The Thirty-Third International
FLAIRS Conference (FLAIRS-33)

Context-Based Inferences from Probabilistic Conditionals with Default Negation at Maximum Entropy

Marco Wilhelm, Gabriele Kern-Isberner
Department of Computer Science, TU Dortmund, Dortmund, Germany

Abstract

The principle of maximum entropy (MaxEnt) constitutes a powerful formalism for nonmonotonic reasoning based on probabilistic conditionals. Conditionals are defeasible rules which allow one to express that certain subclasses of some broader concept behave exceptional. In the (common) probabilistic semantics of conditional statements, these exceptions are formalized only implicitly: The conditional $(B|A)[p]$ expresses that if A holds, then B is typically true, namely with probability p, but without explicitly talking about the subclass of A for which B does not hold. There is no possibility to express within the conditional that a subclass C of A is excluded from the inference to B because one is unaware of the probability of B given C. In this paper, we apply the concept of default negation to probabilistic MaxEnt reasoning in order to formalize this kind of unawareness and propose a context-based inference formalism. We exemplify the usefulness of this inference relation, and show that it satisfies basic formal properties of probabilistic reasoning.

Introduction

Probabilistic reasoning (Halpern 2003; Pearl 1988) combines probability theory with classical logic in order to handle uncertainty in knowledge. It is often based on rules of the form "if A holds, then B follows with probability p" which are called *conditionals* and are formally written as $(B|A)[p]$. Conditionals are interpreted by probability distributions that assign each possible state of the real world a degree of belief. Based on this methodology, it is possible to express and reason about subclasses of individuals which behave differently to some broader concept, like penguins that form an exceptional subclass of non-flying birds. In order to model these exceptions by means of probability distributions, it is however necessary to state that they behave contrarily regarding the considered property. Therefore, this approach lacks the opportunity to treat a subclass as exceptional because one is unsure about its compliance with the property.

In this paper, we propose a novel knowledge representation language $\mathcal{PCL}^{\mathsf{not}}$ in which it is possible to exclude certain subclasses from a conditional statement. Formally, this is done by enriching conditionals with *default negated* propositions within their premises that act as disqualifiers (cf. (Clark 1977; Gelfond and Lifschitz 2000) for an introduction to the concept of default negation, also known as negation as failure). Such a conditional is blocked when a disqualifier D holds in a concrete context and remains unconsidered when drawing inferences. We further apply the *principle of maximum entropy* (Paris 2006) to our approach in order to select a meaningful probabilistic model which leads to a novel inference relation $\mid\!\sim_{\mathsf{ME}}$ that is convenient to handle unawareness.

$\mathcal{PCL}^{\mathsf{not}}$ is a natural extension of qualitative conditionals with default negation (Wilhelm et al. 2017) to the probabilistic setting. Although conditionals, default negation, and probabilities each are well-known concepts in nonmonotonic reasoning, to our knowledge, there is no work published yet which combines all three of them. We illustrate the usefulness of probabilistic conditionals with default negation for context-based reasoning with many examples that prove their flexibility when dealing with implicit and explicit exceptions at the same time. We further prove that drawing inferences from such conditionals by using the inference relation $\mid\!\sim_{\mathsf{ME}}$ satisfies a number of desired inference properties of probabilistic reasoning.

The rest of the paper is organized as follows. First, we define the logical foundations and recall some basics about conditionals, negation as failure, as well as probabilities with the focus on how these concepts handle exceptions. Afterwards, we introduce the probabilistic conditional language with default negation $\mathcal{PCL}^{\mathsf{not}}(\Sigma)$ and define the nonmonotonic inference relation $\mid\!\sim_{\mathsf{ME}}$ between knowledge bases in $\mathcal{PCL}^{\mathsf{not}}(\Sigma)$ and query conditionals. We discuss the properties of $\mid\!\sim_{\mathsf{ME}}$ by means of examples and by a formal analysis and conclude with an outlook.

Preliminaries

Let $\mathcal{L}(\Sigma)$ be a *propositional language* defined over a finite set of *atoms* Σ. *Propositions* in $\mathcal{L}(\Sigma)$ are defined by using the connectives $\neg A$ (negation), $A \wedge B$ (conjunction), and $A \vee B$ (disjunction) where A and B are propositions and are interpreted by mappings $\mathcal{I} : \mathcal{L}(\Sigma) \rightarrow \{0, 1\}$ as usual. The set of all such *interpretations* is denoted by $\mathfrak{I}(\Sigma)$. An interpretation $\mathcal{I}$ is a *model* of a proposition A iff $\mathcal{I}(A) = 1$. A proposition A *entails* another proposition B, written $A \models B$, iff every model of A is a model of B. The entailment relation is extended to sets $\mathcal{A}, \mathcal{B}$ of propositions: $\mathcal{A} \models \mathcal{B}$ iff $A \models B$ for all $A \in \mathcal{A}, B \in \mathcal{B}$. If A and B have the same models, they are *logically equivalent*, $A \equiv B$ in symbol. In order to shorten mathematical expressions, we abbreviate $\neg A$ with $\overline{A}$, $A \wedge B$ with AB, $\overline{A} \vee B$ with $A \to B$, and $A \vee \overline{A}$ with $\top$.

As propositions are incontrovertibly true or false, they are well-suited for representing *factual* knowledge. When *uncertainty* comes into play, however, they fail. Therefore, many efforts were and are made to extend propositional logic with a view to handle uncertainty of any kind. Major concepts in this research direction are *conditionals* (Adams 1965), *negation as failure* (Clark 1977; Gelfond and Lifschitz 2000), and *probabilities* (Halpern 2003; Pearl 1988) which we will combine within a unified language in this paper.

Conditionals. A *conditional* $(B|A)$ with $A, B \in \mathcal{L}(\Sigma)$ is a formal representation of the statement "If A holds, then typically B follows." In other words, in the presence of A the proposition B is more plausible than its negation $\overline{B}$. For a formal semantics of conditionals, a (partial) ordering on the set of interpretations is needed so that it is possible to compare interpretations, and hence propositions, with respect to their plausibility. In this context, interpretations are also called *possible worlds*, each describing a possible and more or less plausible way the real world might be. In this paper, possible worlds are represented as complete conjunctions of *literals*, i.e. atoms or negated atoms. The set of all possible worlds is denoted by $\Omega(\Sigma)$.

Conditionals provide the opportunity to formalize exceptions *implicitly*: "If A holds, then B *typically* follows" implies that there might be an exceptional case in which B does *not* follow from A. For example, birds are typically able to fly, formalizable as $(\mathsf{flies}|\mathsf{bird})$, unless they have a broken wing. The reason for the exception, here the broken wing, however, cannot be formalized within the same conditional but requires an additional one. The example could be extended by the conditional $(\neg\mathsf{flies}|\mathsf{bird} \wedge \mathsf{brokenWing})$, for instance.

With conditionals, those exceptions can be formalized that have a property which is in conflict with the superclass (here, $\neg\mathsf{flies}$ against flies). However, sometimes it is desired to express unawareness about the properties of a certain subclass. For instance, this could be the case if one discovers a new species of birds whose ability to fly is unknown. In this case, one might want to waive one's knowledge about the flight behavior of birds in order to be unbiased. Hence, there is the need to *disregard* a conditional in some cases which is not possible when taking account of the common semantics of conditionals.

Negation as Failure. In contrast to conditionals, the concept of *negation as failure*, also known as *default negation*, can be used to formalize exceptions *explicitly*. In logic programming, the rule $(B \leftarrow A, \mathit{not}\ C)$ with $A, B, C \in \mathcal{L}(\Sigma)$ states that "If A holds and C cannot be proven, then B follows." Whether such a rule applies or not is tested against a set of interpretations, e.g. the models of a set of factual knowledge. Given such a set of interpretations $\mathfrak{I}' \subseteq \mathfrak{I}(\Sigma)$, B can be inferred from $(A, \mathit{not}\ C)$ iff A is true in every interpretation in $\mathfrak{I}'$ and there is at least one interpretation in $\mathfrak{I}'$ in which C is not true, i.e., a proof of C fails. Note that this concept of negation differs from classical, also called *strong* negation: $(B \leftarrow A, \neg C)$ applies only if C is false in *every* interpretation in $\mathfrak{I}'$. Hence, $\mathit{not}\ C$ does not imply $\neg C$ but the other way around $\neg C$ implies $\mathit{not}\ C$.

To illustrate the usage of default negation, we take up our example. The rule $(\mathsf{flies} \leftarrow \mathsf{bird}, \mathit{not}\ \mathsf{newSpecies})$ expresses that birds are able to fly unless they are of a new species. In the latter case, the rule is blocked and does not state anything. The drawback of default negated rules is that all exceptions have to be noted down explicitly. This makes programs that consist of rules of the above kind inflexible against changes. For example, the fact that penguins form another exceptional subclass of birds with respect to their flight behavior cannot be formalized in another rule but the existing rule has to be adjusted.

There have already been efforts to combine conditionals and default negation in order to benefit from both ways of handling exceptions. A semantics for conditionals with default negation can be found in (Wilhelm et al. 2017).

Probabilities. While conditionals and default negation are qualitative concepts, probabilities add quantitative uncertainty to propositions. A *probabilistic proposition* $A[p]$ with $A \in \mathcal{L}(\Sigma)$ and $p \in [0,1]$ states that "A holds with probability p" and is interpreted by probability distributions over possible worlds. Semantically, such a probability distribution $\mathcal{P} : \Omega(\Sigma) \to [0,1]$ represents a reasoner's *belief state* and is a *model* of $A[p]$ iff $\mathcal{P}(A) = p$ where $\mathcal{P}(A) = \sum_{\omega \models A} \mathcal{P}(\omega)$. Probabilistic propositions are implicit formalizations of exceptions: If $A[p]$ with $p < 1$ holds, then there is at least one possible world with positive probability in which $\overline{A}$ is true. Hence, the reasoner with belief $A[p]$, $p < 1$, does not completely deny $\overline{A}$ either. The probabilistic semantics of propositions can be easily extended to conditionals: A probability distribution $\mathcal{P} : \Omega(\Sigma) \to [0,1]$ is a *model* of a *probabilistic conditional* $(B|A)[p]$ with $A, B \in \mathcal{L}(\Sigma)$ and $p \in [0,1]$ iff $\mathcal{P}(A) > 0$ and $\mathcal{P}(B|A) = p$. It holds that $A[p] = (A|\top)[p]$.

There have also been some attempts on combining *answer set programming*, which builds on rules with default negation, and probabilities (Baral, Gelfond, and Rushton 2009; Cozman 2019). However, in these approaches the probabilities are not applied to rules but to atoms. Hence, the probabilities are not used to quantify rules with exceptions but to include random events.

Principle of Maximum Entropy. Due to the vast number of probability distributions over $\Omega(\Sigma)$, reasoning over all models of a set of probabilistic conditionals is often very uninformative. For example, let $a, b, c \in \Sigma$, and consider the set of conditionals $\mathcal{R}_{\mathsf{ex}} = \{(c|a)[0.7], (c|b)[0.9]\}$, i.e., a and b are evidence for c. Then, nothing can be said about the likelihood of c in the presence of a *and* b. More precisely, for every $p \in [0,1]$ there is a model of $\mathcal{R}_{\mathsf{ex}}$ in which

$(c|ab)[p]$ is true as well. At the same time, it is reasonable to assume that the probability p of $(c|ab)[p]$ is at least 0.7 (unless a and b weaken each other their strength of evidence for c, which is possible but certainly not most obvious). Hence, for reasoning tasks, it is useful to select a single model of $\mathcal{R}_{\mathsf{ex}}$. Of course, this model should reflect a reasoner's *belief state* appropriately when the set of conditionals constitutes the reasoner's beliefs. From a commonsense point of view, the *maximum entropy distribution* (MaxEnt distribution) fits best to the model selection task (Paris 2006). It is defined by

$$\mathsf{ME}(\mathcal{R}) = \arg\max_{\mathcal{P} \models \mathcal{R}} - \sum_{\omega \in \Omega(\Sigma)} \mathcal{P}(\omega) \cdot \log \mathcal{P}(\omega),$$

where the convention $0 \cdot \log 0 = 0$ applies, and is the unique model of $\mathcal{R}$ which adds as few information as possible. In order to compute $\mathsf{ME}(\mathcal{R})$, one has to solve a nonlinear optimization problem. We recommend (Boyd and Vandenberghe 2014) for the theoretical background of MaxEnt calculations and the software tool SPIRIT (Rödder and Meyer 1996) for practical applications. Note that $\mathsf{ME}(\mathcal{R})$ always exists if $\mathcal{R}$ is consistent, i.e., there is a probability distribution that models all conditionals in $\mathcal{R}$.

Probabilistic Conditionals with Default Negation

We now define the *probabilistic conditional language with default negation* $\mathcal{PCL}^{\mathsf{not}}(\Sigma)$ which combines the benefits of conditionals, negation as failure as well as probabilities. For this, let $A, B \in \mathcal{L}(\Sigma)$ be propositions, let $\mathcal{D} \subseteq \mathcal{L}(\Sigma)$ be a set of propositions, and let $p \in [0,1]$. A *default negated probabilistic conditional* is an expression of the form $(B|A, \mathit{not}\ \mathcal{D})[p]$ with the meaning "If A holds, then B follows with probability p unless any proposition $D \in \mathcal{D}$ is provably true." In the rest of the paper, we will call these expressions *conditionals* for short. Elements in $\mathcal{D}$ are called *disqualifiers* as their validity blocks the whole conditional. The language $\mathcal{PCL}^{\mathsf{not}}(\Sigma)$ consists of all conditionals that can be built over Σ.

A conditional $(B|A, \mathit{not}\ \mathcal{D})[p]$ without default negated part, i.e. with $\mathcal{D} = \emptyset$, is a classical probabilistic conditional and simply written as $(B|A)[p]$. If, in addition, $p = 1$, then the conditional becomes *factual* as it forces every possible world in which $A\overline{B}$ is true to have zero probability (conditionals of the form $(B|A)[0]$ are equivalent to $(\overline{B}|A)[1]$ and, thus, also factual but redundant). In order to have a clear separation between factual and uncertain knowledge, we write factual conditionals $(B|A)[1]$ in the form of material implications, $A \rightarrow B$, or more general in form of propositions $P \equiv A \rightarrow B$. Hence, a set of propositions without annotated probabilities $\mathcal{F}$ defines a subsets of possible worlds

$$\Omega_{\mathcal{F}}(\Sigma) = \{\omega \in \Omega(\Sigma) \mid \forall F \in \mathcal{F} : \omega \models F\}$$

so that all remaining possible worlds in $\Omega(\Sigma) \setminus \Omega_{\mathcal{F}}(\Sigma)$ have zero probability.

A *knowledge base* is a tuple $\mathcal{R} = (\mathcal{F}_{\mathcal{R}}, \mathcal{B}_{\mathcal{R}})$ consisting of a finite set of propositions $\mathcal{F}_{\mathcal{R}}$ which represent factual knowledge and a finite set of conditionals $\mathcal{B}_{\mathcal{R}}$ with a probability $p \in (0,1)$ expressing the reasoner's beliefs.

Example 1. *Consider a reasoner who formalizes her knowledge about birds. She knows that antarctic birds are birds and that birds are typically able to fly. However, she is unsure about the flight capacity of antarctic birds because she knows that there are living different species in the antarctic in contrast to the rest of the world. Therefore, she wants to exclude antarctic birds from her beliefs about the flight capacity of birds. Her formalized knowledge could look like*

$$\mathcal{R}_{\mathsf{brd}} = (\{A \rightarrow B\},\ \{(F|B, \mathit{not}\ \{A\})[0.95]\})$$

with the abbreviations A = antarctic bird, B = bird, and F = able to fly.

The formal semantics of probabilistic conditionals with default negation, and more general of knowledge bases, is based on probability distributions that are defined for *reducts* of conditionals. The idea is to evaluate conditionals in the light of a concrete context which is formalized by a proposition C. The reduct represents the reaonser's beliefs that hold in this context. If any disqualifier $D \in \mathcal{D}$ of a conditional holds in the context C, i.e $C \models D$, the conditional is blocked and therefore ignored when determining a probability distribution as a model. If no disqualifier holds in the context, then the conditional reduced by its default negated part is considered. When considering a whole knowledge base, the context is unified with the factual knowledge for the evaluation of the default negated parts of the conditionals.

Definition 1. *Let $\mathcal{R} = (\mathcal{F}_{\mathcal{R}}, \mathcal{B}_{\mathcal{R}})$ be a knowledge base, and let C be a proposition. The* reduct *$\mathcal{R}^C = (\mathcal{F}_{\mathcal{R}}, \mathcal{B}^C_{\mathcal{R}})$ is the knowledge base that consists of the same set of facts $\mathcal{F}_{\mathcal{R}}$ as in $\mathcal{R}$ and the set of beliefs*

$$\mathcal{B}^C_{\mathcal{R}} = \{(B|A) \mid (B|A, \mathit{not}\ \mathcal{D}) \in \mathcal{R} \ \ \text{and} \\ \forall D \in \mathcal{D} : \mathcal{F}_{\mathcal{R}} \cup \{C\} \not\models D\}.$$

Hence, reducts are knowledge bases that are free of default negation and can be interpreted in the usual way.

Definition 2. *Let $\mathcal{R} = (\mathcal{F}_{\mathcal{R}}, \mathcal{B}_{\mathcal{R}})$ be a knowledge base, and let $C \in \mathcal{L}(\Sigma)$. A probability distribution $\mathcal{P} : \Omega(\Sigma) \rightarrow [0,1]$ is a* model *of $\mathcal{R}$ in the context C iff $\mathcal{P}(\omega) = 0$ for $\omega \in \Omega(\Sigma) \setminus \Omega_{\mathcal{F}}(\Sigma)$ and $\mathcal{P}(B|A) = p$ for $(B|A)[p] \in \mathcal{B}^C_{\mathcal{R}}$.*

We illustrate this concept of context-dependent models based on Example 1.

Example 2. *The knowledge base $\mathcal{R}_{\mathsf{brd}}$ from Example 1 has two different reducts: $\mathcal{R}^C_{\mathsf{brd}} = (\{A \rightarrow B\}, \emptyset)$ for any context C with $C \models A$ and $\mathcal{R}^{C'}_{\mathsf{brd}} = (\{A \rightarrow B\},\ \{(F|B)[0.95]\})$ for any context C' with $C' \not\models A$. Each probability distribution $\mathcal{P}$ on $\Omega(\Sigma)$ which satisfies $\mathcal{P}(A\overline{B}) = 0$ is a model of $\mathcal{R}^C_{\mathsf{brd}}$. Models $\mathcal{P}'$ of $\mathcal{R}^{C'}_{\mathsf{brd}}$ have to satisfy $\mathcal{P}'(F|B) = 0.95$ in addition. For example, in the context B, i.e., the individual under consideration is a bird, each model states that the individual is able to fly with probability 0.95. If the context is A, i.e., the individual is an antarctic bird, there is no constraint on the probability with which the individual is able to fly, instead.*

In the next section, we discuss how inferences can be drawn from knowledge bases in $\mathcal{PCL}^{\mathsf{not}}$.

Context-Based Inferences at Maximum Entropy

Let $\mathcal{R}$ be a knowledge base. Once a context C as well as a model $\mathcal{P}$ of $\mathcal{R}^C$ is fixed, one obtains the inference relation

$$\mathcal{R}, C \mid\!\sim_{\mathcal{P}} (B|A, \mathit{not}\ \mathcal{D})[p] \quad \text{iff} \quad \mathcal{P}(B|A) = p \\ \vee\ \exists D \in \mathcal{D} : \mathcal{F}_{\mathcal{R}} \cup \{C\} \models D.$$

Hence, a conditional is inferred from $\mathcal{R}$ if it is blocked by a disqualifier or the conditional reduced by its default negated part holds. In order to draw meaningful inferences from $\mathcal{R}$, it is necessary to select a reasonable context and model. For the context, we suggest a query-dependent choice: If one asks for the probability p of the conditional $(B|A)[p]$, one is interested in the likelihood of observing B in the presence of A. That is, one assumes that A is true which should conform with the context. The model, instead, should reflect the inference behavior of the reasoner. In general, it varies from reasoner to reasoner and could also vary from context to context. However, if nothing is known about the inference behavior of a particular reasoner or one is interested in a commonsense point of view, the MaxEnt distribution is a good choice for the model as mentioned in the preliminaries. This leads to the nonmonotonic inference relation

$$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B|A, \mathit{not}\ \mathcal{D})[p] \quad \text{iff} \quad \mathsf{ME}(\mathcal{R}^A)(B|A) = p \\ \vee\ \exists D \in \mathcal{D} : \mathcal{F}_{\mathcal{R}} \cup \{A\} \models D.$$

As this inference relation particularly holds for probabilities $p \in \{0, 1\}$, it follows that $\mathcal{R} \mid\!\sim_{\mathsf{ME}} A$ iff $\mathsf{ME}(\mathcal{R}^\top)(A) = 1$ for propositions A, i.e., for factual knowledge. Consequently, if $\mathcal{F}_{\mathcal{R}} \models A$ classical logically, then $\mathcal{R} \mid\!\sim_{\mathsf{ME}} A$. One can show that $\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B|A, \mathit{not}\ \mathcal{D})[p]$ holds, too, if $(B|A, \mathit{not}\ \mathcal{D})[p] \in \mathcal{B}_{\mathcal{R}}$, i.e. $\mid\!\sim_{\mathsf{ME}}$ satisfies *Direct Inference* (see the proof of Proposition 1).

Before we analyze the inference relation $\mid\!\sim_{\mathsf{ME}}$ in detail, it remains to discuss for which knowledge bases it is well-defined, i.e., for which knowledge bases $\mathcal{R}$ the MaxEnt distributions $\mathsf{ME}(\mathcal{R}^C)$ for all reducts $\mathcal{R}^C$ of $\mathcal{R}$ exist. Fortunately, this is easy to answer. Obviously, it holds that $\mathcal{B}_{\mathcal{R}}^C$ is a subset of $\mathcal{B}_{\mathcal{R}}^\top$ for any context C since classical entailment is monotonous ($\forall D \in \mathcal{D} : \mathcal{F}_{\mathcal{R}} \cup \{C\} \not\models D$ implies $\forall D \in \mathcal{D} : \mathcal{F}_{\mathcal{R}} \cup \{\top\} \not\models D$). Hence, if the MaxEnt distribution for $\mathcal{R}^\top$ exists, then it also exists for all other reducts of $\mathcal{R}$. Since $\mathsf{ME}(\mathcal{R}^\top)$ exists iff $\mathcal{R}^\top$ is consistent, one only has to check whether $\mathcal{R}^\top$ is consistent or not. Consequently, we may call $\mathcal{R}$ *consistent* iff $\mathcal{R}^\top$ is consistent.

If the knowledge base $\mathcal{R}$ does not mention default negations at all, for any context C, the reduct $\mathcal{R}^C$ equals the original knowledge base $\mathcal{R}$ and the inference relation $\mid\!\sim_{\mathsf{ME}}$ is the standard MaxEnt inference relation for probabilistic conditional knowledge (cf., e.g., (Kern-Isberner 2001)). In the following we investigate the properties of the inference relation $\mid\!\sim_{\mathsf{ME}}$ in the case where default negations are present, instead, both by means of examples and based on formal inference properties.

Example 3. *We recall that the knowledge base $\mathcal{R}_{\mathsf{brd}}$ from Example 1 has the two reducts $\mathcal{R}_{\mathsf{brd}}^C = (\{A \to B\}, \emptyset)$ for C with $C \models A$ and $\mathcal{R}_{\mathsf{brd}}^{C'} = (\{A \to B\}, \{(F|B)[0.95]\})$ for C' with $C' \not\models A$. If one is interested in the likelihood that an arbitrary bird is able to fly, one would assume that the probability is 0.95. And, indeed, $\mathcal{R}_{\mathsf{brd}} \mid\!\sim_{\mathsf{ME}} (F|B)[p]$ iff $p = 0.95$ as one has to infer $(F|B)[p]$ from the reduct $\mathcal{R}_{\mathsf{brd}}^{C'}$ because $\{A \to B, B\} \not\models A$ and $\mathcal{R}_{\mathsf{brd}}^{C'}$ explicitly states that birds are able to fly with probability 0.95.*

On the contrary, the likelihood that an antarctic bird is able to fly should most probably be 0.5, as the reasoner with knowledge base $\mathcal{R}_{\mathsf{brd}}$ has excluded antarctic birds from the conditional that states that birds are typically able to fly. And again, the supposed inference $\mathcal{R}_{\mathsf{brd}} \mid\!\sim_{\mathsf{ME}} (F|A)[0.5]$ holds: $\{A \to B, A\} \models A$ and, thus, the relevant reduct is $\mathcal{R}_{\mathsf{brd}}^C$ this time. Further, $\mathsf{ME}(\mathcal{R}_{\mathsf{brd}}^C)(F|A) = 0.5$ holds, as the MaxEnt distribution for a reduct with empty set of beliefs is the uniform distribution on $\Omega_{\mathcal{F}_{\mathcal{R}_{\mathsf{brd}}}}(\Sigma)$.

$\mathcal{R}_i$	$\mathsf{ME}(\mathcal{R}_i)(F\|B)$	$\mathsf{ME}(\mathcal{R}_i)(F\|A)$
$\mathcal{R}_{\mathsf{brd}}$	0.95	0.5
$\mathcal{R}_1$	0.95	0.95
$\mathcal{R}_2$	0.670	0.5
$\mathcal{R}_3$	0.95	0.5
$\mathcal{R}_{3,b}$	0.95	0.5
$\mathcal{R}_{\mathsf{brd},b}$	0.95	0.366

Table 1: Comparison of the example knowledge bases with respect to the inferences about the flight behavior of (antarctic) birds that can be drawn from the knowledge bases.

As a preliminary discussion of the differences between explicit and implicit formalizations of exceptions, we now investigate whether the same inferences as from $\mathcal{R}_{\mathsf{brd}}$ can be drawn from knowledge bases without default negation. See Table 1 for a brief overview.

Example 4. *We consider the knowledge base*

$$\mathcal{R}_1 = (\{A \to B\},\ \{(F|B)[0.95]\})$$

which is the same knowledge base as $\mathcal{R}_{\mathsf{brd}}$ aside from the fact that the reasoner has ignored her undecidedness concerning her valuation of the flight capability of antarctic birds. Of course, one still infers $(F|B)[0.95]$ from $\mathcal{R}_1$ with respect to any model of $\mathcal{R}_1$ and, hence, also with respect to the MaxEnt distribution. However, in contrast to $\mathcal{R}_{\mathsf{brd}}$, antarctic birds inherit the flight capability from birds, and $\mathsf{ME}(\mathcal{R}_1)(F|A) = 0.95$ holds, too, while for $\mathcal{R}_{\mathsf{brd}}$ the respective probability is 0.5.

Example 5. *Another way of dealing with the fact that one does not believe that antarctic birds fly with the same probability than birds is to limit the scope of the conditional about the flight behavior in $\mathcal{R}_{\mathsf{brd}}$ to birds that are not from the antarctic at all. One obtains*

$$\mathcal{R}_2 = (\{A \to B\},\ \{(F|B\overline{A})[0.95]\}).$$

Since antarctic birds are excluded from the conditional, one obtains $\mathsf{ME}(\mathcal{R}_2)(F|B) = 0.5$ as for $\mathcal{R}_{\mathsf{brd}}$. However, the likelihood that an arbitrary bird is able to fly decreases from the intended probability 0.95 to $\mathsf{ME}(\mathcal{R}_2)(F|B) \approx 0.670$.

Name of Property	Property			
Reflexivity				$\mathcal{R} \mid\!\sim_{\mathsf{ME}} A \to A$
Left Logical Equivalence	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|A, \mathit{not}\ \mathcal{D})[p]$	$A \equiv C$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|C, \mathit{not}\ \mathcal{D})[p]$
Cut	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|AC, \mathit{not}\ \mathcal{D})[p]$,	$A \models C$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|A, \mathit{not}\ \mathcal{D})[p]$
Cautious Monotonicity	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|A, \mathit{not}\ \mathcal{D})[p]$,	$A \models C$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|AC, \mathit{not}\ \mathcal{D})[p]$
Right Weakening	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|A, \mathit{not}\ \mathcal{D})[p]$,	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} B \to C$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (C\|A, \mathit{not}\ \mathcal{D})[q]$ with $q \geq p$
Or	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} A \to B$,	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} C \to B$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} A \vee C \to B$
Inclusion / Direct Inference	$(B\|A, \mathit{not}\ \mathcal{D})[p] \in \mathcal{B}_{\mathcal{R}}$		$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|A, \mathit{not}\ \mathcal{D})[p]$
Conditioning	$(B\|A, \mathit{not}\ \mathcal{D})[p] \in \mathcal{B}_{\mathcal{R}}$,	$A \equiv C$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|C, \mathit{not}\ \mathcal{D})[p]$
Irrelevance	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|A, \mathit{not}\ \mathcal{D})[p]$, no atom from $\mathcal{R}$ and $(B\|A, \mathit{not}\ \mathcal{D})[p]$ is in C		$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} (B\|AC, \mathit{not}\ \mathcal{D})[p]$
Rational Monotonicity	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} A \to B$,	$\mathcal{R} \not\mid\!\sim_{\mathsf{ME}} A \to \overline{C}$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} AC \to B$
Inheritance of Logical Knowledge	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} A \to B$,	$C \models A$	$\vdash$	$\mathcal{R} \mid\!\sim_{\mathsf{ME}} C \to B$

Table 2: Inference properties that are satisfied by $\mid\!\sim_{\mathsf{ME}}$. $\mathcal{R}$ is assumed to be a consistent knowledge base.

This is because there is no information about the flight capability of antarctic birds in $\mathcal{R}_2$ and the principle of maximum entropy tends to add missing information in a most cautious way, i.e., unknown probabilities tend to 0.5*, and* 0.670 *is some kind of mean of* 0.5 *(antarctic birds) and* 0.95 *(non-antarctic birds).*

Example 6. *The last and probably most adequate way of imitating the conditional $(F|B, \mathit{not}\ \{A\})[0.95]$ without default negation is to split the information about birds and about antarctic birds into two separate conditionals:*

$$\mathcal{R}_3 = (\{A \to B\},\ \{(F|B)[0.95], (F|A)[0.5]\}).$$

As both probabilities $\mathcal{P}(F|B) = 0.95$ and $\mathcal{P}(F|A) = 0.5$ are explicitly stated in the knowledge base $\mathcal{R}_3$, they can be trivially inferred from $\mathcal{R}_3$ following the principle of maximum entropy, too. This strategy is in accordance with the remark about handling explicit exceptions with conditionals in the preliminaries.

The difference between $\mathcal{R}_{\mathsf{brd}}$ and $\mathcal{R}_3$ becomes visible if additional knowledge comes into play that implicitly affects the view on (antarctic) birds.

Example 7. *Assume that a reasoner has acquired knowledge about birds that goes beyond the knowledge stated in Example 1. More precisely, she has studied the species of penguins and now concludes that penguins are birds that are not able to fly. Further, she believes that penguins are typically antarctic birds and that penguins are quite rare. Depending on how she deals with her beliefs about the flight capability of birds and of antarctic birds (following the strategy from $\mathcal{R}_{\mathsf{brd}}$ or $\mathcal{R}_3$), her knowledge could be formalized by*

$$\mathcal{R}_{\mathsf{brd},b} = (\{A \to B, P \to B, P \to \overline{F}\},$$
$$\{(P|B)[0.001], (A|P)[0.8], (F|B, \mathit{not}\ \{A\})[0.95]\}) \text{ or}$$
$$\mathcal{R}_{3,b} = (\{A \to B, P \to B, P \to \overline{F}\},$$
$$\{(P|B)[0.001], (A|P)[0.8], (F|B)[0.95], (F|A)[0.5]\}).$$

When asking for the probabilities of $(F|B)[p]$ and $(F|A)[q]$, nothing changes if one takes the extended knowledge base $\mathcal{R}_{3,b}$ instead of $\mathcal{R}_3$ into account. As the queried probabilities p and q are explicitly stated in $\mathcal{R}_{3,b}$, one obtains $\mathsf{ME}(\mathcal{R}_{3,b})(F|B) = 0.95$ and $\mathsf{ME}(\mathcal{R}_{3,b})(F|A) = 0.5$. The information about the flight behavior of penguins that partially form a subclass of antarctic birds does not interfere with the information about antarctic birds. In other words, the probability of $(F|A)[0.5]$ is strict and does not really reflect the reasoner's uncertainty about the flight behavior of antarctic birds. In fact, it states that half of the antarctic birds are able to fly and the other half is not.

In contrast to that, the view on antarctic birds based on $\mathcal{R}_{\mathsf{brd},b}$ is more flexible as it allows one to incorporate the knowledge about the class of penguins into one's opinion about antarctic birds. One gets, $\mathsf{ME}(\mathcal{R}_{\mathsf{brd},b})(F|B) = 0.95$ and $\mathsf{ME}(\mathcal{R}_{\mathsf{brd},b})(F|A) \approx 0.366$. Hence, the likelihood that an antarctic bird is able to fly is lowered from 0.5 to 0.366 since penguins are typically antarctic birds that are not able to fly.

To briefly summarize the benefit of probabilistic conditionals with default negation, one can say that they allow one to exclude subclasses from a conditional statement without the need to give an explanation for the exclusion, while it remains possible to make implicit statements about the exceptionality of the subclass.

We now investigate formal properties of $\mid\!\sim_{\mathsf{ME}}$. The rationality postulates of System P proposed in (Kraus, Lehmann, and Magidor 1990) are commonly regarded as quality criteria for nonmonotonic inferences. In (Lukasiewicz 2005) probabilistic versions of these postulates are formulated which we adopt here in order to analyze $\mid\!\sim_{\mathsf{ME}}$. To call the postulates by name, they are *Reflexivity*, *Left Logical Equivalence*, *Cut*, *Cautious Monotonicity*, *Right Weakening*, and *Or*. Their probabilistic versions that are considered here are specified in Table 2 and differ from those in (Lukasiewicz 2005) in that we do not consider interval probabilities. Further properties of inference relations that are stated in (Lukasiewicz 2005) are *Inclusion*, *Conditioning*, *Irrelevance*, *Rational Monotonicity*, and *Inheritance of Logical Knowledge* that can also be found in Table 2.

Proposition 1. *The inference relation $\mid\!\sim_{\mathsf{ME}}$ satisfies all inference properties stated in Table 2, in particular the probabilistic version of System P.*

Proof Sketch. First, one observes that the mentioned inference properties hold for MaxEnt reasoning without default negation, i.e. if $\mathcal{D} = \emptyset$. It immediately follows that *Reflexivity*, *Or*, *Rational Monotonicity*, and *Inheritance of Logical Knowledge* are satisfied. The remaining properties follow by case analysis what we exemplarily discuss for the property *Conditioning*. For this, note that $A \equiv C$ implies $\mathcal{R}^A = \mathcal{R}^C$. Hence, in plain words, reasoning remains in the same reduct and either both $(B|A, not\ \mathcal{D})[p]$ and $(B|C, not\ \mathcal{D})[p]$ are blocked or both are not. In the first case the property is trivially satisfied and in the second case it is reduced to MaxEnt reasoning without default negation. The proof of the other properties is analogous as $A \models C$ implies $\mathcal{R}^A = \mathcal{R}^{AC}$ and $\mathcal{R}^A = \mathcal{R}^A$ is trivially true. With respect to *Irrelevance*, $\mathcal{R}^A = \mathcal{R}^{AC}$ holds due to the condition that no atom from $\mathcal{R}$ and $(B|A)[p]$ occurs in C. □

A knowledge base is *p-consistent* (Finthammer 2016) iff only factual knowledge leads to $\{0, 1\}$-probabilities, i.e., $\mathsf{ME}(\mathcal{R}^T)(\omega) = 0$ iff $\omega \in \Omega(\Sigma) \setminus \Omega_{\mathcal{F}}(\Sigma)$. For p-consistent knowledge bases, $\mathrel{|\!\!\sim}_{\mathsf{ME}}$ satisfies stronger versions of *Left Logical Equivalence*, *Cut*, *Cautious Monotonicity*, and *Conditioning* that are obtained by replacing $A \equiv C$ by $\mathcal{R} \mathrel{|\!\!\sim}_{\mathsf{ME}} AC \vee \overline{A}\overline{C}$ and $A \models C$ by $\mathcal{R} \mathrel{|\!\!\sim}_{\mathsf{ME}} A \rightarrow C$, respectively. That is, $AC \vee \overline{A}\overline{C}$ and $A \vee C$ do not have to be logically valid but it is sufficient that these propositions can be inferred from the reasoner's beliefs.

A further inference property that is stated in (Lukasiewicz 2005) is *Inheritance of Probabilistic Knowledge*:

$$\mathcal{R} \mathrel{|\!\!\sim}_{\mathsf{ME}} (B|A, not\ \mathcal{D})[p],\ C \rightarrow A \vdash \mathcal{R} \mathrel{|\!\!\sim}_{\mathsf{ME}} (B|C, not\ \mathcal{D})[p].$$

However, compliance with this property is explicitly not desired in nonmonotonic reasoning and *not* satisfied by $\mathrel{|\!\!\sim}_{\mathsf{ME}}$. If *Inheritance of Probabilistic Knowledge* would hold, the knowledge base $(\{P \rightarrow B, P \rightarrow \overline{F}\}, \{(F|B)[0.95]\})$ stating that penguins are non-flying birds and birds typically do fly with probability 0.95 would be inconsistent, for example.

To conclude, one can say that the inference relation $\mathrel{|\!\!\sim}_{\mathsf{ME}}$ satisfies a number of desired inference properties, albeit one has to say that some of the inference properties could be translated into the probabilistic setting in a more general fashion. For example, *Rational Monotonicity* in the version from (Lukasiewicz 2005) and in ours is trivially satisfied in probabilistic frameworks. This leaves room for further formal analysis of $\mathrel{|\!\!\sim}_{\mathsf{ME}}$.

Conclusion and Future Work

We combined the concepts of conditionals, negation as failure, and probabilities in order to define the language of probabilistic conditionals with default negation $\mathcal{PCL}^{\mathsf{not}}(\Sigma)$ that has it advantages when it is necessary to handle both implicit and explicit exceptions at the same time. We further defined the context-based inference relation $\mathrel{|\!\!\sim}_{\mathsf{ME}}$ following the principle of maximum entropy with which it is possible to infer implicit knowledge from a knowledge base that mentions facts and beliefs from $\mathcal{PCL}^{\mathsf{not}}(\Sigma)$. We highlighted the benefits of our approach by means of some illustrating examples and proved that $\mathrel{|\!\!\sim}_{\mathsf{ME}}$ satisfies a number of prominent inference properties from nonmonotonic reasoning.

In future work, we want to test our approach on real data and intensify the analysis of the differences between explicit and implicit exceptions. A more detailed analysis of Example 7 shows that the population sizes of classes and their exceptional subclasses play an important role in probabilistic reasoning. Therefore, we also want to extend our approach to probabilistic first order conditionals in order to incorporate domain sizes properly, e.g. in the context of the probabilistic Description Logic $\mathcal{ALC}^{\mathsf{ME}}$ (Baader et al. 2019).

Acknowledgments. This research was supported by the German National Science Foundation (DFG) Research Unit FOR 1513 on Hybrid Reasoning for Intelligent Systems.

References

Adams, E. 1965. The logic of conditionals. *Inquiry* 8(1-4):166–197.

Baader, F.; Ecke, A.; Kern-Isberner, G.; and Wilhelm, M. 2019. The complexity of the consistency problem in the probabilistic description logic $\mathcal{ALC}^{\mathsf{ME}}$. In *Proceedings of FroCos 2019*, 167–184.

Baral, C.; Gelfond, M.; and Rushton, J. N. 2009. Probabilistic reasoning with answer sets. *TPLP* 9(1):57–144.

Boyd, S. P., and Vandenberghe, L. 2014. *Convex Optimization*. Cambridge University Press.

Clark, K. L. 1977. Negation as failure. In *Symposium on Logic and Data Bases*, 293–322.

Cozman, F. G. 2019. The joy of probabilistic answer set programming. In *Proceedings of ISIPTA 2019*, 91–101.

Finthammer, M. 2016. *Concepts and algorithms for computing maximum entropy distributions for knowledge bases with relational probabilistic conditionals*. Ph.D. Dissertation, University of Hagen.

Gelfond, M., and Lifschitz, V. 2000. The stable model semantics for logic programming. *Logic Programming* 2.

Halpern, J. Y. 2003. *Reasoning about Uncertainty*. MIT Press.

Kern-Isberner, G. 2001. *Conditionals in Nonmonotonic Reasoning and Belief Revision*. Springer.

Kraus, S.; Lehmann, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.* 44(1-2):167–207.

Lukasiewicz, T. 2005. Weak nonmonotonic probabilistic logics. *Artif. Intell.* 168(1-2):119–161.

Paris, J. B. 2006. *The Uncertain Reasoner's Companion: A Mathematical Perspective*. Cambridge University Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

Rödder, W., and Meyer, C. 1996. Coherent knowledge processing at maximum entropy by SPIRIT. In *Proceedings of UAI 1996*, 470–476.

Wilhelm, M.; Eichhorn, C.; Niland, R.; and Kern-Isberner, G. 2017. A semantics for conditionals with default negation. In *Proceedings of ECSQARU 2017*, 257–267.