Although I cannot name them individually, I also want to thank all staff members of the "Sächsische Landesbibliothek -Staats- und Universitätsbibliothek" and "Studentenwerk Dresden" for offering a decent environment to study, live and eat.

Amongst others, hopefully, this work can also serve as an answer to all of my friends, who have been trying to figure out what I have been doing in the last two years. On the personal side, I want to say my earnest thanks to my parents, Osman and Anakadın Ceylan, for their support throughout this process; and to my brother Ramazan Volkan Ceylan, for always being there to motivate me. Lastly, I want to thank Bahar, who harvested my life, and this work in particular, with the new words she has nourished me with.

Contents

Dedicated to my family...

Chapter 1

Introduction

Knowledge Representation (KR) KR is the field of studies that aims to "develop formalisms for providing high-level descriptions of the world that can be effectively used to build intelligent applications" as described by Brachman and Nardi [1]. There are a number of questions that can be raised with respect to this quotation, such as what is meant by a formalism, what do the high level descriptions correspond to and most importantly when can an application be classified as intelligent. In KR, *intelligent applications* are restricted to those which are able to reason about the knowledge and infer implicit knowledge from the explicitly represented one. A *formalism* in KR is realized by means of a language that has a well-defined syntax and an unambiguous semantics. Lastly, a *high-level description* refers to concentrating only on the relevant aspects of the domain of interest.

The domain of interest can cover any part of the real world as well as any hypothetical system about which one desires to represent knowledge. Typically, KR formalisms enable us to form statements. A set of statements that describes a domain is then called a knowledge base or ontology, upon which reasoning can be performed by making use of the formal semantics of the statements.

Description Logics (DLs) DLs [1] comprise a widely-studied family of knowledge representation formalisms. This family is characterized by the use of concepts, roles and individuals. In DLs, concepts describe an abstract or generic idea generalised from particular individuals and roles describe the binary relations between such individuals.

Informally, the syntax of DLs is based on named concepts and named roles which are given a priori. Concept descriptions, in short concepts, are defined inductively from named concepts and named roles by making use of the constructors provided by the particular DL. Consequently, the particular DL is mainly a result of the choice over the constructors to be used. Basic constructors in DLs are conjunction ($\sqcap$), disjunction ($\sqcup$), negation ($\neg$), existential restriction ($\exists$) and value restriction ($\forall$). Top concept ($\top$) and bottom concept ($\bot$) are special constructors, that denote concepts on their own.

The semantics of DLs is defined by interpreting the named concepts as sets and the named roles as relations. The interpretations are then generalised to all concepts by additionally setting the intended meaning of the constructors. Intuitively, given two concepts C and D, the conjunct $(C \sqcap D)$ defines a new concept that contains only the individuals that belong to both C and D, that is, an intersection in set theory. Analogously, disjunction is interpreted as union and negation as complement. $\exists r.C$ describes the set of individuals that are related to an individual in concept C via the role r. $\forall r.C$ describes the set of individuals x where for all y that is related to x via r, y belongs to the concept C. Finally, $\top$ is the set consisting of all individuals in the domain and $\bot$ is the empty set.

The language described so far is called the concept language and is mainly a result of the choices on the constructors to be used. However, even if the concept language is fixed, it is still possible to form different types of ontologies in DLs. This stems from the fact that there exists different operators for forming statements, or axioms, as is widely-known, that aid in realizing different types of ontologies. DL ontologies mostly consist of a terminological box (TBox) and an assertional box (ABox), where the former contains axioms in the concept level, and the latter on the individual level.

ABox axioms, usually called assertions, are units of statements about individuals. Considering the concept *Parent*, the assertion *Parent(Hans)* states that the individual *Hans* is a *Parent*. TBox axioms, on the other hand, are units of statements on the concept level, i.e., they allow us to form statements about concepts. General TBoxes are finite collections of axioms, known as general concept inclusions, that are built with the subsumption operator ($\sqsubseteq$). For instance, ($Father \sqsubseteq Parent$) is a simple general concept inclusion that is interpreted as a sub-class relation between *Father* and *Parent*, meaning that all individuals belonging to the concept *Father* also belong to the concept *Parent*.

Besides being theoretical formalisms, DLs are also promising in practice. Given their numerous applications in various domains of knowledge, it would be safe to say that DLs have achieved a considerable success in KR. For instance, DLs are the underlying theory behind the Semantic Web, which is the effort to structure the content on the web a priori in a machine-processable way. To date, various DL formalisms have been proposed based on the different combinations of the constructors. Several reasoning problems that allow to deduce consequences such as subclass and instance relationships from the DL ontologies have been defined and solved by providing various algorithms. Furthermore, plenty of tools that perform reasoning over DL ontologies have been developed such as Hermit [2], Fact++ [3] and CEL [4].

Uncertainty Having partial knowledge over a domain is in most cases, if not in all, unavoidable. As a result, the question of whether and to what extent this fact is taken into account in KR is important. A number of probabilistic approaches, dating back to the mathematical investigation of the probabilistic logics in 1854 [5], have been developed to address uncertainty. Since then, probabilistic extensions of different logics have been studied widely. Nilsson [6] provides a probabilistic logic which can be considered as the basis for its successors. Fagin et al. [7] introduce a probabilistic logic, closely related to Nilsson's probabilistic logic, that allows to specify linear inequalities over events. Frisch and Haddawy [8] generalise the propositional version of Nilsson's probabilistic logic by incorporating conditional probabilities and they introduce inference rules upon which producing proofs to explain how conclusions are drawn is possible. Lukasiewicz [9–12] introduces locally complete inference rules for probabilistic deduction from taxonomic and probabilistic knowledge-bases over conjunctive events.

Rather recently, probabilistic DLs are being investigated, see [13] for a survey. Although there are plenty of different approaches, some properties of probabilistic DLs are mostly shared. Most importantly, all formalisms need to satisfy the requirements of the probability calculus. This is commonly realized through the so-called probabilistic world semantics. In this semantics, the unknown part of the knowledge domain is instantiated, which then corresponds to a world. Each world is associated with a probability that determines its likelihood. Exhaustively doing this for every possible world enables us to draw conclusions over the domain as a whole. Intuitively, worlds are abstractions through which we complete our partial knowledge over the domain.

Probabilistic DLs have a larger degree of freedom than classic DLs in the sense that there are other choices than the underlying DL. We recall the related work in the relevant chapters by putting an emphasis on these choices.

Context-sensitivity Another challenge in KR is to distinguish the context of the knowledge domain. Suppose we want to state that plants make photosynthesis. Assuming that existential restrictions are available in the DL we can form an axiom as follows.

$$\mathit{Plant} \sqsubseteq \exists \mathit{make}.\mathit{Photosynthesis}$$

Now suppose that we want to put restrictions on this axiom; for instance, we want to say that this is the case only if there light is available. Naïvely, this can be done by manipulating the axiom. On the other hand, the main intention here is to describe a context for the axiom. Hence, it is important to keep the axiom as general as possible and annotate it with certain context, which also enables reuse of a particular axiom in different contexts. In his "Notes on formalizing context" [14], McCarthy has discussed the importance of contexts in knowledge domains and proposed a theory to formalize them. The basis of his theory is to be able to make assertions in the form of $ist(c, p)$ which is interpreted as the proposition p is true in the context c, that is; p is true if the context is c. Adopting this to our example we can form an axiom as follows.

$$\mathit{ist}(\mathit{Plant} \sqsubseteq \exists \mathit{makes}.\mathit{Photosynthesis}, \{\mathit{Light\ is\ available}\})$$

A context in this particular example consists of one proposition, i.e., light being available. It is worth to noting that we may as well benefit from a set of propositions to describe a context. Such KR formalisms that are able to take into account the context-dependent character of knowledge are called context-sensitive formalisms. Accordingly, reasoning in context-sensitive formalisms is called context-sensitive reasoning or context-based reasoning. Context-based reasoning in DLs is recently being investigated; see Klarman [15] for a framework on context based reasoning in DLs and related work. Intuitively, contexts describe another dimension for knowledge domains. Then, the main questions are how to represent the new dimension and what properties are required to be met by this representation.

Context-sensitivity over uncertain domains This work aims to provide a framework that is context-sensitive where the context is uncertain. We can illustrate this further by building on the previous example of the axiom stating that plants make photosynthesis provided that there is light in the environment. On the other hand, we do not have certain information about the appearance of light in the environment, that is, having light in the environment comes with some probability.

The main approach held in this work is to combine two KR formalisms, namely, DLs and Bayesian networks (BNs) with a dimensional perspective. We provide axioms with two dimensions, the first dimension being the standard DL axiom, and the second one being the context, which is a set defined w.r.t. a BN. Semantically, these are connected with an "if condition" as in McCarthy's formalism.

As an application scenario we consider the lightweight DL $\mathcal{EL}$ and extend it to Bayesian $\mathcal{EL}$, which we abbreviate in the remainder of the text as $\mathcal{BEL}$. The thesis is organised in the following way: In Chapter 2 we define the preliminaries where we introduce individual formalisms that are combined. This is followed by Chapter 3, in which the syntax and semantics of $\mathcal{BEL}$ is given. Next, we define reasoning problems for $\mathcal{BEL}$ and provide algorithms to solve them. These results together with the computational complexity results of the algorithms are collected in Chapter 4. In Chapter 5, we generalise $\mathcal{BEL}$ to the DL family, denoted as $\mathcal{BDL}$, and discuss reasoning in $\mathcal{BDL}$. We conclude by summarising our results and contributions to the existing literature as well as discussing potential directions for future work.

Chapter 2

Preliminaries

We propose a new formalism that combines the DL $\mathcal{EL}$ with BNs. It is important to know the insights of these formalisms before moving on to the new logic, where these formalisms represent the dimensions.

The first formalism that we concentrate on is $\mathcal{EL}$. $\mathcal{EL}$ uses only the constructors $\top$, $\sqcap$ and $\exists$. It is a lightweight DL and reasoning in $\mathcal{EL}$ is known to be tractable, i.e., polynomial. It has drawn the attention of researchers because tractable extensions of $\mathcal{EL}$, particularly $\mathcal{EL}^{++}$ [16], are already sufficient for knowledge representation and reasoning tasks over various knowledge domains. These extensions have been used in bio-medical ontologies such as the Systematized Nomenclature of Medicine [17], The Gene Ontology [18] and large parts of the Galen Medical Knowledge Base [19]. Furthermore, $\mathcal{EL}$ underlies the OWL 2 EL profile, which has been standardised by World Wide Web Consortium (W3W) in 2009.

The term "Bayesian" comes from Thomas Bayes who is the pioneer in proposing to update beliefs. His ideas, however, have not been published during his life-time. After Bayes's death, Richard Price has significantly edited and published his notes [20]. What we name "Bayesian" in the modern sense has been mostly formulated by Laplace in his work titled "Théorie analytique des probabilités" [21]. In his "Essai philosophique sur les probabilités" [22], Laplace has derived the general form of "Bayes's theorem" that we know today. Furthermore, he suggested a system for inductive reasoning based on probability, which laid the ground for today's "Bayesian statistics". BNs are "Bayesian" in the sense that they employ the principles of Bayesian inference (based on Bayes's theorem) but they do not necessarily imply a commitment to "Bayesian statistics". Indeed, it is common to use frequentists methods. BNs can be seen as an automated mechanism for applying the principles of Bayesian inference to more complex problems by making use of directed acyclic graphs.

The idea of combining DLs with BNs goes back to Koller et. al. [23], where authors extend the old description logic CLASSIC to P-Classic by making use of BNs. The reasoning problems defined for P-Classic are then reduced to inference in BNs. In addition to the lack of support for assertional knowledge, P-Classic introduces certain restrictions, which make the reasoning easier, even polynomial if BNs are restricted to poly-trees as their underlying data structure. Our framework differs from this not only w.r.t. the underlying DL, but also in the contextual setting that we define, upon which the two formalisms are semantically linked.

2.1 $\mathcal{EL}$

$\mathcal{EL}$ is a DL, which uses only $\top$, $\sqcap$ and $\exists$ as constructors. Assume that a countably infinite supply of concept names, usually denoted as A and B, and of role names, usually denoted as r and s, are available. Let $\mathcal{N_C}$ and $\mathcal{N_R}$ be disjoint sets of concept and role names, respectively. Then the syntax and semantics of $\mathcal{EL}$ are defined as follows:

Definition 2.1 (Syntax)**.** The set of $\mathcal{EL}$-*concepts* is inductively defined as follows:

- $\top$ and $A \in \mathcal{N_C}$ are concepts.
- If C and D are concepts then so is $C \sqcap D$.
- If C is a concept and $r \in \mathcal{N_R}$ then $\exists r.C$ is a concept.

The semantics of $\mathcal{EL}$ is given in terms of interpretations. An interpretation consists of an interpretation function and a non-empty interpretation domain. We first define the interpretation function over the elements of the sets $\mathcal{N_C}$ and $\mathcal{N_R}$ and then extend it to all concepts.

Definition 2.2 (Semantics)**.** An *interpretation* is a pair $(\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where $\Delta^\mathcal{I}$ is a non-empty *domain* and $\cdot^\mathcal{I}$ is an *interpretation function* such that:

- $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ for all $A \in \mathcal{N_C}$
- $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$ for all $r \in \mathcal{N_R}$

The interpretation function $\cdot^\mathcal{I}$ is extended to all concepts as follows:

- $\top^\mathcal{I} = \Delta^\mathcal{I}$
- $(C \sqcap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$
- $(\exists r.C)^\mathcal{I} = \{x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I} : (x, y) \in r^\mathcal{I} \wedge y \in C^\mathcal{I}\}$

We have defined the concept language of $\mathcal{EL}$. In $\mathcal{EL}$ it is also important to capture the terminological knowledge of application domains in a structured way. This is achieved through terminological boxes.

Definition 2.3 (TBox)**.** A *GCI* is an expression of the form $C \sqsubseteq D$, where C, D are concepts. An interpretation $\mathcal{I}$ satisfies the GCI $C \sqsubseteq D$ iff $C^\mathcal{I} \subseteq D^\mathcal{I}$. An $\mathcal{EL}$ *terminological box (TBox)* $\mathcal{T}$ is a finite set of GCIs. An interpretation $\mathcal{I}$ is a *model* of the TBox $\mathcal{T}$ iff it satisfies all the GCIs in $\mathcal{T}$.

The main reasoning service in $\mathcal{EL}$ is subsumption checking, i.e., deciding the sub-concept relations between given concepts based on their semantic definitions. Subsumption is formally defined as follows.

Definition 2.4 (Subsumption)**.** C is *subsumed by* D w.r.t. the TBox $\mathcal{T}$ ($C \sqsubseteq_{\mathcal{T}} D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$.

It has been shown that subsumption can be decided in $\mathcal{EL}$ in polynomial time [16] by an algorithm (known as completion algorithm), which we will refer to in Chapter 4. This concludes our remarks on the description logic $\mathcal{EL}$. In Section 2.2, we elaborate on Bayesian networks which is the other constituent of our hybrid formalism.

2.2 Bayesian Networks

Bayesian networks [24], also known as Belief networks, provide a probabilistic graphical model, which has a directed acyclic graph (DAG) as its underlying data structure. Each node in the graph represents a *random variable*, and the set of edges in the graph represent probabilistic dependencies among random variables.

Definition 2.5 (Bayesian network)**.** A *Bayesian network* is a pair $\mathcal{BN} = (G, \Theta)$ where

- $G = (\mathcal{V}, E)$ is a *directed graph* with $\mathcal{V}$ as the set of random variables and E as the set of dependencies among the random variables
- Θ is a set of *conditional probability distributions* $\mathcal{P}_{\mathcal{BN}}$, one for each node $X \in \mathcal{V}$ given its parents:

$$\Theta = \{\mathcal{P}_{\mathcal{BN}}(X = x | Pa(X) = x') | X \in \mathcal{V}\}$$

  where x and x' represent the valuations of X and $Pa(X)$ (parents of X) respectively.

This definition of BNs encodes the *local Markov property*, which states that each variable is conditionally independent of its non-descendants given its parent variables. This is an important property of BNs since it suffices to check the parental variables to determine a conditional probability distribution.

We assume that the random variables are discrete. For BNs with discrete random variables, each *conditional probability* is represented by a *conditional probability table (CPT).* Each such CPT stores the conditional probabilities of a random variable for each possible combination of the values of its parent nodes. For each random variable X and its valuations $\mathcal{D}(X)$ in the BN, a conditional probability table is formed as follows:

- For each combination of the values of parental nodes a row is introduced.
- For each row in the table, exactly as many columns as $|\mathcal{D}(X)|$ is introduced.
- Each cell in the table contains a probability value such that the sum of the probabilities in a row add up to 1.

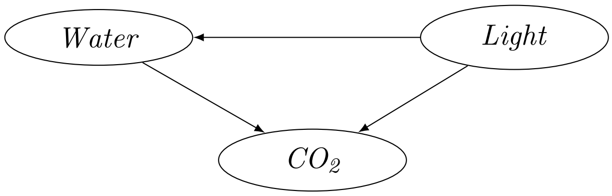
In what follows, we describe a situation that can be modelled by BNs.

Example 2.6. *Suppose*[1] *that we are in a hypothetical environment in which Light, Water and CO_2 are the parameters. We know the conditional probability distributions and we want*

[1]The content as well as the probabilities provided in Example 2.6 are for illustrative purposes. Nothing further is intended.

Light	*Water=t*	*Water=f*
t	0.70	0.30
f	0.60	0.40

Light=t	*Light=f*
0.60	0.40



Light	*Water*	CO_2*=t*	CO_2*=f*
t	*t*	0.90	0.10
t	*f*	0.80	0.20
f	*t*	0.70	0.30
f	*f*	0.50	0.50

FIGURE 2.1: A simple Bayesian network

to compute the probability of CO_2 *being available in the environment. Figure 2.1 shows a Bayesian network motivated by this situation.*

The network consists of three random variables. For the random variable CO_2*, we have a domain of values* $\mathcal{D}(CO_2) = \{t, f\}$*. The probability of* CO_2 *being true is denoted as* $\mathcal{P}_{\mathcal{BN}}(CO_2 = t)$*. There are conditional probability tables next to each random variable. For example, the probability of* CO_2 *being true given that there exists both Light and Water is 0.90. To calculate the probability of* CO_2 *being true, we need to introduce the full joint probability distribution and the worlds.*

A valuation of a random variable is usually called an *event* and the probability of every possible event can be calculated w.r.t the full joint probability distribution.

Definition 2.7. (Full joint probability distribution) Given a Bayesian network $\mathcal{BN}$, the *full joint probability distribution* is the probability of all possible events represented by the random variables and their valuations in $\mathcal{BN}$. *Full joint probability distribution* can be calculated [25] as:

$$\mathcal{P}_{\mathcal{F}}(V = v) = \prod_{X \in \mathcal{V}} \mathcal{P}_{\mathcal{BN}}(X = v | Pa(X) = v)$$

where v is restricted to the valuations of X and to the valuations of $Pa(X)$ respectively.

Given Definition 2.7, we can calculate the probability of any set of events. Calculating the probability of a given set of events, also called *Bayesian inference*, is the main inference problem in Bayesian networks. Given a set of variables E from a Bayesian network $\mathcal{BN}$ and their valuations $e \in \mathcal{D}(E)$ it holds that:

$$\mathcal{P}_{\mathcal{BN}}(E = e) = \sum_{E=e} \mathcal{P}_{\mathcal{F}}(V = v)$$

It has been shown that Bayesian inference is NP-hard [26]. On the other hand, if the set of events is complete, i.e., it contains a valuation for each random variable then inference can be done in polynomial time.

Worlds	*Light*	*Water*	CO_2	*Value*
1	t	t	t	0.378
2	t	f	t	0.144
3	f	t	t	0.168
4	f	f	t	0.080
5	t	t	f	0.042
6	t	f	f	0.036
7	f	t	f	0.072
8	f	f	f	0.080

TABLE 2.1: Full joint probability distribution

Definition 2.8. (World) A *World* is a set of events such that for all $X_i \in \mathcal{V}$ it contains exactly one valuation $X_i = x_i$ where $x_i \in \mathcal{D}(X_i)$.

For every world, the calculation of the probability of the world can be done simply by applying the *chain rule* over the specified valuations of the random variables. Prior to applying the chain rule it is required to order the variables such that ancestor variables appear before the successors, which can be done in polynomial time. Suppose an ordered world $(X_1 = x_1, ..., X_n = x_n)$ is given where $\mathcal{V} = \{X_i | 1 \leq i \leq n\}$ and $x_i \in \mathcal{D}(X_i)$ then:

$$\mathcal{P}_{\mathcal{BN}}(X_1 = x_1, ..., X_n = x_n) = \prod_{1 \leq i < j \leq n} \mathcal{P}_{\mathcal{BN}}(X_i = x_i | X_j = x_j)$$

There are exponentially many worlds in BNs. Suppose that all the worlds are given. Since BNs provide complete probabilistic models and there is no other possible world, the following is an immediate result

Lemma 2.9. *Given a* $\mathcal{BN} = (G, \Theta)$ *and all of its worlds* $(\mathcal{W}_1, ..., \mathcal{W}_k)$ *it holds that:*

$$\sum_{1 \leq i \leq k} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) = 1.$$

Given these we can return to our motivating example and calculate the probability of CO_2 being true.

Example 2.10 (Continuation of Example 2.6)**.** *The full joint probability distribution can be seen as a large CPT that results from the multiplication of all CPTs. Hence, we get the full joint probability distribution shown in Table 2.1 for our minimal example. Summing the rows where* CO_2*=t yields the probability 0.77, which denotes the probability of* CO_2 *being available in the environment.*

We have revisited the basic definitions of BNs. Lastly, we define the notion of a context as a set of events. As a consequence, the probabilities of this set of events determines the likelihood of the context.

Definition 2.11. (Context) A *Context* is a set of events $(X_1 = x_1, ..., X_n = x_n)$.

A world is a special context that contains a valuation for every random variable. Essentially, each context describes a set of worlds. The joint probability of each world can be calculated by applying the chain rule as explained. This calculation can be performed in polynomial time and will later be important in defining reasoning procedures for the new logic. Additionally, it is important to note that for the contexts that are not worlds this does not hold. This stems from the fact that the contexts that are not worlds are partial, i.e., there are variables that are not mapped to a valuation which causes the calculation to be propagated over all possible valuations.

This concludes Chapter 2. We can build now on this background and define the new logic $\mathcal{BEL}$, which aims to extend the description logic $\mathcal{EL}$ with uncertain contexts, provided by the Bayesian network. This will be elaborated in Chapter 3 in detail.

Chapter 3

Bayesian $\mathcal{EL}$

After the establishment of P-Classic [23], combining probabilistic graphical models such as BNs and Markov Networks (MN) with DLs has been spread. A similar attempt to P-Classic has been made by Yelland [27] to extend the DL $\mathcal{FL}$ which uses the constructors $\top$, $\sqcup$ and $\forall$. This approach has also had reflections in the lightweight DLs such as $\mathcal{EL}$ and DL-Lite. Niepert, Noessner and Stuckenschmidt present a probabilistic extension of $\mathcal{EL}^{++}$ [28] by making use of the Markov logic that is based on Markov Logic Networks [29]. Another formalism that extends $\mathcal{EL}^{++}$ with MNs is given in [30], which basically translates $\mathcal{EL}^{++}$ to first order logic (FOL) to make use of the reasoning services that have been defined for the first order Markov logic. Our work differs from these previous attempts at least in two significant ways. Firstly, as mentioned above, we use BNs. In addition, Markov Logic Networks provide a model in which the first order axioms are annotated with weights (probabilities) only whereas we have contexts as annotation which evaluate to a probability.

Another proposal for a probabilistic $\mathcal{EL}$ has been given in [31]. This framework is applied to different logics as well as to $\mathcal{EL}$ which is named *Prob-*$\mathcal{EL}$. This is closely related to the one in [30] mostly in the sense that they both use the principles of an existing probabilistic FOL developed in [7]. Yet, differently from the $\mathcal{BEL}$ proposed in this thesis, *Prob-*$\mathcal{EL}$ does not make use of any probabilistic graphical model. Additionally, it is different from our proposal in the sense that it extends the concepts with probabilities whereas we extend the axioms.

It will be fruitful at this point to briefly discuss two previous works from the literature, which can be considered as closest to the framework we propose here. The first one, called BDL-Lite [32] extends the lightweight DL DL-Lite with BNs. BDL-Lite ontologies contain axioms that are combinations of DL-Lite axioms and a set of random variable valuations, taken from a BN. This is close to our dimensionality perspective and its syntax can be easily generalised towards other DLs. On the other hand, these two dimensions are connected to each other via an “if and only if condition”, which does not suit to our contextual setting. This “if and only if condition” leads to some inconsistencies since it forces a set of random variables to be satisfied given that a certain axiom is satisfied. The second work that we highly benefit from is by Bellodi et. al. [33]. They use a semantics which they call DIstribution Semantics for Probabilistic ONTologiEs (DISPONTE) and provide a framework that can be applied to the DL family. The key idea here is to annotate each axiom with a probability as in BDL-Lite. As opposed to BDL-Lite they combine the dimensions with an "if condition" and they consider all random variables to be independent. In effect, they do not benefit from conditional probabilities or any graphical model. They also provide a system called BUNDLE [34] that performs inference over probabilistic OWL DL ontologies by making use of other DL reasoners as oracles.

To the best of our knowledge, there is no work on context-based reasoning over uncertain domains. On the other hand, DISPONTE provides a semantics which is capable of taking contexts into account. It is sufficient to treat annotations as contexts. Instead of using annotations that are assumed to be independent we make use of BNs as in BDL-Lite. Yet, in contrast to BDL-Lite, we consider DISPONTE as the underlying semantics. As a result of this, our framework allows forming axioms which hold provided that a certain context holds. As a basis to our framework we extend the description logic $\mathcal{EL}$ to $\mathcal{BEL}$. The syntax and semantics of $\mathcal{BEL}$ is defined in a way that it can be extended to more expressive logics.

3.1 Syntax and Semantics of $\mathcal{BEL}$

The $\mathcal{BEL}$ concept language is defined exactly as the $\mathcal{EL}$ concept language. The semantics of $\mathcal{BEL}$ is based on probabilistic world semantics, which defines a probability distribution over a set of contextual interpretations. In the following, we first define the contextual interpretations and then introduce the probabilistic world semantics of $\mathcal{BEL}$. The contextual interpretations extend the standard $\mathcal{EL}$ interpretations by additionally mapping the random variables from the Bayesian network to their domain of values.

Definition 3.1 (Contextual interpretations)**.** Given a Bayesian network $\mathcal{BN} = (G, \Theta)$ defined over the set $G = (\mathcal{V}, E)$, a *contextual interpretation* is a triple $(\Delta^{\mathcal{I}_\mathcal{C}}, \mathcal{V}^{\mathcal{I}_\mathcal{C}}, \cdot^{\mathcal{I}_\mathcal{C}})$ where $\Delta^{\mathcal{I}_\mathcal{C}}$ is a non-empty domain disjoint from the domain $\mathcal{V}^{\mathcal{I}_\mathcal{C}}$ and $\cdot^{\mathcal{I}_\mathcal{C}}$ is an *interpretation function* such that:

- $A^{\mathcal{I}_\mathcal{C}} \subseteq \Delta^{\mathcal{I}_\mathcal{C}}$ for all $A \in \mathcal{N}_\mathcal{C}$
- $r^{\mathcal{I}_\mathcal{C}} \subseteq \Delta^{\mathcal{I}_\mathcal{C}} \times \Delta^{\mathcal{I}_\mathcal{C}}$ for all $r \in \mathcal{N}_\mathcal{R}$
- $X^{\mathcal{I}_\mathcal{C}} \in \mathcal{V}^{\mathcal{I}_\mathcal{C}}$ for all random variables $X \in \mathcal{V}$.

The *interpretation function* $\cdot^{\mathcal{I}_\mathcal{C}}$ is extended to all concepts as follows:

- $\top^{\mathcal{I}_\mathcal{C}} = \Delta^{\mathcal{I}_\mathcal{C}}$
- $(C \sqcap D)^{\mathcal{I}_\mathcal{C}} = C^{\mathcal{I}_\mathcal{C}} \cap D^{\mathcal{I}_\mathcal{C}}$
- $(\exists r.C)^{\mathcal{I}_\mathcal{C}} = \{x \in \Delta^{\mathcal{I}_\mathcal{C}} \mid \exists y \in \Delta^{\mathcal{I}_\mathcal{C}} : (x, y) \in r^{\mathcal{I}_\mathcal{C}} \wedge y \in C^{\mathcal{I}_\mathcal{C}}\}$

The intuition behind introducing the contextual interpretations is to be able to distinguish the contexts. Hence, we define the satisfiability of a context w.r.t. a given contextual interpretation.

Definition 3.2 (Satisfiability of a context)**.** Let $\mathcal{BN} = (G, \Theta)$ be a Bayesian network over the graph $G = (\mathcal{V}, E)$, $\psi = (X_1 = x_1, ..., X_n = x_n)$ a context, and $\mathcal{I_C}$ a contextual interpretation. $\mathcal{I_C}$ satisfies the context ψ, denoted as $\mathcal{I_C} \models \psi$, iff $\{X_1^{\mathcal{I_C}} = x_1, ..., X_n^{\mathcal{I_C}} = x_n\}$

In $\mathcal{BEL}$, GCIs are replaced with probabilistic general concept inclusions (PGCIs), which is a pair consisting of a standard GCI and a possibly empty context. In what follows, we formally define the notion of PGCIs, give the satisfaction condition of a PGCI and define the $\mathcal{BEL}$ TBoxes and ontologies.

Definition 3.3 (*PGCI, TBox and Ontology*)**.** A *probabilistic general concept inclusion (PGCI)* is an expression of the form $(C \sqsubseteq D : \psi)$, where C, D are concepts and ψ is a context. A contextual interpretation $\mathcal{I_C}$ satisfies the probabilistic general concept inclusion $(C \sqsubseteq D : \psi)$, denoted as $\mathcal{I_C} \models (C \sqsubseteq D : \psi)$, iff the following implication holds:

$$(\mathcal{I_C} \models \psi) \rightarrow (\mathcal{I_C} \models (C \sqsubseteq D))$$

A $\mathcal{BEL}$ *TBox* $\mathcal{T}$ is a finite set of *PGCIs* that extend the standard GCIs with the contexts. A $\mathcal{BEL}$ *ontology* $\mathcal{K}$ is a pair $(\mathcal{T}, \mathcal{BN})$ where $\mathcal{BN}$ is a Bayesian network and $\mathcal{T}$ is a $\mathcal{BEL}$ TBox which has contexts that contain only events from $\mathcal{BN}$.

Note that the contextual interpretations stem from McCarthy's *ist* function. Example 3.4 demonstrates the contextual semantics of $\mathcal{BEL}$.

Example 3.4. *(Contextual interpretations of PGCIs) It is possible to formulate an axiom α which states that plants make photosynthesis provided that light, CO_2 and water are available as:*

$$\alpha = (Plant \sqsubseteq \exists make.Photosynthesis, \{Light{=}t,\ CO_2{=}t,\ Water{=}t\ \})$$

Suppose that the three contextual interpretations $\mathcal{I_C}_1$, $\mathcal{I_C}_2$ and $\mathcal{I_C}_3$ are given as specified below.

- $\mathcal{I_C}_1 = (\Delta^{\mathcal{I_C}_1}, \mathcal{V}^{\mathcal{I_C}_1}, \cdot^{\mathcal{I_C}_1})$, $\mathcal{I_C}_2 = (\Delta^{\mathcal{I_C}_2}, \mathcal{V}^{\mathcal{I_C}_2}, \cdot^{\mathcal{I_C}_2})$, $\mathcal{I_C}_3 = (\Delta^{\mathcal{I_C}_3}, \mathcal{V}^{\mathcal{I_C}_3}, \cdot^{\mathcal{I_C}_3})$
- $\Delta^{\mathcal{I_C}_1} = \Delta^{\mathcal{I_C}_2} = \Delta^{\mathcal{I_C}_3} = \{a, v\}$
- $Plant^{\mathcal{I_C}_1} = Plant^{\mathcal{I_C}_2} = Plant^{\mathcal{I_C}_3} = \{a\}$
- $Photosynthesis^{\mathcal{I_C}_1} = Photosynthesis^{\mathcal{I_C}_2} = Photosynthesis^{\mathcal{I_C}_3} = \{v\}$
- $make^{\mathcal{I_C}_1} = \{(a, v)\}$, $make^{\mathcal{I_C}_2} = make^{\mathcal{I_C}_3} = \{\}$
- $Light^{\mathcal{I_C}_1} = Light^{\mathcal{I_C}_2} = Light^{\mathcal{I_C}_3} = t$
- $CO_2^{\mathcal{I_C}_1} = CO_2^{\mathcal{I_C}_2} = CO_2^{\mathcal{I_C}_3} = t$
- $Water^{\mathcal{I_C}_1} = Water^{\mathcal{I_C}_2} = t$, $Water^{\mathcal{I_C}_3} = f$

We analyse whether the given contextual interpretations satisfy α or not. By definition, a contextual interpretation $\mathcal{I_C}$ satisfies α iff it satisfies the following implication:

$$(\mathcal{I_C} \models \{Light = t, CO_2 = t, Water = t\}) \rightarrow (\mathcal{I_C} \models (Plant \sqsubseteq \exists make.Photosynthesis))$$

The contextual interpretation $\mathcal{I_C}_1$ satisfies this implication since it satisfies both the classical part of the axiom and the context. Hence, we conclude that $\mathcal{I_C}_1$ satisfies α. On the other hand, $\mathcal{I_C}_2$ does satisfy the context but not the classical part since there is no individual in Plant that is connected to Photosynthesis with make relation. Therefore, $\mathcal{I_C}_2$ does not *satisfy α. Because of the same reason as in $\mathcal{I}_{\mathcal{C}2}$, $\mathcal{I}_{\mathcal{C}3}$ does not satisfy the classical part of the axiom either. However, it also does not satisfy the context since Water is interpreted as false. Hence, the implication holds, i.e., $\mathcal{I}_{\mathcal{C}3}$ satisfies α.*

Having defined the contextual interpretations it is possible to define the semantics of $\mathcal{BEL}$. The semantics of $\mathcal{BEL}$ is based on the probabilistic world semantics which defines a probability distribution over a set of contextual interpretations.

Definition 3.5 (Probabilistic interpretations)**.** A *probabilistic interpretation* $\mathcal{I}_{\mathcal{P}}$ is a pair $(\Phi, \mathcal{P}r)$ where Φ is a set of contextual interpretations and $\mathcal{P}r$ is a probability distribution over Φ such that: $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\} \subseteq \Phi$, $1 \leq i \leq n$ for a finite n and for every contextual interpretation $\mathcal{I}_{\mathcal{C}} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\}$ it holds that $\mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = 0$. A *probabilistic interpretation* $\mathcal{I}_{\mathcal{P}}$ is a *model of* $\mathcal{T}$ iff for each $(C \sqsubseteq D : \psi) \in \mathcal{T}$ it satisfies:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1$$

$\mathcal{I}_{\mathcal{P}}$ is *consistent with a Bayesian network* $\mathcal{BN}$ iff for every world $\mathcal{W}$ in $\mathcal{BN}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models \mathcal{W}} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W})$$

A probabilistic interpretation $\mathcal{I}_{\mathcal{P}}$ is a *model of a $\mathcal{BEL}$ ontology* $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ iff $\mathcal{I}_{\mathcal{P}}$ is a model of $\mathcal{T}$ and consistent with $\mathcal{BN}$. A $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ is *satisfiable* iff it has a probabilistic model.

Note that we allow infinitely many contextual interpretations but only finitely many of them are assigned a positive probability by the probability distribution. Suppose that the Bayesian network in Figure 2.1 is given. We extend Example 3.4 and show a construction of a probabilistic interpretation w.r.t. a $\mathcal{BEL}$ ontology. Furthermore, we show that the constructed probabilistic interpretation is a model of the given ontology.

Example 3.6. *(A model construction) Let $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be a $\mathcal{BEL}$ ontology where $\mathcal{T} = \{\alpha\}$ and $\mathcal{BN}$ be the Bayesian network shown in Figure 2.1. We define a set of contextual interpretations $\Phi = \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}8}\}$ in the following way. Let $\mathcal{I}_{\mathcal{C}j} = (\Delta^{\mathcal{I}_{\mathcal{C}j}}, \mathcal{V}^{\mathcal{I}_{\mathcal{C}j}}, \cdot^{\mathcal{I}_{\mathcal{C}j}})$ where $\Delta^{\mathcal{I}_{\mathcal{C}j}} = \{a, v\}$, $Plant^{\mathcal{I}_{\mathcal{C}j}} = \{a\}$, $Photosynthesis^{\mathcal{I}_{\mathcal{C}j}} = \{v\}$, $make^{\mathcal{I}_{\mathcal{C}j}} = \{(a, v)\}$ for $1 \leq j \leq 8$. $\mathcal{V}^{\mathcal{I}_{\mathcal{C}j}}$, $1 \leq j \leq 8$, and the probability distribution $\mathcal{P}r$ over $\mathcal{I}_{\mathcal{C}j}$ are as follows.*

- $Light^{\mathcal{I}_{\mathcal{C}1}} = t, Water^{\mathcal{I}_{\mathcal{C}1}} = t, {CO_2}^{\mathcal{I}_{\mathcal{C}1}} = t$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}1}) = 0.378$
- $Light^{\mathcal{I}_{\mathcal{C}2}} = t, Water^{\mathcal{I}_{\mathcal{C}2}} = f, {CO_2}^{\mathcal{I}_{\mathcal{C}2}} = t$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}2}) = 0.144$
- $Light^{\mathcal{I}_{\mathcal{C}3}} = f, Water^{\mathcal{I}_{\mathcal{C}3}} = t, {CO_2}^{\mathcal{I}_{\mathcal{C}3}} = t$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}3}) = 0.168$
- $Light^{\mathcal{I}_{\mathcal{C}4}} = f, Water^{\mathcal{I}_{\mathcal{C}4}} = f, {CO_2}^{\mathcal{I}_{\mathcal{C}4}} = t$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}4}) = 0.080$
- $Light^{\mathcal{I}_{\mathcal{C}5}} = t, Water^{\mathcal{I}_{\mathcal{C}5}} = t, {CO_2}^{\mathcal{I}_{\mathcal{C}5}} = f$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}5}) = 0.042$
- $Light^{\mathcal{I}_{\mathcal{C}6}} = t, Water^{\mathcal{I}_{\mathcal{C}6}} = f, {CO_2}^{\mathcal{I}_{\mathcal{C}6}} = f$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}6}) = 0.036$
- $Light^{\mathcal{I}_{\mathcal{C}7}} = f, Water^{\mathcal{I}_{\mathcal{C}7}} = t, {CO_2}^{\mathcal{I}_{\mathcal{C}7}} = f$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}7}) = 0.072$
- $Light^{\mathcal{I}_{\mathcal{C}8}} = f, Water^{\mathcal{I}_{\mathcal{C}8}} = f, {CO_2}^{\mathcal{I}_{\mathcal{C}8}} = f$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}8}) = 0.080$

This construction ensures that every contextual interpretation in Φ satisfies the classical part of α, i.e., ($Plant \sqsubseteq \exists make.Photosynthesis$). We define $\mathcal{I_P} = (\Phi, \mathcal{P}r)$ as a probabilistic interpretation. We first check whether $\mathcal{I_P}$ is a model of $\mathcal{T}$. Since there is only one axiom in $\mathcal{T}$ it is enough to check whether the sum of the probabilities of the contextual interpretations satisfying α add up to 1. The following shows that this is the case:

$$\sum_{\mathcal{I_C}_i \models \alpha} \mathcal{P}r(\mathcal{I_C}_i) = \mathcal{P}r(\mathcal{I_C}_1) + ... + \mathcal{P}r(\mathcal{I_C}_8) = 1$$

Hence, $\mathcal{I_P}$ is a model of $\mathcal{T}$. Next, we check whether the probabilistic interpretation $\mathcal{I_P}$ is consistent with the Bayesian network $\mathcal{BN}$, i.e., for every world $\mathcal{W}$ in $\mathcal{BN}$ it holds that:

$$\sum_{\mathcal{I_C}_i \models \mathcal{W}} \mathcal{P}r(\mathcal{I_C}_i) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W})$$

Each of the contextual interpretations that we have defined correspond to one world in $\mathcal{BN}$ and these are all of the worlds. We have assigned the probabilities to the contextual interpretations such that they comply with the probabilities of the respective worlds. Therefore we get:

$$\sum_{\mathcal{I_C}_i \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{\mathcal{I_C}_j \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I_C}_j) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$$

Hence, $\mathcal{I_P}$ is consistent with $\mathcal{BN}$. As a result we conclude that $\mathcal{I_P}$ is a model of the ontology.

For every model, we have forced the probability distributions of the models to be consistent with the probabilities of the worlds in the Bayesian network. This ensures that the joint probability distribution defined by the Bayesian network and the probability distribution defined by the model are consistent. This is proven in Lemma 3.7.

Lemma 3.7. *Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, for every context ψ and for all models $\mathcal{I_P} = (\Phi, \mathcal{P}r)$ of $\mathcal{K}$ with $\mathcal{I_C}_i \in \Phi$ where $\mathcal{P}r(\mathcal{I_C}_i) > 0$ it holds that:*

$$\sum_{\mathcal{I_C}_i \models \psi} \mathcal{P}r(\mathcal{I_C}_i) = \mathcal{P}_{\mathcal{BN}}(\psi)$$

Proof. Result follows from the fact that the probability of a context is the sum of its probabilities in each world. Hence, we get:

$$\sum_{\mathcal{I_C}_i \models \psi} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{\mathcal{I_C}_i \models \mathcal{W}, \psi \subseteq \mathcal{W}} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{\psi \subseteq \mathcal{W}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}) = \mathcal{P}_{\mathcal{BN}}(\psi)$$

□

So far, we introduced the syntax and semantics of $\mathcal{BEL}$. Moreover, we have shown that it is possible to represent contextual knowledge over uncertain domains in $\mathcal{BEL}$. For instance, we have formulated an axiom which states that "plants make photosynthesis provided that there exists light, water and CO_2 in the environment", where the likelihood of the environment is determined by the Bayesian network. In Chapter 4, we introduce several reasoning problems for $\mathcal{BEL}$ and provide algorithms for solving them.

Chapter 4

Reasoning in $\mathcal{BEL}$

Most DLs deal with two basic reasoning procedures, *consistency checking* and *subsumption checking*. In $\mathcal{EL}$, on the other hand, the main reasoning procedure is subsumption checking since any $\mathcal{EL}$ ontology is known to be consistent. We show that this property of $\mathcal{EL}$ is preserved when extending to $\mathcal{BEL}$, i.e., every $\mathcal{BEL}$ ontology is consistent.

We introduce subsumption checking as a reasoning problem in $\mathcal{BEL}$. Furthermore, we define different types of subsumption problems w.r.t. the probabilities and provide algorithms for solving them.

4.1 Consistency of any $\mathcal{BEL}$ ontology

A $\mathcal{BEL}$ ontology is consistent iff it has a model. We hereby show a construction of such a model $\mathcal{I_P}$ for a given $\mathcal{BEL}$ ontology. Intuitively, a Bayesian network describes a world by each possible valuation of all of its variables. There are only finitely many different worlds, which can be enumerated as $(\mathcal{W}_1, ..., \mathcal{W}_k)$. Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ we define the probabilistic interpretation $\mathcal{I_P} = (\{\mathcal{I_C}_1, ..., \mathcal{I_C}_k\}, \mathcal{P}r)$ where:

- $\mathcal{I_C}_i = (\{a_i\}, \mathcal{V}^{\mathcal{I_C}_i}, \cdot^{\mathcal{I_C}_i})$
- $A^{\mathcal{I_C}_i} = \{a_i\}$, $r^{\mathcal{I_C}_i} = \{(a_i, a_i)\}$, $\{X = X^{\mathcal{I_C}_i} | X \in \mathcal{V}\} = \mathcal{W}_i$ for all $A \in \mathcal{N_C}$ and $r \in \mathcal{N_R}$
- $\mathcal{P}r(\mathcal{I_C}_i)$=$\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i)$

We prove the consistency of any $\mathcal{BEL}$ ontology by showing that $\mathcal{I_P}$ is indeed a model of the given ontology.

Theorem 4.1. *Every $\mathcal{BEL}$ ontology is consistent.*

Proof. Consider the probabilistic interpretation $\mathcal{I_P} = (\{\mathcal{I_C}_1, ..., \mathcal{I_C}_k\}, \mathcal{P}r)$ that has been constructed from an arbitrary $\mathcal{BEL}$ ontology $\mathcal{K}$. Bayesian networks providing complete probabilistic models ensure that the sum of the probabilities of all the worlds add up to 1:

$$\sum_{1 \leq i \leq k} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{1 \leq i \leq k} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) = 1$$

Each $\mathcal{I}_{\mathcal{C}i}$ is an interpretation that satisfies every axiom. As a result of this we get that for every axiom $(C \sqsubseteq D : \psi) \in \mathcal{T}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \sum_{1 \leq i \leq k} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1.$$

Hence $\mathcal{I}_{\mathcal{P}}$ is a model of $\mathcal{T}$. By construction, for every world $\mathcal{W}_j$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \sum_{\mathcal{I}_{\mathcal{C}j} \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I}_{\mathcal{C}j}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$$

Hence $\mathcal{I}_{\mathcal{P}}$ is consistent with $\mathcal{BN}$. As a result, we get that $\mathcal{I}_{\mathcal{P}}$ is a model of the given ontology $\mathcal{K}$. Since we have shown a construction of a model for an arbitrary $\mathcal{BEL}$ ontology, we get that every $\mathcal{BEL}$ ontology is consistent. □

4.2 Subsumption in $\mathcal{BEL}$

This section introduces subsumption-based reasoning services. In $\mathcal{BEL}$, we are able to express probabilities of subsumptions upon which the different types of subsumption problems are defined.

Definition 4.2 (Probability of a subsumption)**.** Let $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be a $\mathcal{BEL}$ ontology and $\mathcal{I}_{\mathcal{P}} = (\Phi, \mathcal{P}r)$ a probabilistic interpretation. The probability of the subsumption $(C \sqsubseteq D : \psi)$ w.r.t. $\mathcal{I}_{\mathcal{P}}$, denoted as $\mathcal{P}(C \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} D : \psi)$ is:

$$\mathcal{P}(C \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} D : \psi) = \sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i})$$

The probability of $(C \sqsubseteq D : \psi)$ w.r.t. $\mathcal{K}$, denoted as $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D : \psi)$ is defined as the infimum over the probabilities of the subsumption w.r.t. all models:

$$\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D : \psi) = \inf\{\mathcal{P}(C \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} D : \psi) | \mathcal{I}_{\mathcal{P}} \models \mathcal{K}\}$$

$\mathcal{P}(C \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} D)$ is written short for $\mathcal{P}(C \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} D : \{\})$ and $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D)$ for $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D : \{\})$.

We can also determine whether a given axiom $(C \sqsubseteq D : \psi)$ is a consequence of an ontology. Intuitively, a consequence is an exact entailment, i.e., for all models of the ontology an axiom is a consequence iff every contextual interpretation which is assigned a positive probability in the model satisfies the axiom.

Definition 4.3 (PGCI as a consequence)**.** Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, $(C \sqsubseteq D : \psi)$ is a consequence of $\mathcal{K}$, denoted as $(C \sqsubseteq_{\mathcal{K}} D : \psi)$ iff it holds that $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D : \psi) = 1$.

We also define the notion of a partial interpretation w.r.t. a world and a subsumption. The intuition is to detect the interpretations that partially satisfy a subsumption in a given world. This notion is particularly useful in proving that the probability of a subsumption w.r.t. an ontology must be a sum of the probabilities of a set of worlds.

Definition 4.4 (Partial interpretations)**.** A probabilistic interpretation $\mathcal{I_P} = (\Phi, \mathcal{P}r)$ is *partial* w.r.t. a world $\mathcal{W}$ and a subsumption $(C \sqsubseteq D)$ if there exists two conceptual interpretations $\mathcal{I_C}_a, \mathcal{I_C}_b \in \Phi$ satisfying the following conditions:

$$\begin{aligned} &\mathcal{I_C}_a \models \mathcal{W},\ \mathcal{I_C}_a \models (C \sqsubseteq D),\ \mathcal{P}r(\mathcal{I_C}_a) > 0 \\ &\mathcal{I_C}_b \models \mathcal{W},\ \mathcal{I_C}_b \not\models (C \sqsubseteq D),\ \mathcal{P}r(\mathcal{I_C}_b) > 0 \end{aligned} \tag{4.1}$$

We use the notion of partial interpretations to show that the infimum in the definition of the probability of the subsumption w.r.t. an ontology is also a minimum.

Lemma 4.5. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be an ontology C, D concepts. There exists a model $\mathcal{I_P}$ of $\mathcal{K}$ such that $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = \mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D)$.*

Proof. Suppose that there are k worlds $\mathcal{W}_j$, $1 \leq j \leq k$. We first show that $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D)$ has the form $\sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$ where $\overline{\mathcal{W}}$ is a set of worlds . By the definition of subsumption w.r.t. an ontology we know that $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = \inf\{\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) | \mathcal{I_P} \models \mathcal{K}\}$. Hence, by the definition of infimum, we can find a model $\mathcal{I_P}^+$ of $\mathcal{K}$ such that:

$$\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}^+} D) < \sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j) + \min\{\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0 | 1 \leq i \leq k\}$$

Let $\mathcal{I_P}^+ = (\Phi^+, \mathcal{P}r^+)$ with $\{\mathcal{I_C}_1^+, ..., \mathcal{I_C}_m^+\} \subseteq \Phi^+$ and for every contextual interpretation $\mathcal{I_C}^+ \notin \{\mathcal{I_C}_1^+, ..., \mathcal{I_C}_m^+\}$ it holds that $\mathcal{P}r(\mathcal{I_C}^+) = 0$. This is well-defined since there can be infinitely many contextual interpretations but only finitely many of them (in this case $\{\mathcal{I_C}_1^+, ..., \mathcal{I_C}_m^+\}$) can be assigned a positive probability by the probability distribution $\mathcal{P}r$. We rewrite $\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}^+} D)$ as follows:

$$\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}^+} D) = \sum_{\mathcal{I_C}_i^+ \models \mathcal{W}_1,\ \mathcal{I_C}_i^+ \models (C \sqsubseteq D)} \mathcal{P}r(\mathcal{I_C}_i^+) + \cdots + \sum_{\mathcal{I_C}_i^+ \models \mathcal{W}_k,\ \mathcal{I_C}_i^+ \models (C \sqsubseteq D)} \mathcal{P}r(\mathcal{I_C}_i^+)$$

We show that we can construct a model $\mathcal{I_P}$ such that the value $\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D)$ equals to $\sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$. We show this by induction on the number of the worlds $\mathcal{W}$ for which $\mathcal{I_P}^+$ is partial w.r.t. the subsumption $(C \sqsubseteq D)$, i.e., for which there exists two conceptual interpretations $\mathcal{I_C}_a^+$, $\mathcal{I_C}_b^+ \in \Phi^+$ satisfying the conditions in (4.1). We assume that the worlds are enumerated such that the worlds satisfying (4.1) precede all the worlds that do not satisfy (4.1).

Induction Base: For $(s = 0)$, we know that there is no world satisfying (4.1). Hence we take $\mathcal{I_P} = \mathcal{I_P}^+$ and get that:

$$\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) = \sum_{\mathcal{I_C}_i \models C \sqsubseteq D, \mathcal{I_C}_i \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{\mathcal{I_C}_i \models C \sqsubseteq D, \mathcal{I_C}_i \models \mathcal{W}_j} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j) = \sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$$

Induction Hypothesis: For $(s = l - 1)$, there exists a model $\mathcal{I_P}' = (\Phi', \mathcal{P}r')$ where

- $\{\mathcal{I_C}_1', ..., \mathcal{I_C}_n'\} \subseteq \Phi'$, $1 \leq i \leq n$ and
- for every contextual interpretation $\mathcal{I_C}' \notin \{\mathcal{I_C}_1', ..., \mathcal{I_C}_n'\}$ it holds that $\mathcal{P}r(\mathcal{I_C}') = 0$.

such that $\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) = \sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$.

Induction: Take $(s = l)$. We know by the induction hypothesis that there exists a model $\mathcal{I_P}'$ such that $\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}'} D) = \sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$. If the model $\mathcal{I_P}'$ does not satisfy (4.1) w.r.t. $(C \sqsubseteq D)$ and $\mathcal{W}_l$, then the result follows directly since $\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_l)$ is either added or not, where both of the cases yield the result. Assume that it satisfies (4.1), i.e., there exists at least two conceptual interpretations $\mathcal{I_C}_a^{'}, \mathcal{I_C}_b^{'} \in \Phi'$ satisfying the conditions set in (4.1). This means that for $\mathcal{W}_l$ there exists at least one contextual interpretation $\mathcal{I_C}_b^{'} \in \Phi'$ with $\mathcal{P}_{\mathcal{BN}}(\mathcal{I_C}_b^{'}) > 0$ where $(\mathcal{I_C}_b^{'} \models \mathcal{W})$ and $(\mathcal{I_C}_b^{'} \not\models C \sqsubseteq D)$. This shows that the subsumption $(C \sqsubseteq D)$ does not need to hold in $\mathcal{W}_l$. We use this fact to construct a probabilistic interpretation $\mathcal{I_P}$ of the ontology. Let $\mathcal{I_P} = (\Phi, \mathcal{P}r)$ where:

- $\Phi = \{\mathcal{I_C}_1, ..., \mathcal{I_C}_n\}$,
- $\mathcal{I_C}_i = \mathcal{I_C}_i^{'}$ for all $\mathcal{I_C}_i^{'} \not\models \mathcal{W}$
- $\mathcal{I_C}_i = \mathcal{I_C}_i^{'}$ for all $\mathcal{I_C}_i^{'} \models \mathcal{W}$ where $\mathcal{W} \neq \mathcal{W}_l$,
- $\mathcal{I_C}_i = \mathcal{I_C}_b^{'}$ for all $\mathcal{I_C}_i^{'} \models \mathcal{W}_l$,
- $\mathcal{P}r(\mathcal{I_C}_i) = \mathcal{P}r(\mathcal{I_C}_i^{'})$ for all $1 \leq i \leq n$.

This construction ensures that every contextual interpretation with positive probability satisfying $\mathcal{W}_l$ is equal to $\mathcal{I_C}_b^{'}$. Since $\mathcal{I_C}_b^{'}$ does not satisfy the subsumption $(C \sqsubseteq D)$ we get:

$$\sum_{\mathcal{I_C}_i \models \mathcal{W}_l, \mathcal{I_C}_i \models (C \sqsubseteq D)} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{\mathcal{I_C}_i = \mathcal{I_C}_b^{'}, \mathcal{I_C}_i \models (C \sqsubseteq D)} \mathcal{P}r(\mathcal{I_C}_i) = 0.$$

Hence, the subsumption does not hold in $\mathcal{W}_l$. As a result, we get that $\mathcal{P}r(C \sqsubseteq_{\mathcal{I_P}} D)$ has the form $\sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$. We still need to show that the probabilistic interpretation $\mathcal{I_P}$ is a model of the ontology. We know that $\mathcal{P}r(\mathcal{I_C}_i) = \mathcal{P}r(\mathcal{I_C}_i^{'})$ for $1 \leq i \leq n$ and $\mathcal{I_P}'$ is a model. Hence, for every world $\mathcal{W}_j$ it holds that:

$$\sum_{\mathcal{I_C}_i \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{\mathcal{I_C}_i^{'} \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I_C}_i^{'}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j) \tag{4.2}$$

While constructing $\mathcal{I_P}$, we have replaced the contextual interpretations in $\mathcal{I_P}'$ that satisfy $\mathcal{W}_l$ and $(C \sqsubseteq D)$ with $\mathcal{I_C}_b^{'}$. We know that $\mathcal{I_C}_b^{'}$ does satisfy $\mathcal{W}_l$ but does not satisfy the subsumption $(C \sqsubseteq D)$. Additionally, $\mathcal{I_C}_b^{'}$ has a positive probability. Therefore, it must satisfy all axioms that hold in $\mathcal{W}_l$ since otherwise $\mathcal{I_P}'$ would not be a model. We use this fact to show that $\mathcal{I_P}$ is a model of the TBox $\mathcal{T}$. Since $\mathcal{I_P}'$ is a model, for all axioms $(E \sqsubseteq F : \psi) \in \mathcal{T}$ it holds that:

$$\sum_{\mathcal{I_C}_i^{'} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i^{'}) = \sum_{\mathcal{I_C}_i^{'} \not\models \mathcal{W},\ \mathcal{I_C}_i^{'} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i^{'}) + \sum_{\mathcal{I_C}_i^{'} \models \mathcal{W},\ \mathcal{I_C}_i^{'} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i^{'}) = 1 \tag{4.3}$$

Since for all $\mathcal{I_C}_i^{'} \not\models \mathcal{W}$ it holds that $\mathcal{I_C}_i = \mathcal{I_C}_i^{'}$ and $\mathcal{P}r(\mathcal{I_C}_i) = \mathcal{P}r(\mathcal{I_C}_i^{'})$, we can rewrite the first summand in Equation 4.3 as:

$$\sum_{\mathcal{I_C}_i \not\models \mathcal{W},\ \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) \tag{4.4}$$

Likewise, we can rewrite the second summand in Equation 4.3 as:

$$\sum_{\mathcal{I_C}_i \models \mathcal{W},\, \mathcal{W} \neq \mathcal{W}_l,\, \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) + \sum_{\mathcal{I_C}_i \models \mathcal{W}_l,\, \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) \tag{4.5}$$

We know that $\mathcal{I_C}_i = \mathcal{I_C}'_b$ for all $\mathcal{I_C}'_i \models \mathcal{W}_l$. Since all $\mathcal{I_C}'_b$ must satisfy all axioms in $\mathcal{W}_l$, we can rewrite (4.5) as:

$$\sum_{\mathcal{I_C}_i \models \mathcal{W},\, \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) \tag{4.6}$$

Since the summands in Equation 4.3 can be replaced by (4.4) and (4.6) respectively we get the following:

$$\sum_{\mathcal{I_C}_i \not\models \mathcal{W},\, \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) + \sum_{\mathcal{I_C}_i \models \mathcal{W},\, \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) = 1. \tag{4.7}$$

Then for all axioms $(E \sqsubseteq F : \psi) \in \mathcal{T}$ the following holds:

$$\sum_{\mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) = \sum_{\mathcal{I_C}_i \not\models \mathcal{W},\, \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) + \sum_{\mathcal{I_C}_i \models \mathcal{W},\, \mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}_i) = 1 \tag{4.8}$$

From (4.8), we get that $\mathcal{I_P}$ is a model of $\mathcal{T}$. Combining this with the statement in Equation (4.2) yields that $\mathcal{I_P}$ is a model of $\mathcal{K}$. This shows that $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D)$ has the form $\sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$ where $\overline{\mathcal{W}}$ is a subset of the set of all worlds. Given this, Lemma 4.5 holds as a result of the following facts.

We know that there are finitely many worlds. Therefore, there are only finitely many sums over finitely many worlds where $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = \sum_{\mathcal{W}_j \in \overline{\mathcal{W}}} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j)$. Hence we get that the infimum in the definition of subsumption w.r.t. an ontology is indeed a minimum, i.e., $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = \inf\{\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) | \mathcal{I_P} \models \mathcal{K}\} = \min\{\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) | \mathcal{I_P} \models \mathcal{K}\}$. This shows that there exists a model $\mathcal{I_P}$ of $\mathcal{K}$ where $\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) = \mathcal{P}(C \sqsubseteq_{\mathcal{K}} D)$. □

As it has been mentioned, it is possible to define various types of subsumption problems in $\mathcal{BEL}$. This stems from the fact that we can represent uncertainty and hence we can have degrees of a subsumption to hold. We define the following four types of reasoning problems w.r.t. these degrees.

Definition 4.6 (Subsumption types)**.** Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and the concepts C, D we say that C is *precisely subsumed* by D w.r.t. $\mathcal{K}$ with probability p if $p = \mathcal{P}(C \sqsubseteq_{\mathcal{K}} D)$. Additionally, we define two special cases of precise subsumption as follows. C is *positively subsumed* by D w.r.t. $\mathcal{K}$ if $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) > 0$ and C is *certainly subsumed* by D w.r.t. $\mathcal{K}$ if $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = 1$. Given that C is *positively subsumed* by D, we say that ψ is the *most likely context* for the subsumption $(C \sqsubseteq D)$ if $\mathcal{P}_{\mathcal{BN}}(\psi) = sup\{\mathcal{P}_{\mathcal{BN}}(\psi') | (C \sqsubseteq_{\mathcal{K}} D : \psi')\}$.

The rest of this chapter is divided in subsections, each of which provides a solution to one of the reasoning problems that has been defined.

4.2.1 Precise Subsumption

Precise subsumption is to calculate the precise probability of a subsumption w.r.t. a given ontology. We provide an algorithm, named Algorithm 1, that solves the precise subsumption

Algorithm 1 Precise subsumption

Input: $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and A, $B \in \mathcal{N_C}$

Output: The precise probability p of the subsumption $A \sqsubseteq B$

1: $p = 0$ ▶ Initialize the global variable p to 0
2: **for** every world $\mathcal{W}_i = \{X = x | X \in \mathcal{V}, x \in \mathcal{D}(X)\}$ **do**
3: $\quad \mathcal{T}_{\mathcal{W}_i} = \{\}$ ▶ Take an empty $\mathcal{EL}$ ontology
4: $\quad$ **for** every $(E \sqsubseteq F : \psi) \in \mathcal{T}$ **do** ▶ Create an $\mathcal{EL}$ ontology w.r.t. $\mathcal{W}_i$
5: $\qquad$ **if** $\psi \subseteq \mathcal{W}_i$ **then** ▶ If the current world includes ψ
6: $\qquad\quad$ add $(E \sqsubseteq F)$ to $\mathcal{T}_{\mathcal{W}_i}$
7: $\quad$ **if** $(\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B$ and $\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0)$ **then**
8: $\qquad p = p + \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i)$ ▶ Add the worlds probability to p
9: return p

where the subsumption is restricted to named concepts. Later, we show that this can easily be extended to subsumptions with arbitrary concepts. Algorithm 1 takes one world $\mathcal{W}_i$ of the given $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ at a time. Depending on the world $\mathcal{W}_i$ it constructs an $\mathcal{EL}$ ontology $\mathcal{T}_{\mathcal{W}_i}$ upon which the subsumption $(A \sqsubseteq B)$ is checked by the standard $\mathcal{EL}$ completion algorithm. The probability of the world is added to a global variable p if $\mathcal{T}_{\mathcal{W}_i}$ entails the subsumption $(A \sqsubseteq B)$ and the world has a positive probability w.r.t. the given Bayesian network. After repeating this for all of the worlds, Algorithm 1 returns p, in which the precise probability of the subsumption is stored.

Lemma 4.7. *Algorithm 1 is in PSPACE.*

Proof. We need to show that the algorithm uses at most polynomial space w.r.t. the size of the input. Although the algorithm needs to check all the worlds, which are exponentially many, only one world $\mathcal{W}_i$ is kept in memory at a time where $|\mathcal{W}_i|$ is bound with the size of the random variables given by $\mathcal{BN}$. Inside of the for-loop (2-8) consists of two parts, i.e., another for-loop (4-6) and an if-clause (7-8). The space used in the for-loop (4-6) is bound with $|\mathcal{T}| + |\mathcal{W}_i|$, which is linear in the size of the input. The space used in if-clause (7-8) also has a polynomial bound since checking both $\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B$ and $\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0$ can be done in polynomial space, i.e., the former by calling the $\mathcal{EL}$ completion algorithm and the latter by applying the chain rule. Hence, Algorithm 1 is in PSPACE. □

Lemma 4.8 (Soundness)**.** *Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, and the named concepts $A, B \in \mathcal{N_C}$ as inputs, Algorithm 1 returns p such that $p \leq \mathcal{P}(A \sqsubseteq_{\mathcal{K}} B)$.*

Proof. We assume that there are k worlds. For each world $\mathcal{W}_i$, Algorithm 1 constructs an $\mathcal{EL}$ ontology $\mathcal{T}_{\mathcal{W}_i} = \{(E \sqsubseteq F) | (E \sqsubseteq F : \psi) \in \mathcal{T}, \psi \subseteq \mathcal{W}_i\}$. For every $\mathcal{W}_i$ with $\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0$ and its corresponding $\mathcal{EL}$ ontology $\mathcal{T}_{\mathcal{W}_i}$, Algorithm 1 returns p such that:

$$p = \sum_{\mathcal{T}_{\mathcal{W}_i} \models (A \sqsubseteq B)} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) \tag{4.9}$$

By Lemma 4.5, we know that if $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) = p$ then there exists a model $\mathcal{I_P} = (\Phi, \mathcal{P}r)$ of $\mathcal{K}$ where $\mathcal{P}(A \sqsubseteq_{\mathcal{I_P}} B) = p$. There are only finitely many interpretations with positive probability in $\mathcal{I_P}$. Let these contextual interpretations be $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\} \subseteq \Phi$. We use $\mathcal{I_P}$ to construct another probabilistic interpretation $\mathcal{I_P}'$ as follows. For each world $\mathcal{W}_j$, $1 \leq j \leq k$, we define a contextual interpretation $\mathcal{I_C}_j'$ such that:

$$\mathcal{I_C}_j' = \dot{\bigcup_{\mathcal{I_C}_i \models \mathcal{W}_j}} \mathcal{I_C}_i$$

For the probabilistic interpretation $\mathcal{I_P}'$ it holds that:

$$\mathcal{P}(A \sqsubseteq_{\mathcal{I_P}'} B) = \sum_{\mathcal{I_C}_i' \models (A \sqsubseteq B : \mathcal{W}_i)} \mathcal{P}r(\mathcal{I_C}') = \sum_{\mathcal{I_C}_i' \models (A \sqsubseteq B), \mathcal{I_C}_i' \models \mathcal{W}_i} \mathcal{P}r(\mathcal{I_C}') + \sum_{\mathcal{I_C}_i' \not\models \mathcal{W}_i} \mathcal{P}r(\mathcal{I_C}') \quad (4.10)$$

Furthermore, for arbitrary concepts C, D our construction ensures that for any world $\mathcal{W}_i$ and its corresponding contextual interpretation $\mathcal{I_C}_i'$ it holds that:

$$\mathcal{I_C}_i' \models (C \sqsubseteq D) \textit{ iff } \mathcal{T}_{\mathcal{W}_i} \models (C \sqsubseteq D) \quad (4.11)$$

Rewriting $\sum_{\mathcal{I_C}_i' \models (A \sqsubseteq B), \mathcal{I_C}_i' \models \mathcal{W}_i} \mathcal{P}r(\mathcal{I_C}')$ as $\sum_{\mathcal{T}_{\mathcal{W}_i} \models (A \sqsubseteq B)} \mathcal{P_{BN}}(\mathcal{W}_i)$ in Equation 4.10 yields:

$$\mathcal{P}(A \sqsubseteq_{\mathcal{I_P}'} B) = \sum_{\mathcal{T}_{\mathcal{W}_i} \models (A \sqsubseteq B)} \mathcal{P_{BN}}(\mathcal{W}_i) + \sum_{\mathcal{I_C}_i' \not\models \mathcal{W}_i} \mathcal{P}r(\mathcal{I_C}') \quad (4.12)$$

Since $\sum_{\mathcal{I_C}_i' \not\models \mathcal{W}_i} \mathcal{P}r(\mathcal{I_C}') \geq 0$ we get:

$$\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) = \mathcal{P}(A \sqsubseteq_{\mathcal{I_P}'} B) \geq \sum_{\mathcal{T}_{\mathcal{W}_i} \models (A \sqsubseteq B)} \mathcal{P_{BN}}(\mathcal{W}_i) = p$$

It is only left to show that $\mathcal{I_P}$ is indeed a model of the $\mathcal{BEL}$ ontology $\mathcal{K}$. By the construction the following holds for every world $\mathcal{W}_j$:

$$\sum_{\mathcal{I_C}_i' \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I_C}_i') = \sum_{\mathcal{I_C}_j' \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I_C}_j') = \mathcal{P_{BN}}(\mathcal{W}_j) \quad (4.13)$$

For every $(E \sqsubseteq F : \psi) \in \mathcal{T}$ it holds that:

$$\sum_{\mathcal{I_C}_i \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}) = \sum_{\mathcal{I_C}_i \models \psi,\ \mathcal{I_C}_i \models (E \sqsubseteq F)} \mathcal{P}r(\mathcal{I_C}_i) + \sum_{\mathcal{I_C}_i \not\models \psi} \mathcal{P}r(\mathcal{I_C}_i) \quad (4.14)$$

For every contextual interpretation with positive probability we know that $\mathcal{I_C}_i \models \psi$ iff for $\mathcal{W}_i$ it holds that $\psi \subseteq \mathcal{W}_i$. Combining this with (4.11) we get that ($\mathcal{I_C}_i \models \psi$ and $\mathcal{I_C}_i \models (E \sqsubseteq F)$) iff $\psi \subseteq \mathcal{W}_i$. We use this fact to rewrite the Equation 4.14:

$$\sum_{\mathcal{I_C} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I_C}) = \sum_{\psi \subseteq \mathcal{W}_i} \mathcal{P_{BN}}(\mathcal{W}_i) + \sum_{\psi \not\subseteq \mathcal{W}_i} \mathcal{P_{BN}}(\mathcal{W}_i) = \sum_{\mathcal{W}_i} \mathcal{P_{BN}}(\mathcal{W}_i) = 1 \quad (4.15)$$

From (4.15) we get that $\mathcal{I_P}$ is a model of $\mathcal{T}$. Together with (4.13) we get that $\mathcal{I_P}$ is a model of $\mathcal{K}$. □

Lemma 4.9 (Completeness)**.** *Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, and the named concepts $A, B \in \mathcal{N_C}$ as inputs, Algorithm 1 returns p such that $p \geq \mathcal{P}(A \sqsubseteq_{\mathcal{K}} B)$.*

Proof. We assume that there are k worlds. For each world $\mathcal{W}_i$, Algorithm 1 constructs an $\mathcal{EL}$ ontology $\mathcal{T}_{\mathcal{W}_i} = \{(E \sqsubseteq F) | (E \sqsubseteq F : \psi) \in \mathcal{T}, \psi \subseteq \mathcal{W}_i\}$. For every world $\mathcal{W}_i$ with $\mathcal{P_{BN}}(\mathcal{W}_i) > 0$ and its corresponding $\mathcal{EL}$ ontology $\mathcal{T}_{\mathcal{W}_i}$, Algorithm 1 returns p such that:

$$p = \sum_{\mathcal{T}_{\mathcal{W}_i} \models (A \sqsubseteq B)} \mathcal{P_{BN}}(\mathcal{W}_i) \quad (4.16)$$

Let $\mathcal{I}_i$ be a model of $\mathcal{T}_{\mathcal{W}_i}$ such that for arbitrary concepts C and D it holds that:

$$\mathcal{I}_i \models (C \sqsubseteq D) \text{ iff } \mathcal{T}_{\mathcal{W}_i} \models (C \sqsubseteq D) \tag{4.17}$$

We define a set of contextual interpretations $\mathcal{I}_{\mathcal{C}i}$ as a triple $\mathcal{I}_{\mathcal{C}i} = (\Delta^{\mathcal{I}_i}, \mathcal{V}^{\mathcal{I}_{\mathcal{C}i}}, \cdot^{\mathcal{I}_{\mathcal{C}i}})$ where $\cdot^{\mathcal{I}_{\mathcal{C}i}}$ is a disjoint union of mappings:

$$\cdot^{\mathcal{I}_{\mathcal{C}i}} = \cdot^{\mathcal{I}_i} \uplus (X \mapsto x) \text{ where } \{X = x | X \in \mathcal{V}\} = \mathcal{W}_i.$$

Since $\Delta^{\mathcal{I}_i}$ and $\mathcal{V}^{\mathcal{I}_{\mathcal{C}i}}$ are disjoint sets $\cdot^{\mathcal{I}_{\mathcal{C}i}}$ is well defined. We define a probabilistic interpretation $\mathcal{I}_{\mathcal{P}} = (\Phi, \mathcal{P}r)$ where $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}k}\} \subseteq \Phi$ with $\mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i)$ for $1 \leq i \leq k$. For every contextual interpretation $\mathcal{I}_{\mathcal{C}} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}k}\}$ it holds that $\mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = 0$. Since every conceptual interpretation $\mathcal{I}_{\mathcal{C}i}$ is constructed from $\mathcal{I}_i$ by an additional disjoint mapping determined by $\mathcal{W}_i$, we get an analogous result as (4.17):

$$\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D) \text{ iff } \mathcal{T}_{\mathcal{W}_i} \models (C \sqsubseteq D) \tag{4.18}$$

By the definition of a subsumption w.r.t. a probabilistic interpretation we get:

$$\mathcal{P}(A \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} B) = \sum_{\mathcal{I}_{\mathcal{C}1} \models \mathcal{W}_1,\, \mathcal{I}_{\mathcal{C}1} \models (A \sqsubseteq B)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}1}) + \cdots + \sum_{\mathcal{I}_{\mathcal{C}k} \models \mathcal{W}_k,\, \mathcal{I}_{\mathcal{C}k} \models (A \sqsubseteq B)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}k}) \tag{4.19}$$

Each term on the right hand side of Equation 4.19 either equals to $\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i)$ or to 0 depending on whether the contextual interpretation entails the subsumption or not . Hence, the Equation 4.19 can be rewritten as:

$$\mathcal{P}(A \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} B) = \sum_{\mathcal{I}_{\mathcal{C}i} \models (A \sqsubseteq B)} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) \tag{4.20}$$

Given (4.18) we can replace the subscript $\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D)$ on the right hand side of Equation 4.20 with $\mathcal{T}_{\mathcal{W}_i} \models (A \sqsubseteq B)$ which yields:

$$\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) \leq \mathcal{P}(A \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} B) = \sum_{\mathcal{T}_{\mathcal{W}_i} \models (A \sqsubseteq B)} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) = p \tag{4.21}$$

It is only left to show that $\mathcal{I}_{\mathcal{P}}$ is indeed a model of the $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$. By the construction the following holds for every world $\mathcal{W}_j$:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = \sum_{\mathcal{I}_{\mathcal{C}j} \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I}_{\mathcal{C}j}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j) \tag{4.22}$$

For every $(E \sqsubseteq F : \psi) \in \mathcal{T}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \sum_{\mathcal{I}_{\mathcal{C}i} \models \psi,\, \mathcal{I}_{\mathcal{C}i} \models (E \sqsubseteq F)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) + \sum_{\mathcal{I}_{\mathcal{C}i} \not\models \psi} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) \tag{4.23}$$

For the contextual interpretations with positive probability we know that $\mathcal{I}_{\mathcal{C}i} \models \psi$ iff for $\mathcal{W}_i$ it holds that $\psi \subseteq \mathcal{W}_i$. Combining this with (4.18), we get that ($\mathcal{I}_{\mathcal{C}i} \models \psi$ and $\mathcal{I}_{\mathcal{C}i} \models (E \sqsubseteq F)$ iff $\psi \subseteq \mathcal{W}_i$. We use this fact to rewrite Equation 4.23 as follows:

$$\sum_{\mathcal{I}_{\mathcal{C}} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = \sum_{\psi \subseteq \mathcal{W}_i} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) + \sum_{\psi \not\subseteq \mathcal{W}_i} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) = \sum_{\mathcal{W}_i} \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) = 1 \tag{4.24}$$

From (4.24) we get that $\mathcal{I}_{\mathcal{P}}$ is a model of $\mathcal{T}$. Together with (4.22) we get that $\mathcal{I}_{\mathcal{P}}$ is a model of $\mathcal{K}$. □

We have provided Algorithm 1 to solve precise subsumption for a given pair of named concepts. By Lemma 4.8 and Lemma 4.9 we get the correctness of the algorithm. Next, we show that this can be generalised to all concepts. Let C, D be arbitrary concepts, A, B new named concepts, $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and $\mathcal{K}' = (\{\mathcal{T} \cup \{(A \sqsubseteq C : \{\}), (D \sqsubseteq B : \{\})\}\}, \mathcal{BN})$ $\mathcal{BEL}$ ontologies. Then, Lemma 4.10 holds.

Lemma 4.10. *C is precisely subsumed by D w.r.t. $\mathcal{K}$ with probability p only if A is precisely subsumed by B w.r.t. $\mathcal{K}^{'}$ with probability p.*

Proof. Assume by contradiction that A is most likely subsumed by B w.r.t. $\mathcal{K}'$ with probability $p^{'} > p$. Then, for all models $\mathcal{I_P}^{'}$ of $\mathcal{K}^{'}$ it holds that $\mathcal{P}(A \sqsubseteq_{\mathcal{I_P}^{'}} B) \geq p^{'}$. But then, for all models $\mathcal{I_P}$ of $\mathcal{K}$ it holds that $\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) \geq p^{'}$ since empty context is satisfied by any contextual interpretation. Hence $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) \geq p^{'} > p$, which leads to a contradiction.

Assume by contradiction that A is most likely subsumed by B w.r.t. $\mathcal{K}'$ with probability $p^{'} < p$. Then, there exists a model $\mathcal{I_P}^{'}$ of $\mathcal{K}^{'}$ such that $\mathcal{P}(A \sqsubseteq_{\mathcal{I_P}^{'}} B) = p^{'}$ by Lemma 4.5. But then, $\mathcal{I_P}^{'}$ is also a model of the sub-ontology $\mathcal{K}$ such that $\mathcal{P}(C \sqsubseteq_{\mathcal{I_P}} D) = p^{'}$. Hence, $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) \leq p^{'} < p$, which leads to a contradiction. □

From Lemmas 4.7-4.10, Theorem 4.11 is an immediate result.

Theorem 4.11. *Precise subsumption can be decided in PSPACE.*

In the next two subsections we look into two special instances of precise subsumption, i.e., positive subsumption and certain subsumption.

4.2.2 Positive Subsumption

Positive subsumption can be solved by a non-deterministic algorithm, Algorithm 2, which first guesses a world $\mathcal{W}_i$. Depending on $\mathcal{W}_i$, it constructs an $\mathcal{EL}$ ontology $\mathcal{T}_{\mathcal{W}_i}$ upon which the subsumption $(A \sqsubseteq B)$ is checked by the standard $\mathcal{EL}$ completion algorithm. If the subsumption holds and at the same time the world has a positive probability w.r.t. the given Bayesian network, Algorithm 2 answers true.

Lemma 4.12. *Algorithm 2 is a non-deterministic algorithm that terminates in polynomial time.*

Proof. To prove that the algorithm is in NP, we need to show the following: Both guessing a world and checking whether the guess was a correct one can be done in polynomial time. Producing a world can be done in polynomial time by picking one value from $\mathcal{D}(X)$ for each random variable X, which is linear in the size of the given input.

The sub-procedure in the algorithm consists of 2 basic parts. First part is the for-loop (3-5) which is bound with the size of the $\mathcal{T}$ and subset checking (4) inside the for-loop is linear in the size of $\mathcal{W}_i$. Hence, first part stays in polynomial-time. The second part (6-9) checks both: $\mathcal{T}_{\mathcal{EL}} \models C \sqsubseteq D$ and $\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0$, where the former can be done by calling the $\mathcal{EL}$ completion algorithm in polynomial time and the latter by applying the chain rule. Hence, the combined complexity of the whole sub-procedure is then polynomial. □

Algorithm 2 Positive subsumption

Input: $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and $A,\ B \in \mathcal{N_C}$
Output: `true if A is positively subsumbed by B, false otherwise`
$\mathcal{W}_i$=$\{X = x | X \in \mathcal{V}, x \in \mathcal{D}(X)\}$ ▶ Guess a valuation for all random variables
$\mathcal{T}_{\mathcal{W}_i} = \{\}$ ▶ Take an empty $\mathcal{EL}$ ontology
for every $(E \sqsubseteq F : \psi) \in \mathcal{T}$ **do** ▶ Create an $\mathcal{EL}$ ontology w.r.t. $\mathcal{W}_i$
  if $\psi \subseteq \mathcal{W}_i$ **then** ▶ If the current world includes ψ
    `add` $(E \sqsubseteq F)$ `to` $\mathcal{T}_{\mathcal{W}_i}$
if $(\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B)$ and $(\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0)$ **then**
  `return true`
else
  `return false`

Lemma 4.13 (Soundness)**.** *Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, and the named concepts $A, B \in \mathcal{N_C}$ as inputs, Algorithm 2 returns true only if $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) > 0$.*

Proof. As a result of Lemma 4.8 stating the soundness of Algorithm 1 we know that all the worlds with positive probability w.r.t. which the entailment $\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B$ holds contribute to the probability of the subsumption. Since Algorithm 2 answers true only if it finds such a world $\mathcal{W}_i$ we get $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) > 0$. □

Lemma 4.14 (Completeness)**.** *Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, and the named concepts $A, B \in \mathcal{N_C}$ as inputs, Algorithm 2 returns true if $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) > 0$.*

Proof. We have proved Lemma 4.9 stating the completeness of Algorithm 1. Hence, we know that the precise probability of the subsumption can be calculated by only taking into account the worlds with positive probability w.r.t. which the entailment $\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B$ holds. Hence, if $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) > 0$ then there must exist such a world. Assuming that the guess for a world done by Algorithm 2 is a correct one Algorithm 2 answers true. □

We have shown the correctness of the algorithm in Lemma 4.13 and Lemma 4.14. We still need to show that this can be generalised to arbitrary concepts. Let C, D be concepts, A, B new named concepts, $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and $\mathcal{K}' = (\mathcal{T}', \mathcal{BN})$ $\mathcal{BEL}$ ontologies where $\mathcal{T}' = \mathcal{T} \cup \{(A \sqsubseteq C : \{\}), (B \sqsubseteq D : \{\})\}$. Since Algorithm 2 is a special case of Algorithm 1 and we have shown the generalisation of Algorithm 1 in Lemma 4.10, Lemma 4.15 is an immediate result.

Lemma 4.15. *C is positively subsumed by D w.r.t. $\mathcal{K}$ iff A is positively subsumed by B w.r.t. $\mathcal{K}'$.*

From Lemma 4.12 we know that Algorithm 2 is a non-deterministic algorithm that terminates in polynomial time. From Lemmas 4.13-4.15 we get that the positive subsumption problem is decidable.

Lemma 4.16. *Positive subsumption can be decided in NP.*

It has been shown that the problem can be solved in NP. In the rest of this section, NP is shown to be also a lower bound for the problem. Lemma 4.17 shows that inferencing in Bayesian networks can be reduced to positive subsumption w.r.t. a $\mathcal{BEL}$ ontology, that is constructed in polynomial time.

Lemma 4.17 (Hardness)**.** *Let $\mathcal{BN}$ be a Bayesian network and $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be a $\mathcal{BEL}$ ontology where $\mathcal{T} = \{(C \sqsubseteq D : \psi)\}$. Then it holds that $\mathcal{P}_{\mathcal{BN}}(\psi) > 0$ iff $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) > 0$.*

Proof. $\Rightarrow$ Let $\mathcal{I}_{\mathcal{P}} = (\Phi, \mathcal{P}r)$ be an arbitrary model of $\mathcal{K}$ where $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\} \subseteq \Phi$, $1 \leq i \leq n$ and for every contextual interpretation $\mathcal{I}_{\mathcal{C}} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\}$ it holds that $\mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = 0$. By the definition of a model, we know that for $(C \sqsubseteq D : \psi) \in \mathcal{T}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1$$

Since $\mathcal{P}_{\mathcal{BN}}(\psi) > 0$ there is at least one contextual interpretation $\mathcal{I}_{\mathcal{C}i}$ with positive probability that satisfies ψ, which as a consequence, needs also to satisfy $C \sqsubseteq D$. Hence, $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) > 0$.

$\Leftarrow$ For all models $\mathcal{I}_{\mathcal{P}} = (\Phi, \mathcal{P}r)$ of $\mathcal{K}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1, \tag{4.25}$$

where $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\} \subseteq \Phi$, $1 \leq i \leq n$ and for every contextual interpretation $\mathcal{I}_{\mathcal{C}} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\}$ it holds that $\mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = 0$. Equation 4.25 holds for all models of $\mathcal{K}$ only if the followings hold:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models \psi} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) > 0, \quad \sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) > 0 \tag{4.26}$$

Combining (4.25) with Lemma 3.7, which asserts that the sum of the contextual interpretations satisfying a context ψ must equal to the probability of ψ in the Bayesian network, we get that $\mathcal{P}_{\mathcal{BN}}(\psi) > 0$. ☐

By Lemma 4.16 and Lemma 4.17, Theorem 4.18 is an immediate result.

Theorem 4.18. *Positive subsumption is NP-complete.*

In Section 4.2.3 we analyse the dual decision problem, i.e., certain subsumption, which is another special case of precise subsumption.

4.2.3 Certain Subsumption

Certain subsumption can be solved analogously to the positive subsumption. Instead of the certain subsumption problem, an answer to the dual problem is produced. As in Algorithm 2, first a world $\mathcal{W}_i$ is guessed. Depending on $\mathcal{W}_i$, the algorithm constructs an $\mathcal{EL}$ ontology $\mathcal{T}_{\mathcal{W}_i}$ upon which the subsumption $A \sqsubseteq B$ needs to be checked by the standard $\mathcal{EL}$ completion algorithm. If the subsumption does not hold and at the same time the context has a positive probability w.r.t. the given Bayesian network the algorithm answers false. The intuition is to look for a world $\mathcal{W}_i$ with positive probability that does not satisfy the subsumption, which guarantees that the subsumption has a probability strictly less than 1. In the following, Algorithm 3 is given to solve the certain subsumption problem as explained.

Algorithm 3 Certain subsumption

Input: $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ `and` A, $B \in \mathcal{N_C}$
Output: `true if A is certainly subsumbed by B, false otherwise`
1: $\mathcal{W}_i$=$\{X = x | X \in \mathcal{V}, x \in \mathcal{D}(X)\}$ ▶ Guess a valuation for all random variables
2: $\mathcal{T}_{\mathcal{W}_i} = \{\}$ ▶ Take an empty $\mathcal{EL}$ ontology
3: **for** `every` $(E \sqsubseteq F : \psi) \in \mathcal{T}$ **do** ▶ Create an $\mathcal{EL}$ ontology w.r.t. $\mathcal{W}_i$
4:     **if** $\psi \subseteq \mathcal{W}_i$ **then** ▶ If the current world includes ψ
5:         `add` $(E \sqsubseteq F)$ `to` $\mathcal{T}_{\mathcal{W}_i}$
6: **if** $(\mathcal{T}_{\mathcal{W}_i} \not\models (A \sqsubseteq B)) \wedge (\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0)$ **then**
7:     `return false`
8: **else**
9:     `return true`

Lemma 4.19. *Algorithm 3 is a non-deterministic algorithm that returns the answer false in polynomial time.*

Proof. We show that Algorithm 3 returns the answer false in NP. As in Algorithm 2 we need to show the following: Both guessing a world and checking whether the guess was a correct one can be done in polynomial time. Producing a world can be done in polynomial time by picking one value from $\mathcal{D}(X)$ for each random variable X, which is linear in the size of the given input.

The sub-procedure in the algorithm consists of 2 basic parts. First part is the for-loop (3-5) which is bound with the size of the $\mathcal{T}$ and subset checking (4) inside the for-loop is linear in the size of $\mathcal{W}_i$. Hence, first part stays in polynomial-time. The second part (6-9) checks both: $\mathcal{T}_{\mathcal{W}_i} \not\models A \sqsubseteq B$ and $\mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i) > 0$, where the former can be done by calling the $\mathcal{EL}$ completion algorithm in polynomial time and the latter by applying the chain rule. Hence, the combined complexity of the whole sub-procedure is then polynomial. □

Lemma 4.20 (Soundness)**.** *Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, and the named concepts $A, B \in \mathcal{N_C}$ as inputs, Algorithm 3 returns true only if $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) = 1$.*

Proof. If Algorithm 3 returns true then we know that there is no world $\mathcal{W}_i$ with positive probability satisfying $\mathcal{T}_{\mathcal{W}_i} \not\models A \sqsubseteq B$, i.e., for all of the worlds $\mathcal{W}_i$ with positive probability we get $\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B$. By the soundness result of Algorithm 1 given in Lemma 4.8, we know that all of the worlds $\mathcal{W}_i$ with positive probability upon which the entailment $\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B$ holds contribute to the probability of the subsumption. Since the sum of probabilities of all worlds add up to 1, $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) = 1$ follows. □

Lemma 4.21 (Completeness)**.** *Let $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be a $\mathcal{BEL}$ ontology and A, $B \in \mathcal{N_C}$. Given $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, A and B as inputs, Algorithm 3 returns true if $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) = 1$.*

Proof. We proved the completeness of Algorithm 1 in Lemma 4.9. Hence, we know that the precise probability of the subsumption can be calculated by only taking into account the worlds with positive probability upon which the entailment $\mathcal{T}_{\mathcal{W}_i} \models A \sqsubseteq B$ holds. Since the sum over the probabilities of all worlds equals to 1 and $\mathcal{P}(A \sqsubseteq_{\mathcal{K}} B) = 1$, we know that there is no world with positive probability upon which $\mathcal{T}_{\mathcal{W}_i} \not\models A \sqsubseteq B$ holds. As a result, Algorithm 3 returns true. □

Let C, D be concepts, A, B new named concepts, $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and $\mathcal{K}' = (\mathcal{T}', \mathcal{BN})$ $\mathcal{BEL}$ ontologies where $\mathcal{T}' = \mathcal{T} \cup \{(A \sqsubseteq C : \{\}), (B \sqsubseteq D : \{\})\}$. With the same argument as in Lemma 4.15 in positive subsumption, Lemma 4.22 is an immediate result.

Lemma 4.22. *C is certainly subsumed by D w.r.t. $\mathcal{K}$ iff A is certainly subsumed by B w.r.t. $\mathcal{K}'$.*

From Lemma 4.19 we know that Algorithm 3 is a non-deterministic algorithm that produces an answer to the dual certain subsumption problem in polynomial time. Together with Lemma 4.22 and the correctness results shown in Lemma 4.20 and Lemma 4.21 we get the decidability result.

Lemma 4.23. *Certain subsumption can be decided in coNP.*

It has been shown that the problem is in coNP. We prove the completeness of the problem by a reduction from inferencing in Bayesian networks to certain subsumption problem w.r.t. a $\mathcal{BEL}$ ontology, that is constructed in polynomial time.

Lemma 4.24 (Hardness)**.** *Let $\mathcal{BN}$ be a Bayesian network and $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be a $\mathcal{BEL}$ ontology where $\mathcal{T} = \{(C \sqsubseteq D : \psi)\}$. Then it holds that $\mathcal{P}_{\mathcal{BN}}(\psi) > 0$ iff $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = 1$.*

Proof. $\Rightarrow$ Let $\mathcal{I}_{\mathcal{P}} = (\Phi, \mathcal{P}r)$ be an arbitrary model of $\mathcal{K}$ where $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\} \subseteq \Phi$, $1 \leq i \leq n$ and for every contextual interpretation $\mathcal{I}_{\mathcal{C}} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\}$ it holds that $\mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = 0$. By definition of a model, we know that for $(C \sqsubseteq D : \psi) \in \mathcal{T}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1$$

Since $\mathcal{P}_{\mathcal{BN}}(\psi) = 1$, all of the contextual interpretations $\mathcal{I}_{\mathcal{C}i}$ with positive probability satisfy ψ. Consequently, all of the contextual interpretations $\mathcal{I}_{\mathcal{C}i}$ with positive probability also satisfy $C \sqsubseteq D$. Hence, $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = 1$.

$\Leftarrow$ Since $\mathcal{P}(C \sqsubseteq_{\mathcal{K}} D) = 1$, for all models $\mathcal{I}_{\mathcal{P}} = (\Phi, \mathcal{P}r)$ of $\mathcal{K}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (C \sqsubseteq D)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1, \tag{4.27}$$

where $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\} \subseteq \Phi$, $1 \leq i \leq n$ and for every contextual interpretation $\mathcal{I}_{\mathcal{C}} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\}$ it holds that $\mathcal{P}r(\mathcal{I}_{\mathcal{C}}) = 0$. This implies that the subsumption holds in every world. Hence we get:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models \psi} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1 \tag{4.28}$$

Combining (4.28) with Lemma 3.7 we get that $\mathcal{P}_{\mathcal{BN}}(\psi) = 1$. □

By Lemma 4.23 and Lemma 4.24 we get the completeness result for the certain subsumption problem.

Theorem 4.25. *Certain subsumption is coNP-complete.*

Together with certain subsumption, we have provided procedures to solve three of the reasoning problems. These reasoning problems are closely related. Positive and certain subsumption are dual reasoning problems that are special instances of precise subsumption.

In Section 4.2.4, we concentrate on the last reasoning problem and solve it with an approach that is analogous to finding justifications for consequences in standard DLs.

4.2.4 Most Likely Context

Most likely context for a subsumption is an interesting decision problem that stems from our contextual approach. In DL ontologies given a consequence, it is very important to find justifications for this consequence. The task of finding justifications, minimal subsets, of a DL ontology that have a given consequence is called axiom pinpointing [35]. Axiom pinpointing in $\mathcal{EL}$ together with its complexity results are presented by Peñaloza and Sertkaya in [36]. The idea in axiom pinpointing is to label every axiom in an ontology and keep track of these labels while computing the consequences. At the end, the set of labels are used to find the set of minimal axioms. We use this idea in finding the most likely context for a subsumption. The $\mathcal{BEL}$ contexts can be seen as labels in axiom pinpointing. Differently, the contexts have probabilities that provide the likelihoods.

To decide the most likely context for a subsumption in $\mathcal{BEL}$, we introduce an algorithm that follows basically the same lines as the existing graph completion algorithm for $\mathcal{EL}$. In $\mathcal{EL}$, the completion algorithm for computing subsumption w.r.t. a TBox classifies the whole TBox, i.e., it computes the subsumption relationships between all named concepts of a given TBox simultaneously. This algorithm proceeds in four steps:

1. Normalisation of the TBox.
2. Translation of the normalised TBox into completion sets.
3. Completion of these sets using completion rules.
4. Reading off the subsumption relationships from the normalised graph.

In the sequel, we present the $\mathcal{BEL}$ *completion* algorithm, which follows the same steps as the $\mathcal{EL}$*-completion* algorithm. We begin with the normalisation step, which enables us to convert a given $\mathcal{BEL}$ TBox to an equivalent normalised $\mathcal{BEL}$ TBox.

1. Normalisation. The idea behind the normalisation is to replace the concepts that are not named concepts with the named ones. The aim is to convert the TBox to an equivalent TBox w.r.t. the subsumption relation upon which the classification can be done in polynomial time. We say that a $\mathcal{BEL}$ ontology is in normal form if the TBox in the ontology contains only axioms in a special form.

Definition 4.26 (Normal form)**.** A $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ is in *normal form* iff every axiom in $\mathcal{T}$ is one of the following forms:

$$(A \sqsubseteq B : \psi), (A_1 \sqcap A_2 \sqsubseteq B : \psi), (A \sqsubseteq \exists r.B : \psi), (\exists r.A \sqsubseteq B : \psi)$$

where A, A_1, A_2, $B \in (\mathcal{N}_\mathcal{C} \cup \top)$ and ψ is a context.

We introduce the *normalisation rules* which enable us to rewrite a given $\mathcal{BEL}$ ontology into a normalised one.

NR1	$((C \sqcap \widehat{D} \sqsubseteq E) : \psi)$	$\dashrightarrow$	$\{(\widehat{D} \sqsubseteq A : \psi_1), ((C \sqcap A \sqsubseteq E) : \psi_2)\}$
NR2	$(\exists r.\widehat{C} \sqsubseteq D) : \psi)$	$\dashrightarrow$	$\{(\widehat{C} \sqsubseteq A : \psi_1), (\exists r.A \sqsubseteq D) : \psi_2)\}$
NR3	$(\widehat{C} \sqsubseteq \widehat{D}) : \psi)$	$\dashrightarrow$	$\{((\widehat{C} \sqsubseteq A) : \psi_1), ((A \sqsubseteq \widehat{D}) : \psi_2)\}$
NR4	$(B \sqsubseteq \exists r.\widehat{C}) : \psi)$	$\dashrightarrow$	$\{(B \sqsubseteq \exists r.A : \psi_1), (A \sqsubseteq \widehat{C}) : \psi_2)\}$
NR5	$((B \sqsubseteq C \sqcap D) : \psi)$	$\dashrightarrow$	$\{((B \sqsubseteq C) : \psi), ((B \sqsubseteq D) : \psi)\}$
where $\widehat{C}$, $\widehat{D} \notin \{\mathcal{N}_C \cup \top\}$, A is a new concept name and $\psi = \psi_1 \cup \psi_2$			

TABLE 4.1: $\mathcal{BEL}$ normalisation rules

$\mathcal{BEL}$ ontologies can be transformed into a normal form by exhaustively applying the normalisation rules given in Table 4.1. Suppose that we are given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$. Applying a rule updates the TBox $\mathcal{T}$ by replacing the axioms having one of the forms listed on the left hand side of Table 4.1 by the two axioms on the right hand side. We first apply the rules (NR1)-(NR4) exhaustively and then apply the rule (NR5). The number of possible applications of rules (NF1) to (NF4) is limited linearly in the size of $\mathcal{T}$. Each of these rules increases the size of $\mathcal{T}$ only by a constant. Hence, applying the rules (NF1) to (NF4) exhaustively increases the size of $\mathcal{T}$ only polynomially. Then, we begin to apply the rule (NF5). Therefore, single application of the rule (NF5) also increases the size of $\mathcal{T}$ only by a constant. Hence, by exhaustively applying (NF5) the size of $\mathcal{T}$ increased only linearly. As a consequence, normalisation takes only polynomial time. It is also easy to verify that these rules yield a TBox in normal form.

Furthermore, the rule applications are sound, meaning that, for each rule it is possible to derive the initial axiom from the two new axioms. It is also easy to see that any contextual interpretation satisfying the right hand side must satisfy the left hand side by our semantics. As a result, any model of the normalised ontology is also a model of the original ontology. We collect these results in Lemma 4.27.

Lemma 4.27. *Let A, $B \in \{\mathcal{N}_C \cup \top\}$. Every $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ can be transformed in polynomial time into a normalised ontology $\mathcal{K}' = (\mathcal{T}', \mathcal{BN})$ such that:*

$$(A \sqsubseteq_{\mathcal{K}} B : \psi) \text{ iff } (A \sqsubseteq_{\mathcal{K}'} B : \psi)$$

In the rest, we assume that the ontology is in normal form. We first define the completion sets w.r.t. the normalised ontology and then construct the so-called completion graph. We define the completion rules that modify the completions sets represented by the completion graph in a way that the subsumption relations can be read directly from the graph.

2. Completion Sets. Completion sets are the basic elements of the completion algorithm. There are two kinds of completion sets used in the algorithm: $S(A)$ and $R(A, B)$ for each concept name A, B occurring in the normalised TBox. Given a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ the set labels $S(A)$ and $R(A, B)$ w.r.t. $\mathcal{K}$ are initialised as follows:

– $S(A) = \{(A : \{\}), (\top : \{\})\}$ for each concept name $A \in \mathcal{T}$

CR1.	If $(A_1 \sqsubseteq B : \psi_1) \in \mathcal{T}$ and it holds that: $(A_1 : \psi_2) \in S(A)$ then add $(B : \psi_1 \cup \psi_2)$ to $S(A)$.
CR2.	If $(A_1 \sqcap A_2 \sqsubseteq B : \psi_1) \in \mathcal{T}$ and it holds that $(A_1 : \psi_2) \in S(A)$ and $(A_2 : \psi_3) \in S(A)$ then add $(B : (\psi_1 \cup \psi_2 \cup \psi_3)))$ to $S(A)$.
CR3.	If $(A_1 \sqsubseteq \exists r.B : \psi_1) \in \mathcal{T}$ and $(A_1 : \psi_2) \in S(A)$ then add $(r : (\psi_1 \cup \psi_2))$ to $R(A, B)$.
CR4.	If $(\exists r.B_1 \sqsubseteq A_1 : \psi_1) \in \mathcal{T}$ and it holds that $(B_1 : \psi_2) \in S(B)$, $(r : \psi_3) \in R(A, B)$ then add $(A_1 : ((\psi_1 \cup \psi_2 \cup \psi_3))$ to $S(A)$.

TABLE 4.2: $\mathcal{BEL}$ completion rules

– $R(A, B) = \{\}$ for each concept name A, $B \in \mathcal{T}$

We represent the completion sets with a graph, named completion graph. We manipulate the set labels in the graph with the so-called completion rules.

3. Completion Graph. Intuitively the sets $S(A)$ represent the nodes that contain set labels, while the sets $R(A, B)$ represent the edges of the completion graph. Formally the completion graph $G = (V, V \times V, S, R)$ is defined over a $\mathcal{BEL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ as follows:

– $V = A$ where $A \in (\mathcal{N_C} \cup \top)$
– $S = V \mapsto \mathcal{N_{BC}}$ where $\mathcal{N_{BC}} = \{(A, \psi) | A \in \mathcal{N_C}\}$
– $R = V \times V \mapsto \mathcal{N_{BR}}$ where $\mathcal{N_{BR}} = \{(r, \psi) | A \in \mathcal{N_R}\}$

The completion graph is constructed with the initialised completion sets. Next, we define the *completion rules* given in Table 4.2 that change the set labels in the nodes of the graph in a certain way. The completion rules are applied only if they change a set label in the completion graph and the algorithm continues until no more rule is applicable. Intuitively, the completion rules make the subsumption relationships in the TBox explicit so that they can be read directly from the completion graph. In fact, the completion rules preserve the following invariants that guarantee this:

– $(B : \psi) \in S(A) \rightarrow (A \sqsubseteq_{\mathcal{K}} B : \psi)$
– $(r : \psi) \in r(A, B) \rightarrow (A \sqsubseteq_{\mathcal{K}} \exists r.B : \psi)$

Lemma 4.28. *The completion rules preserve invariants.*

Proof. After a successful application of the completion rule $R1$, $(B : (\psi_1 \cup \psi_2))$ is added to $S(A)$. To show that the rule preserves the invariants $(A \sqsubseteq_{\mathcal{K}} B : (\psi_1 \cup \psi_2))$ needs to be shown. $(A_1 \sqsubseteq_{\mathcal{K}} B : \psi_1)$ holds since it is in $\mathcal{T}$. Furthermore, $(A_1 : \psi_2) \in S(A)$ is given. Assuming that the invariants hold before the rule application, $(A_1 : \psi_2) \in S(A)$ implies $(A \sqsubseteq_{\mathcal{K}} A_1 : \psi_2)$. This means that all of the contextual interpretations with positive probability satisfy both $(A_1 \sqsubseteq_{\mathcal{K}} B : \psi_1)$ and $(A \sqsubseteq_{\mathcal{K}} A_1 : \psi_2)$ for all models $\mathcal{I_P}$ of $\mathcal{K}$. As a result, we get $(A \sqsubseteq_{\mathcal{K}} B : (\psi_1 \cup \psi_2))$.

After a successful application of the completion rule $R2$, $(B : (\psi_1 \cup \psi_2 \cup \psi_3))$ is added to $S(A)$. To show that the rule preserves the invariants $(A \sqsubseteq_{\mathcal{K}} B : (\psi_1 \cup \psi_2 \cup \psi_3))$ needs to be shown. $(A_1 \sqcap A_2 \sqsubseteq_{\mathcal{K}} B : \psi_1)$ holds since it is in $\mathcal{T}$. $(A_1 : \psi_2) \in S(A)$ and

$(A_2 : \psi_3) \in S(A)$ are given. Assuming that the invariants hold before the rule application, we get both $(A \sqsubseteq_{\mathcal{K}} A_1 : \psi_2)$ and $(A \sqsubseteq_{\mathcal{K}} A_2 : \psi_3)$. This means that all of the contextual interpretations with positive probability satisfy both $(A \sqsubseteq_{\mathcal{K}} A_1 : \psi_2)$ and $(A \sqsubseteq_{\mathcal{K}} A_2 : \psi_3)$ for all models $\mathcal{I}_{\mathcal{P}}$ of $\mathcal{K}$. Hence we get $(A \sqsubseteq_{\mathcal{K}} A_1 \sqcap A_2 : (\psi_2 \cup \psi_3))$. Similarly combining this with $(A_1 \sqcap A_2 \sqsubseteq_{\mathcal{K}} B : \psi_1)$ we get $(A \sqsubseteq_{\mathcal{K}} B : (\psi_1 \cup \psi_2 \cup \psi_3))$.

After a successful application of the completion rule $R3$, $(r : (\psi_1 \cup \psi_2))$ is added to $R(A, B)$. Hence, $(A \sqsubseteq_{\mathcal{K}} \exists r.B : (\psi_1 \cup \psi_2))$ needs to be shown. Since $(A_1 \sqsubseteq \exists r.B : \psi_1) \in \mathcal{T}$ it holds that $(A_1 \sqsubseteq_{\mathcal{K}} \exists r.B : \psi_1)$ and $(A_1 : \psi_2) \in S(A)$ is given. Assuming that the invariants hold prior to rule application $(A_1 : \psi_2) \in S(A)$ implies $(A \sqsubseteq_{\mathcal{K}} A_1 : \psi_2)$. Since all of the contextual interpretations with positive probability satisfy both $(A \sqsubseteq_{\mathcal{K}} A_1 : \psi_2)$ and $(A_1 \sqsubseteq_{\mathcal{K}} \exists r.B : \psi_1)$ for all models $\mathcal{I}_{\mathcal{P}}$ of $\mathcal{K}$ it follows that all of the contextual interpretations with positive probability satisfy $(A \sqsubseteq \exists r.B : (\psi_1 \cup \psi_2))$.

After a successful application of the completion rule $R4$, $(A_1 : (\psi_1 \cup \psi_2 \cup \psi_3))$ is added to $S(A)$. $(A \sqsubseteq_{\mathcal{K}} A_1 : (\psi_1 \cup \psi_2 \cup \psi_3))$ needs to be shown. $(\exists r.B_1 \sqsubseteq A_1 : \psi_1) \in \mathcal{T}$ and hence $(\exists r.B_1 \sqsubseteq_{\mathcal{K}} A_1 : \psi_1)$ holds. Furthermore, $(B_1 : \psi_2) \in S(B)$ and $(r : \psi_3) \in R(A, B)$ are given. Analogously, by the invariants we get $(B \sqsubseteq_{\mathcal{K}} B_1 : \psi_2)$ and $(A \sqsubseteq_{\mathcal{K}} \exists r.B : \psi_3)$ respectively given that the invariants hold prior to the rule application. $(A \sqsubseteq_{\mathcal{K}} \exists r.B : \psi_3)$ and $((B \sqsubseteq_{\mathcal{K}} B_1) : \psi_2)$ imply $(A \sqsubseteq_{\mathcal{K}} \exists r.B_1 : (\psi_2 \cup \psi_3))$ by the same argument over all of the contextual interpretations of all models. Combining this with $(\exists r.B_1 \sqsubseteq_{\mathcal{K}} A_1 : \psi_1)$ we get $(A \sqsubseteq_{\mathcal{K}} A_1 : (\psi_1 \cup \psi_2 \cup \psi_3))$. □

Lemma 4.29. *Completion algorithm terminates after an exponential number of steps.*

Proof. The size of V in G is linear in the size of $(\mathcal{N}_{\mathcal{C}} \cup \top)$. The size of the edges of the graph on the other hand is quadratic in $|V|$. Hence, the size of the graph is polynomial in the size of the input. On the other hand, both $\mathcal{N}_{\mathcal{BC}}$ and $\mathcal{N}_{\mathcal{BR}}$ are potentially exponential sets since there are exponentially many contexts in the size of the random variables. Furthermore, since each rule application adds one element to the graph, and the rules are applied only if they change anything, it follows that rule application terminates after an exponential number of steps. □

Since the initial configuration of the graph satisfies the invariants and since rule applications do satisfy invariants we get the soundness of the algorithm.

Lemma 4.30 (Soundness)**.** *Let $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be a $\mathcal{BEL}$ ontology. After running the completion algorithm for $\mathcal{K}$, it holds for all A, $B \in \mathcal{N}_{\mathcal{C}}$ and the contexts ψ that:*

$$(A \sqsubseteq_{\mathcal{K}} B : \psi) \text{ if } (B : \psi) \in S(A)$$

To prove the completeness of the algorithm we construct a probabilistic interpretation based on the completion graph.

Lemma 4.31 (Completeness)**.** *Let $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ be a $\mathcal{BEL}$ ontology. After running the completion algorithm for $\mathcal{K}$, it holds for all A, $B \in \mathcal{N}_{\mathcal{C}}$ and contexts ψ that:*

$$(A \sqsubseteq_{\mathcal{K}} B : \psi) \text{ only if } (B : \psi) \in S(A)$$

Proof. Suppose that $(B : \psi) \notin S(A)$ when no more rule is applicable. Suppose that there are k worlds. We construct a probabilistic interpretation $\mathcal{I}_\mathcal{P} = (\Phi, \mathcal{P}r)$ of $\mathcal{K}$ where $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}k}\} \subseteq \Phi$ and for every contextual interpretation $\mathcal{I}_\mathcal{C} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}k}\}$ it holds that $\mathcal{P}r(\mathcal{I}_\mathcal{C}) = 0$. Furthermore, we define:

- $\Delta^{\mathcal{I}_{\mathcal{C}i}} = V$, $\mathcal{I}_{\mathcal{C}i} \models \mathcal{W}_i$, $\mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_i)$
- $B'^{\mathcal{I}_{\mathcal{C}i}} = \{A' \mid \psi' \subseteq \mathcal{W}_i, (B' : \psi') \in S(A')\}$ for all $B' \in \mathcal{N}_\mathcal{C}$
- $r'^{\mathcal{I}_{\mathcal{C}i}} = \{(A', B') \mid \psi' \subseteq \mathcal{W}_i, (r' : \psi') \in R(A', B')\}$ for all $r' \in \mathcal{N}_\mathcal{R}$

Since $(B : \psi) \notin S(A)$ we get that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (A \sqsubseteq B : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) < 1$$

Therefore, we know that there exists at least one world with positive probability that does not entail this axiom. It is only left to show that $\mathcal{I}_\mathcal{P}$ is a model of $\mathcal{K}$. For every world $\mathcal{W}_j$, the following follows from the construction of $\mathcal{I}_\mathcal{P}$:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models \mathcal{W}_j} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \mathcal{P}r(\mathcal{I}_{\mathcal{C}j}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W}_j) \tag{4.29}$$

Furthermore, we need to show that $\mathcal{I}_\mathcal{P}$ is a *model of* $\mathcal{T}$, i.e., for each $(E \sqsubseteq F : \psi) \in \mathcal{T}$ it satisfies:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1$$

The following holds:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (E \sqsubseteq F : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \sum_{\mathcal{I}_{\mathcal{C}i} \models \psi, \mathcal{I}_{\mathcal{C}i} \models E \sqsubseteq F} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) + \sum_{\mathcal{I}_{\mathcal{C}i} \not\models \psi} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i})$$

Therefore, it is enough to show that every contextual interpretation $\mathcal{I}_{\mathcal{C}i}$, $1 \leq i \leq k$ with positive probability satisfying ψ also satisfies the subsumption. We show that this is indeed the case by showing this over the four types of axioms that may occur in the TBox:

- Suppose the axiom is of type $((A_1 \sqsubseteq B) : \psi) \in \mathcal{T}$. Let $\mathcal{I}_{\mathcal{C}i}$ be a contextual interpretation that satisfies the context ψ. Suppose that $C \in A_1^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi \subseteq \mathcal{W}_i$, i.e., $(C : \psi) \in S(A)$. Since (CR1) no longer applies it holds that $(C : \psi) \in S(B)$. This implies that $C \in B^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi \subseteq \mathcal{W}_i$. Since $\mathcal{I}_{\mathcal{C}i}$ was selected arbitrarily we get that every contextual interpretation with positive probability that satisfies ψ also satisfies the subsumption $(A_1 \sqsubseteq B)$.
- Suppose the axiom is of type $((A_1 \sqcap A_2 \sqsubseteq B) : \psi) \in \mathcal{T}$. Let $\mathcal{I}_{\mathcal{C}i}$ be a contextual interpretation that satisfies the context ψ. Suppose that $C \in (A_1 \sqcap A_2)^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi_1 \subseteq \mathcal{W}_i$, i.e., $(C : \psi_1) \in S(A)$, $(C : \psi_1) \in S(A_2)$. Since (CR2) no longer applies it holds that $(C : \psi_2) \in S(B)$ with $\psi = \psi_1 \cup \psi_2$. This implies that $C \in B^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi_2 \subseteq \mathcal{W}_i$. Since $\mathcal{I}_{\mathcal{C}i}$ was selected arbitrarily we get that every contextual interpretation with positive probability that satisfies ψ also satisfies the subsumption $(A_1 \sqcap A_2 \sqsubseteq B)$.
- Suppose the axiom is of type $(A_1 \sqsubseteq \exists r.B : \psi) \in \mathcal{T}$. Let $\mathcal{I}_{\mathcal{C}i}$ be a contextual interpretation that satisfies the context ψ. Suppose that $C \in A_1^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi_1 \subseteq \mathcal{W}_i$,

i.e., $(C : \psi_1) \in S(A)$. Since (CR3) no longer applies it holds that $(r : \psi_2) \in R(C, B)$ with $\psi = \psi_1 \cup \psi_2$. This implies that $C \in (\exists r.B)^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi_2 \subseteq \mathcal{W}_i$. Since $\mathcal{I}_{\mathcal{C}i}$ was selected arbitrarily we get that every contextual interpretation with positive probability that satisfies ψ also satisfies the subsumption $(A_1 \sqsubseteq \exists r.B)$.

– Suppose the axiom is of type $(\exists r.B_1 \sqsubseteq A_1 : \psi) \in \mathcal{T}$. Let $\mathcal{I}_{\mathcal{C}i}$ be a contextual interpretation that satisfies the context ψ. Suppose that $C \in (\exists r.B_1)^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi_1 \subseteq \mathcal{W}_i$, i.e., $(r : \psi_1) \in S(B)$. Since (CR4) no longer applies it holds that $(A_1 : \psi_2) \in R(C, B)$ with $\psi = \psi_1 \cup \psi_2$. This implies that $C \in A_1^{\mathcal{I}_{\mathcal{C}i}}$ w.r.t. a context $\psi_2 \subseteq \mathcal{W}_i$. Since $\mathcal{I}_{\mathcal{C}i}$ was selected arbitrarily we get that every contextual interpretation with positive probability that satisfies ψ also satisfies the subsumption $(\exists r.B_1 \sqsubseteq A_1)$.

As a result we get that $\mathcal{I}_{\mathcal{P}}$ is a model of the ontology with $\sum_{\mathcal{I}_{\mathcal{C}i} \models (A \sqsubseteq B : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) < 1$. Hence, $\inf\{A \sqsubseteq_{\mathcal{I}_{\mathcal{P}}} B : \psi \mid \mathcal{I}_{\mathcal{P}} \models \mathcal{K}\} < 1$. Consequently, $(A \sqsubseteq_{\mathcal{K}} B : \psi)$ does not hold. □

Given Lemma 4.30 and Lemma 4.31 we get the correctness of the algorithm. As a result, most likely context can be determined as follows.

4. Reading the Subsumptions We have shown that the completion graph contains all subsumption relationships in the given ontology. Hence, ψ is the most likely context in $\mathcal{K}$ for the subsumption $(A \sqsubseteq B)$ iff $\mathcal{P}_{\mathcal{BN}}(\psi) = sup\{\mathcal{P}_{\mathcal{BN}}(\psi')|(B : \psi') \in S(A)\}$. We generalise this to arbitrary concepts C and D as follows. Let C, D be concepts, A, B new named concepts, $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and $\mathcal{K}' = (\{\mathcal{T} \cup \{(A \sqsubseteq C : \{\}), (D \sqsubseteq B : \{\})\}\}, \mathcal{BN})$ a $\mathcal{BEL}$ ontology. Then, Lemma 4.32 holds.

Lemma 4.32. *ψ is the most likely context in $\mathcal{K}$ for the subsumption $(C \sqsubseteq D)$ only if ψ is the most likely context in $\mathcal{K}$ for the subsumption $(A \sqsubseteq B)$.*

Proof. Assume by contradiction that ψ^* is the most likely context in $\mathcal{K}$ for $(A \sqsubseteq B)$ such that $\mathcal{P}_{\mathcal{BN}}(\psi^*) > \mathcal{P}_{\mathcal{BN}}(\psi)$. Then by definition, there exists a consequence $(A \sqsubseteq_{\mathcal{K}'} B : \psi^*)$. But then, $(C \sqsubseteq_{\mathcal{K}} D : \psi^*)$ is also a consequence. Therefore, ψ is not the most likely context for the subsumption since $\mathcal{P}_{\mathcal{BN}}(\psi)$ is not a supremum. Hence, contradiction.

Assume by contradiction that ψ^* is the most likely context in $\mathcal{K}$ for $(A \sqsubseteq B)$ such that $\mathcal{P}_{\mathcal{BN}}(\psi^*) < \mathcal{P}_{\mathcal{BN}}(\psi)$. By definition, we know that $(C \sqsubseteq_{\mathcal{K}} D : \psi)$ is a consequence. But then, $(A \sqsubseteq_{\mathcal{K}'} B : \psi)$ is also a consequence. Therefore, ψ^* is not the most likely context for the subsumption since $\mathcal{P}_{\mathcal{BN}}(\psi^*)$ is not a supremum. Hence, contradiction. □

Theorem 4.33. *Most likely context can be decided in EXPTIME.*

Proof. It has already been shown that the completion algorithm terminates in exponential time. For each consequence that has been generated by the modified completion algorithm we need to check how likely it is, which requires making inference in Bayesian networks, which is NP-hard. Hence, the combined complexity is EXPTIME. □

Chapter 5

Generalising the framework to DL family: $\mathcal{B}$ + DL

So far, we have provided a Bayesian extension of the description logic $\mathcal{EL}$. Even though we have restricted the scenario to the ontologies with TBoxes, we now postulate that this restriction can be removed easily. Furthermore, the scenario can also be extended to more expressive DLs. For our framework to be extended, it is sufficient to have a classic-valued, monotonic logic, namely, one where the consequence-based entailment can be checked.

We will briefly review expressive probabilistic DLs before presenting our generalisation. $\mathcal{ALC}$ is the classic DL that uses the constructors $\top$, $\bot$, $\sqcap$, $\sqcup$, $\neg$, $\exists$, $\forall$. Two of the early probabilistic DLs, by Heinsohn [37] and Jaeger [38] propose a probabilistic extension of $\mathcal{ALC}$ and define probabilistic inference services for reasoning. Both of them support probabilistic concepts at terminological level, but do not fully support assertions over individuals. A more recent work that extends $\mathcal{ALC}$ has been conducted by Dürig and Studer. Their formalism allows assertional probabilistic knowledge, but does not allow terminological probabilistic knowledge. The main reasoning problem is consistency w.r.t. the assertional knowledge, for which no decidability result is given. Lukasiewicz [39] provides a framework for more expressive probabilistic DLs, that contain additional data, concept and axiom types. His framework is mainly based on the notion of semantic division of the probabilistic ontologies. Based on this assumption some reasoning problems have been defined. The reasoning problems are based on finding intervals of probabilities which is different from our approach. A set of algorithms to solve these problems are defined and a set of complexity results are provided. Importantly, [39] is one of the few formalisms that allows probabilistic extensions at both assertional and terminological level. Different from $\mathcal{BDL}$, the theory in the background is probabilistic lexicographic entailment taken from probabilistic default reasoning.

In this section we do not provide the individual syntax and semantics of the Bayesian DLs. Instead we define a general syntax and semantics that is built on top of the semantics of the concept language of the underlying DL. We additionally show that first three decision problems that we have defined for $\mathcal{BEL}$ can be extended towards Bayesian DLs, since these algorithms use the classic DL reasoners as oracles.

5.1 Towards More Expressive Bayesian DLs

The syntax and semantics of $\mathcal{BEL}$ are defined in a way that it can easily be extended towards other description logics. In defining the syntax and semantics of Bayesian DLs we have the following strategy. First, we take a DL that is classic-valued and monotonic. Then, we define its Bayesian extension without specifying the underlying logic. We denote the classic-valued, monotonic description logics as $\mathcal{DL}$ and their Bayesian extensions as $\mathcal{BDL}$.

The $\mathcal{BDL}$ concept language is defined exactly as the $\mathcal{DL}$ concept language. The contextual interpretations extend the standard $\mathcal{DL}$ interpretations by additionally mapping the random variables to their domain of values. The $\mathcal{BDL}$ syntax extends the $\mathcal{DL}$ syntax with probabilistic axioms. Furthermore, instead of defining different sets for different types of axioms, as it is common in description logics, we define a general set, named general box, which contains all types of axioms.

Definition 5.1 (*Axiom, General Box and Ontology*)**.** Let $\mathcal{DL}$ be a classic-valued description logic that is monotonic. A *probabilistic axiom* is an expression of the form $(\alpha : \psi)$, where α is an axiom of any type that can be formed in $\mathcal{DL}$ and ψ a context. A contextual interpretation $\mathcal{I}_\mathcal{C}$ satisfies the axiom $(\alpha : \psi)$, denoted as $\mathcal{I}_\mathcal{C} \models (\alpha : \psi)$, iff it holds that:

$$(\mathcal{I}_\mathcal{C} \models \psi) \rightarrow (\mathcal{I}_\mathcal{C} \models \alpha)$$

A $\mathcal{DL}$ *general box* $\mathcal{T}$ is a finite set of axioms that can be formed in $\mathcal{DL}$. A $\mathcal{DL}$ *ontology* $\mathcal{K}$ is a pair $(\mathcal{T}, \mathcal{BN})$ where $\mathcal{T}$ is a $\mathcal{DL}$ general box and $\mathcal{BN}$ is a Bayesian network.

The semantics of $\mathcal{BDL}$ is defined analogously to the semantics of $\mathcal{BEL}$. The only difference is that we have a general box instead of a TBox.

Definition 5.2 (Semantics of $\mathcal{BDL}$)**.** A *probabilistic interpretation* $\mathcal{I}_\mathcal{P}$ is a pair $(\Phi, \mathcal{P}r)$ where Φ is a set of contextual interpretations and $\mathcal{P}r$ is a probability distribution over Φ such that: $\{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\} \subseteq \Phi$, $1 \leq i \leq n$ where n is finite and for every contextual interpretation $\mathcal{I}_\mathcal{C} \notin \{\mathcal{I}_{\mathcal{C}1}, ..., \mathcal{I}_{\mathcal{C}n}\}$ it holds that $\mathcal{P}r(\mathcal{I}_\mathcal{C}) = 0$. A *probabilistic interpretation* $\mathcal{I}_\mathcal{P}$ is a *model of* $\mathcal{T}$ iff for each $(\alpha : \psi) \in \mathcal{T}$ it satisfies:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models (\alpha:\psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = 1$$

$\mathcal{I}_\mathcal{P}$ is *consistent with a Bayesian network* $\mathcal{BN}$ iff for every world $\mathcal{W}$ in $\mathcal{BN}$ it holds that:

$$\sum_{\mathcal{I}_{\mathcal{C}i} \models \mathcal{W}} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i}) = \mathcal{P}_{\mathcal{BN}}(\mathcal{W})$$

A probabilistic interpretation $\mathcal{I}_\mathcal{P}$ is a *model of a* $\mathcal{DL}$ *ontology* $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ iff $\mathcal{I}_\mathcal{P}$ is a model of $\mathcal{T}$ and consistent with $\mathcal{BN}$. A $\mathcal{DL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ is *satisfiable* iff it has a probabilistic model.

We introduced the syntax and semantics of $\mathcal{BDL}$ w.r.t a description logic $\mathcal{DL}$, which is classic-valued and monotonic. In Section 5.2, we lift the reasoning problems defined for $\mathcal{BEL}$ to $\mathcal{BDL}$.

5.2 Reasoning in $\mathcal{BDL}$

Reasoning procedures for $\mathcal{BDL}$ can be seen as the lifted reasoning procedures that have been defined for $\mathcal{BEL}$. Since $\mathcal{DL}$ is not specified other than being classic-valued and monotonic, we generalise the subsumption to any kind of axiom that can be formed in $\mathcal{DL}$. Hence, we define the probabilities of axioms instead of subsumptions.

Definition 5.3 (Probabilities of axioms)**.** Given a $\mathcal{BDL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and a probabilistic interpretation $\mathcal{I}_{\mathcal{P}} = (\Phi, \mathcal{P}r)$. The probability of the axiom $(\alpha : \psi)$ w.r.t. $\mathcal{I}_{\mathcal{P}}$, denoted as $\mathcal{P}(\alpha_{\mathcal{I}_{\mathcal{P}}} : \psi)$ is:

$$\mathcal{P}(\alpha_{\mathcal{I}_{\mathcal{P}}} : \psi) = \sum_{\mathcal{I}_{\mathcal{C}i} \models (\alpha_{\mathcal{I}_{\mathcal{P}}} : \psi)} \mathcal{P}r(\mathcal{I}_{\mathcal{C}i})$$

The probability of α w.r.t. $\mathcal{K}$, denoted as $\mathcal{P}(\alpha_{\mathcal{K}} : \psi)$ is:

$$\mathcal{P}(\alpha_{\mathcal{K}} : \psi) = \inf\{\mathcal{P}(\alpha_{\mathcal{I}_{\mathcal{P}}} : \psi) | \mathcal{I}_{\mathcal{P}} \models \mathcal{K}\}$$

$\mathcal{P}(\alpha_{\mathcal{I}_{\mathcal{P}}})$ is written short for $\mathcal{P}(\alpha_{\mathcal{I}_{\mathcal{P}}} : \{\})$ and $\mathcal{P}(\alpha_{\mathcal{K}})$ for $\mathcal{P}(\alpha_{\mathcal{K}} : \{\})$.

Analogously, we can determine whether a given probabilistic axiom $(\alpha_{\mathcal{K}} : \psi)$ is a consequence of an ontology.

Definition 5.4 (Axiom as a consequence)**.** Given a $\mathcal{BDL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$, $(\alpha_{\mathcal{K}} : \psi)$ is a consequence of $\mathcal{K}$, denoted as $(\alpha_{\mathcal{K}} : \psi)$ iff it holds that $\mathcal{P}(\alpha_{\mathcal{K}} : \psi) = 1$.

It is now possible to define the reasoning procedures. Since subsumptions are replaced with general axioms, instead of checking whether a subsumption is entailed we check whether an axiom is entailed.

Definition 5.5 (Reasoning problems)**.** Given a $\mathcal{BDL}$ ontology $\mathcal{K} = (\mathcal{T}, \mathcal{BN})$ and an axiom α, we say that α *is precisely entailed with probability* p w.r.t. $\mathcal{K}$ if $p = \mathcal{P}(\alpha_{\mathcal{K}})$. We define two special cases of precise entailment, i.e., positive entailment and certain entailment as follows. α *is entailed positively* w.r.t. $\mathcal{K}$ if $\mathcal{P}(\alpha_{\mathcal{K}}) > 0$ and α *is entailed certainly* w.r.t. $\mathcal{K}$ if $\mathcal{P}(\alpha_{\mathcal{K}}) = 1$. Given an axiom α that is positively entailed, we say that ψ *is the most likely context* for the entailment α in $\mathcal{K}$ if $\mathcal{P}_{\mathcal{BN}}(\psi) = sup\{\mathcal{P}_{\mathcal{BN}}(\psi') | \mathcal{P}(\alpha_{\mathcal{K}} : \psi')\}$.

Generally, there are two approaches in developing algorithms to reason over the extensions of DLs. First approach, known as the black-box approach, is to use the results of the standard reasoners developed for the underlying logic to develop a new reasoning procedure over the extended logic. This is also called using DL reasoners as oracles. The second approach, known as the glass-box approach, is to modify the internals of a DL reasoner to develop a new reasoning procedure. Both of the approaches has its own advantages and disadvantages. The main advantage of the black-box approach is that it can be generalised easily. On the other hand, the glass box approach concentrates on a more specific problem by taking into account its internal structure which usually leads to more optimal procedures. However, since it highly depends on the structure of the particular problem it cannot be generalised to other structures easily.

Algorithm 1, Algorithm 2 and Algorithm 3 are based on the black box approach since they do not make use of any particular internal structure of $\mathcal{EL}$. The algorithms basically use an $\mathcal{EL}$ reasoner as an oracle to check whether an axiom is entailed from the constructed $\mathcal{EL}$ TBox. Therefore, it can be easily extended to DLs where entailment checking can be done.

Note that most likely context cannot be generalised since it modifies the internals of a DL reasoner such that it yields the results during the computation. Within the scope of this thesis, we leave most likely context in $\mathcal{BDL}$ open.

We assumed that the description logic $\mathcal{DL}$ that underlies $\mathcal{BDL}$ is a monotonic logic. Therefore, we know that we can check the entailment for a given axiom. Hence, the algorithms are changed so that they use the reasoners developed for the underlying $\mathcal{DL}$ to check entailment. These yield in sound and complete algorithms for $\mathcal{BDL}$. The only difference is of course in the complexity results, since the underlying DLs may have different computational complexities for entailment checking.

It is important to note the fact that any $\mathcal{BDL}$ is decidable provided that entailment checking in $\mathcal{DL}$ is decidable. Moreover, precise entailment can be decided in $\mathrm{PSPACE}^{\mathrm{C}}$, positive entailment in NP^{C} and certain entailment in $\mathrm{coNP}^{\mathrm{C}}$, where C is the complexity of deciding the entailment in $\mathcal{DL}$.

Let us explain these w.r.t. some known DLs. We have already given the complexities for $\mathcal{BEL}$. One interesting question is how far $\mathcal{BEL}$ can be pushed without any change in the complexity result of the three reasoning problems. It is known that $\mathcal{EL}^{++}$ has polynomial time entailment checking. As a result, $\mathcal{BEL}^{++}$ has the same complexities for the same reasoning problems. $\mathcal{EL}^{++}$ as well as other DLs in which entailment checking can be done better than PSPACE, these complexity results have the following meaning. The computational complexities are w.r.t. the size of the random variables given by the Bayesian network. Hence, it is possible to have a large set of axioms, as it is usually desired in DLs, while keeping the Bayesian network minimal, which will lead to efficient reasoning procedures.

Consider the DL $\mathcal{ALC}$ in which entailment can be checked in EXPTIME. As a result, the complexity of reasoning in Bayesian $\mathcal{ALC}$ stays same as the complexity of reasoning in standard $\mathcal{ALC}$. Indeed, this is also the fact for other expressive DLs.

We have shown that our scenario can be generalised towards more expressive logics. Furthermore, the reasoning problems, except most likely context, can be solved analogously to $\mathcal{BEL}$ since the algorithms use a black-box approach. In Chapter 6, we summarize the results and offer some future research directions.

Chapter 6

Further Research and Summary

In this work, we provided a combination of two widely studied KR formalisms in order to achieve context based reasoning over uncertain contexts. The main contributions can be listed as follows.

- We provided a context-sensitive extension of DLs that is different from existing approaches. The current work is especially marked by its considering uncertain domains for the contexts, an area that was not explored before. Furthermore, it also differs from existing probabilistic DLs in a number of other ways aforementioned in relevant sections of the text.
- We have provided decision problems w.r.t. the probabilities of the contexts and developed algorithms to solve these. We have given decidability results and a set of complexity results. For $\mathcal{BEL}$, we have shown that precise subsumption is in PSPACE, positive and certain subsumption is NP-complete. Lastly, we have provided an exponential algorithm to solve most likely context.
- We generalised our scenario to classic-valued, monotonic DLs. Precise, positive and certain subsumption problems are lifted and complexity results for these generalisations w.r.t. the existing algorithms are provided.

There are several open avenues for future work to consider. Some of them can be listed as follows.

- The hardness results for precise subsumption and most likely context are not given. This is left as an open problem.
- In the scope of this thesis, we do not provide any implementation for any of the algorithms provided. It is important to implement these algorithms considering real application domains, for instance the bio-medical domain. This points to yet another future dimension; one that needs to be discussed in close collaboration with the experts of such knowledge domains.
- A downside of Algorithm 1 is that it is complexity-wise strict, i.e., it does not perform better in the best case. We strongly think that practically-oriented algorithms can

be developed, which may have bad worst-case behaviour, but could provide good average-case behaviour as it is the case for many DL reasoners.

- The completion-like algorithm does not suffer from strictness. It may have a good average case performance. We think that this algorithm can be optimised further. For instance, one can choose set labels prior to adding them, and ignore some irrelevant ones. This is also left as a future work.
- Although we have only considered ontologies, it might be possible to adopt this formalism to Databases, which gives rise to another future work.
- We think that this formalism can also attract researchers from the Bayesian networks community. It is an open question enquiry whether and how the framework can be used for learning tasks in Bayesian networks from the perspective of Bayesian network community.

Context-sensitive formalisms as well as probabilistic formalisms address important characteristics of the knowledge domains. We think that this work proposes promising questions which needs to be further studied. These questions may offer new horizons for KR which can address some problems of KR as well as open new ones.

Bibliography

[1] F. Baader. Description logic terminology. In F. Baader, D. Calvanese, D. McGuinness, Daniele Nardi, and P. F. Patel-Schneider, editors, *The Description Logic Handbook: Theory, Implementation, and Applications*, pages 485–495. Cambridge University Press, 2003.

[2] R. Shearer, B. Motik, and I. Horrocks. Hermit: A highly-efficient owl reasoner. In C. Dolbear, A. Ruttenberg, and U. Sattler, editors, *OWLED*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.

[3] D. Tsarkov and I. Horrocks. Description logic reasoner: System description. In U. Furbach and N. Shankar, editors, *International Joint Conference on Automated Reasoning, IJCAR 2006. Proceedings*, volume 4130 of *Lecture Notes in Computer Science*, pages 292–297, Berlin, Heidelberg, 2006. Springer.

[4] F. Baader, C. Lutz, and B. Suntisrivaraporn. CEL—a polynomial-time reasoner for life science ontologies. In U. Furbach and N. Shankar, editors, *Proceedings of the 3rd International Joint Conference on Automated Reasoning (IJCAR'06)*, volume 4130 of *Lecture Notes in Artificial Intelligence*, pages 287–291. Springer-Verlag, 2006.

[5] G. Boole. *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities.* (reprint: Dover Publications, New York, 1958), 1854.

[6] N. J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28(1):71–88, February 1986.

[7] R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87:78–128, 1990.

[8] A. Frisch and P. Haddawy. Anytime deduction for probabilistic logic. *Artificial Intelligence*, 69:93–122, 1994.

[9] T. Lukasiewicz. Local probabilistic deduction from taxonomic and probabilistic knowledge-bases over conjunctive events. *International Journal of Approximate Reasoning*, 21(1):23 – 61, 1999.

[10] T. Lukasiewicz. Probabilistic deduction with conditional constraints over basic events. *Journal of Artificial Intelligence Researh*, 10:380–391, 1999.

[11] T. Lukasiewicz. Probabilistic logic programming with conditional constraints. *ACM Trans. Comput. Logic*, 2(3):289–339, July 2001.

[12] T. Lukasiewicz. Probabilistic default reasoning with conditional constraints. *Annual Mathematical Artificial Intelligence*, 34:200–2, 2000.

[13] T. Lukasiewicz and U. Straccia. Managing uncertainty and vagueness in description logics for the semantic web. *Web Semant.*, 6(4):291–308, November 2008.

[14] J. McCarthy. Notes on formalizing context. In *Proceedings of the 13th international joint conference on Artifical intelligence - Volume 1*, IJCAI'93, pages 555–560, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.

[15] S. Klarman and V. Gutirrez-Basulto. $\mathcal{ALC}_{\mathcal{ALC}}$: A context description logic. In T. Janhunen and I. Niemelä, editors, *Logics in Artificial Intelligence*, volume 6341 of *Lecture Notes in Computer Science*, pages 208–220. Springer Berlin Heidelberg, 2010.

[16] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05*, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.

[17] K.A. Spackman. Managing clinical terminology hierarchies using algorithmic calculation of subsumption: Experience with SNOMED-RT. *J. of the American Medical Informatics Association*, 2000.

[18] M. Ashburner. Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.

[19] A. L. Rector and I. R. Horrocks. Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions. In *In Proceedings of the Workshop on Ontological Engineering, AAAI Spring Symposium.* AAAI Press, 1997.

[20] T. Bayes and R. Price. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions (1683-1775)*, 53:370–418, 1763.

[21] P. S. Laplace. *Theorie Analytique des Probabilites.* Courcier, Paris, 1812.

[22] P. S. Laplace. *Essai Philosophique Sur Les Probabilites.* Cambridge Univ Pr, 2011.

[23] D. Koller, A. Levy, and A. Pfeffer. P-classic: A tractable probabilistic description logic. In *Proceedings of AAAI-97*, pages 390–397, 1997.

[24] J. Pearl. Bayesian networks. pages 149–153, 1998.

[25] A. Darwiche. *Modeling and Reasoning with Bayesian Networks.* Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[26] G. F. Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 42(23):393 – 405, 1990.

[27] P. M. Yelland. An alternative combination of bayesian networks and description logics. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *KR*, pages 225–234. Morgan Kaufmann, 2000.

[28] M. Niepert, J. Noessner, and H. Stuckenschmidt. Log-linear description logics. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*, IJCAI'11, pages 2153–2158. AAAI Press, 2011.

[29] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, February 2006.

[30] T. Lukasiewicz, M. V. Martinez, G. Orsi, and G. I. Simari. Heuristic ranking in tightly coupled probabilistic description logics. *CoRR*, abs/1210.4894, 2012.

[31] C. Lutz and L. Schröder. Probabilistic description logics for subjective uncertainty. In F. Lin, U. Sattler, and M. Truszczynski, editors, *KR*. AAAI Press, 2010.

[32] C. dAmato, N. Fanizzi, and T. Lukasiewicz. Tractable reasoning with bayesian description logics. In S. Greco and T. Lukasiewicz, editors, *Scalable Uncertainty Management*, volume 5291 of *Lecture Notes in Computer Science*, pages 146–159. Springer Berlin Heidelberg, 2008.

[33] E. Bellodi, E. Lamma, F. Riguzzi, and S. Albani. A distribution semantics for probabilistic ontologies. In F. Bobillo, R. N. Carvalho, P.C. G. da Costa, C. d'Amato, N. Fanizzi, K. B. Laskey, T. Lukasiewicz, T. Martin, and M. Nickles, editors, *URSW*, volume 778 of *CEUR Workshop Proceedings*, pages 75–86. CEUR-WS.org, 2011.

[34] F. Riguzzi, E. Bellodi, E. Lamma, and R. Zese. Bundle: A reasoner for probabilistic ontologies. In W. Faber and D. Lembo, editors, *Web Reasoning and Rule Systems*, volume 7994 of *Lecture Notes in Computer Science*, pages 183–197. Springer Berlin Heidelberg, 2013.

[35] F. Baader, R. Peñaloza, and B. Suntisrivaraporn. Pinpointing in the description logic $\mathcal{EL}$. In *Proceedings of the 30th German Conference on Artificial Intelligence (KI2007)*, volume 4667 of *Lecture Notes in Artificial Intelligence*, pages 52–67, Osnabrück, Germany, 2007. Springer-Verlag.

[36] R. Peñaloza and B. Sertkaya. On the complexity of axiom pinpointing in the el family of description logics. In F. Lin, U. Sattler, and M. Truszczynski, editors, *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*. AAAI Press, 2010.

[37] J. Heinsohn. Probabilistic description logics. In *Proceedings of UAI-94*, pages 311–318, 1994.

[38] M. Jaeger. Probabilistic reasoning in terminological logics. In J. Doyle, E. Sandewall, and P. Torasso, editors, *KR*, pages 305–316. Morgan Kaufmann, 1994.

[39] T. Lukasiewicz. Expressive probabilistic description logics. *Artificial Intelligence*, 172(67):852 – 883, 2008.