

# Ampliative Inference Under Varied Entropy Levels

Paul D. Thorn<sup>1</sup>, and Gerhard Schurz<sup>1</sup>

<sup>1</sup> Heinrich-Heine-University, Institute for Philosophy, Universitaetsstr. 1, 40204  
Duesseldorf, Germany  
{thorn, schurz}@phil-fak.uni-duesseldorf.de

**Abstract.** Systems of *logico-probabilistic* (LP) reasoning characterize inference from conditional assertions that are interpreted as expressing high conditional probabilities. In previous work, we studied four well known LP systems (namely, systems O, P, Z, and QC), and presented data from computer simulations in an attempt to illustrate the performance of the four systems. These simulations evaluated the four systems in terms of their tendency to license inference to accurate and informative lower probability bounds, given incomplete information about a randomly selected probability distribution (where this probability distribution may be understood as representing the true stochastic state of the world). In our earlier work, the procedure used in generating the unknown probability distribution (i.e., the true stochastic state of the world) tended to yield probability distributions with moderately high entropy levels. In the present article, we present data charting the performance of the four systems in reasoning about probability distributions with various entropy levels. The results allow for a more inclusive assessment of the reliability and robustness of the four LP systems.

**Keywords:** ampliative inference, default reasoning, non-monotonic reasoning, probability logic.

## 1 LP Reasoning: Systems O, P, Z, and QC

We represent the four LP systems considered here (**O**, **P**, **Z**, and **QC**; described below) using a simple propositional language  $L$ , with the usual connectives  $\neg$ ,  $\wedge$ ,  $\vee$ , and  $\supset$ , and  $A$ ,  $B$ ,  $C$ , etc. as meta-logical variables ranging over arbitrary sentences of  $L$ . Our main interest will be in extensions of  $L$  by means of a default (or uncertain) conditional operator:  $\Rightarrow$ . In particular, we will concern ourselves with extensions of  $L$  by *simple* uncertain conditionals of the form  $A \Rightarrow B$ . Throughout the paper,  $\alpha$  and  $\beta$  will serve as meta-variables ranging over such simple conditional formulas, while  $\Gamma$  ranges over sets of them. “ $\vdash$ ” is used to denote derivability in classical logic, and “ $\perp$ ” to denote an arbitrary contradiction. The four LP systems that we consider are ordered in terms of the number of inferences they license ( $\mathbf{O} \subset \mathbf{P} \subset \mathbf{Z} \subset \mathbf{QC}$ ). We proceed by considering the weakest system first.

## 1.1 System **O**

System **O** is of interest because of its close connection to the following consequence relation:

(1) *Strict Preservation*:  $A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n \Vdash_{s.p.} C \Rightarrow D$  iff for all probability functions  $P$  (over  $L$ ):  $P(D|C) \geq \min(\{P(B_i|A_i) : 1 \leq i \leq n\})$ .

System **O** was developed by Hawthorne [1] and Hawthorne and Makinson [2] as an inferential calculus for  $\Vdash_{s.p.}$ . Throughout the present article, “ $\vdash_{\mathbf{O}}$ ” denotes the syntactical notion of derivability in system **O**.

**System O** (after Hawthorne):

REF (reflexivity):  $\vdash_{\mathbf{O}} A \Rightarrow A$

LLE (left logical equivalence): if  $\vdash (A \supset B) \wedge (B \supset A)$ , then  $A \Rightarrow C \vdash_{\mathbf{O}} B \Rightarrow C$

RW (right weakening): if  $\vdash B \supset C$ , then  $A \Rightarrow B \vdash_{\mathbf{O}} A \Rightarrow C$

VCM (very cautious monotony):  $A \Rightarrow B \wedge C \vdash_{\mathbf{O}} A \wedge B \Rightarrow C$

XOR (exclusive Or): if  $\vdash \neg(A \wedge B)$ , then  $A \Rightarrow C, B \Rightarrow C \vdash_{\mathbf{O}} A \vee B \Rightarrow C$

WAND (weak And):  $A \Rightarrow B, A \wedge \neg C \Rightarrow \perp \vdash_{\mathbf{O}} A \Rightarrow B \wedge C$

It is easy to see that all of the rules of system **O** are correct with respect to  $\Vdash_{s.p.}$ , i.e.,  $\Gamma \vdash_{\mathbf{O}} A \Rightarrow B$  implies  $\Gamma \Vdash_{s.p.} A \Rightarrow B$ . It was the hope of Hawthorne and Makinson [2] that  $\vdash_{\mathbf{O}}$  was also complete with respect to  $\Vdash_{s.p.}$ , but Paris and Simmonds [3] have shown that this not the case.

Following [5], we propose a marriage of system **O**, and a rule for inferring lower probability bounds that corresponds to the correctness of system **O** for  $\Vdash_{s.p.}$ . To make sense of such inferences, we employ statements of the form “ $A \Rightarrow_r B$ ” to express that  $P(B|A) \geq r$ , and say that system **O** licenses the (valid) inference to  $C \Rightarrow_{\min(\{r_i : 1 \leq i \leq n\})} D$  from  $A_1 \Rightarrow_{r_1} B_1, \dots, A_n \Rightarrow_{r_n} B_n$ , in cases where  $A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n \vdash_{\mathbf{O}} C \Rightarrow D$ .

A remarkable fact about **O** is its weakness compared to standard systems of conditional logic. According to Segerberg [4], the weakest ‘reasonable’ system of conditional logic includes REF, LLE, and RW, along with the following rule:

(AND): from  $A \Rightarrow B$  and  $A \Rightarrow C$  infer  $A \Rightarrow B \wedge C$ .

The inferential power of AND is quite significant. By adding AND to the system **O**, we obtain (in one step) the well known system **P**. In comparison with WAND, we note that  $A \Rightarrow C$  is derivable from  $A \wedge \neg C \Rightarrow \perp$ , given RW, REF, and XOR, while

$A \wedge \neg C \Rightarrow \perp$  is not derivable from  $A \Rightarrow C$ , given these rules. It is in this respect that WAND is weaker than the rule AND.<sup>1</sup>

## 1.2 System P

As described in [5], system **P** represents the confluence of a number of different semantic criteria. But the feature of system **P** that is of greatest interest here is its connection with the following consequence relation (cf. [6]):

(2) *Improbability-Sum Preservation*:  $A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n \Vdash_{\text{i.s.p.}} C \Rightarrow D$  iff for all probability functions over  $L$ :  $I(D|C) \leq \Sigma\{I(B_i|A_i) : 1 \leq i \leq n\}$ , where  $I(A|B)$  is defined as  $1 - P(A|B)$ .

Adams demonstrated that the following calculus (denoted by  $\vdash_{\mathbf{P}}$ ) is correct and complete for  $\Vdash_{\text{i.s.p.}}$ .

**System P** (after Adams):

REF }  
LLE } as with system **O**  
RW }

AND: as above

CC (cautious cut):  $A \Rightarrow B, A \wedge B \Rightarrow C \vdash_{\mathbf{P}} A \Rightarrow C$

CM (cautious monotony):  $A \Rightarrow B, A \Rightarrow C \vdash_{\mathbf{P}} A \wedge B \Rightarrow C$

OR:  $A \Rightarrow C, B \Rightarrow C \vdash_{\mathbf{P}} A \wedge B \Rightarrow C$

Following [5], we propose a marriage of system **P**, and a rule for inferring lower probability bounds that corresponds to the correctness of system **P** for  $\Vdash_{\text{i.s.p.}}$ . In particular, we say that system **P** licenses the (valid) inference to  $C \Rightarrow_{1 - \Sigma\{1 - r_i; 1 \leq i \leq n\}} D$  from  $A_1 \Rightarrow_{r_1} B_1, \dots, A_n \Rightarrow_{r_n} B_n$ , in cases where  $A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n \vdash_{\mathbf{P}} C \Rightarrow D$ .

## 1.3 System Z

While system **P** sanctions more inferences than system **O**, it still sanctions fewer inferences than one might reasonably accept. For instance, **P** does not licence inference via *subclass inheritance* based on default assumptions of *irrelevance* (or *independence*). For example, if we know that this animal is a male bird ( $B \wedge M$ ) and that birds can normally fly ( $B \Rightarrow F$ ), and nothing else of relevance, then we would intuitively draw the conclusion that this male bird can fly ( $F$ ). However,  $B \wedge M \Rightarrow F$  is not **P**-

---

<sup>1</sup> We also observe that XOR is weaker than the rule OR (introduced below), and that VCM implies CM (below), in the presence of AND, while CM implies VCM (below), in the presence of RW ([2], 251).

entailed by  $B \Rightarrow F$ , because there are possible probability distributions in which  $P(F|B \wedge M)$  is much smaller than  $P(F|B)$ . If we do infer  $B \wedge M \Rightarrow F$  from  $B \Rightarrow F$ , in such cases, then we assume, *by default*, that the additional factor  $M$  (in this case the gender of a bird) is *irrelevant* to its ability to fly (or in other words,  $M$  and  $F$  are assumed to be probabilistically independent given  $B$ ). A straightforward means of enlarging the set of LP-derivable conditionals, in order to include such default inferences, is to give up the requirement that a reasonable inference be valid for *all possible* probability distributions, and consider only ‘normal’ probability distributions, i.e., those distributions which satisfy the default assumption of irrelevance. An early suggestion for realizing this idea was the *maximum entropy* approach to default inference (cf. [7]; [8], 491-3). By selecting a probability distribution that maximizes entropy, one minimizes probabilistic dependences. Despite having some attractive features, the maximum entropy approach is rather complicated, and has some further disadvantages, such as language dependence.

System  $\mathbf{Z}$  of Pearl [9] and Goldszmidt and Pearl [10] maintains many of the advantages of the maximum entropy approach, while overcoming its disadvantages. Like the maximum entropy approach, inference in system  $\mathbf{Z}$  proceeds via the construction of a semantic model of the premise conditionals that maximizes probabilistic independences. In system  $\mathbf{Z}$ , this is achieved by maximizing the *degree-of-normality* of the set of possible worlds represented by a ranked model, according to the following definition:

(3) *Definition* (cf. [10], 68, def. 15; [11], 308f): A ranked model  $(W, r)$  is *as least as normal* as a ranked model  $(W, r^*)$  (with the same world set), in short  $(W, r) \geq_N (W, r^*)$ , *iff* for all  $w \in W$ ,  $r(w) \leq r^*(w)$ .

As has been shown (cf. [5]; [9]), every set of worlds  $W$  (which is constructed over the language of the conditional knowledge base  $\Gamma$ ) has a unique *most normal* ranked model, the so called *z-model*. In order to define the notion of a *z-model*, we first define the notion of a *z-rank*.

(4) *Definition* ([9], section 1; [10], 65, fig. 2): For every (finite)  $\mathbf{P}$ -consistent<sup>2</sup> set of conditionals  $\Gamma = \{A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n\}$ , the *z-rank* of the elements of  $\Gamma$  is defined by the following *z-algorithm*:

(i) *Initial step*: Set  $i = 0$ . Set  $\Delta = \Gamma$ .

(ii) *Iterative step*:

(1) If  $\Delta$  is nonempty, let  $\Delta_i \subseteq \Delta$  consist of all conditionals  $\alpha$  in  $\Delta$  which are *tolerated* by  $\Delta$ , otherwise go to (iii).<sup>3</sup>

(2) If  $\Delta_i$  is nonempty, set  $\Delta = \Delta - \Delta_i$ , and  $i = i + 1$ .

(3) If  $\Delta_i$  is empty, set  $\Delta_\infty = \Delta$ , and set  $\Delta = \emptyset$ .

<sup>2</sup> A set of conditionals  $\Gamma$  is called  $\mathbf{P}$ -consistent *iff*  $\Gamma$  does not  $\mathbf{P}$ -entail  $\neg \perp \Rightarrow \perp$ .

<sup>3</sup> A conditional  $A \Rightarrow B$  is tolerated by  $\Delta$ , if there is a possible world over the propositional atoms appearing in  $\Delta$  that verifies  $A \wedge B$  and does not falsify any conditional in  $\Delta$ .

(iii) *Output*: The  $z$ -partition  $(\Delta_0, \dots, \Delta_k, \Delta_\infty)$ .

The  $z$ -rank of a conditional  $\alpha$  in a  $\mathbf{P}$ -consistent  $\Gamma$ , written “ $z_\Gamma(\alpha)$ ”, is defined as the index  $i$  of that set  $\Delta_i$  in the  $z$ -partition of  $\Gamma$  in which  $\alpha$  occurs.

The assumption of the preceding definition, that  $\Gamma$  is  $\mathbf{P}$ -consistent, guarantees that there is a  $z$ -model for  $\Gamma$ , according to the following definition:

(4) *Definition* ([9], 123-5, Eq. 5, 6, and 10): The  $z$ -model of a  $\mathbf{P}$ -consistent  $\Gamma$ ,  $(W_\Gamma, z_\Gamma)$ , is defined as follows:

For each  $w$  among the set of logically possible worlds over the propositional atoms appearing in  $\Gamma$ :

(i) If  $w$  falsifies  $\Delta_\infty$ , then  $w \notin W_\Gamma$ .

*Else*: (ii)  $w \in W_\Gamma$ , and  $z_\Gamma(w) = 0$ , if  $w$  doesn't falsify any  $\alpha$  in  $\Gamma$ ; otherwise  $z_\Gamma(w) = \max(\{z_\Gamma(\alpha) : w \text{ falsifies } \alpha\}) + 1$ .

(iii) The  $z$ -rank of an arbitrary formula  $C$  relative to  $(W_\Gamma, z_\Gamma)$  is defined as  $z_\Gamma(C) = \min(\{z_\Gamma(w) : w \in W_\Gamma \text{ and } w \text{ verifies } C\})$ , with  $\min(\emptyset) = \infty$ .

(iv) For all  $\Gamma$ :  $\Gamma \Vdash_{\sim\sim\mathbf{Z}} C \Rightarrow D$  ( $\Gamma$   $\mathbf{Z}$ -entails  $C \Rightarrow D$ ) iff either (a)  $\Gamma$  is  $\mathbf{P}$ -inconsistent, or (b)  $C \Rightarrow D$  is satisfied in  $(W_\Gamma, z_\Gamma)$  (i.e., all worlds with rank  $z_\Gamma(C)$  verify  $D$ ).

$\mathbf{Z}$ -entailment validates inference by *default inheritance* (i.e.,  $A \Rightarrow B \Vdash_{\sim\sim\mathbf{Z}} A \wedge C \Rightarrow B$ ) as well as *default contraposition* (i.e.,  $A \Rightarrow B \Vdash_{\sim\sim\mathbf{Z}} \neg B \Rightarrow \neg A$ ). That these inferences hold ‘by default’ means that they hold under the condition that the conditional knowledge base doesn't contain further conditionals that are  $\varepsilon$ -inconsistent<sup>4</sup> with the conclusions of these inferences (cf. Adams 1975). The relation  $\Vdash_{\sim\sim\mathbf{Z}}$  is thus *non-monotonic*, since, for example, whether  $\Gamma \cup \{A \Rightarrow B\} \Vdash_{\sim\sim\mathbf{Z}} \neg B \Rightarrow \neg A$ , depends on whether  $\Gamma \cup \{A \Rightarrow B\} \cup \{\neg B \Rightarrow \neg A\}$  is  $\varepsilon$ -inconsistent.

One disadvantage of  $\mathbf{Z}$ -entailment is that (in the absence of further assumptions) it does not automatically provide information concerning probabilistic reliability, such as provided by the improbability-sum semantics for system  $\mathbf{P}$ . However, in [5] it is shown how to obtain this desideratum (based on work in [12]):

**Theorem 1** If  $A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n \Vdash_{\sim\sim\mathbf{Z}} C \Rightarrow D$  holds, then improbability-sum preservation ( $I(D|C) \leq \sum\{I(B_i|A_i) : 1 \leq i \leq n\}$ ) holds for all probability functions  $\mathbf{P}$  that satisfy the default assumptions  $P(A_i \supset B_i | C) \geq P(B_i | A_i)$ , for all  $1 \leq i \leq n$ .

Proof: See [5], theorem 4 (5).

We proceed here as if the default assumptions specified in theorem 1 hold, and say that system  $\mathbf{Z}$  licenses the inference to  $C \Rightarrow_{1-\sum\{1-r_i : 1 \leq i \leq n\}} D$  from  $A_1 \Rightarrow_{r_1} B_1, \dots, A_n \Rightarrow_{r_n} B_n$ , in cases where  $A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n \Vdash_{\sim\sim\mathbf{Z}} C \Rightarrow D$ . As with the evaluations conducted in

<sup>4</sup> A set of conditionals is  $\varepsilon$ -consistent just in case the corresponding conditional probabilities can be simultaneously made arbitrarily close to 1.

[5], a central question concerns whether inference in accordance with the preceding principle tends to yield accurate conclusions.

## 1.4 System QC

**Z**-entailment is not the strongest (minimally reasonable) inference calculus for ‘risky’ default inference among uncertain conditionals. An even stronger and extremely simple calculus is *quasi-classical* reasoning. Here one reasons with uncertain conditionals as if they were material implications:

$$(5) \Gamma \vdash_{\text{QC}} C \supset D \text{ iff } \{ A \supset B : A \Rightarrow B \in \Gamma \} \vdash C \supset D.$$

Improbability-sum preservation holds for inferences between material conditionals, or more generally, between formulas of propositional logic, as was shown by Suppes ([13], 54). In particular,  $\{A_1, \dots, A_n\} \vdash B$  iff it holds for all probability distributions that  $I(B) \leq \sum \{I(A_i) : 1 \leq i \leq n\}$ . Beyond the result of Suppes, it is possible to formulate probabilistic conditions under which **QC**-reasoning approximately satisfies improbability-sum preservation. In particular, it is shown in ([5], sec. 2.5, (13)) that a **QC** inference from a given set of premises is guaranteed to preserve probability in the manner of system **P** iff the improbability-sum of the premises is very small, and some decimal powers smaller than the probability of the conclusion’s antecedent. Following [5], we proceed as if these conditions hold, and say that system **QC** licenses the inference to  $C \Rightarrow_{1-\sum\{1-r_i : 1 \leq i \leq n\}} D$  from  $A_1 \Rightarrow_{r_1} B_1, \dots, A_n \Rightarrow_{r_n} B_n$ , in cases where  $A_1 \Rightarrow B_1, \dots, A_n \Rightarrow B_n \vdash_{\text{QC}} C \supset D$ . The question remains of whether inference in accordance with the preceding principle tends to yield accurate conclusions.

## 2 The Simulations

Following [5], our simulations operate over a simple language with four two-valued variables: a, b, c, and d. Similarly, we assume a probability distribution over the sixteen possible worlds describable in this language. For all of our simulations, we generated a probability distribution over these worlds by setting the values of the following fifteen independently variable probabilities:  $P(a)$ ,  $P(b|a)$ ,  $P(b|\neg a)$ ,  $P(c|a \wedge b)$ ,  $P(c|a \wedge \neg b)$ ,  $P(c|\neg a \wedge b)$ ,  $P(c|\neg a \wedge \neg b)$ ,  $P(d|a \wedge b \wedge c)$ ,  $P(d|a \wedge b \wedge \neg c)$ ,  $P(d|a \wedge \neg b \wedge c)$ ,  $P(d|a \wedge \neg b \wedge \neg c)$ ,  $P(d|\neg a \wedge b \wedge c)$ ,  $P(d|\neg a \wedge b \wedge \neg c)$ ,  $P(d|\neg a \wedge \neg b \wedge c)$ , and  $P(d|\neg a \wedge \neg b \wedge \neg c)$ . Within [5], the probability distributions over the sixteen worlds were selected for each simulation, by setting the above fifteen conditional probabilities according to a uniform probability distribution on the unit interval. Diverging from [5], we controlled the entropy level of the probability distributions over the sixteen worlds. For each simulation, we chose a particular entropy level  $\delta$ . Our program then proceeded by generating probability distributions in the manner of [5] until a distribution was generated whose entropy resided in the interval  $[\delta-0.05, \delta+0.05]$ .

To manage the search space in assessing the four LP systems, we restricted our attention to conditionals whose antecedent and consequent consist in conjunctions of literals. We also assumed that no propositional atom appears twice in any *premise conditional* or *inferred conditional*. These restrictions effectively limited the language under consideration to 464 conditionals (cf. [5]). We call the language composed of this set of 464 conditionals “ $L_4$ ”. Drawing from  $L_4$ , we assumed that a small number of conditionals, so-called *premise conditionals*, together with their associated probabilities, were known to the reasoning systems. We further required that the probability associated with each premise conditional was at least 0.9. We chose the cut-off 0.9, since cases where the probability of the premise conditionals is relatively high represent a significant challenge for systems **Z** and **QC** (cf. [5]). In each simulation, the three premise conditionals were selected at random from among the sentences of  $L_4$  whose probability was at least 0.9. We then allowed each LP reasoning system to infer, from the given premise conditionals, all of the conditionals,  $C \Rightarrow_r D$ , that follow according to the respective systems. For systems **P**, **Z**, and **QC**, the value  $r$ , for each inferred conditional, was set to be one minus the sum of the improbabilities of the premise conditionals needed in deriving the conclusion. For system **O**,  $r$  was set to be the probability value of the least probable premise conditional needed for the derivation of  $C \Rightarrow D$  in **O**.

After determining which conclusions were inferred by the four systems, each system was assigned numeric scores for each of the conclusions that it inferred. The first scoring measure that we applied is called the *advantage-compared-to-guessing* measure. The idea behind this measure derives from the fact that the mean difference between two random choices of two real values  $r$  and  $s$  from the unit interval is (provably)  $1/3$ . Based on this fact, we assessed each system by counting a judged lower probability bound that differs from the true probability by more than one-third *negatively*, and counting a judged lower probability bound that differs from the true probability by less than one-third *positively*. We scored the judged lower probability bounds by a simple *linear* measure of their distance from the true probabilities:

(6) The *advantage-compared-to-guessing* (ACG) score for derived conditionals:

$$\text{Score}_{\text{ACG}}(C \Rightarrow_r D, P) := 1/3 - |r - P(D|C)|.$$

For reasons elaborated in [5], the ACG measure does not provide a fully adequate means of evaluating LP systems. In order to take a broad view of the advantages and disadvantages of reasoning in accordance with the four systems, we considered two other scoring measures.

We call the second measure that we considered the *subtle-price-is-right* measure. This measure assigns a positive score to any inferred lower probability bound that does not exceed the true probability, and penalizes inferred bounds that exceed the true probability by a simple *linear* measure of their distance above the true probability:

(7) The *subtle-price-is-right* score for derived conditionals<sup>5</sup>:

$$\begin{aligned} \text{Score}_{\text{SPIR}}(C \Rightarrow_r D, P) &:= r, \text{ if } r \leq P(D|C), \\ &:= P(D|C) - r, \text{ otherwise.} \end{aligned}$$

We call the final scoring measure that we considered the *expected utility* measure:

(8) The *expected utility* score for derived conditionals:

$$\text{Score}_{\text{EU}}(C \Rightarrow_r D, P) := (P(D|C)^2 - (P(D|C) - r)^2) \cdot P(C)/2.$$

The EU measure scores an inferred conditional,  $C \Rightarrow_r D$ , by evaluating the expected value of the decisions licensed by the acceptance of such a conditional (i.e., a conditional whose content is  $P(D|C) \geq r$ ). In particular, we assume that a judged greatest lower conditional probability bound has the following behavioral import: If  $r$  is the greatest lower probability bound that a given agent accepts for  $D$  given  $C$ , then (if she is prudent and has sufficient wealth) she will purchase all wagers on  $D$ , conditional on  $C$ , at price  $\$s$ , so long as  $s < r$ , and refuse to accept such wagers for  $s \geq r$ . Given this behavioral interpretation of inferred conditionals, we considered an environment in which a respective agent is offered a single opportunity to purchase a wager on  $D$  conditional on  $C$  with a stake  $s$ , where  $s$  is determined at random, according to a uniform probability distribution over the interval  $[0, 1]$ . In that environment, the expected value of accepting the greatest lower probability bound  $r$  on  $P(D|C)$  is *provably*:  $(P(D|C)^2 - (P(D|C) - r)^2) \cdot P(C)/2$  (cf. [14]).

### 3 The Results

The entropy of a probability distribution,  $P$ , over a finite set of possible worlds,  $W$ , is defined as  $E(P) = -\sum_i P(w_i) \cdot \log(P(w_i))$  (for  $w_i \in W$ ). So in the case where  $W$  contains sixteen worlds (as is the case in our simulations)  $E(P)$  will be in  $[0, 4]$ , where  $E(P) = 4$  means that  $P$  is a uniform probability distribution over  $W$ , and  $E(P) = 0$  means that  $P$  is a standard valuation function (assigning the value 1 to exactly one world, and the value 0 to all others). Since the four LP systems that we consider are ordered in terms of the number of inferences they license ( $\mathbf{O} \subset \mathbf{P} \subset \mathbf{Z} \subset \mathbf{QC}$ ), our focus here is on the ‘new’ inferences licensed by each system as one proceeds from system  $\mathbf{O}$  to system  $\mathbf{QC}$ , i.e., the inferences licensed by system  $\mathbf{O}$ , the inferences licensed by system  $\mathbf{P}$  that are not licensed by system  $\mathbf{O}$  ( $\mathbf{P}-\mathbf{O}$ ), the inferences licensed by system  $\mathbf{Z}$  that are not licensed by system  $\mathbf{P}$  ( $\mathbf{Z}-\mathbf{P}$ ), and the inferences licensed by system  $\mathbf{QC}$  that are not licensed by system  $\mathbf{Z}$  ( $\mathbf{QC}-\mathbf{Z}$ ). Table 1 lists the average number of conclusions inferred by each (sub)system, across varied entropy levels, and the average number erroneous inferences among the  $\mathbf{Z}-\mathbf{P}$  and  $\mathbf{QC}-\mathbf{Z}$  inferences, i.e., those instances

---

<sup>5</sup> The name of the measure derives from the long running American game show where contestants must guess the price of items, and succeed by have the most accurate guess that does not exceed the price of the relevant item.



where the inferred lower bound exceeded the actual probability. (These average values are based on a sample of one thousand simulations at each listed entropy level.)

**Table 1.** Mean number of inferences and errors

Entropy Level	Mean Number of Inferences				Mean Number of Errors	
	O	P-O	Z-P	QC-Z	Z-P	QC-Z
3.5	3.02	0.06	10.31	3.62	8.2	3.55
3.0	3.1	0.25	23.33	20.8	16.04	20.03
2.5	3.28	0.4	30.88	35.33	21.21	34.03
2.0	3.56	0.76	34.65	42.61	24.41	40.85
1.5	4.01	1.23	38.4	49.33	27.77	46.98
1.0	4.3	1.84	40.06	51.62	29.94	49.1
0.5	4.9	2.82	40.61	53.75	31.3	51.56

The most obvious pattern exhibited in table 1 is that the number of inferences drawn by each system is a decreasing function of the entropy level. This pattern was expected, since lower entropy levels imply a less evenly distributed probability function, and in turn a greater number of possible premise conditionals with multiple conjuncts in their consequents. Such conditionals support a greater number of inferences in all of the systems considered.

We now consider the average scores earned by the respective systems for the full set of conclusions drawn within a single simulation. Tables 2, 3, and 4 list the results.

**Table 2.** Mean ACG scores

Entropy Level	Mean ACG Scores			
	O	P-O	Z-P	QC-Z
3.5	1.01	0.02	2.41	-0.49
3.0	1.03	0.07	4.47	-3.61
2.5	1.08	0.11	4.02	-6.76
2.0	1.17	0.21	2.59	-8.48
1.5	1.31	0.34	1.10	-9.98
1.0	1.40	0.51	-1.06	-10.67
0.5	1.600	0.85	-2.78	-11.91

**Table 3.** Mean sPIR scores

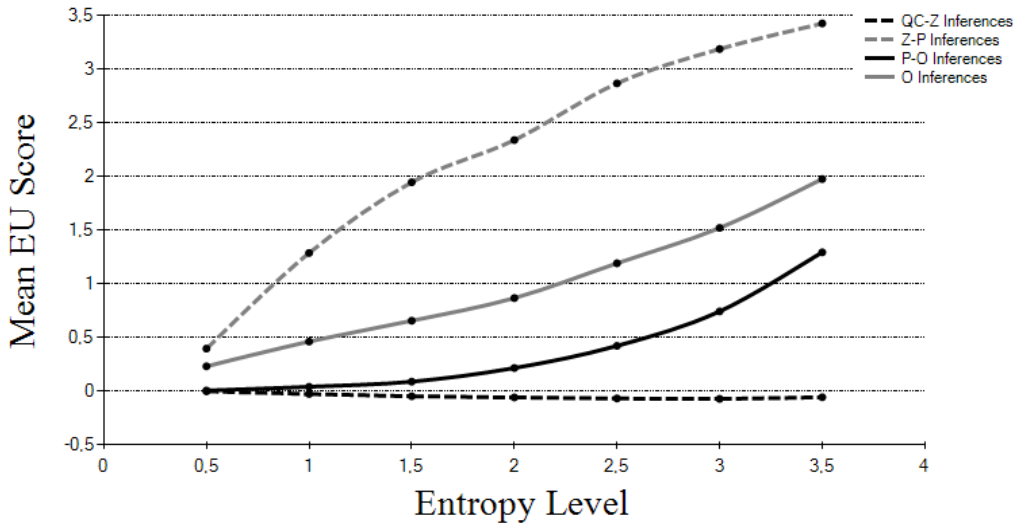
Entropy Level	Mean sPIR Scores			
	O	P-O	Z-P	QC-Z
3.5	2.83	0.06	1.01	-1.62
3.0	2.92	0.22	3.73	-9.80
2.5	3.11	0.35	3.11	-17.28
2.0	3.40	0.68	1.00	-20.98
1.5	3.83	1.10	-1.34	-24.14
1.0	4.12	1.67	-4.55	-25.43
0.5	4.76	2.68	-7.16	-27.67

**Table 4.** EU scores

Entropy Level	Mean EU Scores			
	O	P-O	Z-P	QC-Z
3.5	0.232	0.006	0.396	-0.001
3.0	0.462	0.041	1.287	-0.026
2.5	0.656	0.089	1.945	-0.047
2.0	0.867	0.216	2.338	-0.058
1.5	1.191	0.422	2.865	-0.067
1.0	1.520	0.743	3.185	-0.071
0.5	1.975	1.293	3.423	-0.056

Examining tables 2, 3, and 4, we see that the **QC-Z** inferences earn negative scores at every entropy level, according to all three scoring rules. This provides a relatively good reason for concluding that we should not reason in accordance with system **QC**, if our concern is to draw conclusions that are accurate and informative. On the other hand, we see that **O** and **P-O** inferences earn positive scores at every entropy level, according to all three scoring rules. So it pretty clear that it is reasonable to make these inferences. In fact, the present conclusion is unsurprising given (1) and (2), above, that characterize the ability of systems **O** and **P** to preserve premise probability.

It is only when we turn to evaluate the quality of **Z-P** inferences that the data from tables 1, 2, and 3 is equivocal. When considering the ACG and sPIR scores for the **Z-P** inferences, we observe a peak in performance, when the entropy level of the underlying probability distribution is relatively high ( $\approx 3.00$ ), but thereafter decreasing entropy correlates with decreasing ACG and sPIR scores. On the other hand, decreasing entropy correlates with increasing EU scores. Figure 1 provides a graphical representation of that pattern.

**Fig. 1.** Mean EU scores

In order to get a clearer idea of what's going on, it is helpful to look at the average scores earned for single inferences across varied entropy levels. Tables 5, 6, and 7 list the results, and figure 2 provides a graphic presentation of the information presented in table 7.

**Table 5.** Mean ACG scores per inference

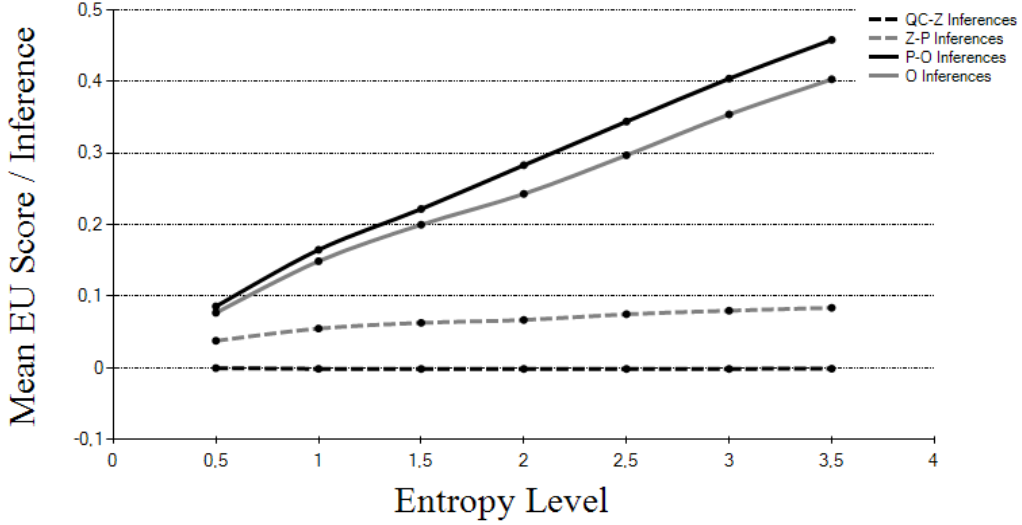
Entropy Level	Mean ACG Score per Inference			
	O	P-O	Z-P	QC-Z
3,5	0.333	0.275	0.234	-0.134
3,0	0.332	0.269	0.192	-0.173
2,5	0.331	0.275	0.130	-0.191
2,0	0.329	0.273	0.075	-0.199
1,5	0.326	0.274	0.029	-0.202
1,0	0.325	0.278	-0.026	-0.207
0,5	0.327	0.302	-0.069	-0.221

**Table 6.** Mean sPIR scores per inference

Entropy Level	Mean sPIR Score per Inference			
	O	P-O	Z-P	QC-Z
3,5	0.939	0.863	0.098	-0.447
3,0	0.944	0.866	0.160	-0.471
2,5	0.950	0.878	0.101	-0.489
2,0	0.953	0.887	0.029	-0.492
1,5	0.954	0.898	-0.035	-0.489
1,0	0.958	0.906	-0.114	-0.493
0,5	0.972	0.949	-0.176	-0.515

**Table 7.** Mean EU scores per inference

Entropy Level	Mean EU Score per Inference			
	O	P-O	Z-P	QC-Z
3,5	0.077	0.086	0.038	-0.00015
3,0	0.149	0.165	0.055	-0.00124
2,5	0.200	0.222	0.063	-0.00133
2,0	0.243	0.283	0.067	-0.00137
1,5	0.297	0.344	0.075	-0.00137
1,0	0.354	0.404	0.080	-0.00138
0,5	0.403	0.458	0.084	-0.00103



**Fig. 2.** Mean EU scores per inference

Our main remaining concern is to evaluate the quality of **Z–P** inferences. Tables 5 shows that **Z–P** inferences lead to bounds that are relatively close to the true probabilities, when entropy high. However, when the entropy level is very low, the distance between the judged bounds and the true probability tends to be rather great. For example, when the entropy of the underlying distribution is 0.5, the inferred lower bound for an average **Z–P** inference differs from the true probability by about 0.4. Similarly, while **Z–P** inferences are expected to yield relatively high sPIR scores, when an inferred bound is not in error (ranging from about 23 to 38 percent of cases, depending on the entropy level), we see that when the entropy level is low, a typical erroneous inferred bound exceeds the true probability by a significant margin. In contrast, the EU scores for **Z–P** inferences (as with **O** and **P–O** inferences) increases with decreases in the entropy of the underlying probability distribution.<sup>6</sup> The latter result marks one positive sign in favor of the quality of **Z–P** inferences. And we maintain that the latter result *does* reflect a significant capacity of **Z–P** inferences to exploit information about an environment to draw *helpful* conclusions about that environment. Indeed, if we consider plausible *aprioristic* methods of assigning lower probability bounds, such as the ones considered in [14], i.e., methods of assigning lower probability bounds to the elements of  $L_4$  without exploiting the information that was supplied to the four LP systems (in the form of premise conditionals), then we see that the EU scores earned for **Z–P** inferences tend to be much higher than the scores earned by aprioristic methods. For example, the most successful aprioristic method considered in [14] assigned the lower bounds 1/2, 1/4, and 1/8, respectively, to conditionals with one, two, or three conjuncts in their consequent (as the values 1/2, 1/4, and 1/8 are the average probabilities for conditionals with the corresponding number of conjuncts in their consequents). In the case where entropy was not con-

<sup>6</sup> The present effect is the result of inferred conclusions with more probable antecedents, when the entropy level is low.

trolled (and the mean entropy of the underlying probability distributions was about 2.88), this aprioristic method earned an EU score of about 0.0204 per inference, which is far lower than the average scores earned by **Z–P** inferences (across all entropy levels).

## 4 Conclusions

It almost goes without saying that it is reasonable to accept the conclusions of **O** and **O–P** inferences, so long as our goal is to accept accurate and informative probability statements. It is also quite clear that we should not accept the conclusions of **QC–Z** inferences. The difficult choice is whether to accept the conclusions of **Z–P** inferences. In cases where it is *known* that the entropy of the underlying distribution is not low (or probably not low), it will usually be reasonable to shoulder the risk inherent in accepting the conclusions of **Z–P** inferences. More generally, the tendency of **Z–P** inferences to deliver significant positive EU scores (even when the entropy of the underlying distribution is very low) indicates the value of these inferences as a basis for decision making.

In considering whether it is reasonable to accept the conclusions of **Z–P** inferences, we think it is reasonable to consider whether there are alternatives that would support better probability judgments. Since we know that our present method of associating lower probability bounds with **Z–P** inferences is prone to overestimation, we conjecture that a more optimal method would make a downward correction to these assigned bounds. It would also make sense to vary the size of this correction, in cases where the entropy of the underlying distribution is known. The exploration of this idea is left to future work.

**Acknowledgments.** Work on this paper was supported by the DFG-Grant SCHU1566/5-1 as part of EuroCores LogiCCC project *The Logic of Causal and Probabilistic Reasoning in Uncertain Environments* (LcpR), and the DFG-Grant SCHU1566/9-1 as part of the SPP 1516 priority program “New Frameworks of Rationality”.

## References

1. Hawthorne, J.: On the Logic of Non-Monotonic Conditionals and Conditional Probabilities, *Journal of Philosophical Logic* 25, 185-218 (1996)
2. Hawthorne, J., and Makinson, D.: The Quantitative / Qualitative Watershed for Rules of Uncertain Inference, *Studia Logica* 86, 247-297 (2007)
3. Paris, J.B., and Simmonds, R.: O is not Enough, *Review of Symbolic Logic* 2(2), 298-309 (2009)
4. Segerberg, K.: Notes on Conditional Logic, *Studia Logica* 48, 157-168 (1989)
5. Schurz, G., and Thorn, P.: Reward versus Risk in Uncertain Inference: Theorems and Simulations, *Review of Symbolic Logic* 4(2), 574-612 (2012)

6. Adams, E.W.: *The Logic of Conditionals*, Reidel, Dordrecht (1975)
7. Jaynes, E.: Prior Probabilities, *IEEE Transactions On Systems Science and Cybernetics* 4(3), 227-241 (1968)
8. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, Santa Mateo, California (1988)
9. Pearl, J.: System Z, *Proceedings of Theoretical Aspects of Reasoning about Knowledge*, Santa Mateo, California, 121-135 (1990)
10. Goldszmidt, M., and Pearl, J.: Qualitative Probabilities for Default Reasoning, Belief Revision and Causal Modeling, *Artificial Intelligence* 84, 57-112 (1996)
11. Halpern, J.: *Reasoning about Uncertainty*, MIT Press, Cambridge, Massachusetts (2003)
12. Schurz, G.: Probabilistic Default Reasoning Based on Relevance and Irrelevance Assumptions. In: D. Gabbay et al. (eds.), *Qualitative and Quantitative Practical Reasoning (LNAI 1244)*, Springer, Berlin, 536-553 (1997)
13. Suppes, P.: Probabilistic Inference and the Concept of Total Evidence. In: Hintikka, J., and Suppes, P. (eds.), *Aspects of Inductive Logic*, North-Holland Publ. Comp., Amsterdam, 49-65 (1966)
14. Thorn, P. and Schurz, G.: A Utility Based Evaluation of Logico-Probabilistic Systems. Manuscript submitted for publication