Propagating probabilities in System P

Rachel Bourne[1] and Simon Parsons
Department of Electronic Engineering.
Queen Mary and Westfield College
University of London.
London E1 4NS. UK.

Abstract

In this paper we suggest a way of using the rules of System P to propagate lower bounds on conditional probabilities. Using a knowledge base of default rules which are considered to be constraints on a probability distribution, the result of applying the rules of P gives us new constraints that were implicit in the knowledge base and their associated lower bounds.

Introduction

Default reasoning has been widely studied in artificial intelligence, and a number of formalisms have been proposed as a means of capturing such reasoning (Ginsberg 1987), most prominent among which are default logic (Reiter 1980) and circumscription (McCarthy 1980). Many of these systems, including default logic and circumscription, have proposed particular mechanisms for default reasoning, and might therefore be considered quite specialised. However, there has also been work on more general approaches which attempt to analyse in broader terms what default reasoning involves. An early attempt to do this was Shoham's (1987) proposal that all non-monotonic systems could be characterised in terms of the preference order over their models. A more proof-theoretic strand of this research has investigated the formalisation of the underlying requirements for any non-monotonic consequence relation. Perhaps the most influential piece of work within this area is that of Kraus et al. (1990).

Kraus et al. investigated the properties of different sets of Gentzen-style proof rules for non-monotonic consequence relations, and related these sets of rules to the model-theoretic properties of the associated logics. Their major result was that a particular set of proof rules had the same model-theoretic properties that Shoham had identified for logics in which there is a preference order over models. This system of proof rules was termed System P by Kraus et al., the P

standing for "preferential". System P has been the subject of much research, and is now widely accepted as the weakest interesting non-monotonic system: it sanctions the smallest acceptable set of conclusions from a set of default statements.

The reason that we are interested in the rules of System P is that, in addition to a semantics in terms of a preference order over models. they also have a probabilistic semantics. In particular, Pearl (1988), following work by Adams (1975), showed that a semantics based on infinitesimal probabilities satisfies the rules of System P. While the use of infinitesimal probabilities is theoretically interesting. it lacks something in practical terms. If we are to use System P to reason about the real world we will have to write defaults which summarise our knowledge about it, and we may well be unhappy making statements whose validity depends upon infinitesimal values. To overcome this difficulty, we suggest using real probabilities along with the rules of System P, giving each default statement a lower bounded probability, and showing that proofs in the System P can be used to propagate these bounds to find out something concrete about the probability of the derived results.

Entailment in System P

The rules of inference for the System P (see Figure 1) may be applied to a knowledge base made up of conditional assertions of the form $\alpha \mathbin{\vdash\!\!\!\sim} \beta$. In this context α and β are well-formed formulae of classical propositional logic. and $\mathbin{\vdash\!\!\!\sim}$ is a binary relation between pairs of formulae. The connectives $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$ have their usual meanings. The inference rules are written in the usual Gentzen style, with antecedents above the line and consequents below it. Thus the rule 'And' says that if it is possible to derive $\alpha \mathbin{\vdash\!\!\!\sim} \beta$ and it is possible to derive $\alpha \mathbin{\vdash\!\!\!\sim} \gamma$. then it is possible to derive $\alpha \mathbin{\vdash\!\!\!\sim} \beta \wedge \gamma$. The inference rules can thus be viewed as a means of obtaining new conclusions from current knowledge; from an initial set of conditional assertions,

$$\alpha \mathrel{\vert\!\sim} \alpha \qquad \text{Reflexivity}$$

$$\frac{\models \alpha \leftrightarrow \beta,\ \alpha \mathrel{\vert\!\sim} \gamma}{\beta \mathrel{\vert\!\sim} \gamma} \qquad \text{Left Logical Equivalence}$$

$$\frac{\models \beta \to \gamma,\ \alpha \mathrel{\vert\!\sim} \beta}{\alpha \mathrel{\vert\!\sim} \gamma} \qquad \text{Right Weakening}$$

$$\frac{\alpha \mathrel{\vert\!\sim} \beta,\ \alpha \mathrel{\vert\!\sim} \gamma}{\alpha \mathrel{\vert\!\sim} \beta \wedge \gamma} \qquad \text{And}$$

$$\frac{\alpha \mathrel{\vert\!\sim} \beta,\ \alpha \mathrel{\vert\!\sim} \gamma}{\alpha \wedge \beta \mathrel{\vert\!\sim} \gamma} \qquad \text{Cautious Monotonicity}$$

$$\frac{\alpha \mathrel{\vert\!\sim} \gamma,\ \beta \mathrel{\vert\!\sim} \gamma}{\alpha \vee \beta \mathrel{\vert\!\sim} \gamma} \qquad \text{Or}$$

Figure 1: Rules of System P

further conditional assertions may be obtained by applying the rules.

Two things should be noted about the set of rules in Figure 1. Firstly, they only tell us how to derive new conditional assertions. If we wish to know whether we are justified in inferring a new fact, say γ, given that we currently know some other fact, say α, *and this is all we know*, it is necessary to determine whether $\alpha \mathrel{\vert\!\sim} \gamma$ is derivable from our knowledge base of conditional assertions. Secondly, the proof rules in Figure 1 form a minimal set sufficient to characterise System P. Other rules may be derived from them in much the same way that new conditional assertions are derived. Two such rules are given in Figure 2; Cut which allows the elimination of a conjunct from the antecedent side, and S which allows the derivation of a material implication. Both of these (as we shall see later in the paper) may be derived directly by the application of the basic rules.

The semantics for System P introduced by Adams makes the assumption that the propositional variables are the basis of an unspecified joint probability distribution which is constrained by the conditional assertions. These conditionals are taken to represent conditional probabilities of the consequent given the antecedent being greater than or equal to $1 - \epsilon$ for any $\epsilon > 0$, that is:

$$\alpha \mathrel{\vert\!\sim} \beta \text{ represents } P(\beta|\alpha) \geq 1 - \epsilon \text{ for all } \epsilon > 0. \quad (1)$$

Probabilistic consistency is defined as the existence of at least one probability distribution which satisfies these constraints (Adams 1975); probabilistic entailment of a further conditional is defined as probabilistic

$$\frac{\alpha \wedge \beta \mathrel{\vert\!\sim} \gamma,\ \alpha \mathrel{\vert\!\sim} \beta}{\alpha \mathrel{\vert\!\sim} \gamma} \qquad \text{Cut}$$

$$\frac{\alpha \wedge \beta \mathrel{\vert\!\sim} \gamma}{\alpha \mathrel{\vert\!\sim} \beta \to \gamma} \qquad \text{S}$$

Figure 2: Two derived rules of System P

inconsistency of its counterpart, that is:

$\alpha \mathrel{\vert\!\sim} \beta$ is p-entailed by Δ iff $\Delta \cup \{\alpha \mathrel{\vert\!\sim} \neg\beta\}$ is p-inconsistent

This implies that all probability distributions that satisfy Δ also satisfy $\alpha \mathrel{\vert\!\sim} \beta$. However this result may only be achieved by using infinitesimal analysis so that the derived conditional will be constrained to be greater than $1 - \delta$ for any $\delta > 0$ if the ϵ of the original conditionals is made small enough. This can be paraphrased as saying that System P allows us to make our conclusions as close to certainty as we like, provided the conditional probabilities associated with the conditional assertions are sufficiently close to certainty. In the literature this is used to justify the conclusions drawn by using System P; if we are sure of the conditional assertions and so are willing to give them high conditional probabilities, then the conditional assertions derived from them will also have high probabilities.

However, using this interpretation of the rules means assuming that we are able to give the conditional assertions arbitrarily high conditional probabilities. This is fine in the case that the assertions are pieces of default knowledge which are felt to hold almost all of the time. However, with less reliable information, for which ϵ is not infinitesimal. it seems less justifiable to accept the infinitesimal analysis. In particular, if a set of conditional assertions are used to derive new assertions and these new assertions are themselves used as the basis for new deductions, then it seems likely that some ϵ values will be far from infinitesimal. Because of this concern, this paper investigates the impact of non-infinitesimal ϵ values by considering what happens to values of ϵ and δ when the rules of P are applied. The result is twofold. First it is possible to track the effect of non-infinitesimal values, and second it becomes possible to identify bounds on the actual conditional probability of derived assertions.

Using real ϵ-values

We associate with each conditional assertion an ϵ-value which represents, for $\alpha \mathrel{\vert\!\sim} \beta$, an upper bound on the conditional probability $P(\neg\beta|\alpha)$. We demonstrate how using these values for each original conditional, we can

$$
\begin{array}{ccccc}
0 & \leq & P(\alpha) & \leq & 1 \\
0 & \leq & P(\beta) & \leq & 1 \\
P(\alpha)(1 - \epsilon_1) & \leq & P(\alpha \wedge \gamma) & \leq & P(\alpha) \\
P(\beta)(1 - \epsilon_2) & \leq & P(\beta \wedge \gamma) & \leq & P(\beta) \\
max\{0, P(\alpha) + P(\beta) - 1\} & \leq & P(\alpha \wedge \beta) & \leq & min\{P(\alpha), P(\beta)\} \\
max\{0, P(\alpha \wedge \gamma) + P(\beta \wedge \gamma) - 1\} & \leq & P(\alpha \wedge \beta \wedge \gamma) & \leq & min\{P(\alpha \wedge \gamma), P(\beta \wedge \gamma), P(\alpha \wedge \beta)\}
\end{array}
$$

Figure 3: The constraints for Or.

generate δ values for the output conclusions. This enables us to calculate the lower bound on the probability of a conclusion based on the proof steps used to derive it. We consider first the six basic rules of System P, and then use the results obtained for those rules to obtain results for S and Cut.

Reflexivity: A reflexive conditional assertion may be introduced at any stage in a proof, and, since $P(\alpha|\alpha) = 1$ for all formulae α, any such conditional will have an ϵ-value of zero.

Left Logical Equivalence: This rule means that we may take any conditional assertion and replace its antecedent with a logically equivalent expression. Clearly, the derived conditional will have the same ϵ-value as the original one.

Right Weakening: Right Weakening involves replacing the consequent of a conditional with any expression classically derivable from it. Now, $\beta \rightarrow \gamma$ means the models of β are a subset of the models of γ and hence:

$$P(\gamma, \alpha) \geq P(\beta, \alpha)$$

Now, since:

$$P(\gamma|\alpha) = \frac{P(\gamma, \alpha)}{P(\alpha)}, \qquad P(\beta|\alpha) = \frac{P(\beta, \alpha)}{P(\alpha)}$$

it follows that:

$$P(\gamma|\alpha) \geq P(\beta|\alpha) \tag{2}$$

and therefore the ϵ-value of a rule obtained by Right Weakening will not be larger than the ϵ-value of the rule from which it was obtained. Since we are dealing with lower bounds, we may use the same value for the derived rule.

Cautious Monotonicity: Here we are interested in the value of $P(\gamma|\alpha, \beta)$. Now:

$$P(\gamma|\alpha) = P(\gamma|\alpha, \beta)P(\beta|\alpha) + P(\gamma|\alpha, \neg\beta)P(\neg\beta|\alpha) \tag{3}$$

Substituting $1 - P(\beta|\alpha)$ for $P(\neg\beta|\alpha)$ and rearranging, we obtain

$$P(\gamma|\alpha, \beta) = \frac{P(\gamma|\alpha) - (1 - P(\beta|\alpha))P(\gamma|\alpha, \neg\beta)}{P(\beta|\alpha)} \tag{4}$$

We are required to minimize this expression subject to the constraints:

$$
\begin{array}{ccccc}
1 - \epsilon_1 & \leq & P(\beta|\alpha) & \leq & 1 \\
1 - \epsilon_2 & \leq & P(\gamma|\alpha) & \leq & 1 \\
0 & \leq & P(\gamma|\alpha, \neg\beta) & \leq & 1
\end{array}
$$

Equation (4) is linear in $P(\gamma|\alpha)$ and $P(\gamma|\alpha, \neg\beta)$ and will therefore attain its minimum when $P(\gamma|\alpha)$ is minimum and $P(\gamma|\alpha, \neg\beta)$ is maximum. This gives us:

$$P(\gamma|\alpha, \beta) \geq \frac{(1 - \epsilon_1) - (1 - P(\beta|\alpha))}{P(\beta|\alpha)}$$
$$\geq 1 - \frac{\epsilon_1}{P(\beta|\alpha)}$$

which will be minimum when $P(\beta|\alpha)$ is minimum. This gives us an ϵ-value for the derived rule $\alpha \wedge \beta \mid\!\sim \gamma$ of:

$$\frac{\epsilon_2}{1 - \epsilon_1}$$

And: This time we are interested in $P(\beta, \gamma|\alpha)$. Consider:

$$P(\beta, \gamma|\alpha) = \frac{P(\alpha, \beta, \gamma)}{P(\alpha)}$$
$$= \frac{P(\alpha, \beta, \gamma)P(\alpha, \beta)}{P(\alpha, \beta)P(\alpha)}$$
$$= P(\gamma|\alpha, \beta)P(\beta|\alpha) \tag{5}$$

We are required to minimize this expression subject to the constraints:

$$
\begin{array}{ccccc}
1 - \epsilon_1 & \leq & P(\beta|\alpha) & \leq & 1 \\
1 - \epsilon_2 & \leq & P(\gamma|\alpha) & \leq & 1
\end{array}
$$

and in the previous case we saw that these constraints imply that:

$$1 - \frac{\epsilon_2}{1 - \epsilon_1} \leq P(\gamma|\alpha, \beta) \leq 1$$

Equation (5) will be minimum when both factors in the product on the right-hand side are, so that

$$P(\beta, \gamma|\alpha) \geq \left(1 - \frac{\epsilon_2}{1 - \epsilon_1}\right)(1 - \epsilon_1)$$
$$= 1 - (\epsilon_1 + \epsilon_2) \tag{6}$$

$$\alpha \mid\!\sim_0 \alpha \qquad \text{Reflexivity}$$

$$\frac{\models \beta \to \gamma, \alpha \mid\!\sim_{\epsilon_1} \beta}{\alpha \mid\!\sim_{\epsilon_1} \gamma} \qquad \text{Right Weakening}$$

$$\frac{\alpha \mid\!\sim_{\epsilon_1} \beta, \alpha \mid\!\sim_{\epsilon_2} \gamma}{\alpha \wedge \beta \mid\!\sim_{\frac{\epsilon_2}{1-\epsilon_1}} \gamma} \qquad \text{Cautious Monotonicity}$$

$$\frac{\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \gamma, \alpha \mid\!\sim_{\epsilon_2} \beta}{\alpha \mid\!\sim_{\epsilon_1 + \epsilon_2} \gamma} \qquad \text{Cut}$$

$$\frac{\models \alpha \leftrightarrow \beta, \alpha \mid\!\sim_{\epsilon_1} \gamma}{\beta \mid\!\sim_{\epsilon_1} \gamma} \qquad \text{Left Logical Equivalence}$$

$$\frac{\alpha \mid\!\sim_{\epsilon_1} \beta, \alpha \mid\!\sim_{\epsilon_2} \gamma}{\alpha \mid\!\sim_{\epsilon_1+\epsilon_2} \beta \wedge \gamma} \qquad \text{And}$$

$$\frac{\alpha \mid\!\sim_{\epsilon_1} \gamma, \beta \mid\!\sim_{\epsilon_2} \gamma}{\alpha \vee \beta \mid\!\sim_{\epsilon_1+\epsilon_2-\epsilon_1\epsilon_2} \gamma} \qquad \text{Or}$$

$$\frac{\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \gamma}{\alpha \mid\!\sim_{\epsilon_1} \beta \to \gamma} \qquad \text{S}$$

Figure 4: Rules with associated bounds

which, as we would expect, is symmetrical in ϵ_1 and ϵ_2. This gives us an ϵ-value for the derived rule of $\epsilon_1 + \epsilon_2$.

Or: For this rule we want to know the value of $P(\gamma | \alpha \vee \beta)$. This is slightly more tricky than previous cases since the necessary equation cannot be solved analytically. However there is a function which approximates the lower bound closely for small ϵ-values and is no worse than 10% lower than it even for large values. Consider the following:

$$P(\gamma | \alpha \vee \beta) = \frac{P(\alpha \wedge \gamma) + P(\beta \wedge \gamma) - P(\alpha \wedge \beta \wedge \gamma)}{P(\alpha) + P(\beta) - P(\alpha \wedge \beta)}$$

We are required to minimize this expression subject to the constraints of Figure 3. This is a non-linear optimisation problem which we can easily solve numerically over the specified range. A simple brute force search shows that the function is bounded below by $(1 - \epsilon_1)(1 - \epsilon_2)$ which gives us an ϵ-value for the derived rule of $\epsilon_1 + \epsilon_2 - \epsilon_1\epsilon_2$.

For completeness sake, we examine the derived rules Cut and S since they are the most useful rules when proving things. To make the presentation clearer we have denoted by $\mid\!\sim_{\epsilon_1}$ a conditional with ϵ-value of ϵ_1.

S: For S we need to derive $\alpha \mid\!\sim_{\epsilon_{new}} \beta \to \gamma$ and the value of ϵ_{new} from $\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \gamma$ just using the basic rules. This can be done as follows. First apply Right Weakening to $\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \gamma$ to get:

$$\frac{\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \gamma, \models \gamma \to (\beta \to \gamma)}{\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \beta \to \gamma} \tag{7}$$

We then apply Reflexivity followed by Right Weakening (twice) to $\alpha \wedge \neg\beta$ to get:

$$\frac{\alpha \wedge \neg\beta \mid\!\sim_0 \alpha \wedge \neg\beta, \models \neg\beta \to (\beta \to \gamma)}{\alpha \wedge \neg\beta \mid\!\sim_0 \beta \to \gamma} \tag{8}$$

Then we combine (7) and (8) using Or and apply Left Logical Equivalence to get:

$$\frac{\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \beta \to \gamma, \alpha \wedge \neg\beta \mid\!\sim_0 \beta \to \gamma}{\alpha \mid\!\sim_{\epsilon_1} \beta \to \gamma}$$

The consequent of this last derivation is the consequent of S, and comparing this with the antecedent, we can see that applying S has no effect on the ϵ-value; the value for the derived conditional assertion is the same as for the original assertion.

Cut: For Cut, we need to discover how $\alpha \mid\!\sim_{\epsilon_{new}} \gamma$ may be derived from $\alpha \wedge \beta \mid\!\sim_{\epsilon_1} \gamma$ and $\alpha \mid\!\sim_{\epsilon_2} \beta$. This turns out to be easy given the result for S. S tells us that the ϵ-value of $\alpha \mid\!\sim \beta \to \gamma$ is the same as that of $\alpha \wedge \beta \mid\!\sim \gamma$, so we have $\alpha \mid\!\sim_{\epsilon_1} \beta \to \gamma$ and applying And to $\alpha \mid\!\sim_{\epsilon_1} \beta \to \gamma$ and $\alpha \mid\!\sim_{\epsilon_2} \beta$, followed by Right Weakening gives:

$$\frac{\alpha \mid\!\sim_{\epsilon_1} \beta \to \gamma, \alpha \mid\!\sim_{\epsilon_2} \beta}{\alpha \mid\!\sim_{\epsilon_1+\epsilon_2} \gamma}$$

Cut is thus proved, and the ϵ-value for its consequent established.

In obtaining these results, we have shown that using each of the rules of P, and hence any derived rules, we can obtain lower bounds on the conditional probability of the conclusion given those of the antecedents. Figure 4 shows the basic rules plus S and Cut annotated with lower probability bounds on antecedents and consequents. It is clear that these lower bounds never improve. Using rules And and Or, or rules derived from these, means adding the ϵ-values so that after only a few proof steps our conclusions may attain high ϵ-values. A high ϵ-value means that the lower bound on the associated conditional probability is low and if this becomes too low then we don't have much information about the probability since the upper bound is always 1. Clearly, therefore, our input values must either be

$$\text{(i)} \quad \frac{linda \wedge steve \; \vdash_{0.05} \neg noisy, \; linda \; \vdash_{0.1} steve}{linda \; \vdash_{0.15} \neg noisy} \qquad \text{Cut 4.5}$$

$$\text{(ii)} \quad \frac{linda \; \vdash_{0.15} \neg noisy, \; linda \; \vdash_{0.01} great}{linda \; \vdash_{0.16} great \wedge \neg noisy} \qquad \text{And (i),2}$$

$$\text{(iii)} \quad \frac{\top \; \vdash_{0.16} \neg linda \vee (great \wedge \neg noisy), \; \top \; \vdash_{0.1} noisy \vee \neg great}{\top \; \vdash_{0.26} \neg linda} \qquad \text{S. And (ii),7, RW}$$

$$\text{(i)} \quad \frac{linda \; \vdash_{0.1} steve, \; linda \; \vdash_{0.01} great}{linda \wedge steve \; \vdash_{0.011} great} \qquad \text{CM 4.2}$$

$$\text{(ii)} \quad \frac{linda \wedge steve \; \vdash_{0.05} \neg noisy, \; linda \wedge steve \; \vdash_{0.011} great}{linda \wedge steve \; \vdash_{0.061} great \wedge \neg noisy} \qquad \text{And 5.(i)}$$

Figure 5: Proofs of $\top \; \vdash_{0.26} \neg linda$ and $linda \wedge steve \; \vdash_{0.061} great \wedge \neg noisy$

extremely small or our proofs short in order to obtain useful results. However, as our example shows, these conditions can be met without too much imagination.

Example

The following was inspired by examples given by Kraus et al. (Kraus, Lehmann, & Magidor 1990).

Brian and Linda are two happy-go-lucky people who are normally the life and soul of any party (so if either go to a party it will normally be great). Until recently Brian and Linda were married, but then Linda ran off with a mime artist, Steve. As a result, if both Brian and Linda go to the same party they will probably have a screaming row and ruin it (so it will not be great and it will be noisy).

If Linda goes to a party she will probably take her new boyfriend Steve and get him to entertain the guests with his marvellous miming. Thus if Linda goes to a party, Steve will probably go to the same party and if Linda and Steve go to a party together it will normally not be noisy because everyone will be watching his miming. Normally parties that great are noisy and those that are not noisy are not great.

We represent this by the following rules and ϵ-values: It should be understood that we are trying to ascertain the likelihood of any given party having various attributes (*brian* is present, it is *noisy*, and so on).

1. $brian \; \vdash_{0.01} great$
2. $linda \; \vdash_{0.01} great$
3. $brian \wedge linda \; \vdash_{0.15} \neg great \wedge noisy$
4. $linda \; \vdash_{0.1} steve$
5. $linda \wedge steve \; \vdash_{0.05} \neg noisy$
6. $great \; \vdash_{0.1} noisy$
7. $\neg noisy \; \vdash_{0.1} \neg great$

Figure 5 gives proofs for two new conditionals with justifications on the right and ϵ-values propagated along the way. $\top$ denotes any tautology. The first of these gives us the lower bound on the probability of Linda not attending any particular party given what we know (which is what the $\top$ on the antecedent side means). Remembering the denotation of (1), $\top \; \vdash_{0.26} \neg linda$ means that:

$$P(\neg linda | \top) \geq 1 - 0.26$$
$$P(\neg linda | \top) \geq 0.74$$

from which we can conclude that Linda is unlikely to go to any particular party. However, this does not stop us drawing conclusions about parties which we know that Linda does go to. Indeed, Figure 5 shows such a conclusion:

$$linda \wedge steve \; \vdash_{0.061} great \wedge \neg noisy$$

means that if Linda and Steve go to a party, then the probability that it is both *great* and *¬noisy* is greater than 0.94. These results show that using our method we can obtain usable numerical results from an initial set of defaults.

This is particularly useful in unexpected or unlikely circumstances (as in the case that we know Linda goes to a party) since such cases could not be captured in classical logic. To see this point consider what would happen if we translated part of our knowledge about parties into propositional logic:

1. *linda ⊃ great*
2. *great ⊃ noisy*
3. *linda ⊃ steve*
4. *linda ∧ steve ⊃ ¬noisy*

From the first two sentences we could conclude *noisy*, from the first, third and fourth sentences we could conclude ¬*noisy* and with the two we would have an inconsistency. What System P does is to balance the effects that Linda and Steve have on a party and our approach extends this to allow us to predict just how likely the balanced outcome is.

Conclusion

We have shown that given the assumption that conditional assertions may be treated as conditional probabilities with lower bounds, we can obtain lower bounds for the derived conclusions. Thus if we know the lower bounds on the conditional probabilities of a set of input assertions, we can establish the lower bounds on the conditional probabilities of the derived assertions. Moreover these are given by simple functions of the initial bounds calculated for each proof step in System P. One advantage of this approach is that it allows us to use real rather than infinitesimal probabilities since by keeping track of the bounds we can tell which conclusions are justified;clearly any conclusion with a low lower bound might be considered suspect. Another advantage of this method is that only a lower bound conditional probability is required for each default rule rather than a point probability, and this may mean that the numerical values are easier to assess. Clearly we still require these values to be high or the results obtained will be useless since derived conditionals will only be known to have an associated conditional probability that is greater than some small value.

There are two drawbacks to this approach which should be mentioned. Firstly System P is accepted as being a sceptical reasoning mechanism, that is, only conservative (and completely sound) conclusions can be obtained. This is insufficient for most purposes since we will often want to draw more tenuous conclusions. However, specialisations have been suggested (Goldszmidt, Morris, & Pearl 1993; Pearl 1990) and it may be possible for our approach to be extended in this direction. Secondly, despite the appearance of the rules of System P, it is not known whether a feasible procedure for proof generation exists; a model-theoretic procedure for determining p-entailment (which guarantees a proof) exists but the complexity prevents realistic applications.

Naturally for our approach to be acceptable, it is necessary to accept that defaults may be reasonably

taken to be statements about probabilities, in particular constraints on a joint probability distribution which describes the state of the world. While this is natural from a Bayesian perspective, it might seem more questionable to some. However, we feel that the position can be justified from a pragmatic perspective as well. Using probability to say what defaults mean gives them a semantics which is comparatively simple, at least at an informal level, and one which could, with our extension, be used in conjunction with real data about the world. This seems a good justification for accepting the semantics; in our opinion it is clearer and therefore more plausible than other default reasoning systems. Any reasoning mechanism that is required to treat defaults in a uniform way must have some representation for them. Humans may use default rules in different ways in differing contexts, but if we wish to build practical reasoning systems, we must find systematic ways to approximate defaults. If this leads to making some assumptions then, so long as they are clear, so be it. We will achieve nothing unless we have a firm basis for saying what we mean by a default.

References

Adams, E. 1975. *The Logic of Conditionals*. Dordrecht, Netherlands: Reidel.

Ginsberg, M. L., ed. 1987. *Readings in Nonmonotonic Reasoning*. San Mateo, CA: Morgan Kaufmann.

Goldszmidt, M.; Morris, P.; and Pearl, J. 1993. A maximum entropy approach to nonmonotonic reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15:220 232.

Kraus, S.; Lehmann, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44:167–207.

McCarthy, J. 1980. Circumscription--a form of nonmonotonic reasoning. *Artificial Intelligence* 13:27-39.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. 1990. System Z: a natural ordering of defaults with tractable applications to default reasoning. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, 121-135.

Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13:81-132.

Shoham, Y. 1987. Nonmonotonic logics: meaning and utility. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 388 393.