Maximum Entropy and Variable Strength Defaults

Rachel A. Bourne and **Simon Parsons**
Department of Electronic Engineering
Queen Mary & Westfield College
University of London
London E1 4NS, UK
r.a.bourne,s.d.parsons@elec.qmw.ac.uk

Abstract

A new algorithm for computing the maximum entropy ranking over models is presented. The algorithm handles arbitrary sets of propositional defaults with associated strength assignments and succeeds whenever the set satisfies a robustness condition. Failure of this condition implies the problem may not be sufficiently specified for a unique solution to exist. This work extends the applicability of the maximum entropy approach detailed in [Goldszmidt *et al.*, 1993], and clarifies the assumptions on which the method is based.

1 Introduction

There have been several suggestions of what might constitute the best consequence relation to be associated with a set of propositional defaults. The weakest, and most widely accepted, is System P [Adams, 1975], [Kraus *et al.*, 1990]. Of those which handle the more complex default interactions, such as exceptional inheritance, correctly, the maximum entropy approach (me) has, arguably, the clearest objective justification being derived from a well understood principle of indifference. In this paper, the me-approach of [Goldszmidt *et al.*, 1993] is extended so that the me-ranking for an arbitrary set of variable strength defaults can be found. A new algorithm is presented along with a sufficient condition for its successful computation. As well as handling the usual examples from the literature in a satisfactory way, this extended framework provides a flexible method for handling default knowledge through its use of variable strength defaults which sheds some light on previously ambiguous examples. Indeed, the results suggest that some examples are inherently ambiguous. However, the clear underlying principle of the me-approach clarifies why this ambiguity arises, and suggests how it might be resolved.

2 Deriving the me-ranking

Consider a set of defaults, $\Delta = \{r_i : a_i \Rightarrow b_i\}$ where a, b, c, are formulæ of a finite propositional language, $\mathcal{L}$, with the usual connectives $\wedge$, $\vee$, $\neg$, $\rightarrow$. The symbol, $\Rightarrow$, denotes a default connective. The models of $\mathcal{L}$ are contained in the set $\mathcal{M}$. A model, $m \in \mathcal{M}$, is said to *verify* a default, $a \Rightarrow b$, if $m \models a \wedge b$. Conversely, a model, m, is said to *falsify* a default, $a \Rightarrow b$, if $m \models a \wedge \neg b$.

The semantics of defaults is given in terms of conditional probabilities. Each default $a \Rightarrow b$ is supposed to constrain a set of probability distributions. For example, if it were assumed that $P(\neg b|a) = 0.05$, then the set $\{a \Rightarrow b\}$ would define all those probability distributions which satisfied the constraint imposed by the default. However, in this context no actual conditional probabilities are specified only the (fixed) relationships between the defaults in a given set.

The entropy of a probability distribution over a set of models $\mathcal{M}$ is given by

$$H[P] = -\sum_{m \in \mathcal{M}} P(m) \log P(m) \qquad (1)$$

The problem is to select that probability distribution which maximises (1) subject to constraints imposed by the defaults. The main supposition underlying this formalism is that specifying relative orders of magnitude for the conditional probabilities corresponding to each default implies a similar order of magnitude description of the probabilities of each model. This is achieved by parameterising the conditional probabilities and examining the behaviour as the parameter tends to zero. Intuitively, this can be thought of as taking a set of assumptions to the extreme in order to ascertain what other information is implied. The intuitive interpretation of the relative orders of magnitude between defaults is that one is required to specify their relative *strengths*; that is, numerically higher strength defaults can be thought of as holding more strongly than, or as having priority over, those of lower strength. Note that the symbol $\sim$ will be used to denote asymptotic equality since, for the purposes of this analysis, it is only the asymptotic behaviour of the probabilities that is important not their actual values, nor indeed the actual value of entropy.

Goldszmidt *et al.* [1993] originally chose to use inequalities for the default constraints but were unable to obtain results except for a small class of default sets, termed minimal core sets, which were guaranteed to satisfy the constraints as equalities. As they pointed out, for minimal core sets, their algorithm is easily adapted to cater for variable strength defaults. For arbitrary sets, however, the algorithm is unsound and an analysis of the behaviour of the me-approach applied to variable strength defaults was not provided.

In this revised analysis, the maximum entropy approach is extended by insisting on working with strict equality constraints, at least up to asymptotic equivalence. Specifying relative strengths for all defaults limits the region of possible probability distributions from which the maximum entropy distribution is taken. Although this requires a firmer commitment, that is, more information, from the knowledge engineer, it leads to me-solutions in a much larger number of cases.

Each default is assigned an associated *strength*, or order of magnitude, relative to the other defaults. Asymptotically, the coefficients of conditional probabilities can be ignored and so only the relative orders of magnitude between models will be relevant. The strength of each default is therefore expressed as some power of a parameter ε which has no significance other than linking all defaults together. Thus a default $a \Rightarrow b$ will be said to have relative strength s iff $P(\neg b|a) \sim \varepsilon^s$ for some integer $s > 0$. Letting $\varepsilon \to 0$, the term $\varepsilon^s \to 0$ and so $P(b|a) \to 1$, and the default becomes arbitrarily certain. In specifying a default, it is assumed that the knowledge engineer is encoding information which he takes to be almost certain. Similarly, the probability of each model m will be assumed to be asymptotically equivalent to some non-negative integer power $\kappa(m)$ of ε, so that $P(m) \sim \varepsilon^{\kappa(m)}$ for $\kappa(m) \geq 0$. The constraints imposed on P by the defaults $\{r_i\}$ can be written as:

$$\sum_{m \models a_i \wedge \neg b_i} P(m) \quad \sim \quad \frac{\varepsilon^{s_i}}{1-\varepsilon^{s_i}} \sum_{m \models a_i \wedge b_i} P(m) \qquad (2)$$

Using these constraints and the Lagrange multiplier technique to find the point of maximum entropy, Goldszmidt *et al.* [1993] derived the following elegant and simple approximation for the probability of each model :

$$P(m) \quad \sim \quad \prod_{\substack{r_i \\ m \models a_i \wedge \neg b_i}} \alpha_i \qquad (3)$$

where the α_i are related to the Lagrange multipliers for each rule. Making a further assumption that the α_i can also be approximated by a relative order of magnitude, thus writing $\alpha_i \sim \varepsilon^{\kappa(r_i)}$, the probability expressions (3) are substituted back into the constraints (2) yielding $|\Delta|$ simultaneous equations with $|\Delta|$ unknowns, the $\kappa(r_i)$.

In the limit as $\varepsilon \to 0$ those models with the lowest powers of ε will dominate, and the constraints reduce to:

$$\kappa(r_i) + \min_{m \models a_i \wedge \neg b_i} \sum_{\substack{r_j, j \neq i \\ m \models a_j \wedge \neg b_j}} \kappa(r_j) \quad = \quad s_i + \min_{m \models a_i \wedge b_i} \sum_{\substack{r_j, j \neq i \\ m \models a_j \wedge \neg b_j}} \kappa(r_j) \qquad (4)$$

Given this ranking, $\kappa(r_i)$, over defaults, the me-ranking over models, $\kappa(m)$, can be found using equation (3). The me-rank of each model is given by the sum of the me-ranks of those defaults it falsifies:

$$\kappa(m) = \sum_{\substack{r_i \\ m \models a_i \wedge \neg b_i}} \kappa(r_i) \qquad (5)$$

This completes the derivation of the maximum entropy ranking with $\kappa(m)$ defining the me-consequence relation. The following section looks at conditions under which the assumptions used to find equations (4) and (5) are valid.

3 Robustness of rankings

In the above analysis, it was assumed that it was only necessary to consider the asymptotic behaviour of the defaults, and that fixing the relative strength of defaults in this way uniquely determined the me-ranking. It turns out that while this is not true in general, it is true for a useful class of problems which this section characterises. As an example of a case in which the assumptions are not true, consider the following where the probabilities (3) are used to consider what may happen when all defaults have the same strength but their coefficients are allowed to vary.

Example 3.1

$$\Delta = \{r_1 : a \Rightarrow b, r_2 : a \Rightarrow c, r_3 : a \wedge b \Rightarrow c\}$$

The table shows whether a model falsifies or verifies each default and gives its (unnormalised) probability using equation (3):

m	a	b	c	r_1	r_2	r_3	$P(m)$
m_1	0	0	0	-	-	-	1
m_2	0	0	1	-	-	-	1
m_3	0	1	0	-	-	-	1
m_4	0	1	1	-	-	-	1
m_5	1	0	0	f	f	-	$\alpha_1\alpha_2$
m_6	1	0	1	f	v	-	α_1
m_7	1	1	0	v	f	f	$\alpha_2\alpha_3$
m_8	1	1	1	v	v	v	1

Using the substitution $u = \frac{\varepsilon}{1-\varepsilon}$, with all defaults having equal strength of 1, and letting their coefficients be c_1, c_2, c_3, respectively, the constraint equations (2) give rise to three simultaneous equations:

$$\begin{aligned} \alpha_1\alpha_2 + \alpha_1 &= c_1 u(1 + \alpha_2\alpha_3) \\ \alpha_1\alpha_2 + \alpha_2\alpha_3 &= c_2 u(1 + \alpha_1) \\ \alpha_2\alpha_3 &= c_3 u \end{aligned}$$

Solving these for the α_i in terms of u gives:

$$\begin{aligned} \alpha_1 &= \frac{u(c_1 + c_1c_3u - c_2 + c_3)}{1 + c_2u} \\ \alpha_2 &= \frac{c_1c_2u + c_1c_2c_3u^2 + c_2 - c_3}{c_1 + c_1c_2u - c_2 + c_3} \\ \alpha_3 &= \frac{c_3u(c_1 + c_1c_2u - c_2 + c_3)}{c_1c_2u + c_1c_2c_3u^2 + c_2 - c_3} \end{aligned}$$

Now consider what happens asymptotically for various values of the coefficients.

Case 1: Let $c_1 = 2(c_2 - c_3)$ (for $c_2 > c_3$). This gives a solution of $\alpha_1 \sim u$, $\alpha_2 \sim 1$, $\alpha_3 \sim u$ and leads to an me-ranking over defaults of (1,0,1). The corresponding me-ranking over models is given in the table below.

Case 2: Let $c_2 = c_3$. This gives a solution of $\alpha_1 \sim u$, $\alpha_2 \sim u$, $\alpha_3 \sim 1$, and an me-ranking over defaults of (1,1,0). The corresponding me-ranking is given in the table below.

Case 3: Let $c_1 + c_3 = c_2$. This gives a solution of $\alpha_1 \sim u^2$, $\alpha_2 \sim \frac{1}{u}$, $\alpha_3 \sim u^2$ and an me-ranking over defaults of (2,-1,2). The corresponding me-ranking is given in the table below.

m	a	b	c	$(1,0,1)$	$(1,1,0)$	$(2,-1,2)$
m_1	0	0	0	0	0	0
m_2	0	0	1	0	0	0
m_3	0	1	0	0	0	0
m_4	0	1	1	0	0	0
m_5	1	0	0	1	2	1
m_6	1	0	1	1	1	2
m_7	1	1	0	1	1	1
m_8	1	1	1	0	0	0

Different choices for the coefficients clearly lead to completely different me-rankings over the defaults and, more importantly, over the models. This is because there are multiple solutions to the non-linear simultaneous equations given by (4). In addition to having many solutions, these equations may have no solution at all if the strength assignments represent inconsistent probabilistic constraints. However, for maximum entropy entailment to be well-defined, it is desirable to be able to determine when a unique solution to these equations can be found. This is guaranteed whenever the me-ranking is robust.

Definition 3.2 *An integer ranking, κ, over models is said to be* robust [1] *with respect to a set of defaults, $\{r_i\}$, with associated strengths, $\{s_i\}$, if no two defaults share a common minimal falsifying model in κ.*

In the sequel, v_r (respectively, f_r) represent minimal verifying (respectively, falsifying) models of r in κ. Similarly, $v'_{r'}$ (respectively, $f'_{r'}$) represent minimal verifying (respectively, falsifying) models of r' in κ', and so on.

Definition 3.3 *An integer ranking, κ, over a set of defaults, $\{r_i\}$, with associated strengths, $\{s_i\}$, is said to be* me-valid *with respect to that set if it satisfies (5) and for all r*

$$\kappa(v_r) + s_r = \kappa(f_r) \tag{6}$$

Definition 3.4 *Two me-valid rankings, κ and κ', are said to be* distinct *iff $\kappa(r) \neq \kappa'(r)$ for some default r. Such a default is said to be* distinctly ranked.

The following lemma shows that any distinctly ranked default, r, which has minimal $\kappa(f_r)$ among distinctly ranked defaults, also has minimal $\kappa'(f'_r)$ among distinctly ranked defaults.

Lemma 3.5 *Given two distinct me-valid rankings, κ and κ', if r is such that $\kappa(r) \neq \kappa'(r)$ and for all r' with $\kappa(r') \neq \kappa'(r')$, $\kappa(f_{r'}) \geq \kappa(f_r)$, then $\kappa'(f'_{r'}) \geq \kappa'(f'_r)$.*

Proof. Suppose otherwise, that is, there exists $r' \neq r$, such that $\kappa(r') \neq \kappa'(r')$ with $\kappa(f_{r'}) \geq \kappa(f_r)$ but $\kappa'(f'_r) > \kappa'(f'_{r'})$. Without loss of generality, suppose that r' has minimal $\kappa'(f'_{r'})$ among distinctly ranked defaults. Now, because κ is me-valid, $\kappa(v_r) + s_r = \kappa(f_r)$, and v_r can only falsify defaults, s, for which $\kappa(s) = \kappa'(s)$, so that $\kappa(v_r) = \kappa'(v_r)$. It follows that

$$\kappa(f_r) = \kappa(v_r) + s_r = \kappa'(v_r) + s_r \geq \kappa'(v'_r) + s_r = \kappa'(f'_r) > \kappa'(f'_{r'}) \tag{7}$$

Similarly, since r' was chosen to have minimal $\kappa'(f'_{r'})$ among distinctly ranked defaults, $\kappa'(v'_{r'}) + s_{r'} = \kappa'(f'_{r'})$, and $v'_{r'}$ can only falsify defaults, s, for which $\kappa'(s) = \kappa(s)$, and $\kappa'(v'_{r'}) = \kappa(v'_{r'})$. It follows that

$$\kappa'(f'_{r'}) = \kappa'(v'_{r'}) + s_{r'} = \kappa(v'_{r'}) + s_{r'} \geq \kappa(v_{r'}) + s_{r'} = \kappa(f_{r'}) \geq \kappa(f_r) \tag{8}$$

Putting (7) and (8) together, $\kappa(f_r) \geq \kappa'(f'_r) > \kappa'(f'_{r'}) \geq \kappa(f_{r'}) \geq \kappa(f_r)$, which contradiction implies that $\kappa'(f'_{r'}) \geq \kappa'(f'_r)$, as required. •

Theorem 3.6 *Given a finite set of defaults, $\{r_i\}$, with associated strengths, $\{s_i\}$, if a robust me-valid ranking, κ, exists then it is unique.*

Proof. Let κ and κ' be distinct me-valid rankings and r be a distinctly ranked default with minimal $\kappa(f_r)$ among distinctly ranked defaults and, by Lemma 3.5, minimal $\kappa'(f'_r)$. Suppose that κ is robust. Then f_r falsifies only r and other defaults, s, with $\kappa(s) = \kappa'(s)$; also $\kappa(v_r) = \kappa'(v'_r)$ since they only falsify non-distinctly ranked defaults, and, since both κ and κ' are me-valid, it follows that $\kappa(f_r) = \kappa'(f'_r)$ with $\kappa(r) \neq \kappa'(r)$.

Consider $\kappa'(f_r)$ for which $\kappa'(f_r) \geq \kappa'(f'_r)$. But $\kappa'(f'_r) = \kappa(f_r)$ and f_r falsifies only non-distinctly ranked defaults and r itself, for which $\kappa(r) \neq \kappa'(r)$. Therefore $\kappa'(f_r) > \kappa'(f'_r)$ and hence $\kappa'(r) > \kappa(r)$.

Now, if f'_r falsified no other distinctly ranked default, $\kappa(f'_r) < \kappa'(f'_r) = \kappa(f_r)$, which contradicts f_r being minimal in κ. This implies that f'_r must falsify some other distinctly ranked defaults and hence κ' is not robust. Let these be $r_1, r_2, \ldots, r_n$; since all these r_i are also minimal distinctly ranked defaults in κ', by Lemma 3.5, they are also minimal in κ and there must exist $f_{r_1}, f_{r_2}, \ldots, f_{r_n}$, minimally ranked falsifying models in κ such that $\kappa(f_r) = \kappa(f_{r_i})$ for all r_i. Further, because κ is robust, none of the f_{r_i} can falsify any other distinctly ranked defaults.

But, by the same argument as above, this implies that for all r_i, $\kappa(r_i) < \kappa'(r_i)$. However, this in turn implies that f'_r which falsifies r, all the r_i, and non-distinctly ranked defaults, must have a lower rank than f_r in κ, i.e., $\kappa(f'_r) < \kappa'(f'_r) = \kappa(f_r)$, which contradicts f_r being the minimal falsifying model of r in κ. Hence, κ cannot be robust either. It follows that, if two distinct me-rankings exist, neither can be robust, and any robust me-valid ranking is unique. •

[1] Adopting the use of “robustness” to indicate existence of a unique solution from [Bacchus *et al.*, 1996].

Note that given two distinct rankings, κ and κ', it may still be the case that $\kappa(m) = \kappa'(m)$ for all m, i.e., the ranking over models may be unique despite there being multiple solutions for the $\kappa(r_i)$ to the constraint equations (5) and (6). For example, the set $\{r_1 : a \Rightarrow b, r_2 : \neg b \Rightarrow \neg a\}$, produces the two equations

$$\begin{aligned} \kappa(r_1) + \kappa(r_2) &= s_1 \\ \kappa(r_1) + \kappa(r_2) &= s_2 \end{aligned}$$

which have no solution unless $s_1 = s_2$ in which case there are an infinite number of solutions. However, all solutions lead to the same unique ranking over models. Refining the robustness condition and understanding its significance in such cases is the subject of ongoing research.

4 Computing the me-ranking

Using the robustness condition and equation (4), it is possible to determine the me-ranking over defaults one by one. Robustness guarantees that for at least one default the currently computed minimal ranks of models are indeed their genuine ranks in the me-ranking.

Let the function MINV(r) (respectively, MINF(r)) be defined so that it returns the rank of the current minimal verifying model of r (respectively, the rank of the current minimal falsifying model of r *excluding its own contribution*) using equation (5). Then equation (6) can be re-written as

$$\kappa(r) = s_r + \kappa(v_r) - (\kappa(f_r) - \kappa(r)) \tag{9}$$

which in the algorithm is used to assign the rank of a default using

$$\kappa(r) := s_r + \mathrm{MINV}(r) - \mathrm{MINF}(r) \tag{10}$$

```
Algorithm to compute me-ranking

Input:  a set of defaults, $\{r_i\}$, and associated
strength assignments, $\{s_i\}$.
Output:  the me-ranking, $\kappa$, or an exception if the
set is p-inconsistent, or if the robustness
condition is violated.

[1] Initialise all $\kappa(r_i) = \infty$.
[2] From all $r_i$ with $\kappa(r_i) = \infty$, find the minimal
    value of $s_i + \mathtt{MINV}(r_i)$ and select any $r_i$ for which
    this holds, say $r$.
[3] If $\mathtt{MINV}(r) = \infty$ then the input set is
    p-inconsistent.  Output an exception.
[4] Find $\mathtt{MINF}(r)$.
[5] If $\mathtt{MINF}(r) = \infty$ the robustness condition
    is violated.  Output an exception.
[6] Let $\kappa(r) := s_r + \mathtt{MINV}(r) - \mathtt{MINF}(r)$.
[7] If any $\kappa(r_i) = \infty$ goto step 2.
[8] Assign ranks to models using equation (5).
[9] Validate the ranking by ensuring both that the
    constraints (4) and that the robustness
    condition are satisfied.  Output either the
    me-ranking or an exception.
```

This algorithm clearly terminates at step 3, if the input set is probabilistically inconsistent, or at step 5, or at step 9. Termination does not guarantee that a valid ranking has been found but this is checked for and reported at step 9. The following theorem proves that, provided the robustness condition is satisfied, the algorithm will compute the unique me-ranking. That the algorithm works given certain pre-conditions can be verified if the two ranks in the assignment (10) can be shown to be valid. This requires that the ranks selected for MINV(r) and MINF(r) when the assignment is made are indeed the minimal ranks for r.

Theorem 4.1 *Given a finite set of defaults, $\{r_i\}$, with associated strengths, $\{s_i\}$, the algorithm computes the unique me-ranking, κ, if it is robust.*

Proof. The theorem is proved by induction. On the first pass of the loop no rules have been ranked and so the ranks of each rule ranked (i.e. none) are correct. The inductive hypothesis assumes that at the nth pass of the loop all rules ranked in the previous $(n-1)$ passes have been assigned their correct me-ranks. Consider that on the nth pass of the loop, rule r, with minimal $s_i + \mathrm{MINV}(r_i)$, is selected to be ranked.

Let v_c be a verifying model of r such that $\kappa(v_c) = \mathrm{MINV}(r)$. Suppose that v_c is not a minimal verifying model of r, so there exists v_r, such that $\kappa(v_r) < \kappa(v_c)$. Now, $\kappa(v_r) < \mathrm{MINV}(r)$, the computed minimal verifying rank for r, so it must be the case that v_r falsifies some rule, $r' \neq r$, which has not yet been ranked, and since r' was not selected to be ranked in this pass of the loop it follows that

$$s_r + \mathrm{MINV}(r) \leq s_{r'} + \mathrm{MINV}(r')$$

Then, since v_r falsifies r', $\kappa(f_{r'}) \leq \kappa(v_r)$, in particular, using (6)

$$\begin{aligned} s_{r'} + \kappa(v_{r'}) = \kappa(f_{r'}) < s_r + \kappa(v_r) <& \\ s_r + \mathrm{MINV}(r) \leq s_{r'} + \mathrm{MINV}(r')& \end{aligned}$$

so that $\kappa(v_{r'}) < \mathrm{MINV}(r')$. It follows that $v_{r'}$, too, must falsify some rule, $r'' \neq r' \neq r$, which has not yet been ranked. Then, since $v_{r'}$ falsifies r'', $\kappa(f_{r''}) \leq \kappa(v_{r'})$. Continuing in this way, an infinite descending chain of distinct unranked rules is constructed. This contradicts the finite size of the original default set, and therefore v_c must be a minimal verifying model of r.

Let f_c be a falsifying model of r such that $\kappa(f_c) = \kappa(r) + \mathrm{MINF}(r)$. Suppose that f_c is not a minimal falsifying model of r, so there exists f_r, such that $\kappa(f_r) < \kappa(f_c)$. Now, since $\kappa(f_r) - \kappa(r) < \mathrm{MINF}(r)$, the computed minimal falsifying rank for r, it must be the case that f_r falsifies some rule, $r' \neq r$, which has not yet been ranked, and since r' was not selected to be ranked in this pass of the loop it follows that

$$s_r + \mathrm{MINV}(r) \leq s_{r'} + \mathrm{MINV}(r')$$

Now, f_r falsifies r', and under the assumption that the robustness condition holds, no two defaults share a common minimal falsifying model in the me-ranking. Therefore, $\kappa(f_{r'}) < \kappa(f_r)$, and the following inequality holds

$$s_{r'} + \kappa(v_{r'}) = \kappa(f_{r'}) < \kappa(f_r) = s_r + \mathrm{MINV}(r) \leq s_{r'} + \mathrm{MINV}(r')$$

so that $\kappa(v_{r'}) < \mathrm{MINV}(r')$, the computed minimal verifying rank for r'. It follows that $v_{r'}$, too, must falsify some rule, $r'' \neq r' \neq r$, which has not yet been ranked. Then, since $v_{r'}$ falsifies r'', $\kappa(f_{r''}) \leq \kappa(v_{r'})$. Continuing in this way, an infinite descending chain of distinct unranked rules is constructed. This contradicts the finite size of the original default set and therefore f_c must be a minimal falsifying model of r.

Given that for the selected rule, r, the values $\mathrm{MINV}(r)$ and $\mathrm{MINF}(r)$ calculated at this pass of the loop represent the me-ranks of its minimal verifying and falsifying models (excluding its own contribution), respectively, it follows that the assignment

$$\kappa(r) := s_r + \mathrm{MINV}(r) - \mathrm{MINF}(r) \qquad (11)$$

is valid and r is assigned its correct me-rank. The theorem follows by induction. •

5 Examples

In the first example, the solution is tabulated explictly to illustrate the method of finding the me-ranking but later this is omitted to save space.

Example 5.1 (Exceptional inheritance)

$$\Delta = \{r_1 : b \Rightarrow f, r_2 : p \Rightarrow b, r_3 : p \Rightarrow \neg f, r_4 : b \Rightarrow w\}$$

The intended interpretation of this knowledge base is that birds fly, penguins are birds, penguins do not fly and birds have wings; each r_i has strength s_i. The table shows whether a model falsifies or verifies each default. The column headed $\kappa(m)$ gives the me-rank of each model in terms of the $\kappa(r_i)$ using equation (5).

m	b	f	p	w	r_1	r_2	r_3	r_4	$\kappa(m)$
m_1	0	0	0	0	-	-	-	-	0
m_2	0	0	0	1	-	-	-	-	0
m_3	0	0	1	0	-	f	v	-	$\kappa(r_2)$
m_4	0	0	1	1	-	f	v	-	$\kappa(r_2)$
m_5	0	1	0	0	-	-	-	-	0
m_6	0	1	0	1	-	-	-	-	0
m_7	0	1	1	0	-	f	f	-	$\kappa(r_2) + \kappa(r_3)$
m_8	0	1	1	1	-	f	f	-	$\kappa(r_2) + \kappa(r_3)$
m_9	1	0	0	0	f	-	-	f	$\kappa(r_1) + \kappa(r_4)$
m_{10}	1	0	0	1	f	-	-	v	$\kappa(r_1)$
m_{11}	1	0	1	0	f	v	v	f	$\kappa(r_1) + \kappa(r_4)$
m_{12}	1	0	1	1	f	v	v	v	$\kappa(r_1)$
m_{13}	1	1	0	0	v	-	-	f	$\kappa(r_4)$
m_{14}	1	1	0	1	v	-	-	v	0
m_{15}	1	1	1	0	v	v	f	f	$\kappa(r_3) + \kappa(r_4)$
m_{16}	1	1	1	1	v	v	f	v	$\kappa(r_3)$

Substituting the $\kappa(m)$ into the reduced constraint equations (4) gives rise to:

$$\begin{aligned} \kappa(r_1) &= s_1 \\ \kappa(r_2) &= s_2 + \min(\kappa(r_1), \kappa(r_3)) \\ \kappa(r_3) &= s_3 + \min(\kappa(r_1), \kappa(r_2)) \\ \kappa(r_4) &= s_4 \end{aligned}$$

Clearly, the only solution to these equations is $\kappa(r_1) = s_1$, $\kappa(r_2) = s_1 + s_2$, $\kappa(r_3) = s_1 + s_3$, and $\kappa(r_4) = s_4$.

To determine default consequences it is necessary to compare the ranks of a default's minimum verifying and falsifying models. Since the solution holds for any strength assignment (s_1, s_2, s_3, s_4), it follows that some default conclusions may hold in general. In particular, it can be seen that the default $p \wedge b \Rightarrow \neg f$ is me-entailed since

$$\begin{aligned} \kappa(p \wedge b \wedge \neg f) &< \kappa(p \wedge b \wedge f) \\ s_1 &< s_1 + s_3 \end{aligned}$$

This result is unsurprising since $p \wedge b \Rightarrow \neg f$ is a preferential consequence of Δ. A more interesting general conclusion is $p \Rightarrow w$, which follows since

$$\kappa(p \wedge w) = s_1 < \kappa(p \wedge \neg w) = s_1 + \min(s_2, s_4) \qquad (12)$$

Again this result holds regardless of the strength assignments and illustrates that, for this example, the inheritance of w to p via b is uncontroversial. •

Example 5.2 (Nixon diamond)

$$\Delta = \{r_1 : q \Rightarrow p, r_2 : r \Rightarrow \neg p\}$$

The intended interpretation is that quakers are pacificists whereas republicans are not pacifists. Given a strength assignment of (s_1, s_2) is easily shown that $\kappa(r_1) = s_1$ and $\kappa(r_2) = s_2$. The classical problem associated with this knowledge base is to ask whether Nixon, being a quaker and a republican, is pacifist or not. This is represented by the default $r \wedge q \Rightarrow p$. The two relevant models to compare are $r \wedge q \wedge p$ and $r \wedge q \wedge \neg p$ whose me-ranks in the general me-solution are

$$\kappa(r \wedge q \wedge p) = s_2 \quad \text{and} \quad \kappa(r \wedge q \wedge \neg p) = s_1 \qquad (13)$$

Clearly either $r \wedge q \Rightarrow p$ or $r \wedge q \Rightarrow \neg p$, or neither, may be me-entailed depending on the comparative strengths s_1 and s_2. This result is in accordance with the “intuitive” solution that no conclusion should be drawn regarding Nixon's pacifist status unless there is reason to suppose that one default holds more strongly than the other. In the case of one default being stronger, the conclusion favoured by the stronger would prevail. •

Example 5.3 (Royal elephants/marine chaplains)

$$\Delta = \{r_1 : a \Rightarrow b, r_2 : c \Rightarrow b, r_3 : b \Rightarrow d, r_4 : a \Rightarrow \neg d\}$$

There are two interpretations of this knowledge base. In the first, the propositions a, b, c, and d, stand for royal, elephant, african and grey, respectively; in the second, the propositions stand for chaplain, man, marine and beer drinker, respectively. The constraint equations (4) give rise to:

$$\begin{aligned} \kappa(r_1) &= s_1 + \min(\kappa(r_3), \kappa(r_4)) \\ \kappa(r_2) &= s_2 \\ \kappa(r_3) &= s_3 \\ \kappa(r_4) &= s_4 + \min(\kappa(r_1), \kappa(r_3)) \end{aligned}$$

which have the unique solution $\kappa(r_1) = s_1 + s_3$, $\kappa(r_2) = s_2$, $\kappa(r_3) = s_3$, and $\kappa(r_4) = s_3 + s_4$.

The key question relating to this knowledge base is "Are elephants which are both royal and african, not grey?", or alternatively, "Don't marine chaplains drink beer?" This translates into the default $a \wedge c \Rightarrow \neg d$ which is me-entailed in general as can be seen from:

$$\begin{array}{rcl} \kappa(a \wedge c \wedge \neg d) & < & \kappa(a \wedge c \wedge d) \\ s_3 & < & s_3 + \min(s_4, s_1 + s_2 + s_3) \end{array}$$

The result in this example is unambiguous, that is, it holds for all strength assignments[2]. However, [Touretzky *et al.*, 1987] were not entirely happy about the conclusion that marine-chaplains do not drink beer. They argued that if the rate of beer drinking amongst marines was significantly higher than normal, then this might alter the behaviour associated with marine-chaplains.

Now, the default $r_5 : c \Rightarrow d$ (marines drink beer) is in fact me-entailed by Δ, but adding it to the database with all defaults having equal strength violates the robustness condition. If, however, r_5 were added with a higher strength, so that it represented a new constraint for the purposes of maximising entropy, a robust solution would result and the status of the default $a \wedge c \Rightarrow \neg d$ would depend on the relative strengths s_4 and s_5.

So, Touretzky *et al.* were correct to suppose that if marines were heavier drinkers than men in general then it may not be clear whether marine chaplains are beer drinkers or not. However, it seems they were expecting too much of a default reasoning mechanism (a path-based inheritance reasoner in their case) in assuming it could draw such conclusions since this would involve using information *which it had never been told.* •

It is interesting to note that many of the more complex examples from the literature (for example, see [Makinson and Schlechta, 1991]), which have been devised deliberately to overcome any intuitive biases, fail to satisfy the robustness condition when all defaults are assigned equal strengths. If a set is probabilistically consistent it is usually possible to restore robustness by altering the strengths. This suggests that some sets may be too complex for the human intuition to disentangle because they are ambiguous or underspecified. By requiring more information from the knowledge engineer, in terms of a strength assignment over defaults, some of these ambiguities can be cleared up and the hitherto implicit biases made explicit.

6 Conclusions

This paper has refined and extended the work of Goldszmidt *et al.* [1993] on applying the principle of maximum entropy to probabilistic semantics for default rules to enable it to be applied to a much wider class of default sets. A new algorithm was presented which finds the maximum entropy ranking for a set of variable strength defaults that satisfy a sufficient condition for a unique solution to exist. The output is a consequence relation based on a total ordering of models—a *rational consequence relation* in the sense of Lehmann and Magidor [Lehmann and Magidor, 1992]. Some extreme technical cases remain to be investigated.

Using the me-approach for default reasoning provides the same benefits as its use in statistical problems. As Jaynes [1979] suggests, by encoding all known relevant information and finding the maximum entropy distribution, any observations which differ significantly from the result imply that other constraints, in this case defaults, exist. A closer approximation is obtained by adding more defaults or by adjusting the strengths. Instead of questioning the conclusions of a default reasoning system, one should ensure that all relevant information has been encoded — the maximum entropy formalism enables the precise and explicit representation of this as default knowledge. The main disadvantage of the me-approach is its intractability, however, this extension to arbitrary sets has shed some light onto the causes of controversy among classical examples from the literature and pointed to ways of resolving them.

References

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

[Bacchus *et al.*, 1996] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. From statistical knowledge bases to degrees of belief. *Artificial Intelligence*, 87:75–143, 1996.

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

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

[Lehmann and Magidor, 1992] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

[Makinson and Schlechta, 1991] D. Makinson and K. Schlechta. Floating conclusions and zombie paths: two deep difficulties in the "directly skeptical" approach to defeasible inheritance nets. *Artificial Intelligence*, 48:199–209, 1991.

[Touretzky *et al.*, 1987] D. S. Touretzky, J. F. Horty, and R. H. Thomason. A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 476–482, 1987.

[2] In fact all these examples have general solutions since they are minimal core sets as defined by Goldszmidt *et al.* [1993].