Novel Semantical Approaches to Relational Probabilistic Conditionals

Gabriele Kern-Isberner and **Matthias Thimm**
Technische Universität Dortmund
Department of Computer Science
Germany

Abstract

It seems to be a common view that in order to interpret probabilistic first-order sentences, either a statistical approach that counts (tuples of) individuals has to be used, or the knowledge base has to be grounded to make a possible worlds semantics applicable, for a subjective interpretation of probabilities. In this paper, we propose novel semantical perspectives on first-order (or relational) probabilistic conditionals that are motivated by considering them as subjective, but population-based statements. We propose two different semantics for relational probabilistic conditionals, and a set of postulates for suitable inference operators in this framework. Finally, we present two inference operators by applying the maximum entropy principle to the respective model theories. Both operators are shown to yield reasonable inferences according to the postulates.

Introduction

Applying probabilistic reasoning methods to relational[1], resp. first-order representations of knowledge is a very active and controversial area of research. During the past few years the fields of *probabilistic inductive logic programming* and *statistical relational learning* have put forth a lot of proposals that deal with combining traditional probabilistic models of knowledge like Bayes nets or Markov nets (Pearl 1998) with first-order logic, cf. (Getoor and Taskar 2007). The relational structure of many real-world problems such as telecommunication networks, citation analysis, human sciences, bioinformatics, and logistics as well as the presence of uncertainty in these problems demand sophisticated reasoning and learning methods employing both these concepts, see e. g. (Lodhi and Muggleton 2004; Cocura et al. 2006) for some applications. Two of the most prominent approaches for extending propositional approaches to the relational case are Bayesian logic programs (Getoor and Taskar 2007, Ch. 10) and Markov logic networks (Getoor and Taskar 2007, Ch. 12), extending Bayes nets and Markov nets, respectively. Both frameworks employ knowledge-based model construction techniques (Wellman, Breese, and Goldman 1992) to reduce the problem of probabilistic reasoning in a relational context to probabilistic reasoning in a propositional context. In both frameworks—and also in most other approaches—this is done by appropriately grounding the parts of the knowledge base that are needed for answering a particular query and treating this grounded parts as a propositional knowledge base.

Most of these approaches, however, are primarily concerned with machine learning problems, and do not care about logical or formal properties of relational probabilistic reasoning. The following example (inspired by (Delgrande 1998)) illustrates that even defining a proper semantics for first-order probabilistic knowledge bases is not an easy task. Let $el(X)$ denote that X is an elephant, $ke(X)$ means that X is a keeper, and $likes(X,Y)$ denotes that X likes Y.

$$r_1 \quad : \quad (likes(X,Y) \,|\, el(X) \wedge ke(Y))[0.6]$$
$$r_2 \quad : \quad (likes(X,fred) \,|\, el(X) \wedge ke(fred))[0.4]$$
$$r_3 \quad : \quad (likes(clyde,fred) \,|\, el(clyde) \wedge ke(fred))[0.7]$$

Rule r_1 expresses that in some given population, choosing randomly an elephant-keeper-pair, we would expect that the elephant likes the keeper with probability 0.7. However, keeper *Fred* and elephant *Clyde* are exceptional—mostly, elephants do not like *Fred*, but *Clyde* likes (even) *Fred*. Maybe *Clyde* is a particularly good-natured elephant, maybe he is as moody as *Fred* and likes only him. So, *Clyde* is definitely exceptional with respect to r_2, but maybe even with respect to r_1.

However, the example is ambiguous, and its formal interpretation via probabilistic constraints is intricate. Rule r_1 seems to express a belief an agent may hold about a population, while r_3 clearly expresses individual belief: Considering all situations (possible worlds) involving *Clyde* and *Fred* which are imaginable, in 70 % of them *Clyde* likes *Fred*. So, we might think of applying different techniques to r_1 and r_3, but r_2 obviously mixes the two types of knowledge, how should r_2 be dealt with?

In many approaches, e. g. in Bayesian logic programs and for Markov logic networks, the relational rules are grounded, and the probability is attached to each instance. For r_1, this means: $(likes(a,b) \,|\, el(a) \wedge ke(b))[0.6]$ for all $a, b \in U$.

[1]We will use the term "relational" to emphasize that the respective framework is not necessarily full-fledged first-order logic, but a fragment thereof, which, however, allows at least for the expression of statements on objects.

Here U is a properly (or arbitrarily) chosen universe. Besides the question, *how* U should be chosen, there are two other problems. First, grounding turns the relational statement r_1 into a collection of statements of the same type as r_3, i.e. statements about individual beliefs. The population aspect gets lost, more precisely: r_1 is no longer a statement describing a generic behaviour in a population of (possibly very) individualistic individuals, but is understood to be a statement on lemmings which all behave the same. Secondly, naive grounding techniques make the knowledge base inconsistent, as then r_3 collides with the respective instances of r_1 and r_2. So, grounding has to take further constraints into account, to return a consistent knowledge base (cf. (Finthammer, Loh, and Thimm 2009)).

In this paper, we will propose two approaches to giving formal semantics to relational probabilistic knowledge bases that aim at catching properly the common sense intuition and resolving ambiguities. We will focus on model-based inference operators for each of these semantics, in order to improve inferences from knowledge bases which usually represent only partial knowledge. We will make explicit what reasonable inference in this extended framework of relational probabilistic logic means by setting up a set of postulates. Of course, all this should be clearly related to work on probabilistic reasoning in the propositional case. In particular, we expect our semantics to coincide with propositional approaches, if the knowledge base is ground.

Moreover, we will present a model-based inference operator that is based on the principle of maximum entropy for each of the two semantics. This principle is known to ensure several desirable properties for commonsense reasoning in the propositional framework (Grove, Halpern, and Koller 1994; Kern-Isberner 2001). The idea of application is quite simple and similar to the propositional case: Having defined the set of models of a relational probabilistic knowledge base (according to each of the semantics), one chooses the unique probability distribution among these models that has maximal entropy, if possible, and therefore allows us to reason precisely (i.e. with precise probabilities, not based on intervals), but in a most cautious way (see (Grove, Halpern, and Koller 1994; Kern-Isberner 2001) for the theoretical foundations). Examples will illustrate in which respects these inference operators differ, but we will show that both inference operators comply with all postulates.

This paper continues and extends work begun in (Thimm 2009) while the main innovations of the present paper lie in the introduction of the novel aggregating semantics, a more elaborate discussion of desirable properties, and a more thorough evaluation of the inference operators.

The outline for the present paper is as follows. First, we formalize the syntactical details of a probabilistic first-order conditional logic, and propose two different semantics for it, the *averaging* and the *aggregating* semantics. Afterwards we discuss the problem of inference in this logic by developing several desirable properties of rational inference operators. We continue by presenting model-based inference operators that employ the principle of maximum entropy in both semantical frameworks, giving rise to two different inference operators which are exemplified and evaluated by means of the previously stated properties. We conclude with a brief summary and some discussions on related and further work. Proofs of results can be found in the appendix.

Syntax and Semantics of First-Order Conditional Logic

In the following we give an extension of probabilistic conditional logic to the relational case similar as in (Fisseler 2009).

We consider only a fragment of a first-order language, so let $\Sigma = \langle \mathsf{Pred}, D\rangle$ be a first-order signature consisting of a (finite) set of predicate symbols Pred, a finite set of constants D and without functions with arity greater zero. A predicate declaration P/n with a natural number n means that P is a predicate of arity n. Let $\mathcal{L}_\Sigma$ be a first-order language over the signature $\Sigma = \langle \mathsf{Pred}, D\rangle$ that is generated in the usual way using negation, conjunction, and disjunction, but without quantifiers. If appropriate we abbreviate conjunctions $A \wedge B$ by AB. We denote variables with a beginning uppercase, constants with a beginning lowercase letter, and vectors of these with $\vec{X}$ resp. $\vec{a}$.

A formula that contains no variable is called *ground*. Let $\mathsf{ground}_C(A)$ denote the set of ground instances of A with respect to a set of constants $C \subseteq D$, e.g. $\mathsf{ground}_{\{a,b\}}(A(X,Y))$ is $\{A(a,a), A(a,b), A(b,a), A(b,b)\}$.

Definition 1 (Probabilistic Conditional)**.** An expression of the form $(B \mid A)[\alpha]$ with $A, B \in \mathcal{L}_\Sigma$ (not necessarily ground) and a real number $\alpha \in [0,1]$ is called a *probabilistic conditional*. A probabilistic conditional $(B \mid A)[\alpha]$ is *ground* if both A and B are ground. Let $(\mathcal{L}|\mathcal{L})^{prob}$ be the set of all probabilistic conditionals over $\mathcal{L}_\Sigma$.

If the premise of A of a conditional $(B \mid A)[\alpha]$ is ground and tautological, i.e. $A \equiv \top$, we abbreviate $(B \mid \top)[\alpha]$ by $(B)[\alpha]$. A conditional of the form $(B)[\alpha]$ is also called a *probabilistic fact*. Let $\mathsf{ground}_C((B \mid A)[\alpha])$ denote the set of all grounded probabilistic conditionals of a conditional $(B \mid A)[\alpha]$ with respect to the set of constants $C \subseteq D$.

Definition 2 (Knowledge base)**.** A finite set R of probabilistic conditionals is called a *knowledge base*. A knowledge base R is *ground* if every probabilistic conditional in R is ground. Let $\mathfrak{R}$ denote the set of knowledge bases.

Introducing relational aspects in probabilistic statements raises some ambiguity on the understanding of these statements. We illustrate this problem on the example mentioned in the introduction (cf. also (Delgrande 1998)).

Example 1. Consider the knowledge base $R = \{r_1, r_2, r_3\}$ with

$$\begin{aligned}
r_1 &: (likes(X,Y) \mid el(X) \wedge ke(Y))[0.6] \\
r_2 &: (likes(X, fred) \mid el(X) \wedge ke(fred))[0.4] \\
r_3 &: (likes(clyde, fred) \mid el(clyde) \wedge ke(fred))[0.7]
\end{aligned}$$

The knowledge base R describes the relationships between keepers and elephants in a zoo, thereby stating both subjective degrees of belief on the relationship between Clyde and

Fred (r_3), as well as population-based probabilities that involve *all* elephants and keepers (r_1, r_2). So, r_3 should be interpreted via a possible worlds semantics, whereas r_2, r_3 seem to need some statistics (cf. e.g. (Bacchus et al. 1996)). As these two approaches are deemed substantially different, this dilemma is not easily resolved. In this paper, we propose thoroughly subjective approaches to probability even for the relational case, using a possible worlds semantics for all three statements above. This allows an intuitive and coherent interpretation of relational probabilistic statements that takes into account both informations on specific objects and informations on a population.

Formal semantics for first-order probabilistic conditional logic will be given by probability distributions that are defined over possible worlds of the given first-order language $\mathcal{L}_\Sigma$. Here, we use Herbrand interpretations for possible worlds. The Herbrand base $\mathcal{H}$ is the set of all ground atoms that can be built using the predicate symbols and constants in $\Sigma = \langle \mathsf{Pred}, D\rangle$, and a Herbrand interpretation is a subset of $\mathcal{H}$. A Herbrand interpretation ω satisfies a ground atom A, denoted by $\omega \models A$, if $A \in \omega$. The satisfaction relation $\models$ is extended to arbitrary ground formulas in the usual way. While we assume Pred to be fixed, we will allow the set of constants D to vary, in order to be able to investigate the influence of the universe (or respectively, its size) on the probabilistic evaluation of statements. So, we parametrize the set Ω_D of possible worlds by the set of constants, i. e. Ω_D is the set of all Herbrand interpretations on a signature with domain D. Likewise, we will write $\mathcal{L}_D$ and $(\mathcal{L}_D|\mathcal{L}_D)^{prob}$, to make the set of constants explicit.

Let $P : \Omega_D \to [0, 1]$ be a probability distribution over Ω_D, and let Prob_D be the set of all such probability distributions. $P \in \mathsf{Prob}_D$ is extended on ground formulas A by setting $P(A) = \sum_{\omega \models A} P(\omega)$. For the propositional case (Grove, Halpern, and Koller 1994; Kern-Isberner 2001) satisfaction of a conditional is defined via conditonal probabilities. Let $(B \mid A)[\alpha]$ be a ground conditional. Then a probability distribution P satisfies $(B \mid A)[\alpha]$, denoted by $P \models (B \mid A)[\alpha]$, if the following condition holds $P \models (B \mid A)[\alpha]$ iff $P(B \mid A) = \alpha$ and $P(A) > 0$. It remains to define a satisfaction relation for conditionals with variables (see Example 1). Taking a naïve approach by grounding all conditionals in R universally and taking this grounding R' as a propositional knowledge base, we can (usually) not determine any probability distribution that satisfies R' due to its inherent inconsistency (Finthammer, Loh, and Thimm 2009).

In the following, we propose two different approaches for semantics of $\mathcal{L}_\Sigma$ that coincide with the propositional case above on ground conditionals but differ on the interpretation of population-based statements.

Averaging Semantics

Our first approach gives semantics to probabilistic conditionals by averaging conditional probabilities. The entailment relation $\models^{cp}_{\varnothing}$ between distributions from Ω_D and relational probabilistic conditionals over $\mathcal{L}$ is defined by

$$P \models^{cp}_{\varnothing} (B(\vec{X}) \mid A(\vec{X}))[\alpha] \text{ iff}$$

$$\frac{\sum_{(B(\vec{c}) \mid A(\vec{c})) \in \mathsf{ground}_D((B(\vec{X}) \mid A(\vec{X})))} P(B(\vec{c}) \mid A(\vec{c}))}{|\mathsf{ground}_D(B(\vec{X}) \mid A(\vec{X}))|} = \alpha\,. \tag{1}$$

Intuitively spoken, a probability distribution P $\varnothing$-satisfies a conditional $(B \mid A)[\alpha]$ if the average of the individual instantiations of $(B \mid A)[\alpha]$ is α.

Remark 1. For a ground conditional $(G_2 \mid G_1)[\alpha]$ the operator $\models^{cp}_{\varnothing}$ indeed coincides with the propositional case due to $\mathsf{ground}_D(G_2 \mid G_1) = \{(G_2 \mid G_1)\}$.

As usual, a probability distribution P $\varnothing$-satisfies a knowledge base R, denoted $P \models^{cp}_{\varnothing} R$, if P $\varnothing$-satisfies every probabilistic conditional $r \in R$. We say that R is $\varnothing$-consistent iff there is at least one P with $P \models^{cp}_{\varnothing} R$, otherwise R is $\varnothing$-inconsistent.

Aggregating Semantics

Our second semantical approach is inspired by statistical approaches. However, instead of counting objects, or tuples of objects, respectively, that make a formula true, we sum up the probabilities of the correspondingly instantiated formulas. The entailment relation $\models^{cp}_{\odot}$ between distributions from Ω_D and relational probabilistic conditionals over $\mathcal{L}$ is defined by

$$P \models^{cp}_{\odot} (B(\vec{X}) \mid A(\vec{X}))[\alpha] \text{ iff}$$

$$\frac{\sum_{(B((\vec{c}) \mid A(\vec{c})) \in \mathsf{ground}_D((B(\vec{X}) \mid A(\vec{X})))} P(A(\vec{c})B(\vec{c}))}{\sum_{(B(\vec{c}) \mid A(\vec{c})) \in \mathsf{ground}_D((B(\vec{X}) \mid A(\vec{X})))} P(A(\vec{c}))} = \alpha\,. \tag{2}$$

If P is a uniform distribution, we end up with a statistical interpretation of the conditional. However, the probabilities in this paper will be subjective, so $\models^{cp}_{\odot}$ mimicks the statistical view from a subjective perspective.

Remark 2. As for $\models^{cp}_{\varnothing}$, for a ground conditional $(B_2 \mid A_1)[\alpha]$ the operator $\models^{cp}_{\odot}$ coincides with the propositional case due to $\mathsf{ground}_D(B_2 \mid A_1) = \{(B_2 \mid A_1)\}$.

As above, a probability distribution P $\odot$-satisfies a knowledge base R, denoted $P \models^{cp}_{\odot} R$, if P $\odot$-satisfies every probabilistic conditional $r \in R$. We say that R is $\odot$-consistent iff there is at least on P with $P \models^{cp}_{\odot} R$, otherwise R is $\odot$-inconsistent.

Comparing the Semantics

Due to remarks 1 and 2, both semantics agree on ground conditionals. Furthermore, it is straightforward to show that $\models^{cp}_{\varnothing}$ and $\models^{cp}_{\odot}$ also agree on probabilistic facts (that may contain variables).

Proposition 1. *Let $P \in \mathsf{Prob}_D$ be a probability distribution and $(B)[\alpha] \in (\mathcal{L}_D|\mathcal{L}_D)^{prob}$ a probabilistic fact. Then it holds that $P \models^{cp}_{\varnothing} (B)[\alpha]$ iff $P \models^{cp}_{\odot} (B)[\alpha]$.*

For general conditionals in $(\mathcal{L}_D|\mathcal{L}_D)^{prob}$, however, the two semantics turn out to be different, as the following example shows.

Example 2. Let $A/1$ and $B/1$ be two predicates, let the D consist of the five elements $a_1, \ldots, a_5$, and consider the following (ground) knowledge base R:

$$\begin{array}{rr} (A(a_1))[0.5] & (A(a_2))[0.1] \\ (A(a_3))[0.9] & (A(a_4))[0.6] \\ (A(a_5))[0.4] & (B(a_1)A(a_1))[0.5] \\ (B(a_2)A(a_2))[0.1] & (B(a_3)A(a_3))[0.9] \\ (B(a_4)A(a_4))[0.4] & (B(a_5)A(a_5))[0.1] \end{array}$$

In addition, consider the conditional $r = (B(X) \mid A(X))[0.8]$. On the one hand, any probability distribution P with $P \models^{cp}_{\odot} R$ also obeys $P \models^{cp}_{\odot} r$ as

$$\begin{aligned} & \frac{P(B(a_1)A(a_1)) + \cdots + P(B(a_5)A(a_5))}{P(A(a_1)) + \cdots + P(A(a_5))} \\ = \; & \frac{0.5 + 0.1 + 0.9 + 0.4 + 0.1}{0.5 + 0.1 + 0.9 + 0.6 + 0.4} = 0.8 \end{aligned}$$

On the other hand, every probability distribution P with $P \models^{cp}_{\varnothing} R$ does not obey $P \models^{cp}_{\varnothing} r$ due to

$$\begin{aligned} & 1/5\ (P(B(a_1) \mid A(a_1)) + \cdots + P(B(a_5) \mid A(a_5))) \\ = \; & \left(\frac{0.5}{0.5} + \frac{0.1}{0.1} + \frac{0.9}{0.9} + \frac{0.4}{0.6} + \frac{0.1}{0.4}\right)/5 = 0.78\overline{3} \neq 0.8 \end{aligned}$$

As $P \models^{cp}_{\odot} R$ is equivalent to $P \models^{cp}_{\varnothing} R$ due to Proposition 1 the different semantics may lead to different inferences. Furthermore, the two semantics feature a different notion of consistency as $R \cup \{r\}$ is $\varnothing$-inconsistent but $\odot$-consistent.

Inference in First-Order Conditional Logic

We are interested in finding a "good" probability distribution P that satisfies all probabilistic conditionals of a given knowledge base R given one of the two proposed semantics. More specifically, we are interested in an operator $\mathcal{I}(R, D)$ that takes a knowledge base R and as set of constants D as input and returns a probability distribution $P = \mathcal{I}(R, D) \in \mathsf{Prob}_D$ as output such that P describes R "best" in a commonsensical manner. In particular, the resulting distribution should be a model of R. So, let $\models^{cp}_{\circ}$ be any entailment relation between distributions from Ω_D and relational probabilistic conditionals from $(\mathcal{L}_D|\mathcal{L}_D)^{prob}$. In this section, we state some properties that a reasonable model-based $\mathcal{I}$ operator should observe. In the following section, we will present two operators that comply with all postulates.

In order to ease notation and presentation, we will implicitly assume that R is defined over a language $\mathcal{L}_D$ the predicate symbols of which are held fixed, and the set D of constants is to contain all constants appearing in R.

Our first demand for an operator $\mathcal{I}$ to be appropriate is its well-definedness. As an inconsistent knowledge base R has no models and therefore an operator $\mathcal{I}$ cannot determine any model of R for reasoning, let undef be a new symbol for this case. Let $\mathfrak{D}$ be the set of all sets of constants; for each knowledge base R, let $\mathfrak{D}_R$ contain all sets of constants that contain all constants from R. Let $\mathcal{I} : \mathfrak{R} \times \mathfrak{D} \to \mathsf{Prob}_D \cup \{\mathsf{undef}\}$ be an operator that maps a knowledge base $R \in \mathfrak{R}$ and a set $D \in \mathfrak{D}_R$ of constants onto a probability distribution $P \in \mathsf{Prob}_D$, or to undef.

(Well-Definedness) It is $\mathcal{I}(R, D) \in \mathsf{Prob}_D$ such that $\mathcal{I}(R, D) \models^{cp}_{\circ} R$ iff $D \in \mathfrak{D}_R$ and $R \subseteq \mathcal{L}_D$ is $\circ$-consistent.

We need some further notation to go on. For a formula A let $A[d/c]$ denote the formula that is the same as A except that every occurrence of the term c (either a variable or a constant) is substituted with the term d. More generally, let $A[d_1/c_1, \ldots, d_n/c_n]$ denote the formula that is the same as A except that every occurrence of c_i is substituted with d_i for $1 \leq i \leq n$ simultaneously. Furthermore, let $A[c \leftrightarrow d]$ be an abbreviation for $A[c/d, d/c]$. The substitution operator $[\cdot]$ is extended on sets of formulas, conditionals, and knowledge bases in the usual way.

When considering knowledge bases based on a relational language the beliefs one obtains for specific individuals is of special interest. An important demand to be made is that for indistinguishable individuals, the same information should be obtained. Here, indistinguishability is defined with respect to the information expressed by R. More specifically, if the explicit information encoded in R for two different individuals $c_1, c_2 \in D$ is the same, the probability distribution $P = \mathcal{I}(R, D)$ should treat them as indistinguishable. We formalize this indistinguishability by introducing an equivalence relation on constants.

Definition 3 (Syntactical Equivalence)**.** Let R be a knowledge base. The constants $c_1, c_2 \in D$ are *syntactically equivalent* with respect to R, denoted by $c_1 \equiv_R c_2$, iff $R = R[c_1 \leftrightarrow c_2]$.

Observe that $\equiv_R$ is indeed an equivalence relation, i. e., it is reflexive, transitive, and symmetric. The equivalence classes of $\equiv_R$ are called *R-equivalence classes* and the set of all R-equivalence classes is denoted by $\mathcal{S}_R$. Note, that the notion of syntactical equivalence bears a resemblance with the notion of *reference classes* (Bacchus et al. 1996) but on a pure syntactical level.

Using syntactical equivalence we can state our demand for equal treatment of indistinguishable individuals as follows.

(Prototypical Indifference) Let R be a knowledge base on $\mathcal{L}_D$ and A a ground sentence. For any $c_1, c_2 \in D$ with $c_1 \equiv_R c_2$ it is $\mathcal{I}(R, D)(A) = \mathcal{I}(R, D)(A[c_1 \leftrightarrow c_2])$.

Even more basically, renaming an individual should have no impact on the information that can be derived for it.

(Name Irrelevance) Let R be a knowledge base on $\mathcal{L}_D$, $d \notin D$ a new constant, and $A \in \mathcal{L}_D$ a ground sentence. For every $c \in D$, it holds that $\mathcal{I}(R, D)(A) = \mathcal{I}(R[d/c], (D \cup \{d\}) \setminus \{c\})(A[d/c])$ where $R[d/c]$ is a knowledge base on $\mathcal{L}_{(D \cup \{d\}) \setminus \{c\}}$.

As can easily be seen, every function $\mathcal{I}$ satisfying (Name Irrelevance) also satisfies (Prototypical Indifference).

Proposition 2. *If $\mathcal{I}$ satisfies (Name Irrelevance) then $\mathcal{I}$ satisfies (Prototypical Indifference).*

It is clear that (Prototypical Indifference) can be verified by considering all Herbrand interpretations, respectively. Thus, the following proposition is given without proof.

Proposition 3. *$\mathcal{I}$ satisfies (Prototypical Indifference) iff for all knowledge bases $R \in \mathfrak{R}$, and for all constants $c_1, c_2 \in D$ that do not occur in R, and for all interpretations $\omega \in \Omega_D$, $\mathcal{I}(R, D)(\omega) = \mathcal{I}(R, D)(\omega[c_1 \leftrightarrow c_2])$.*

On the other hand, from (Prototypical Indifference) some generalizations follow immediately.

Proposition 4. *Let $\mathcal{I}$ satisfy (Prototypical Indifference). Let R be a knowledge base on $\mathcal{L}_D$.*

1. *Let G_1, G_2 be two ground sentences. For $c_1, c_2 \in D$ with $c_1 \equiv_R c_2$ it holds $\mathcal{I}(R, D)(G_2|G_1) = \mathcal{I}(R, D)(G_2[c_1 \leftrightarrow c_2] \mid G_1[c_1 \leftrightarrow c_2])$.*
2. *Let $S \in \mathcal{S}_R$, $c_1, \ldots, c_n \in S$, and $\sigma : S \to S$ a permutation on S, i. e. a bijective function on S. Then it holds $\mathcal{I}(R, D)(A) = \mathcal{I}(R, D)(A[\sigma(c_1)/c_1, \ldots, \sigma(c_n)/c_n])$.*

The following postulate focusses on the implications that a population-based statement $r = (B(\vec{X}) \mid A(\vec{X}))[\alpha]$ should have for the probability of a proper instantiation $P(B(\vec{c})|A(\vec{c})$. Our intention about r is that in general, the conditional probability of $B(\vec{c})$ given $A(\vec{c})$ "should" be (around) α. But surely, we cannot guarantee that every possible instantiation r' of r will conform to a strict interpretation of this demand. This follows mainly from the fact, that using ground conditionals we should be able to give exceptions to this rule, cf. Example 1. What we are really want to describe when representing a population-based statement r is that *given an adequate large domain*, the respective conditional probability for constant tuples that may serve as prototypes will converge towards α. This behavior resembles the intuition behind the "Law of Large Numbers".

(Conditional Probability in the Limit (CPL)) Let D be a set of constants that contain all constants from R, and assume an increasing sequence of sets of constants $D = D_1 \subset D_2 \subset \ldots$. Let R be a $\circ$-consistent knowledge base on $\mathcal{L}_D$; then R can also be considered as a knowledge base on $\mathcal{L}_{D_1}, \mathcal{L}_{D_1}, \ldots$. For a conditional $r = (B(\vec{X}) \mid A(\vec{X}))[\alpha] \in R$, let $(B(\vec{c}) \mid A(\vec{c}))[\alpha]$ be a proper instantiation of r with constants $\vec{c}$ that do not appear in R. Then it holds that

$$\lim_{i\to\infty} \mathcal{I}(R, D_i)(B(\vec{c}) \mid A(\vec{c})) = \alpha \quad .$$

The important aspect of population-based statements is their capability of expressing a general behaviour within a population while allowing for exceptions. So, population-based statements are to reflect some kind of *expected value* over the set of individual instantiations that aggregates individual behaviours. As such, if the probability of one instantiation of a population-based statements lies below the probability assigned to the statement there has to be another instantiation with a probability higher than this probability value in order to compensate for the other exception (remember that the universe D is assumed to be finite).

(Compensation) Let R be a $\circ$-consistent knowledge base and $(B(\vec{X}) \mid A(\vec{X}))[\alpha] \in R$ a non-ground conditional with $0 < \alpha < 1$. If $\vec{c_1}$ is a vector of constants such that $\mathcal{I}(R, D)(B(\vec{c_1}) \mid A(\vec{c_1})) < \alpha$ then there is another vector of constants $\vec{c_2}$ with $\mathcal{I}(R, D)(B(\vec{c_2}) \mid A(\vec{c_2})) > \alpha$.

On the other hand, when considering non-ground conditionals $(B(\vec{X}) \mid A(\vec{X}))[\alpha]$ with $\alpha \in \{0, 1\}$ no compensation for exceptions is possible thus requiring *direct inference* (Bacchus et al. 1996) for this particular case.

(Strict Inference) Let R be a $\circ$-consistent knowledge base and $(B(\vec{X}) \mid A(\vec{X}))[\alpha] \in R$ a non-ground conditional with $\alpha \in \{0, 1\}$. Then for any $(B(\vec{c}) \mid A(\vec{c})) \in \mathsf{ground}_D(A(\vec{X}) \mid B(\vec{X}))$, $\mathcal{I}(R, D)(B(\vec{c}) \mid A(\vec{c})) = \alpha$.

In the following section, we present two operators that satisfy all postulates given above.

Relational Maximum Entropy Reasoning

In the propositional case, ME-inference (*Maximum Entropy*) has proven to be a suitable approach for commonsense reasoning as it features several nice properties (Grove, Halpern, and Koller 1994; Kern-Isberner 2001). The entropy $H(P)$ of a probability distribution P is defined as $H(P) = -\sum_{\omega \in \Omega_{\mathcal{V}}} P(\omega) \log P(\omega)$, and measures the amount of indeterminateness inherent in P. By selecting the unique probability distribution P^* among all probabilistic models of a (propositional) set of formulas S that has maximal entropy, i. e. by computing the solution to the optimization problem $P^* := \mathsf{ME}(S) = \arg\max_{P \models S} H(P)$, we get the one probability distribution that satisfies S and adds as little information as necessary. For further details, we refer to (Grove, Halpern, and Koller 1994; Kern-Isberner 2001).

As we are interested in generalizing the propositional ME-operator to the first-order case, we will postulate a proper form of compatibility to the propositional ME-inference, in addition to the postulates stated for general inference operators in the previous section. For ground knowledge bases (which can be considered as propositional knowledge bases), the operation $\mathcal{I}$ should coincide with the ME operator on propositional knowledge bases.

(ME-Compatibility) Let R be a ground knowledge base and D the constants appearing in R. If A is a ground sentence then it is $\mathsf{ME}(R)(A) = \mathcal{I}(R, D)(A)$.

After having introduced the averaging and the aggregating semantics for relational probabilistic knowledge bases, now we apply the maximum entropy principle to the respective model sets to single out "best" models.

Relational Maximum Entropy Inference by Averaging Probabilities

In the following we define our first variant of an ME-inference $\mathcal{I}_\varnothing : \mathfrak{R} \times \mathfrak{D} \to \mathsf{Prob}_D \cup \{\mathsf{undef}\}$ in a relational context, that is based upon the semantics $\models^{cp}_\varnothing$. A preliminary discussion of this operator can also be found in (Thimm 2009). As (1) yields a set of non-convex constraints we define $\mathcal{I}_\varnothing(R, D)$ as

$$\mathcal{I}_\varnothing(R, D) = \begin{cases} \arg\max\limits_{P \models^{cp}_\varnothing R} H(P) & \text{if unique and } D \in \mathfrak{D}_R \\ \mathsf{undef} & \text{otherwise} \end{cases} \qquad (3)$$

The second case catches scenarios where either R is $\varnothing$-inconsistent or the optimization problem of the first case is not uniquely solvable. Obviously, $\mathcal{I}_{\varnothing}$ is a model-based inference operator using semantics $\models^{cp}_{\varnothing}$. In particular, if R is $\varnothing$-consistent there is at least one probability distribution with maximum entropy that can be chosen in Equation (3).

Example 3. We continue Example 1. Let $\mathcal{L}_D$ be a first-order language with predicates $el/1$, $ke/1$, and $likes/2$ and domain $D = \{clyde, dumbo, giddy, fred, dave\}$. Let R be given by

$$(el(clyde))[1] \qquad (el(giddy))[1]$$
$$(ke(fred))[1] \qquad (ke(dave))[1]$$
$$(likes(X,Y)\,|\,el(X) \wedge ke(Y))[0.6] \tag{4}$$
$$(likes(X,fred)\,|\,el(X) \wedge ke(fred))[0.4] \tag{5}$$
$$(likes(clyde,fred)\,|\,el(clyde) \wedge ke(fred))[0.7] \tag{6}$$

Notice, that we have no knowledge of Dumbo being an elephant. In the following we give the probabilities of several instantiations of $likes$ in $\mathcal{I}_{\varnothing}(R,D)$.

$$\mathcal{I}_{\varnothing}(R,D)(likes(clyde,dave)) \approx 0.723 \tag{7}$$
$$\mathcal{I}_{\varnothing}(R,D)(likes(dumbo,dave)) \approx 0.642 \tag{8}$$
$$\mathcal{I}_{\varnothing}(R,D)(likes(giddy,dave)) \approx 0.723 \tag{9}$$
$$\mathcal{I}_{\varnothing}(R,D)(likes(clyde,fred)) = 0.7 \tag{10}$$
$$\mathcal{I}_{\varnothing}(R,D)(likes(dumbo,fred)) \approx 0.387 \tag{11}$$
$$\mathcal{I}_{\varnothing}(R,D)(likes(giddy,fred)) \approx 0.36 \tag{12}$$
$$\mathcal{I}_{\varnothing}(R,D)(el(dumbo)) \approx 0.312 \tag{13}$$

Notice, how the deviations brought about by the exceptional individuals Clyde and Fred have to be balanced out by the other individuals. For example, the probabilities of the individual elephants liking Dave are greater than conditional (4) specified them to be. This is because the probabilities of the elephants liking Fred is considerably smaller as demanded by conditional (5). Nonetheless, the average of the conditional probabilities do indeed satisfy the conditionals in R. Notice furthermore, that the probability of Dumbo being an elephant is very small—see (13)—considering that maximum entropy is achieved by deviating only as little as possible from the uniform distribution. But due to the interaction of the conditionals in R, a smaller probability of Dumbo being an elephant is necessary in order to achieve the correct average conditional probabilities defined in the knowledge base. Thus, the belief of Dumbo being an elephant alleviates due to the premise of believing in the defined conditionals.

In the following we give some theoretical results that the proposed operator $\mathcal{I}_{\varnothing}$ indeed fulfills the desired properties discussed in in the previous section. Due to the non-convexity of the optimization problem defined by (3) $\mathcal{I}_{\varnothing}$ satisfies (Well-Definedness) only for the case that (3) is uniquely solvable. However, all examples considered so far were indeed uniquely solvable.

In order to show that $\mathcal{I}_{\varnothing}$ satisfies (CPL) we need the following lemma.

Lemma 1. *Let R be a knowledge base on $\mathcal{L}_D$ such that $\mathcal{I}_{\varnothing}(R,D) \neq \mathsf{undef}$. Then for any D' with $D \subseteq D'$ it is $\mathcal{I}_{\varnothing}(R,D') \neq \mathsf{undef}$.*

Proposition 5. *$\mathcal{I}_{\varnothing}$ satisfies (Name Irrelevance), (Prototypical Indifference), (ME-Compatibility), (Conditional Probability in the Limit), (Compensation), and (Strict Inference).*

Relational Maximum Entropy Inference by Aggregating Probabilities

In a similar manner, we define the ME-inference operator $\mathcal{I}_{\odot} : \mathfrak{R} \times \mathfrak{D} \to \mathsf{Prob}_D \cup \{\mathsf{undef}\}$ that is based upon the semantics $\models^{cp}_{\odot}$. Let

$$\mathcal{I}_{\odot}(R,D) = \begin{cases} \arg\max_{P \models^{cp}_{\odot} R} H(P) & \text{if } R \text{ } \odot\text{-consistent and } D \in \mathfrak{D}_R \\ \mathsf{undef} & \text{otherwise} \end{cases} \tag{14}$$

Obviously, $\mathcal{I}_{\odot}$ is a model-based inference operator using semantics $\models^{cp}_{\odot}$. In this semantical context, the conditionals from R induce linear constraints on the probabilities of the possible worlds so that the set of probability distributions satisfying R forms a convex set. This makes the solution to the optimization problem (14) unique (if a solution exists). The following lemma is given without proof.

Lemma 2. *Let $r = (B(\vec{X})\,|\,A(\vec{X}))[\alpha]$ be a probabilistic conditional and Sol_r the set of probability distributions that satisfy r, i. e. it is $Sol_r = \{P \mid P \models^{cp}_{\odot} (B(\vec{X})\,|\,A(\vec{X}))[\alpha]\}$. Then Sol_r is convex.*

Proposition 6. *The probability distribution defined by (14) is uniquely determined if R is $\odot$-consistent.*

Example 4. We apply $\mathcal{I}_{\odot}$ onto the knowledge base in Example 3. This yields the following inferences:

$$\mathcal{I}_{\odot}(R,D)(likes(clyde,dave)) \approx 0.8 \tag{15}$$
$$\mathcal{I}_{\odot}(R,D)(likes(dumbo,dave)) \approx 0.64 \tag{16}$$
$$\mathcal{I}_{\odot}(R,D)(likes(giddy,dave)) \approx 0.8 \tag{17}$$
$$\mathcal{I}_{\odot}(R,D)(likes(clyde,fred)) = 0.7 \tag{18}$$
$$\mathcal{I}_{\odot}(R,D)(likes(dumbo,fred)) \approx 0.356 \tag{19}$$
$$\mathcal{I}_{\odot}(R,D)(likes(giddy,fred)) \approx 0.196 \tag{20}$$
$$\mathcal{I}_{\odot}(R,D)(el(dumbo)) \approx 0.475 \tag{21}$$

The results are similar to those computed by using $\mathcal{I}_{\varnothing}$ in the example above. In particular, with regard to liking Dave, both approaches calculate very similar probabilities for all individuals mentioned in the queries. Here, Dumbo—the individual not known to be an elephant—likes Dave with a lower probability than the elephants Clyde and Giddy, cf. (15), (16), and (17). More substantial differences can be noticed with respect to the elephants' liking the moody keeper Fred. For Giddy liking Fred, $\mathcal{I}_{\odot}$ returns a considerably lower probability than $\mathcal{I}_{\varnothing}$, see (20). On the other hand, $\mathcal{I}_{\odot}$ is more cautious when processing information on Dumbo, its probability of being an elephant is nearly 0.5 (21), while $\mathcal{I}_{\varnothing}$ suggests that Dumbo is not an elephant.

We will now show that $\mathcal{I}_{\odot}$ satisfies all postulates listed in the previous Section.

Proposition 7. $\mathcal{I}_\odot$ *satisfies (Well-Definedness), (Name Irrelevance), (Prototypical Indifference), (*ME*-Compatibility), (Conditional Probability in the Limit), (Strict Inference), and (Compensation).*

Discussion – Related and Further Work

In this paper, we developed a first-order conditional logic and proposed two different semantics for it. We deviced a set of desirable properties of inference operators on this logic, and for each of the semantics we proposed an inference operator that extends proposition inference on maximum entropy to the relational case. Both operators fulfill in principle the catalogue of desired properties.

From a computational point of view the operator, $\mathcal{I}_\odot$ and thus the semantics $\models_\odot^{cp}$ seems to be the favorable choice for reasoning in first-order conditional logic. Although, for a straightforward implementation of the optimization problems (3) and (14), both operators need an exponential transformation (as the probabilities for all Herbrand interpretations have to be considered). But for solving the (convex) optimization problem (14) efficient algorithms are available (Boyd and Vandenberghe 2004). However, except for these computational issues, both semantics seem to be reasonable choices for interpreting first-order conditional logic. Part of our current research is an even deeper investigation and comparison of the proposed semantical approaches. Another interesting issue for future work lies in investigating the possibility if fast algorithms for computing the ME-distribution in propositional frameworks, such as SPIRIT (Rödder and Meyer 1996), can be adapted to the relational case.

Some approaches to define a proper probabilistic logic for first-order fragments have been proposed in previous works, and some of them even apply the principle of maximum entropy for inferences. The papers by Halpern and colleagues (Bacchus et al. 1996; Grove, Halpern, and Koller 1994) aim at bridging statistical and subjective views on probabilistic beliefs by showing how subjective beliefs arise from statistical information by considering approximative probabilities and limits. The principle of maximum entropy plays a prominent role in these frameworks, too, but the authors mention problems when applying this principle to knowledge bases with n-ary predicates with $n > 1$. As our semantical approaches are thoroughly subjective by choosing subjective probabilities throughout, we did not encounter most of the problems that those authors have to struggle with. For instance, in statistical approaches to probabilities, the size of the universe determines the probabilities that can be realized, so approximations of probabilities have to be considered. This is not the case in our approaches, as no frequentistic interpretation underlies the probabilities. Moreover, the application of the maximum entropy principle to knowledge bases with arbitrary predicates seems to be unproblematic, but this has to be investigated in more detail in further work.

Most approaches that make use of a subjective, possible worlds semantics for first-order probabilistic logic, as e. g. (Kern-Isberner and Lukasiewicz 2004; Fisseler 2009; Finthammer, Loh, and Thimm 2009), but also Markov logic networks (Getoor and Taskar 2007, Ch. 12) and Bayesian logic programs (Getoor and Taskar 2007, Ch. 10), apply their techniques to a grounded version of the knowledge base. Grounding probabilistic relational conditionals with precise probabilities and using them as a set of constraints, however, may give rise easily to conflicts and inconsistencies. Such problems were circumvented by the approaches cited above by considering imprecise, interval-valued probabilities, by restricting syntactically the grounding of the formulas, by using external combination functions for different instantiations of conditionals, or by not considering the knowledge base as a set of constraints. Our approaches aim at reflecting an overall behaviour within a population to which each individual contributes, while at the same time allowing individuals to defer drastically from that behaviour. In this way, both class knowledge and individual, maybe exceptional knowledge can be represented and processed within one framework. As the satisfaction of the (CPL) postulate shows, the overall behaviour might also be interpreted as a prototypical behaviour in universes which are large enough. Nevertheless, in a companion paper (Loh, Thimm, and Kern-Isberner 2010), we propose different grounding strategies for relational probabilistic knowledge bases that avoid inconsistencies, and apply the maximum entropy principle to the arising propositional knowledge bases. A thorough comparison with the results from this work is also part of our ongoing research.

Acknowledgements. The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant KE 1413/2-1). We also thank Adam Chachaj for implementing prototypes of the presented inference operators that were used in computing the examples in this paper.

References

Bacchus, F.; Grove, A. J.; Halpern, J. Y.; and Koller, D. 1996. From Statistical Knowledge Bases to Degrees of Belief. *Artificial Intelligence* 87(1–2):75–143.

Beachy, J. A., and Blair, W. D. 2005. *Abstract Algebra*. Long Grove, Illinois, USA: Waveland Press, Inc., third edition.

Boyd, S. P., and Vandenberghe, L. 2004. *Convex Optimization*. Cambridge University Press.

Cocura, A.; Kersting, K.; Plageman, C.; Burgard, W.; and Raedt, L. D. 2006. Learning relational navigation policies. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-06)*.

Delgrande, J. P. 1998. On First-Order Conditional Logics. *Artificial Intelligence* 105(1–2):105–137.

Finthammer, M.; Loh, S.; and Thimm, M. 2009. Towards a Toolbox for Relational Probabilistic Knowledge Representation, Reasoning, and Learning. In *Workshop on Relational Approaches to Knowledge Representation and Learning, Proceedings*.

Fisseler, J. 2009. *Learning and Modeling with Probabilistic Conditional Logic*. Ph.D. Dissertation, FernUniversität in Hagen, Germany.

Getoor, L., and Taskar, B., eds. 2007. *Introduction to Statistical Relational Learning*. MIT Press.

Grove, A. J.; Halpern, J. Y.; and Koller, D. 1994. Random Worlds and Maximum Entropy. *Journal of Artificial Intelligence Research (JAIR)* 2:33–88.

Kern-Isberner, G., and Lukasiewicz, T. 2004. Combining probabilistic logic programming with the power of maximum entropy. *Artificial Intelligence, Special Issue on Nonmonotonic Reasoning* 157(1-2):139–202.

Kern-Isberner, G. 2001. *Conditionals in Nonmonotonic Reasoning and Belief Revision*. Number 2087 in Lecture Notes in Computer Science. Springer.

Lodhi, H., and Muggleton, S. 2004. Modelling metabolic pathways using stochastic logic programs-based ensemble methods. In *Proceedings of the 2nd International Conference on Computational Methods in System Biology*. Springer.

Loh, S.; Thimm, M.; and Kern-Isberner, G. 2010. Grounding techniques for first-order probabilistic conditional logic. (in preparation).

Pearl, J. 1998. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Rödder, W., and Meyer, C.-H. 1996. Coherent Knowledge Processing at Maximum Entropy by SPIRIT. In *Proc. of the Twelfth Conf. on Uncertainty in Artificial Intelligence*, 470–476.

Shore, J., and Johnson, R. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* 26:26–37.

Thimm, M. 2009. Representing Statistical Information and Degrees of Belief in First-Order Probabilistic Conditional Logic. In *Workshop on Relational Approaches to Knowledge Representation and Learning, Proceedings*, 49–63.

Wellman, M. P.; Breese, J. S.; and Goldman, R. P. 1992. From Knowledge Bases to Decision Models. *The Knowledge Engineering Review* 7(1):35–53.

Proofs of Results

Proposition 1. *Let P be a probability distribution and $(B)[\alpha]$ a probabilistic fact. Then it holds $P \models^{cp}_{\varnothing} (B)[\alpha]$ iff $P \models^{cp}_{\odot} (B)[\alpha]$.*

Proof. It holds $P \models^{cp}_{\varnothing} (B)[\alpha]$ iff

$$\frac{\sum_{B' \in \mathsf{ground}_D(B)} P(B')}{|\mathsf{ground}_D(B)|} = \alpha$$

by definition and furthermore due to $P(\top) = 1$ it holds

$$\begin{aligned} P \models^{cp}_{\odot} (B)[\alpha] \quad &\Leftrightarrow \quad \frac{\sum_{B' \in \mathsf{ground}_D(B)} P(B')}{\sum_{B' \in \mathsf{ground}_D(B)} P(\top)} = \alpha \\ &\Leftrightarrow \quad \frac{\sum_{B' \in \mathsf{ground}_D(B)} P(B')}{|\mathsf{ground}_D(B)|} = \alpha \end{aligned}$$

□

Proposition 2. *If $\mathcal{I}$ satisfies (Name Irrelevance) then $\mathcal{I}$ satisfies (Prototypical Indifference).*

Proof. Let R be a knowledge base on $\mathcal{L}_D$ and $d_1, d_2 \notin D$. Let furthermore $c_1, c_2 \in D$ with $c_1 \equiv_R c_2$ and $c_1 \neq c_2$. Then it holds for a ground sentence A:

$$\begin{aligned} & \mathcal{I}(R, D)(A) \\ = \ & \mathcal{I}(R[d_1/c_1], (D \cup \{d_1\}) \setminus \{c_1\})(A[d_1/c_1]) \\ = \ & \mathcal{I}(R[d_1/c_1, d_2/c_2], (D \cup \{d_1, d_2\}) \setminus \{c_1, c_2\}) \\ & \qquad (A[d_1/c_1, d_2/c_2]) \end{aligned}$$

As $R[d_1/c_1, d_2/c_2]$ is a knowledge base on $\mathcal{L}_{(D \cup \{d_1,d_2\}) \setminus \{c_1,c_2\}}$ and $c_1, c_2 \notin (D \cup \{d_1, d_2\}) \setminus \{c_1, c_2\}$ it holds

$$\begin{aligned} & \mathcal{I}(R[d_1/c_1, d_2/c_2], (D \cup \{d_1, d_2\}) \setminus \{c_1, c_2\}) \\ & \qquad (A[d_1/c_1, d_2/c_2]) \\ = \ & \mathcal{I}(R[d_1/c_1, d_2/c_2][c_2/d_1, c_1/d_2], \\ & (((D \cup \{d_1, d_2\}) \setminus \{c_1, c_2\}) \cup \{c_1, c_2\}) \setminus \{d_1, d_2\}) \\ & (A[d_1/c_1, d_2/c_2][c_2/d_1, c_1/d_2]) \quad . \end{aligned}$$

Due to

$$\begin{aligned} R[d_1/c_1, d_2/c_2][c_2/d_1, c_1/d_2] = R[c_2/c_1, c_1/c_2] \quad &= \quad R \\ (((D \cup \{d_1, d_2\}) \setminus \{c_1, c_2\}) \cup \{c_1, c_2\}) \setminus \{d_1, d_2\} \quad &= \quad D \end{aligned}$$

and

$$A[d_1/c_1, d_2/c_2][c_2/d_1, c_1/d_2] \quad = \quad A[c_1/c_2, c_2/c_1]$$

this yields $\mathcal{I}(R, D)(A) = \mathcal{I}(R, D)(A[c_1/c_2, c_2/c_1])$. □

Proposition 4. *Let $\mathcal{I}$ satisfy (Prototypical Indifference). Let R be a knowledge base on $\mathcal{L}_D$.*

1. *Let G_1, G_2 be two ground sentences. For $c_1, c_2 \in D$ with $c_1 \equiv_R c_2$ it holds $\mathcal{I}(R, D)(G_2 \mid G_1) = \mathcal{I}(R, D)(G_2[c_1 \leftrightarrow c_2] \mid G_1[c_1 \leftrightarrow c_2])$.*
2. *Let $S \in \mathcal{S}_R$, $c_1, \ldots, c_n \in S$, and $\sigma : S \rightarrow S$ a permutation on S, i. e. a bijective function on S. Then it holds $\mathcal{I}(R, D)(A) = \mathcal{I}(R, D)(A[\sigma(c_1)/c_1, \ldots, \sigma(c_n)/c_n])$.*

Proof.

1. Because of (Prototypical Indifference) it is $\mathcal{I}(R, D)(G_1) = \mathcal{I}(R, D)\ (G_1[c_1 \leftrightarrow c_2])$ and $\mathcal{I}(R, D)(G_2 \wedge G_2) = \mathcal{I}(R, D)((A_1 \wedge A_2)[c_1 \leftrightarrow c_2])$ and hence
$$\begin{aligned} & \mathcal{I}(R, D)(G_2 \mid G_1) \\ = \ & \frac{\mathcal{I}(R, D)(G_2 \wedge G_1)}{\mathcal{I}(R, D)(G_1)} \\ = \ & \frac{\mathcal{I}(R, D)((G_2 \wedge G_1)[c_1 \leftrightarrow c_2])}{\mathcal{I}(R, D)(G_1[c_1 \leftrightarrow c_2])} \\ = \ & \mathcal{I}(R, D)(G_2[c_1 \leftrightarrow c_2] \mid G_1[c_1 \leftrightarrow c_2]) \end{aligned}$$
due to $(G_2 \wedge G_1)[x_i/y_i]_{i=1,\ldots,n} = G_2[x_i/y_i]_{i=1,\ldots,n} \wedge G_1[x_i/y_i]_{i=1,\ldots,n}$.

2. This follows from the fact that every permutation can be represented as a product of transpositions (Beachy and Blair 2005), i. e. permutations that exactly transpose two elements. Let $\sigma_1, \ldots, \sigma_m$ be these transpositions of σ and let $\sigma_{1\ldots i} = \sigma_i \circ \ldots \circ \sigma_1$ for $i = 1, \ldots, m$. Note, that $\sigma_{1\ldots 1} = \sigma_1$ and $\sigma_{1\ldots m} = \sigma$. Due to (Prototypical Indifference) it holds $\mathcal{I}(R,D)(A) = \mathcal{I}(R,D)(A[\sigma_1(c_1)/c_1, \ldots, \sigma_1(c_n)/c_n])$ and for any $i = 2, \ldots, m$ it is

$$\begin{aligned} & \mathcal{I}(R,D)(A[\sigma_{1\ldots i-1}(c_1)/c_1, \ldots, \sigma_{1\ldots i-1}(c_n)/c_n]) \\ = \; & \mathcal{I}(R,D)(A[\sigma_{1\ldots i}(c_1)/c_1, \ldots, \sigma_{1\ldots i}(c_n)/c_n]) \quad . \end{aligned}$$

Via transitivity and $\sigma_{1\ldots m} = \sigma$ it follows $\mathcal{I}(R,D)(A) = \mathcal{I}(R)(A[\sigma(c_1)/c_1, \ldots, \sigma(c_n)/c_n])$. □

We only give a proof sketch for Lemma 1.

Lemma 1. *Let R be a knowledge base on $\mathcal{L}_D$ such that $\mathcal{I}_\varnothing(R,D) \neq$ undef. Then for any D' with $D \subseteq D'$ it is $\mathcal{I}_\varnothing(R,D') \neq$ undef.*

Proof. (Sketch) Let $P_0 = \mathcal{I}_\varnothing(R,D)$. Consider the (ground) knowledge base R' that consists of all ground conditionals $(B' \mid A')[x]$ such that R contains a conditional $(B \mid A)[y]$, $(B' \mid A')[y]$ is an instantiation of $(B \mid A)[y]$ with respect to D, and $x = P_0(B' \mid A')$. So, R' contains all instances of conditionals in R with their actual probabilites in the ME-model of R. Then, clearly, $P_0 = \mathsf{ME}(R')$ in the propositional sense. Let now $C_0 \subseteq D$ be the set of constants that do not appear in R, let $d \notin D$ be a new constant, and let $D_1 = D \cup \{d\}$. The set C_0 is also an R-equivalence class and introducing d into the language of R obviously yields that $C_0 \cup \{d\}$ is an R-equivalence class under the new language. Let $S = \{(B_1 \mid A_1)[x_1], \ldots, (B_n \mid A_n)[x_n]\} \subseteq R'$ be a maximal subset of R', such that $(B_i \mid A_i)[x_i][c_1 \leftrightarrow c_2] = (B_j \mid A_j)[x_j]$ for any $1 \leq i < j \leq n$ and some $c_1, c_2 \in C_0$. This means that S contains all ground conditionals of equivalent structure that just differ in prototypical elements of C_0; in the following we call S an *equivalence set*. Due to (Prototypical Indifference) it follows $x_1 = \ldots = x_n$. For any such set S let $S_1 = \{(B_1 \mid A_1)[y], \ldots, (B_{n+m} \mid A_{n+m})[y]\}$ be the corresponding set with respect to $C_0 \cup \{d\}$ and set $y = n * x/(n+m)$. By extending D to D_1 this yields a balancing out of the probabilities that are distributed over the prototypical elements of C_0. Let now R'_1 be a knowledge base that is the same as R' except that equivalence set S is replaced by S_1. Then $P_1 = \mathsf{ME}(R'_1)$ is a distribution on $\mathcal{L}_{D_1}$ and with a similar argumentation as above it follows $P_1 = \mathcal{I}_\varnothing(R, D_1)$. This extends iteratively to arbitrary D' with $D \subseteq D'$. □

We illustrate the argumentation in the proof of Lemma 1 by means of a simple example.

Example 5. Let R be given by

$$\begin{aligned} r_1 &: (flies(X))[0.7] \\ r_2 &: (flies(tweety)[0.9] \end{aligned}$$

and consider the constants $D = \{tweety, huey, dewey\}$. There are two R-equivalence classes, $T_1 = \{huey, dewey\}$ and $T_2 = \{tweety\}$. It follows that $\mathcal{I}_\varnothing(R,D)(flies(huey)) = \mathcal{I}_\varnothing(R,D)(flies(dewey)) = 0.6$ and obviously $\mathcal{I}_\varnothing(R,D)(flies(tweety)) = 0.9$. By assigning these probabilities to the grounded versions of the conditionals we yield R' with

$$\begin{aligned} r_{1,1} &: (flies(huey))[0.6] \\ r_{1,2} &: (flies(dewey)[0.6] \\ r_{1,3} &: (flies(tweety))[0.9] \\ r_2 &: (flies(tweety)[0.9] \end{aligned}$$

Observe that $r_{1,3}$ and r_2 are equivalent and that $r_{1,3}$ is just added to illustrate the general approach of the argumentation in the proof of Lemma 1. Now let $D_1 = D \cup \{louie\}$, so T_1 becomes $T'_1 = T_1 \cup \{louie\}$ and T'_1 and T_2 are the R-equivalence classes with respect to D_1. Neglecting conditional $r_{1,3}$ we have two equivalence sets in R', namely $S^1 = \{r_{1,1}, r_{1,2}\}$ and $S^2 = \{r_2\}$. While $S^2 = S^2_1$ because S^2 does not mention any element in T_1, for S^1 we get S^1_1 with

$$\begin{aligned} r'_{1,1} &: (flies(huey))[y] \\ r'_{1,2} &: (flies(dewey)[y] \\ r'_{1,3} &: (flies(louie))[y] \end{aligned}$$

and $y = 0.6 * 2/3 = 0.4$. Let $R'_1 = S^1_1 \cup S^1_2$. It follows that $\mathsf{ME}(R'_1) \models^{cp}_\varnothing R$ with respect to D_1 and due to the strategy applied when modifying the probabilities it is $\mathsf{ME}(R'_1) = \mathcal{I}_\varnothing(R, D_1)$.

Proposition 5. *$\mathcal{I}_\varnothing$ satisfies (Name Irrelevance), (Prototypical Indifference), (ME-Compatibility), (Conditional Probability in the Limit), (Compensation), and (Strict Inference).*

Proof.

(Name Irrelevance) This is obvious as the principle of maximum entropy is unbiased to renaming of constants, cf. (Shore and Johnson 1980).

(Prototypical Indifference) This follows from Proposition 2.

(ME-Compatibility) Let R be a ground knowledge base. Due to Remark 2 the operator $\models^{cp}_\varnothing$ is equivalent to $\models$ in the propositional case. Then Equation (3) also becomes equivalent to the propositional case and is in particular uniquely solvable. Hence, it is $\mathsf{ME}(R')(A) = \mathcal{I}_\varnothing(R,D)(A)$ for any ground sentence A.

(Conditional Probability in the Limit) Let R be a knowledge base on $\mathcal{L}_D$ such that $P^* := \mathcal{I}_\varnothing(R,D) \neq$ undef. Let $r = (B(\vec{X}) \mid A(\vec{X}))[\alpha] \in R$ with $\vec{X} = (X_1, \ldots, X_h)$ and $c_1, \ldots, c_n$ the constants that appear in R. Let furthermore $\{d_1, \ldots, d_m\} = D \setminus \{c_1, \ldots, c_n\}$, so it is $|D| = n + m$. Let $\vec{d}_1, \ldots, \vec{d}_k$ be all vectors of constants of $d_1, \ldots, d_m$ with length h such that for any $\vec{d}_i$ with $1 \leq i \leq k$ no two elements are the same. Let $\vec{c}_1, \ldots, \vec{c}_l$ be all remaining vectors of constants in D. It follows that $(l+k) = (|D|)^h = (n+m)^h$ and $k = m^{\underline{h}} =$

$m(m-1)\dots(m-h+1)$ (the falling factorial) and thus $l = (n+m)^h - m^{\underline{h}}$. Let $P^*_{\vec{c}}$ denote $P^*(B(\vec{c}) \mid A(\vec{c}))$ for a vector $\vec{c}$. In order to have $P^* \models^{cp}_{\varnothing} r$ it must hold $P^*_{\vec{c}_1} + \dots + P^*_{\vec{c}_l} + P^*_{\vec{d}_1} + \dots + P^*_{\vec{d}_k} = \alpha \cdot (k+l)$. From (Prototypical Indifference) and Proposition 4 it follows that $P^*_{\vec{d}_1} = \dots = P^*_{\vec{d}_k}$. Define $P^*_k := P^*_{\vec{d}_1}$, so it is $P^*_{\vec{d}_1} + \dots + P^*_{\vec{d}_k} = kP^*_k$. It follows

$$\begin{aligned} P^*_k &= \frac{\alpha\cdot(k+l) - P^*_{\vec{c}_1} - \dots - P^*_{\vec{c}_l}}{k} \\ &\leq \frac{\alpha\cdot(k+l)}{k} \\ &= \alpha \underbrace{\frac{(n+m)^h}{m^{\underline{h}}}}_{\stackrel{m\to\infty}{\to} 1} \\ &\stackrel{m\to\infty}{\to} \alpha \end{aligned}$$

Similarly it holds

$$\begin{aligned} P^*_k &= \frac{\alpha\cdot(k+l) - P^*_{\vec{c}_1} - \dots - P^*_{\vec{c}_l}}{k} \\ &\geq \frac{\alpha\cdot(k+l) - l}{k} \\ &= \alpha \underbrace{\frac{(n+m)^h}{m^{\underline{h}}}}_{\stackrel{m\to\infty}{\to} 1} - \underbrace{\frac{(n+m)^h - m^{\underline{h}}}{m^{\underline{h}}}}_{\stackrel{m\to\infty}{\to} 0} \\ &\stackrel{m\to\infty}{\to} \alpha \quad . \end{aligned}$$

Due to Lemma 1 the probability distributions P^*_k are well-defined for any k and it follows $P^*_k \to \alpha$ for $m \to \infty$.

(Compensation) Let R be a knowledge base and $(B(\vec{X}) \mid A(\vec{X}))[\alpha] \in R$ a non-ground conditional with $\alpha \in (0,1)$. Suppose $\mathcal{I}_{\varnothing}(R,D)(B(\vec{c}) \mid A(\vec{c})) < \alpha$ for all $(B(\vec{c}) \mid A(\vec{c}))[\alpha] \in \mathsf{ground}_D(B(\vec{c}) \mid A(\vec{c}))$. Then (for finite D) it is

$$\begin{aligned} &\frac{\sum_{(B(\vec{c})|A(\vec{c}))\in\mathsf{ground}_D((B(\vec{X})|A(\vec{X})))} P(B(\vec{c}) \mid A(\vec{c}))}{|\mathsf{ground}_D(B(\vec{X}) \mid A(\vec{X}))|} \\ <\ &\frac{\alpha\cdot|\mathsf{ground}_D(B(\vec{X}) \mid A(\vec{X}))|}{|\mathsf{ground}_D(B(\vec{X}) \mid A(\vec{X}))|} = \alpha \end{aligned}$$

contradicting $\mathcal{I}_{\varnothing}(R,D) \models^{cp}_{\varnothing} R$.

(Strict Inference) Let R be a knowledge base and $(B(\vec{X}) \mid A(\vec{X}))[\alpha] \in R$ a non-ground conditional with $\alpha = 1$ (the case of $\alpha = 0$ can be shown analogously). Suppose $\mathcal{I}_{\varnothing}(R,D)(B(\vec{c}) \mid A(\vec{c})) < 1$ for some $(B(\vec{c}) \mid A(\vec{c}))[\alpha] \in \mathsf{ground}_D(B(\vec{c}) \mid A(\vec{c}))$. Then (for finite D) it is

$$\begin{aligned} &\frac{\sum_{(B(\vec{c})|A(\vec{c}))\in\mathsf{ground}_D((B(\vec{X})|A(\vec{X})))} P(B(\vec{c}) \mid A(\vec{c}))}{|\mathsf{ground}_D(B(\vec{X}) \mid A(\vec{X}))|} \\ <\ &\frac{|\mathsf{ground}_D(B(\vec{X}) \mid A(\vec{X}))|}{|\mathsf{ground}_D(B(\vec{X}) \mid A(\vec{X}))|} = 1 \end{aligned}$$

contradicting $\mathcal{I}_{\varnothing}(R,D) \models^{cp}_{\varnothing} R$. □

Proposition 6. *The probability distribution defined by (14) is uniquely determined if R is $\odot$-consistent.*

Proof. For any knowledge base R the set of probability distributions that satisfy R is a convex set due to Lemma 2 and the fact that the intersection of two convex sets is again a convex set. The entropy is a strict concave function and maximization of a strict concave function over a convex set has a unique solution (Boyd and Vandenberghe 2004). □

Proposition 7. *$\mathcal{I}_{\odot}$ satisfies (Well-Definedness), (Name Irrelevance), (Prototypical Indifference), (Compatibility), (Conditional Probability in the Limit), (Strict Inference), and (ME-Compensation).*

Proof.

(Well-Definedness) This is true due to Proposition 6.

(Name Irrelevance) This is obvious as the principle of maximum entropy is unbiased to renaming of constants, cf. (Shore and Johnson 1980).

(Prototypical Indifference) This follows directly from Proposition 2.

(ME-Compatibility) For ground conditional knowledge bases R, the semantics is the same as for the propositional case, so $ME(R) = \mathcal{I}_{\odot}(R,D)$.

(Conditional Probability in the Limit) Let R be a relational conditional knowledge base on $\langle Pred, D = C_R \cup C_0\rangle$, where C_R contains all constants occurring in R, and $C_0 = D \backslash C_R$. Let $r = (B(\vec{X}) \mid A(\vec{X}))[\alpha] \in R$ be a relational conditional in R with free variables, and let $r_g = (B(\vec{c}) \mid A(\vec{c}))[\alpha]$ be a proper instantiation of r with constants $\vec{c}$ from C_0. Let $D_n = C_R \cup C_0 \cup C_n$ with $C_n \subset C_{n+1}$ and $|C_0 \cup C_n| = n$, $n \in \mathbb{N}, n \geq |C_0|$, be a sequence of sets of constants. Let $P^*_n = \mathcal{I}_{\odot}(R, D_n)$ be the ME-distribution of R that takes the constants from D_n into account. Due to (Prototypical Indifference), any constant from C_0 can be replaced by any constant from C_n when calculating $P^*_n(B(\vec{a})|A(\vec{a}))$, since neither of them appears in R. So, all probabilities of instantiations $P^*_n(B(\vec{a})|A(\vec{a}))$ are given by instantiations over $C_R \cup C_0$, but we have to take proper multiplicities into regard. Let $(B(\vec{x})|A(\vec{x}))$ have arity s, and let $(B(\vec{a})|A(\vec{a}))$ be a proper instantiation. Then $\vec{a}$ is a vector of arity s that might have m components from $E_n = C_0 \cup C_n$, $0 \leq m \leq s$, and $s-m$ components from C_R. W.l.o.g., we assume $t = |C_R|, |C_0| \geq s$. Since the positions of these components can make a difference, we have $\binom{s}{m}$ non-indifferent instantiations $(B(\vec{a^m_{k_m,l_m}})|A(\vec{a^m_{k_m,l_m}})), 1 \leq k_m \leq \binom{s}{m}, 1 \leq l_m \leq t^{s-m}$, with vectors $a^m_{k_m,l_m}$ over $C_R \cup C_0$ such that $P^*_n(B(\vec{a^m_{k_m,l_m}})|A(\vec{a^m_{k_m,l_m}}))$ occurs n^m times among the instantiations over D_n. In particular, for $m = s$, all n^s instantiations over E_n are individually indifferent with respect to P^*_n, one of them being the instantiation for $\vec{c}$, so $P^*_n(B(\vec{c})|A(\vec{c}))$ can serve as a representative for these n^s probabilities. Similar statements hold for all instantiations of $P^*_n(A(\vec{x})B(\vec{x}))$ and $P^*_n(A(\vec{x}))$.

$P_n^* \models_{\odot}^{cp} R$, so in particular, $P_n^* \models_{\odot}^{cp} r$, which means that

$$\begin{aligned}\alpha &= \frac{\sum\limits_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D_n}((B(\vec{X})|A(\vec{X})))} P(A(\vec{a})B(\vec{a}))}{\sum\limits_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D_n}((B(\vec{X})|A(\vec{X})))} P(A(\vec{a}))} \\ &= \frac{\Sigma(A(\vec{a})B(\vec{a}), C_n)}{\Sigma(A(\vec{a}), C_n)} \quad .\end{aligned}$$

For the nominator, we obtain

$$\begin{aligned}&\sum(A(\vec{a})B(\vec{a}), D_n) \\ =\;& \sum_{l_0=1}^{t^s} P_n^*(A(\vec{a}_{l_0})B(\vec{a}_{l_0})) + \\ & n\sum_{k_1=1}^{s}\sum_{l_1=1}^{t^{s-1}} P_n^*(A(\vec{a}^1_{k_1,l_1})B(\vec{a}^1_{k_1,l_1})) + \\ & n^2\sum_{k_2=1}^{\binom{s}{2}}\sum_{l_2=1}^{t^{s-2}} P_n^*(A(\vec{a}^2_{k_2,l_2})B(\vec{a}^2_{k_2,l_2})) + \ldots + \\ & n^{s-1}\sum_{k_{s-1}=1}^{\binom{s}{s-1}}\sum_{l_{s-1}=1}^{t} \\ &\qquad P_n^*(A(\vec{a}^{s-1}_{k_{s-1},l_{s-1}})B(\vec{a}^{s-1}_{k_{s-1},l_{s-1}})) + \\ & +n^s P_n^*(A(\vec{c})B(\vec{c})) \\ =\;& n^s[\,\frac{1}{n^s}\sum_{l_0=1}^{t^s} P_n^*(A(\vec{a}_{l_0})B(\vec{a}_{l_0})) + \\ & \frac{1}{n^{s-1}}\sum_{k_1=1}^{s}\sum_{l_1=1}^{t^{s-1}} P_n^*(A(\vec{a}^1_{k_1,l_1})B(\vec{a}^1_{k_1,l_1})) + \\ & \frac{1}{n^{s-2}}\sum_{k_2=1}^{\binom{s}{2}}\sum_{l_2=1}^{t^{s-2}} P_n^*(A(\vec{a}^2_{k_2,l_2})B(\vec{a}^2_{k_2,l_2})) + \ldots + \\ & \frac{1}{n}\sum_{k_{s-1}=1}^{\binom{s}{s-1}}\sum_{l_{s-1}=1}^{t} \\ &\qquad P_n^*(A(\vec{a}^{s-1}_{k_{s-1},l_{s-1}})B(\vec{a}^{s-1}_{k_{s-1},l_{s-1}})) + \\ & P_n^*(A(\vec{c})B(\vec{c}))\,] = n^s[\,\epsilon_1(n) + P_n^*(A(\vec{c})B(\vec{c}))\,]\end{aligned}$$

with $\epsilon_1(n) \leq \frac{1}{n^s}t^s + \frac{1}{n^{s-1}}st^{s-1} + \frac{1}{n^{s-2}}\binom{s}{2}t^{s-2} + \ldots + \frac{1}{n}st = O(\frac{1}{n})$. In the same way, for the denominator, we have $\Sigma(A(\vec{a}), C_n) = n^s[\epsilon_2(n) + P_n^*(A(\vec{c}))]$ with $\epsilon_2(n) = O(\frac{1}{n})$. This shows that for $n \to \infty$, $P_n^*(A(\vec{c})B(\vec{c}))$ and $P_n^*(A(\vec{c}))$ are the dominant terms, hence $\alpha = \lim_{n\to\infty} \frac{P_n^*(A(\vec{c})B(\vec{c}))}{P_n^*(A(\vec{c}))} = \lim_{n\to\infty} P_n^*(B(\vec{c})|A(\vec{c}))$, what was to be shown.

(Compensation) Let R be a knowledge base and $(B(\vec{X})\,|\,A(\vec{X}))[\alpha] \in R$ a population-based statement with $\alpha \in (0,1)$, and let $\vec{c_1}$ be a vector of constants such that $\mathcal{I}_{\odot}(R,D)(B(\vec{c_1}) \mid A(\vec{c_1})) < \alpha$. Let $P^* = \mathcal{I}_{\odot}(R)$. From the presupposition $(B(\vec{X})\,|\,A(\vec{X}))[\alpha] \in R$ and $P^* \models_{\odot}^{cp} R$, in particular, we have $P^* \models_{\odot}^{cp} (B(\vec{X})\mid A(\vec{X}))[\alpha]$, which means

$$\alpha = \frac{\sum\limits_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X})))} P^*(A(\vec{a})B(\vec{a}))}{\sum\limits_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X})))} P^*(A(\vec{a}))}.$$

Assume that for all (proper) instantiations $\vec{a} \neq \vec{c_1}$, $P^*(B(\vec{a}) \mid A(\vec{a})) \leq \alpha$. Then we had

$$\begin{aligned}&\sum_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X})))} P^*(A(\vec{a})B(\vec{a})) \\ =\;& P^*(A(\vec{c_1})B(\vec{c_1})) + \\ &\sum_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X}))),\vec{a}\neq\vec{c_1}} P^*(A(\vec{a})B(\vec{a})) \\ <\;& \alpha P^*(A(\vec{c_1})) + \\ &\sum_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X}))),\vec{a}\neq\vec{c_1}} \alpha P^*(A(\vec{a})) \\ =\;& \alpha \sum_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X})))} P^*(A(\vec{a})),\end{aligned}$$

hence

$$\frac{\sum\limits_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X})))} P^*(A(\vec{a})B(\vec{a}))}{\sum\limits_{(B(\vec{a})|A(\vec{a}))\in \mathsf{ground}_{D}((B(\vec{X})|A(\vec{X})))} P^*(A(\vec{a}))} < \alpha,$$

which contradicts $P^* \models_{\odot}^{cp} (B(\vec{X})\,|\,A(\vec{X}))[\alpha]$. So, there must be another vector of constants $\vec{c_2}$ with $P^*(A(\vec{c_2}) \mid B(\vec{c_2})) > \alpha$.

(Strict Inference) Let R be a $\circ$-consistent knowledge base and $(B(\vec{X})\,|\,A(\vec{X}))[\alpha] \in R$ a non-ground conditional with $\alpha \in \{0,1\}$. Let $(B(\vec{c}) \mid A(\vec{c})) \in \mathsf{ground}_D(A(\vec{X}) \mid B(\vec{X}))$. It is to be shown that $\mathcal{I}(R,D)(B(\vec{c}) \mid A(\vec{c})) = \alpha$.

Suppose that $\alpha = 0$. Since $P^* = \mathcal{I}_{\odot}(R,D)$ is a model of R, in particular, we have $P^* \models_{\odot}^{cp} (B(\vec{X})\,|\,A(\vec{X}))[0]$. This implies that $\sum_{(B((\vec{c})|A(\vec{c}))\in \mathsf{ground}_D((B(\vec{X})|A(\vec{X})))} P(B(\vec{c}) \wedge A(\vec{c})) = 0$, so for all $(B(\vec{c}) \mid A(\vec{c})) \in \mathsf{ground}_D((B(\vec{X}) \mid A(\vec{X})))$, $P(B(\vec{c}) \wedge A(\vec{c})) = 0$.

In case that $\alpha = 1$, $P^* \models_{\odot}^{cp} (B(\vec{X})\,|\,A(\vec{X}))[1]$ yields $\sum_{(B(\vec{c})|A(\vec{c}))\in \mathsf{ground}_D((B(\vec{X})|A(\vec{X})))} P(B(\vec{c}) \wedge A(\vec{c})) = \sum_{(B(\vec{c})|A(\vec{c}))\in \mathsf{ground}_D((B(\vec{X})|A(\vec{X})))} P(A(\vec{c}))$. If there were $\vec{c}$ such that $P(B(\vec{c})| \wedge A(\vec{c})) < 1$, i.e. $P(B(\vec{c}) \wedge A(\vec{c})) < P(A(\vec{c}))$, this would result in $\sum_{(B(\vec{c})|A(\vec{c}))\in \mathsf{ground}_D((B(\vec{X})|A(\vec{X})))} P(B(\vec{c}) \wedge A(\vec{c})) < \sum_{(B(\vec{c})|A(\vec{c}))\in \mathsf{ground}_D((B(\vec{X})|A(\vec{X})))} P(A(\vec{c}))$, a contradiction. Hence, (Strict Inference) is also satisfied in this case.

$\square$