Qualitative Probabilistic Inference with Default Inheritance

Paul Thorn[*,a], Christian Eichhorn[†,b], Gabriele Kern-Isberner[†], and Gerhard Schurz[*]

[*]: Institute for Philosophy, Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany
[†]: Department of Computer Science, Technische Universität Dortmund, Dortmund, Germany
[a]:thorn@phil.hhu.de, [b]:christian.eichhorn@tu-dortmund.de

Abstract. There are numerous formal systems that allow inference of new conditionals based on a conditional knowledge base. Many of these systems have been analysed theoretically and some have been tested against human reasoning in psychological studies, but experiments evaluating the performance of such systems are rare. In this article, we extend the experiments in [19] in order to evaluate the inferential properties of c-representations in comparison to the well-known Systems P and Z. Since it is known that System Z and c-representations mainly differ in the sorts of inheritance inferences they allow, we discuss subclass inheritance and present experimental data for this type of inference in particular.

1 Introduction

There are systems of conditional reasoning (such as Adams' System P [2]) that can be used to make *valid* (i.e., truth preserving) inferences about conditional probabilities. More generally, there are systems of conditional reasoning where it is plausible to adopt a *probabilistic interpretation* of conditionals, where conditionals of the form $(\psi|\phi)$ are *interpreted* as expressing that the corresponding conditional probability, $P(\psi|\phi)$, is high. In some cases, it may be plausible to adopt a *probabilistic interpretation* of conditionals, for a given system, even when the inferences licensed by the respective system are ampliative, and not truth preserving, given the probabilistic interpretation. For example, although inheritance inference (i.e., from $(\psi|\phi)$ infer $(\psi|\phi\wedge\chi)$) may fail to preserve high probability in many cases, inheritance inference is a reasonable form of inference that one might like to codify within a system of conditional reasoning.

In the present paper, we compare and evaluate the behaviour of two systems of conditional reasoning that are stronger than System P, but admit of a probabilistic interpretation, namely: System Z [16], and System MinC (which we define based on the inductive method of c-representations [8,9]). The two systems are of interest, since they both license a number of desirable inference patterns, such as inheritance inference and contraposition, that are not licensed by System P. Nevertheless the two systems differ in some important respects, such as in their treatment of inheritance reasoning.

Within a system where conditionals are treated as expressing defaults, it is desirable that *subclass inheritance* among defaults be licensed, defeasibly. For example, from the default that birds usually can fly we would like to infer that crows (a subclass of birds) are usually capable of flight, in the case where we have no background knowledge indicating that crows are exceptional birds. Such inferences are assumed to be defeasible, meaning that there are conditions under which such inferences are defeated (i.e., conditions under which the inference is not licensed).

Beyond defeasible subclass inheritance, it is controversial whether *inheritance in the case of exceptional subclasses* should be licensed, defeasibly [6]. For example, notice that penguins are exceptional birds inasmuch as they lack the capacity of flight. Given that penguins represent an exceptional subclass of the class of birds, it is controversial whether the subclass, penguins, should inherit other characteristics typical of birds. For example, assuming that birds usually have wings, it is controversial whether it is reasonable to infer that penguins usually have wings, given that they are (usually) incapable of flight.

A principal difference between System Z and System MinC is that the latter, and not the former, permits inheritance inference in the case of exceptional subclasses. Prima Facie, this fact speaks in favor of System MinC, a point which we briefly discuss in Section 5. However, as our primary means of evaluation, our paper reports the results of experiments which test the behaviour of Systems Z and MinC in reasoning about a simulated stochastic environment. For additional perspective, we also tested the behavior of System P and System QC [19,23]. The results show that while System MinC makes *many* inferences that are not drawn by System Z, System Z *rarely* makes an inference that is not drawn by System MinC. Since the two systems are both ampliative with respect to the *probabilistic interpretation* of conditionals (in contrast to System P), it is clear that the conclusions drawn by Systems Z and MinC are more risky than the ones drawn by System P. It is also plausible to think that conclusions that are drawn by System MinC and not System Z are more *risky* than the conclusions that are drawn by both systems, since such conclusions go “farther out on a limb”. The results presented here vindicate this thought, and provide a clearer picture of just how risky these inferences are.

The paper is organised as follows: After introducing the necessary formal preliminaries in Section 2, we introduce Systems P, Z, and QC in Section 3. We define System MinC via c-representations in Section 4. In Section 5 we discuss subclass inheritance for exceptional subclasses. We present the experimental setup and the results of the experiments in Sections 6 and 7, and conclude in Section 8.

2 Preliminaries

Let $\Sigma = \{V_1, ..., V_m\}$ be a propositional alphabet where a *literal* is a variable V interpreted to $true$ (v) or $false$ ($\overline{v}$). From these we obtain the propositional language $\mathfrak{L}$ as the set of formulas of Σ closed under negation $\neg$, conjunction $\wedge$, and disjunction $\vee$, as usual; for shorter formulas, we abbreviate conjunction by juxtaposition (i.e., ab is equivalent to $a \wedge b$), and negation by overlining (i.e., $\overline{a}$ is equivalent to $\neg a$). We write the material implication as $\phi \to \psi$ which is, as usual, equivalent to $\overline{\phi} \vee \psi$. *Interpretations* or *possible worlds* are also defined in the usual way; the set of all possible worlds is denoted by Ω. We often take advantage of the 1-1 association between worlds and *complete conjunctions*, i.e., conjunctions of literals where every $V_i \in \Sigma$ appears exactly once.

Table 1. Evaluation of conditionals in the penguin example (+ indicates verification, − falsification, an empty cell inapplicability) (above). Two OCF for the penguin example (below).

	$pbfw$	$pbf\overline{w}$	$pb\overline{f}w$	$pb\overline{f}\overline{w}$	$p\overline{b}fw$	$p\overline{b}f\overline{w}$	$p\overline{b}\overline{f}w$	$p\overline{b}\overline{f}\overline{w}$	$\overline{p}bfw$	$\overline{p}bf\overline{w}$	$\overline{p}b\overline{f}w$	$\overline{p}b\overline{f}\overline{w}$	$\overline{p}\overline{b}fw$	$\overline{p}\overline{b}f\overline{w}$	$\overline{p}\overline{b}\overline{f}w$	$\overline{p}\overline{b}\overline{f}\overline{w}$
$(f\|b)$	+	+	−	−					+	+	−	−				
$(\overline{f}\|p)$	−	−	+	+	−	−	+	+								
$(b\|p)$	+	+	+	+	−	−	−	−								
$(w\|b)$	+	−	+	−					+	−	+	−				
$\kappa^Z_\Delta(\omega)$	2	2	1	1	2	2	2	2	0	1	1	1	0	0	0	0
$\kappa^{c'}_\Delta(\omega)$	2	3	1	2	4	4	2	2	0	1	1	2	0	0	0	0

A *conditional* $(\psi|\phi)$, $\phi, \psi \in \mathfrak{L}$, is trivalent, with the evaluation: $(\psi|\phi)$ is *verified* iff $\omega \models \phi\psi$, $(\psi|\phi)$ is *falsified* iff $\omega \models \phi\overline{\psi}$, and $(\psi|\phi)$ is *inapplicable* iff $\omega \models \overline{\phi}$ [5,8]. A finite set of conditionals $\Delta = \{(\psi_1|\phi_1), \ldots, (\psi_n|\phi_n)\}$ is called a *knowledge base*.

An *Ordinal Conditional Function* (OCF, ranking function [21,20]) is a function $\Omega \to \mathbb{N}_0 \cup \{\infty\}$ that assigns to each world an implausibility rank, such that $\kappa^{-1}(0)$, the preimage of 0, is non-empty. The rank of a formula $\psi \in \mathfrak{L}$ is the rank of the lowest ranked world that satisfies the formula, formally: $\kappa(\phi) = \min_{\omega \models \phi}\{\kappa(\phi)\}$. The rank of a conditional $(\psi|\phi)$ is defined as: $\kappa(\psi|\phi) = \kappa(\phi\psi) - \kappa(\phi)$. A ranking function *accepts* a conditional $(\psi|\phi)$ (written $\kappa \models (\psi|\phi)$) iff $\kappa(\phi\psi) < \kappa(\phi\overline{\psi})$; κ is *admissible* with respect to a knowledge base Δ if and only if $\kappa \models (\psi|\phi)$ for all $(\psi|\phi) \in \Delta$.

Example 1. We illustrate these preliminaries with the well-known penguin example. Let B indicate whether something is a bird (b) or not ($\overline{b}$), let P indicate whether something is a penguin (p) or not ($\overline{p}$), let F indicate whether something is capable of flying (f) or not ($\overline{f}$), and let W indicate whether something has wings (w) or not ($\overline{w}$). This gives us the alphabet $\Sigma = \{P, B, F, W\}$ with a set of worlds given in the top row of Table 1. We use the conditionals "birds usually can fly" $(f|b)$, "penguins usually cannot fly" $(\overline{f}|p)$, "penguins usually are birds" $(b|p)$, and "birds usually have wings" $(w|b)$ to compose the knowledge base $\Delta = \{(f|b), (\overline{f}|p), (b|p), (w|b)\}$. Table 1 displays the evaluation of these conditionals within the worlds $\omega \in \Omega$. Table 1 also displays two ranking functions, κ^Z_Δ and $\kappa^{c'}_\Delta$ that are admissible with respect to this knowledge base. We discuss the two ranking functions, especially how they are generated inductively from the above knowledge base, later in the paper.

3 Overview of Systems P, Z, and QC

As described in [7], System P represents the confluence of a number of different semantic criteria. One feature of System P that is of interest here is its connection with the following consequence relation (cf. [2]):

Improbability-Sum Preservation: $(\psi_1|\phi_1), ..., (\psi_n|\phi_n) \models_{i.s.p.} (\xi|\chi)$ iff for all probability functions, P, over the appropriate language: $I(\xi|\chi) \leq \Sigma^n_{i=1} I(\psi_i|\phi_i)$, where $I(\psi|\phi)$, the *improbability* of ψ given ϕ, is defined as $1 - P(\psi|\phi)$.

As Adams [2] demonstrated, the following calculus (denoted by $\vdash_P$) is correct and complete for $\models_{i.s.p.}$:

(REF) (reflexivity) $\vdash_P (\psi|\phi)$
(LLE) (left logical equivalence) if $\models (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$, then $(\chi|\phi) \vdash_P (\chi|\psi)$
(RW) (right weakening) if $\models \phi \rightarrow \psi$, then $(\phi|\chi) \vdash_P (\psi|\chi)$
(CC) (cautious cut) $(\psi|\phi), (\chi|\psi \wedge \phi) \vdash_P (\chi|\phi)$
(CM) (cautious monotony) $(\psi|\phi), (\chi|\phi) \vdash_P (\chi|\psi \wedge \phi)$
(AND) $(\phi|\chi), (\psi|\chi) \vdash_P (\phi \wedge \psi|\chi)$
(OR) $(\chi|\phi), (\chi|\psi) \vdash_P (\chi|\psi \vee \phi)$

System P also has a semantics expressible in terms of ranking functions. In particular, $\Delta \vdash_P (\psi|\phi)$ if and only if every ranking function that is *admissible* for Δ *accepts* $(\psi|\phi)$ [1,13]. Apart from being characterized by reasonable (if mininal) principles, and plausible semantic theories, empirical studies show that human reasoning makes use of the principles of System P (c.f. [17,11]), which renders the study of System P especially worthwhile.

Inference by System Z [16] is based upon the unique ranking function κ^Z_Δ, among the admissible ranking functions for (consistent) $\Delta = \{(\psi_1|\phi_1), \ldots, (\psi_n|\phi_n)\}$, that minimizes the rank of each world in the set of possible worlds Ω_Δ defined over the propositional atoms appearing in Δ. This is achieved by forming an ordered partition $(\Delta_0, ..., \Delta_m)$ of Δ, where each Δ_i is the maximal subset of $\bigcup_{j=i}^m \Delta_j$ that is tolerated by $\bigcup_{j=i}^m \Delta_j$ (where a conditional, $(\psi|\phi)$, is *tolerated* by a set of conditionals, Δ, iff $\exists\omega$: $\omega \models \phi\psi$ and $\forall(\psi_i|\phi_i) \in \Delta$: $\omega \models \phi_i \rightarrow \psi_i$). Due to maximality, such partitions are unique for every Δ. Given the respective partition, κ^Z_Δ is defined as the OCF that assigns the value 0 to a world, ω, if no elment of Δ falsified at ω, and otherwise assigns the value $i + 1$, where i is index of the rightmost element of $(\Delta_0, ..., \Delta_m)$ that contains a conditional falsified by ω. Table 1 (above) presents κ^Z_Δ for the knowledge base described in Example 1. Inference by System Z is characterized by the relation $\vdash_Z$, which is defined in terms of the conditionals *accepted* by κ^Z_Δ:

$$\Delta \vdash_Z (\psi|\phi) \qquad \text{iff} \qquad \kappa^Z_\Delta \models (\psi|\phi). \qquad \text{(System Z)}$$

By adding the rule *Monotony*, i.e., $(\psi|\phi)$ implies $(\psi|\phi \wedge \chi)$, to System Z (or merely to System P), we obtain System QC. We here follow [19], and implement System QC by reasoning with conditionals *as if* they were material implications, and define System QC as follows:

$$\Delta \vdash_{QC} (\psi|\phi) \qquad \text{iff} \qquad \{\ \phi_i \rightarrow \psi_i \mid (\psi_i|\phi_i) \in \Delta\ \} \models \phi \rightarrow \psi \qquad \text{(System QC)}$$

4 System MinC

System MinC is defined in terms of ranking functions known as c-representations [8,9]. A c-representation assigns an individual impact value $\kappa_i^- \in \mathbb{N}_0$ to each conditional $(\psi_i|\phi_i) \in \Delta$. Using these impact values, a ranking function, κ^c_Δ, is defined, where each world ω is assigned the rank $\kappa^c_\Delta(\omega)$, which is the sum of the impacts of the conditionals falsified by ω:

$$\kappa^c_\Delta(\omega) = \sum_{i:\omega \models \phi_i \overline{\psi}_i} \kappa_i^- . \qquad (1)$$

The impacts of the conditionals are chosen so that $\kappa_\Delta^c \models \Delta$, which is the case if

$$\kappa_i^- > \min_{\omega \models \psi_i \phi_i} \Big\{ \sum_{\substack{j:\omega \models \phi_j \overline{\psi}_j \\ i \neq j}} \kappa_j^- \Big\} - \min_{\omega \models \psi_i \overline{\phi}_i} \Big\{ \sum_{\substack{j:\omega \models \phi_j \overline{\psi}_j \\ i \neq j}} \kappa_j^- \Big\}. \quad (2)$$

Entailment with respect to a c-representation is defined, as usual, via the OCF κ_Δ^c.

$$\kappa_\Delta^c \models (\psi|\phi) \qquad \text{iff} \qquad \kappa_\Delta^c(\phi\psi) < \kappa_\Delta^c(\phi\overline{\psi}). \quad (3)$$

Example 2. We illustrate c-representations with the penguin example (Example 1). Table 1 shows the verification/falsification behaviour of the worlds and the conditionals in this example, where (2) gives us:

$$\begin{aligned}
\kappa_1^- &= \min\{\kappa_2^-, \kappa_2^- + \kappa_4^-, 0, \kappa_4^-\} - \min\{0, \kappa_4^-, 0, \kappa_4^-\} \\
\kappa_2^- &= \min\{\kappa_1^-, \kappa_1^- + \kappa_4^-, \kappa_3^-, \kappa_3^-\} - \min\{0, \kappa_4^-, \kappa_3^-, \kappa_3^-\} \\
\kappa_2^- &= \min\{\kappa_2^-, \kappa_2^- + \kappa_4^-, \kappa_1^-, \kappa_1^- + \kappa_4^-\} - \min\{\kappa_2^-, \kappa_2^-, 0, 0\} \\
\kappa_1^- &= \min\{\kappa_2^-, \kappa_1^-, 0, \kappa_1^-\} - \min\{\kappa_2^-, \kappa_1^-, 0, \kappa_1^-\}
\end{aligned}$$

This can be solved via the minimal solution $\kappa_1^- = 1$, $\kappa_2^- = 2$, $\kappa_3^- = 2$, $\kappa_4^- = 1$, which, with (1) gives us the c-representation $\kappa_\Delta^{c'}$ shown in Table 1.

The defining system (2) is a system of *inequalities*. The system defines a schema for all c-representations of a given knowledge base Δ rather than a unique ranking function for Δ. To apply the method of c-representations to define a system of conditional infererence, we introduced an algorythm for selecting a unique c-representation for each knowledge base. We call the resulting system "*MinC*" (minimal c-representation). Following the idea of System Z being the pareto-minimal ranking function admissible to a knowledge base Δ, we define System MinC via a *minimal* c-representation that assigns the smallest possible rank to each world. Since there are no straightforward criteria for identifying a *unique* minimal c-representation, we opted for the following hierarchy of criteria (cf. [15]):

(a) minimising the combined rank $\sum_{\omega \in \Omega} \kappa(\omega)$,
(b) minimising the maximal rank $\max_{\omega \in \Omega}\{\kappa(\omega)\}$,
(c) minimising the combined impacts $\sum_{i=1}^{n} \kappa_i^-$, and
(d) minimising the maximal impact $\max_{1 \leq i \leq n}\{\kappa_i^-\}$.

Most of the time, these criteria, in this order, select a minimal c-representation within one or two steps.

To determine our designated minimal c-representation, we order c-representations by (a), the ones indistinguishable by (a) are then ordered by (b), the ones indistinguishable by (b) are then ordered by (c), followed by (d). Since ordering by (a) through (d) does not always yield a unique minimal c-representation, we implemented a practical measure for identifying our designated c-representation as that c-representation having the lexicographically smallest vector $(\kappa_1^-, ..., \kappa_n^-)$ among the minimal solutions ordered by (a) though (d). To distinguish this uinque c-representation from the general κ_Δ^c, we call the c-representation that is chosen according to the preceding tests, for respective Δ, $\kappa^{c'}_{\Delta}$, and define a corresponding inference system as follows:

$$\Delta \vdash_{MinC} (\psi|\phi) \quad \text{iff} \quad \kappa^{c'}_{\Delta} \models (\psi|\phi). \qquad \text{(System MinC)}$$

Note that while their is no known axiomatic characterization of $\vdash_{MinC}$ or $\vdash_Z$, both satisfy all of the principles that characterize $\vdash_P$. In addition, both $\vdash_{MinC}$ and $\vdash_Z$ satisfy *rational monotony* [12]: from $(\psi|\phi)$ and the non-validity of $(\overline{\chi}|\phi)$ infer $(\psi|\phi\chi)$.

5 Exceptionality and subclass inheritance

A principal difference between System Z and System MinC is that the latter, and not the former, permits inheritance inference in the case of exceptional subclasses. This fact is illustrated by the ranking functions $\kappa^{Z}_{\Delta}(\omega)$ and $\kappa^{c'}_{\Delta}(\omega)$ of Table 1, concering Example 1. In this case, System minC permits the conclusion that $(w|p)$, whiles System Z does not. Prima Facie, this behavior speaks in favor of System MinC. Indeed, the range of possible inheritance inferences to exceptional subclasses is very broad – broader than generally recognized – and encompasses many inferences that are generally, and correctly, regarded as reasonable. As a consequence, it appears that abandoning inheritance inference to exceptional subclasses, as a default, would forsake too much, i.e., too many reasonable inferences. Systems that do abandon these inferences are described of having a *Drowning Problem* [4].

The fact that a prohibition of inheritance inference to exceptional subclasses would forsake too much can be seen by considering a range of typical inheritance inferences, where the relevant subclass represents a small proportion of the respective superclass. For example, suppose it is given that $(f|b)$ (birds are usually able to fly), and we would like to infer $(f|jb)$ (j-birds are usually able to fly). Assume that we possess no special information regarding the class j, save that j corresponds to a relatively small (or improbable) subclass of b. In that case, we are in a position to conclude that j is exceptional relative to b, since we are in a position to accept $(\overline{j}|b)$. But it is clear that the proposed inference should be permitted. Indeed, the proposed inference is no less reasonable than the most reasonable instances of inheritance reasoning. Moreover, the fact that j corresponds to a small subclass of b does not speak against the inference. The latter point is particularly important when we consider cases of classical direct inference, where inheritance reasoning is used in order to draw a conclusion about a particular individual (see [18,3]).

6 Experiments

We here extend the experiments conducted in [19], with the aim of evaluating the performance of System MinC in comparison to System Z. To make the search space manageable, we restricted the experiments to an alphabet $\Sigma = \{A, B, C, D\}$ with a language $\mathfrak{L}^{\wedge}$ restricted to conjunctions of literals. The language of conditionals $(\mathfrak{L}^{\wedge}|\mathfrak{L}^{\wedge})$ is further restricted so that no variable may appear in both the antecedent and the consequent of a conditional. This means that $(b|a)$ and $(cd|a\bar{b})$ are in $(\mathfrak{L}^\wedge|\mathfrak{L}^\wedge)$, but $(bcd|ab)$ is not.

To generate a stochastic environment, we randomly assigned values from the real-valued interval $[0,1]$ to the probabilities: $(a|\top)$,$(b|\dot{a})$, $(c|\dot{a}\dot{b})$, and $(d|\dot{a}\dot{b}\dot{c})$ with $\dot{a} \in \{a, \overline{a}\}$, $\dot{b} \in \{b, \overline{b}\}$, and $\dot{c} \in \{c, \overline{c}\}$. We then generated the probability distribution $P : \Omega \to [0,1]$ by the so called "chain rule". Based on this distribution, ***four*** conditionals $(\psi|\phi)$ with $P(\psi|\phi) \geq mp$ (the ***minimum probability*** of the conditionals in the knowledge base for the respective simulation) were chosen randomly from $(\mathfrak{L}^\wedge|\mathfrak{L}^\wedge)$. Given this knowledge base Δ, the sets of all entailed conditionals $\mathcal{C}^X(\Delta) = \{(\psi|\phi)|(\psi|\phi) \in (\mathfrak{L}^\wedge|\mathfrak{L}^\wedge), \Delta \vdash_X (\psi|\phi)\}$, for $X \in \{\mathrm{P}, \mathrm{Z}, \mathrm{MinC}, \mathrm{QC}\}$, were computed. The restriction of our simulations to cases where the systems are provided with four premise conditionals expressed within $(\mathfrak{L}^\wedge|\mathfrak{L}^\wedge)$ partly limits the scope of our results. For some explanation concerning why these limitations are not so significant, see [24].

The accuracy of the inferences drawn by the four systems was assessed by treating the systems as asserting that the probability of the inferred conditional was at least the sum of the improbabilities of the premises upon which the inference was based. This amounts to treating the systems as licensing inference to inferred lower probability bounds. According to the present assumption, the precise bound licensed by a respective system, X, relative to a given knowledge base, Δ, and a probability function, P, is as follows, where Δ' ranges over the subsets of Δ such that $\Delta' \vdash_X (\psi|\phi)$:

$$X(\psi|\phi) = \max_{\Delta' \subseteq \Delta} \Big\{ 1 - \sum_{(\psi_i|\phi_i) \in \Delta'} (1 - P(\psi_i|\phi_i)) \Big\}. \quad (4)$$

While the present assumption is 'correct' in the case of System P, it may lead to overestimation when applied to the other three systems. Precisely, we say that inference made by a system counts as an *overestimation*, whenever $X(\psi|\phi) > P(\psi|\phi)$. For the moment, we will proceed as if it is reasonable to evaluate the accuracy of inferred conditionals in the present manner, bearing in mind that any charge of "overestimation" is based on the assumption that it is correct to propogate lower probability bounds in the manner of improbability sums. In the conclusion of the paper, when we consider what to make of our experminental results, we will briefly revisit this assumption.

Beyond attending to cases where a respective system overestimates respective conditional probabilities, our interest is in comparing the accuracy of the bounds licenced by the Systems Z and MinC. Unfortunately, there are no established and uncontroversial measures for scoring the accuracy of lower probability bounds. For this reason, we report the results of a scoring method that has a principled motivation and is pertinent to assessing accuracy, namely, the advantage-compared-to-guessing measure (ACG) [19]:

$$ACG(X(\psi|\phi), P) = \frac{1}{3} - |P(\psi|\phi) - X(\psi|\phi)|. \quad \text{(ACG)}$$

The idea behind this measure derives from the fact that the mean difference between two random choices of real values r and s from the unit interval is, provably, $\frac{1}{3}$. This means that the 'strategy' of setting lower probability bounds by randomly choosing numbers in $[0,1]$ is expected to yield an ACG score of zero, on average (assuming that the true probabilities are also selected randomly from $[0, 1]$). Reporting ACG scores, rather than the *linear distance* of inferred bounds from the true probabilities, has heuristic value, since the measure assigns *positive* scores to judgments that are 'better than guessing', and *negative* scores to judgments that are 'worse than guessing'. Given the appearance of X in the calculation of ACG scores, we once again observe that our proposed evaluation assumes that it is correct to propogate lower probability bounds in the manner of improbability sums.

7 Experimental results

The results presented here, regarding systems P, Z, and QC, are similar to those presented in [19]. The results of this paper are novel inasmuch they permit a comparison of the performance of Systems Z and MinC. All tested systems do satisfy certain quality criteria, as noted in Sections 3 and 4, and hence the inferences drawn are sensible with respect to those criteria.

Table 2 presents the number of inferences made by each of the four systems over the course of 5,000 simulations, for each of the listed values of mp (the minimum probability of the conditionals in the knowledge base). Table 2 illustrates that System MinC permits more inferences than System Z, while both systems permit quite a few more inferences than System P, and far fewer inferences than System QC. It may also be observed that the difference between the number of System MinC and System Z inferences decreases with increases in the value of mp. Indeed, if we exclude those inferences that are made by System P, then we see that System MinC licenses about 10% more inferences than System Z, when mp = 0.5. At mp = 0.99, System MinC licenses about 5% more inferences than System Z. At present, we cannot say whether the behavior of System Z and System MinC converge as mp goes to 1.

Every inference licensed by System P is included in each of the other systems. On the other hand, it has been demonstrated that the set of inferences licensed by a minimal c-representation does not generally include those licensed by System Z, and similarly the set of inferences licensed by System Z does not generally include those licensed by a minimal c-representation [10]. Our experiments expand upon this finding, showing that although there are inferences that are licensed by System Z that are not licensed by System MinC, such inferences are rare. Indeed, in addition to licensing more conclusions than System Z, the set of conclusions licensed by System MinC frequently includes the set of conclusions licensed by System Z, as presented in the right most column of Tbl 2.

Example 3. To show that System Z and System MinC are different in general we use an Example from [10]. By applying System Z and System MinC to the knowledge base $\Delta = \{(a|b), (\overline{a}|c), (b|c), (d|b)\}$, we obtain that $((\overline{c}b \vee c\overline{b}) \wedge \overline{a}\overline{d}) \vdash_Z \overline{c}b$ and $((\overline{c}b \vee c\overline{b}) \wedge \overline{a}\overline{d}) \nvdash_{MinC} \overline{c}b$, whereas $c\overline{b}d \vdash_{MinC} \overline{a}$ and $c\overline{b}d \nvdash_Z \overline{a}$.

Table 3 shows that both System Z and MinC are somewhat prone to overestimation, which characterizes the majority of System Z and MinC inferences when the value of mp is high. Table 3 also shows that the inferences made by System MinC tend to be less accurate than those of System Z, as measured by the ACG measure. This fact is partially obscured by the fact that the sets of inferences made by systems Z and MinC

Table 2. Total number of inferred conditionals.

mp	Number of inferred conditionals							Cases where
	P	Z	MinC	QC	$\|Z \cap$ MinC$\|$	$\|Z\setminus$MinC$\|$	$\|$MinC$\setminus$Z$\|$	Z$\subseteq$MinC
0.5	65 777	258 400	278 366	612 815	257 671	729	20 695	4 535
0.6	51 368	232 926	249 612	508 811	232 404	522	17 208	4 627
0.7	39 354	206 108	218 756	423 102	205 783	325	12 973	4 744
0.8	29 899	175 412	184 751	338 832	175 201	211	9 550	4 813
0.9	24 296	133 602	139 197	218 434	133 566	36	5 631	4 965
0.99	20 690	74 368	76 000	92 904	74 368	0	1 632	5 000

Table 3. Aggregate ACG scores and number of overestimations.

mp	Aggregate ACG scores				Overestimations		
	P	Z	MinC	QC	Z	MinC	QC
0.5	9 413.7	31 385.6	33 019.9	33 880.1	94 512	106 767	368 061
0.6	10 556.5	32 564.0	33 852.3	24 691.5	98 270	109 754	329 044
0.7	10 078.2	30 887.3	31 715.8	15 609.1	101 240	110 764	294 497
0.8	8 973.8	27 711.8	28 144.2	7 497.0	100 573	108 520	255 494
0.9	7 888.6	22 808.0	22 823.5	8 987.9	90 520	95 958	174 652
0.99	6 893.0	19 076.3	19 324.0	16 015.8	56 659	58 291	75 195

both include the inferences made by System P, and by the fact that the set of System Z inferences is 'practically' included in the set of System MinC inferences. In order to present a clearer picture, Table 4 presents the average ACG score earned for *individual* inferences made by System P, inferences made by System Z that were not made by System P (Z\P), inferences made by System MinC that were not made by System Z (MinC\Z), and inferences made by System QC that were not made by System MinC (QC\MinC). Here we see that inferences proper to System MinC (MinC\Z) earned positive ACG across all values of mp, as with the inferences proper to System Z (Z\P), and unlike the inferences proper to System QC (QC\MinC).

Finally, since one of our primary concerns was to assess the reasonableness of inheritance inference in the case of exceptional subclasses, we compared the accuracy of the inheritance inferences licensed by System Z (which only involve unexceptional subclasses) with the accuracy of the inheritance inferences licensed by System MinC (which may involve exceptional subclasses). As a means of assessment, we counted an inference to a conditional $(\psi|\phi)$ as inferred by inheritance from a given premise set if and only if (i) $(\psi|\phi)$ was neither a member of the premise set nor inferred from the premise set by System P, and (ii) some conditional $(\psi|\xi)$ was also inferred, where $\phi \models \xi$ and $\xi \not\models \phi$. Information regarding such inferences is recorded in Table 5. Here we see that inheritance inferences make up the majority of the inferences licensed by Systems Z and MinC that are not also licensed by System P, ranging from just over 75% of the total inferences, for mp = 0.5, to just over 50% of the total inferences, for mp = 0.99. It also follows from results of Table 5 that the accuracy of the inheritance inferences licensed by System Z, as measured the average ACG scores per inference, is nearly identical to

Table 4. Mean ACG scores per inference.

mp	P	Z\P	MinC\Z	QC\MinC
0.5	0.143	0.114	0.082	0.003
0.6	0.206	0.121	0.077	−0.035
0.7	0.256	0.125	0.066	−0.079
0.8	0.300	0.129	0.046	−0.134
0.9	0.325	0.136	0.003	−0.175
0.99	0.333	0.227	0.152	−0.196

Table 5. Number of inheritance inferences and their aggregate ACG scores.

mp	Number of inferences			Aggregate ACG scores		
	Z	MinC	QC	Z	MinC	QC
0.5	145 763	163 532	384 009	15 827.1	17 276.2	20 253.2
0.6	134 338	149 252	314 733	16 587.1	17 749.1	15 043.7
0.7	120 407	131 738	259 069	16 103.8	16 867.5	10 200.4
0.8	102 610	111 088	204 322	14 992.7	15 395.3	6 017.1
0.9	72 671	77 896	120 005	11 752.3	11 770.0	6 291.0
0.99	27 021	28 649	35 261	6 758.9	7 006.0	6 315.4

the accuracy of the non-inheritance inferences among Z\P. Similarly, the accuracy of the inheritance inferences licensed by System MinC, and not by System Z, is nearly identical to the accuracy of the non-inheritance inferences among MinC\Z.

In summary, the results of our simulations are as follows (where accuracy claims assume the correctness of probability propogation by improbability sums):

1. System MinC licensed significantly more inferences than System Z, with a decreasing margin proportional to the value of mp.
2. While neither System Z nor System MinC strictly includes the other (as shown in [10]), the set of System Z inferences was a subset of the set of System MinC inferences within a vast majority of our simulations.
3. The accuracy of inferences licensed by System MinC was somewhat less than the accuracy of inferences drawn by System Z. We also observed that the accuracy of System MinC inferences tended to decrease with increasing values of mp (excluding the case where $mp = 0.99$, whose exceptionality is discussed at length in [19, § 2.5]).
4. The accuracy of the inheritance inferences licensed by System Z was nearly identical to that of the other inferences licensed by System Z that were not licensed by System P. Similarly, the accuracy of the inheritance inferences licensed by System MinC was nearly identical to that of the other inferences licensed by System MinC that were not licensed by System Z.

8 Conclusion

Our results show that for practical purposes, System MinC represents a stronger system of inference than System Z. Our results also show that inference by System MinC (and inheritance for exceptional subclasses, as licensed by System MinC) is more risky than inference by System Z. These results accord with existing theoretical analyses of the systems studied here [12,14,10]. Indeed, as measured by the type of monotony that characterize the systems, we see that inferential strength increases as we proceed from System P to Systems Z and MinC, and finally to System QC: *cautious* monotony holds for System P, *rational* monotony holds for Systems Z and MinC, and "full" monotony holds for System QC. As measured by the type of subclass inheritance supported by the systems, we see that inferential strength increases as we proceed from System P to System Z to System MinC and finally to System QC: no inheritance inference is permitted in System P, inheritance inference in the case of unexceptional subclasses is permitted in System Z, defeasible inheritance for exceptional subclasses is permitted in System MinC, and unrestricted inheritance inference is permitted in System QC. Our experimental results show that increasing inferential strength, as described, comes at the risk of decreased accuracy. Assuming the risk associated with such inferential strength is too high in the case of System QC (as argued in [19,22,23]), the question remains of whether inference by System MinC should be favored over inference by System Z.

While inference by System MinC carries greater risk than inference by System Z, the same claim can be made in comparing inference by System Z to inference by System P. In the latter case, the riskiness of inference by System Z appears to be small enough, so that inference by System Z should be preferred to inference by System P (as argued in [19,22,23]), or better: One should perform the inferences licensed by System Z *in addition* to those licensed by System P. Assuming such arguments are cogent in the case of System Z, are similar arguments cogent in the case of System MinC? In other words, should one perform the inferences licensed by System MinC in addition to those licensed by System Z? While we grant that the risks (of overestimation and inaccurate judgment) are greater in the case of System MinC (in comparison to System Z), we also observe that inference by System MinC generally yields positive accuracy scores according to the ACG measure, in the case where probability propogation is determined by improbability sums.

In addition to evaluating the performance of System MinC, we were keen to evaluate the accuracy of inheritance inference in the case of exceptional subclasses. In Section 5, we offered conceptual reasons for rejecting a blanket prohibition of such inferences. Our argument there proceeded from the fact that the class of inheritance inferences to exceptional subclasses is very broad and encompasses many inferences that are generally, and correctly (we maintain), regarded as reasonable. Of course, we do not endorse the wholesale adoption of all inheritance inferences, which would be tantamount to reasoning in accordance with System QC. Our hope is rather that there is some systematic way to move beyond System Z, and a blanket prohibition of inheritance inference in the case of exceptional subclasses. Our motivation for evaluating the performance of System MinC experimentally was to determine whether inference by System MinC might serve as an appropriate means of moving beyond System Z. As things stand (and for the reasons adduced in the preceding paragraphs), we think that inference by System MinC represents a promising option.

Finally, it should be mentioned, once again, that the overestimations and accuracy scores attributed to the studied systems are premised on treating the systems as inferring lower probability bounds in accordance with (4), above. We think that application of (4) is reasonable, since such inferences are valid (i.e., guarenteed to be truth preserving) for System P, and the other systems represent incremental strengthenings of System P. Moreover, the fact that such inferences are invalid in the case of Systems Z, MinC, and QC, is not a decisive objection to the proposed application of (4), since these three systems all license inheritance inference, for which there is never a guarantee that high premise probability is preserved (i.e., $\forall \phi, \psi, \chi, r$: $(r < 1$ and $\phi \nvDash \chi) \Rightarrow (\exists P$: $P(\psi|\phi) = r$ and $P(\psi|\phi \wedge \chi) = 0)$. Nevertheless, while applying (4) yields a plausible means of evaluating the four systems, there are certainly possible alternatives. Exploring such alternatives is an object of present and future research.

Acknowledgment: We thank the anonymous referees for their valuable suggestions that helped us improve the paper. This work was supported by DFG-Grant KI1413/5-1 of Prof. Dr. Gabriele Kern-Isberner, and DFG-Grant SCHU1566/9-1 of Prof. Dr. Gerhard Schurz, both as part of the priority program "New Frameworks of Rationality" (SPP 1516).

References

1. Adams, E.: A note on comparing probabilistic and modal logics of conditionals. Theoria 43, 186–194 (1971)
2. Adams, E.: The Logic of Conditionals: An Application of Probability to Deductive Logic. Synthese Library, Springer (1975)
3. Bacchus, F.: Representing and reasoning with probabilistic knowledge. MIT Press, Cambridge, Massachusetts (1990)
4. Benferhat, S., Cayrol, C., Dubois, D., Lang, J., Prade, H.: Inconsistency management and prioritized syntax-based entailment. In: Proceedings of the International Joint Conferences on Artificial Intelligence. pp. 640–647 (1993)
5. De Finetti, B.: Theory of Probability, vol. 1,2. John Wiley and Sons, New York, USA (1974)
6. Geffner, H., Pearl, J.: Conditional entailment: Bridging two approaches to default reasoning. Artificial Intelligence 53(2–3), 209–244 (1992)
7. Hawthorne, J., Makinson, D.: The quantitative/qualitative watershed for rules of uncertain inference. Studia Logica 86(2), 247–297 (2007)
8. Kern-Isberner, G.: Conditionals in Nonmonotonic Reasoning and Belief Revision – Considering Conditionals as Agents. No. 2087 in LNCS, Springer (2001)
9. Kern-Isberner, G.: A thorough axiomatization of a principle of conditional preservation in belief revision. Annals of Mathematics and Artificial Intelligence 40, 127–164 (2004)
10. Kern-Isberner, G., Eichhorn, C.: Structural inference from conditional knowledge bases. In: Unterhuber, M., Schurz, G. (eds.) Studia Logica Special Issue Logic and Probability: Reasoning in Uncertain Environments, pp. 751–769. No. 102 (4), Springer Science+Business Media, Dordrecht, NL (August 2014)
11. Kuhnmünch, G., Ragni, M.: Can Formal Non-monotonic Systems Properly Describe Human Reasoning? In: Proceedings of the Cognitive Science Conference (COGSCI2014). pp. 1806–1811 (2014)
12. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial Intelligence 55(1), 1–60 (1992)
13. Lewis, D.: Counterfactuals. Blackwell, Oxford, UK (1973)

14. Makinson, D.: General patterns in nonmonotonic reasoning. In: Gabbay, D.M., Hogger, C.J., Robinson, J.A. (eds.) Handbook of logic in artificial intelligence and logic programming, vol. 3, pp. 35–110. Oxford University Press, Inc., New York, NY, USA (1994)
15. Müller, C.: Optimierte Wissensdarstellung auf der Basis von Default-Regeln. Masters thesis, FernUniversität Hagen, Hagen, DE (2003), (in German)
16. Pearl, J.: System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In: Proceedings of the 3rd conference on Theoretical aspects of reasoning about knowledge. pp. 121–135. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1990)
17. Pfeifer, N., Kleiter, G.D.: Coherence and Nonmonotonicity in Human Reasoning. Synthese 146(1–2), 93–109 (2005)
18. Pollock, J.L.: Nomic Probability and the Foundations of Induction. Oxford University Press (1990)
19. Schurz, G., Thorn, P.D.: Reward versus risk in uncertain inference: Theorems and simulations. The Review of Symbolic Logic 5, 574–612 (11 2012)
20. Spohn, W.: The Laws of Belief: Ranking Theory and Its Philosophical Applications. Oxford University Press (2012)
21. Spohn, W.: Ordinal Conditional Functions: A Dynamic Theory of Epistemic States. In: Causation in Decision, Belief Change and Statistics: Proceedings of the Irvine Conference on Probability and Causation. The Western Ontario Series in Philosophy of Science, vol. 42, pp. 105–134. Springer Science+Business Media, Dordrecht, NL (1988)
22. Thorn, P.D., Schurz, G.: Ampliative Inference Under Varied Entropy Levels. In: Beierle, C., Kern-Isberner, G. (eds.) Proceedings of the 4th Workshop on Dynamics of Knowledge and Belief (DBK-2013). pp. 46–60. FernUniversität in Hagen (2013)
23. Thorn, P.D., Schurz, G.: A Utility Based Evaluation of Logico-probabilistic Systems. Studia Logica 102(4), 867–890 (2014)
24. Thorn, P.D., Schurz, G.: Qualitative probabilistic inference under varied entropy levels. Journal of Applied Logic (forthcoming)