Extending the Maximum Entropy Approach to Variable Strength Defaults

Rachel A. Bourne [a,*] Simon Parsons [b]

[a] *Department of Electronic Engineering,*
Queen Mary, University of London,
London E1 4NS, UK
E-mail: r.a.bourne@elec.qmw.ac.uk
[b] *Department of Computer Science,*
University of Liverpool,
Liverpool L69 7ZF, UK
E-mail: s.d.parsons@csc.liv.ac.uk

A generalisation of the maximum entropy (ME) approach to default reasoning [7,8] to cater for variable strength defaults is presented. The assumptions on which the original work was based are reviewed and revised. A new algorithm is presented that is shown to compute the ME-ranking under these more general conditions. The limitations of the revised approach are discussed and a test for the uniqueness of the ME-solution is given. The ME-solutions to several illustrative examples of default reasoning are given, and the approach is shown to handle them appropriately. The conclusion is that the ME-approach can be regarded as providing a benchmark theory of default reasoning against which default intuitions and other default systems may be assessed.

Keywords: nonmonotonic reasoning, default reasoning, consequence relations

1. Introduction

The general requirements of default reasoning have mainly been laid down with the help of illustrative examples that demonstrate behaviours such as respect for specificity, inheritance to exceptional subclasses and maintenance of ambiguity. While there is consensus regarding the most basic requirements—preferential reasoning [10] is accepted as core behaviour for nonmonotonic reasoning systems

[*] Corresponding author.

[6,17]—there is no general theory that provides a satisfactory formalisation of what underlies the default intuitions themselves. Although some default systems have captured the required behaviours, e.g., lexicographic entailment [11], there has been little objective justification of the reasons behind them.

This paper aims to take a step towards the development of such a general theory by extending an existing approach [8] that does have a well-established foundation but which was previously limited in its applicability. The ε-semantics for defaults [1,16], which exhibits the core behaviour cited above, is grounded in probability theory—the most common tool for reasoning under uncertainty. The ε-consequences of a set of defaults are those satisfied in all probability distributions compatible with that set. By selecting just one of those compatible distributions using a well-known principle of indifference—maximising entropy—one arrives at an extension to the set of ε-consequences that can be shown to satisfy not only the core behaviour but other, more sophisticated default requirements. The principle of maximum entropy is to select the probability distribution that contains the most uncertainty while still satisfying the original defaults, and this method leads to the unique least committed, or least biased, distribution among all compatible ones [18]. The use of this principle has been shown to be the only consistent method of inductive inference under uncertainty [15]. By extending the ME-approach to handle differing priorities or strengths among a set of defaults, a new and more general algorithm for computing the ME-solution is given and the implications of using this extended framework are discussed.

The paper is organised as follows: section 2 gives some preliminary definitions and notation; section 3 reviews the original work on the ME-approach and compares it with the work presented here; section 4 gives a derivation of the equations that constrain the ME-ranking; section 5 presents the algorithm along with proof of its correctness under the stated assumptions; section 6 details a condition for the uniqueness of the solution obtained using the algorithm; section 7 discusses how to determine whether the assumptions made are valid, and what to do if they are not; section 8 gives the ME-solutions to several illustrative examples of default reasoning; and, finally, section 9 concludes by arguing for the adoption of the revised ME-approach as a general theory of default reasoning that can be used as a benchmark for evaluating both default intuitions and other default reasoning systems. This work builds on results originally reported in [3].

2. Preliminaries

First some preliminary definitions and notation are given. A finite propositional language $\mathcal{L}$ is made up of propositions a, b, c, ... and the usual connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$. A *default rule*, e.g., $a \Rightarrow b$, is a pair of propositions or formulæ joined by a new default connective $\Rightarrow$, which should not be confused with material implication $\rightarrow$. The language $\mathcal{L}$ has a finite set of models, $\mathcal{M}$. A model, m, is said to *verify* a default, $a \Rightarrow b$, if $m \models a \wedge b$, where $\models$ is classical entailment, and is said to *falsify* it if $m \models a \wedge \neg b$.

The ε-semantics equates each default with a probabilistic constraint such that $a \Rightarrow b$ means that $P(\neg b|a) < \varepsilon$ for some small, real $\varepsilon > 0$. A set of defaults, Δ, is ε-consistent if at least one probability distribution exists that satisfies these constraints for all defaults in Δ. A default $a \Rightarrow b$ is ε-entailed by Δ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $P(\neg b|a) < \varepsilon$ if $P(\neg b_i|a_i) < \delta$ for all $a_i \Rightarrow b_i \in \Delta$. The set of all defaults ε-entailed by Δ is equivalent to its preferential closure [10].

A ranking function, κ, over the models of $\mathcal{M}$ is an assignment of non-negative integer ranks such that for at least one $m \in \mathcal{M}$, $\kappa(m) = 0$. A formula, a, takes the rank of its minimal satisfying model(s), $\kappa(a) = \min_{m \models a}[\kappa(m)]$. A default, $a \Rightarrow b$, is entailed by a ranking function, κ, iff the rank of its minimal verifying model is strictly lower than that of its minimal falsifying model. That is:

$$\mid\!\sim_{\kappa} a \Rightarrow b \qquad \text{iff} \qquad \kappa(a \wedge b) < \kappa(a \wedge \neg b)$$

Ranking functions induce rational consequence relations [12]. A ranking function is *admissible* with respect to a set of defaults if it entails all defaults in that set; the rational consequence relation associated with an admissible ranking function for a given set of defaults is an extension of the preferential closure of that set. This paper examines the rational consequence relations induced by ME-rankings, i.e., those obtained by applying the principle of maximum entropy to the ε-semantics.

3. Comparison with original work

The original definition of the ME-approach to default reasoning was given by Goldszmidt *et al.* [8]. Their approach involved finding the probability distribution that maximised entropy subject to default constraints given by the ε-semantics. That is, each default, $a_i \Rightarrow b_i$, is constrained by:

$$P(b_i|a_i) \geq 1 - \varepsilon \tag{1}$$

This leads to a unique ME-distribution for any ε-consistent set of defaults. By using ε as a parameter, and under certain restrictions, Goldszmidt *et al.* derive an algorithm that finds the ranking function abstraction of this ME-distribution for *minimal core sets* of defaults, which are guaranteed to satisfy their imposed restrictions. However, there are several reasons why this definition is not wholly satisfactory due mainly to the overly restrictive nature of the constraints and to the system's inability to represent priorities among defaults. These two shortcomings are now addressed in turn.

Firstly, if the aim of the ME-approach is to find a ranking function abstraction of a set of probability distributions, then the exact function of ε used to constrain a default in (1) is unnecessarily precise. As the following section will show, it is the *order of magnitude* of ε in these constraints that appears in the equations that determine the ME-solution[1]. It would be more appropriate for these constraints to be of the form:

$$P(b_i|a_i) \geq 1 - C_i\varepsilon \tag{2}$$

where C_i is some unspecified *convergence coefficient* that is allowed to vary from default to default. In fact, this is more in keeping with the definition of entailment in the ε-semantics, since the conditional probabilities of ε-entailed defaults are required to tend to zero at least at the same rate as ε, but not necessarily precisely as ε. In many cases, in particular for minimal core sets, such a relaxation of the constraints does not lead to any difference in the ME-ranking. This means that someone specifying a set of defaults need only commit himself to all defaults being constrained to the same order of magnitude and this is enough to determine the ME-solution.

However, there are cases for which this imprecision causes multiple ME-solutions to occur. Consider the following example:

Example 1.

$$\Delta_1 = \{a \Rightarrow b, a \Rightarrow c\} \qquad \Delta_2 = \{a \Rightarrow b, a \wedge b \Rightarrow c\}$$

Both these sets are minimal core. Table 1 indicates the models that verify and falsify each of these defaults, and gives the ME-rankings for Δ_1 and Δ_2. It is easily seen that Δ_1 ME-entails $a \wedge b \Rightarrow c$, and that Δ_2 ME-entails $a \Rightarrow c$. The

[1] As will subsequently become clear, the term ME-solution refers to both the ME-ranking and a special ranking over the defaults themselves.

Table 1
The ME-rankings for Δ_1 and Δ_2

m	a	b	c	$a \Rightarrow b$	$a \Rightarrow c$	$a \wedge b \Rightarrow c$	Δ_1	Δ_2
m_1	0	0	0	-	-	-	0	0
m_2	0	0	1	-	-	-	0	0
m_3	0	1	0	-	-	-	0	0
m_4	0	1	1	-	-	-	0	0
m_5	1	0	0	f	f	-	2	1
m_6	1	0	1	f	v	-	1	1
m_7	1	1	0	v	f	f	1	1
m_8	1	1	1	v	v	v	0	0

question arises: what is the ME-solution for the combined set $\Delta_3 = \{a \Rightarrow b, a \Rightarrow c, a \wedge b \Rightarrow c\}$? Since the set Δ_3 is non-minimal core, it is not possible to apply the algorithm given in [8]. The original definition of the ME-approach dictates that just one ME-ranking exists, for any set of defaults; however, it is not clear which, if either, of the above rankings represents the ME-solution for Δ_3.

This situation arises because sets of defaults may contain redundant information in the sense that some defaults are already ME-entailed by the other defaults. Since a redundant default is already satisfied in the ME-distribution, the constraint associated with it will not have any effect; in fact, the constraint will be satisfied as a strict inequality. The constraints of active defaults are all satisfied as strict equalities in the ME-distribution.

In the example above, it is not clear which default in the set Δ_3 is causing the redundancy. This occurs because the convergence coefficients have been left imprecise: different values for these coefficients may alter which default constraints are active and which are redundant. Under the proposed new interpretation, both the ME-rankings are valid; the limits of the analysis of defaults using only orders of magnitude have been reached in such cases; only the user can decide which default should be treated as redundant. In other words, multiple ME-solutions arise because of the lack of precise specification of the convergence coefficients. However undesirable this may appear, it reflects the fact that the designer of the default set has specified an ambiguous situation according to the revised semantics; the ME-ranking critically depends on which default is redundant and, as it stands, such a default set represents two or more slightly different points of view. This can be equated with a situation in which one default is “explained away" by the others. Identifying default sets that contain redundancy turns out to be important for the interpretation of ME-solutions. However, it should be stressed that these situations are unlikely to be what is intended by the user of a default system, and can be resolved by adding priorities to defaults, which is now described.

The second shortcoming of the original ME-approach is that it is unable to represent different priorities among defaults. It is often useful to consider that some default holds more strongly than another, as a means of resolving ambiguity. In the case above, for example, it may be felt that one of the potentially redundant defaults in fact holds more strongly than its counterparts; perhaps, rather than removing it, it should be strengthened relative to them. Although this problem is recognised by Goldszmidt *et al.*, unfortunately, their original definition implies that any set of defaults can have just one ME-solution, regardless of any priorities the designer may feel exist or may wish to represent.

To overcome this, the ε-semantics can be extended to cater for *variable strength defaults*, or defaults that have been assigned specific strengths by the user. This is represented in the current framework by allowing defaults not only to have individual convergence coefficients, but also to converge at different rates of ε. Each variable strength default, $a_i \overset{s_i}{\Rightarrow} b_i$, is required to satisfy an *asymptotic constraint*:

$$P(b_i|a_i) \geq 1 - O_i(\varepsilon^{s_i}) \tag{3}$$

where $O_i(\varepsilon^{s_i})$ is some unspecified *convergence function* of ε that satisfies:

$$\lim_{\varepsilon \to 0} \frac{O_i(\varepsilon^{s_i})}{\varepsilon^{s_i}} = C_i \tag{4}$$

C_i being the convergence coefficient of r_i. Note that since ε is merely a parameter, the strengths associated with defaults are relative and have no objective meaning; a simple change of parameter could lead to any rational assignment of strengths.

Goldszmidt and Pearl show in [7] that adding strengths to defaults in this way does not affect the ε-consistency of a set. The constraint (3) translates quite naturally into the ranking function representation as:

$$s_i + \kappa(a_i \wedge b_i) \leq \kappa(a_i \wedge \neg b_i) \tag{5}$$

A ranking function that satisfies this constraint for all variable strength defaults in a set $\Delta^+ = \{a_i \overset{s_i}{\Rightarrow} b_i\}$ will be called ε^+-admissible with respect to that set for that particular assignment of strengths. Under this extended semantics, ranking functions can also be considered to entail a default to a certain degree, being the difference between its minimal verifying and falsifying models.

Since the constraints associated with the defaults in a given set can now be altered, by assigning different strengths, it turns out that the ME-solution with respect to that set also varies. In turn this leads to different ME-consequence relations also depending on the strengths assigned. This allows for greater expressiveness in specifying defaults and enables potentially ambiguous situations—including those involving redundancy—to be resolved by adjusting the strengths.

These changes to the ME-approach involve a shift in the commitment required of the user: while he must now specify the relative strengths of defaults, he need not be committed to any particular convergence coefficients. This fits more neatly into the ranking function representation of consequence relations, which uses ranks to represent degrees of disbelief—higher ranks imply lower degrees of belief [19,7]. What is, perhaps, remarkable about this new framework, is that in almost all cases, merely specifying the order of magnitude strengths is sufficient to determine a unique ME-solution; the only occasions when multiple ME-solutions can arise are those that contain redundancy, and even then only for a specific type of redundancy. This means, for example, that practically all illustrative examples of default reasoning have unique ME-rankings, and those that do not are invariably those that have caused "clashes of intuition" among researchers (see example 16 below). The related redundancy problem, which occurs for sets that contain redundancy through an underspecification of strength[2], is troublesome only insofar as the optimisation technique used to derive the ME-distribution requires that all defaults are active; clearly, an understrength default should be ignored, but this may not be recognised until a solution has been computed, and such a solution may therefore be invalid.

The problems created by redundancy are overcome by making an assumption that the problem is well-posed and that all constraints are active, i.e., that a given set of defaults has a unique ME-solution that satisfies the assigned strengths and that is independent of the convergence coefficients. The validity of this assumption will need to be verified after a solution has been found, and, if it does not hold, the ME-solution may need to be recomputed after removing any redundant defaults. This assumption will be called the *validity assumption* and

[2] That is, when the strength assigned to a default does not provide a constraint, since other defaults already constrain it to a greater degree.

how to assess whether it holds will be discussed in section 7.

The advantages of the revised ME-approach can be summarised as follows:

- Relaxation of the default constraints requires only order of magnitude commitments about default convergence from the user.
- The inputs to the system reflect the same type of knowledge as the outputs from it, i.e., variable strength defaults go in, and defaults plus degrees of entailment come out.
- Defaults can be prioritised using strength assignments that may allow more accurate modelling of default intuitions.
- In cases of ambiguity, the user can identify and remove potentially redundant defaults.

4. Deriving the maximum entropy ranking

The ME-ranking over the models of $\mathcal{L}$ is an asymptotic abstraction of the exponents of their probabilities in the ME-distribution. The basic idea is to find the ME-distribution for a fixed ε, and to consider what happens to it as $\varepsilon \to 0$. The assumption is made that all variables in this problem can be represented by expressions of the form $O(\varepsilon^s)$, that is, as $\varepsilon \to 0$ each variable asymptotically approaches some function of ε of some order s. In this way, the equations that determine the ME-distribution can be abstracted into integer equations that determine the ranks, or exponents of ε, for the variables in the ME-solution.

The ME-distribution is found using the Lagrange multiplier technique which can be used to optimise an objective function subject to a set of active constraints; in this case the entropy function is maximised subject to the conditional probability constraints associated with the defaults. The entropy of a probability distribution over a set of models, $\mathcal{M}$, is given by:

$$H[P] = -\sum_{m \in \mathcal{M}} P(m) \log P(m) \tag{6}$$

As discussed in the previous section, under the validity assumption, each default, r_i, satisfies an asymptotic constraint of the form:

$$P(b_i|a_i) = 1 - O_i(\varepsilon^{s_i}) \tag{7}$$

where the strengths, s_i, are specified for each default but the convergence functions, $O_i(\varepsilon^{s_i})$, may vary. The strengths, s_i, can be interpreted intuitively as representing relative priorities between defaults with numerically higher strength defaults holding more strongly than those of lower strength.

Given a set of variable strength defaults, $\Delta^+ = \{r_i : a_i \overset{s_i}{\Rightarrow} b_i\}$, the constraints (7) imposed on P for each default can be rewritten:

$$\sum_{m \models a_i \wedge \neg b_i} P(m) \quad - \quad \frac{O_i(\varepsilon^{s_i})}{1 - O_i(\varepsilon^{s_i})} \sum_{m \models a_i \wedge b_i} P(m) \quad = \quad 0 \tag{8}$$

Each constraint is multiplied by a Lagrange multiplier, λ_i, and added to the objective function, H, to give H':

$$H'[P] = -\sum_{m \in \mathcal{M}} P(m) \log P(m) + \sum_{r_i} \lambda_i \left[P(a_i \wedge \neg b_i) - \frac{O_i(\varepsilon^{s_i})}{1 - O_i(\varepsilon^{s_i})} P(a_i \wedge b_i) \right] \tag{9}$$

When the constraints are satisfied as equalities, the additional summands are effectively zero, in which case $H' \equiv H$. To find the point of maximum entropy subject to the constraints imposed, the function is differentiated with respect to each $P(m)$, and the derivative is set to zero; this gives $|\mathcal{M}|$ simultaneous equations of the form:

$$\frac{\partial H'[P]}{\partial P(m)} = -1 - \log P(m) + \sum_{\substack{r_i \\ \overline{m \models a_i \wedge \neg b_i}}} \lambda_i - \sum_{\substack{r_i \\ \overline{m \models a_i \wedge b_i}}} \frac{O_i(\varepsilon^{s_i})}{1 - O_i(\varepsilon^{s_i})} \lambda_i = 0 \tag{10}$$

where the first sum ranges over those defaults that m falsifies and the second over those that it verifies. Note that there is another constraint on P, since it is a probability distribution, that requires it to sum to one. However, this would merely be represented by some normalisation factor, common to each model's probability, so it can be safely ignored; the distribution found will in fact represent the unnormalised ME-distribution.

Introducing the substitution $\alpha_i = e^{\lambda_i}$, and taking antilogs of (10), yields expressions for the probabilities of each model in terms of the α_i and the $O_i(\varepsilon^{s_i})$:

$$P(m) = e^{-1} \prod_{\substack{r_i \\ \overline{m \models a_i \wedge \neg b_i}}} \alpha_i \prod_{\substack{r_i \\ \overline{m \models a_i \wedge b_i}}} \alpha_i^{-\frac{O_i(\mathcal{E}^{s_i})}{1 - O_i(\mathcal{E}^{s_i})}} \tag{11}$$

This analytic solution for the probability of each model in the unnormalised ME-distribution contains two unknowns for each default: α_i, associated with the Lagrange multipler, λ_i, and $O_i(\varepsilon^{s_i})$, the convergence function for r_i. By finding a solution for the α_i, the probabilities of each model can be determined from

(11). Assuming that all these variables are of the form $O(\varepsilon^s)$, introduce the substitutions:

$$\alpha_i = O_{r_i}(\varepsilon^{\phi(r_i)}) \qquad\qquad P(m) = O_m(\varepsilon^{\kappa(m)})$$

where $\phi(r_i)$ and $\kappa(m)$ represent the integer ranks (i.e., the exponents) of the defaults and of the probability of each model in the ME-solution, respectively. Note that, under these assumptions, the constant factor e^{-1} and the second product in (11) both represent functions of order zero[3], and can therefore be replaced by a function c_m that will tend to a constant as $\varepsilon \to 0$. The expression for the probability of each model (11) reduces to:

$$O_m(\varepsilon^{\kappa(m)}) = c_m \prod_{\frac{r_i}{m \models a_i \wedge \neg b_i}} O_{r_i}(\varepsilon^{\phi(r_i)}) \tag{12}$$

which, by comparing exponents on both sides of the equation, reduces to the integer equation:

$$\kappa(m) = \sum_{\frac{r_i}{m \models a_i \wedge \neg b_i}} \phi(r_i) \tag{13}$$

Under these same assumptions and substitutions, the constraint equations (8), reduce to the integer equations:

$$\min_{m \models a_i \wedge \neg b_i} [\kappa(m)] = s_i + \min_{m \models a_i \wedge b_i} [\kappa(m)] \tag{14}$$

The solutions to this system of equations, (13) and (14), take the form of functional mappings from defaults to integers, $\Phi = \{\phi(r_i)\}$, such that the ranking over models, κ, determined by Φ using (13), satisfies (14) for each default. Any integer mapping, Φ, that satisfies this condition for a particular set of variable strength defaults, Δ^+, will be called a *solution-set* for Δ^+.

Several remarks need to be made about these solutions. Firstly, although a solution-set, Φ, uniquely determines the ranking over models, κ, the inverse relationship may be one-to-many, i.e., the same ranking over models may be determined by many different solution-sets. However, it is often the case that there is just one solution-set giving rise to a unique ranking over models. Secondly, there may be many rankings, κ, that satisfy (14); since the constraints only include the minimal verifying and falsifying models of defaults, non-minimal models are unconstrained; such rankings may or may not be determined by a solution-set.

[3] Note that $f(\varepsilon, x, y) = \varepsilon^{-x\varepsilon^y} \to 1$ as $\varepsilon \to 0$ for fixed real x and fixed real $y > 0$.

Thirdly, there may be no rankings, κ, for which equations (14) are satisfiable, which means that no solution-sets exist. Finally, it is important to remember that solution-sets only lead to ME-rankings if the validity assumption holds; care is therefore needed when interpreting the solutions obtained. This will be discussed in more detail in section 7, after an algorithm for finding solution-sets is given.

The following simple example illustrates what an ME-solution may look like.

Example 2.

$$\Delta^{+} = \{r : a \overset{s}{\Rightarrow} b\}$$

Of the 4 models of $\mathcal{L}$, only one falsifies the default. The ME-ranks of models are therefore given by:

$$\begin{aligned}\kappa(a \wedge b) &= 0\\ \kappa(a \wedge \neg b) &= \phi(r)\\ \kappa(\neg a \wedge b) &= 0\\ \kappa(\neg a \wedge \neg b) &= 0\end{aligned}$$

There is just one constraint:

$$\kappa(a \wedge \neg b) = s + \kappa(a \wedge b)$$

which implies that:

$$\phi(r) = s$$

So the solution-set for Δ^{+} is $\{s\}$. The validity assumption holds trivially for a singleton default set. The ranking κ is therefore the unique ME-ranking for Δ^{+}. The default $\neg b \Rightarrow \neg a$ is ME-entailed by Δ^{+} to degree s.

5. The algorithm

This section presents an algorithm that can be used to compute a solution-set, Φ, and its corresponding ranking, κ, for a given set of variable strength defaults, Δ^{+}. The algorithm succeeds when a unique solution-set exists, but may fail when the set contains redundancy through being either underspecified or overspecified. However, it will be shown that the algorithm always computes an ε^{+}-admissible ranking over the set; furthermore, section 7 will discuss how to use the results produced by the algorithm to determine, and hence eliminate, the nature of the redundancy, should it be present.

Equations (13) and (14) represent a set of non-linear simultaneous equations, for which there is no guarantee that a solution exists, nor that a given solution is unique; since no general method exists to solve such equations, an algorithmic approach is taken. Before describing the algorithm itself, the key ideas behind how it works are explained.

Firstly, let v_r (respectively, f_r) represent a minimal verifying (respectively, falsifying) model of r in some ranking κ. Now, a ranking derived from a solution-set, Φ, has the form:

$$\kappa(f_r) = s_r + \kappa(v_r) \tag{15}$$

and since each falsifying model of a default contains a contribution from its own integer rank, $\phi(r)$, equation (15) can be rewritten as:

$$\phi(r) + (\kappa(f_r) - \phi(r)) = s_r + \kappa(v_r) \tag{16}$$

Now, if the integer ranks in the solution-set for defaults with lesser ranked minimal falsifying models were already known, equation (16) could be used to determine the value of $\phi(r)$. Expanding (16) gives:

$$\phi(r_i) + \min_{m \models a_i \wedge \neg b_i} \left[\sum_{\substack{r_j, j \neq i \\ \overline{m \models a_j \wedge \neg b_j}}} \phi(r_j) \right] = s_i + \min_{m \models a_i \wedge b_i} \left[\sum_{\substack{r_j, j \neq i \\ \overline{m \models a_j \wedge \neg b_j}}} \phi(r_j) \right] \tag{17}$$

The algorithm to compute a solution-set works as follows. Initially, each default is assigned an infinite integer rank. Two functions are defined, MINV(r) and MINF(r), that compute, respectively, the minimal rank of all verifying models of r, and the minimal rank of all falsifying models of r *excluding its own contribution*, using the current integer ranks of each default. The algorithm assigns new ranks to defaults, one by one, using these functions to determine which one to rank next, via the assignment:

$$\phi(r) := s_r + \mathrm{MINV}(r) - \mathrm{MINF}(r) \tag{18}$$

In this way, after $|\Delta^+|$ passes, all defaults will have been assigned appropriate integer ranks from which a ranking over models is then computed. Finally, each default is checked to determine whether equation (14) is satisfied as an equality (i.e., a solution-set has been found) or otherwise (i.e., only an ε^+-admissible ranking has been found).

Algorithm 3. Input: a set of variable strength defaults, $\Delta^+ = \{r_i : a_i \overset{s_i}{\Rightarrow} b_i\}$. Output: an ε^+-admissible ranking, κ, and if successful, a solution-set, Φ.

```
[1] Initialise all $\phi(r_i) = \mathrm{INF}$.
[2] While any $\phi(r_i) = \mathrm{INF}$ do:
    (a) For all $r_i$ with $\phi(r_i) = \mathrm{INF}$, compute $s_i + \mathrm{MINV}(r_i)$.
    (b) For all such $r_i$ with minimal $s_i + \mathrm{MINV}(r_i)$, compute $\mathrm{MINF}(r_i)$.
    (c) Select $r$ with minimal $\mathrm{MINF}(r_i)$.
    (d) If $\mathrm{MINF}(r) = \mathrm{INF}$ let $\phi(r) := 0$
        else let $\phi(r) := s + \mathrm{MINV}(r) - \mathrm{MINF}(r)$.
[3] Assign ranks to models using equation (13).
[4] If equations (14) are satisfied strictly for all defaults return $\Phi$ and $\kappa$;
    else return $\kappa$.
```

The remainder of this section will demonstrate that this algorithm either computes a solution-set, Φ, for Δ^+, or an ε^+-admissible ranking, κ, provided the set is ε-consistent. In the latter case, it will be shown that at least one default is assigned a zero rank; it is claimed that such a default is redundant and should be removed from the set.

The first lemma shows that the algorithm always assigns each default some finite rank.

Lemma 4. Given an ε-consistent set of variable strength defaults, the algorithm assigns a finite rank to each default.

Proof. Provided the minimal computed value for the function $\mathrm{MINV}(r)$ is finite at each pass of the loop, then the rank assigned to the chosen default will also be finite: zero, if the computed value of $\mathrm{MINF}(r)$ is infinite; and $s_r + \mathrm{MINV}(r) - \mathrm{MINF}(r)$, otherwise. Suppose therefore that at some pass of the loop the minimal computed value for $\mathrm{MINV}(r)$ is infinite for all unranked r. This means that all verifying models of each unranked default also falsify an unranked default; by

Thm 3.3 in [9], this contradicts the ε-consistency of the original set and hence each default will be assigned a finite rank. □

Given an ε-consistent set of defaults, therefore, some set of finite integer ranks, Φ, will be produced, which in turn implies a finite set of ranks over models, κ. The next lemma shows that κ represents a ranking function over models, i.e., that all ranks for models are non-negative and that at least one has zero rank.

Lemma 5. Given an ε-consistent set of variable strength defaults, the algorithm assigns a non-negative rank to each model.

Proof. This is shown by induction. The rank of each model at any given stage equals the sum of the current ranks of those defaults it falsifies. At the start, as all defaults have infinite rank, the current rank of a model is either zero, if it falsifies no defaults, or infinite. Moreover, since the set is ε-consistent, there exists at least one model which falsifies no defaults and therefore has rank zero. Assume that at some intermediate stage all models have non-negative rank before the chosen default, r, is assigned a rank. Now, if the computed value of $\mathrm{MINF}(r)$ is infinite, the default is assigned a rank of 0 but this will not change the current rank of any model since all its falsifying models also falsify other unranked defaults. If, on the other hand, $\mathrm{MINF}(r)$ is finite then $\phi(r) := s_r + \mathrm{MINV}(r) - \mathrm{MINF}(r)$. Let m be a falsifying model of r, and suppose that the current rank of m, without the contribution of r, is i. Clearly, $\mathrm{MINF}(r) \leq i$. Now the current rank of m is $\phi(r) + i = s_r + \mathrm{MINV}(r) - \mathrm{MINF}(r) + i \geq s_r + \mathrm{MINV}(r)$. By the inductive hypothesis, $\mathrm{MINV}(r)$ is non-negative. Therefore the rank of m is also non-negative. The lemma follows by induction. □

This lemma does not preclude defaults from having negative ranks, only models. Note that, at this stage, there is no guarantee that the computed ranking over models is admissible, only that it represents a ranking. The following lemma shows that the defaults are ranked in an order corresponding to the ascending order of their $s_r + \kappa(v_r)$ in the final ranking.

Lemma 6. Given an ε-consistent set of variable strength defaults, the algorithm assigns ranks to defaults in ascending order of the final ranks of their minimal verifying models plus their strengths.

Proof. Suppose r' is the next rule to be ranked after r. If $\mathrm{MINV}(r')$ is not affected by the ranking of r, then $s_r + \mathrm{MINV}(r) \leq s_{r'} + \mathrm{MINV}(r')$ by minimality in the selection of r. Otherwise, the minimal model of $\mathrm{MINV}(r')$ became finite during the ranking of r, and is therefore a falsifying model of r. By the proof of lemma 5, $s_r + \mathrm{MINV}(r) \leq \mathrm{MINV}(r')$ and hence $s_r + \mathrm{MINV}(r) < s_{r'} + \mathrm{MINV}(r')$. Thus $s_r + \mathrm{MINV}(r)$ increases monotonically for successive r. □

Corollary 7. κ is ε^+-admissible, that is, for all r

$$s_r + \kappa(v_r) \leq \kappa(f_r)$$

Proof. Note that all falsifying models of a default have infinite rank when it is being ranked and so cannot have a final rank of less than $s_r + \kappa(v_r)$. □

So the ranking produced by the algorithm is ε^+-admissible. Because the ranks of the models are computed from the ranks of the defaults, the equations (13) are guaranteed to be satisfied, although the same cannot necessarily be said for equations (14). However, the following lemma shows that, if for some default, equation (14) is satisfied as a strict inequality, that is, if $s_r + \kappa(v_r) < \kappa(f_r)$ in the computed ranking, then that default will have been assigned a rank of zero. In section 7, it is argued that these cases represent default sets containing redundancy, which need to be handled carefully since the assumption that all defaults are active is no longer valid.

Lemma 8. Given an ε-consistent set of variable strength defaults, if for some default, r, in the ranking computing by the ME algorithm $s_r + \kappa(v_r) < \kappa(f_r)$, then that default will have been assigned a rank of zero (i.e., $\phi(r) = 0$).

Proof. If the ranking computed by the ME algorithm, κ, is such that $s_r + \kappa(v_r) < \kappa(f_r)$ for some r, then, when r was selected to be ranked, it cannot be the case that $\mathrm{MINF}(r)$ was finite; if it were then the assignment $\phi(r_j) := s_j + \mathrm{MINV}(r_j) - \mathrm{MINF}(r_j)$ would mean that at least one falsifying model of r satisfied $s_r + \kappa(v_r) = \kappa(f_r)$ in the final ranking. Thus, since $\mathrm{MINF}(r)$ was infinite, r was assigned rank zero. □

Lemma 8 shows that, should the algorithm fail to find a solution-set, any defaults which satisfy (14) as a strict inequality will be assigned a zero rank. The resultant ranking will entail such a default to a degree greater than its assigned strength, indicating that the default is redundant, that is, it has no effect on the ranking produced. This type of redundancy is discussed further in section 7.

Even when a solution-set is discovered, care needs to be taken before claiming that the ranking it produces represents the ME-ranking. In order to establish whether or not the algorithm has found a solution-set that does lead to an ME-ranking, it is necessary firstly to check whether or not it is unique, and secondly, to verify the validity assumption. The following section highlights a sufficient condition that guarantees that a particular ranking corresponds to a unique solution-set and a discussion of how to use this is given in section 7.

6. Unique solution-sets

This section identifies a sufficient condition for a ranking, κ, determined by a solution-set, Φ, that guarantees the uniqueness of that solution-set. This is the initial step towards verifying that the validity assumption holds, since if there are multiple solution-sets it is unlikely that the assumption is valid, unless they all lead to the same ranking over models.

The following results will show that the uniqueness of a solution-set, Φ, can be guaranteed if the ranking, κ, determined by Φ, satisfies the following condition:

Definition 9. A ranking, κ, over models is said to be *robust* with respect to a set of variable strength defaults if no two defaults share a common minimal falsifying model in κ.

To demonstrate uniqueness of a solution-set it must be possible to distinguish between two arbitrary solution-sets:

Definition 10. Two solution-sets, Φ and Φ', are said to be *distinct* iff $\phi(r) \neq \phi'(r)$ for some default r. Such a default is said to be *distinctly ranked.*

The following notation will be useful. Let κ and κ' be rankings determined by solution-sets Φ and Φ', respectively, using (13). As before, let v_r, $v'_{r'}$ represent minimal verifying models of r, r' in κ, κ', respectively, and similarly let f_r, $f'_{r'}$ represent minimal falsifying models of r, r' in κ, κ', respectively, and so on. The following lemma is required for the main theorem; the lemma shows that any default, r, that is distinctly ranked in two solution-sets, Φ and Φ', and has minimal $\kappa(f_r)$ among distinctly ranked defaults, also has minimal $\kappa'(f'_r)$ among distinctly ranked defaults.

Lemma 11. Given two distinct solution-sets, Φ and Φ', with corresponding rankings, κ and κ'. If r is such that $\phi(r) \neq \phi'(r)$ and for all r' with $\phi(r') \neq \phi'(r')$, $\kappa(f_{r'}) \geq \kappa(f_r)$, then $\kappa'(f'_{r'}) \geq \kappa'(f'_r)$.

Proof. Suppose otherwise, that is, there exists $r' \neq r$, such that $\phi(r') \neq \phi'(r')$ 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 Φ is a solution-set, $s_r + \kappa(v_r) = \kappa(f_r)$, and v_r can only falsify defaults, d, for which $\phi(d) = \phi'(d)$, so that $\kappa(v_r) = \kappa'(v_r)$. It follows that:

$$\begin{aligned}\kappa(f_r) = s_r + \kappa(v_r) = s_r + \kappa'(v_r) \geq \\ s_r + \kappa'(v'_r) = \kappa'(f'_r) > \kappa'(f'_{r'})\end{aligned} \tag{19}$$

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

$$\begin{aligned}\kappa'(f'_{r'}) = s_{r'} + \kappa'(v'_{r'}) = s_{r'} + \kappa(v'_{r'}) \geq \\ s_r + \kappa(v_{r'}) = \kappa(f_{r'}) \geq \kappa(f_r)\end{aligned} \tag{20}$$

Putting (19) and (20) together, $\kappa(f_r) \geq \kappa'(f'_r) > \kappa'(f'_{r'}) \geq \kappa(f_{r'}) \geq \kappa(f_r)$, which by contradiction implies that $\kappa'(f'_{r'}) \geq \kappa'(f'_r)$, as required. □

The following theorem connects the robustness of a ranking with the uniqueness of its determining solution-set.

Theorem 12. Given a finite set of variable strength defaults, $\Delta^+ = \{r_i : a_i \overset{s_i}{\Rightarrow} b_i\}$, if a solution-set, Φ, determines a ranking, κ, that is robust, then Φ is the unique solution-set for Δ^+.

Proof. Let Φ and Φ' be distinct solution-sets for Δ^+ with corresponding rankings, κ and κ', and let r be a distinctly ranked default with minimal $\kappa(f_r)$ among distinctly ranked defaults and, by lemma 11, minimal $\kappa'(f'_r)$. Suppose further that κ is robust. Then f_r falsifies only r and other defaults, d, with $\phi(d) = \phi'(d)$; also $\kappa(v_r) = \kappa'(v'_r)$, since they only falsify non-distinctly ranked defaults, and since both Φ and Φ' are solution-sets; it follows that $\kappa(f_r) = \kappa'(f'_r)$ with $\phi(r) \neq \phi'(r)$.

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 $\phi(r) \neq \phi'(r)$. Hence $\kappa'(f_r) \neq \kappa(f_r)$ and so $\kappa'(f_r) \neq \kappa'(f'_r)$. Therefore $\kappa'(f_r) > \kappa'(f'_r)$ and hence $\phi'(r) > \phi(r)$.

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 11, 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, $\phi(r_i) < \phi'(r_i)$. 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 solution-sets exist, neither of their corresponding rankings can be robust, and hence any solution-set that determines a robust ranking must be unique. □

Theorem 12 provides a test for the uniqueness of a solution-set by testing for the robustness of the ranking determined by it. However, the robustness condition only gives a sufficient condition for uniqueness; it is possible that a non-robust ranking may be determined from a unique solution-set, although no examples have yet been found by the authors.

Given that there may be many solution-sets for a given set of defaults, one might be tempted to adapt algorithm 3 so as to find all possible solutions. There are several reasons why this is not recommended. Firstly, there may be an infinite number of solutions; at step [2](d) of the algorithm, assigning the value zero to the selected default is arbitrary, any integer would suffice. The value zero reflects our interpretation that such a default is a candidate—though not necessarily the only one—for removal from the set. Secondly, although there may be a choice of default to which to assign this zero rank, there is no guarantee that the outcome will be a solution-set, let alone that the ranking will represent the ME-ranking; computing many solution-sets does not alleviate the need to verify the validity assumption, indeed multiple solution-sets usually point to the fact that it has been violated. Thirdly, although it is useful to recognise the defaults that cause redundancy, our aim is to enable the design of unambiguous default knowledge bases along with their unique ME-interpretations; therefore, we concentrate of identifying redundancy and correcting it by the removal or strengthening of defaults.

7. Verifying the validity assumption

In the previous sections, equations (13) and (14) were derived and a method for finding solutions to them was given. The validity assumption clearly fails when no solution-set is found, since at least one default is not active; however, even when a solution-set has been found, it is still necessary to verify the validity assumption to determine whether the ranking produced is the unique ME-solution. How to establish whether it does indeed hold, and what to do when it does not, is the subject of this section.

The problem with trying to verify that the validity assumption holds is evident from its definition, which requires that *a given set of defaults has a unique* ME*-solution that satisfies the assigned strengths and that is independent of the convergence coefficients.* Thus the ME-solution is required in order to test the validity assumption. This circularity causes a problem since it is not possible to test a solution for validity until it has been verified that it is a valid ME-solution!

Under what circumstances may the validity assumption fail to hold? There are two possibilities, both of which are related to redundancy in the default information. The first type of redundancy occurs when, although all defaults are satisfied to their correct strengths, differences in their convergence coefficients can lead to different ME-solutions; in this situation there are multiple, equally valid ME-rankings that depend on which defaults are considered to be redundant. The second type of redundancy occurs when one default is satisfied in the ME-solution to a degree greater than its assigned strength; in this case solution-sets to equations (13) and (14), if they exist at all, may not even be valid ME-solutions.

To understand what is happening in these cases it is helpful to imagine the space of all probability distributions that is constrained by the defaults. Normally, the solution to the maximum entropy problem will occur at the edge of this space, i.e., with all constraints active, since by tightening a constraint a new solution needs to be found. However, in some cases, a constraint is already satisfied in the ME-solution given by the other defaults; tightening its constraint will have no effect until a critical point when it becomes active is reached. Redundant defaults should therefore be ignored when computing the ME-solution since their constraints are satisfied as strict inequalities and do not affect the solution.

For the first type of redundancy, the solution-set produced by the algorithm can be of some help: defaults with zero ranks usually point to multiple solutions. A further check is to test the resulting ranking for robustness, indicating whether or not the solution-set is unique. If the solution-set is not unique, it is not always clear which default is the redundant one. Clearly, any default with zero rank is a candidate, but others may also be; this reflects the fact that this type of redundancy is often ambiguous, and only the designer can know exactly which default he intends to be redundant. Once such a default has been removed, the algorithm (and verification) should be re-applied to the remaining defaults. Alternatively, if it is felt that all defaults ought to be active, a redundant default may be strengthened to allow it to represent an active constraint.

For the second type of redundancy, the algorithm may or may not find a solution-set. Failing to find a solution-set indicates clearly that some default has been assigned an inadequate strength, and is therefore redundant. However, in other cases, this type of redundancy may be harder to recognise; sometimes a solution-set exists, but because the validity assumption has failed, it does not represent an ME-solution. In this case, the assignment at step 2[d] of the algorithm is invalid; in fact, such a default will obtain a *negative* rank, and this is the way to identify these "improper" solution-sets. If a solution is found that contains a negative integer rank, the default should be removed and the algorithm re-applied to the remaining defaults. However, in the former case when no solution-set was found, it turns out that the ranking produced may well represent the ME-ranking. This is because any default which has a zero rank does not affect the ranking; by assigning a rank of zero to a default involved, it is being ignored as it should be, and the computed ε^+-admissible ranking does represent an ME-solution, provided no other redundancy is present.

It should be borne in mind that not only are cases of redundancy rare, but also default sets that contain redundancy are unlikely to be of much use in practice. Nevertheless these technical difficulties with the ME-solution are very real, and further work needs to be done to address them. For practical purposes, the following conjecture is used:

Conjecture. If a solution-set, Φ, for a given set of variable strength defaults,

Δ^+, is unique and contains only strictly positive integer ranks, then the ranking, κ, determined by Φ is the unique ME-ranking.

The truth of this conjecture remains an open research question; however, the following algorithm can be used to determine whether a given set of defaults contains either type of redundancy.

Algorithm 13. Input: $\Delta^+ = \{r_i : a_i \overset{s_i}{\Rightarrow} b_i\}$.
Output: $\Delta_{redundant}$ containing candidates for redundancy (if $\Delta_{redundant} = \emptyset$, Δ^+ contains no redundancy).

```
[1] Let $\Delta_{redundant} = \emptyset$.
[2] For each $r_i \in \Delta^+$.
    (a) Let $\Delta' = \Delta^+ - \{r_i\}$.
    (b) Let $\kappa'$ be the ranking produced by applying algorithm 3 to $\Delta'$.
    (c) If $\kappa'$ entails $r_i$ to degree $s_i$ or higher, $\Delta_{redundant} := \Delta_{redundant} \cup \{r_i\}$.
[3] Return $\Delta_{redundant}$.
```

If no redundancy exists, all defaults actively constrain the ME-distribution and hence it is unique. Thus, to establish that a set of defaults, Δ, does have a unique ME-solution, the ME-algorithm needs to be applied $|\Delta| + 1$ times.

Recent research by Eiter and Lukasiewicz [4] has shown that computing the ME-default ranking using the ME-algorithm is $\mathrm{FP}^{\mathrm{NP}}$-complete, while deciding ME-entailment is $\mathrm{NP}^{\mathrm{NP}}$-complete. This complexity result does not improve by restriction to Horn defaults; however, at least one tractable subset of defaults, termed feedback-free defaults, has been identified.

8. The ME-solutions to illustrative examples of default reasoning

This section gives the ME-solutions to some classic problems from default reasoning. It will be seen that not only do the ME-solutions produce the desired results but also help to clarify why some examples have led to disagreements among researchers. 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.

Table 2
The ME-ranking for the penguin example.

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	-	$\phi(r_2)$
m_4	0	0	1	1	-	f	v	-	$\phi(r_2)$
m_5	0	1	0	0	-	-	-	-	0
m_6	0	1	0	1	-	-	-	-	0
m_7	0	1	1	0	-	f	f	-	$\phi(r_2) + \phi(r_3)$
m_8	0	1	1	1	-	f	f	-	$\phi(r_2) + \phi(r_3)$
m_9	1	0	0	0	f	-	-	f	$\phi(r_1) + \phi(r_4)$
m_{10}	1	0	0	1	f	-	-	v	$\phi(r_1)$
m_{11}	1	0	1	0	f	v	v	f	$\phi(r_1) + \phi(r_4)$
m_{12}	1	0	1	1	f	v	v	v	$\phi(r_1)$
m_{13}	1	1	0	0	v	-	-	f	$\phi(r_4)$
m_{14}	1	1	0	1	v	-	-	v	0
m_{15}	1	1	1	0	v	v	f	f	$\phi(r_3) + \phi(r_4)$
m_{16}	1	1	1	1	v	v	f	v	$\phi(r_3)$

Example 14 Exceptional inheritance.

$$\Delta^+ = \{r_1 : b \overset{s_1}{\Rightarrow} f, r_2 : p \overset{s_2}{\Rightarrow} b, r_3 : p \overset{s_3}{\Rightarrow} \neg f, r_4 : b \overset{s_4}{\Rightarrow} w\}$$

The intended interpretation of this knowledge base is that birds fly, penguins are birds, penguins do not fly and birds have wings. Table 2 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 $\phi(r_i)$ using equation (13). Substituting the $\kappa(m)$ into the reduced constraint equations (14) gives rise to:

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

Clearly, the only solution to these equations is $\phi(r_1) = s_1$, $\phi(r_2) = s_1 + s_2$, $\phi(r_3) = s_1 + s_3$, and $\phi(r_4) = s_4$, all of which are strictly positive.

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

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

This result is unsurprising since $p \wedge b \Rightarrow \neg f$ is an ε-consequence of Δ^+, and so is bound to be satisfied in all ME-solutions. A more interesting conclusion is $p \Rightarrow w$, which follows since:

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

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 15 Nixon diamond.

$$\Delta^+ = \{r_1 : q \overset{s_1}{\Rightarrow} p, r_2 : r \overset{s_2}{\Rightarrow} \neg p\}$$

The intended interpretation is that quakers are pacifists whereas republicans are not pacifists. Given a strength assignment of (s_1, s_2) it is easily shown that $\phi(r_1) = s_1$ and $\phi(r_2) = s_2$. The classical problem associated with this knowledge base is to ask whether Nixon, being a republican and a quaker, 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 ME-solution are:

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

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 pacifism 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. This demonstrates the flexibility of the ME-approach in that it allows different default conclusions to be obtained depending on the strengths assigned.

As the examples so far have shown, under the ME-approach there are some conclusions that occur for any strength assignment and others that vary according to the strengths assigned to defaults. The fact that some default sets may

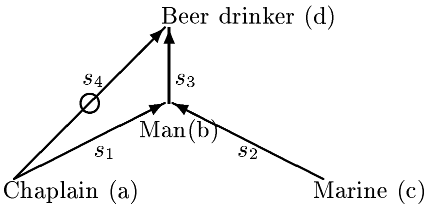

Figure 1. Marine chaplains (Δ).

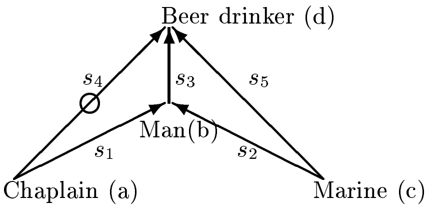

Figure 2. Marine chaplains (Δ').

sanction two opposite conclusions, i.e., a default and its converse, depending on the strengths assigned, is an interesting development for default reasoning. Historically, it was thought that there were “intuitively correct” outcomes which corresponded to commonsense reasoning but under this new ME-approach some conclusions depend critically on the strength assignment. Indeed, this is necessary if default sets like the Nixon diamond are going to be handled intuitively through prioritisation of defaults. The distinction between assignment-dependent ME-consequences and uncontroversial ones (i.e., those which hold under any strength assignment), may prove a useful way of explaining the disagreements among researchers regarding the more ambiguous, and less intuitively predictable, examples of default reasoning.

The following default set, an example that demonstrates multiple inheritance, is an extension of a well-known controversial example from the field of inheritance hierarchies. The original version appeared in several papers, and caused much debate [13,14,20].

Example 16 Marine chaplains.

$$\Delta = \{r_1 : a \overset{s_1}{\Rightarrow} b, r_2 : c \overset{s_2}{\Rightarrow} b, r_3 : b \overset{s_3}{\Rightarrow} d, r_4 : a \overset{s_4}{\Rightarrow} \neg d\}$$

The intended interpretation is that chaplains are men, marines are men, men are beer drinkers, and chaplains are not beer drinkers. The strength assignment (s_1, s_2, s_3, s_4) leads to a unique solution-set of $\phi(r_1) = s_1 + s_3$, $\phi(r_2) = s_2$, $\phi(r_3) = s_3$ and $\phi(r_4) = s_3 + s_4$.

The controversy surrounding this example, depicted graphically in figure 1, involves whether or not the default “Marine chaplains are not beer drinkers” ($a \wedge c \Rightarrow \neg d$) should be a default conclusion. The ME-ranks of the relevant models:

$$\kappa(a \wedge c \wedge \neg d) < \kappa(a \wedge c \wedge d)$$
$$s_3 < s_3 + \min(s_4, s_1 + s_2 + s_3)$$

show that in this, the original example, with no direct link from Marine to beer drinker, there is an uncontroversial ME-consequence of $a \wedge c \Rightarrow \neg d$ ("Marine chaplains are *not* beer drinkers"). This result should be unsurprising: the link between chaplain and beer drinker is clearly more specific than that to beer drinker from Marine via man. In other words, chaplains are known to be men who are known to be beer drinkers, and this fails to outweigh the direct link from "chaplain" to "not beer drinker"; the fact that a chaplain is also a Marine should not affect the conclusion that he does not drink beer; after all, Marines are only known to be beer drinkers by virtue of being men, at least as represented in the original problem.

However, Touretzky *et al.* [20] speculated that if Marines were known to be heavier drinkers than men in general, then this could affect the conclusion for Marine chaplains. To represent this, an extra default $r_5 : c \overset{s_5}{\Rightarrow} d$ is included, creating a direct link between Marines and beer drinkers (depicted graphically in figure 2).

$$\Delta' = \Delta \cup \{r_5 : c \overset{s_5}{\Rightarrow} d\}$$

Now this default, r_5, is already an ME-consequence of the original set and is ME-entailed to degree $\min(s_2, s_3)$. Table 3 shows whether a model falsifies or verifies each default, and the unknown integer ranks for each model are given in the final column according to equation (13). Substituting the $\kappa(m)$ into equations (14) gives rise to:

$$\begin{aligned}
\phi(r_1) &= s_1 + \min(\phi(r_3), \phi(r_4)) \\
\phi(r_2) &= s_2 \\
\phi(r_3) &= s_3 \\
\phi(r_4) &= s_4 + \min(\phi(r_1), \phi(r_3)) \\
\phi(r_5) &= s_5 - \min(\phi(r_2), \phi(r_3))
\end{aligned}$$

which, if $s_5 > \min(s_2, s_3)$, has a solution of $\phi(r_1) = s_1 + s_3$, $\phi(r_2) = s_2$, $\phi(r_3) = s_3$, $\phi(r_4) = s_3 + s_4$ and $\phi(r_5) = s_5 - \min(s_2, s_3)$. If, however, $s_5 < \min(s_2, s_3)$, the default r_5 is effectively redundant and the equations cannot all be solved as

Table 3
The ME-ranking for the Marine/chaplain example.

m	a	b	c	d	r_1	r_2	r_3	r_4	r_5	$\kappa(m)$
m_1	0	0	0	0	-	-	-	-	-	0
m_2	0	0	0	1	-	-	-	-	-	0
m_3	0	0	1	0	-	f	-	-	f	$\phi(r_2) + \phi(r_5)$
m_4	0	0	1	1	-	f	-	-	v	$\phi(r_2)$
m_5	0	1	0	0	-	-	f	-	-	$\phi(r_3)$
m_6	0	1	0	1	-	-	v	-	-	0
m_7	0	1	1	0	-	v	f	-	f	$\phi(r_3) + \phi(r_5)$
m_8	0	1	1	1	-	v	v	-	v	0
m_9	1	0	0	0	f	-	-	v	-	$\phi(r_1)$
m_{10}	1	0	0	1	f	-	-	f	-	$\phi(r_1) + \phi(r_4)$
m_{11}	1	0	1	0	f	f	-	v	f	$\phi(r_1) + \phi(r_2) + \phi(r_5)$
m_{12}	1	0	1	1	f	f	-	f	v	$\phi(r_1) + \phi(r_2) + \phi(r_4)$
m_{13}	1	1	0	0	v	-	f	v	-	$\phi(r_3)$
m_{14}	1	1	0	1	v	-	v	f	-	$\phi(r_4)$
m_{15}	1	1	1	0	v	v	f	v	f	$\phi(r_3) + \phi(r_5)$
m_{16}	1	1	1	1	v	v	v	f	v	$\phi(r_4)$

equalities. By assigning r_5 an integer rank of zero, the ME-ranking for the original problem is recovered. For the in-between case when $s_5 = \min(s_2, s_3)$, the ranking computed by the algorithm is non-robust, and there are multiple solution-sets indicating the presence of redundancy.

Looking only at cases for which a unique solution can be found, i.e., when the default r_5 is not redundant and does not cause multiple solutions, the conclusion regarding whether or not Marine chaplains are beer drinkers is indeed a controversial one. The minimal verifying and falsifying models of $a \wedge c \Rightarrow \neg d$ are:

$$\begin{aligned} \kappa(a \wedge c \wedge \neg d) &: \kappa(a \wedge c \wedge d) \\ s_3 + s_5 - \min(s_2, s_3) &: s_4 \end{aligned}$$

Clearly the default conclusion obtained from the ME-approach depends on the strengths s_2, s_3, s_4, s_5. It is therefore unsurprising that examples like this one have led to controversy—multiple inheritance is bound to lead to ambiguous situations[4]. Indeed, in some ways this can be seen as an extended and more

[4] After all, look at the problems this concept has caused in object-oriented programming languages (see [5], p.77, for example).

complex case of what occurs in the Nixon diamond.

This example has demonstrated that the ME-approach can be used to clarify the ambiguities that arise in multiple inheritance situations, and, at the same time, it can help to identify both the causes of controversy and how to resolve them.

9. Conclusion

This paper has introduced a refinement on the maximum entropy approach to default reasoning. By making slightly different assumptions from those of Goldszmidt [7,8], in particular, by requiring the user to specify the order of magnitude at which defaults converge, a more flexible means of representing default information and of computing the ME-ranking has been developed. To the extent that these two approaches overlap, that is, for minimal core sets of defaults of equal strength, the ME-rankings found by both methods coincide. However, while Goldszmidt's version defines a single solution for any set of defaults and is restricted to minimal core sets, this refinement makes the ME-approach both more flexible and more widely applicable. It is now possible to obtain different ME-rankings corresponding to different strength assignments over a given set of defaults. In fact, some defaults are ME-entailed regardless of a strength assignment, whereas others depend on the strengths assigned to the extent that both a default and its converse may be ME-entailed by the same set under different assignments. But is this useful?

There are two reasons that suggest that this more general ME-approach gives a very realistic account of what is meant by default reasoning. Firstly, it enables conflict among defaults to be resolved both definitively and flexibly. That is, although one has the freedom to alter the priorities between defaults, the effect this has is determined by the structure of the problem. This means that some default conclusions are susceptible to different strengths while others are not. The fact that this new approach can model both "intuitively" correct default conclusions (those which are uncontentious) and ambiguous conclusions (those which depend on different strengths), makes it a strong candidate for being recognised as the definitive theory of default reasoning. As such the ME-approach can be used to analyse the structure of default reasoning itself and hence enable a better understanding of what underlies it.

Secondly, the fact that a given set may be represented by many different ME-rankings suggests that some of these may have already been proposed as default consequence relations. From the point of view that the ME-approach represents the least biased estimate of what should be entailed by a set of defaults, the underlying meaning and biases of other default systems can be examined through comparison with it (see [2] for a comparison with lexicographic entailment). Thus the revised ME-approach can be used as a benchmark system in its own right from which to assess other formalisms.

In theory, using the ME-approach offers a well-motivated account of default reasoning that satisfies all the default intuitions incorporated in the illustrative examples. However, from the practical perspective, it is less than ideal. The main problem is, of course, one of complexity; although recent research has identified at least one type of default set for which the ME-algorithm is tractable [4], the general case has a lower bound complexity in $\mathrm{FP}^{\mathrm{NP}}$. But the other seemingly practical problem of the ME-approach, that of obtaining multiple solutions, can be seen as offering some insight into the nature of default reasoning itself. The redundancy that causes this problem is extremely rare, which means that in most cases, by merely specifying the relative strengths of defaults, a unique solution can be found. If the user is unwilling to commit himself to specifying their relative strengths, uniform strengths can be assigned to defaults. As was seen in the final example, genuine redundancy can lead to clashes of intuitions, but the use of this semantics to model default knowledge can help to identify and resolve these ambiguities.

Acknowledgements

This work was partly funded by the EPSRC under grant GR/L84117 and was carried out while the first author was funded by an EPSRC Studentship. We are very grateful to the anonymous reviewers for carefully reading through the manuscript and for many useful comments and suggestions.

References

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

[2] R. A. Bourne and S. Parsons. Connecting lexicographic with maximum entropy entailment. In A. Hunter and S. Parsons, editors, *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Lecture Notes in Artificial Intelligence 1638)*, pages 80–91. Springer, 1999.

[3] R. A. Bourne and S. Parsons. Maximum entropy and variable strength defaults. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 50–55, 1999.

[4] T. Eiter and T. Lukasiewicz. Default reasoning from conditional knowledge bases: Complexity and tractable cases. *Artificial Intelligence*, 124:169–241, 2000.

[5] D. Flanagan. *Java in a Nutshell.* O'Reilly, Sebastopol, CA, 1997.

[6] H. Geffner. *Default reasoning: causal and conditional theories.* MIT Press, Cambridge, MA, 1992.

[7] M. Goldszmidt. *Qualitative Probabilities: A Normative Framework for Commonsense Reasoning.* PhD thesis: Technical report R-190, Cognitive Systems Laboratory, UCLA, Los Angeles, 1992.

[8] 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.

[9] M. Goldszmidt and J. Pearl. On the consistency of defeasible databases. *Artificial Intelligence*, 52:121–149, 1991.

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

[11] D. Lehmann. Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence*, 15:61–82, 1995.

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

[13] 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.

[14] E. Neufeld. Notes on "A clash of intuitions". *Artificial Intelligence*, 48:225–240, 1991.

[15] J. Paris and A. Vencovská. A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 4:183–224, 1990.

[16] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: a survey. In *Knowledge Representation*, pages 505–515, 1989.

[17] J. Pearl. 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*, pages 121–135, 1990.

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

[19] W. Spohn. A general non-probabilistic theory of inductive reasoning. In *Uncertainty in Artificial Intelligence 4*, pages 149–159, 1990.

[20] 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.