Making Decisions over Contextual Ontologies

Erman Acar[1], Rafael Peñaloza[2], and Livia Predoiu[3]

[1] Vrije Universiteit Amsterdam `erman.acar@vu.nl`
[2] University of Milano-Bicocca `rafael.penaloza@unimib.it`
[3] Free University of Bozen-Bolzano `livia.predoiu@unibz.it`

Abstract. Various probabilistic description logics (DLs) have been proposed for dealing with the uncertainty endemic to many domain knowledge representation scenarios. A particular class of such formalisms focuses on representing knowledge that is certain, but holds only in some uncertain contexts. In this paper, we consider an extension of those formalisms that allows an agent to influence the choice of the context and minimise its subjective cost. This is achieved through a combination of the light-weight DL $\mathcal{EL}$ and influence diagrams, a graphical model for representing decision situations, and their potential costs, under uncertainty.

1 Introduction

A well-recognised limitation of classical description logics (DLs) [4] is their inability to deal with uncertainty. In order to model different aspects of knowledge domains where uncertainty is unavoidable, such as in the bio-medical sciences, many probabilistic extensions of DLs have been proposed in the literature [11,13,14,16]. Among them, a prominent example are Bayesian DLs [6,8–10], which provide a means for expressing complex probabilistic and logical dependencies between axioms. For example, in these logics it is easy to express that two axioms must always appear together, or that if one axiom holds, then the likelihood of another one holding is some probability p.

The expressive power of Bayesian DL arises from combining a set of (classical) DL ontologies (called contexts) with a Bayesian network (BN) [15] representing the joint probability distribution of these ontologies to hold. This allows to reason about the likelihood of a consequence to hold, but also update the beliefs about the probabilities of the contexts. However, this remains as a passive attitude towards knowledge. In practice, one would like an agent to be able to make choices depending on its knowledge and observations. Some earlier works [1, 2] that employed (probabilistic) DLs in a decision-theoretic setting, however, neither addressed observations, nor contextual reasoning. Hence, they stay completely orthogonal to our work.

Influence diagrams (IDs) [17] are a generalisation of BNs aimed at modelling potential decisions made by an agent and their associated cost. As a typical toy example, an agent has to decide whether to go for a picnic or not, based on the weather forecast that depends (unreliably) on the actual weather. The overall

cost to the agent will depend on their choice and on the state of the weather (see Figure 1). In this paper we propose an extension of the Bayesian DL $\mathcal{BEL}$ [7] to allow for agent decisions combining influence diagrams with the light-weight DL $\mathcal{EL}$ [3]. We call this logic ID-$\mathcal{EL}$.

In ID-$\mathcal{EL}$, the contexts consider the uncertainty in the network, as well as the potential choices from the agent and, obviously, also their associated cost. More importantly, the ontological knowledge can be used as evidence about the potential context, thus modifying the underlying probabilities. We study the reasoning problems associated with the selection, by the agent, of a strategy that minimises its expected cost given such evidence.

2   Influence Diagrams

Influence diagrams (IDs) [17] are a generalisation of Bayesian networks (BNs) [15], which contain three types of nodes: *chance nodes* that reflect the uncertainty of the environment as in BNs; *decision nodes*, which express the choices made by an agent in response to the environment; and a *cost node* (sometimes also called a *utility node*), which reflects the cost (or utility) of a given outcome. From a formal perspective, each of these nodes represents a discrete random variable, and the main difference is how this variable is interpreted or used within the network.

Formally, an *influence diagram* is a pair $\mathcal{D} = (G, \Phi)$ where $G = (V \cup \{c\}, E)$ is a directed acyclic graph (DAG), whose nodes V are partitioned into two sets B and D of *chance nodes* (or *Bayesian nodes*), and *decision nodes*, respectively, and c is a single *cost node*. For simplicity, we assume w.l.o.g. that all nodes in V are Boolean random variables.[4] The cost node c has no outgoing edges, and represents a cost function from the valuations of its parent nodes to a finite set $\mathsf{val}(c) \subseteq \mathbb{R}$ of values. For every node $v \in V \cup \{c\}$, $\pi(v)$ denotes the parents of v. Given a decision node $d \in D$, $\mathsf{d\text{-}anc}(d)$ is the set of all decision ancestors of d, and

$$\mathsf{infl}(d) := \mathsf{d\text{-}anc}(d) \cup \pi(d).$$

is the *influence set* of d. When the set of nodes in $\mathcal{D}$ is $V \cup \{c\}$, we say that $\mathcal{D}$ is an ID *over* V. The second part of the tuple $\mathcal{D}$, Φ, is a class of conditional probability distribution tables $P(v \mid \pi(v))$, one for each chance node $v \in B$ given its parents. Notice that there is no probability distribution associated with decision nodes, and recall that the node c represents a function from the class of all valuations of $\pi(c)$ to $\mathbb{R}$.

IDs are represented graphically using circles to denote chance nodes, squares for decision nodes, and a diamond for the cost node; see Figure 1. Seen in this

---

[4] In general, chance and decision nodes can be arbitrary finite random variables. By considering only the Boolean case, we greatly simplify the notation and presentation, without affecting the generality of the approach.
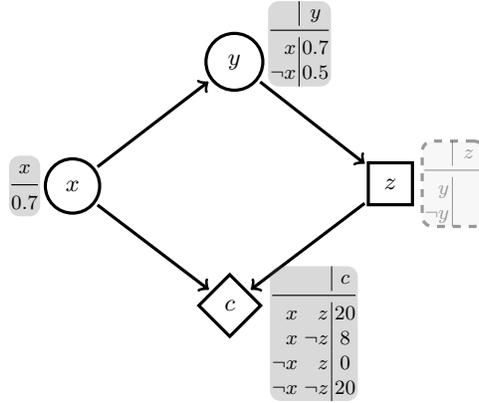
Fig. 1. An influence diagram; x, y are choice nodes, z is a decision node, and c is the only cost node with $\mathsf{val}(c) = \{0, 8, 20\}$. The probability table for the choice node z is not specified (represented through transparency).

way, an ID can be thought of as an *incomplete* BN where some of the nodes are missing their conditional probability tables, given their parents.[5]

If these tables are added to the ID, then one could derive the joint probability distribution of all the variables in V using the standard chain rule from BNs

$$P_{\mathcal{D}}(V) = \prod_{v \in V} P(v \mid \pi(v)).$$

Instead, in an ID, the decision nodes correspond to choices that an agent can make based on the information available. The actual response of the agent is called a strategy.

Definition 1 (Strategy). *A (local)* strategy *on a decision node $d \in D$ is a conditional probability distribution of d given its influence set* $\mathsf{infl}(d)$. *A (global)* strategy *on the ID $\mathcal{D}$ is a set of local strategies, containing one for each $d \in D$. A local or global strategy is* pure *if it only assigns probabilities 0 or 1.*

Note that the strategy at a decision node does not depend on its parents only, but on its whole influence set; that is, it depends on its decision ancestors. Intuitively, we can see the direction of the DAG as a precedence in the choices made. Hence, every decision depends also on the choices made earlier. This can be seen as having implicit connections between the node d and its influence set. This assumption, known as no-forgetting, is commonly used in IDs, thus we include it in our formalism. However, removing it would have no effect over the results in this work. In the ID from Figure 1, one possible pure strategy would be to assign $P(z \mid \neg y) = P(\neg z \mid y) = 1$. To distinguish pure and general

---

[5] Note that the utility function can be seen as a special kind of probability distribution over the set $\mathsf{val}(c)$, where probabilities are always 0 or 1.

strategies, we often call the former an *action*. We denote by $P_{\mathcal{D}(S)}$ the probability distribution obtained by adopting the strategy S in $\mathcal{D}$.

Clearly, an agent has a large class of strategies from which to choose. Which one is better depends on the probability of paying different costs given the chosen strategy. One usual approach is to try to minimise the expected cost.

Definition 2 (Expected cost). *Given a global strategy S on the ID $\mathcal{D}$, the expected cost of S w.r.t. $\mathcal{D}$ is*

$$E[\mathcal{D} \mid S] := \sum_{r \in \mathsf{val}(c)} r \cdot P_{\mathcal{D}(S)}(c = r).$$

For example, if we use the aforementioned strategy S in the ID $\mathcal{D}$ from Figure 1, we obtain that

$$P_{\mathcal{D}(S)} = \begin{cases} 0.15 & r = 0 \\ 0.49 & r = 8 \\ 0.36 & r = 20. \end{cases}$$

Hence, expected cost $E[\mathcal{D} \mid S] = 0 \cdot 0.15 + 8 \cdot 0.49 + 20 \cdot 0.36 = 11.12$.

As mentioned, strategies in IDs are often targeted to minimising the expected cost on the resulting network. However, other kinds of problems can also be considered over such a network, such as finding the most likely cost, or maximising the probability of the minimum cost. If we limit ourselves to pure strategies only, then one can verify that the strategy S in our running example is in fact the one which minimises the expected cost. On the other hand, the strategy S' which assigns $P(z \mid y) = P(z \mid \neg y) = 1$ maximises the probability of observing the least possible cost 0: $P_{\mathcal{D}(S')}(c = 0) = 0.3$.

In the next section, we propose a combination of IDs with the description logic $\mathcal{EL}$ and later study some of its reasoning problems.

3 Influence Diagrams with Contextual $\mathcal{EL}$ Ontologies

We now introduce a new logic that combines the light-weight DL $\mathcal{EL}$ with an influence diagram to allow reasoning and deriving strategies according to observed knowledge. The connection between the two formalisms is based on adding a contextual annotation to every axiom, expressing in which circumstances it is required to hold. We formalise this next.

Definition 3 (KB). *Consider a finite set V of* contextual variables *(or variables for short) and two disjoint sets N_C and N_R or* concept *and* role names, *also disjoint with V.* ID-$\mathcal{EL}$ concepts *are constructed through the grammar rule $C ::= A \mid \top \mid C \sqcap C \mid \exists r.C$, where $A \in N_C$ and $r \in N_R$. A (contextual)* general concept inclusion *(V-GCI) is an expression of the form $\langle C \sqsubseteq D : \varphi \rangle$ where C, D are two ID-$\mathcal{EL}$ concepts and φ is a propositional formula over V. A V-TBox is a finite set of V-GCIs. An ID-$\mathcal{EL}$* knowledge base *(KB) is a pair $\mathcal{K} = (\mathcal{D}, \mathcal{T})$, where $\mathcal{D}$ is an ID over V and $\mathcal{T}$ is a V-TBox.*

As with other context-based DLs introduced in the past [5,9], the idea is that a V-GCI is only required to hold when its context φ is satisfied. This intuition is formalised via a possible world semantics that uses so-called V-interpretations. These combine classical DL interpretations with propositional valuations to link the GCIs with their contexts.

Definition 4 (Semantics). *A V-interpretation is a triple $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set called the* domain, $\mathcal{V}^{\mathcal{I}} : V \to \{0,1\}$ *is a* valuation *of V, and $\cdot^{\mathcal{I}}$ is the* interpretation function *that maps every $A \in N_C$ to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and every $r \in N_R$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. This function is extended to complex concepts as usual in $\mathcal{EL}$.*

The V-interpretation $\mathcal{I}$ satisfies the V-GCI $\langle C \sqsubseteq D : \varphi \rangle$ ($\mathcal{I} \models \langle C \sqsubseteq D : \varphi \rangle$) iff $\mathcal{V}^{\mathcal{I}} \not\models \varphi$ or $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. It is a model *of the V-TBox $\mathcal{T}$ iff it satisfies all V-GCIs in $\mathcal{T}$.*

When there is no ambiguity, we omit the prefix V and speak of e.g., interpretations or TBoxes.

Clearly, the probabilistic DL $\mathcal{BEL}$ [9]—which combines a contextual ontology with a Bayesian network—is a special case of ID-$\mathcal{EL}$, where there are no decision nodes, and the cost node is ignored (e.g., it may be disconnected from the rest of the DAG). As in that special case, it is often useful to consider the classical $\mathcal{EL}$ TBox induced by a valuation of the variables in V.

Definition 5 (Restricted KB). *Let $\mathcal{K} = (\mathcal{D}, \mathcal{T})$ be a KB, and $\mathcal{W}$ a valuation of the variables in V. The* restriction of $\mathcal{T}$ to $\mathcal{W}$ *is the $\mathcal{EL}$ TBox*

$$\mathcal{T}_{\mathcal{W}} := \{C \sqsubseteq D \mid \langle C \sqsubseteq D : \varphi \rangle \in \mathcal{T}, \mathcal{W} \models \varphi\}.$$

To consider the uncertainty associated with the contexts, $\mathcal{BEL}$ defines a possible world semantics where each world is associated with a probability that needs to be compatible with the probability distribution of the nodes. In ID-$\mathcal{EL}$ this definition cannot be applied directly, because the actual probability distribution is underspecified, and depends on the strategy chosen by the agent. Thus, we consider probabilistic models that are parameterised w.r.t. a strategy.

Definition 6 (Probabilistic model). *A* probabilistic interpretation *is a pair $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$, where $\mathfrak{I}$ is a finite set of V-interpretations and $P_{\mathfrak{I}}$ is a probability distribution over $\mathfrak{I}$. This probabilistic interpretation is a* model of the TBox $\mathcal{T}$ *if every $\mathcal{I} \in \mathfrak{I}$ is a model of $\mathcal{T}$.*

Given an ID $\mathcal{D}$ and a strategy S on $\mathcal{D}$, the probabilistic interpretation $\mathcal{P}$ is consistent *with $\mathcal{D}$ w.r.t. S if for every possible valuation $\mathcal{W}$ of the variables in V it holds that $P_{\mathcal{D}(S)}(\mathcal{W}) = \sum_{\mathcal{I} \in \mathfrak{I}, \mathcal{V}^{\mathcal{I}} = \mathcal{W}} P_{\mathfrak{I}}(\mathcal{I})$.*

$\mathcal{P}$ is a model *of the KB $\mathcal{K} = (\mathcal{D}, \mathcal{T})$ w.r.t. the strategy S (denoted as $\mathcal{P} \models_S \mathcal{K}$) iff it is a model of $\mathcal{T}$ and consistent with $\mathcal{D}$ w.r.t. S.*

Example 7. Let $\mathcal{K}_{\mathsf{exa}} = (\mathcal{D}, \mathcal{T}_{\mathsf{exa}})$ be the ID-$\mathcal{EL}$ KB where $\mathcal{D}$ is the ID in Figure 1, and

$$\mathcal{T}_{\mathsf{exa}} := \{\langle A \sqsubseteq B : \neg x \vee z \rangle, \ \langle B \sqsubseteq C : \neg z \vee x \rangle, \ \langle B \sqsubseteq D : \neg x \wedge z \rangle\}.$$
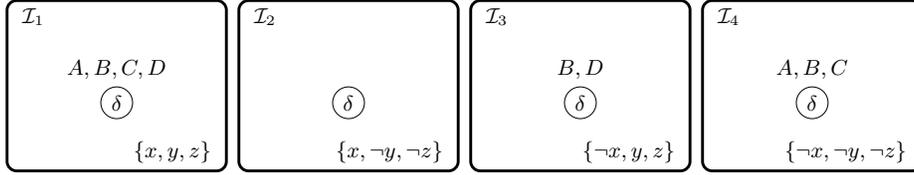
Fig. 2. Four V-interpretations that satisfy the TBox $\mathcal{T}_{\mathsf{exa}}$ from Example 7.

One possiblevaluation of the variables in V is $\mathcal{W}_{\mathsf{exa}} = \{\neg x, y, \neg z\}$. The interpretation $\mathcal{I}_{\mathsf{exa}} = (\{\delta\}, \cdot^{\mathcal{I}}_{\mathsf{exa}}, \mathcal{W}_{\mathsf{exa}})$ where $A^{\mathcal{I}}_{\mathsf{exa}} = B^{\mathcal{I}}_{\mathsf{exa}} = \{\delta\}$ and $C^{\mathcal{I}}_{\mathsf{exa}} = D^{\mathcal{I}}_{\mathsf{exa}} = \emptyset$ satisfies the first and the last GCIs, but not of the second.

Consider now the interpretations $\mathcal{I}_i := (\{\delta\}, \cdot^{\mathcal{I}_i}, \mathcal{W}_i), 1 \leq i \leq 4$, where

$$A^{\mathcal{I}_1} = \{\delta\} \qquad B^{\mathcal{I}_1} = \{\delta\} \qquad C^{\mathcal{I}_1} = \{\delta\} \qquad D^{\mathcal{I}_1} = \{\delta\}$$
$$A^{\mathcal{I}_2} = \emptyset \qquad B^{\mathcal{I}_2} = \emptyset \qquad C^{\mathcal{I}_2} = \emptyset \qquad D^{\mathcal{I}_2} = \emptyset$$
$$A^{\mathcal{I}_3} = \emptyset \qquad B^{\mathcal{I}_3} = \{\delta\} \qquad C^{\mathcal{I}_3} = \emptyset \qquad D^{\mathcal{I}_3} = \{\delta\}$$
$$A^{\mathcal{I}_4} = \{\delta\} \qquad B^{\mathcal{I}_4} = \{\delta\} \qquad C^{\mathcal{I}_4} = \{\delta\} \qquad D^{\mathcal{I}_4} = \emptyset$$

and $\mathcal{W}_1 = \{x, y, z\}$, $\mathcal{W}_2 = \{x, \neg y, \neg z\}$, $\mathcal{W}_3 = \{\neg x, y, z\}$, $\mathcal{W}_4 = \{\neg x, \neg y, \neg z\}$, depicted in Figure 2. It is easy to verify that the probabilistic interpretation $\mathcal{P}_{\mathsf{exa}} = (\mathfrak{I}, P_{\mathfrak{I}})$ with $\mathfrak{I} = \{\mathcal{I}_1, \dots, \mathcal{I}_4\}$ and $P_{\mathcal{I}_1} = 0.49$, $P_{\mathcal{I}_2} = 0.21$, $P_{\mathcal{I}_3} = 0.15$, $P_{\mathcal{I}_4} = 0.15$ is a model of $\mathcal{T}_{\mathsf{exa}}$ and is consistent with the strategy S that assigns $P(z \mid \neg y) = P(\neg z \mid y) = 1$. Hence $\mathcal{P}_{\mathsf{exa}}$ is a model of $\mathcal{K}_{\mathsf{exa}}$ w.r.t. S.

We emphasise once more that the notion of a model is always dependent on a given strategy chosen by the agent. This is in line with our understanding of IDs. For instance, the strategy of an agent could be such that some contexts become impossible. Then, a model of the knowledge of this agent should disallow to have those contexts with any positive probability. As a consequence, the basic reasoning tasks in ID-$\mathcal{EL}$ must also be parameterised on the chosen strategy. We also note that the requirement for $\mathfrak{I}$ to be finite can be relaxed by imposing some additional constraints in the probability distribution $P_{\mathfrak{I}}$. To avoid unnecessary technicalities, we simply focus on the finite case.

We can now define the cost associated with V-interpretations and probabilistic models.

Definition 8 (Expected cost). *Given an ID $\mathcal{D}$ over V, the cost of the V-interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \mathcal{V}^{\mathcal{I}})$ is $\mathsf{c}(\mathcal{I}) := c(\mathcal{V}^{\mathcal{I}}|_{\pi(c)})$, where $\mathcal{V}^{\mathcal{I}}|_{\pi(c)}$ denotes the restriction of the valuation $\mathcal{V}^{\mathcal{I}}$ to the parents of c.*

Given a strategy S on $\mathcal{D}$ and a probabilistic interpretation $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$ which is consistent with $\mathcal{D}$ w.r.t. S, the expected cost of $\mathcal{P}$ (w.r.t. S) is

$$E[\mathcal{P} \mid S] := \sum_{\mathcal{I} \in \mathfrak{I}} P_{\mathfrak{I}}(\mathcal{I}) \cdot \mathsf{c}(\mathcal{I}).$$

Since the probability distribution in a probabilistic model must be consistent with the distribution induced by the strategy S, the expected cost of any model of a KB $\mathcal{K} = (\mathcal{D}, \mathcal{T})$ w.r.t. S corresponds exactly to the expected cost of $\mathcal{D}$ w.r.t. S. That is, once that the strategy has been chosen, the expected cost does not depend on the specific model of $\mathcal{K}$. Thus, we can define the *expected cost* of $\mathcal{K}$ w.r.t. S as $E[\mathcal{K} \mid S] := E[\mathcal{P} \mid S]$, where $\mathcal{P}$ is any model of $\mathcal{K}$.

Rather than defining a cost function directly on the nodes of the network, it sometimes makes sense to consider this function to be implicitly defined by the properties of the contexts that the node c can observe. In the extreme case, all nodes in V are parents of c and defining the cost function in terms of the contexts obtained by each valuation avoids having to represent the exponentially large mapping. A natural choice for such a cost function is the size of the context. Intuitively, this function allows us to express that a smaller context is preferred over a larger one. Using this function makes sense, for instance, when the context needs to be transferred or manipulated over an unreliable channel. A smaller ontology is preferred to reduce the risk of errors. However, many other functions can be considered depending on the application. As an additional example, if the contexts refer to different levels of granularity of access, then considering the size of the vocabulary as cost is more relevant. We emphasise, however, that ID-$\mathcal{EL}$ does not require the use of any of these cost functions, or even that the node c is influenced by all nodes in V. These are just given as concrete examples with an application-oriented motivation.

4   Reasoning in ID-$\mathcal{EL}$

Before delving in detail into the reasoning tasks for ID-$\mathcal{EL}$, we notice that just as in the special cases of $\mathcal{EL}$ and $\mathcal{BEL}$, every ID-$\mathcal{EL}$ ontology is consistent. More precisely, for every ID-$\mathcal{EL}$ KB and strategy S, there is a model of $\mathcal{K}$ w.r.t. S. Hence, we are more interested in reasoning problems related to subsumptions, their probabilities, and more importantly, their costs.

The first reasoning task that we consider in this setting corresponds to computing bounds on the expected costs associated with the models of a given KB $\mathcal{K}$. That is, we would like to compute an *optimal strategy*, which minimises the expected cost w.r.t. $\mathcal{D}$, and a *pessimal strategy*, which maximises this cost. From the previous discussion, it follows that these bounds correspond exactly to the bounds on the expected cost of the ID $\mathcal{D}$ from $\mathcal{K}$. To speak about complexity, we consider their associated decision problem versions.

Problem 9 (Optimal/Pessimal strategy). Given an ID $\mathcal{D}$ and $b \in \mathbb{R}$, the *optimal strategy problem* (D-Opt) is to decide whether there is a strategy S such that $E[\mathcal{D} \mid S] < b$. Dually, the *pessimal strategy problem* (D-Pes) is to decide whether there is a strategy S such that $E[\mathcal{D} \mid S] > b$.

Both of these problems are PSPACE-complete [12]. It follows that the analogous problems defined for ID-$\mathcal{EL}$ KBs are PSPACE-complete as well.

Theorem 10. *Given an ID-$\mathcal{EL}$ KB $\mathcal{K} = (\mathcal{D}, \mathcal{T})$ and $b \in \mathbb{R}$, deciding whether there exists a strategy S such that $E[\mathcal{K} \mid S] < b$ or $E[\mathcal{K} \mid S] > b$ is* PSPACE-*complete.*

Notice, however, that in general we cannot expect a polynomial-space algorithm to enumerate an optimal strategy. Indeed, even if we limit the search to pure strategies, we should observe that a pure local strategy is merely a Boolean function over the parent variables. It is well known that for every $n \geq 2$ there exist Boolean functions (and hence, local strategies) that cannot be expressed with circuits of size smaller or equal to $2^n/2n$ [18].

One can also consider the problem of entailment of a contextual subsumption, or computing the probability of a subsumption relation to hold. For the latter, as already explained, one must first instantiate the chosen strategy.

Definition 11 (Probabilistic subsumption). *Let $\mathcal{K} = (\mathcal{D}, \mathcal{T})$ be a KB, α a context, and C, D two ID-$\mathcal{EL}$ concepts. Given the probabilistic interpretation $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$, the* probability *of $\langle C \sqsubseteq D : \alpha \rangle$ w.r.t. $\mathcal{P}$ and w.r.t. the strategy S over $\mathcal{D}$ are defined, respectively, as*

$$P(\langle C \sqsubseteq_{\mathcal{P}} D : \alpha \rangle) := \sum_{\mathcal{I} \in \mathfrak{I}, \mathfrak{I} \models \langle C \sqsubseteq D : \alpha \rangle} P_{\mathfrak{I}}(\mathcal{I}), \qquad and$$

$$P(\langle C \sqsubseteq_{\mathcal{K}, S} D : \alpha \rangle) := \inf_{\mathcal{P} \models_S \mathcal{K}} P(\langle C \sqsubseteq_{\mathcal{P}} D : \alpha \rangle).$$

In particular, we denote as $P(C \sqsubseteq_{\mathcal{P}} D)$ the case where $\alpha = \mathsf{true}$ is the universal context satisfied by all propositional valuations.

Recall that an ID together with a strategy forms a BN, and hence after choosing the strategy, the probability of each instantiation of all the variables in V is fully specified. Still, one can choose different models for the KB w.r.t. this strategy. Indeed, note that the universal $\mathcal{EL}$ model which contains only one element belonging to all concepts and connected to itself via all roles, can always be used to build a probabilistic model $\mathcal{P}$ such that $P(\langle C \sqsubseteq_{\mathcal{P}} D : \alpha \rangle) = 1$ for all concepts C, D. Choosing the infimum in the definition of the probability of a subsumption is the natural cautious bound that is guaranteed to hold in *all* models.

In a decision situation, an agent might observe a fact, and try to act upon it with the best available strategy. In IDs, this is modelled through the introduction of *evidence*; formally, the instantiation of one of the chance nodes. In our setting, we are more interested in observing facts that arise from the ontological perspective. Hence, rather than observing the behaviour of the ID, we observe a fact that provides information about the possible contexts that can still hold, hence also influencing the probability distribution over the underlying ID. In practice, when we observe a consequence, we can immediately exclude some cases which contradict our observation. The probabilities of the remaining cases need to be updated accordingly.

Definition 12 (Conditional expected cost). *Let $\mathcal{K} = (\mathcal{D}, \mathcal{T})$ be an ID-$\mathcal{EL}$ KB, S a strategy on $\mathcal{D}$, $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$ a probabilistic model of $\mathcal{D}$ w.r.t. S, and*

C, D two concepts such that $P(C \sqsubseteq_{\mathcal{P}} D) > 0$. The conditional probability *of the interpretation* $\mathcal{I} \in \mathfrak{J}$ *given the subsumption* $C \sqsubseteq D$ *is*

$$P_{\mathfrak{J}}(\mathcal{I} \mid C \sqsubseteq D) := \begin{cases} 0 & \text{if } \mathcal{I} \not\models C \sqsubseteq D \\ \frac{P_{\mathfrak{J}}(\mathcal{I})}{P(C \sqsubseteq_{\mathcal{P}} D)} & \text{otherwise.} \end{cases}$$

The conditional expected cost *of* $\mathcal{P}$ *given* $C \sqsubseteq D$ *w.r.t.* S *is*

$$E[\mathcal{P} \mid S, C \sqsubseteq D] := \sum_{\mathcal{I} \in \mathfrak{J}} P_{\mathfrak{J}}(\mathcal{I} \mid C \sqsubseteq D) \cdot c(\mathcal{I}).$$

As when dealing with probabilities alone, when trying to understand the expected cost given an observation it is important to consider all the possible models of the KB. Accordingly, we can consider an optimistic or a pessimistic approach depending on whether we try to maximise or minimise this expected cost.

Definition 13. *Let $\mathcal{K}$ be an ID-$\mathcal{EL}$ KB, S a strategy, and C, D two concepts. The* optimistic expected cost $\underline{E}$ *and the* pessimistic expected cost $\overline{E}$ *of $\mathcal{K}$ w.r.t. S given $C \sqsubseteq D$ are defined, respectively, by*

$$\underline{E}[\mathcal{K} \mid S, C \sqsubseteq D] := \inf_{\mathcal{P} \models_S \mathcal{K}} E[\mathcal{P} \mid S, C \sqsubseteq D],$$

$$\overline{E}[\mathcal{K} \mid S, C \sqsubseteq D] := \sup_{\mathcal{P} \models_S \mathcal{K}} E[\mathcal{P} \mid S, C \sqsubseteq D].$$

Note that, as mentioned already, for every context it is always possible to construct an $\mathcal{EL}$ model of the context that satisfies also the GCI $C \sqsubseteq D$. In such a model $\mathcal{P} = (\mathfrak{J}, P_{\mathfrak{J}})$, it always holds that $P_{\mathfrak{J}}(\mathcal{I} \mid C \sqsubseteq D) = P_{\mathfrak{J}}(\mathcal{I})$ for all $\mathcal{I} \in \mathfrak{J}$. In particular, this also means that $E[\mathcal{P} \mid S, C \sqsubseteq D] = E[\mathcal{P} \mid S]$ for all strategies S. This yields the following result.

Proposition 14. *For every ID-$\mathcal{EL}$ KB $\mathcal{K}$, strategy S, and concepts C, D, it holds that $\underline{E}[\mathcal{K} \mid S, C \sqsubseteq D] \leq E[\mathcal{K} \mid S] \leq \overline{E}[\mathcal{K} \mid S, C \sqsubseteq D]$.*

Example 15. Consider again the KB $\mathcal{K}_{\mathsf{exa}}$ from Example 7. We have already seen that under the pure strategy $P(z \mid \neg y) = P(\neg z \mid y) = 1$, it holds that $E[\mathcal{K}_{\mathsf{exa}} \mid S] = 11.12$. Consider now the evidence $A \sqsubseteq C$. It is easy to see that there exists a probabilistic model $\mathcal{P}$ such that $E[\mathcal{P} \mid S, A \sqsubseteq C] = 20$. Similarly, if the evidence is $A \sqsubseteq D$, there is a model $\mathcal{P}'$ such that $E[\mathcal{P}' \mid S, A \sqsubseteq D] = 0$. Hence, in general the pessimistic and optimistic expected costs given an evidence do not coincide with the expected cost of the KB. This example also shows that different models may reduce or increase the expected cost, in manners that may not be obvious at first sight.

Theorem 16. *Optimistic and pessimistic expected costs given $C \sqsubseteq D$ can be computed in polynomial space on the size of V.*

Proof. There are exponentially many valuations of the variables in V. For each valuation $\mathcal{W}$, we construct the TBox $\mathcal{T}_{\mathcal{W}}$. Let n be the smallest value in $\mathsf{val}(c)$. We construct a probabilistic model $\mathcal{P} = (\mathfrak{I}, P_{\mathfrak{I}})$ as follows. For each valuation $\mathcal{W}$, $\mathfrak{I}$ contains a V-interpretation $\mathcal{I}_{\mathcal{W}} = (\Delta^{\mathcal{I}_{\mathcal{W}}}, \cdot^{\mathcal{I}_{\mathcal{W}}}, \mathcal{W})$ s.t. (i) $(\Delta^{\mathcal{I}_{\mathcal{W}}}, \cdot^{\mathcal{I}_{\mathcal{W}}}) \models \mathcal{T}_{\mathcal{W}}$, (ii) if $\mathsf{c}(\mathcal{I}) = n$ then $(\Delta^{\mathcal{I}_{\mathcal{W}}}, \cdot^{\mathcal{I}_{\mathcal{W}}}) \models C \sqsubseteq D$, and (iii) if $\mathsf{c}(\mathcal{I}) \neq n$ and $\mathcal{T}_{\mathcal{W}} \not\models C \sqsubseteq D$, then $(\Delta^{\mathcal{I}_{\mathcal{W}}}, \cdot^{\mathcal{I}_{\mathcal{W}}}) \not\models C \sqsubseteq D$. Moreover, $P_{\mathfrak{I}}(\mathcal{I}_{\mathcal{W}}) = P_{\mathcal{D}(S)}(\mathcal{W})$. It is easy to verify that this is a model, constructed in exponential time, and minimises the expected cost. To compute this cost in polynomial space, we store only one interpretation at a time, and accumulate the relative cost of each interpretation iteratively. For the pessimistic expected cost, the proof is analogous. □

However, we are not interested in the expected costs *per se*, but rather as a means to identify the best strategy that the agent can follow under the evidence. In this case, we have the choices to minimise or maximise the optimistic or pessimistic expected costs, which yields four different notions. To reduce the overhead of the definition, we focus only on minimising these costs; notice however that maximisation can be treated analogously.

Definition 17 (Dominant strategies). *Let $\mathcal{K}$ be an ID-$\mathcal{EL}$ KB and C, D two concepts. The strategy S is* dominant optimistic *if for every strategy S' it holds that $\underline{E}[\mathcal{K} \mid S, C \sqsubseteq D] \leq \underline{E}[\mathcal{K} \mid S', C \sqsubseteq D]$. It is* dominant pessimistic *if for all strategies S', $\overline{E}[\mathcal{K} \mid S, C \sqsubseteq D] \leq \overline{E}[\mathcal{K} \mid S', C \sqsubseteq D]$.*

A naïve approach for finding *pure* dominant strategies is to enumerate all possible options, and preserve those that yield the lowest expected costs. In the worst case, there are doubly-exponentially many such strategies on the size of V, which makes this naïve approach infeasible, despite its effectiveness. On the other hand, it is easy to see that the optimal strategy for the whole network is a special case of Definition 17, where $C \sqsubseteq D$ corresponds to any $\mathcal{EL}$ tautology (e.g., $A \sqsubseteq A$).

Consider the decision problems (D-Dom-Opt and D-Dom-Pes, respectively) associated with Definition 17: given a KB $\mathcal{K}$, two concepts C, D and $b \in \mathbb{R}$, decide whether there are strategies S, S' such that $\underline{E}[\mathcal{K} \mid S, C \sqsubseteq D] < b$, and $\overline{E}[\mathcal{K} \mid S', C \sqsubseteq D] < b$, respectively. Using an approach similar to Theorem 16, we can build a polynomial-space algorithm for deciding D-Dom-Opt under pure strategies by enumerating all valuations of the chance nodes, guessing for each of them a valuation of the decision variables and computing the minimal cost that arises from each of them. The only issue is that this needs to be done in a specific order to guarantee that for equal parent nodes, the same guess is made always in a decision variable.

Theorem 18. *The problems D-Dom-Opt and D-Dom-Pes are* PSPACE-*complete for pure strategies.*

Obviously, the lower bound holds also for arbitrary strategies. The upper bound can be extended to non-pure strategies, as long as they are representable in exponential space; otherwise, we would not be able to guess them in exponential time.

5 Conclusions

We have introduced ID-$\mathcal{EL}$, a new extension of the DL $\mathcal{EL}$ capable of modeling and dealing with decision situations under uncertainty. This is achieved by integrating an influence diagram to represent the uncertainty, potential decisions, and the overall costs of a choice. The ontological ($\mathcal{EL}$) portion and the influence representation are combined through contexts. From an abstract point of view, we build a collection of ontologies, which hold only in specific contexts. These ontologies contain only certain knowledge, but the specific context under consideration is uncertain.

Extending the basic idea of $\mathcal{BEL}$, our framework allows for an agent to influence its context by making choices in specific nodes of the network. The agent is motivated to make choices that minimise the overall expected cost. Intuitively, this means minimising the probability of large costs, and maximising the probability of low costs.

We studied the basic reasoning problems in this logic, and gave tight complexity bounds for all of them. Interestingly, the decision problem associated with finding a dominating optimal strategy, in which the agent should find the best strategy conditioned on an ontological observation, remains PSPACE-complete. A practical algorithm for solving this problem is left for future work. As future work we will also consider other decision-based reasoning tasks, and complexity classes. Notably, we will study whether optimal strategies or costs can be approximated efficiently. Moreover, we will consider the task of building strategies iteratively, as a response to the environment; this is justified by the no-forgetting assumption of IDs. Another interesting issue to resolve is how to dislodge the strategies from the underlying ID, and allow the agent to select consequences (rather than direct contexts) instead.

To conclude, we note that the choice of $\mathcal{EL}$ as a logical formalism is motivated by its polynomial-time reasoning problems, which allow us to understand complexity issues better. However, our framework can be combined with other (potentially more expressive) logics. Building those extensions introduces further problems (e.g., consistency) that would need to be studied in detail as well.

References

1. Acar, E., Fink, M., Meilicke, C., Thorne, C., Stuckenschmidt, H.: Multi-attribute decision making with weighted description logics. IfCoLog Journal of Logics and its Applications **4**(7), 1973–1996 (2017)
2. Acar, E., Thorne, C., Stuckenschmidt, H.: Towards decision making via expressive probabilistic ontologies. In: International Conference on Algorithmic Decision Theory. pp. 52–68. Springer (2015)
3. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Kaelbling, L., Saffiotti, A. (eds.) Proc. 19th Int. Joint Conf. on Artificial Intelligence (IJCAI'05). pp. 364–369. Professional Book Center (2005)
4. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, second edn. (2007)

5. Baader, F., Knechtel, M., Peñaloza, R.: Context-dependent views to axioms and consequences of semantic web ontologies. Journal of Web Semantics **12**, 22–40 (2012). https://doi.org/10.1016/j.websem.2011.11.006

6. Botha, L., Meyer, T., Peñaloza, R.: A Bayesian extension of the description logic ALC. In: Proceedings of the 16th European Conference on Logics in Artificial Intelligence (JELIA'19). Lecture Notes in Computer Science, vol. 11468, pp. 339–354. Springer (2019)

7. Ceylan, İ.İ., Peñaloza, R.: The Bayesian description logic BEL. In: Demri, S., Kapur, D., Weidenbach, C. (eds.) Proceedings of the 7th International Joint Conference on Automated Reasoning (IJCAR'14). Lecture Notes in Computer Science, vol. 8562, pp. 480–494. Springer (2014)

8. Ceylan, İ.İ., Peñaloza, R.: Bayesian description logics. In: Proceedings of the 27th International Workshop on Description Logics (DL' 14). CEUR Workshop Proceedings, vol. 1193, pp. 447–458. CEUR-WS.org (2014), http://ceur-ws.org/Vol-1193/paper_5.pdf

9. Ceylan, İ.İ., Peñaloza, R.: The Bayesian ontology language BEL. Journal of Automated Reasoning **58**(1), 67–95 (2017). https://doi.org/10.1007/s10817-016-9386-0

10. d'Amato, C., Fanizzi, N., Lukasiewicz, T.: Tractable reasoning with bayesian description logics. In: Greco, S., Lukasiewicz, T. (eds.) Proceedings of the Second International Conference on Scalable Uncertainty Management (SUM 2008). Lecture Notes in Computer Science, vol. 5291, pp. 146–159. Springer (2008). https://doi.org/10.1007/978-3-540-87993-0_13

11. Gutiérrez-Basulto, V., Jung, J.C., Lutz, C., Schröder, L.: Probabilistic description logics for subjective uncertainty. Journal of Artificial Intelligence Research **58**, 1–66 (2017). https://doi.org/10.1613/jair.5222

12. Littman, M.L., Majercik, S.M., Pitassi, T.: Stochastic Boolean satisfiability. Journal of Automated Reasoning **27**(3), 251–296 (2001)

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

14. Niepert, M., Noessner, J., Stuckenschmidt, H.: Log-linear description logics. In: Twenty-Second International Joint Conference on Artificial Intelligence (2011)

15. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1988)

16. Riguzzi, F., Bellodi, E., Lamma, E., Zese, R.: Probabilistic description logics under the distribution semantics. Semantic Web **6**(5), 477–501 (2015)

17. Shachter, R.D.: Evaluating influence diagrams. Operations Research **34**(6), 871–882 (1986). https://doi.org/10.1287/opre.34.6.871

18. Shannon, C.E.: The synthesis of two-terminal switching circuits. Bell System Technical Journal **28**(1), 59–98 (1949). https://doi.org/10.1002/j.1538-7305.1949.tb03624.x