Explaining Default Intuitions Using Maximum Entropy

Rachel A. Bourne
Electronic Engineering Department
Queen Mary, University of London
London E1 4NS, UK
r.a.bourne@elec.qmul.ac.uk

March 13, 2003

Abstract

While research into default reasoning is extensive and many default intuitions are commonly held, no one system has yet captured all these intuitions nor given a formal account to motivate them. This paper argues that the extended maximum entropy approach which incorporates variable strength defaults provides a benchmark for default reasoning that is not only objectively motivated but also satisfies all the accepted default intuitions. It is shown that the behaviour of the approach coincides with a wide range of default intuitions taken from examples in the literature, and can be used to explain why some examples have led to confusion. Moreover, analysing the solutions produced by the maximum entropy approach enables clearer differentiation between the default knowledge they contain and the default inferences required of the reasoning system. This suggests that the maximum entropy approach can be used as a benchmark both for eliciting default knowledge when building a knowledge base and, by comparison, for clarifying the underlying biases of other default reasoning systems.

1 Introduction

A common intuition is that a default represents a general rule that may admit exceptions, e.g., “women are bad drivers”. Rather than having a truth value, such rules can be thought of as existing as part of some reasoner's background knowledge. In conjunction with known factual information about the world, defaults are used to derive new, defeasible beliefs about the most plausible state of the world, subject to retraction if conflicting factual information comes to light. The question arises: how does the reasoner use its default knowledge? That is, how can a set of defaults be systematically processed to ensure that the output indeed represents the most plausible state of the world? Perhaps more importantly, what does "most plausible" actually mean?

This paper attempts to answers these questions by arguing that an existing system—the extended maximum entropy approach or ME^{+}approach [5, 11]—provides a formal theory of default reasoning that can be used as a benchmark. This is not to imply that other default systems are "wrong", but when used as a tool for comparison, the ME^{+}approach can help to clarify the assumptions on which other systems are based, and hence assist in their evaluation. The paper is intended to give a high level argument, rather than a very technical one, demonstrating both that the ME^{+}approach indeed corresponds to default intuitions, which have been so hard to capture precisely, and that it provides a model from which to explain what these intuitions may involve.

The paper is arranged as follows: in the next section, an overview of the default reasoning process is followed by justification of the use of maximum entropy to establish a benchmark default reasoning mechanism; in section 3, an overview of the extended maximum entropy approach is given; in section 4, the common intuitions of default reasoning are analysed using the ME^{+}approach; in section 5, the ME^{+}approach is applied in the context of building a default knowledge base to demonstrate its use as a means of eliciting default knowledge; section 6 concludes.

2 Defaults, default reasoning and maximum entropy

Default reasoning is broadly a process by which a set of default rules are manipulated to arrive at defeasible beliefs. There are various frameworks which accomplish this including default logic [32], circumscription [24], preferential reasoning [17], and reasoning using rational consequence relations [19]; other closely related systems include inheritance hierarchies [13, 38] and conditional logics [26, 16]. The meaning of defaults in these frameworks differs, and they may or may not involve the reasoner incorporating facts about the world in order to come to conclusions. However, one way or another, all these systems require that the user input some set of defaults representing his background knowledge, and result in further default conclusions being produced. A distinction therefore needs to be drawn between the inputs—the default knowledge—and the outputs—the default inferences: the former represent the user's firm beliefs about how the world usually behaves, while the latter represent other, defeasible beliefs about how one may suppose the world to behave based on the original defaults. It is not always easy to distinguish these different types of information since they often have an identical structure (i.e., they are both defaults), but by doing so, it is possible to come to a clearer understanding of the process of default reasoning and hence have realistic expectations of what it should be capable of achieving.

The framework for default reasoning which paints a clear picture of this process, is that described by Kraus, Lehmann and Magidor in their influential paper on cumulative and preferential reasoning [17]; indeed, preferential logics were originally proposed as a framework for capturing the behaviour of any system of nonmonotonic reasoning [34]. By considering a system of default inference as a means of *extending* a set of defaults, it becomes obvious that the original set represents the background knowledge and the extra defaults in the extension are the default inferences. Thus, any method for extending a set of defaults defines a system of default reasoning. The question becomes: which system best represents the default intuitions so widely held, and why do some examples confound these intuitions and lead to disagreements?

Certainly, many systems, including preferential reasoning itself, fail to exhibit some intuitions which are widely considered necessary to fully account for default reasoning. While preferential conclusions are rarely disputed, more complex behaviours appear to be required; where these come from, and exactly what they involve, has never been defined precisely. These extra behaviours have mainly been illustrated through examples with intuitively obvious default conclusions; while such examples do illustrate common patterns of default reasoning, they do not help to explain how such patterns arise, nor how more complex interactions among defaults should be resolved.

Rather than designing a default reasoning system in order to satisfy the examples, what is needed is a means of capturing the underlying concepts which cause default intuitions to arise. It has been shown, separately by Shore and Johnson [36, 35] and by Paris and Vencovská [28, 29], that maximum entropy (ME) inference is the only sound and coherent model of inductive inference. From simple assumptions of consistency and independence, and using different methods of analysis, two distinct derivations of the uniqueness of the maximum entropy method are possible. Kern-Isberner derives a third, more specific, characterisation of inference using ME, in the context of quantified probabilistic conditionals [16]. Again it is shown that ME inference is the only solution to the update problem. Kern-Isberner also uses ME updating to demonstrate the validity of some deduction rules of conditional logic [26]—for example, transitivity, specificity and reasoning by cases [15]. However, since ME updating is a global inference strategy, and the deduction rules apply to subsets of probabilistic conditionals, the results are only valid in isolation as other conditionals may affect the updating process. The sanctioning of common patterns of plausible inference under ME, albeit invalid if applied only locally, offers some evidence that commonsense reasoning follows the underlying principle of indifference which ME incorporates, as has been argued elsewhere [27].

Given these findings that maximum entropy provides a logical and consistent means of inference from uncertainty [36, 14]—indeed the only one—and given that there is no general agreement on a benchmark system of default reasoning, it is useful to consider the validity of default intuitions in the context of maximum entropy. The extended maximum entropy approach [5, 6] is derived by applying ME to the ε-semantics [1, 30], a sound and clear semantics for preferential reasoning based on probability theory. This paper shows that the resultant system, motivated soley by probability theory and ME, allows the whole gamut of default intuitions to be recovered and, hence, explained.

3 The extended maximum entropy approach

First some preliminary definitions and notation. A finite propositional language $\mathcal{L}$ is made up of propositions a, b, c, $\ldots$ and the usual connectives $\neg$,

$\wedge$, $\vee$, $\rightarrow$. A *default rule*, e.g., $a \Rightarrow b$, is a pair of propositions or propositional formulas joined by a 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* or *violate* it if $m \models a \wedge \neg b$. The default $a \Rightarrow \neg b$ is called the *converse* of $a \Rightarrow b$.

The ε-semantics [1, 30] for a default is that it represents a constraint on a probability distribution such that the conditional probability associated with a default, $a \Rightarrow b$, is constrained to be greater than $1 - \varepsilon$ for some infinitesimal parameter $\varepsilon > 0$.

$$a \Rightarrow b \qquad \equiv \qquad P(b|a) \geq 1 - \varepsilon \tag{1}$$

The exact value of ε is not relevant since it is merely a parameter used to link together the constraints associated with a set of defaults. Given that default information is intended to represent general rules of the form *"if a then, normally, b"*, the associated conditional probability is assumed to be relatively high and so the parameter ε is taken to be a real number close to zero.

In original ME approach, developed by Goldszmidt [11], the ε-semantics is extending using maximum entropy to select one probability distribution from all those that satisfy the default constraints. The unique maximum entropy distribution so defined is then abstracted to a ranking function that determines the rational consequences of a set of defaults. The ε-semantics sanctions default conclusions whose order of magnitude constraints exist in the limit as the uncertainties are made arbitrarily small; more colloquially, this can be thought of as taking one's assumptions to the extreme. Using maximum entropy this process is further refined: the entropy of a probability distribution is a measure of the uncertainty inherent in it [14]; by choosing the ME distribution, the uncertainty is maximised, leading to the most indifferent distribution satisfying the defaults; any distribution with a lower entropy can be thought of as being biased towards one default or another, hence the ME ranking is the least biased [36].

The extended ME approach [5, 6] adapts the original ME approach in two ways. The first adaptation is that, using the abstraction to ranking functions, it extends the semantics of individual defaults to incorporate variable priorities or relative strengths. Under the ε-semantics, when a user specifies a default, he believes that the probability that the default will be violated is extremely small, but not zero—he allows a small amount of uncertainty that the default may not be verified. Using defaults with relative strengths means that the uncertainties associated with some defaults are *orders of magnitude* smaller than those of other defaults; with no intuitions about defaults' relative strengths, all are simply assigned the same strength. It should be noted that when all defaults have equal strength, the ME$^+$approach subsumes the original ME approach and both systems sanction the same conclusions.

The second adaptation is to the constraints used to represent defaults, which are, in a sense, loosened to better reflect the fact that they will be abstracted. Thus for a default $a_i \Rightarrow b_i$, Goldszmidt uses the constraint:

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

to represent the fact that as $\varepsilon \rightarrow 0$, the probability of finding b_i to be true when a_i is true tends to one. In the ME$^+$approach, each default is assigned a strength, s_i, written $a_i \overset{s_i}{\Rightarrow} b_i$, and the constraint becomes:

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

where the function $O_i(\varepsilon^{s_i})$, termed a *convergence function*, has a specific order of magnitude corresponding to the strength of the default, but an unspecified convergence coefficient. The motivation for this is that since the coefficients will be lost when the probability distribution is abstracted into a ranking function, it is inappropriate that they be specified in the defaults. This different interpretation of defaults has an impact on the result of applying the maximum entropy technique. Whereas, the original ME approach determines a unique solution, in the ME$^+$approach there are different solutions corresponding to the different strengths assigned to the defaults. However, this comes at a price; because the constraints associated with defaults are only determined up to their order of magnitude, there may also be multiple solutions *even for the same default strength assignment*. Such cases occur when the set of defaults contains redundancy which can potentially be ambiguous. In practice, such pathological cases turn out to be unusual—where they do occur it can indicate that the default set itself needs to be more clearly specified. An example of this occurs in section 4.4.

After the principle of maximum entropy has been applied, by allowing $\varepsilon \rightarrow 0$, a ranking function abstraction to the ME$^+$solution is obtained. The ME$^+$ranking for a set of variable strength defaults, $\Delta = \{r_i : a_i \overset{s_i}{\Rightarrow} b_i\}$ is given by solving the coupled equations:

$$\textsc{me}^+(m) = \sum_{\substack{r_i \\ m \models a_i \wedge \neg b_i}} \textsc{me}^+(r_i) \tag{4}$$

$$\textsc{me}^+(a_i \wedge b_i) + s_i \leq \textsc{me}^+(a_i \wedge \neg b_i) \tag{5}$$

in which each default, r_i, influences the ranking according to its integer weight, or ME$^+$rank, $\textsc{me}^+(r_i)$. A mathematical derivation of these equations and an algorithm for solving them is given in [6].

From (4), it can be seen that the ME$^+$rank of any model (or state) is determined by summing the weights of those defaults it falsifies. If a model falsifies no defaults, it will have rank 0, and, in general, the more defaults a model falsifies, the more abnormal (higher ranked) it becomes. The weights associated with the ME$^+$solution reflect both a default's structural relevance to the other defaults in Δ and its strength; because of the indifference achieved by using ME, in some sense the weights reflect the least biased estimate of a default's relative importance. A default is ME$^+$entailed if its lowest ranked verifying model is strictly lower than its lowest ranked falsifying model.

Interestingly, many other default systems have used similar methods which assign some sort of weight to a default, and which use the default violations to assess each model's worth; examples include systems Z and Z+ [12], lexicographic entailment [18], LCD belief functions [3] and penalty logics [8]. The difference between these systems and the ME$^+$approach is that the latter is derived in a mathematically sound way directly from the ε-semantics rather than being an intuitively motivated, but ultimately ad hoc, invention.

The application of the ME$^+$approach can be illustrated through a simple example. Consider the following default knowledge base which represent the two defaults that birds usually fly and that parrots are usually birds.

$$\Delta_1 = \{r_1 : b \overset{s_1}{\Rightarrow} f, r_2 : p \overset{s_2}{\Rightarrow} b\}$$

This results in two constraint equations:

$$\textsc{me}^+(b \wedge f) + s_1 \leq \textsc{me}^+(b \wedge \neg f) \tag{6}$$

m	b	f	p	r_1	r_2	$\text{ME}^+(m)$
m_1	0	0	0	-	-	0
m_2	0	0	1	-	f	s_2
m_3	0	1	0	-	-	0
m_4	0	1	1	-	f	s_2
m_5	1	0	0	f	-	s_1
m_6	1	0	1	f	v	s_1
m_7	1	1	0	v	-	0
m_8	1	1	1	v	v	0

Table 1: The ME$^+$ranks for models using Δ_1.

$$\text{ME}^+(p \wedge b) + s_2 \leq \text{ME}^+(p \wedge \neg b) \qquad (7)$$

along with 8 equations for the ranks of the models. Table 1 shows each model and whether or not it falsifies each of the defaults. The final column gives the ME$^+$rank of the model.

For this example, the ME$^+$solution is can easily be calculated to find that the ME$^+$rank of each default is simply its strength (this is because the defaults do not interfere in any way, i.e., they are independent).

$$\text{ME}^+(r_1) = s_1 \qquad \text{ME}^+(r_2) = s_2$$

To determine whether or not any particular default is ME$^+$entailed, that is, belongs to the ME$^+$extension of Δ_1, one examines the target default's minimal verifying and falsifying models. Take the default $p \Rightarrow f$, that is, "do parrots usually fly?".

$$\text{ME}^+(p \wedge f) = 0 < \text{ME}^+(p \wedge \neg f) = \min(s_1, s_2)$$

Because the verifying model is strictly lower, the default is indeed ME$^+$entailed. It is possible to determine the *strength* of the entailment which is just the difference between these two ranks; in this case the default $p \Rightarrow f$ is ME$^+$entailed with strength $\min(s_1, s_2)$.

4 Analysis of default intuitions

In this section, the ME$^+$approach is applied to examples that illustrate the intuitive patterns of default behaviour which are considered to be general requirements of default reasoning systems, though no system has yet captured all such behaviours. Unfortunately, one of the problems with designing default reasoning systems is that attempting to define precisely what these so-called benchmark behaviours should be is rather a circular argument—if the behaviours could be specified precisely, then that specification would represent a blueprint from which to build a system; on the other hand, without a precise specification, there is plenty of room for differences of opinion since intuitions can be highly subjective.

Despite these rather negative remarks, several authors have attempted to draw up a list of requirements (see, for example, [21, 3]). While those presented here roughly follow the "desirable properties" listed in [3], some license has been taken with the description of what those properties represent. Thus each behaviour is itself subjected to a short analysis before the behaviour of the ME$^+$solutions are discussed. In the main, these behaviours relate to some form of inheritance, e.g., when should a property be inherited by a subclass, when should it be blocked, on what does this depend, etc. By the end of this section, it is hoped that the reader will at least agree that the ME$^+$approach handles all default intuitions satisfactorily; further, the reader may accept that using the ME$^+$approach is a candidate for a normative model of default reasoning.

4.1 Property inheritance, transitivity and irrelevance

The logical property of transitivity, that is, from $a \to b$ and $b \to c$ deduce $a \to c$, has been a thorn in the side of nonmonotonic reasoning. One of the main uses of defaults is to encode generalised knowledge concisely in terms of rules (e.g., normally birds fly), but the fact that these are not logical and admit exceptions means that the transitive transfer of properties, or normal inheritance, must sometimes be blocked. Clearly for nonmonotonic reasoning, transitivity does not hold unilaterally.

Makinson pointed out that transitivity can be separated into two more basic inference rules: cumulative transitivity[1] and monotony [22]. He suggested that nonmonotonic reasoning processes need only satisfy the first of these conditions. Makinson's analysis, along with that of Gabbay [9], ultimately led to the formulation of the rule system P, as core behaviour for nonmonotonic reasoning systems [17]. But it appears that this set of rules is the limit in terms of attempting to formalise nonmonotonic behaviour using rules of inference. As Lehmann and Magidor subsequently found, more sophisticated rules such as rational monotonicity lead to multiple solutions [19].

The real difficulty lies in attempting to impose a property such as transitivity as a rule of inference, or as a constraint. It is far better to view it as a property one would expect to find unless an exceptional circumstance exists, i.e., an observable phenomenon whose absence indicates an exception has occurred.

The role of transitivity has led to much confusion both in the design of default systems and in how to represent default knowledge. An example of this was the refusal to accept a transitive conclusion from default logic in the following case [33]. Reiter and Criscuolo argued that from the two defaults "typically high school dropouts are adults" and "typically adults are employed", it was undesirable to conclude that "typically high school dropouts are employed". They went on to say:

> Nor would we want to conclude that "Typically high school dropouts are not employed." Rather we would remain agnostic about the employment status of a typical high school dropout.

For whatever reasons, presumably preconceived ideas about high school dropouts, the results of the reasoning process were prejudged. By requiring that it remain agnostic about a particular default conclusion, an additional constraint was unwittingly imposed. This neatly illustrates the importance of distinguishing between default knowledge and default inference: wishing to remain agnostic was problem specific information that was not supplied to the reasoning mechanism.

To resolve their dilemma, Reiter and Criscuolo resorted to using "semi-normal" defaults so that the desired conclusions could be obtained. In effect,

[1] Equivalent to cautious monotonicity.

they were having to block the transitive conclusions which default logic naturally produced. This led to unwanted side effects both in terms of extra, counterintuitive conclusions being sanctioned, and of multiple or non-existent extensions to default theories. A similar situation occurs with circumscription, where it becomes necessary to introduce abnormality predicates so that the "correct" results can be obtained [24, 20]. But this presupposes the default inferences required, ultimately invalidating the use of a system of default reasoning! After all, if one knew all the conclusions in advance, in theory one could simply program a system to reproduce them. By using defaults, the hope is to provide a concise representation of some domain and a mechanism for extracting a plausible view of the whole picture. Although this often involves what looks like transitive inference, it is a mistake to force a default system to behave in this way, or to prevent it, since this will undoubtedly lead to incorrect conclusions in some cases.

In the previous section, the example used to illustrate the ME$^+$approach in practice was a simple case of transitivity, that is, from the database:

$$\Delta_1 = \{r_1 : b \overset{s_1}{\Rightarrow} f, r_2 : p \overset{s_2}{\Rightarrow} b\}$$

it was shown that $p \Rightarrow f$ is an ME$^+$consequence regardless of the strengths assigned to the defaults (though s_1 and s_2 do affect the degree of the inference). Another way to view this is to notice that both falsifying models of $p \Rightarrow f$, must either falsify $p \Rightarrow b$ or $b \Rightarrow f$, whereas $p \Rightarrow f$ has a verifying model, $p \wedge b \wedge f$, that does not falsify either; so $p \Rightarrow f$ is unconditionally ME$^+$entailed by Δ. The uncontroversial conclusion given by EME is that the ability to fly is normally inherited in a transitive way. Whether defaults are viewed as some kind of inference rule, or as constraints, it seems hard to argue that $p \Rightarrow f$ should *not* be a default conclusion of this set, at least when the abstract symbols are not loaded with intuitive interpretations.

What this shows is that, in isolation, defaults *do* chain transitively under the ME$^+$approach. A similar observation is made by Kern-Isberner in her paper on probabilistic conditionals [15]. Obviously, other defaults which deal with the same propositions or formulas, may cause interference which prevents this from happening, but, other things being equal, transitivity is to be expected.

Since transitive behaviour is normal in unexceptional circumstances, it is not surprising that it is a property which has been isolated and considered important. This simple example also demonstrates that, if it is accepted that ME$^+$entailment represents the least biased consequence relation, it can be used to *discover* the hidden biases which exist in one's knowledge. Any unusual conclusions or side effects reflect differences in the problem as it has been encoded and the implicit constraints which the user has failed to encode. An example of this process is given in section 5.

4.1.1 Irrelevance

Related to the problem of being able to correctly perform property inheritance is the ability to do so in the presence of irrelevant information. Some default systems, especially those based on the ε-semantics, have suffered from the inability to ignore extraneous information. For example, the default $a \wedge c \Rightarrow b$ is not ε-entailed by the singleton set $\{a \Rightarrow b\}$. This problem arises because the ε-semantics sanctions only defaults whose constraints are satisfied by all distributions compatible with the initial set. Some of these distributions may also satisfy the set $\{a \Rightarrow b, a \wedge c \Rightarrow \neg b\}$, so $a \wedge c \Rightarrow b$ cannot be ε-entailed, even when c is a completely irrelevant proposition. In contrast, nonmonotonic reasoning systems which allow a form of transitivity, such as default logic and circumscription, do not suffer from this problem exactly because inheritance is blocked explicitly by the relevant formulas.

Under the ME$^+$approach, models in which some irrelevant formula is true and those in which it is false are treated equally. Returning to the example above, Δ_1 ME$^+$entails not only $p \Rightarrow f$ but also $p \wedge x \Rightarrow f$ and $p \wedge \neg x \Rightarrow f$, for any x not previously mentioned, under any strength assignment. Thus the requirement of ignoring irrelevant information is satisfied by the ME$^+$approach.

4.2 Conflicting inheritance and specificity

A primary requirement of default reasoning is surely that it only sanction default conclusions which are, in some way, justifiable. The complement of this is that, when there is no reasonable conclusion to come to, it should remain ambivalent. The classic example of this behaviour is given by the so-called Nixon diamond:

$$\Delta_2 = \{r_1 : a \overset{s_1}{\Rightarrow} c, r_2 : b \overset{s_2}{\Rightarrow} \neg c\}$$

In this example, two defaults with differing antecedents point to opposite conclusions. While the defaults do not conflict directly, if both their antecedents are satisfied simultaneously, it becomes unclear whether or not any default conclusion can be reached. The question is, if an object exhibits both properties a and b, should it inherit property c, $\neg c$, or neither?

The ME$^+$solution to this example is straightforward; in fact, the ME$^+$rank of each default is, again, simply its strength: $\text{ME}^+(r_1) = s_1$ and $\text{ME}^+(r_2) = s_2$. This leads to behaviour which captures the intuition that, when the problem is symmetrical (when $s_1 = s_2$), ambiguity is preserved and neither $a \wedge b \Rightarrow c$ nor $a \wedge b \Rightarrow \neg c$ is ME$^+$entailed; however, if either strength dominates the other, then that default takes priority and the corresponding conclusion is sanctioned; for example, if $s_2 > s_1$, then $a \wedge b \Rightarrow \neg c$ is ME$^+$entailed. Note that in such a case the strength of the inference is weakened by the presence of conflict so that the conclusion is only ME$^+$entailed to degree $s_2 - s_1$.

This result seems appropriate: when there is no reason to prefer one conclusion over another, no firm conclusion can be reached, but if one default is stronger its conclusion will prevail. In this case, the ME$^+$solution resolves the conflict because it treats each default in the same way while taking account of their assigned strengths. This allows the system to come to completely different default conclusions when the strengths are altered. Of course, this does not mean that *any* default conclusion can be obtained by altering the strengths, just that by tipping the balance in favour of one default or the other, this particularly simple type of conflict can be resolved in an even-handed manner.

Whereas in the above example, the strengths were the deciding factor, in other circumstances, it may be the structure of the interaction between defaults which forces the conclusion one way or another, regardless of relative strengths. This structural prioritisation of defaults has been termed specificity [31] since one default may relate to a more specific situation than another, and therefore override the application of less specific, conflicting defaults.

Consider the addition of an extra default, $a \overset{s_3}{\Rightarrow} b$, to Δ_2 above:

$$\Delta_3 = \{r_1 : a \overset{s_1}{\Rightarrow} c, r_2 : b \overset{s_2}{\Rightarrow} \neg c, r_3 : a \overset{s_3}{\Rightarrow} b\}$$

Again, in this example, the ME$^+$solution is particularly simple; the ME$^+$ranks of defaults are given by $\text{ME}^+(r_1) = s_1 + s_2$, $\text{ME}^+(r_2) = s_2$, $\text{ME}^+(r_3) = s_2 + s_3$.

Now consider again whether an object which exhibits both properties a and b, inherits property c, $\neg c$, or neither. The ME$^+$ranks of the relevant models are:

$$\text{ME}^+(a \wedge b \wedge c) = s_2 \quad \text{and} \quad \text{ME}^+(a \wedge b \wedge \neg c) = s_1 + s_2$$

The default conclusion $a \wedge b \Rightarrow c$ is ME$^+$entailed to degree s_1, regardless of the values for the strengths s_1 and s_2; in particular, it makes no difference if $s_2 > s_1$; such a conclusion might be said to be *unconditionally* ME$^+$entailed[2]. This is a good example of how the ME$^+$approach handles specificity: because a is effectively a subclass of b (as witnessed by the default $a \Rightarrow b$), the default which refers to a specifically in its antecedent takes priority over that which refers only to b, the superclass. An interesting observation for this particular example is that, because it is the more specific default, $a \Rightarrow c$, which is active, the derived default conclusion is ME$^+$entailed to the same degree as that which caused it. The way the conflict is handled under the ME$^+$approach illustrates neatly exactly which defaults are involved in the inheritance and which are blocked.

Together, the Nixon diamond and its extension[3], illustrate two different ways in which the ME$^+$approach handles conflict resolution among default interactions. When defaults are of equal status, conflicts can be resolved by examining the relative strengths of the defaults involved, with the possibility for preserving ambiguity when there is no bias one way or the other. When there is an implicit structural priority over defaults, with one being applicable in a more specific circumstance than the other, the relative strengths are immaterial and the conflict is resolved in favour of the more specific default. The fact that both types of conflict resolution are handled naturally by the ME$^+$approach—that is, they were not design specifications but are purely a result of the chosen semantics—leads one to expect that these conflicts will be resolved in a similarly reasonable manner for larger and more complicated default interactions.

[2] In fact, $a \wedge b \Rightarrow c$ is an ε-consequence of Δ_2.

[3] Isomorphic to the classic "penguins do not fly" example (substitute a for "penguin", b for "bird", c for "not fly").

4.3 Exceptional inheritance

Inheritance to exceptional subclasses has been one of the most difficult behaviours to obtain from default systems. The intuition is that property inheritance should not be blocked for exceptional subclasses except for those properties which make them exceptional. This can be seen either as just a special case of transitivity which occurs for the unexceptional cases but not the exceptional ones, or as a presumption that defaults hold *unless* there is information to the contrary. The assumption is that the information about an exceptional feature should not affect other features of the same status or at the same level.

Another argument in favour of exceptional inheritance is that, although exceptions to defaults are known to be a possibility, the reason that a superclass exists at all is because objects have been classified according to their common features or because they exhibit similar properties. Thus objects which belong to the same class should be similar in all features unless they are known explicitly not to be, meaning that as many typical features as possible should be inherited.

This is another case where, although illustrative examples may have intuitively appealing conclusions, in more complex cases, it is not clear whether or not exceptional properties should be inherited. While the ME$^+$approach handles this "correctly" for the following example, it is also flexible in the way it achieves this, leaving room for nonmonotonic changes in inference if conflicting defaults are subsequently added.

$$\Delta_4 = \{r_1 : b \overset{s_1}{\Rightarrow} f, r_2 : p \overset{s_2}{\Rightarrow} b, r_3 : p \overset{s_3}{\Rightarrow} \neg f, r_4 : b \overset{s_4}{\Rightarrow} w\}$$

The intended interpretation of these defaults is that birds normally fly, penguins are normally birds, penguins normally do not fly and birds normally have wings. The question is whether penguins can inherit the wing attribute of birds despite being exceptional.

This problem again has a simple ME$^+$solution for any strength assignment given by $\textsc{me}^+(r_1) = s_1$, $\textsc{me}^+(r_2) = s_1 + s_2$, $\textsc{me}^+(r_3) = s_1 + s_3$, and $\textsc{me}^+(r_4) = s_4$. It follows that some default conclusions hold unconditionally, in particular, the default $p \Rightarrow w$, is unconditionally ME$^+$entailed (there are always minimal verifying models of it which are more normal than its falsifying models), although the strength assignment may dictate which defaults are actually responsible for the inheritance. If, on the one hand, $s_2 < s_4$ then the fact that falsifying $p \Rightarrow b$ is more serious than falsifying $b \Rightarrow f$ leads to $p \Rightarrow w$ being ME$^+$entailed; if, on the other hand, $s_4 < s_2$ then the reason for inheritance is simply because fewer defaults are violated. Interestingly, the degree to which $p \Rightarrow w$ is ME$^+$entailed depends on the strength of either $p \Rightarrow b$ or $b \Rightarrow w$, whichever is weaker, giving support to the view that an argument is only as strong as its weakest link. Thus the ME$^+$approach sanctions the transitive conclusion $p \Rightarrow w$ but prohibits $p \Rightarrow f$.

Further examination of this default database indicates that there comes a point when, rather than exceptional inheritance occurring, it is possible to doubt that an unusual object is even a member of the class to which it is purported to belong. For example, consider whether a wingless penguin should still be considered a bird despite being unusual in a number of ways. That is consider whether the default $p \wedge \neg w \Rightarrow b$ is ME$^+$entailed. The ME$^+$ranks of the relevant models are:

$$\text{ME}^+(p \wedge \neg w \wedge b) = s_1 + s_4 \quad \text{and} \quad \text{ME}^+(p \wedge \neg w \wedge \neg b) = s_1 + s_2$$

Thus the default $p \wedge \neg w \Rightarrow b$ is a controversial ME$^+$conclusion—it may be ME$^+$entailed but so may its converse. The resolution of this controversy depends solely on the strengths assigned to the defaults r_2 and r_4, i.e., how strongly are penguins birds relative to birds having wings.

The reason the ME$^+$approach handles inheritance to exceptional subclasses so elegantly, is because it assesses the models on the basis of a weighted sum of all defaults falsified. But, as the ME$^+$solution to this example indicates, there may be different reasons for the inheritance, dependent on the strengths used, implying that the causes of exceptional inheritance are not clearcut. Perhaps this goes some way to explaining why exceptional inheritance has been so difficult to formalise, and why Weydert has derived an "impossibility" result for systems which are forced to satisfy it [39].

4.4 Multiple inheritance

In some sense, multiple inheritance is an extension of the conflict resolution already examined in section 4.2. This section demonstrated that the

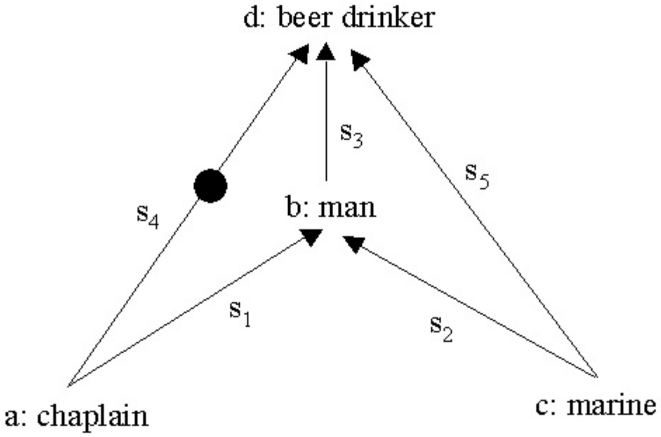

Figure 1: Illustration of marine-chaplain example.

ME$^+$approach “weighs” up these conflicts, balancing both the structure of the problem and the relative strengths of defaults. This leads to some interesting insights into this complex issue, much debated in the field of inheritance hierarchies [37]. The main point is that while a default set may lead to unconditional default conclusions, these may still be nonmonotonically retracted if the addition of new defaults changes the structure of the problem.

To illustrate this phenomenon, an extension of a well-known controversial example is analysed. The original version is discussed at length in several papers [23, 25, 38]. The default set is given by:

$$\begin{aligned}\Delta_5 = \quad \{ \quad & r_1 : a \overset{s_1}{\Rightarrow} b, r_2 : c \overset{s_2}{\Rightarrow} b, \\ & r_3 : b \overset{s_3}{\Rightarrow} d, r_4 : a \overset{s_4}{\Rightarrow} \neg d, r_5 : c \overset{s_5}{\Rightarrow} d \quad \}\end{aligned}$$

In view of its history, an illustration appears in figure 1.

The controversy surrounding this example involves whether or not the default “marine-chaplains are beer drinkers” ($a \wedge c \Rightarrow d$), or its converse, is a default conclusion. In the original example, the direct link from marine to beer drinker, r_5, is omitted and the ME$^+$solution is unique giving an unconditional ME$^+$consequence of $a \wedge c \Rightarrow \neg d$ (i.e., “marine-chaplains are *not* beer drinkers”). This result is unsurprising since the link between chaplains and beer drinker is clearly more specific than that between beer drinker and marine which is via man. In other words, chaplains are known to be men who are known to be beer drinkers, so the fact that a chaplain is also a marine should not affect the conclusion that he doesn't drink beer. After all, marines are only known to be beer drinkers by virtue of being men, at least as represented in the original example.

However, Touretzky *et al.* [38] 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, the additional extra default $r_5 : c \overset{s_5}{\Rightarrow} d$ is added, creating a direct link between marine and beer drinker. Now this default is already an ME$^+$consequence of the original set and is ME$^+$entailed to degree $\min(s_2, s_3)$. This means that the ME$^+$solution to the extended example is a complex one, reflecting the fact that some rules become redundant under some strength assignments; in this case whenever $s_5 le \min(s_2, s_3)$. However, if $s_5 > \min(s_2, s_3)$, all defaults are active giving an ME$^+$solution of $\text{ME}^+(r_1) = s_1 + s_3$, $\text{ME}^+(r_2) = s_2$, $\text{ME}^+(r_3) = s_3$, $\text{ME}^+(r_4) = s_3 + s_4$ and $\text{ME}^+(r_5) = s_5 - \min(s_2, s_3)$.

In this case, the ME$^+$conclusion is still a controversial one. The minimal verifying and falsifying models of $a \wedge c \Rightarrow \neg d$ are:

$$\begin{array}{rcl} \text{ME}^+(a \wedge c \wedge \neg d) & : & \text{ME}^+(a \wedge c \wedge d) \\ s_3 + s_5 - \min(s_2, s_3) & : & s_4 \end{array}$$

and the result depends critically on the strengths.

It appears that multiple inheritance is bound to lead to ambiguous situations; however, this example demonstrates that the ME$^+$approach can be used to disentangle the ambiguities which arise, since it can help to identify both the causes of controversy and how to resolve them.

5 Eliciting default knowledge

Using the ME$^+$approach necessarily involves coming to some quite specific conclusions, mainly because the result is a ranking function that totally orders the set of possible world models. In fact, the use of rational consequence relations to represent default knowledge has been criticised as too committed to ranking worlds by several researchers [10, 2]. However, the conclusions which result from the ME$^+$ranking can be justified as those most likely to pertain if the defaults supplied are the *only* constraints which exist for the given domain. In reality, of course, this is a crude and simplistic model, but despite this, it can be used to elicit default knowledge from users since any significant deviations from the conclusions obtained using EME, imply that extra, or different, constraints exist. This use of ME was suggested by Jaynes for finding physical constraints [14], but is equally applicable to the more abstract problem of eliciting default knowledge.

There has already been an example of this process in action in section 4.1: Reiter and Criscuolo rejected the ME$^+$consequence that "typically high school dropouts are employed" but admitted that they wished to remain "agnostic" on this point. This agnosticism amounts to a further constraint on the problem that was not represented in the original default set. Rather than altering the reasoning mechanism to "correct" this problem, a more accurate model of the background knowledge is required.

By constructing a set of defaults and examining its ME$^+$consequences, usually by initially assigning all defaults equal strengths, it is often possible to obtain a better understanding of the intuitions of the user both in terms of how the defaults interact and whether any hold more strongly than others. This leads to a better translation of background knowledge into default rules. The following construction of a knowledge base from some background information illustrates how the ME$^+$approach can be put to work in practice. The example is taken from Brewka [7]:

> *Usually one has to go to a project meeting.*
> *This rule does not apply if somebody is sick, unless he only has a cold.*
> *The rule is also not applicable if somebody is on vacation.*

Firstly, it should be noted that there are several ways in which one might choose to encode this information. In particular, it is not obvious that the phrase *unless he only has a cold* implies that having a cold is a type of sickness, although common sense tells us that this is the case. There may be situations in which "unless" means only "if [something] happens to be the case as well". So the user must be aware of those of his intuitions which relate to the semantics of everyday language and therefore need to be represented explicitly. Taking this point on board, the following set is one way to represent the given information:

$$\Delta_6 = \{\texttt{True} \overset{1}{\Rightarrow} m, s \overset{1}{\Rightarrow} \neg m, c \overset{1}{\Rightarrow} s, c \overset{1}{\Rightarrow} m, v \overset{1}{\Rightarrow} \neg m\}$$

with the symbols standing for m meeting, s sick, c cold, and v vacation, and the strengths of all defaults being equal, initially. (So the default $\texttt{True} \Rightarrow m$ is a direct translation of "usually one has to go to a project meeting".)

Secondly, the information as it stands does not indicate whether or not one should attend the meeting if one has a cold but is on vacation. Although, intuitively, being on vacation overrides going to work, this is not made explicit in the information above. This leads to an interesting point. Is this conclusion a semantic intuition or a structural one? That is, should it be represented explicitly as an extra default, or should it be a derivable conclusion? This is the type of decision that the ME$^+$approach can help the user to make when building his knowledge base.

Given Δ_6 as it stands, the default $c \wedge v \Rightarrow \neg m$ is an ME$^+$consequence, but it is a controversial one. By increasing the strength of the default $c \Rightarrow m$ to 2, the default is no longer ME$^+$entailed, while increasing it further, to 3 or higher, means that the converse, i.e., $c \wedge v \Rightarrow m$, is ME$^+$entailed. To ensure that the "intuitive" conclusion holds, it is necessary for the strength of the default $v \Rightarrow \neg m$ to be greater than or equal to that of $c \Rightarrow m$. The user must therefore decide whether the default set is sufficient as it stands, so that altering the strengths, or adding extra defaults, might lead to a different conclusion; or whether, in fact, this intuition is a further constraint which needs to be made explicit and added to the default set.

What the example illustrates is that it is important, as a user, to be able to distinguish between different types of intuition: structural and semantic. While it is the responsibility of the default reasoning mechanism to handle the structural interactions of defaults, i.e., to satisfy the requirements of default reasoning as discussed in section 4, this will only produce the "correct" answers if the user has correctly encoded his semantic intuitions about the propositions. The ME$^+$approach can assist him in clarifying his intuitions because it treats all defaults equally, giving unbiased conclusions enabling the user to determine both the nature and extent of his own biases.

6 Conclusion

This paper has shown that the ME$^+$approach can represent many forms of default intuitions, including complex behaviours like multiple inheritance, despite the simplicity of its underlying motivation and semantics. Using the system as a benchmark, it is possible to differentiate between the default knowledge encoded in standard examples and the behaviour which is being asked of a default inference system. Furthermore, the system itself can be used in comparison with other default systems, clarifying the implicit semantics which they assign to defaults and default inference.

Because of this, it is argued that the default conclusions sanctioned under the ME$^+$approach represent the fundamental default inferences obtainable from a given set of defaults. A user who expects different default inferences must be using other constraints not encoded in his set of defaults; unbeknown to him, he has failed to encode implicit information he is using. This makes the extended ME$^+$approach an ideal tool for examining default intuitions themselves.

References

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

[2] 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.

[3] S. Benferhat, A. Saffioti, and P. Smets. Belief functions and default reasoning. In *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence*, pages 19–26, 1995.

[4] R. A. Bourne. *Default reasoning using maximum entropy and variable strength defaults.* PhD thesis, University of London, 1999.

[5] 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.

[6] R. A. Bourne and S. Parsons. Extending the maximum entropy approach to variable strength defaults. *Annals of Mathematics and Artificial Intelligence*, page To appear, 2003.

[7] G. Brewka. Preferred subtheories: an extended logical framework for default reasoning. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1043–1048, 1989.

[8] F. Dupin de Saint Cyr, J. Lang, and N. Schieux. Penalty logic and its link with Dempster-Shafer theory. In *Proceedings of the 10th Annual Conference on Uncertainty in Artificial Intelligence*, pages 204–211, 1994.

[9] D. M. Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In K. R. Apt, editor, *Logics and Models of Concurrent Systems*, pages 439–457, Berlin, 1985. NATO NSI Series, Springer.

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

[11] 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.

[12] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84:57–112, 1996.

[13] J. F. Horty, R. H. Thomason, and D. S. Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence*, 42:311–348, 1990.

[14] E. Jaynes. Where do we stand on maximum entropy? In R. Levine and M. Tribus, editors, *The Maximum Entropy Formalism*, pages 15–118, Cambridge, MA, 1979. MIT Press.

[15] G. Kern-Isberner. A logically sound method for uncertain reasoning with quantified conditionals. In D. M. Gabbay, R. Kruse, A. Nonnengart, and

H. J. Ohlbach, editors, *Qualitative and Quantitative Practical Reasoning (Lecture Notes in Artificial Intelligence 1244)*, pages 365–379, Berlin, 1997. Springer.

[16] G. Kern-Isberner. Characterizing the principle of minimum cross entropy in the conditional logic framework. *Artificial Intelligence*, 86:169–208, 1998.

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

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

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

[20] V. Lifschitz. Pointwise circumscription. In M. Ginsberg, editor, *Readings in nonmonotonic reasoning*, pages 179–193, San Mateo, 1987. Morgan Kaufmann.

[21] V. Lifschitz. Benchmark problems for formal non-monotonic reasoning, version 2.00. In M. Reinfrank, J. de Kleer, M. L. Ginsberg, and E. Sandewall, editors, *Non-monotonic reasoning (Lecture Notes in Artificial Intelligence 346)*, pages 202–219, Berlin, 1988. Springer.

[22] D. Makinson. General theory of cumulative inference. In M. Reinfrank, J. de Kleer, M. L. Ginsberg, and E. Sandewall, editors, *Nonmonotonic reasoning (Lecture Notes in Artificial Intelligence 346)*, pages 1–18, Berlin, 1988. Springer.

[23] 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.

[24] J. McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[25] E. Neufeld. Notes on “a clash of intuitions”. *Artificial Intelligence*, 48:225–240, 1991.

[26] D. Nute. *Topics in conditional logic.* Reidel, Dordrecht, Netherlands, 1980.

[27] J. Paris. Common sense and maximum entropy. *Synthese*, 117:75–93, 1998.

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

[29] J. Paris and A. Vencovská. In defense of the maximum entropy inference process. *International Journal of Approximate Reasoning*, 17:77–103, 1997.

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

[31] D. L. Poole. On the comparison of theories: preferring the most specific explanation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 144–147, Los Angeles, CA, 1985. Morgan Kaufmann.

[32] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[33] R. Reiter and G. Criscuolo. Some representational issues in default reasoning. *Computers & Mathematics with Applications*, 9:15–27, 1983.

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

[35] J. E. Shore. Relative entropy, probabilistic inference, and ai. In *Proceedings of the 2nd Annual Conference on Uncertainty in Artificial Intelligence*, pages 211–215, 1986.

[36] 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.

[37] D. S. Touretzky. *The mathematics of inheritance systems.* Morgan Kaufmann, San Mateo, 1986.

[38] 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.

[39] E. Weydert. SYSTEM JZ how to build a canonical ranking model of a default knowledge base. In *Proceedings of the Sixth International Conference on the Principles of Knowledge Representation and Reasoning*, pages 190–201, 1998.