View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Queen Mary Research Online

Queen Mary
University of London

Default reasoning using maximum entropy and variable strength defaults

Bourne, Rachel A

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/jspui/handle/123456789/3808

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

Default reasoning using maximum entropy and variable strength defaults

Rachel A Bourne

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of the
University of London.

Department of Electronic Engineering
Queen Mary & Westfield College

1999

Abstract

The thesis presents a computational model for reasoning with partial information which uses default rules or information about what normally happens. The idea is to provide a means of filling the gaps in an incomplete world view with the most plausible assumptions while allowing for the retraction of conclusions should they subsequently turn out to be incorrect. The model can be used both to reason from a given knowledge base of default rules, and to aid in the construction of such knowledge bases by allowing their designer to compare the consequences of his design with his own default assumptions. The conclusions supported by the proposed model are justified by the use of a probabilistic semantics for default rules in conjunction with the application of a rational means of inference from incomplete knowledge—the principle of maximum entropy (ME). The thesis develops both the theory and algorithms for the ME approach and argues that it should be considered as a general theory of default reasoning.

The argument supporting the thesis has two main threads. Firstly, the ME approach is tested on the benchmark examples required of nonmonotonic behaviour, and it is found to handle them appropriately. Moreover, these patterns of commonsense reasoning emerge as consequences of the chosen semantics rather than being design features. It is argued that this makes the ME approach more objective, and its conclusions more justifiable, than other default systems. Secondly, the ME approach is compared with two existing systems: the lexicographic approach (LEX) and system Z^+. It is shown that the former can be equated with ME under suitable conditions making it strictly less expressive, while the latter is too crude to perform the subtle resolution of default conflict which the ME approach allows. Finally, a program called DRS is described which implements all systems discussed in the thesis and provides a tool for testing their behaviours.

Acknowledgements

I would like to thank my supervisor, Simon Parsons, for all his help and encouragement over the past three years. Everyone in the Intelligent Systems Group and, indeed, in the Department of Electronic Engineering as a whole, has helped make my time here both enjoyable and productive. I am also grateful for the financial support which I received in the form of an EPSRC studentship.

Contents

1 Introduction **6**

1.1 Motivation 6

1.2 Requirements of default reasoning 8

1.3 Inference using maximum entropy 12

1.4 Overview of thesis and main contributions 13

2 Background **16**

2.1 Circumscription 16

2.2 Default logic 17

2.3 Inheritance hierarchies 19

2.4 Preferential reasoning 20

2.5 Qualitative probabilities and ranking functions 22

2.6 Infinitesimals 25

2.7 Maximum entropy inference in AI 29

3 Systems of default reasoning **33**

3.1 System P and the ε-semantics 33

3.2 Rational closure and system Z 38

3.3 Variable strength defaults 43

3.4 System Z^+ 46

3.5 Lexicographical closure 49

3.6 Goldszmidt's maximum entropy approach 53

3.7 Discussion 56

4 Maximum entropy and variable strength defaults **58**

4.1 Review of original ME assumptions 59

4.2 Deriving the maximum entropy ranking 61

4.3 Case study 65

4.4 The ME algorithm 68

4.4.1 Complexity 71
4.5 Uniqueness condition 72
4.6 Interpreting computed rankings as ME-rankings 75
4.7 Discussion 77

5 Analysis of benchmark problems 79
5.1 Property inheritance and transitivity 80
5.1.1 Irrelevance 82
5.2 Conflicting inheritance and specificity 82
5.3 Exceptional inheritance 85
5.4 Multiple inheritance 88
5.5 Discussion 91

6 Comparing LEX and Z^+ with ME 92
6.1 Comparison of LEX with ME 93
6.2 Comparison of Z^+ with ME 98
6.3 Dynamic behaviour 102
6.4 Summary 105

7 Constructing and testing default knowledge bases 106
7.1 Creating a default knowledge base 106
7.2 How to use DRS 108
7.3 Complexity of DRS 110

8 Conclusion 113
8.1 Review of thesis 113
8.2 Future uses of the ME approach 116

Chapter 1

Introduction

1.1 Motivation

One of the defining features of human intelligence is its adaptability and robustness in an uncertain world. The ability to think on one's feet and cope with unexpected events or dynamic environments is taken for granted. The behaviour of computers or software agents, on the other hand, is deterministic—they are simply machines following instructions. The subject of this thesis is how to simulate patterns of commonsense human reasoning which, firstly, can be justified as a rational way to reason from incomplete information, and, secondly, can be implemented as a formal computational theory of reasoning.

As computer programs, or software agents, are being given more and more responsibility, and as the tasks they perform become more complex, it is increasingly difficult to enumerate all possible situations they may encounter; nor indeed may it be possible to prescribe exactly what their behaviour should be in some highly exceptional or unforeseen circumstance. Ideally, it would be desirable to be able to give a high level specification of behaviour for a set of known or common occurrences, i.e., some general rules, and allow the agent to make decisions which incorporate these "boundary conditions" while giving it the freedom to decide what to do "on the hoof". Such instructions would provide a concise method of specifying behaviour and would allow additional requirements that override normal behaviours to be added independently.

Ultimately, one would like to be able to give general instructions to an agent and feel confident that, while not every eventuality has been covered explicitly, the agent is capable of exhibiting "common sense" when it encounters something unusual and of behaving appropriately. While varying degrees of robustness may be required of agents—some types of mistakes may be acceptable—it may also be

necessary to guarantee that an agent will not do something completely stupid. In order to have confidence that this will not happen, it is necessary to understand and to be able to predict the reasoning processes of the agent.

The area of artificial intelligence which focuses on these issues is known as nonmonotonic reasoning, so-called because beliefs do not necessarily increase with extra knowledge. Nonmonotonic reasoning is an attempt to capture the sort of everyday reasoning which humans perform and which cannot be modelled using classical logic alone. For example, suppose that an agent has a piece of background knowledge which says that bears usually live in the woods. If the agent encounters a bear it may jump to the conclusion that the animal lives in the woods. However, suppose the agent encounters a bear which lives in captivity in a zoo, known to be situated in a city nowhere near any woods. Now, this extra information interferes with the normal conclusion that the bear lives in the woods; in fact it specifically contradicts it. What belief should the agent come to regarding the bear? Does it live in the woods or not? The consensus is that the agent should retract its belief that the bear lives in the woods and substitute the belief that it does not, since it has more specific information about where the bear lives. The fact that learning *more* about the bear has led to a *retraction* of beliefs is what makes this form of reasoning different from classical logic. Classical logic is monotonic: if something can be proved from a given set of formulæ, then adding to these formulæ cannot lead to that belief being retracted (although it can lead to an inconsistent set of beliefs). In contrast, commonsense reasoning appears to be nonmonotonic, so that learning more information can lead to radical changes in beliefs—including retraction. The subject of this thesis is how to formalise this type of reasoning.

An agent may reason about an uncertain or incomplete environment and come to hold beliefs which are useful for making decisions under uncertainty. However, if these assumptions about the state of the world, or defeasible beliefs, subsequently turn out to be incorrect, the agent must be able to recognise what has happened and revise its beliefs and actions accordingly. The beliefs of such an agent are not necessarily truths about the world, but its own version of reality based on what it knows to be true and what it assumes to be true based on its background knowledge. As in the real world, different agents may come to hold different beliefs given the same information since they may be using different models of the world as background knowledge. In this thesis, the background knowledge of an agent will be encoded using default rules, that is, general rules which normally hold but which may be overridden should exceptional circum-

stances arise, and the beliefs of an agent will be logical formulæ which it accepts until forced through new circumstances to revise.

Now, while most people may agree on what constitutes common sense in some obvious cases—as in the example above—this is nowhere near a formal specification of how to perform commonsense reasoning. In fact, one of the main difficulties in this area is that, though different types of nonmonotonic reasoning are easily identified, a general formalisation is lacking since it is hard to specify precisely what behaviour is required. Problems appear to arise because systems have been designed to satisfy high level, but possibly incomplete, specifications. Though some systems have captured many of the perceived behaviours, they have subsequently been found either to fail to sanction some intuitively correct conclusions, or to have counterintuitive side effects which are hard to justify. Many different ways of attacking the problem have been attempted, some of which are surveyed in chapter 2.

The approach taken in this thesis is somewhat different. Instead of trying to reproduce high level behaviours, a less direct but more semantically transparent route is taken. The idea is that by providing a clear interpretation for what default rules, defeasible beliefs and inference mean, it is possible to produce a system of default reasoning whose foundations can be independently assessed so that the resultant default consequences are both explainable and predictable. The result is a sound and logically consistent framework for performing default reasoning from which patterns of commonsense reasoning emerge as properties rather than being design features. This thesis argues that such a framework provides an objective account of default reasoning which can be used to generate justifiable default inferences and, therefore, as a benchmark against which other default systems can be compared and judged.

1.2 Requirements of default reasoning

Default reasoning as a model of commonsense reasoning is not logical deduction. While logic deals with truth and certainty, common sense deals with uncertainty: it allows people to come to some conclusion in the absence of complete information. For example, if the sky looks dark and cloudy, one would usually choose to wear a raincoat and carry an umbrella—common sense says that it is very likely to rain and therefore one takes precautions, although it is by no means certain that it will rain. Attempting to model this type of reasoning using default rules, or rules with exceptions, requires both a formal model for the rules themselves and a

sound procedure for manipulating them. This section looks at what meaning can be ascribed to defaults and establishes the general requirements against which the soundness of a given procedure can be assessed.

In broad terms, there are two different approaches to default semantics: the extensional approach and the conditional approach. The extensional approach treats defaults as specialised rules of inference and provides procedures which determine when a default should be applied and its consequences accepted. This might involve explicitly checking that adding the conclusion of one default to a theory does not render that theory inconsistent, or providing pre-conditions in the antecedent of a default to prevent its application in abnormal circumstances. The advantage of using this approach is that defaults may be represented by some extension of first order logic, allowing the default mechanism to be implemented using existing theorem-proving techniques. The disadvantage is that the reasoning process needs to be guided, using exceptions explicitly to obtain the desired default conclusions. Moreover, when new defaults are added it may be necessary to recode the old ones to take account of new exceptions. This approach to default reasoning seems contrary to its original purpose, however, since if the "correct" conclusions were known in advance then using defaults would be unnecessary. Examples of extensional systems include default logic and circumscription, both of which are reviewed in the following chapter.

The conditional approach to defaults treats them as constraints on a consequence relation, that is, the relation must contain at least the defaults themselves. By extending the consequence relation to contain other defaults, more sophisticated default reasoning can be obtained, but there may be many different ways of extending a set of defaults, leading to different systems of default reasoning. The advantage of the constraint-based approach is that it is not necessary to make explicit reference to exceptional circumstances; it is the way the consequence relation is extended that determines default conclusions. This means that new defaults can be added modularly, as and when required, to obtain more sophisticated consequence relations. This is the approach to defaults that will be taken in this thesis with clear definitions for the meanings of default constraints and default inference.

The difficulty in developing a general theory of default reasoning is that to check whether it is a good model one needs something against which to test it. But such a benchmark is exactly what is being sought from the general theory, leading to rather a circular argument. However, there are some very general default behaviours which have been widely accepted as requirements which any default

reasoning system should satisfy. The remainder of this section characterises these behaviours so that at least some properties of default systems can be tested.

One of the most obvious uses for default rules is that they allow one to make generalised statements about groups of individuals. For example, the default "humans have two legs" should allow one to conclude that an arbitrary human has two legs. It should also allow one to conclude that an extension of the default applies for an arbitrary subgroup of humans, e.g., "female humans have two legs". In this way the original default is a concise way of representing information about a wide group of individuals. The reasoning process uses defaults to make more specific inferences as and when necessary. In exceptional cases, when a default does not apply, any incompatible beliefs can be retracted since they are defeasible; encountering an exceptional individual, e.g., a one-legged human, should not cause any problems for the reasoning process. Indeed, a whole subgroup may be exceptional, e.g., land-mine victims whose legs have been blown off. For this exceptional subgroup another default, which conflicts with the original but is more specific, may be applicable, e.g., "human land-mine victims do not normally have two legs". This default should override the original one by virtue of it being more specific. The idea of using general rules is that a concise representation of certain features can be given with exceptional cases being governed by more specific rules. A successful default reasoning system must be capable of applying the appropriate rule or of resolving the conflict which arises.

This example illustrates two important behavioural requirements of default reasoning: *property inheritance*, or the ability to inherit features from a superclass; and respect for *specificity*, or the ability to override general rules when more specific ones are available. Specificity has a role which is complementary to property inheritance: in a more specific situation properties are normally inherited unless there is a special rule which explicitly overrides the more general one. For example, if one can infer that "female humans have two legs", then one should also be able to infer that "Egyptian female humans have two legs", since there is no reason to suppose that being Egyptian has any relevance. In the same way, the default "Egyptian female human land-mine victims do not have two legs" should also be inferred, extending the more specific default with irrelevant information. So the default reasoning process must be capable of discounting features about the environment which are *irrelevant* and ignoring them, as well as being able to identify when some feature is relevant and leads to an exceptional or more specific rule being applicable.

The example of two-legged humans is a simple one and the default inferences to be drawn from it are intuitively obvious. More complicated interactions occur when two defaults are both applicable but have contradictory conclusions. The classic example of this is given by the Nixon diamond. Two defaults state that "quakers are pacifists" and "republicans are non-pacifists": given Nixon, who is both a quaker and a republican, what can be assumed about his stance on pacifism? If this is the only information available, one cannot draw any clear-cut conclusion since this would imply that one default dominates the other. Nixon may be a pacifist or not, or he may be ambivalent, but no fair reasoning process can decide this based only on the information supplied. Any sound default reasoning mechanism must preserve this *indifference* since an arbitrary resolution one way or the other implies that the defaults are not being treated equally. Since, ultimately, these rules will be reduced to symbols manipulated by the reasoner, a re-coding of the problem could lead to a different conclusion thus leading to an inconsistency in the inferences sanctioned on different occasions. Clearly, such behaviour would be undesirable.

However, in the Nixon diamond, it may be the case that some rules do hold more strongly than others. For example, while both rules may be valid, it may be felt that quakers have a stronger propensity for pacifism than republicans do for non-pacifism—perhaps religious beliefs are held more firmly than political ones. In such cases it would seem reasonable that the rule for quakers should be applied rather than the rule for republicans, and Nixon should be assumed to be a pacifist. Contrast this with a situation in which both rules were thought to hold equally strongly but it was also known that quakers were usually republicans. In this case the rule for quakers is more specific than that for republicans since they are already known to be a subclass of republicans; so again, Nixon should be assumed to be a pacifist, but for a different reason. This illustrates two different ways in which defaults may override each other and potential conflicts may be resolved. In one case it is the relative *strength* of the defaults which is important, whereas in the other their strengths are irrelevant since one rule is clearly more *specific* than another and will override it regardless of their relative strengths.

This section has identified some fairly high level behaviours which a default reasoning process should exhibit. These include property inheritance, priority for more specific or stronger defaults, irrelevance, and indifference. Although these have been described quite loosely, they still provide a basic specification for a system which reasons with default information. Originally, these behaviours were

mainly identified in association with benchmark problems which researchers in nonmonotonic reasoning have addressed. In chapter 5, those benchmark problems and the behaviours they represent are examined in more detail with reference both to existing systems and the general theory proposed in this thesis.

1.3 Inference using maximum entropy

In attempting to develop a general theory of default reasoning, one must avoid at all costs making arbitrary assumptions. However one must start somewhere, and the starting point of this thesis has been to select a representation for defaults according to the conditional interpretation. In fact, there is an extremely successful conditional semantics for defaults, called the ε-semantics, based on a non-standard probabilistic interpretation[1]. Using probability theory to model this type of reasoning allows the logical laws of probability to be applied in order to generate default inferences. The ε-semantics has been widely accepted as providing some of the core behaviour required of default reasoning, in particular, probability is naturally nonmonotonic. Unfortunately, however, it does not capture all of the requirements as described in the previous section. Clearly, in order to extend the semantics further so as to obtain more inferences, some further assumption is required. But how can making such an assumption be justified?

A similar problem, which has occurred across several fields of scientific research, involves selecting a probability distribution which is constrained to some extent but not sufficiently to determine it precisely. Now, if just one distribution is required from all those consistent with the constraints, which one should be selected? Moreover, if this procedure is to be applied repeatedly, is there a way of doing so that guarantees that the choice made is not arbitrary? The answer comes from the use of a quantity called *entropy*, which can be described as measuring the uncertainty inherent in a probability distribution. The entropy, H, of a probability distribution, P, is given by:

$$H[P] = -\sum_{p_i \in P} p_i \log p_i \tag{1.1}$$

Consider the probabilities associated with throwing a die. If the die is fair, then each face has an equal chance of coming up, so with six faces each has probability $\frac{1}{6}$. The probability distribution which represents this is given by $P(i) = \frac{1}{6}$ for $1 \leq i \leq 6$ which has an entropy of $H[P] = 1.79176$. Now this distribution leaves the thrower in the most uncertain state as to what face will come up, since any face

[1]The ε-semantics will be described in detail in sections 2.6 and 3.1.

is as likely as any other. The reader can easily verify that any other distribution of probabilities leads to a lower value of entropy: maximum entropy equates to maximum uncertainty. In particular, if it is known for sure that one face will come up, i.e., if it has a probability of 1 and all the other faces have a probability of 0, then the entropy is 0 and the outcome is certain.

The entropy function was derived by Shannon and Weaver in the context of information theory as the unique solution to a set of functional equations which describe the necessary behaviour of a measure of "information" (Shannon & Weaver 1949). In fact, as Jaynes points out, it is easier to think of entropy as measuring the "degree of ignorance", i.e., the uncertainty, of an observer confronted with a probability distribution (Jaynes 1979).

But how is this measure to be used as a method of inference? The answer is simple. Given a set of constraints which are known to hold, compute and select that distribution which maximises the entropy function; this distribution is the most uncertain and therefore represents the least biased estimate of the true distribution. This procedure has become known as applying the *Principle of Maximum Entropy* (Jaynes 1979). The soundness and consistency of using it as a method of inference has been shown separately by Shore and Johnson (1980), and by Paris and Vencovská (1990, 1997). In fact it is claimed that ME inference "provides the only consistent model of inductive inference" (Paris & Vencovská 1990).

Given that the ε-semantics for defaults provides a core for the perceived requirements of default reasoning but fails to satisfy some of the more adventurous ones, e.g., the assumption of indifference, a good test of these requirements would be to compare them with solutions obtained using maximum entropy inference. Since ME inferences are the least biased, any deviation from them would imply that some additional assumptions underlie the requirements. Against this, it may turn out that the requirements are consistent with ME inference, in which case it seems reasonable to conclude that the ε-semantics extended using ME provides a general theory of default reasoning—the ultimate goal of this thesis.

1.4 Overview of thesis and main contributions

This chapter has given the general background motivation for the work of the thesis. In fact what will be presented is the derivation of a system of default reasoning which, it is argued, represents the most acceptable means of extending a set of defaults given the probabilistic (ε) semantics for defaults. However, since there are already several proposals which extend the ε-semantics one way or another, it is

certainly possible to compare and contrast the proposed system against these alternatives, therefore producing a (hopefully) convincing argument for its superiority. This is the aim of the thesis which proceeds as follows.

In chapter 2, the work of the thesis is put in context. A review of other basic approaches to nonmonotonic reasoning is followed by a more detailed look at proposals which use similar concepts to those found throughout the thesis. While these alternatives may not be directly comparable to the ε-semantics systems, they do use some similar motivations and representations. Some other applications of maximum entropy to artificial intelligence are also reviewed.

In chapter 3, systems of default reasoning which are based directly on the ε-semantics are described in detail. The purpose is to provide all the technical background required for the development of the new approach. Theorems and algorithms for these systems are given where appropriate, although proofs are omitted. A benchmark example of default reasoning, which incorporates most of the desirable behaviours, is applied to all these systems to demonstrate where they succeed and where they founder. This illustrates how each system successively comes closer to attaining the full requirements of a default reasoning system.

In chapter 4, the main new work of the thesis commences with a discussion of the assumptions underlying the original work of Goldszmidt *et al.* (1993) and how these differ from those of the new approach. Under revised assumptions, mainly involving a specific requirement that each default be assigned a strength relative to the others, the new ME approach is derived and its limitations are analysed. A new algorithm is given which is shown to be sound and a condition which can be used to test for the uniqueness of its output is identified.

In chapter 5, the new ME approach using variable strength defaults is applied to the appropriate benchmark examples from the nonmonotonic literature and is shown to handle them in ways which match commonsense intuitions. At the same time, the meaning and purpose of the examples is critically assessed with respect to the answers obtained from ME. Bearing in mind that the ME approach is a formal system with a clear semantics, in some cases it is possible to use it in a systematic way in order to assess in which directions intuitions should lead. Comparing the intuitive answers with the ME answers can lead to a reassessment of ways that problems are encoded, enabling a clearer description of the problem itself.

In chapter 6, the ME approach is compared with two existing systems which were presented in chapter 3. The comparisons look for similarities between the systems to see in which ways they are related. It is shown that the LEX system is

a close relation of the ME approach but that the latter subsumes the former. It is also shown that the similarities between system Z^+ and ME are only superficial. Finally, all three systems are assessed for reasonableness at the meta-level. These three quite different comparisons are used to argue that ME is a superior system of default reasoning.

In chapter 7, a short guide to encoding background knowledge as defaults using the ME approach is given. Following this is a user-guide for a program called DRS which can be used to test default knowledge bases for entailment under the default systems based on the ε-semantics given in chapters 3 and 4. The complexity of the program for each of these systems prohibits its application to larger problems, but it is capable of handling all examples described herein.

Chapter 8 concludes by arguing in support of the main thesis and discussing directions for future research.

Chapter 2

Background

This chapter gives some general background material along with more specific recent research which is more or less related to the work of this thesis. The following chapter will describe in detail the default reasoning systems which are more directly relevant. Firstly, some basic nonmonotonic formalisms are described; then, the generalisation of such systems into the preferential model semantics; and then, other uses of qualitative probabilities and infinitesimals. Finally, some uses of maximum entropy in AI are described.

2.1 Circumscription

One of the first formal theories of nonmonotonic reasoning was McCarthy's *circumscription* (McCarthy 1980). The ideas it incorporates can be traced to some fundamental concepts used in computer science including *negation as failure*, which arose during the development of logic programming (Clark 1978), and the *closed world assumption*, which arose in database theory (Reiter 1978). Underlying these ideas was a central theme: in the absence of positive information that some fact holds, assume that it does not. The justification for such an assumption is that, if a problem is soluble, all relevant information required to solve it must have been given. This rule of thumb can be used to reason logically. McCarthy described his original definition as follows:

> Circumscription is a rule of conjecture that can be used by a person or a program for "jumping to certain conclusions". Namely, *the objects that can be shown to have a certain property P by reasoning from certain facts A are all the objects that satisfy P.* (McCarthy 1980)

The procedure is to use a second order formula to extend a first order theory by minimising the extensions of certain predicates. This means that circumscription

does not involve a new logic, unlike other approaches. However, while the axiom of circumscription offers a general rule for filling in gaps in a logical knowledge base, it requires the selection of particular predicates as parameters, and says nothing about which should be selected nor in what order the process should be applied to several predicates. So how can circumscription be used to perform default reasoning?

To apply circumscription to problems which require "common sense", McCarthy later proposed modelling default rules as implicative formulæ which included an abnormality predicate that could be used to block the right-hand side of the implication (McCarthy 1986). A default formula would then be of the form:

$$\texttt{bird(x)} \wedge \neg\texttt{ab(x)} \rightarrow \texttt{fly(x)}$$

and, by circumscribing the abnormality predicate `ab(x)`, those objects which were not known to be abnormal would be assumed to be normal. The implication would then be applied and objects which were not abnormal would exhibit normal bird attributes. When explicitly abnormal objects were encountered, the implication would be blocked. However, as was suggested in the introduction, a theory of default reasoning needs to be capable of resolving conflict among defaults and to handle exceptions with respect for specificity. This may require numerous abnormality predicates, e.g., $\texttt{ab}_1\texttt{(x)}$, $\texttt{ab}_2\texttt{(x)}$, $\texttt{ab}_3\texttt{(x)}$, etc., and the question of which to circumscribe, and in what order, can lead to different outcomes. Although using circumscription it is easy to obtain the intuitively correct solutions using defaults and abnormality predicates, this is accomplished by guiding the process in a heavy-handed way which rather obfuscates its use as a theory of default reasoning.

Nevertheless, circumscription is widely used and accepted as a useful mechanism for obtaining nonmonotonic behaviour, mainly because it is based on first order logic and can be incorporated into many other formalisms. It has been used in conjunction with the situation calculus (McCarthy 1968), the event calculus (Kowalski & Sergot 1986) and temporal logics (Shoham 1988), and there have been many variations on the original predicate circumscription including formula and prioritised circumscription (McCarthy 1986), and pointwise circumscription (Lifschitz 1987).

2.2 Default logic

Default logic was invented by Reiter (1980), who proposed treating a default rule as a licence to accept a conclusion given some evidence providing some criteria of consistency is met. This has sometimes been called a *presumptive* reading for

defaults because the conclusion is presumed to hold unless there is some indication to the contrary. A *default rule* is written:

$$\frac{\alpha : \beta_1, \beta_2, \ldots \beta_n}{\gamma}$$

and is a rule of inference triggered when the *prerequisite* (or *antecedent*), α, is known to hold, in which case the *consequent*, γ, is accepted unless the *justifications*, $\beta_1, \beta_2, \ldots \beta_n$ are inconsistent with the existing extension. A *default theory* is a pair, (D, W), where D is a set of defaults and W is a set of formulæ. An extension to a default theory is obtained from the set of formulæ W using the defaults and adding their consequents to the extension. The semantics of a default theory is given by its extensions—since there are no restrictions on the order in which defaults are applied, a default theory may have a unique extension, no extension at all (if the theory is inconsistent), or multiple extensions. There are also no restrictions on the formulæ which make up a default, so that the default $\frac{\alpha:\neg\alpha}{\beta}$ would be permitted though it is blatantly useless.

Default logic provides a simple, formal mechanism for default reasoning, but in isolation, it does not satisfy the requirements discussed in chapter 1 since the conclusions to be drawn from a default theory depend on how the extensions are interpreted. Generally, the logical consequences which are common to every extension are known as the *sceptical* or *cautious* conclusions while those which appear in just one extension are the *credulous* or *adventurous* conclusions. Since neither of these definitions provides satisfactory results from default logic alone, there have been many proposals which build on the framework.

Some proposals have restricted the type of defaults permitted in order to obtain more acceptable behaviour. For example, by using only normal defaults of the form $\frac{\alpha:\beta}{\beta}$, that is, defaults for which the consequent and justification coincide, the existence of a unique extension is guaranteed. Other restrictions in this vein include justified default logic (Lukaszewicz 1988), constrained default logic (Delgrande, Schaub, & Jackson 1994) and rational default logic (Mikitiuk 1996).

Other proposals have focused on attempts to obtain the "correct" default inferences using default logic. The first suggestion was to use semi-normal defaults, that is, defaults for which the justification logically implies the consequent (Reiter & Criscuolo 1983). An example of a semi-normal default might be:

$$\frac{\texttt{bird} : \texttt{flies} \wedge \neg\texttt{penguin}}{\texttt{flies}}$$

which is read as "if there is a bird and it is consistent both that it flies and that it is not a penguin, then assume that it flies". But using semi-normal defaults, like

using abnormality predicates, is impractical since it requires that the exceptional cases, which block the application of the default, be encoded explicitly. This rather neutralises the concept of using a default in the first place; to quote Pearl "[it] defies the very purpose of nonmonotonic reasoning" (see Pearl (1988), p. 516 and also Touretzky (1984)).

More interesting proposals have involved prioritising defaults to guide the order in which they are applied. For example, in prioritised default logic (Brewka 1989), the user explicitly gives a priority ordering over defaults. If they are completely pre-ordered, it is clear that a unique extension will be obtained. But even providing just a partial order may substantially reduce the number of extensions.

In fact, Brewka extended this idea even further by incorporating default priorities into the logical language itself (Brewka 1994). This system is capable of actually reasoning *about* the default priorities. By explicitly naming defaults, d_1, d_2, ..., he added a new type of formula, $d_1 \prec d_2$, which reads "d_1 has priority over d_2". Priority extensions are then those extensions which respect the extra constraints imposed by the new type of formula. However, the semantics of such theories becomes much more complex.

Antoniou gives details of all these variants of default logic and an interpretation of the systems in terms of their operational semantics (Antoniou 1997).

2.3 Inheritance hierarchies

Circumscription and default logic are both fairly general mechanisms for performing nonmonotonic and default reasoning. Others have used defaults more specifically to capture relations between objects. Inheritance hierarchies are used to perform reasoning about class characteristics. The hierarchies are modelled using directed graphs with nodes representing classes or individuals, and links taking the role of defaults. There are two types of links: positive "IS-A" links, which represent class inclusion, e.g., $bird \rightsquigarrow fly$, $penguin \rightsquigarrow bird$; and negative, "IS-NOT-A" links, which represent inclusion in the complement of a class, e.g., $penguin \not\rightsquigarrow fly$. The conclusions sanctioned by an inheritance hierarchy relate to which paths in a graph are acceptable. When there are several paths joining nodes which point to different conclusions, one path may *pre-empt* another. For example, given the links mentioned above, the path $penguin \not\rightsquigarrow fly$ would pre-empt the path $penguin \rightsquigarrow bird \rightsquigarrow fly$ because it is shorter and therefore more specific.

The theory of inheritance hierarchies has focused on finding reasonable strategies for resolving conflicts to determine which paths should pre-empt others. What

"reasonable" means in this context, however, has been the subject of some debate (Horty, Thomason, & Touretzky 1990, Touretzky, Horty, & Thomason 1987). Touretzky's *inferential path distance* (1986) provides a mathematically sound and coherent method for resolving conflicts, but it seems that without a formal semantics from which to assess them, the relative merits of different strategies can only be based on intuitions. It is not clear that there is a correct way to resolve multiple inheritance issues (Sandewall 1986). This lack of consensus, though, has led to some interesting dilemmas, some of which will be examined in chapter 5.

More recently, it has been suggested that links could be graded as a means of guiding a more general strategy for path validation (Neumann 1996). This leads to a more flexible interpretation of inheritance hierarchies which allows for different conclusions based on the grades applied.

2.4 Preferential reasoning

In parallel with the development of formal systems for performing nonmonotonic reasoning, others were looking at more theoretical properties of these new systems. Gabbay (1985) was the first to attempt to formalise the high level behaviour of expert systems. He suggested that while, clearly, not all of the axioms of classical logic hold for a nonmonotonic system, nevertheless one can characterise its behaviour according to some subset of these rules. For example, it seems reasonable to expect that from any formal system the contents of its knowledge base should be deducible. This can be formalised as the rule of *reflexivity*:

$$A \mid\!\sim A$$

Gabbay also introduced the symbol $\mid\!\sim$, denoting nonmonotonic deducibility or consequence, to replace the symbol $\vdash$ which denotes classical provability. He hoped to develop a framework by which the many new nonmonotonic systems could be compared and classified.

Along similar lines, Makinson (1988) defined rules which govern *cumulative inference operations*. These he referred to as inclusion, cumulative transitivity and cumulative monotony. He was careful to point out that these definitions made no reference to the object language or its connectives. Thus Makinson's analysis was an abstract means of classifying the behaviour of a great variety of formalisms. As will be discussed later, some behaviours which are intuitively central to systems which use logic as the object language may seem less important for systems which handle more complex objects (see section 6.3).

The work of Gabbay and Makinson could be classified as looking at the proof theoretic properties of nonmonotonic systems, but others looked at more model theoretic considerations. Shoham's *preferential logics* (1987) could be used to capture most nonmonotonic systems in a single unifying framework. He noticed that these systems could be represented by associating a preference relation over the models of a standard logic. The nonmonotonic behaviour of the systems came from the selection of particular preferred models of the underlying logic, be it classical, first order or modal, rather than having to consider all satisfying models.

This idea was extended by Kraus *et al.* in what has become one of the seminal papers on nonmonotonic reasoning (Kraus, Lehmann, & Magidor 1990). In it, the authors bring together, and prove the equivalence of, the preferential model semantics (a slight adaptation of Shoham's) and the proof rules for cumulative inference of Makinson. This remarkable result laid a foundation for nonmonotonic reasoning in the form of a sound and complete axiomatisation of nonmonotonic behaviour as defined by preferential models; this system of rules is now known as P, standing for preferential. Interestingly, the rules of P are also equivalent to those defined by Adams (1975) for his *probabilistic reasoning*, so P can also be thought of as standing for probabilistic (see sections 2.6 and 3.1). Preferential reasoning is very safe since it sanctions only those conclusions which hold in all preferential orderings consistent with a set of defaults. In particular, the rule system P satisfies the requirement of specificity so that defaults applying to a subclass automatically override those applying to its superclass. Because of this, preferential consequences have become known as the *core* of acceptable nonmonotonic behaviour (Geffner 1992). However, these consequences turn out to be insufficient to account fully for a theory of default reasoning since they fail to include some of the required common patterns of default inference, e.g., concluding that "red birds normally fly" given only that "birds normally fly" is not sound with respect to preferential reasoning, leading to a failure of the irrelevance requirement.

In search of a more complete system which was capable of fully realising the requirements, Lehmann and Magidor later examined what happened when further rules, in particular one called *rational monotonicity*, were added to P (Lehmann & Magidor 1992). They found that rational behaviour of a preferential model could only be obtained by restricting the preference relation to be a total ordering over models of the underlying language. However, such *ranked preferential models*, resulting in *rational consequence relations*, were not uniquely defined for a given set of defaults. Their search for an "ideal" ranked model resulted in the *rational clo-*

sure which was justified on the grounds that it was the ranked model for which all ranks were minimal. While this improved on the basic preferential inferences, it still failed to account for many behaviours, in particular exceptional inheritance.

Since the preferential and rational model semantics have been fully formalised by Kraus *et al.* (1990), a wide branch of nonmonotonic reasoning research has focused on consequence relations which satisfy these models, several of which are detailed in the following chapter.

2.5 Qualitative probabilities and ranking functions

Quantitative methods for handling uncertainty, e.g., probability theory (De Finetti 1974) and its application to Bayesian networks (Pearl 1988), possibility theory (Dubois & Prade 1988), and Dempster-Shafer theory (Shafer 1976), can be effective only when access to the underlying numerical values is available. Knowledge engineers have long been uncomfortable with extracting such numbers representing probabilities and conditional probabilities, even from expert practioners (Doyle 1990), since psychological studies have demonstrated that most people have only a rudimentary understanding of how probability works (Tversky & Kahneman 1981). Although follow-up studies have indicated that people who deal frequently with similar situations use more coherent models of numerical reasoning with respect to probability theory (Beach & Braun 1994), any estimates given by experts are likely to be incomplete or inconsistent. Moreover, there are indications that the outcome of expert systems may be relatively insensitive to the precise numbers used (Pradhan *et al.* 1996). For these reasons, qualitative and semi-qualitative approaches to uncertainty were considered to be viable alternatives. In particular, qualitative probabilities, and the closely related ranking functions, have been widely used to model beliefs and belief change.

Spohn was interested in modelling epistemic beliefs and how they change when revised (Spohn 1988, 1990). His argument was that numerical probabilities were inadequate to model *plain belief*, a concept best described in his own words:

> Intuitively, we have the notion of *plain belief*; we believe propositions to be true (or to be false or neither). Probability theory, however, offers no counterpart to this notion. Believing A is not the same as having probability 1 for A, because probability 1 is incorrigible; but plain belief is clearly corrigible. And believing A is not the same as giving A a probability larger than some $1 - \varepsilon$, because believing A and believing B is usually taken to be equivalent to believing $A\&B$. (Spohn 1990)

Spohn acknowledged, however, that an acceptable theory of belief change required a well-defined concept of conditionalisation. He proposed modelling epistemic states by *natural conditional functions*, though he now accepts that *ranking function* is a better term (Spohn 1998).

In Spohn's theory, a ranking function, κ, maps each proposition, a, b, c, ..., to a non-negative integer with at least one proposition having rank 0. These integers are intended to represent *degrees of disbelief*: a proposition with rank 0 is not disbelieved at all, while one with rank 1 is disbelieved to degree 1, etc. Belief in a disjunction of propositions is thus taken to be their minimal rank for, if one disbelieves a proposition to some degree, then one cannot have a greater disbelief in its disjunction with any other proposition:

$$\kappa(a \vee b) = \min(\kappa(a), \kappa(b))$$

The degree of disbelief in one proposition, b, conditioned on another, a, is then:

$$\kappa(b|a) = \kappa(a \wedge b) - \kappa(a)$$

These definitions for ranking functions form the *kappa calculus* which corresponds to an order of magnitude abstraction of probability measures over propositions, with minimisation replacing addition and addition replacing multiplication (Goldszmidt 1992).

To model the changes in epistemic states, which correspond to conditioning the ranking function with respect to some new evidence, Spohn provided a theory of conditionalisation which allows a proposition to be shifted by some degree producing a new ranking function. These rules correspond to Jeffrey's conditionalisation for probability distributions (Jeffrey 1965).

Although Spohn called his work a "non-probabilistic theory of inductive reasoning", he clearly recognises the connection to non-standard probabilities. This connection is more clearly explicated in Goldszmidt's thesis (Goldszmidt 1992), where he provides two methods for updating ranking functions to be used in conjunction with his system Z^+ (see section 3.4). J-conditioning is equivalent to Spohn's theory and is intended to shift (dis)belief in a proposition with "all things considered", i.e., after conditioning on the belief its new rank will be equal to the one specified. L-conditioning, on the other hand, is intended to represent shifting (dis)belief in the proposition with "nothing else considered" so that its new rank will shift by some degree rather than to a given level.

These theories of belief change use ranking functions to represent dynamic epistemic states so that, as beliefs are revised, new rankings are produced. Dar-

wiche and Pearl (1997) provided an extension of the belief revision postulates of AGM theory (Gärdenfors 1988, Katsuno & Mendelzon 1991) which allows for the preservation of some conditional beliefs when epistemic states are revised. They show that Spohn's theory of conditionalisation satisfies all their extended postulates and is therefore a model for belief revision based entirely on qualitative probabilistic reasoning.

The use of qualitative probabilities and ranking functions in belief revision is slightly different from their application to consequence relations (Kraus, Lehmann, & Magidor 1990). A consequence relation (if it is rational) can be represented by a single ranking function which is therefore a fixed component of an epistemic state (unless new *conditional beliefs* are learned), rather than one which changes as new beliefs arise. Thus ranking functions can be used to model both static and dynamic epistemic states.

Weydert's approach to ranked models combines the static and dynamic elements of belief change by building a canonical ranked model using J-conditionalisation and the notion of constructibility (Weydert 1996, 1998). Starting from the uniform ranking in which all worlds are ranked zero, incremental adjustments accommodate the increase in disbelief associated with those worlds which violate defaults. The defaults can be graded by assigning them real-valued positive strengths. Since there may be whole families of constructible ranked models, Weydert proposes an algorithm which constructs a canonical ranked model from the "bottom up", (i.e., most normal defaults added first) and applying a principle of maximising uniformity when shifting ranks. The system so obtained, named JZ, produces the same inferences as Goldszmidt's maximum entropy approach (Goldszmidt, Morris, & Pearl 1993) for "sufficiently simple" default sets, but the systems disagree on how redundant default information is handled. Weydert uses real numbers as ranks claiming that κ rankings "are not fine-grained enough to capture ME-inference on a semi-qualitatitive level" (Weydert 1998), since they use integers. However, it is not clear to this author that Weydert's framework requires the full positive reals since, if κ rankings are used to represent relative rather than absolute degrees of (dis)belief, they can easily be equated with rational rankings, allowing arbitrarily fine differences in beliefs to be represented.

Qualitative probabilities have also been applied in other areas of symbolic reasoning. Darwiche and Ginsberg showed that it is possible to generalise probability theory by an abstract symbolic representation which retains the "desirable features of the Bayesian approach for representing and changing states of belief"

(Darwiche & Ginsberg 1992). Their framework generalises probability theory, possibility theory and κ-calculus as well as more abstract states of belief. Darwiche extended the theory and algorithms of Bayesian networks to handle abstract symbolic beliefs (Darwiche 1992) and implemented a system which is capable of propagating κ-values, possibilities or probabilities (Darwiche 1994). This system has enabled research into the impact of substituting κ-values for real probabilities with results which suggest that the approximation works reasonably well for small (< 0.02) probabilities (Darwiche & Goldszmidt 1994, Henrion *et al.* 1994, Pradhan *et al.* 1996).

Goldszmidt and Pearl looked into how defaults could be used to model causal rules by mapping them into a Bayesian network (Goldszmidt & Pearl 1992). They defined *stratified rankings* as those which satisfied further constraints imposed by the structure of the network. Causal entailment is defined as entailment in all admissible stratified rankings. This idea was further explored by Geffner who proposed a refinement which produces just one canonical, stratified ranking determined by a user-defined parameter (Geffner 1996).

2.6 Infinitesimals

Adams was the first to propose that default inference could be modelled using a non-standard (infinitesimal) analysis of probabilities. In his *logic of conditionals*, a default is taken to be a statement of "high" conditional probability, e.g., the statement "birds normally fly" is taken to mean that the probability of any individual flying is high, given that it is a bird. What "high" means in this context is only relevant when one comes to consider default inference. In isolation, the actual conditional probability associated with a default is irrelevant, what is important is its connection to the conditional probabilities associated with other defaults. To determine whether a default is entailed, or inferred, from a given set, one examines whether its associated conditional probability can be made arbitrarily high by making those of the original defaults sufficiently high; this is what Adams called *probabilistic entailment* (Adams 1975).

In separate research, Pearl (1988) formalised the theory—which he called ε-semantics—in the following way. A default of the form $a \Rightarrow b$ represents the fact that the conditional probability of b given a is close to certainty, that is,

$$P(b|a) \geq 1 - \varepsilon \tag{2.1}$$

where ε, the infinitesimal parameter, is a real number close to zero. Thus as ε tends to zero so does the probability of the default being found to be false (since

$P(\neg b|a) < \varepsilon$). A default, $c \Rightarrow d$, is then ε-entailed by a set of defaults, Δ, if in all probability distributions which satisfy (2.1) for all defaults in Δ, $P(d|c) \geq 1 - O(\varepsilon)$, where $O(\varepsilon)$ is some function of the same order of magnitude as ε. That is, if for any $\delta > 0$, there exists $\varepsilon > 0$ such that $P(d|c) \geq 1 - \delta$, whenever (2.1) is satisfied for all defaults in Δ. To quote Pearl:

> In essence, this definition guarantees that an ε-entailed statement S is rendered highly probable whenever all the defaults in Δ are highly probable. (Pearl 1989)

This definition demonstrates why the exact value of what is meant by "high" is not material. Presumably, by specifying default information of any kind, one accepts that it holds for some threshold value of ε. Any inferences are relative to this value and can be thought of as holding with the same order of magnitude as the originals. Intuitively, the infinitesimal analysis amounts to pushing one's assumptions to the limit in order to determine what else they imply—a case of *taking one's ideas to the extremes*.

As has already been remarked with respect to the preferential model semantics, the ε-semantics sanctions some of the basic patterns of default reasoning; in particular, being founded in probability theory, it is naturally nonmonotonic which means that the specificity requirement is met through conditioning. Because this semantics is based on conditional probabilities, there is no problem with subclasses having features which are atypical of their superclasses and conditioning on such subjects guarantees that specificity will be respected. This is in sharp contrast with other systems such as default logic (Reiter 1980), where this conflict between subclasses and superclasses often leads to more than one possible interpretation or extension of a default theory. With ε-semantics there is no conflict as subclass attributes are inherited naturally. As Pearl said:

> ...it appears that the machinery of plausible reasoning is more in line with the rules of "almost-all" logic than with those of "support" or "majority" logics. (Pearl 1988) p.496

Since the ε-semantics forms the foundation for much of the work of this thesis, some other uses of infinitesimal analysis in this context are briefly reviewed.

Bacchus *et al.* developed an extremely rich and expressive language which can be used to model knowledge bases containing statistical information and default rules, among other things (Bacchus *et al.* 1996). Their *random worlds* method assumes that a knowledge base represents all that an agent knows about the world

and, according to the theory at least, by enumerating all possible first order models (the N random worlds) and assuming them to be equiprobable, a degree of belief in any given formula can be obtained from the proportion of those worlds in which it holds.

The relevant part of the random worlds model, viz-a viz this thesis, is the use of infinitesimals to model statistical and default knowledge. Since ratios of random worlds are rational, it is necessary for some degree of flexibility in defining which worlds satisfy statistical statements like $\|Hep(x)|Jaun(x)\|_x = 0.8$, meaning the conditional probability of any individual in a world having hepatitis given that he is known to have jaundice is 0.8. In particular, such a statement implies that the number of worlds in which individuals have jaundice must be a multiple of 5! To overcome this, Bacchus *et al.* allowed these statements to be true with respect to a class of *approximate equality relations*, denoted $\approx_i$, where each statement has its own relation, e.g., $\|Hep(x)|Jaun(x)\|_x \approx_1 0.8$. Default rules can be naturally accommodated in this framework, for example, the default "birds fly" becomes $\|Fly(x)|Bird(x)\|_x \approx_2 1$. The semantics for approximate equality is given in terms of a tolerance vector, $\vec{\tau}$, containing real numbers for each relation which effectively determine the level of approximation permitted. Since the actual number of random worlds N, and the actual values in $\vec{\tau}$, are not usually known, the degree of belief in a formula with respect to a knowledge base is defined in terms of the limit as N grows infinitely large and the values in $\vec{\tau}$ approach zero. However, there are occasions on which this limit does not exist—they call the limit *non-robust*—because it may depend on the manner in which the tolerance vector tends to zero. In effect, some knowledge bases are found to have multiple interpretations.

While the random worlds model provides a powerful language for expressing knowledge, and is shown to possess many desirable properties, the complexity is prohibitive of a realistic implementation. Despite this, it can be shown that for a simple language containing only unary predicates and constants, the degrees of belief can be found using maximum entropy as a computational tool (Grove *et al.* 1994). Indeed, they claim that in this case Goldszmidt *et al.*'s maximum entropy approach can be embedded in the random worlds framework. Because of this, the work of this thesis, which extends that of (Goldszmidt, Morris, & Pearl 1993), provides some insight into the interpretation of the random worlds semantics (see section 4.5).

Benferhat *et al.* (Benferhat, Saffioti, & Smets 1995) extended the ε-semantics to Shafer belief functions, rather than probabilities. They defined ε-belief functions

(εbfs) which are combined using Dempster's rule of combination. Each default, d, is treated as an item of evidence from a distinct source, with each having its own associated infinitesimal ε_d. They chose the *least committed* (Smets 1988) belief function to represent each default and cited an approximate expression for the plausibility of each model, m, derived from Dempster's rule of combination:

$$\mathtt{pl}_{\bigoplus}(m) \approx \prod_{\frac{d \in \Delta}{m \not\models d}} \varepsilon_d \tag{2.2}$$

That is, the plausibility of each model is approximately equal to the product of the infinitesimal values of the defaults that it falsifies (i.e., in which the antecedent of the default is true but its consequent is false).

Using this framework, Benferhat *et al.* were able to define different consequence relations which depend on more specific assumptions. In fact, using this framework they obtained systems which recover the simplest preferential consequence relation, P, and system Z (see section 3.2). The most successful of their suggested consequence relations is LCD consequence, standing for least commitment plus Dempster's rule. It turns out that LCD is a preferential consequence relation but not a rational one.

The interesting difference in the use of infinitesimals within the εbf framework is that a different one is associated with each default. As such, this approach differs from that of Adams, in which the parameters which constrain defaults tend to zero *at the same rate*. The approach to be taken in this thesis assigns different relative strengths to defaults, which represent the exponents of just one infinitesimal parameter. This makes it hard to compare with the εbf framework for which the infinitesimal parameters are independent and therefore incomparable. This, presumably, makes the consequence relations generated by εbf models more general, but it also makes it difficult to know how to interpret or assess the system.

The expression for the plausibility of a model given by (2.2) bears a striking resemblance to that derived for the probability of a model in the maximum entropy distribution (see section 4.2). However, plausibility is a more general belief measure than probability and it is not yet clear how, or whether, these two formalisms (ME and LCD) are related.

Weydert described a general framework for defeasible inference which also uses infinitesimals (Weydert 1995). He suggested several different interpretations of default constraints and attached individual infinitesimal parameters to each default. However, the entailment relations he defined are independent of these parameters, rather like in the random worlds approach mentioned above. Depending

on the type of constraint, the framework is capable of representing standard preferential entailment (Kraus, Lehmann, & Magidor 1990) and Goldszmidt's maximum entropy entailment (Goldszmidt, Morris, & Pearl 1993).

2.7 Maximum entropy inference in AI

Goldszmidt's maximum entropy (ME) approach to default reasoning (Goldszmidt 1992, Goldszmidt, Morris, & Pearl 1993), which is the starting point of the work of this thesis, will be described in detail in section 3.6. In the final section of this background chapter, other uses of ME in AI, and some of the criticisms it has received, are reviewed.

One of the first applications of maximum entropy to AI was proposed by Cheeseman (1983). Expert systems which store probabilistic knowledge are used to infer or make predictions about the probabilities of arbitrary combinations of propositional variables. The probabilities obtained from experts will usually underconstrain the full joint probability distribution to the extent that only a range of possible values for the probabilities can be found, or further assumptions are necessary to obtain a reasonable estimate of the distribution. In the theory of Bayesian networks, the causal structure of the domain is exploited by assuming causality can be equated with conditional independence (Neapolitan 1990, Pearl 1988). In contrast, Cheeseman proposed using ME updating to obtain the least biased estimate of the joint probability distribution. Although computationally intractable in the general case, his update method offered some possibilities for reducing the size of the problem.

However, some confusion arose between proponents of Bayesian networks and those of ME updating. In particular, suggestions that ME updating is inconsistent with the concept of causality led to an interesting paper by Hunter (1989). Pearl had argued against ME inference citing an experiment in which the result of tossing two coins is used to determine some event, which leads to an ME distribution in which the two coins become probabilistically dependent (see Pearl (1988), p. 463)—something counterintuitive from the standpoint of causality. Hunter gives an analysis of this puzzle and clears up the confusion which arises from the inadequacy of modelling causal information using conditional probabilities alone. By representing this type of information using *counterfactual conditionals* (Ginsberg 1986, Stalnaker 1975), it is possible to obtain results from ME updating which correspond to causal intuitions (Hunter 1989). Hunter's main argument is that the naive application of ME updating, or indeed other methods such as Bayesian up-

dating, will often lead to counterintuitive results. The problem lies in formalising the underlying situation rather than in choosing a method of inference.

It has been shown, separately by Shore and Johnson (1980, 1986) and by Paris and Vencovská (1990, 1997), that 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 but related derivations of the uniqueness of the maximum entropy method are possible. A third and more specific characterisation of inference using ME is derived by Kern-Isberner in the context of updating a probability distribution using only quantified probabilistic conditionals[1] as constraints (Kern-Isberner 1998). Using ideas which originate in the theory of conditional logics (Nute 1980), she considers the principal of conditional preservation as paramount. Using this as a postulate along with three other requirements of a functional concept, logical consistency and representation invariance, Kern-Isberner recovers maximum entropy inference as the only solution to the update problem. She also uses ME updating to demonstrate the validity of some deduction rules of conditional logic—for example, transitivity, specificity and reasoning by cases (Kern-Isberner 1997). 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 (Paris 1998). More recently, Lukasiewicz and Kern-Isberner have used approximate ME-models to search for computationally tractable techniques for solving probabilistic logic programming problems (Lukasiewicz & Kern-Isberner 1999).

Another use of ME is suggested by Rhodes in the context of incomplete causal (Bayesian) networks. Since the complete causal information for the network may not always be available, the missing data can be supplied by computing the ME distribution (Rhodes & Garside 1995). By analysing the algebraic structure of the ME distribution for various types of causal tree, Rhodes and others have found that the same conditional independence assumptions which underlie the Bayesian propagation algorithms can be used to develop algorithms which propagate the ME probabilities. In fact, for many cases, the ME distribution can be computed "bottom-up" from the leaves of the tree (Garside & Rhodes 1996,

[1] That is, conditionals of the form $A \rightsquigarrow B[x]$ where A and B are propositions and $x \in (0, 1]$.

Holmes & Rhodes 1998, Rhodes & Garside 1998). This means that the probabilities for each node can be computed independently of higher nodes, allowing iterative and computationally tractable (in some cases linear time) algorithms to be developed. The work of this thesis attempts to find abstractions of ME distributions rather than their actual numerical values, but it is interesting to note that the ME algorithm developed in chapter 4 is also based on an iterative approach.

The ME approach has also been criticised as being "representation dependent" (see, for example, (Halpern & Koller 1995, Jaeger 1996)). To see what this criticism involves, consider the following example, taken from Halpern and Koller (1995), in which each knowledge base is intended to represent the information that one in every two birds is capable of flight:

$$KB_1 = [Prob(fly|bird) = 1/2]$$

$$KB_2 = [flyingBird \rightarrow bird, Prob(flyingBird|bird) = 1/2]$$

The ME distribution for KB_1 gives $Prob(bird) = 1/2$ while that for KB_2 gives $Prob(bird) = 2/3$. This may appear, at first glance, to contradict the fact that ME is a consistent method of inference since both knowledge bases are intended to represent the same information. But this criticism is unfounded. Inference using ME is consistent—from identical knowledge bases identical results will be obtained. Clearly KB_1 and KB_2 do not contain the same information, although superficially they may appear to. In KB_2, flying non-birds do not arise, and this restricts the number of worlds to 3 rather than 4, leading to the absolute probability of being a bird being higher for KB_2.

The use of maximum entropy is valid under the assumption that absolutely all the available information is contained in the knowledge base; if this is the case, the fact that each possible world is assigned an equal prior probability is the only reasonable place to start. But there is a more general variant of maximum entropy inference which is used when a prior probability distribution needs to be adjusted to account for extra information. This is the principle of *minimum cross-entropy*, which can be thought of as minimising the "distance" between the prior and posterior distributions (Shore & Johnson 1980). Maximum entropy is just a special case of applying this principle when the prior distribution is uniform. Halpern and Koller suggest that using minimum cross-entropy instead of maximum entropy alleviates the problem of representation dependence (Halpern & Koller 1995).

Clearly all inference procedures are sensitive to the information with which they are supplied. In particular, it is important for their users to understand that,

although according to their subjective interpretation two encodings of a problem may be identical, it does not follow that they are semantically equivalent from the perspective of the inference procedure. This appears to be the cause of the confusion surrounding the representation dependence of ME inference which Paris and Vencovská termed a "non-problem", saying

> [using minimum cross-entropy] in no sense "solves" the so-called problem of representation dependence; it merely provides a safety net for the careless. (Paris & Vencovská 1997)

Chapter 3

Systems of default reasoning

This chapter provides the technical background for the thesis. The basic ε-semantics for defaults and its common extensions are presented in a unified format along with their related theorems[1] and algorithms. Most of the material covered is merely reproduced from the literature with the exception of the introduction to variable strength defaults (section 3.3), which is new, and the slight adaptation of system Z^+ and its associated algorithm, to make it fit more naturally with the other systems (section 3.4). A running example is used throughout the chapter to assess each system against the requirements for default reasoning given in chapter 1. The systems are presented roughly in order of their conception to illustrate how progress was made towards satisfying these requirements. At the end of the chapter, the objectives for the remainder of the thesis are set out.

3.1 System P and the ε-semantics

First some preliminary definitions and notation. 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 $m \models a \wedge \neg b$. A default rule, r, is said to *tolerate* a set of defaults, Δ, if and only if it has a verifying model which does not falsify any defaults in Δ; such a model will be called a *confirming* model of r with respect to Δ; a set which contains at least one default which tolerates it will be called *confirmable*. The default $a \Rightarrow \neg b$ is called the *converse* of $a \Rightarrow b$.

The ε-semantics (Adams 1975, Pearl 1988) for a default is that it represents a

[1] The theorems are only cited. For their proofs the reader is referred to the source material.

constraint on a probability distribution (PD) such that the conditional probability associated with a default, $a \Rightarrow b$, is constrained to be greater than $1 - \varepsilon$ for some infinitesimal parameter $\varepsilon > 0$.

$$a \Rightarrow b \qquad \equiv \qquad P(b|a) \geq 1 - \varepsilon \tag{3.1}$$

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.

Given a set of defaults as background knowledge, how can this be used to infer further information? Since the knowledge is encoded in default form, and since defaults represent constraints, it seems appropriate to look for other constraints which are implied by the original set. All systems described in this chapter provide a means of inferring whether or not arbitrary defaults are entailed from some original set of defaults. This can also be thought of as a means of extending some set of defaults into a larger superset of all those defaults entailed under the given system.

The most basic form of entailment, called ε-entailment, is that sanctioned purely by the laws of probability (De Finetti 1974). Probabilistic axioms can be used to derive default constraints of a similar nature to those of the original defaults. A default, $a \Rightarrow b$, is ε-entailed by a set of defaults, $\Delta = \{a_i \Rightarrow b_i\}$, if its associated conditional probability, $P(b|a)$, can be made arbitrarily close to 1 when those conditional probabilities associated with the original defaults are made sufficiently close to 1. More formally:

Definition 3.1.1 *A default, $a \Rightarrow b$, is* ε-entailed *by a set of defaults, $\Delta = \{a_i \Rightarrow b_i\}$, iff for all $\delta > 0$ there exists $\varepsilon > 0$ such that $P(b|a) \geq 1-\delta$ if, for all $a_i \Rightarrow b_i$, $P(b_i|a_i) \geq 1-\varepsilon$.*

For example, let $\Delta = \{a \Rightarrow b, a \Rightarrow c\}$ which gives rise to the two constraints:

$$P(b|a) \geq 1 - \varepsilon \qquad\qquad P(c|a) \geq 1 - \varepsilon$$

Consider the default $a \wedge b \Rightarrow c$. The conditional probability associated with this default, $P(c|a \wedge b)$, can be found by conditioning $P(c|a)$ on b to give:

$$P(c|a) = P(c|a \wedge b)P(b|a) + P(c|a \wedge \neg b)P(\neg b|a) \tag{3.2}$$

Substituting $1 - P(b|a)$ for $P(\neg b|a)$ and rearranging, gives:

$$P(c|a \wedge b) = \frac{P(c|a) - (1 - P(b|a))P(c|a \wedge \neg b)}{P(b|a)} \tag{3.3}$$

As $\varepsilon \to 0$, clearly $P(c|a \wedge b) \to 1$. Thus Δ ε-entails $a \wedge b \Rightarrow c$ and, by a symmetrical analysis, $a \wedge c \Rightarrow b$.

This result can be expressed in the style of a rule of inference so that from the two defaults, $a \Rightarrow b$ and $a \Rightarrow c$, the new default $a \wedge b \Rightarrow c$ may be inferred. This rule has become known as *cautious monotonicity*, so named to reflect the idea that learning a fact, b, that was already presumed to hold, should not lead to the retraction of any other previously inferred beliefs, c.

$$\frac{a \Rightarrow b, a \Rightarrow c}{a \wedge b \Rightarrow c} \quad \text{Cautious Monotonicity}$$

Using similar analysis it can easily be shown that the following other rules of inference are also probabilistically sound[2], and so lead to defaults which are ε-entailed, called ε-consequences.

$$\frac{\forall a}{a \Rightarrow a} \quad \text{Reflexivity}$$

$$\frac{\models b \to c, a \Rightarrow b}{a \Rightarrow c} \quad \text{Right Weakening}$$

$$\frac{\models a \leftrightarrow b, a \Rightarrow c}{b \Rightarrow c} \quad \text{Left Logical Equivalence}$$

$$\frac{a \Rightarrow b, a \Rightarrow c}{a \Rightarrow b \wedge c} \quad \text{And}$$

$$\frac{a \Rightarrow c, b \Rightarrow c}{a \vee b \Rightarrow c} \quad \text{Or}$$

An equivalent axiomatisation of these probabilistically sound rules of inference are shown by Adams (1975) to be complete with respect to ε-entailment[3]. That is, for any default which is ε-entailed by a set of defaults there exists a proof sequence using the above rules of inference from that set to the entailed default. Kraus, Lehmann and Magidor named this rule system P because it represents a sound and complete set of axioms for preferential reasoning (Kraus, Lehmann, & Magidor 1990) (see section 2.4). Thus preferential consequences, P-consequences and ε-consequences coincide. The set of all ε-consequences of a set of defaults, Δ, is known as the P-closure of Δ and denoted Δ^{P}.

Adams gives two important theorems which allow ε-entailed defaults to

[2]Lower bound functions for the conditional probabilities associated with the derived defaults are given in Bourne and Parsons (1998).

[3]Adams used the term P-entailment for *probabilistic* entailment.

```
consistency-check algorithm
Input:  a set of defaults, Δ.
Output:  true or false.
 [1] Set Γ = Δ, answer = true, i = 0.
 [2] While Γ is non-empty and answer is true:
     (a) Let Δ_i be those defaults in Γ which tolerate Γ.
     (b) If Δ_i is empty let answer = false.
     (c) Let Γ = Γ − Δ_i and i = i + 1.
 [3] Return answer.
```

Figure 3.1: The consistency-check algorithm

be determined algorithmically. The first theorem relates to his definition of ε-consistency:

Definition 3.1.2 *A set of defaults, $\Delta = \{a_i \Rightarrow b_i\}$, is* ε-consistent *iff for all $\varepsilon > 0$ there exists a probability distribution, P, such that, for all defaults $a_i \Rightarrow b_i$, $P(b_i|a_i) \geq 1 - \varepsilon$.*

The theorem connects ε-consistency with confirmable subsets of Δ.

Theorem 3.1.3 (ε-consistency) *(Adams 1975) A set of defaults, $\Delta = \{a_i \Rightarrow b_i\}$, is ε-consistent iff every non-empty subset of Δ is confirmable.*

This theorem leads to a simple algorithm for testing the ε-consistency of any set of defaults which is given in figure 3.1. Given theorem 3.1.3, it is possible to test for ε-consistency by constructing an ordered partition of the set. If such a partition can be formed, the set is ε-consistent. The algorithm works by repeatedly finding and removing all defaults which tolerate the set, and forming a new partition set with them. The process is repeated until a set is reached in which all defaults tolerate each other. If no such set is reached, the original set is ε-inconsistent. Thus, when Δ is ε-consistent, a by-product of the algorithm is the constructed partition, $\Delta_0 \cup \Delta_1 \ldots \cup \Delta_n$, called the Z-*partition* by Pearl (1990). The Z-partition will be discussed in more detail in section 3.2.

The second theorem shows that a default, $a \Rightarrow b$, is ε-entailed whenever adding its converse, $a \Rightarrow \neg b$, to a set renders it ε-inconsistent.

Theorem 3.1.4 (ε-entailment) *(Adams 1975) A set of defaults, $\Delta = \{a_i \Rightarrow b_i\}$, ε-entails a default, $a \Rightarrow b$, iff the set $\Delta \cup \{a \Rightarrow \neg b\}$ is ε-inconsistent.*

Theorem 3.1.4 provides a method for testing whether an arbitrary default is ε-entailed by a set: simply add its converse and test for ε-consistency.

The ε-semantics therefore has two distinct strands: the rule system P provides a proof theory from which ε-entailed defaults can be constructed, while the consistency-check algorithm can be used to test arbitrary defaults for ε-entailment.

The following example demonstrates the kind of inferences possible with ε-entailment.

Example 3.1.5 (Penguins)

$$\Delta = \{b \Rightarrow f, b \Rightarrow w, p \Rightarrow b, p \Rightarrow \neg f\}$$

(the intended interpretation of this database is that birds normally fly, birds normally have wings, penguins are normally birds[4] but penguins do not normally fly).

Firstly, it can be seen that ε-entailment satisfies the requirement of specificity because the default $p \wedge b \Rightarrow \neg f$ is ε-entailed. This can be shown in two ways. By applying the rule of cautious monotonicity to $p \Rightarrow b$ and $p \Rightarrow \neg f$, the default $p \wedge b \Rightarrow \neg f$ can be derived directly; and, by applying the consistency-check algorithm to the set $\Delta' = \{b \Rightarrow f, b \Rightarrow w, p \Rightarrow b, p \Rightarrow \neg f, p \wedge b \Rightarrow f\}$, it is easily shown that $p \wedge b \Rightarrow \neg f$ is ε-entailed by Δ: $b \Rightarrow f$ and $b \Rightarrow w$ tolerate the whole set so $\Delta_0 = \{b \Rightarrow f, b \Rightarrow w\}$ and $\Gamma = \{p \Rightarrow b, p \Rightarrow \neg f, p \wedge b \Rightarrow f\}$; no default in Γ tolerates it so Δ_1 is empty and therefore Δ' is ε-inconsistent and $p \wedge b \Rightarrow \neg f$ is ε-entailed by Δ. Thus, in the case of two conflicting possibilities for penguin-birds, they may fly, because they are birds, or not, because they are penguins, the ε-semantics automatically selects that default conclusion favoured by the more specific default; in this case, since penguins are known to be a subclass of birds, the default relating specifically to penguins applies.

Secondly, it can be seen that ordinary property inheritance does not occur in the presence of irrelevant information. Consider whether red birds fly or not. According to the requirement of property inheritance, red birds should inherit the flying attribute since there is no reason to suppose that redness interferes with flying. However, $\Delta \cup \{r \wedge b \Rightarrow \neg f\}$ is ε-consistent and therefore $r \wedge b \Rightarrow f$ is *not* ε-entailed. The reason for this failure is that some probability distributions exist which are consistent with a default which states that "red birds normally do not

[4]Some may argue that penguins *are* birds, however this thesis is concerned with defaults rather than strict rules. Goldszmidt considered mixed knowledge bases with both strict and defeasible rules (Goldszmidt 1992); in this framework, a strict rule is simply a propositional formula.

fly". This means that ε-consequences, which are only those which hold in *all* PDs consistent with the original defaults, do not contain all the default conclusions theoretically required of default reasoning. In order to obtain these extra inferences, the condition of absolute probabilistic soundness will need to be relaxed.

Thirdly, and unsurprisingly given the second point, it can be seen that ε-entailment does not handle exceptional inheritance in a satisfactory way. Ideally, one would like the wing attribute of birds to be inherited by penguins, a subclass of birds, despite the fact that penguins are exceptional in the flying attribute. However, the set $\Delta \cup \{p \wedge b \Rightarrow \neg w\}$ is ε-consistent so the default $p \wedge b \Rightarrow w$ is not ε-entailed. □

So, despite its firm foundation in probability theory, the basic ε-semantics is clearly not sufficient to fully capture the kind of reasoning required of default systems. Being probabilistically sound, all ε-consequences are acceptable as default conclusions, but there are other defaults which, though not probabilistically sound, nevertheless ought, intuitively, to be entailed. These correspond to commonsense guidelines such as ignoring irrelevant information and assuming that the only exceptions to defaults which exist are those explicitly represented. One of the main difficulties in formalising this type of reasoning lies in the fact that these intuitions are hard to define precisely.

The ε-semantics therefore needs extending if it is to fully capture all the default reasoning requirements, but for this to be successful it must be done in such a way that can be seen to be objective and reasonable. Rather than attempting to satisfy ill-defined requirements, it would be preferable to find a rational method of extension. The following section looks at the initial attempts to achieve this.

3.2 Rational closure and system Z

The rule system P was studied extensively by Kraus, Lehmann and Magidor (1990) in their influential paper which describes the *preferential model semantics*. Preferential models[5] provide an alternative characterisation of default reasoning based on strict partial orders over models of a language (Shoham 1987), which sanctions exactly the same inferences as the ε-semantics.

Lehmann and Magidor (1992) proposed extending the proof theory associated with their semantics with an additional rule called *rational monotonicity*:

[5]To avoid confusion, preferential models are mathematical structures consisting of a language, $\mathcal{L}$, and a strict partial order, $\prec$, over its models, $\mathcal{M}$. The word "model" when unqualified, refers to members of $\mathcal{M}$, i.e., to models of the underlying language, $\mathcal{L}$.

$$\frac{a \Rightarrow c, \ \texttt{not}(a \Rightarrow \neg b)}{a \wedge b \Rightarrow c} \quad \text{Rational Monotonicity}$$

This rule is more adventurous than that of cautious monotonicity, described above, and it reflects the intuition that learning information which is not entirely unexpected, b, should not cause the retraction of any previously inferred beliefs, c (though obviously new beliefs may arise). The rule of cautious monotonicity requires that beliefs be maintained only in the presence of anticipated information.

The rule of rational monotonicity is not of the same form as those of system P, since it is a proof rule which requires that some default *cannot* be proved. As such, it can hardly be used to generate new inferences since these may subsequently block the validity of its own application[6]. However, an extension to a set of defaults may be termed *rational* if it satisfies this rule as a constraint. It turns out that there may be several rational extensions to a set of defaults. For example, given $\{a \Rightarrow c, b \Rightarrow \neg c\}$, a rational extension which does not contain $a \Rightarrow \neg b$ will contain $a \wedge b \Rightarrow c$, whereas a rational extension which does not contain $b \Rightarrow \neg a$ will contain $a \wedge b \Rightarrow \neg c$. Since both these inferences cannot belong to the same extension (it would be ε-inconsistent), there must be different rational extensions of the same set.

The work on this rule yielded interesting developments. Firstly, for any preferential model which satisfies rational monotonicity, the preference ordering over models is strict, which means that they can be totally ranked. This led to the definition of *ranked preferential models* (RPMs) in which each model is assigned an integer rank and the preference ordering is simply the less-than relation. Each RPM determines a so-called *rational consequence relation*. Secondly, it has been shown that the intersection of all inferences from all rational consequence relations is just the P-closure itself, that is, just those inferences sanctioned by P and the ε-semantics.

This indicates that there is nothing exceptional about rational inferences, i.e., none are satisfied in every rational consequence relation. In order to obtain more inferences, therefore, it will be necessary to select some subclass of RPMs, or just one, using further assumptions as well as the requirement that rational monotonicity be satisfied[7].

---

[6]In this sense it is similar to default logic (see section 2.2) and, like default logic, it too results in multiple extensions.

[7]Note that not all researchers agree on the adoption of rational monotonicity. Geffner (1992) specifically rejects RPMs in favour of partial orders which he uses to define his own system called *conditional entailment*. Since this requires giving up the underlying probabilistic semantics, it is not entirely clear how to interpret his argument-based system, and so conditional entailment stands at some distance from those systems described in the current chapter.

However, Lehmann and Magidor (1992) did find a distinguished RPM which they termed the *rational closure*. They identified a preference relation over RPMs and showed that for any finite set of defaults there is a unique RPM which is preferable to all others. They justify their definition of preference by using the idea that one situation is more unusual than another if the integer ranks assigned to models are lower in one RPM that in another. In effect they then find the least unusual RPM which turns out to be the one which assigns every model a minimal rank—this is what they call the rational closure. This criterion of minimality is the additional assumption that they make in order to arrive at a uniquely specified rational consequence relation for every set of defaults.

The mechanics of rational closure are more easily described with reference to system Z which was developed by Pearl (1990). System Z sanctions exactly the same inferences as those in the rational closure as discussed in (Goldszmidt & Pearl 1990). First, though, it is necessary to describe the ranking function representation which is used as an abstraction of the probability distributions of the ε-semantics. Ranking functions represent the total ordering of an RPM which gives rise to its rational consequence relation.

A ranking function can be thought of as an abstraction of a probability distribution under the ε-semantics. The rank of a model or formula corresponds to the exponent of ε of its probability in that PD. The default constraint on $a \Rightarrow b$ can be rewritten:

$$P(b|a) \geq 1 - \varepsilon \quad \equiv \quad P(\neg b|a) < \varepsilon \quad \equiv \quad P(a \wedge \neg b) < \varepsilon P(a)$$

that is, the probability of $a \wedge \neg b$ must be at least one order of ε less than that of a itself and, in particular, of $a \wedge b$.

Definition 3.2.1 *A* ranking function, κ, *is a mapping from* $\mathcal{M}$ *to the non-negative integers for which at least one model,* m, *has* $\kappa(m) = 0$. *This determines a preference ordering over models, so that*

$$\kappa(m) < \kappa(m')$$

means that m *is preferred to, or more normal than,* m'.

This function, κ, in turn determines a preference ordering over the formulæ of $\mathcal{L}$, where a formula is as preferred as its most preferred model, so that

$$\kappa(a) = \min_{m \models a} [\kappa(m)] \tag{3.4}$$

Equivalently, $\kappa(a) < \kappa(b)$ means that there exists an m such that $m \models a$ and for all m' such that $m' \models b$, $\kappa(m) < \kappa(m')$.

A default constrains a ranking function so that it is more normal to verify the default than to falsify it. A ranking function which satisfies the constraint is said to be admissible with respect to that default. More formally:

Definition 3.2.2 *A ranking function,* κ*,* satisfies *a default,* $a \Rightarrow b$*, or is* admissible *with respect to it, iff*

$$\kappa(a \wedge b) < \kappa(a \wedge \neg b) \tag{3.5}$$

Note that definition 3.2.2 leaves unspecified the exact difference between the ranks of the default's minimal verifying and falsifying models but clearly, since the ranks are integers, this difference must be greater than or equal to one.

As mentioned above, a ranking function over models corresponds to a rational consequence relation. To determine whether some consequent, b, is a consequence of some antecedent, a, with respect to a ranking function, κ, it is necessary to check whether κ satisfies $a \Rightarrow b$. If $\mid\!\sim_\kappa$ represents the consequence relation, then:

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

A given set of defaults will have infinitely many admissible ranking functions. Since rational consequence relations extend the preferential consequence relation of a given set, any ranking function which is admissible with respect to that set corresponds to a rational consequence relation which is a direct extension of its P-closure.

System Z is defined as follows: associated with each ε-consistent set of defaults, Δ, is its unique Z-partition, $\Delta_0 \cup \Delta_1 \ldots \cup \Delta_n$, being the by-product of the consistency-check algorithm (see figure 3.1). Let each default in Δ be assigned a Z-rank equal to the index of that partition set to which it belongs, that is, if $r_i \in \Delta_j$ then $\mathrm{Z}(r_i) = j$. The Z-ranking is defined as follows:

Definition 3.2.3 (Z-ranking) *A model is assigned a* Z*-rank of 1 plus the highest* Z*-rank of all defaults it falsifies or zero if it falsifies no defaults, that is:*

$$\mathrm{Z}(m) = \begin{cases} 0 & \textit{if } m \textit{ falsifies no defaults in } \Delta \\ 1 + \max\limits_{m \models a_i \wedge \neg b_i} \{\mathrm{Z}(r_i)\} & \textit{otherwise} \end{cases} \tag{3.6}$$

Because of the method by which the Z-partition is constructed, it easily follows that the Z-ranking is admissible: each default, r, in partition-set Δ_i has a minimal verifying model which confirms it with respect to $\Delta_i \cup \Delta_{i+1} \ldots \cup \Delta_n$; as it does not falsify any other defaults, this is a minimal verifying model of r and has a Z-rank of i. All falsifying models of r have a Z-rank of $i + 1$, or higher, hence $\mathrm{Z}(a \wedge b) < \mathrm{Z}(a \wedge \neg b)$ for all defaults, and the Z-ranking is admissible.

m	b	f	p	w	Z
m_1	0	0	0	0	0
m_2	0	0	0	1	0
m_3	0	0	1	0	2
m_4	0	0	1	1	2
m_5	0	1	0	0	0
m_6	0	1	0	1	0
m_7	0	1	1	0	2
m_8	0	1	1	1	2

m	b	f	p	w	Z
m_9	1	0	0	0	1
m_{10}	1	0	0	1	1
m_{11}	1	0	1	0	1
m_{12}	1	0	1	1	1
m_{13}	1	1	0	0	1
m_{14}	1	1	0	1	0
m_{15}	1	1	1	0	2
m_{16}	1	1	1	1	2

Figure 3.2: The Z-rankings for the penguin example.

Pearl (1990) proves that the Z-ranking is uniquely defined for any ε-consistent set, Δ. He also shows that it is minimal in the sense that no model can attain a lower rank in any other admissible ranking function.

Theorem 3.2.4 *(Pearl 1990) Given an ε-consistent set of defaults, Δ, the Z-ranking given by definition 3.2.3 is unique and minimal.*

The following example demonstrates the kind of inferences possible under system Z.

Example 3.2.5 (Penguins (cont'd))

$$\Delta = \{b \Rightarrow f, b \Rightarrow w, p \Rightarrow b, p \Rightarrow \neg f\}$$

The Z-partition of this database has two partition-sets:

$$\Delta_0 = \{b \Rightarrow f, b \Rightarrow w\} \qquad \text{and} \qquad \Delta_1 = \{p \Rightarrow b, p \Rightarrow \neg f\}$$

Here $\mathcal{L}$ has four atoms and $\mathcal{M}$ therefore has 16 models. Figure 3.2 enumerates these models along with their Z-ranks. Firstly, consider whether the default "red birds fly" is Z-entailed. Note that the proposition r (standing for red) is an addition to $\mathcal{L}$ but, while it doubles the number of models in $\mathcal{M}$, it has no effect on the Z-ranks of the models. Since no default in Δ refers to r, it must be irrelevant to the consequence relation produced. It is necessary to consider the Z-ranks of the formulæ $r \wedge b \wedge f$ and $r \wedge b \wedge \neg f$.

$$\mathrm{Z}(r \wedge b \wedge f) \quad = \quad 0 \quad < \quad 1 \quad = \quad \mathrm{Z}(r \wedge b \wedge \neg f)$$

So the default $r \wedge b \Rightarrow f$ is Z-entailed (as is $\neg r \wedge b \Rightarrow f$). Property inheritance has been handled correctly as system Z has been able to disregard the irrelevant proposition, r.

Secondly, consider whether the default "penguins have wings" is Z-entailed. It is necessary to consider the Z-ranks of the minimal verifying and falsifying models of $p \Rightarrow w$ (m_{12} and m_{11}, respectively):

$$\mathrm{Z}(p \wedge w) \quad = \quad 1 \quad = \quad \mathrm{Z}(p \wedge \neg w)$$

and so $p \Rightarrow w$ is not Z-entailed. Thus, property inheritance fails when the situation at hand is already exceptional. The problem arises because system Z cannot distinguish between the relative abnormality of penguins with or without wings since the relevant default ($b \Rightarrow w$) has the same Z-rank as that already falsified by being a penguin ($b \Rightarrow f$). □

As the example has demonstrated, the minimality of the Z-ranking allows an extension of the P-closure so that some irrelevant information can be discounted—all attributes are assumed to be as normal as they possibly can be, i.e., to have the lowest Z-rank.

However, blindly assuming things to be as normal as possible does not appear to be a reasonable assumption to make. Minimising the ranks of models does not minimise the number of exceptions, just the magnitude of the worse exception. Consider that system Z cannot distinguish between models which falsify just one default of a certain rank or several, and therefore it counts as equal models that, intuitively, ought not to be. Nor does it take into account any lower default violations which might help to distinguish between abnormal situations. Moreover, these are just the distinctions that need to be made if a system is to handle exceptional inheritance correctly. It is exactly when some object is unusual that one wants to assume it to be as normal as possible in *all other respects*. The crude, one dimensional nature of the Z-ranking cannot possibly achieve this. The next section introduces variable strength defaults in an attempt to explicitly capture the intuition that some default violations may be relatively better than others.

3.3 Variable strength defaults

This section looks at an extension to the ε-semantics which enables defaults of different strengths to be represented. Using the ranking function representation, ε-entailment can be expressed in the following way:

Lemma 3.3.1 *Given a set of defaults, Δ, a default, $a \Rightarrow b$, is ε-entailed by Δ iff in all ranking functions, κ, admissible with respect to Δ, $\kappa(a \wedge b) < \kappa(a \wedge \neg b)$.*

This result corresponds to the finding by Lehmann and Magidor that the intersection of all rational consequence relations compatible with a set of defaults is just the

P-closure of that set (Lehmann & Magidor 1992). By adapting the ranking function constraint (3.5), so that there is a minimum degree of separation between a default's minimum verifying and falsifying models, a natural means of representing defaults of differing strengths becomes apparent.

Definition 3.3.2 *A* variable strength default, $a \overset{s}{\Rightarrow} b$, *is a default which has been assigned an extra strength attribute,* s*, which is a positive integer. A set of variable strength defaults will be denoted* Δ^+.

In terms of the probabilistic interpretation of the ε-semantics, each default constrains a probability distribution P by $P(b|a) \geq 1 - \varepsilon$ for some parameter $\varepsilon > 0$. By allowing the constraint associated with each default to differ in the order of ε, defaults can be thought of as having different relative strengths. The constraint associated with a variable strength default becomes $P(b|a) \geq 1 - \varepsilon^s$, where s is the exponent of ε, or the strength assigned to $a \Rightarrow b$. Rearranging the new constraint gives:

$$P(\neg b|a) < \varepsilon^s \quad \equiv \quad P(a \wedge \neg b) < \varepsilon^s P(a)$$

This means that $P(a \wedge \neg b)$ is at least s orders of ε higher than $P(a)$, and, in particular, $P(a \wedge b)$. Abstracting the exponents of ε gives a ranking function constraint of $s + \kappa(a \wedge b) \leq \kappa(a \wedge \neg b)$, and hence a revised definition of satisfaction:

Definition 3.3.3 *A ranking function,* κ*,* satisfies *a variable strength default,* $a \overset{s}{\Rightarrow} b$*, iff*

$$s + \kappa(a \wedge b) \leq \kappa(a \wedge \neg b) \tag{3.7}$$

A ranking function, κ*, will be said to be* ε^+-admissible *with respect to a set of variable strength defaults,* Δ^+*, iff it satisfies all defaults in* Δ^+.

Now, the ε-consistency of a set of standard defaults, Δ, can be equated with the existence of at least one ranking function admissible with respect to Δ. This, too, translates naturally to variable strength defaults:

Definition 3.3.4 *A set of variable strength defaults,* Δ^+*, is* ε^+-consistent *iff there exists a ranking function which is* ε^+*-admissible with respect to* Δ^+.

If Δ is the standard counterpart of Δ^+, i.e., a set which contains the same defaults but without strengths, it turns out that ε^+-consistency of Δ^+ is equivalent to ε-consistency of Δ.

Theorem 3.3.5 *(Goldszmidt & Pearl 1996)* Δ^+ *is* ε^+*-consistent iff* Δ *is* ε*-consistent*[8].

[8]Goldszmidt and Pearl (1996) called this δ-consistency.

Note that, the definition of ε^+-consistency requires only that the difference between the ranks of the minimal verifying and falsifying models of a default be bounded below by its assigned strength. The proof and validity of theorem 3.3.5 crucially depends on the fact that definition 3.3.3 uses an inequality. If the definition were to be tightened to an equality, a more restricted form of probabilistic consistency would result leading to some strength assignments not being satisfiable. In chapter 4, this situation will be examined in more detail.

Consider now the set of all ranking functions which are ε^+-admissible with respect to some set of variable strength defaults, Δ^+, denoted $\mathcal{RF}_{\Delta^+}$. Let $\mathcal{RF}_\Delta$ denote the set of all ranking functions admissible with respect to the standard counterpart of Δ^+. Clearly, since the constraint (3.1) is always satisfied when (3.7) is satisfied, the former set of ranking functions is a subset of the latter, i.e., $\mathcal{RF}_{\Delta^+} \subseteq \mathcal{RF}_\Delta$. Now, an ε-consequence of Δ is one for which (3.1) is satisfied in all members of $\mathcal{RF}_\Delta$; it is natural therefore to define ε^+-consequences in the following way:

Definition 3.3.6 *A default $a \Rightarrow b$ is ε^+-entailed by Δ^+ iff it is satisfied in all ranking functions ε^+-admissible with respect to Δ^+, i.e., for all $\kappa \in \mathcal{RF}_{\Delta^+}$,*

$$\kappa(a \wedge b) < \kappa(a \wedge \neg b)$$

One might suppose that, by restricting the number of admissible ranking functions, there may be some defaults which arise as ε^+-consequences which were not ε-consequences. The following theorem, however, shows that this is not the case.

Theorem 3.3.7 *A default is ε^+-entailed by $\Delta^+ = \{a_i \overset{s_i}{\Rightarrow} b_i\}$, iff it is ε-entailed by Δ.*

Proof. Suppose $a \Rightarrow b$ is ε-entailed by Δ. Then for all $\kappa \in \mathcal{RF}_\Delta$, $\kappa(a \wedge b) < \kappa(a \wedge \neg b)$. But, since $\mathcal{RF}_{\Delta^+} \subseteq \mathcal{RF}_\Delta$, it follows that for all $\kappa \in \mathcal{RF}_{\Delta^+}$, $\kappa(a \wedge b) < \kappa(a \wedge \neg b)$ and hence $a \Rightarrow b$ is ε^+-entailed.

Suppose now that $a \Rightarrow b$ is ε^+-entailed but *not* ε-entailed. This means that a) by theorem 3.1.4, the set $\Delta \cup \{a \Rightarrow \neg b\}$ is ε-consistent; and, b) for all $\kappa \in \mathcal{RF}_{\Delta^+}$, $\kappa(a \wedge b) < \kappa(a \wedge \neg b)$. This case will be proved by showing that it is possible to construct a ranking which is ε^+-admissible with respect to Δ^+ but for which b) cannot hold.

Since $\Delta \cup \{a \Rightarrow \neg b\}$ is ε-consistent it has a Z-partition; let this be $\Delta_0 \cup \Delta_1 \ldots \cup \Delta_n$. Construct a ranking function as follows: for all confirming models of Δ_0 set $\kappa(m) = 0$; let $s_0 = \max_{r_j \in \Delta_0}[s_j]$; for all confirming models of Δ_1 set $\kappa(m) = s_0$; let $s_1 = \max_{r_j \in \Delta_1}[s_j]$; for all confirming models of Δ_2 set $\kappa(m) = s_0 + s_1$; proceed in this way until all models have been assigned a rank. Now for any $a_i \overset{s_i}{\Rightarrow} b_i \in \Delta^+$,

$\kappa(a_i \wedge b_i) + s_i \leq \kappa(a_i \wedge \neg b_i)$, so the constructed κ is ε^+-admissible. But, $\kappa(a \wedge \neg b) < \kappa(a \wedge b)$ which contradicts b) and hence $a \Rightarrow b$ cannot be ε^+-entailed. $\square$.

Theorem 3.3.7 shows that there are no ε^+-consequences which are not also ε-consequences and vice versa. This means that any strength assignment will lead to exactly the same consequences, so what has been gained by assigning strengths to defaults? In fact, what does differ according to different strength assignments is the *degree* to which defaults are ε^+-entailed. This degree, which does not necessarily equate to a default's assigned strength, can be defined as follows:

Definition 3.3.8 *A default $a \Rightarrow b$ is ε^+-entailed by Δ^+ to degree d if for all $\kappa \in \mathcal{RF}_{\Delta^+}$,*

$$\kappa(a \wedge b) + d \leq \kappa(a \wedge \neg b) \tag{3.8}$$

and, for any integer d' which also satisfies (3.8), $d' \leq d$.

That is, d is the minimal degree of separation between the minimum verifying and falsifying models of a default in all ε^+-admissible rankings.

All defaults which have been assigned a strength will be ε^+-entailed to at least this degree by the admissibility of all rankings in $\mathcal{RF}_{\Delta^+}$. If a default in Δ^+ is ε^+-entailed by the others to some degree greater than its assigned strength then it can never attain this strength. This means that, for the strength it has been assigned, this default does not represent a constraint on $\mathcal{RF}_{\Delta^+}$, which in turn means that it is redundant. As yet, no means of determining the degree to which defaults are ε^+-entailed has been found. However, this adaptation of the ε-semantics to handle variable strength defaults is useful as clearly it determines a minimal set of ε^+-entailed defaults and their minimal degrees of entailment in all ε^+-admissible rankings.

3.4 System Z$^+$

System Z$^+$ is an adaptation of system Z which caters for variable strength defaults. In common with system Z, all defaults are assigned ranks which determine a unique ranking over models and the Z$^+$-rank of a model is the Z$^+$-rank of the highest default it falsifies. The Z$^+$-ranking is defined as follows[9]:

Definition 3.4.1 *Let $\Delta^+ = \{r_i | r_i : a_i \overset{s_{r_i}}{\Rightarrow} b\}$ be a ε^+-consistent set of variable strength*

[9]This definition of system Z$^+$ is slightly different from that given in (Goldszmidt 1992, Goldszmidt & Pearl 1996). The Z$^+$-ranks of defaults are lower by 1 but the actual Z$^+$-ranks over models are the same. This modification allows the system to fit more neatly into the variable strength framework.

Z$^+$-ranking algorithm

Input: a ε^+-consistent set of variable strength defaults, $\Delta^+ = \{r_i | r_i : a_i \overset{s_{r_i}}{\Rightarrow} b\}$.

Output: the Z$^+$-ranking.

1. Initialise all $\mathrm{Z}^+(r_i) = \infty$.
2. From all r_i with $\mathrm{Z}^+(r_i) = \infty$, select that r with minimal $s_r + \min_{m \models a \wedge b} [\mathrm{Z}^+(m)]$ using the current values of $\mathrm{Z}^+(r_i)$.
3. Let $\mathrm{Z}^+(r) := s_r + \min_{m \models a \wedge b} [\mathrm{Z}^+(m)]$.
4. If any $\mathrm{Z}^+(r_i) = \infty$ goto step 2.
5. Assign ranks to models using equation (3.9).

Figure 3.3: The Z$^+$-ranking algorithm

defaults. Then:

$$\mathrm{Z}^+(m) = \begin{cases} 0 & \textit{if } m \textit{ falsifies no default in } \Delta^+ \\ \max_{m \models a_i \wedge \neg b_i} [\mathrm{Z}^+(r_i)] & \textit{otherwise} \end{cases} \tag{3.9}$$

where $\mathrm{Z}^+(r_i)$ is a priority ordering on rules, defined by:

$$\mathrm{Z}^+(r_i) = s_{r_i} + \min_{m \models a_i \wedge b_i} [\mathrm{Z}^+(m)] \tag{3.10}$$

Goldszmidt (1992) showed that the Z$^+$-ranking is the unique, minimal ranking satisfying the variable strength constraints (3.7).

Theorem 3.4.2 *(Goldszmidt 1992) Every ε^+-consistent Δ^+ has a unique, minimal ranking given by Z$^+$.*

Z$^+$-entailment is determined, as might be expected, by examining the minimal verifying and falsifying models of a default: if $\mathrm{Z}^+(a \wedge b) < \mathrm{Z}^+(a \wedge \neg b)$, then $a \Rightarrow b$ is Z$^+$-entailed. Because there is a unique Z$^+$-ranking, it is possible to associate a degree of Z$^+$-entailment with defaults, this being the difference in Z$^+$-rank between the minimal verifying and falsifying models of a Z$^+$-entailed default. For example, if $\mathrm{Z}^+(a \wedge \neg b) - \mathrm{Z}^+(a \wedge b) = d$, then $a \Rightarrow b$ is Z$^+$-entailed to degree d.

The Z$^+$-ranking algorithm (see figure 3.3) computes the Z$^+$-ranking. Note that this is a different algorithm from the procedure Z$^+$_order given in (Goldszmidt 1992) which is given because of its comparative simplicity over the original algorithm and because of the slightly modified version of system Z$^+$ presented.

Note that Z$^+$-consequences need not be Z-consequences. Z$^+$-consequences depend on the strength assignment over defaults and, if this assigns all defaults

m	b	f	p	w	Z$^+$
m_1	0	0	0	0	0
m_2	0	0	0	1	0
m_3	0	0	1	0	2
m_4	0	0	1	1	2
m_5	0	1	0	0	0
m_6	0	1	0	1	0
m_7	0	1	1	0	2
m_8	0	1	1	1	2

m	b	f	p	w	Z$^+$
m_9	1	0	0	0	2
m_{10}	1	0	0	1	1
m_{11}	1	0	1	0	2
m_{12}	1	0	1	1	1
m_{13}	1	1	0	0	2
m_{14}	1	1	0	1	0
m_{15}	1	1	1	0	2
m_{16}	1	1	1	1	2

Figure 3.4: The Z$^+$-rankings for the penguin example.

strength 1, the Z$^+$-ranking and the Z-ranking coincide. In this sense, system Z$^+$ subsumes system Z.

The following example shows that Z$^+$-entailment can be used to model different costs of default violations.

Example 3.4.3 (Penguins (cont'd))

$$\Delta = \{b \overset{1}{\Rightarrow} f, b \overset{2}{\Rightarrow} w, p \overset{1}{\Rightarrow} b, p \overset{1}{\Rightarrow} \neg f\}$$

Figure 3.4 enumerates the models along with their Z$^+$-ranks. Consider whether the default "penguins have wings" is Z$^+$-entailed. It is necessary to consider the Z$^+$-ranks of the minimal verifying and falsifying models of $p \Rightarrow w$ (m_{12} and all falsifying models, respectively):

$$\text{Z}^+(p \wedge w) \quad = \quad 1 \quad < \quad 2 \quad = \quad \text{Z}^+(p \wedge \neg w)$$

and so $p \Rightarrow w$ is Z$^+$-entailed to degree 1 □

The example demonstrates that, by increasing the strength of certain defaults, in this case $b \Rightarrow w$, it is possible to obtain the kind of inferences which are desirable according to the requirements outlined in chapter 1. However, as Goldszmidt and Pearl (1996) readily admit, this occurs because the user has deliberately emphasised the strength of one default. In this example, it is necessary to increase the strength of "birds have wings" to allow penguins to inherit the wings attribute. The system itself does not automatically satisfy the requirement for exceptional inheritance: ideally, the wings attribute should be inherited directly from birds to penguins without the user having to specify that it holds more strongly. Although, in some cases, it is useful to be able to represent different strengths explicitly, and in the process obtain different conclusions, exceptional inheritance is a property

that should be satisfied independent of any differing strengths. In this example it seems natural to require that penguins should inherit wings from birds without the need for explicitly making the default stronger. Nevertheless, system Z^+ provides a mechanism for capturing the variability of priorities among defaults, even though it suffers from the same inadequacies of system Z itself. In chapter 6, the similarities, and differences, between system Z^+ and the proposed ME approach will be explored in more detail.

3.5 Lexicographical closure

The lexicographical closure was proposed by Lehmann (1995) who argued that the behaviour of the ideal rational consequence relation should satisfy four presumptions of typicality, independence, priority and specificity. Lehmann argued that the rational closure (Lehmann & Magidor 1992, Pearl 1990) is the "correct formalization" of the prototypical reading of a default in which, once an atypical situation has been identified, it is not possible to make more refined judgements about that situation. For the presumptive reading of a default, as first proposed by Reiter (1980), Lehmann argues that since all defaults are presumed to be active unless there is direct evidence to the contrary, the lexicographical ordering provides a means of resolving conflicts between several defaults which cannot all be active at the same time. Lehmann proposes using the "natural priorities" (Pearl 1990) of defaults given by their Z-ranks as a means of determining which default violations are relatively more serious than others. A more flexible variant on Lehmann's lexicographic closure is given by Benferhat *et al.* (1993) and will be discussed at the end of this section.

The assumptions behind the lexicographical ordering are twofold. Firstly, it is assumed that there is a natural ordering of defaults so that it is always more serious to falsify higher ranked defaults compared with lower ranked ones. Secondly, it is assumed that it should be worse to falsify more defaults at any given level, all other things being equal, so that, in the penguin example for instance, a non-flying bird would be less unusual than a non-flying bird without wings.

Lexicographic (LEX) entailment is defined as follows. The LEX-ordering over the models of $\mathcal{L}$ is based on the Z-partition but takes into account all defaults violated by a model, not just that with the greatest Z-rank. The result is a form of entailment which is a direct extension of system Z in the sense that all Z-entailed defaults are also LEX-entailed.

m	b	f	p	w	LEX	m	b	f	p	w	LEX
m_1	0	0	0	0	(0,0)	m_9	1	0	0	0	(2,0)
m_2	0	0	0	1	(0,0)	m_{10}	1	0	0	1	(1,0)
m_3	0	0	1	0	(0,1)	m_{11}	1	0	1	0	(2,0)
m_4	0	0	1	1	(0,1)	m_{12}	1	0	1	1	(1,0)
m_5	0	1	0	0	(0,0)	m_{13}	1	1	0	0	(1,0)
m_6	0	1	0	1	(0,0)	m_{14}	1	1	0	1	(0,0)
m_7	0	1	1	0	(0,2)	m_{15}	1	1	1	0	(1,1)
m_8	0	1	1	1	(0,2)	m_{16}	1	1	1	1	(0,1)

Figure 3.5: The LEX-tuples for the penguin example.

Given a set of defaults, Δ, and its Z-partition, $\Delta_0 \cup \Delta_1 \ldots \cup \Delta_n$, each model is assigned an $(n+1)$-tuple with the number of defaults it violates in partition-set Δ_i appearing in position i of the tuple. The LEX-ordering of tuples (and hence models) is to consider the last elements of the tuples first. If one tuple has fewer default violations in the highest tuple element, it is lower (or preferred) in the LEX-ordering; otherwise the next highest tuple element is considered. For example, $(1,1,0) \prec (0,0,2)$ and $(2,0,1) \prec (0,1,1)$. From the LEX-ordering, entailment is determined as usual by comparing the LEX-tuples of the minimal verifying and falsifying models of a default.

Example 3.5.1 (Penguins (continued)) Continuing the previous example, figure 3.5 enumerates the models along with their LEX-tuples of default violations. To establish whether the default "penguins have wings" is LEX-entailed, it is necessary to compare the minimal verifying and falsifying models of $p \Rightarrow w$ (m_{12} and m_{11}, respectively):

$$\text{LEX}(p \wedge w) \quad = \quad (1,0) \quad \prec \quad (2,0) \quad = \quad \text{LEX}(p \wedge \neg w)$$

and so $p \Rightarrow w$ is LEX-entailed. □

Being a direct extension of system Z, the lexicographical closure benefits from several nice properties. It is an extension of the P-closure and, being a total ordering over models, corresponds to a rational consequence relation. Since all Z-entailed defaults are also LEX-entailed, it is guaranteed to handle property inheritance, specificity and indifference in the same way that system Z does. Lastly, because it compares default violations by taking into account both their number and their degree, it can handle exceptional inheritance—models which falsify fewer or weaker defaults, according to the LEX-ordering, are more preferred so that more

defaults of a given Z-rank are likely to hold (and more properties to be inherited) while higher ranked defaults are more likely to hold than lower ranked ones. Moreover, it is the system which determines how the defaults are prioritised, the user does not need to specify that any are stronger, nor may he.

While it appears reasonable that all default violations need to be considered when comparing models, the question arises as to whether the "natural ordering" of defaults, i.e., their Z-ranks, should be considered to have any intrinsic significance, and whether *any number* of lower violations should be preferable to a single higher one. Consider again the penguin example and the default "penguins which do not fly and do not have wings are birds" ($p \wedge \neg f \wedge \neg w \Rightarrow b$):

$$\text{LEX}(p \wedge \neg f \wedge \neg w \wedge b) \quad = \quad (2,0) \quad \prec \quad (0,1) \quad = \quad \text{LEX}(p \wedge \neg f \wedge \neg w \wedge \neg b)$$

Clearly, $p \wedge \neg f \wedge \neg w \Rightarrow b$ is LEX-entailed, but why should it be worse to violate the single default $p \Rightarrow b$ than the two defaults $b \Rightarrow f$ and $b \Rightarrow w$? Since the given semantics for defaults is identical, what is the justification for treating one default as higher priority than the others? While one can argue that some defaults may have priority over others, it is hard to justify some as having infinite priority over others so that any number of lower violations are better than a single higher one. There may come a point when it seems more reasonable to abandon accepting the stronger default in favour of accepting a larger number of weaker ones. In terms of the penguin example, it may be more reasonable to reject the belief that a penguin is a bird when it displays no bird attributes rather than insisting that it is a bird which exhibits none of them. Under LEX-entailment, this kind of weighing up of default violations cannot occur, and there is no room for such refinements of judgement, since the priorities are fixed by the Z-partition and the chosen method of ordering LEX-tuples. But the meaning of the Z-partition is that higher ranked defaults can only be verified at the expense of falsifying lower ranked defaults—indeed it is because verifying them incurs the cost of falsifying lower ranked defaults that they have attained a higher rank in the first place. It appears than the LEX-ordering may be penalising some defaults twice and therefore not treating each on its own merits. The intrinsic meaning of the Z-rank of a default is that it represents the minimal order of magnitude for the probability of it being verified in the context of the other defaults, which hardly justifies assigning it a corresponding priority.

In chapter 6, the meaning of this assignment of priorities for LEX-entailment will be examined in more detail. The remainder of this section looks at a more general form of LEX-entailment from a slightly different context.

Benferhat *et al.* separately proposed the lexicographical ordering in a more general setting (Benferhat *et al.* 1993). In their version, knowledge is represented by belief sets which consist of propositional formulæ rather than defaults, and they put no restriction on these, so that they may be inconsistent when taken in conjunction (Lehmann's LEX-entailment is meaningful only when applied to ε-consistent sets of defaults). In fact, Benferhat *et al.* were interested in the syntactic resolution of inconsistency through inducing a preferential ordering over maximally consistent subbases of a belief set, although they do not provide semantics for what a belief base represents. A set of beliefs is partitioned by the user according to his intuitions about the relative priority of the formulæ. The LEX-ordering of subbases prefers to maximise to number of original formulæ contained in a subbase with respect to the user-defined priorities in the same way that default violations are minimised in Lehmann's version. This ordering induces a total pre-ordering of models and a corresponding rational consequence relation. This approach subsumes Lehmann's in the following sense: replace each default by its material implication (i.e., $a \Rightarrow b$ becomes $\neg a \vee b$) and select the user-defined partition according to the Z-partition; subject to these changes, both versions produce the same rational consequence relation. However, since there is no consistency requirement for belief bases, and arbitrary user-defined partitions are permitted, the consequence relations produced by the more general form of LEX-entailment may not be admissible from the default reasoning perspective. That is, defaults in the belief base may not be LEX-entailed by some user-defined partitions, as the example below demonstrates.

Example 3.5.2

$$\Delta = \{\neg b \vee f, \neg b \vee w, \neg p \vee b, \neg p \vee \neg f\}$$

In this example the belief base is flat so that all defaults have the same priority, that is, $\Delta_0 = \Delta$. Consider the minimal verifying and falsifying models for the default $p \Rightarrow f$ in the general LEX-ordering:

$$\text{LEX}(p \wedge f) \quad = \quad (1) \qquad\qquad \text{LEX}(p \wedge \neg f) \quad = \quad (1)$$

so that a default which is part of the belief base is not LEX-entailed for this user-defined partition. □

Since this form of LEX-entailment lacks a clear semantics, it does not fit well with the other systems of default reasoning described in this chapter. Any future references to LEX-entailment therefore refer to Lehmann's version rather than that of Benferhat *et al*.

3.6 Goldszmidt's maximum entropy approach

Pursuing a suggestion originally proposed by Pearl (1988), Goldszmidt applied the principle of maximum entropy to the ε-semantics (Goldszmidt 1992, Goldszmidt, Morris, & Pearl 1993). Because the ε-semantics sanctions conclusions which hold in *all* admissible probability distributions (PDs), the idea is to select that distribution which possesses the highest value of entropy as the most appropriate from which to make inferences. Given a problem in which a probability distribution is constrained to some extent but not uniquely determined, it makes sense to select that PD with the highest entropy since this is guaranteed to contain the most uncertainty, or to be the least biased or committed. In fact, to select any other PD means that additional assumptions have been made which are not justified by the data (Jaynes 1979). The problem, then, becomes one of optimising the entropy function subject to the known constraints, leading to the maximum entropy (ME) distribution. This principle has been widely used across many fields and has been described as "a much needed extension to the established principles of rational inference in the sciences" (Buck & Macaulay 1991).

In the context of default reasoning, and in particular using the ε-semantics, this principle is just what is needed to extend a given set of defaults. As described above, ε-consequences are those which hold in all ranking functions which are admissible with respect to a set of defaults, but these are not detailed enough to fully capture the required default inferences. To obtain these extra inferences, a single ranking function is sought, but how to choose one fairly? By applying the principle of ME, it ought to be possible to obtain a single ranking function—the ME-ranking—which represents the one which is the least committed or biased. This is exactly what Goldszmidt attempted to do.

The entropy function for a probability distribution, P, is given by

$$H[P] = -\sum_{m \in \mathcal{M}} P(m) \log P(m) \tag{3.11}$$

Using the original ε-constraints, $P(b_i|a_i) \geq 1 - \varepsilon$, Goldszmidt looked for the maximum entropy PD for a fixed value of ε and examined what happened as $\varepsilon \to 0$. Making an assumption that all default constraints were active in the ME distribution, he abstracted the exponents of ε and obtained the following equations which the ME-ranking must satisfy[10]:

[10] A full and more general derivation of the ME constraint equations will be given in the following chapter.

minimal-core ME algorithm

Input: a ε-consistent, minimal core set of defaults, $\Delta = \{r_i\}$.

Output: the ME-ranking.

[1] Let Γ be the rules tolerated by Δ.

[2] For each rule $r_i \in \Gamma$, set $\text{ME}(r_i) = 1$.

[3] While $\Gamma \neq \Delta$ do:

(a) Let Ω be the set of models, m, such that m falsifies rules only in Γ and verifies at least one rule in $\Delta - \Gamma$; let Γ_m denote the set of rules in Γ falsified by m.

(b) For each $m \in \Omega$ compute $\kappa(m) = \sum_{r_i \in \Gamma_m} \text{ME}(r_i)$.

(c) Let m^* be the model in Ω with minimum κ. For each rule $r_i \notin \Gamma$ that m^* verifies, compute the following:

$$\text{ME}(r_i) = 1 + \kappa(m^*)$$

and set $\Gamma = \Gamma \cup \{r_i\}$.

Figure 3.6: The minimal-core ME algorithm

$$\text{ME}(r_i) = 1 + \min_{m \models a_i \wedge b_i} [\text{ME}(m)] \tag{3.12}$$

$$\text{ME}(m) = \sum_{\substack{r_i \\ m \models a_i \wedge \neg b_i}} \text{ME}(r_i) \tag{3.13}$$

Where the $\text{ME}(m)$ are the ME-ranks associated with the models and the $\text{ME}(r_i)$ are the ME-ranks associated with the defaults. In order to solve these equations, Goldszmidt had to restrict himself to a class of default sets, called *minimal core sets*, for which his assumption is valid.

Definition 3.6.1 *A set of defaults, Δ, is* minimal core *iff for all defaults $a_i \Rightarrow b_i \in \Delta$, their converse, $a_i \Rightarrow \neg b_i$, is tolerated by $\Delta - \{a_i \Rightarrow b_i\}$. Equivalently, for all defaults $a_i \Rightarrow b_i \in \Delta$ there exists a model which falsifies $a_i \Rightarrow b_i$ and no other default.*

The reason for this restriction is twofold. Firstly, the technique used to derive the ME constraint equations requires that all constraints be active, i.e., satisfied as equalities, in the ME-distribution, and this is guaranteed when the set is minimal core. Secondly, it meant that the minimal falsifying model of each default falsifies only that default allowing its ME-rank to be determined directly from (3.12). Goldszmidt's algorithm, which is valid only for minimal core sets, is given in figure 3.6.

m	b	f	p	w	ME	m	b	f	p	w	ME
m_1	0	0	0	0	0	m_9	1	0	0	0	2
m_2	0	0	0	1	0	m_{10}	1	0	0	1	1
m_3	0	0	1	0	2	m_{11}	1	0	1	0	2
m_4	0	0	1	1	2	m_{12}	1	0	1	1	1
m_5	0	1	0	0	0	m_{13}	1	1	0	0	1
m_6	0	1	0	1	0	m_{14}	1	1	0	1	0
m_7	0	1	1	0	4	m_{15}	1	1	1	0	3
m_8	0	1	1	1	4	m_{16}	1	1	1	1	2

Figure 3.7: The ME-ranks for the penguin example.

The following example computes the ME-ranking for the penguin example.

Example 3.6.2 (Penguins (continued)) The constraint equations (3.12) and (3.13) give an ME-ranking for the defaults of $\text{ME}(r_1) = 1$, $\text{ME}(r_2) = 2$, $\text{ME}(r_3) = 2$, and $\text{ME}(r_4) = 1$. The ME-rankings are given in figure 3.7.

To establish whether the default "penguins have wings" is ME-entailed, again compare the minimal verifying and falsifying models of $p \Rightarrow w$ (m_{12} and m_{11}):

$$\text{ME}(p \wedge w) \quad = \quad 1 \quad < \quad 2 \quad = \quad \text{ME}(p \wedge \neg w)$$

and so $p \Rightarrow w$ is ME-entailed. □

Goldszmidt discusses the possibility of extending his minimal-core ME algorithm to cater for variable strength defaults, and he gives some pointers to non-minimal core sets for which his algorithm is valid if applied in stages (Goldszmidt 1992). However, he did not analyse the implications of these suggestions in any detail. The ME approach as defined by Goldzmidt has a unique solution since the original constraints subsume any tighter rates of convergence, e.g., if $P(\neg b|a) \geq 1 - \varepsilon^2$ then it clearly satisfies $P(\neg b|a) \geq 1 - \varepsilon$ for all $\varepsilon < 1$. The ramifications of allowing defaults to take different strengths need to be examined more closely and the definition of the ME approach needs to be revised. The problem of redundant default information was also not fully explored. While Goldszmidt acknowledges that all defaults in a set ought to be relevant to the resultant consequence relation, he failed to recognise the implication that the same default set may have several valid interpretations, i.e., ME-rankings, depending on exactly which defaults are taken as being redundant. In clarifying the original assumptions of Goldszmidt's ME approach (Goldszmidt 1992,

Goldszmidt, Morris, & Pearl 1993) and by examining the problem in a more flexible way, this thesis aims to provide a fuller characterisation of applying ME to default reasoning and to extend it to arbitrary sets of variable strength defaults.

3.7 Discussion

All the systems described in this chapter have the ε-semantics at their core; indeed, it has been described as the core behaviour which all nonmonotonic reasoning systems should exhibit (Geffner 1992, Pearl 1990). But none of these extensions have captured all of the requirements for default reasoning and a complete theory of default inference is still needed. The remainder of this thesis attempts to provide such a theory and to justify its position as the definitive method of obtaining default conclusions from a given set of defaults.

The difficulty of obtaining a sound characterisation of default reasoning lies in the inappropriateness of designing systems specifically to reproduce certain accepted default behaviours. This has led to a chicken-and-egg situation which cannot be remedied simply by designing systems to exhibit some, or even all, of the required behaviours. For example, system Z adopts a seemingly sensible policy of minimising each model's rank; this solves only the problem of ignoring irrelevant information but cannot hope to differentiate between default interactions involving several exceptions. Even more complex and successful systems, such as LEX-entailment, cannot be justified simply because the requirements of default reasoning are met since they do not provide any semantic interpretations for the behaviour obtained. A less arbitrary solution to the problem of what default reasoning means is required and therefore a more objective approach needs to be taken so that default inferences which meet the requirements can be justified.

Fortunately, because of the underlying probabilistic semantics, a sound and consistent method of inference exists that allows one to choose between admissible probability distributions in an absolutely fair way—maximising entropy (Shore & Johnson 1980). Given the success of the ε-semantics as a seemingly sound but incomplete method of default reasoning, its extension using ME-inference ought to enable further default conclusions to be obtained. Since ME-inference is the least committed way of extending a set of data, these conclusions can be the only extra ones obtainable without making further arbitrary assumptions. Any other conclusions would imply the use of additional information not strictly contained in the problem, i.e., external to the actual defaults. Indeed, ME-inference is the only globally consistent means of extension (Paris & Vencovská 1990).

Adopting the philosophy that using ME is the only justifiable way to extend ε-consequences, the following chapter examines how the work of Goldszmidt *et al.* (Goldszmidt 1992, Goldszmidt, Morris, & Pearl 1993) can be generalised and extended so as to provide the ultimate conclusions from arbitrary sets of defaults. In the chapters which follow, it will be shown that ME-consequences do indeed satisfy all the requirements for default reasoning, and the ME approach will be compared with other extensions to the ε-semantics to reveal their particular biases and further promote its own acceptability.

Chapter 4

Maximum entropy and variable strength defaults

This chapter presents a generalisation of the ME approach to default reasoning first proposed by Goldszmidt (Goldszmidt 1992, Goldszmidt, Morris, & Pearl 1993). The original definition of ME-entailment required that all defaults were constrained identically, leading to a unique ME-solution which could be found in a restricted class of cases. The work presented here takes a slightly different view, allowing the way that defaults are constrained to vary as a means of representing differences in their strengths or priorities. This approach implies that different ME-solutions will arise corresponding to different sets of constraints and hence allows a much greater level of expressiveness for default knowledge bases and the consequence relations they induce. The chapter is organised as follows. Firstly, the differences between the original ME approach and this generalisation are clarified. Using revised assumptions, the equations which determine the ME-ranking are derived. A case study is examined to illustrate the types of solution that occur for the generalised problem. An algorithm for finding a solution to the ME-ranking equations for arbitrary sets of variable strength defaults is then presented. A sufficient condition to guarantee the uniqueness of this solution is given. Finally, connecting the solutions found with ME-rankings is discussed; it is found that the solutions represent unique ME-rankings when no redundancy is present. Earlier versions of this work previously appeared in (Bourne & Parsons 1999a) and (Bourne & Parsons 1999c). The chapter concludes with a discussion about how this more widely applicable version of the ME approach can be used, and which theoretical problems of the formalism remain to be solved.

4.1 Review of original ME assumptions

As described in section 3.6, Goldszmidt extended the ε-semantics by applying the principle of maximum entropy (Goldszmidt 1992, Goldszmidt, Morris, & Pearl 1993). Under the assumption that all defaults satisfied a constraint of the form:

$$P(b_i|a_i) \geq 1 - \varepsilon \tag{4.1}$$

he applied the Lagrange multiplier technique to find the ME-distribution. By restricting the analysis to minimal core sets, a set of equations for the ME-ranking was derived and an algorithm for computing it was given. He noted that:

> ...since the constraints $P(\psi_i|\phi_i) \geq 1 - \varepsilon$ define a convex region, this maximum entropy distribution is unique. (Goldszmidt, Morris, & Pearl 1993)

which means that, under this interpretation of defaults as all satisfying the same constraint, each set of defaults has a unique ME-solution. Goldszmidt's approach is unsatisfactory for several reasons.

1. The ultimate aim of the ME approach is to find a ranking function abstraction of a set of probability distributions that best represents a set of defaults. But this means that the exact function of ε used to constrain a default in (4.1) is unnecessarily precise. It is the order of magnitude of ε in this constraint which appears in the equations which determine the ME-ranking. Therefore, it would be more appropriate for these constraints to be of the form:

   $$P(b_i|a_i) \geq 1 - O(\varepsilon)$$

   where $O(\varepsilon)$ is some unspecified linear function of ε.

2. One of the requirements for default reasoning is that defaults should be able to take on different strengths or priorities which implies that the same set of defaults may have different interpretations, or consequence relations, according to the strength assignment over its elements. The above formulation, which implies a unique solution for every set of defaults, is incapable of exhibiting such behaviour. However, using the semantics of variable strength defaults, it is possible to represent their relative strengths by requiring them to satisfy variable constraints (see sections 3.3 and 3.4).

3. The Lagrange multiplier technique assumes that all constraints are satisfied as equalities. However, this may not always be possible. For example, one

default constraint may strictly imply another so that they cannot both be satisfied as equalities—one will be a strict inequality—in which case it would appear that the second default is redundant. However, so long as the constraints are probabilistically consistent, an ME-solution exists though Goldszmidt's algorithm will not be capable of finding it.

These considerations motivate a reformulation of the ME approach so that the constraints associated with the defaults are more flexible. In fact, the constraints will be specified only up to their relative orders of magnitude compared with the other defaults. Obviously, by changing the constraints used, the ME-distribution will not necessarily be the same as that obtained from Goldszmidt's approach. However, this new approach is more expressive since it allows interactions between defaults to be affected by changes in their priorities. The result is a system for which the ME-ranking varies according to the different strengths assigned to defaults. Thus, in this revised formulation, an ME-ranking exists only with respect to some strength assignment. This new approach subsumes the original version when dealing with minimal core sets of defaults of equal strength.

In order to associate hard constraints with each default, the exact nature of the functions of ε which are associated with each default will be left imprecise. In fact, the user will only be required to specify some *rate of convergence* for a default, or the relative order of magnitude with which its associated conditional probability tends to 1, compared with the other defaults. Each variable strength default, $a_i \overset{s_i}{\Rightarrow} b_i$, is required to satisfy an *asymptotic constraint*:

$$P(b_i|a_i) = 1 - O_i(\varepsilon^{s_i}) \tag{4.2}$$

where $O_i(\varepsilon^{s_i})$ is some unspecified function of ε. For each default with strength s_i, the function $O_i(\varepsilon^{s_i})$ satisfies:

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

where C_i is a positive constant called the *convergence coefficient* for default r_i. Since the objective of the ME approach is to find a ranking function over models, the convergence coefficient is not strictly relevant although it will turn out that other constraints exist between the C_i which determine the type of solution obtained using this revised ME approach (see section 4.3).

There are several advantages of these changes to the ME approach. Firstly, the use of asymptotic constraints over defaults fits more neatly into the ranking function abstraction of the ε-semantics. Consider the following derivation:

$$
\begin{array}{rrcl}
 & P(b_i|a_i) & = & 1 - O_i(\varepsilon^{s_i}) \\
\Rightarrow & P(\neg b_i|a_i) & = & O_i(\varepsilon^{s_i}) \\
\Rightarrow & \kappa(\neg b_i|a_i) & = & s_i \\
\Rightarrow & \kappa(a_i \wedge \neg b_i) - \kappa(a_i) & = & s_i \\
\Rightarrow & \kappa(a_i \wedge \neg b_i) & = & s_i + \kappa(a_i \wedge b_i)
\end{array}
$$

Note that only the exponents of ε are relevant for ranking functions and so it makes sense to allow the convergence coefficients to remain unspecified. The final equation is similar to the constraint used for variable strength defaults (see section 3.3), except that for the ME-ranking, the constraint requires s to be the exact difference between the ranks of a default's minimal verifying and falsifying models, rather than just a lower bound.

Another advantage is that both the user's inputs and the system's outputs will now be of the same type. That is, the user inputs a set of defaults plus their strengths, and receives as output an ME-consequence relation which can determine whether arbitrary defaults are ME-entailed and, if so, to what degree.

Of course, the main advantage is that it will be possible to express default information in a far more detailed way since the strengths of defaults can be adjusted. This change will lead both to different ME-consequences for different strength assignments and, where a particular ME-consequence holds for any strength assignment, for example ε-consequences, its degree of ME-entailment will also vary. This should enable more accurate modelling of situations using default knowledge bases.

In the next section a set of equations are derived which determine the ME-ranking. The main problem with the approach is in interpreting the solutions obtained from these equations. It is well known that the ME-distribution is unique (subject to a particular set of constraints) since the entropy function is convex, but the equations occasionally lead to multiple solutions. Similarly, sometimes the equations are insoluble despite the fact that an ME-solution must exist if the constraints are not probabilistically inconsistent. The reasons for both these cases are discussed in section 4.3, and interpreted in section 4.6.

4.2 Deriving the maximum entropy ranking

Throughout this derivation, it is assumed that all variables can be expressed as analytic functions of ε. That is, all variables have the form $O(\varepsilon^s)$, so that as $\varepsilon \to 0$ each variable asymptotically approaches some function of ε of some order s. For any ε-consistent set of defaults, and for some selection of real convergence functions,

the ME-distribution clearly exists; under this analytic assumption, the ME-ranking represents an asymptotic abstraction of the exponents of the probabilities in the ME-distribution. A crucial question is, in what circumstances does fixing only the orders of magnitude of the convergence functions lead to a unique order of magnitude description of the ME-distribution? In some cases, notably those involving redundancy, it will be seen that multiple ME-rankings exist for some strength assignments; it is unclear whether unique solutions arise because the ranked abstraction of the ME-distribution is independent of the convergence coefficients for a specific strength assignment. Nevertheless, the derivation depends critically on this assumption which is discussed further at the end of this chapter.

The basic idea is to find the ME-distribution for a fixed ε, and to consider what happens to it as $\varepsilon \to 0$. Under the analytic assumption discussed above, this means that all the equations which determine the ME-distribution can be abstracted into integer equations which determine the ranks, or exponents of ε, for the variables in the ME-ranking.

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

$$H[P] = -\sum_{m \in \mathcal{M}} P(m) \log P(m) \tag{4.3}$$

Note that, like Goldszmidt, it is assumed that all constraints are active[1]. As discussed in the previous section, each default, r_i, is supposed to satisfy an asymptotic constraint of the form:

$$P(b_i|a_i) = 1 - O_i(\varepsilon^{s_i}) \tag{4.4}$$

Where the strengths, s_i, are specified for each default but the convergence coefficients of the functions, $O_i(\varepsilon^{s_i})$, are left unspecified. 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 (4.4) imposed on P for each default can be rewritten:

$$\sum_{m \models a_i \wedge \neg b_i} P(m) \quad - \quad \frac{O_i(\varepsilon^{s_i})}{1 - O_i(\varepsilon^{s_i})} \sum_{m \models a_i \wedge b_i} P(m) \quad = \quad 0 \tag{4.5}$$

[1]This assumption will be relaxed in section 4.6 where it will be seen that for ME-redundant defaults, their constraints are satisfied as strict inequalities and their Lagrange multipliers are simply zero.

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

$$H'[P] = -\sum_{m \in \mathcal{M}} P(m) \log P(m) + \sum_{r_i} \lambda_i \left[P(a_i \wedge \neg b_i) - \frac{O_i(\varepsilon^{s_i})}{1 - O_i(\varepsilon^{s_i})} P(a_i \wedge b_i) \right] \tag{4.6}$$

The function H' ranges over all probability distributions for which the constraints are satisfied as equalities and hence the additional sumands are effectively zero and $H' \equiv H$. 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, which gives $|\mathcal{M}|$ simultaneous equations of the form:

$$\frac{\partial H'[P]}{\partial P(m)} = -1 - \log P(m) + \sum_{\substack{r_i \\ m \models a_i \wedge \neg b_i}} \lambda_i - \sum_{\substack{r_i \\ m \models a_i \wedge b_i}} \frac{O_i(\varepsilon^{s_i})}{1 - O_i(\varepsilon^{s_i})} \lambda_i = 0 \tag{4.7}$$

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, which requires it to sum to one. Since this will merely be represented by some normalisation factor, common to each model's probability, it can be safely ignored so that the distribution found will in fact represent the unnormalised ME-distribution.

Introducing the substitution $\alpha_i = e^{\lambda_i}$, and taking antilogs of (4.7), yields expressions for the probabilities of each model in terms of the α_i and the $O_i(\varepsilon^{s_i})$:

$$P(m) = e^{-1} \prod_{\substack{r_i \\ m \models a_i \wedge \neg b_i}} \alpha_i \prod_{\substack{r_i \\ m \models a_i \wedge b_i}} \alpha_i^{-\frac{O_i(\varepsilon^{s_i})}{1 - O_i(\varepsilon^{s_i})}} \tag{4.8}$$

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

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

Where $\kappa(r_i)$ and $\kappa(m)$ represent the integer ME-ranks of the defaults and of the probability of each model, respectively. Note that, under these assumptions, the constant factor e^{-1} and the second product in (4.8) both represent functions of order zero[2], and can therefore be replaced by a function c_m which will tend to a

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

constant as $\varepsilon \to 0$. The expression for the probability of each model (4.8) therefore reduces to:

$$O_m(\varepsilon^{\kappa(m)}) = c_m \prod_{\substack{r_i \\ \overline{m \models a_i \wedge \neg b_i}}} O_{r_i}(\varepsilon^{\kappa(r_i)}) \tag{4.9}$$

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

$$\kappa(m) = \sum_{\substack{r_i \\ \overline{m \models a_i \wedge \neg b_i}}} \kappa(r_i) \tag{4.10}$$

Under these same assumptions and substitutions, the constraint equations (4.5), reduce to the integer equations:

$$\min_{m \models a_i \wedge \neg b_i} [\kappa(m)] = s_i + \min_{m \models a_i \wedge b_i} [\kappa(m)] \tag{4.11}$$

These two sets of equations (4.10) and (4.11) thus define the ranking function abstraction to the ME-distribution, under the given assumptions. It is a contention of this thesis that, if all defaults are active in the ME-distribution, then the solutions to equations (4.10) and (4.11) represent an asymptotic abstraction of the ME-distribution (the unique ME-ranking) *regardless of the coefficients of the convergence functions* $O_i(\varepsilon^{s_i})$. This is the purpose of making the analytic assumption at the beginning of this section; however, verifying or refuting the validity of this assumption is beyond the scope of this thesis, and discussion of it is therefore deferred till the end of this chapter. As will be shown in section 4.3, under some circumstances there are no solutions to (4.10) and (4.11), while in others there are multiple solutions; section 4.4 gives an algorithm which computes either an exact solution to these equations or an ε^+-admissible ranking for arbitrary sets of variable strength defaults; section 4.5 identifies a sufficient condition for this solution to be unique; and, section 4.6 discusses when such a solution can be interpreted as the unique ME-ranking. The following simple example illustrates an ME-solution.

Example 4.2.1

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

Of the 4 models of $\mathcal{M}$, only one falsifies the default; their ME-ranks are given by:

$$\kappa(a \wedge b) = 0 \qquad \kappa(a \wedge \neg b) = \kappa(r)$$
$$\kappa(\neg a \wedge b) = 0 \qquad \kappa(\neg a \wedge \neg b) = 0$$

There is just one constraint, $\kappa(a \wedge \neg b) = s + \kappa(a \wedge b)$, which implies that $\kappa(r) = s$. This ME-ranking is clearly unique. The default $\neg b \Rightarrow \neg a$ is ME-entailed by Δ^+ to degree s. □

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

Table 4.1: Table of unnormalised probabilities for the ME-distribution.

4.3 Case study

This section looks at the solution to the ME problem for one particular example. The example demonstrates that for some choices of strength assignment, there may be no convergence coefficients which satisfy the constraint equations. In other cases, and for a fixed strength assignment, different choices of convergence coefficients lead to different solutions for the ME-ranking. Finally, in some cases, there is a unique solution for the ME-ranking regardless of the convergence coefficients. The example illustrates that the solution depends firstly on the strength assignment, and secondly, for some strength assignments, on the actual convergence functions from which they are abstracted.

Example 4.3.1

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

Using equation (4.8) but substituting the factor c_m for each model's function of order zero, table 4.1 shows whether a model falsifies or verifies each default and gives its (unnormalised) probability in the ME-distribution.

Now, let the convergence function for the defaults be $O_1(\varepsilon^{s_1})$, $O_2(\varepsilon^{s_2})$ and $O_3(\varepsilon^{s_3})$, respectively[3]. The constraint equations (4.5) give rise to three simultaneous equations:

$$\begin{aligned} c_{m_5}\alpha_1\alpha_2 + c_{m_6}\alpha_1 &= O_1(\varepsilon^{s_1})(c_{m_7}\alpha_2\alpha_3 + c_{m_8}) \\ c_{m_5}\alpha_1\alpha_2 + c_{m_7}\alpha_2\alpha_3 &= O_2(\varepsilon^{s_2})(c_{m_6}\alpha_1 + c_{m_8}) \\ c_{m_7}\alpha_2\alpha_3 &= O_3(\varepsilon^{s_3})(c_{m_8}) \end{aligned}$$

[3]Strictly speaking, the convergence functions are $O_1'(\varepsilon^{s_1}) = \frac{O_1(\varepsilon^{s_1})}{1+O_1(\varepsilon^{s_1})}$, but since this substitution would not affect their order of magnitude, it is ignored.

Solving these for the α_i gives:

$$\alpha_1 = \frac{c_{m_8}[O_1(\varepsilon^{s_1})(O_3(\varepsilon^{s_3})+1) - O_2(\varepsilon^{s_2}) + O_3(\varepsilon^{s_3})]}{c_{m_6}(1+O_2(\varepsilon^{s_2}))} \tag{4.12}$$

$$\alpha_2 = \frac{c_{m_6}[O_1(\varepsilon^{s_1})O_2(\varepsilon^{s_2})(O_3(\varepsilon^{s_3})+1) + O_2(\varepsilon^{s_2}) - O_3(\varepsilon^{s_3})]}{c_{m_5}[O_1(\varepsilon^{s_1})(O_3(\varepsilon^{s_3})+1) - O_2(\varepsilon^{s_2}) + O_3(\varepsilon^{s_3})]} \tag{4.13}$$

$$\alpha_3 = \frac{c_{m_5}c_{m_8}O_3(\varepsilon^{s_3})[O_1(\varepsilon^{s_1})(O_3(\varepsilon^{s_3})+1) - O_2(\varepsilon^{s_2}) + O_3(\varepsilon^{s_3})]}{c_{m_6}c_{m_7}[O_1(\varepsilon^{s_1})O_2(\varepsilon^{s_2})(O_3(\varepsilon^{s_3})+1) + O_2(\varepsilon^{s_2}) - O_3(\varepsilon^{s_3})]} \tag{4.14}$$

Now, if each α_i can be represented by some function of ε of unknown order (its ME-rank), the fractions of the right-hand side of equations (4.12) to (4.14) must equate to functions of the same order. Comparing exponents on both sides of these equations will result in the ME constraint equations for this example.

However, by looking at the structure of these solutions, constraints on the strength assignments themselves can be established. For example, since $\alpha_i = e^{\lambda_i}$, it must always be a positive quantity. From the numerator of (4.13), evidently it is necessary that $s_2 \leq s_3$, and, if this holds as a strict inequality, from the denominator, it is necessary that $s_1 \leq s_2$. Any violation of these constraints on the strength assignment will lead to no solution to the equations. Since there must be a solution to the ME problem, this case occurs when the assumption that all constraints are active in the ME-distribution is false (see section 4.6).

Consider the case when $s_1 > s_2 > s_3$. Now all these extra constraints on the strength assignments are satisfied and there will be a unique solution for the ME-ranks (of defaults) given by:

$$\begin{aligned}
\kappa(r_1) &= s_1 \\
\kappa(r_2) &= s_2 - s_1 \\
\kappa(r_3) &= s_3 + s_1 - s_2
\end{aligned}$$

A more interesting situation occurs when some defaults are assigned equal strengths. In this case the actual functions O_i, and not just their orders of magnitude, become relevant. For example, suppose the functions $O_2 = O_3$, so that in the equations above $O_2(\varepsilon^{s_2}) - O_3(\varepsilon^{s_3}) = 0$. In this case the solution for the ME-ranks is given by:

$$\begin{aligned}
\kappa(r_1) &= s_1 \\
\kappa(r_2) &= s_2 \\
\kappa(r_3) &= 0
\end{aligned}$$

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

Table 4.2: Table of different ME-rankings for a single strength assignment.

Suppose, instead, that $O_2 = O_1 + O_3$ (all of order s). In this case the solution for the ME-ranks is given by:

$$\begin{aligned}\kappa(r_1) &= s_1 + s_3 = 2s \\ \kappa(r_2) &= -s_3 = -s \\ \kappa(r_3) &= 2s_3 = 2s\end{aligned}$$

Finally, suppose that $O_1 = 2(O_2 - O_3)$ (all of order s). In this case the solution for the ME-ranks is given by:

$$\begin{aligned}\kappa(r_1) &= s_1 = s \\ \kappa(r_2) &= 0 \\ \kappa(r_3) &= s_3 = s\end{aligned}$$

These three solutions for the ME-ranking are given in table 4.2 where all defaults are assigned an equal strength of $s = 1$. As can be seen, each solution corresponds to a slightly different ME-ranking over the models of $\mathcal{M}$, though each satisfies the ME constraint equations.

This case study illustrates how the solutions for the ME-ranking behave in various cases. It should be noted that the set chosen for this example contains some redundancy. In fact, the defaults represent the system P rule of cautious monotonicity. This means that the set $\{a \Rightarrow b, a \Rightarrow c\}$ ε-entails $a \wedge b \Rightarrow c$, and also the set $\{a \Rightarrow b, a \wedge b \Rightarrow c\}$ ε-entails $a \Rightarrow c$. In fact, as will become evident from the behaviour of the algorithm to be described shortly, it is redundancy of defaults, or their assigned strengths, which causes these troublesome cases to occur. What is meant by redundancy is that a default is already ME-entailed to some degree by the other defaults in the set. If such a default is assigned a strength lower than its

degree of ME-entailment, the assumption used to derive the constraint equations is invalid, while if it is assigned the strength by which it is ME-entailed, the meaning of the default sets is ambiguous so that multiple ME-rankings arise. The issue of redundancy is discussed in more detail in section 4.6.

4.4 The ME algorithm

In this section, the ME algorithm is presented. The algorithm searches for an exact solution to equations (4.10) and (4.11), that is, a set of integer ranks over defaults, $\kappa(r_i)$ that define an integer ranking over models, $\kappa(m)$, which satisfies (4.11) exactly for all defaults. If no such exact solution is found, the algorithm finds a set of ranks which lead to a ranking which is ε^+-admissible with respect to the defaults plus their strength assignment, that is, equations (4.11) hold only as strict inequalities. Sometimes there are multiple exact solutions, in which case the algorithm computes an arbitrary one. Following the algorithm are proofs demonstrating these claims.

Now, if an exact solution to (4.10) and (4.11) exists, it can be found by solving equations (4.10) and (4.11) simultaneously. However, these equations are nonlinear, so 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 solution is sought.

The algorithm works by finding suitable ranks of defaults one by one. Expanding equations (4.10) and (4.11) illustrates how this may be accomplished. Let v_r (respectively, f_r) represent a minimal verifying (respectively, falsifying) model of r in some ranking κ. Then, if an exact solution exists, it will satisfy:

$$\kappa(f_r) = s_r + \kappa(v_r) \tag{4.15}$$

for some suitable assignment of ranks to the defaults, $\kappa(r_i)$. Further, if such an exact solution exists, the rank of each falsifying model of a default will contain a contribution from its own rank. Rewriting equation (4.15) makes this explicit:

$$\kappa(r) + (\kappa(f_r) - \kappa(r)) = s_r + \kappa(v_r) \tag{4.16}$$

Now, if the assignment of ranks to defaults with lesser ranked minimal falsifying models were already known, equation (4.16) could be used to determine the value of $\kappa(r)$. Expanding (4.16) gives:

$$\kappa(r_i) + \min_{m \models a_i \wedge \neg b_i} \left[\sum_{\substack{r_j, j \neq i \\ \overline{m \models a_j \wedge \neg b_j}}} \kappa(r_j) \right] = s_i + \min_{m \models a_i \wedge b_i} \left[\sum_{\substack{r_j, j \neq i \\ \overline{m \models a_j \wedge \neg b_j}}} \kappa(r_j) \right] \tag{4.17}$$

ME algorithm

Input: a set of variable strength defaults, $\{r_i : a_i \overset{s_i}{\Rightarrow} b_i\}$.

Output: an integer ranking, κ, over defaults and models.

[1] Initialise all $\kappa(r_i) = \mathtt{INF}$.

[2] While any $\kappa(r_i) = \mathtt{INF}$ do:

(a) For all r_i with $\kappa(r_i) = \mathtt{INF}$, compute $s_i + \mathtt{MINV}(r_i)$.

(b) For all such r_i with minimal $s_i + \mathtt{MINV}(r_i)$, compute $\mathtt{MINF}(r_i)$.

(c) Select r_j with minimal $\mathtt{MINF}(r_i)$.

(d) If $\mathtt{MINF}(r_j) = \mathtt{INF}$ let $\kappa(r_j) := 0$
else let $\kappa(r_j) := s_j + \mathtt{MINV}(r_j) - \mathtt{MINF}(r_j)$.

[3] Assign ranks to models using equation (4.10).

[4] Check whether equations (4.11) are satisfied as equalities or inequalities.

Figure 4.1: The ME algorithm

The algorithm proceeds as follows. Initially, all defaults are assigned an infinite rank. By defining two functions MINV(r) and MINF(r) which compute, respectively, the current minimal rank of all verifying models of r, and the current minimal rank of all falsifying models of r *excluding its own contribution*, it becomes possible to compute an appropriate rank for each default via the assignment:

$$\kappa(r) := s_r + \mathrm{MINV}(r) - \mathrm{MINF}(r) \tag{4.18}$$

The ME algorithm is given in figure 4.1. The remainder of this section sets out to demonstrate the claim that this algorithm computes an exact solution to equations (4.10) and (4.11), or at least an ε^+-admissible ranking provided the set is ε-consistent. The first lemma shows that the ME algorithm always assigns each default some finite rank.

Lemma 4.4.1 *Given an ε-consistent set of variable strength defaults, the* ME *algorithm assigns a finite rank to each default.*

Proof. Provided the minimal computed value for the function 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 MINF(r) is infinite; and s_r + MINV(r) − MINF(r), otherwise. Suppose therefore that at some pass of the loop the minimal computed value for MINV(r) is infinite for all unranked r. This means that all verifying models of each unranked default also falsify an unranked default, i.e.,

the set of defaults remaining to be ranked is unconfirmable. 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 ranks will be produced, which in turn implies a finite set of ranks over models. The next lemma shows that this 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 4.4.2 *Given an ε-consistent set of variable strength defaults, the* ME *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, it is confirmable and 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 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, MINF(r) is finite then the default is assigned a rank of s_r + MINV(r) − MINF(r). Now any falsifying models of r which only falsify other previously ranked defaults will all have a rank of greater than or equal to s_r + MINV(r) because MINF(r) was minimal among them; by the inductive hypothesis, this is non-negative. Any other falsifying models of r will still have infinite rank. The lemma follows by induction. □

This lemma does not preclude defaults from having negative ranks, only models. 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 4.4.3 *Given an ε-consistent set of variable strength defaults, the* ME *algorithm assigns ranks to defaults in ascending order of the final ranks of their minimal verifying models plus their strengths.*

Proof. The proof of lemma 4.4.2 shows that, at any stage, if a model's rank becomes finite it will be greater than or equal to that of the current default's computed s_r + MINV(r). Since, at each pass of the loop, r is chosen so that this is minimal, it also follows that no model which has infinite rank can subsequently obtain a lower final rank than the current s_r + MINV(r). This implies both that

$\kappa(v_r) = \text{MINV}(r)$ for the current r, and that the defaults are ranked in ascending order of their final $s_r + \kappa(v_r)$. □

Corollary 4.4.4 *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 ME algorithm is ε^+-admissible. Because the ranks of the models are computed from the ranks of the defaults, the equations (4.10) are guaranteed to be satisfied. although the same cannot necessarily be said for equations (4.11). However, the following lemma shows that, if for some default, equation (4.11) 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 4.6, 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 4.4.5 *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., $\kappa(r) = 0$).*

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 $\text{MINF}(r)$ was finite; if it were then the assignment $\kappa(r_j) := s_j + \text{MINV}(r_j) - \text{MINF}(r_j)$ 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 $\text{MINF}(r)$ was infinite, r was assigned rank zero. □

Whether or not the failure of the ME algorithm to find an exact solution implies that no such solution exists is an open question, although no such situation has yet been found. However, there certainly are cases for which equations (4.10) and (4.11) have no solution, in which case finding an ε^+-admissible ranking is all that can be expected. In the following section, a condition which identifies rankings which are the unique exact solutions to (4.10) and (4.11) is given, and section 4.6 inteprets these results in terms of ME-rankings.

4.4.1 Complexity

The main disadvantage of this new algorithm, which it shares with that of (Goldszmidt, Morris, & Pearl 1993), is its intractability. Unfortunately, the algorithm

requires that all models of $\mathcal{L}$ be ranked repeatedly, so, if $\mathcal{L}$ has n propositions, the algorithm will be polynomial in 2^n. The issue of complexity is a severe problem for the ME approach (Ben-Eliyahu 1990), but the intention of this thesis is to expound the theoretical benefits of the maximum entropy ranking rather than its practicability. In chapter 7 details of an implementation of this and other default reasoning systems are given along with a comparative complexity analysis.

4.5 Uniqueness condition

A given solution of the constraint equations (4.10) and (4.11), in terms of a set of integer ranks for defaults, leads to a particular ranking over models. There may be many solution sets for these ranks of defaults which may or may not correspond to the same ranking over models. In section 4.3 it was shown that, for some default sets, there are strength assignments which correspond to no solution, to multiple solutions or to a unique solution. Under the analytic assumption, a unique solution implies that the order of the convergence functions of defaults is enough to determine the ME-ranking uniquely. Hence, it would be desirable to know whether any given solution to equations (4.10) and (4.11) was the unique one, or just one of many. Given that the rational consequence relation corresponding to a unique ME-ranking has the benefit of being the least biased, it is important to know whether this has been uniquely determined. Multiple ME-rankings imply that the strength assignment leads to some ME-redundancy in the default set and a knowledge engineer would, presumably, wish to be informed of this fact. In this section, a sufficient condition to determine uniqueness is given with a discussion about the possibility of identifying a necessary condition given at the end.

In their work on the random worlds semantics for statistical knowledge bases, Bacchus *et al.* (1996) encountered situations for which the convergence of their probabilistic inferences depended on the way in which the other probabilities converged. They termed such knowledge bases *non-robust*. In fact, their approach was shown to be closely related to the ME approach when restricted to knowledge bases with unary predicates (Grove *et al.* 1994). Their non-robust knowledge bases correspond to there being different ME-rankings for different strength assignments. In this approach it has been found that there are multiple solutions *even for just one strength assignment*. The term *robustness* is adopted here for ME-rankings which are unique with respect to a single strength assignment although, since the condition given is only sufficient, there may exist unique ME-rankings which fail to satisfy this robustness condition.

Definition 4.5.1 *An integer ranking, κ, over models is said to be* robust *with respect to a set of defaults if no two defaults share a common minimal falsifying model in κ.*

This definition may be applied to any ranking over models, not just those found using the ME approach.

The remainder of this section sets out to demonstrate that, if a computed ranking is robust with respect to a set of variable strength defaults, then it is also the unique solution for that set. To demonstrate this, the following definition and lemma are necessary before the main theorem can be proved:

Definition 4.5.2 *Two exact solutions to equations (4.10) and (4.11), κ and κ', are said to be* distinct *iff $\kappa(r) \neq \kappa'(r)$ for some default r. Such solutions are called* distinct solutions *for which some defaults are* distinctly ranked.

As before, let v_r, $v'_{r'}$ represent minimal verifying models of r, r' in κ, κ', respectively, and let f_r, $f'_{r'}$ represent minimal falsifying models of r, r' in κ, κ', respectively, and so on. The lemma shows that any distinctly ranked default, r, which has minimal $\kappa(f_r)$ among distinctly ranked defaults, also has minimal $\kappa'(f'_r)$ among distinctly ranked defaults.

Lemma 4.5.3 *Given two distinct solutions, κ and κ', if r is such that $\kappa(r) \neq \kappa'(r)$ and for all r' with $\kappa(r') \neq \kappa'(r')$, $\kappa(f_{r'}) \geq \kappa(f_r)$, then $\kappa'(f'_{r'}) \geq \kappa'(f'_r)$.*

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

$$\begin{aligned}\kappa(f_r) = s_r + \kappa(v_r) = s_r + \kappa'(v_r) \geq \\ s_r + \kappa'(v'_r) = \kappa'(f'_r) > \kappa'(f'_{r'})\end{aligned} \tag{4.19}$$

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, s, for which $\kappa'(s) = \kappa(s)$, and $\kappa'(v'_{r'}) = \kappa(v'_{r'})$. It follows that

$$\begin{aligned}\kappa'(f'_{r'}) = s_{r'} + \kappa'(v'_{r'}) = s_{r'} + \kappa(v'_{r'}) \geq \\ s_r + \kappa(v_{r'}) = \kappa(f_{r'}) \geq \kappa(f_r)\end{aligned} \tag{4.20}$$

Putting (4.19) and (4.20) together, $\kappa(f_r) \geq \kappa'(f'_r) > \kappa'(f'_{r'}) \geq \kappa(f_{r'}) \geq \kappa(f_r)$, which contradiction implies that $\kappa'(f'_{r'}) \geq \kappa'(f'_r)$, as required. □

Theorem 4.5.4 *Given a finite set of variable strength defaults, $\{r_i : a_i \overset{s_i}{\Rightarrow} b_i\}$, if an exact solution, κ, produces a robust ranking over defaults, then it is unique.*

Proof. Let κ and κ' be distinct solutions and r be a distinctly ranked default with minimal $\kappa(f_r)$ among distinctly ranked defaults and, by lemma 4.5.3, minimal $\kappa'(f'_r)$. Suppose that the rankig κ is robust. Then f_r falsifies only r and other defaults, s, with $\kappa(s) = \kappa'(s)$; also $\kappa(v_r) = \kappa'(v'_r)$ since they only falsify non-distinctly ranked defaults, and, since both κ and κ' are exact solutions, it follows that $\kappa(f_r) = \kappa'(f'_r)$ with $\kappa(r) \neq \kappa'(r)$.

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

Now, if f'_r falsified no other distinctly ranked default, $\kappa(f'_r) < \kappa'(f'_r) = \kappa(f_r)$, which contradicts f_r being minimal in κ. This implies that f'_r must falsify some other distinctly ranked defaults and hence κ' is not robust. Let these be $r_1, r_2, \ldots, r_n$; since all these r_i are also minimal distinctly ranked defaults in κ', by lemma 4.5.3, they are also minimal in κ and there must exist $f_{r_1}, f_{r_2}, \ldots, f_{r_n}$, minimally ranked falsifying models in κ such that $\kappa(f_r) = \kappa(f_{r_i})$ for all r_i. Further, because κ is robust, none of the f_{r_i} can falsify any other distinctly ranked defaults.

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

The robustness condition allows one to check whether the ranking produced by the ME algorithm is unique, however, there are two situations for which the robustness condition fails, but the solution may still be unique.

Firstly, given two distinct solutions, κ and κ', it may still be the case that $\kappa(m) = \kappa'(m)$ for all m, that is, the ranking over models may be unique despite there being multiple solutions for the $\kappa(r_i)$ to the constraint equations (4.10) and (4.11). For example, the set $\{r_1 : a \overset{s_1}{\Rightarrow} b, r_2 : \neg b \overset{s_2}{\Rightarrow} \neg a\}$, produces the two equations:

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

which have no solution unless $s_1 = s_2$ in which case there are an infinite number of exact solutions. However, all solutions lead to the same ranking over models

which is therefore unique. Clearly this set does not satisfy the robustness condition.

Secondly, it may be that the robustness condition is violated because two defaults share a common minimal falsifying model but they also have distinct minimal falsifying models. The solution found may still be unique but robustness fails.

There may well be a more precise condition which guarantees that the exact solution found is unique. For completeness, it would be desirable to identify a necessary condition; however, this is as far as this thesis goes in solving the uniqueness problem.

4.6 Interpreting computed rankings as ME-rankings

In the derivation of the ME-ranking equations (4.10) and (4.11), an assumption was made that all defaults are active (in order to apply the Lagrange multiplier technique); the purpose of this section is to clarify when this assumption is valid so that an exact solution to the derived equations can be interpreted as the unique ME-ranking. Further, it is argued that sometimes the ε^+-admissible ranking produced by the ME algorithm can also be interpreted as the unique ME-ranking. It is also argued that in cases where no unique ME-ranking exists, some redundancy is present leading to multiple possible interpretations of a set. Discussion of the other crucial assumption, that all variables are analytic, is deferred till the end of this chapter.

Consider the case of a set of active defaults, determining a unique, computed ME-ranking, which ME-entails another default, r, to some degree d. Now consider what happens when r is added to that set with an assigned strength of s. Bearing in mind that the default represents a new constraint on admissible rankings and, indeed, on the ME-distribution, it becomes clear that the strength s is critical to the effect that r has on the ME-ranking for the extended set.

Firstly, suppose that $s < d$, so that the constraint for r is already satisfied (as a strict inequality) in the original ME-ranking; since the maximum entropy distribution has already been found for the original set, adding this constraint cannot lead to a distribution with a higher value of entropy, so the default is effectively redundant. In such a case the assumption that all defaults are active is clearly violated meaning that equations (4.10) and (4.11) for the extended set do not reflect the ME-ranking,*even if an exact solution to them exists*. In fact, in such a case, the additional default must satisfy the constraint $s_r + \kappa(v_r) < \kappa(f_r)$ in the ME-ranking; its assigned rank should be zero so that it plays no part in shaping the ME-ranking.

Occasionally, exact solutions exist to the extended set of equations, in which case the default r will be assigned a *negative* rank by the algorithm. This occurs when the strictly redundant default comes to be ranked and its MINF(r) is finite; because the value of MINF(r) is greater than that of s_r + MINV(r), it will be assigned a negative rank, though clearly this is an aberration. If such a case arises, the solution computed by the ME algorithm is incorrect since the assumption that all defaults are active is false; the redundant default should be removed from the set and the ME algorithm re-applied to the remaining defaults. Usually, however, the strictly redundant default will have an infinite MINF(r) when it comes to be ranked, and will therefore be assigned a rank of zero automatically; in such cases, provided there are no other redundant defaults in the set, the ε^+-admissible ranking computed by the ME algorithm indeed represents the unique ME-ranking.

Secondly, consider the case when $s > d$. In this case, the additional default clearly represents a further constraint on the set of admissible rankings and hence it will be active. Provided this addition does not make any of the original defaults redundant, the ME algorithm will find the new, unique ME-ranking.

Thirdly, consider the case when $s = d$. In this case, either the additional default is redundant, or its addition will lead to one of the original defaults becoming redundant. These cases usually lead to multiple exact solutions to equations (4.10) and (4.11). It appears that, when there are several possibilities for redundancy, the asymptotic abstraction to the ME-distribution may depend critically on the coefficients of the convergence functions themselves, as discussed in section 4.3. In such cases, only the knowledge base designer can know what it is that he intends, that is, which default should be treated as redundant. Again, removal of the redundant default should restore all defaults to being active resulting in a successful application of the ME algorithm .

It is important, therefore, to treat the results of the ME algorithm carefully. Firstly, if any defaults have a zero or negative rank, then they should be removed and the algorithm re-applied to the remainder; such cases indicate that at least one default is not active, i.e., it is redundant. Secondly, the ranking from an exact solution should be tested for uniqueness; again, multiple exact solutions imply that the set contains redundancy. If it is not clear which defaults are candidates for redundancy, it is always possible to remove each in turn and determine whether the remaining set ME-entails the missing one. In practice it is unlikely that knowledge base designers would need to represent default knowledge bases which contain redundancy. As will be seen in the following chapter, practically all the benchmark

problems of default reasoning lead to exact solutions and unique ME-rankings. Indeed, only in the final example (see section 5.4) does a situation containing redundancy arise, and in this case the intuitions of the original researchers clashed.

4.7 Discussion

This chapter has introduced a refinement on the maximum entropy approach to default reasoning. By making slightly different assumptions from those of Goldszmidt (Goldszmidt, Morris, & Pearl 1993, Goldszmidt 1992), which commit the user to specifying the order of magnitude at which defaults converge, a more flexible means of representing default information and of computing the ME-ranking has been given. 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, it will be seen that some defaults are ME-entailed regardless of a strength assignment (e.g., ε-consequences, trivially, but others as well), 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 which suggest that this more general approach gives a very realistic account of what is meant by default reasoning. Firstly, the approach 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. In the following chapter, the ME-solutions to benchmark problems from the AI literature are explored. It is shown that the ME approach can be used to model these examples in a way that satisfies the requirements of a default reasoning process as laid out in chapter 1. The fact that this new approach can model both uncontentious default conclusions—those which are "intuitively" correct—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 it can be used to analyse the benchmark problems themselves to enable a better understanding of the structure of default reasoning.

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 solutions for default entailment. This indeed turns out to be the case for lexicographic entailment. In chapter 6, the revised ME approach is compared with two other systems of default reasoning. 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 the other systems can be examined. Thus the revised ME approach can be used as a benchmark system in its own right from which to assess other formalisms.

As the following chapters will demonstrate, this revised ME approach is widely applicable; however, several technical details with the formalism remain to be resolved. The previous section highlighted when the assumption that all defaults are active does not apply, and makes suggestions about how to proceed in the face of redundancy. However, perhaps a more critical assumption is that made to derive equations (4.10) and (4.11) in the first place; that is, the assumption that all variables are analytic. While it is clear that from a given set of specific convergence functions, which are defined as analytic, the steps leading to the equations are valid, it is not so clear that from exact solutions to those equations, actual convergence functions could be constructed. In order words, for a given set of strengths over the convergence functions, can all their corresponding ME-distributions be abstracted in this way? These technicalities are the subject of on-going investigation and future work.

Chapter 5

Analysis of benchmark problems

In this chapter the ME solutions to benchmark problems are examined. At least one attempt at standardising these problems was made by Lifschitz (1988), but mainly they form part of the lore of the nonmonotonic and commonsense reasoning communities. One criticism that has often been made of these areas is that they have focused too much attention on trying to solve a small number of toy problems, such as the Nixon diamond, and whether or not certain birds fly. Systems which have been designed to solve these problems may not scale up to larger problems and, even if they do, unwanted or counterintuitive conclusions often appear as side effects. From within the communities, researchers have argued that if a system cannot handle the toy problems correctly, there is little hope of it dealing with larger problems.

Certainly, a general theory of default reasoning must be capable of representing and reasoning correctly about these examples. The main problem with attempting to find such a formal theory is that, without a reference point from which to start, it is difficult to determine what is meant by "correct". Unfortunately this has meant that soundness has mainly been tested against subjective intuitions. While these "toy" examples have arisen through a broad consensus in certain areas, the results that should be obtained from more complex examples are far from clear. Additionally, intuitions about this type of reasoning may be altered by the very process of attempting to formalise it, as recognised by Nute (1980).

On the other hand, any system which solves these problems satisfactorily must at least be a candidate for a general theory of default reasoning. The aim of this thesis is to propose the ME approach as such a candidate. The argument to support this necessarily involves verifying its behaviour with respect to the benchmark problems and that is the aim of this chapter. The argument is supplemented with some observations about the nature and structure of the problems themselves.

5.1 Property inheritance and transitivity

The logical property of transitivity, that is, from $a \rightarrow b$ and $b \rightarrow c$ deduce $a \rightarrow c$, 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 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 (Makinson 1988). He suggested that nonmonotonic reasoning processes need only satisfy the first of these conditions. Makinson's analysis, along with that of Gabbay (1985), ultimately led to the formulation of the rule system P as core behaviour for nonmonotonic reasoning systems (Kraus, Lehmann, & Magidor 1990). 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 (Lehmann & Magidor 1992).

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 Reiter and Criscuolo's refusal to accept a transitive conclusion from default logic in the following case (Reiter & Criscuolo 1983). They 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. (Reiter & Criscuolo 1983)

[1] Equivalent to cautious monotonicity.

For whatever reasons, perhaps their own preconceived ideas about high school dropouts, Reiter and Criscuolo prejudged the results of the reasoning process, and, by requiring that it remain agnostic about a particular default conclusion, they unwittingly imposed an additional constraint. To resolve this, they resorted to introducing semi-normal defaults to default logic so that their desired conclusions could be obtained. But this process 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 when circumscription is applied to the same type of problems. It becomes necessary to introduce abnormality predicates and complicated axiomatisations of simple problems, so that the "correct" results can be obtained (McCarthy 1986, Lifschitz 1987). But this kind of tinkering ultimately leads to an invalidation of the entire use of defaults to represent a given problem. After all, if one knew all the conclusions in advance one could, in theory, simply program a system to reproduce them. By using defaults one hopes 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 this way since it will undoubtedly lead to incorrect conclusions in some cases.

To examine how the ME approach behaves with respect to transitivity, the following simple example is used:

$$\Delta = \{a \Rightarrow b, b \Rightarrow c\}$$

Whether defaults are viewed as some kind of inference rule, or as constraints, it is natural to think of a default as linking two formulæ; in either case, it seems hard to argue that $a \Rightarrow c$ should *not* be a default conclusion of this set, at least when the abstract symbols are not loaded with intuitive interpretations. While it is true that $a \Rightarrow c$ is not an ε-consequence of Δ, since it is not probabilistically sound, both falsifying models of $a \Rightarrow c$, must either falsify $a \Rightarrow b$ or $b \Rightarrow c$; on the other hand, $a \Rightarrow c$ has a verifying model, $a \wedge b \wedge c$, which does not falsify either; this explains why $a \Rightarrow c$ is an ME-consequence of Δ (for any strength assignment).

What this shows is that, under the ME approach, and in isolation, defaults *do* chain transitively. A similar observation is made by Kern-Isberner (1997). Obviously, other defaults which deal with the same propositions or formulæ, may cause interference which prevents this from happening, but, other things being equal, transitivity is to be expected.

What this means for Reiter and Criscuolo's example is that, by choosing not to accept the transitive conclusion that "typically high school dropouts are employed", they are imposing an extra constraint which implies an exceptional circumstance—that they wish to remain agnostic about a particular situation—and clearly this will need to be encoded when attempting to represent the problem using defaults.

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 chapter 7.

5.1.1 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, $a \wedge c \Rightarrow b$ is not an ε-consequence of the singleton set $\Delta = \{a \Rightarrow b\}$. This problem arises because the ε-semantics is based on constraints which hold in all ε-admissible rankings and some of these are ε-admissible with respect to the set $\{a \Rightarrow b, a \wedge c \Rightarrow \neg b\}$. In contrast, non-monotonic reasoning systems which allow a restricted form of transitivity, such as default logic and circumscription, do not suffer from this problem exactly because relevant formulæ explicitly block inheritance.

Under ME, models in which some irrelevant formula is true and those in which it is false are treated equally; therefore, the ME approach also does not suffer from the irrelevance problem. Returning to the example given above, $\Delta = \{a \Rightarrow b, b \Rightarrow c\}$, Δ ME-entails not only $a \Rightarrow c$ but also $a \wedge d \Rightarrow c$ and $a \wedge \neg d \Rightarrow c$, under any strength assignment. Thus the requirement of ignoring irrelevant information is satisfied by the ME approach.

5.2 Conflicting inheritance and specificity

From the last section, it appears that inheritance of properties from one class to another is the normal state of affairs. So what are the causes of a failure of property inheritance or transitivity? In this section, a simple example, and an extension of it,

demonstrate two ways in which inheritance may become blocked and show how the ME approach resolves the conflicts which arise.

One cause of blocked inheritance is when there are two defaults which point to opposite conclusions. While they do not conflict directly, if both their antecedents are satisfied simultaneously, it becomes unclear whether or not any default conclusion can be reached about the contentious consequent. Consider the following default set which corresponds to the Nixon diamond:

$$\Delta = \{a \overset{s_1}{\Rightarrow} c, b \overset{s_2}{\Rightarrow} \neg c\}$$

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 depends on the relative strengths associated with the two defaults. It is easily shown that $\text{ME}(a \Rightarrow c) = s_1$ and $\text{ME}(b \Rightarrow \neg c) = s_2$. So the ME-ranks of the relevant models are:

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

That is, for this example, the abnormality of a model depends directly on the strengths of those defaults it falsifies. This means that either $a \wedge b \Rightarrow c$ or $a \wedge b \Rightarrow \neg c$, or neither, may be ME-entailed depending on the comparative strengths s_1 and s_2. This result seems appropriate: when there is no reason to prefer one conclusion over another the result is ambivalent, but if one default is stronger its conclusion will prevail. The ME solution in this case resolves the conflict by giving equal overall weight to each default, but allows their relative strengths to tip the balance in favour of one or the other. At the same time, a truly ambiguous situation—when the strengths are equal—is also handled appropriately.

In the above example, the strengths were the deciding factor. In other circumstances, it may be the structure of the interaction between defaults which can force the conclusion one way or another, regardless of relative strengths. This structural prioritisation of defaults has been termed specificity 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, $a \Rightarrow b$, to Δ above, giving the set:

$$\Delta' = \{a \overset{s_1}{\Rightarrow} c, b \overset{s_2}{\Rightarrow} \neg c, a \overset{s_3}{\Rightarrow} b\}$$

For illustration, the full ME-solution will be calculated. Table 5.1 shows which defaults are verified and which falsified by each model. The first column under the heading $\text{ME}(m)$ gives the ME-rank of each model in terms of the ME-ranks of

m	a	c	b	$a \Rightarrow c$	$b \Rightarrow \neg c$	$a \Rightarrow b$	$\text{ME}(m)$	
m_1	0	0	0	-	-	-	0	0
m_2	0	0	1	-	v	-	0	0
m_3	0	1	0	-	-	-	0	0
m_4	0	1	1	-	f	-	$\text{ME}(r_2)$	s_2
m_5	1	0	0	f	-	f	$\text{ME}(r_1) + \text{ME}(r_2)$	$s_1 + 2s_2 + s_3$
m_6	1	0	1	f	v	v	$\text{ME}(r_1)$	$s_1 + s_2$
m_7	1	1	0	v	-	f	$\text{ME}(r_3)$	$s_2 + s_3$
m_8	1	1	1	v	f	v	$\text{ME}(r_2)$	s_2

Table 5.1: Calculating the ME-ranking for a conflicting default set.

the rules (which are the unknowns) using equation (4.10); the second column gives the final ME-rank of each model after the ME-ranks of rules have been found (which for this case is fixed since there is a unique solution corresponding to any strength assignment).

Substituting the unknown $\text{ME}(m)$ into the reduced constraint equations (4.11) gives rise to:

$$\begin{aligned}
\text{ME}(r_1) &= s_1 + \min(\text{ME}(r_2), \text{ME}(r_3)) \\
\text{ME}(r_2) &= s_2 + \min(\text{ME}(r_1), 0) \\
\text{ME}(r_3) &= s_3 + \min(\text{ME}(r_1), \text{ME}(r_2))
\end{aligned}$$

There is only one solution to this set of equations given by: $\text{ME}(r_1) = s_1 + s_2$, $\text{ME}(r_2) = s_2$, $\text{ME}(r_3) = s_2 + s_3$. The final ME-rank of each model is given in table 5.1 under the second column of $\text{ME}(m)$.

Now consider again whether an object which exhibits both properties a and b, inherits property c or $\neg c$. The ME-ranks of the relevant models are now:

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

The default conclusion $a \wedge b \Rightarrow c$ is therefore an uncontentious ME-consequence, being 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 $s_2 > s_1$. 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 the one which refers only to b; and because it is the more specific default which is active, the derived default conclusion is ME-entailed to the same degree as that which caused it. In fact, this resolution of conflict is

down to the ε-semantics, rather than its extension using ME, as $a \wedge b \Rightarrow c$ is also an ε-consequence of Δ'.

This example has illustrated 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 ambivalence 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 (indeed, this is illustrated in section 5.4). The following section examines how the ME approach handles inheritance to exceptional subclasses.

5.3 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. In fact, this is just a special case of transitivity holding for unexceptional cases. The example usually cited is that of penguins which, though they are an exceptional type of bird since they cannot fly, should still inherit other bird features like having wings. Again, the point of using defaults is to enable concise representation of a domain using general rules which can then be used to draw defeasible conclusions in the absence of complete information. Obviously defaults should therefore be presumed to hold *unless* there is information to the contrary. In this case the assumption is that the information to the contrary about a particular feature should not affect other features of the same status or at the same level. Another argument in favour of this presumptive use of defaults is that, although exceptions to defaults are known to be a possibility, the reason that superclasses exist at all is that objects can be 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.

m	b	f	p	w	r_1	r_2	r_3	r_4	$\text{ME}(m)$	
m_1	0	0	0	0	-	-	-	-	0	0
m_2	0	0	0	1	-	-	-	-	0	0
m_3	0	0	1	0	-	f	v	-	$\text{ME}(r_2)$	$s_1 + s_2$
m_4	0	0	1	1	-	f	v	-	$\text{ME}(r_2)$	$s_1 + s_2$
m_5	0	1	0	0	-	-	-	-	0	0
m_6	0	1	0	1	-	-	-	-	0	0
m_7	0	1	1	0	-	f	f	-	$\text{ME}(r_2) + \text{ME}(r_3)$	$2s_1 + s_2 + s_3$
m_8	0	1	1	1	-	f	f	-	$\text{ME}(r_2) + \text{ME}(r_3)$	$2s_1 + s_2 + s_3$
m_9	1	0	0	0	f	-	-	f	$\text{ME}(r_1) + \text{ME}(r_4)$	$s_1 + s_4$
m_{10}	1	0	0	1	f	-	-	v	$\text{ME}(r_1)$	s_1
m_{11}	1	0	1	0	f	v	v	f	$\text{ME}(r_1) + \text{ME}(r_4)$	$s_1 + s_4$
m_{12}	1	0	1	1	f	v	v	v	$\text{ME}(r_1)$	s_1
m_{13}	1	1	0	0	v	-	-	f	$\text{ME}(r_4)$	s_4
m_{14}	1	1	0	1	v	-	-	v	0	0
m_{15}	1	1	1	0	v	v	f	f	$\text{ME}(r_3) + \text{ME}(r_4)$	$s_1 + s_3 + s_4$
m_{16}	1	1	1	1	v	v	f	v	$\text{ME}(r_3)$	$s_1 + s_3$

Table 5.2: The ME-ranking for the penguin example.

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

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

Table 5.2 shows whether a model falsifies or verifies each default. The first column under the heading $\text{ME}(m)$ gives the ME-rank of each model in terms of the $\text{ME}(r_i)$ using equation (4.10). Substituting the $\text{ME}(m)$ into the reduced constraint equations (4.11) gives rise to:

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

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

Since this solution holds for any strength assignment, it follows that some default conclusions hold in general, in particular, it can be seen that the default $p \Rightarrow w$, is always an ME-consequence since:

$$\text{ME}(p \wedge w) = s_1 < \text{ME}(p \wedge \neg w) = s_1 + \min(s_2, s_4)$$

In this example, all falsifying models of $p \Rightarrow w$ either violate more defaults than its minimal verifying model, or ones which have higher ME-ranks. If $s_2 < s_4$ then the fact that falsifying $p \Rightarrow b$ is more serious than falsifying $b \Rightarrow f$ leads to the default being ME-entailed; if, on the other hand, $s_4 < s_2$ then the reason for inheritance is simply because it violates fewer defaults. Interestingly, in this case 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; this gives support to the view that an argument is only as strong as its weakest link. However, regardless of the strengths, the inheritance of w to p via b is uncontroversial since it will always be ME-entailed under any strength assignment.

The reason that the ME approach correctly handles inheritance to the exceptional subclass is because it assesses the models on the basis of all defaults they falsify and assigns defaults ME-ranks which incorporate both their relative strengths and their implicit priorities. Thus the ME approach sanctions the transitive conclusion $p \Rightarrow w$ but prohibits $p \Rightarrow f$.

Now suppose an object is encountered which would normally be assumed to be a member of some class but which displays none of its usual traits. Does there come a point at which it is easier to reject the object as being a member of the class than to accept that it is indeed a member but a highly exceptional one? Continuing with the example given above, consider the default "typically penguins without wings are birds", or $p \wedge \neg w \Rightarrow b$. The minimal verifying and falsifying models are:

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

In this case one of the models verifies the penguin/bird rule but violates both the bird/wings and the bird/fly ones; the other model violates the penguin/bird rule and so is not in a position to violate the others. But whether or not the target conclusion holds depends only on the relative strengths of r_2 and r_4, that is, how strongly are penguins birds relative to birds having wings. Note that the ME-ranks of defaults reflect both their strength and their specificity or structural priority. This allows the ME approach to weigh up the default violations fairly in cases where there are several exceptions.

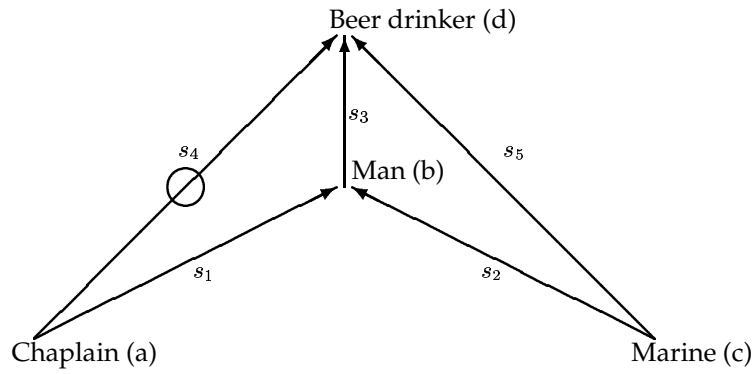

Figure 5.1: Graphical representation of Marine/chaplain example.

5.4 Multiple inheritance

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

The following default set, an example which demonstrates multiple inheritance, is an extension of a well-known controversial example from the field of inheritance hierarchies (see section 2.3). The original version appeared in several papers and caused much debate (Makinson & Schlechta 1991, Neufeld 1991, Touretzky, Horty, & Thomason 1987). The default set is given by:

$$\Delta = \{r_1 : a \overset{s_1}{\Rightarrow} b, r_2 : c \overset{s_2}{\Rightarrow} b, r_3 : b \overset{s_3}{\Rightarrow} d, r_4 : a \overset{s_4}{\Rightarrow} \neg d, r_5 : c \overset{s_5}{\Rightarrow} d\}$$

The controversy surrounding this example, depicted in figure 5.1 above, involves whether or not the default "Marine chaplains are beer drinkers" ($a \wedge c \Rightarrow d$) should be a default conclusion. In the original example, the direct link from Marine to beer

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

Table 5.3: The ME-ranking for the Marine/chaplain example.

drinker is omitted and the ME-solution is unique giving an uncontroversial ME-consequence of $a \wedge c \Rightarrow \neg d$ (i.e., "Marine chaplains are *not* beer drinkers"). This result should be unsurprising since 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 example.

However, Touretzky *et al.* (1987) 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 added, creating a direct link between Marines and beer drinkers. Now this default is already an ME-consequence of the original set and is ME-entailed to degree $\min(s_2, s_3)$. Table 5.3 shows whether a model falsifies or verifies each default, and the unknown ME-ranks for each model are given in the final column according to equation (4.10).

Substituting the $\text{ME}(m)$ into the reduced constraint equations (4.11) gives rise to:

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

which, if $s_5 > \min(s_2, s_3)$, has a 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 $\text{ME}(r_5) = s_5 - \min(s_2, s_3)$. If, however, $s_5 < \min(s_2, s_3)$, the default r_5 is effectively redundant and the equations cannot be solved as equalities. By assigning r_5 an ME-rank of zero, the ME-ranking for the original problem is recovered. For the in-between case when $s_5 = \min(s_2, s_3)$, there are multiple solutions and the ranking computed by the ME algorithm is non-robust.

Looking only at the case 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 ME-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:

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

Clearly the default conclusion obtained under ME will depend critically 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. Indeed, in some ways this can be seen as an extended and more complex case of what occurs in the Nixon diamond.

This example has demonstrated that the ME approach can be used to clarify the ambiguities which 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.

5.5 Discussion

This chapter has seen the ME approach applied to many of the benchmark problems for default reasoning, and it has been shown to satisfy them in an appropriate manner. Apart from actually arriving at the default conclusions which many researchers have considered to be the "intuitively correct" ones, the ME approach has behaved well in cases where there has been confusion and disagreement as to what the correct default conclusions should be. Under ME, some strongly justified behaviour, like respect for specificity, is easily obtained, while other more flexible conclusions, like the ability to switch conclusions according to strengths in the Nixon diamond, also arises quite naturally. It appears that these behaviours are not only "intuitive" but sound with respect to this probabilistic interpretation of defaults in conjunction with a principle of indifference. As such, the ME approach might be used to *validate* the intuitions underlying the benchmark examples rather than simply being seen to satisfy them. This contrasts sharply with other approaches to default reasoning which have used the benchmark behaviour as a requirement that a proposed system should satisfy without any objective reference to where the behaviour comes from. As will be seen in the following chapter, the success of ME in explaining default behaviour is not restricted to the benchmark problems but continues at more abstract levels. It will be shown that ME subsumes another system which has also been shown to satisfy the benchmarks but whose semantics was unclear, and that its treatment of redundancy allows the amalgamation of two different views of how systems should behave under belief revision.

Chapter 6

Comparing LEX and Z^+ with ME

The last chapter looked at how the ME approach handles benchmark examples of default reasoning. In this chapter, the higher level behaviour of the ME approach is compared with two other default systems, LEX and Z^+. Like the ME approach, both systems LEX and Z^+ lead to rational consequence relations and therefore it is interesting to look at the differences and similarities between the systems from a higher level perspective, i.e., not just actual differences in consequences obtained from specific default sets. This thesis argues that applying the principle of maximum entropy to a set of variable strength defaults results in the most acceptable rational consequence relation since it is the least biased. By comparing ME with other systems, it may be possible to describe the ways in which the other systems deviate from the least biased answers and thereby discover the additional assumptions which underlie them.

The lexicographic entailment of Lehmann (1995) uses the Z-partition to prioritise defaults leading to the LEX-ordering over models. In section 6.1 it is shown that for any ε-consistent set of defaults, its LEX-ordering can be translated into a class of ME-rankings with corresponding strength assignments. From the characteristics of these strength assignments, it is shown that LEX-entailment can be viewed as a crude form of ME-entailment in which the order of magnitude of each default's strength corresponds to the Z-rank of its minimal falsifying model. This implies that ME-entailment both subsumes LEX-entailment, and is strictly more expressive than it.

The system Z^+, like ME, produces different consequence relations for different strength assignments over defaults. Indeed the systems appear remarkably similar on a number of levels. In section 6.2, these similarities are highlighted and the relative merits of both systems are assessed.

An important means of judging the reasonableness of a default reasoning system is to examine its behaviour when the default set changes—a form of belief revision. The requirements of belief revision are as nebulous as those of default reasoning itself[1]; however, the main intuition is one of minimal change or continuity of beliefs in the face of incremental changes in knowledge (Gärdenfors 1988). To test how these systems behave, in section 6.3 all three are examined to see the effect on their outputs when previously derivable defaults are added.

The results of these comparisons will be used in chapter 8 to argue in support of the thesis that the ME approach provides the most acceptable consequence relation. Some of the results from this chapter first appeared in (Bourne & Parsons 1999b).

6.1 Comparison of LEX with ME

In section 3.5 the LEX-entailment of Lehmann (1995) which uses the Z-partition to determine the relative priorities of defaults was presented. The LEX consequence relation, being a direct extension of the Z consequence relation, satisfies the requirements of a default reasoning system including inheritance to exceptional subclasses, as example 3.5.1 demonstrated. But the ME approach also satisfies these requirements, and therefore it may be useful to compare the two systems to see whether they are connected semantically. It turns out that LEX-entailment can be viewed as a crude form of ME-entailment with certain implications for the strengths of defaults which are not entirely satisfactory. This section demonstrates this point by showing that the lexicographical ordering induced by any ε-consistent set of defaults can be translated into an ME-ranking.

The similarity between LEX and ME lies in the fact that in both systems the preference relation over models makes use of all defaults violated by each model: the LEX-tuple of a model represents the position and number of defaults it falsifies, while the ME-rank of a model is the sum of the ME-ranks of each default it falsifies. Because of this similarity, it is possible to assign appropriate ME-ranks to defaults so that the ME-ranking over models produces an equivalent consequence relation to that produced by the LEX-ordering[2]. From this ME-ranking, one can compute

[1] Although postulates for belief revision have been formalised in Gärdenfors (1988) and extended for iterated belief revision in Darwiche and Pearl (1997), their justifications are based only on intuitive arguments.

[2] Note that simply assigning ME-ranks to defaults arbitrarily does not necessarily lead to an ME-ranking over models since the ranking induced by the assignment may fail to be admissible. However, since the LEX-ordering is admissible, the assignment in the translation algorithm *will* lead to an ME-ranking over models.

Translation algorithm

Input: The Z-partition of Δ, $\Delta_0 \cup \Delta_1 \ldots \cup \Delta_n$.
Output: The canonical ME-ranking, ME_Δ, plus canonical strength assignment, $\{s_i\}$.

[1] Let $\text{ME}(r_i) = 1$ for all $r_i \in \Delta_0$.

[2] For $k = 1$ to n:

(a) Let $\text{ME}(\Delta_k) = (|\Delta_{k-1}| + 1) * \text{ME}(\Delta_{k-1})$.

(b) Let $\text{ME}(r_i) = \text{ME}(\Delta_k)$ for all $r_i \in \Delta_k$.

[3] For each r_i:

(a) Find the ranks of its minimal verifying and falsifying models, $\text{ME}_\Delta(v_{r_i})$ and $\text{ME}_\Delta(f_{r_i})$, using equation (4.10).

(b) Set $s_i = \text{ME}_\Delta(f_{r_i}) - \text{ME}_\Delta(v_{r_i})$.

Figure 6.1: The translation algorithm

a strength assignment over defaults. Since the LEX-ordering is determined by the Z-partition, it is unique for a given set of defaults, however, it will be seen that the choice of ME-ranks is arbitrary, to an extent, and hence there exists a whole class of ME-rankings which produce consequence relations equivalent to the LEX consequence relation. Each of these ME-rankings implies a corresponding set of strengths for the defaults. The characteristics of these strength assignments serve as a means of interpreting the nature of the priorities that LEX assigns to defaults.

The translation algorithm which finds an ME-ranking equivalent to the LEX-ordering is given in figure 6.1 and is motivated as follows. According to the method of comparison for the LEX-tuples associated with models, it is more costly for a model to violate a default in a higher partition set than one in a lower set, other things being equal, and it is more costly for a model to violate more defaults of a given priority (in the same partition set) than fewer, other things being equal. The ME-rank of a model is determined by summing the ME-ranks of those defaults it falsifies; these must therefore be chosen so as to ensure that the sum reflects both the priority and the number of defaults falsified by each model. Clearly, defaults in higher partition sets will require higher ME-ranks, and defaults in the same partition set will require the same ME-rank, ensuring that whenever two models falsify different defaults which belong to the same partition set, the "penalty" associated with each is the same. In addition, it must always be worse to falsify a single de-

fault in a certain partition set than to falsify *any number* of defaults in lower sets. Thus the ME-rank assigned to defaults in the partition set Δ_i, denoted $\text{ME}(\Delta_i)$, must be greater than the sum of the ME-ranks of all defaults in lower sets; that is, it must be of a *higher order of magnitude*. The algorithm therefore proceeds by, initially, assigning each default in the first partition set an ME-rank of 1. Subsequently, it iteratively computes the minimum ME-rank required for defaults in the next highest partition set and assigns that rank to each of them. When all ME-ranks have been assigned, it computes the ME-ranks of the minimal verifying and falsifying models for each default, from which it can determine their strength.

This translation algorithm contains a certain arbitrariness since at step [2](a), any integer higher than the sum of all previously ranked defaults would suffice. Thus there is a whole class of ME-rankings which are equivalent to the LEX-ordering for any ε-consistent set of defaults. The translation algorithm should therefore be treated as the definition of a particular member of this class which will be called the *canonical* ME-*ranking* and denoted ME_Δ. Similarly, the strength associated with each default by the translation algorithm, being the difference between the ME-ranks of its minimal falsifying and verifying models, will be called the *canonical* ME-*strength* of that default. Note that not just the original defaults, but any default which is LEX-entailed (and hence also canonically ME-entailed), will have an associated canonical ME-strength.

But, as was seen in chapter 4, some strength assignments lead to multiple ME-rankings, so it is pertinent to ask: does the canonical strength assignment always lead to a unique ME-ranking? In fact, it may well be that it does not. If the LEX-ordering is robust then, since the ME-ranking induces the same consequence relation, it too will be robust and the canonical ME-strengths will produce a unique ME-ranking. If, however, the LEX-ordering is not robust, it means that the canonical ME-strength assignment may lead to multiple ME-rankings. But the canonical ME-ranking, as defined by the translation algorithm, will certainly be one of these multiple ME-rankings even though it may not be unique. The point is that, for any LEX-ordering, there will always be an ME-ranking which produces an equivalent consequence relation. Since the purpose of this comparison is to establish that the LEX-ordering is a type of ME-ranking, these other ME-rankings are of no relevance to the analysis.

The following example shows the translation algorithm at work leading to a canonical ME-strength assignment which gives an identical rational consequence relation to that given by the LEX-ordering.

Example 6.1.1 (Bears)

$$\Delta = \{r_1 : b \Rightarrow d, r_2 : t \Rightarrow b, r_3 : t \Rightarrow \neg d, r_4 : b \Rightarrow h, r_5 : t \wedge l \Rightarrow d\}$$

(the intended interpretation of this knowledge base is that bears are dangerous, teddies are bears, teddies are not dangerous, bears like honey, and teddies with loose glass eyes are dangerous). The Z-partition has three partition sets:

$$\Delta_0 = \{b \Rightarrow d, b \Rightarrow h\} \qquad \Delta_1 = \{t \Rightarrow \neg d, t \Rightarrow b\} \qquad \Delta_2 = \{t \wedge l \Rightarrow d\}$$

Following the translation algorithm, set $\text{ME}(r_1) = \text{ME}(r_4) = 1$; then $\text{ME}(\Delta_1) = 3$, so $\text{ME}(r_2) = \text{ME}(r_3) = 3$; finally $\text{ME}(\Delta_2) = 9$, so $\text{ME}(r_5) = 9$. The canonical ME-ranking is robust and corresponds to a canonical ME-strength assignment of $(1, 2, 2, 1, 7)$. The LEX-ordering and canonical ME-ranking both induce the same rational consequence relation. An example of a default entailed by both systems is "teddies which are dangerous and do not like honey are bears", or $t \wedge d \wedge \neg h \Rightarrow b$:

$$\begin{aligned} \text{LEX}(t \wedge d \wedge \neg h \wedge b) = (1,1,0) \quad &\prec \quad \text{LEX}(t \wedge d \wedge \neg h \wedge \neg b) = (0,2,0) \\ \text{ME}_\Delta(t \wedge d \wedge \neg h \wedge b) = 4 \quad &< \quad \text{ME}_\Delta(t \wedge d \wedge \neg h \wedge \neg b) = 6 \end{aligned}$$

and so this default is both LEX-entailed and canonically ME-entailed. □

Because the translation algorithm finds an ME-ranking corresponding to the LEX-ordering for any set of defaults, it appears that the LEX consequence relation is just a special case of the ME consequence relation when applied to a particular class of strength assignments. So what characterises the ME-rankings that simulate the LEX-ordering?

The class of ME-rankings can be characterised by examining the nature of the strength assignments which lead to them. These will reveal what implicit priorities LEX assigns to defaults. As already noted, the assignment at step [2](a) of the translation algorithm is arbitrary to an extent so long as the ME-rank of each partition set is of a *higher order of magnitude* than the ME-rank of the previous one. In fact this is implicit in the way the LEX-tuples are ordered. The Z-rank of a default, which corresponds to the Z-rank of its minimal falsifying model, can be thought of as being the order of magnitude of its associated ME-rank[3]. But if the ME-ranks of defaults increase exponentially with their partition sets, what happens to their corresponding canonical ME-strengths?

[3]Note that this should not be confused with the meaning of the Z-ranks for models which represent abstractions for the exponents of their relative probabilities. A model with a higher Z-rank has a lower absolute probability of being satisfied.

By considering only the order of magnitude of both the ME-ranks and the canonical ME-strengths, one can reason as follows. The nature of the Z-partition (see section 3.2) means that a default in set Δ_i will be verified by some model which falsifies at least one default in Δ_{i-1} but none in Δ_i or higher sets; hence it will have an ME-rank of the same order as $\text{ME}(\Delta_{i-1})$. Its minimal falsifying model, on the other hand, is guaranteed to falsify a default in Δ_i, i.e., itself, and will therefore have an ME-rank of the same order as $\text{ME}(\Delta_i)$, or possibly higher. The canonical ME-strength, being the difference between these two values, will clearly be of the same order as the rank of its minimal falsifying model, since subtracting a quantity of a lower order of magnitude makes little difference to the higher quantity. So the order of magnitude of the canonical ME-strength of a default depends critically on the order of magnitude of the highest priority default which is falsified by its minimal falsifying model. This is interesting since it means the canonical ME-strength of a default, at least up to its order of magnitude, is not determined by the partition set to which it belongs, but by its minimal falsifying model. In short, *the order of magnitude of the canonical* ME*-strength of a default is given by the* Z*-rank of its minimal falsifying model*.

So LEX-entailment can be interpreted as a form of ME-entailment which assigns strengths of higher orders of magnitude to defaults which are less likely to be falsified. This exposes an implicit assumption of LEX-entailment which at first glance might not seem unreasonable—it assigns exponentially higher strengths to defaults which are relatively less likely to be falsified. However, the fact that a default is very unlikely to be falsified, in absolute terms, usually reflects the fact that it may also be very unlikely to be verified, and this occurs because of its interactions with other defaults. The implication is that the meaning of a default changes according to which defaults surround it, i.e., its meaning is highly context dependent. The ramifications of this will be seen more clearly in section 6.3.

In contrast, in the ME approach, the strengths of defaults are assigned *relative* to other defaults but otherwise independently. The strength represents the difference between the likelihood of the default being falsified or verified independent of how likely it is to be verified in absolute terms. It should be easier for a knowledge base designer to use a system in which all defaults have equivalent meanings than one in which their meanings are dynamic—after all, it is a consequence relation which he is trying to construct, a difficult task when the meanings of objects representing his knowledge keep changing.

The restriction imposed by the LEX-ordering not only commits defaults to having strengths inversely proportional to their likelihood of being falsified, but also commits these strengths to increase exponentially as their chance of being falsified decreases. The priorities assumed by LEX-entailment are therefore *doubly* exponential when viewed from their ME translation—a very blunt instrument for resolving conflict between defaults. In contrast, the ME approach offers a much greater degree of control and therefore a more subtle and flexible means of resolving conflict.

Consider again the Nixon diamond:

$$\Delta = \{q \overset{s_1}{\Rightarrow} p, r \overset{s_2}{\Rightarrow} \neg p\}$$

Since the two defaults do not conflict directly, there is just one partition set, which means that LEX-entailment will never be able to distinguish between the two models $q \wedge r \wedge p$ and $q \wedge r \wedge \neg p$. In contrast, under the ME approach either of the defaults $q \wedge r \Rightarrow p$ and $q \wedge r \Rightarrow \neg p$, or neither, may be ME-entailed according to the relative strengths s_1 and s_2 (see the example on page 83). Thus the ME approach is far more expressive and, in consequence, likely to be much more useful.

Another failing of LEX is that there may be cases for which the conclusions it reaches are strength dependent ME-consequences, which means that, under some strength assignments, the converse of a LEX-consequence may be an ME-consequence. For instance, in the penguin example (see page 51), the default $p \wedge \neg f \wedge \neg w \Rightarrow b$ is a LEX-consequence, but under the ME approach, if the default $b \Rightarrow w$ is stronger than $p \Rightarrow b$ then the converse $p \wedge \neg f \wedge \neg w \Rightarrow \neg b$ is ME-entailed (see page 86). The fact that LEX cannot represent this equally valid alternative interpretation makes it less likely to be useful.

6.2 Comparison of Z⁺ with ME

System Z⁺ and the ME approach appear, at least on a structural level, to be even more closely related than LEX and ME. Both use variable strength defaults to constrain a ranking function in the exactly same way, that is, for a default $a \overset{s}{\Rightarrow} b$, both the Z⁺- and ME-rankings satisfy:

$$\kappa(a \wedge b) + s \leq \kappa(a \wedge \neg b) \tag{6.1}$$

Both systems also assign specific ranks to each default which are used to compute the ranking, although in different ways. Indeed, the algorithms used to compute these ranks are themselves very similar (figures 4.1 and 3.3). As discussed in section 4.6, some strength assignments may lead to defaults being redundant under ME, and it turns out that the same occurs for system Z⁺. This section examines

these similarities but shows that the systems are only related superficially. It is also shown that, while the ME-ranks assigned to defaults represent their influence on the consequence relation, the Z⁺-ranks reveal little about a default's significance.

The fact that the constraint (6.1) is an inequality, under both systems, is important because it means that a set of defaults can be considered consistent under any strength assignment. The equivalence of ε-consistency and ε^+-consistency (theorem 3.3.5) was proved by Goldszmidt and Pearl (1996) when defining the system Z^+, but it applies equally well for the more general system of variable strength defaults, ε^+, and the ME approach.

Both systems Z^+ and ME assign ranks to defaults, but these differ both in the values they take and in the way in which they are used. Recall the definition[4] of system Z^+:

$$Z^+(m) = \begin{cases} 0 & \text{if } m \text{ falsifies no default in } \Delta^+ \\ \max\limits_{m \models a_i \wedge \neg b_i} [Z^+(r_i)] & \text{otherwise} \end{cases} \tag{6.2}$$

where $Z^+(r_i)$ is a priority ordering on rules, defined by:

$$Z^+(r_i) = s_i + \min_{m \models a_i \wedge b_i} [Z^+(m)] \tag{6.3}$$

So the Z^+-rank of a default corresponds exactly to the Z^+-rank of its minimal falsifying model. In contrast, since the ME-rank of a model is determined by the sum of the ME-ranks of the defaults it falsifies, the ME-rank of a default signifies a weighted contribution to the ranks of its falsifying models. Thus, provided a default is not redundant, i.e., provided its ME-rank is non-zero, it contributes to *every* model it falsifies. The same cannot be said of the Z^+-ranking since each model is affected only by the highest ranked default it falsifies. Because of this, it should be clear that the Z^+-ranking ignores a lot of information—only maximal Z^+-ranks are used—whereas in the ME-ranking, every default violation has some effect but the effects of some defaults are stronger than others. The ME-rank of a default in some sense measures its significance in shaping the consequence relation produced. A simple example to illustrate this point is given by redundant defaults. Under ME, a redundant default has an ME-rank of zero, whereas under Z^+, a redundant default will take the rank of the highest default falsified in its minimal falsifying model. In effect the ME-rank provides some clue as to the significance of a default, but its Z^+-rank reveals little.

[4]Again note that this thesis uses a slightly different definition for system Z^+ than that given in (Goldszmidt & Pearl 1996).

Both systems Z⁺ and ME are determined from two sets of coupled equations. One set of equations computes the rank of a default using the ranks of models, and the other set computes the rank of a model using the ranks of defaults. As originally pointed out by Goldszmidt *et al.* (1993), this apparent circularity is benign because, in both systems, the ranks for defaults can be computed iteratively. This stems from the fact that for any ε-consistent set of defaults, at least one default can be verified without falsifying any others. Thus it is possible to compute the ranks of defaults "bottom up" starting with the default whose minimal verifying model has rank zero and which has the lowest strength. This accounts for the similarity in the two algorithms, both proceeding in the same fashion, ranking defaults one by one. The algorithm for system Z⁺ is relatively simple since a default's rank is merely the rank of its minimal falsifying model. In contrast the algorithm for ME is a little more complex since it is sometimes required to break cycles. This reflects one semantic difference in the two systems: the Z⁺-ranking is uniquely defined for a given set whereas the same set may have multiple ME-rankings, although this only occurs when a strength assignment implies that more than one default is a candidate for redundancy—not a case that would occur often in practice.

It was seen in section 4.6 that for some strength assignments, an individual default may be redundant under ME, being already ME-entailed by the other defaults to a degree higher than its assigned strength. If this occurs, the constraint for the redundant default is satisfied as a strict inequality in the ME-ranking, and the default itself will have an ME-rank of zero. A similar situation can arise with system Z⁺ as the following example demonstrates.

Example 6.2.1

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

The constraints on the Z⁺-ranking are given by:

$$\text{Z}^+(a \wedge b) + s_1 \quad \leq \quad \text{Z}^+(a \wedge \neg b) \tag{6.4}$$

$$\text{Z}^+(a \wedge b \wedge c) + s_2 \quad \leq \quad \text{Z}^+(a \wedge \neg(b \wedge c)) \tag{6.5}$$

Both falsifying models of r_1, $a \wedge \neg b \wedge c$ and $a \wedge \neg b \wedge \neg c$, also falsify r_2, and r_2 has a third falsifying model, $a \wedge b \wedge \neg c$. Now if $s_1 = s_2$, or if $s_2 < s_1$, it follows that both constraints (6.4) and (6.5) can be satisfied as equalities. If, however, $s_1 < s_2$, it follows that (6.5) is satisfied as an equality but because both falsifying models of r_1 have a Z⁺-rank of at least s_2, (6.4) will be satisfied as a strict inequality. In this latter case, although the Z⁺-rank of r_1 is indeed s_1, no model actually takes this

rank and the default is redundant: removing r_1 from Δ would leave the Z⁺-ranking unchanged. In fact, if $s_1 < s_2$, $\Delta - \{r_2\}$ Z⁺-entails r_1 to degree s_2. □

The results for example 6.2.1 could easily have been anticipated since the first default is an ε-consequence of the second. In such a case one should expect that the derived default must be at least as strong as the default which constrains it. However, this does not mean that ε-consequences are always redundant for if they are assigned higher strengths than those from which they are derived, their own constraints will have an influence on both the Z⁺-ranking and the ME-ranking.

One might speculate that, given the similarity between the two systems, they may be related to the extent that a default which is redundant in one system is also redundant in the other. Indeed this is often the case—for example, if a default were ε^+-entailed to a certain degree but assigned a strength strictly less than this degree, it is obvious from the definition of ε^+-entailment that it must be redundant under both systems. However, redundancy can be system dependent, as the following example demonstrates.

Example 6.2.2

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

As can be seen from table 6.1, the strength assignment $(1, 1, 1, 2)$ leads to a Z⁺-ranking in which all defaults are Z⁺-entailed to the same degree as their strength, but in the ME-ranking r_4 is ME-entailed to degree 3 so it is redundant and has an ME-rank of 0. In the strength assignment $(1, 1, 1, 3)$, one ME-solution is given by $\text{ME}(r_1) = 1$, $\text{ME}(r_2) = 1$, $\text{ME}(r_3) = 3$, $\text{ME}(r_4) = 0$, so that r_4 is redundant, but in the Z⁺-ranking, r_3 is Z⁺-entailed to degree 2, greater than its assigned strength, and it is therefore redundant. □

Although this example is a little contrived, it clearly demonstrates that defaults need not be inherently redundant for a given strength assignment but that this may be system dependent. This result shows that, despite their similarities, Z⁺ and ME cannot be connected at a semantic level.

Perhaps it should not be surprising that the systems are only connected structurally, given their different underlying motivations. Whereas system Z⁺ is based on minimising the ranks of models, ME is based on maximising the uncertainty represented in a distribution. It should be clear that forcing the ranks of models to be as low as possible also reduces the uncertainty in a consequence relation, hence the different results of the two systems. Indeed, this is the additional assumption which underlies system Z⁺; the user must judge whether or not it is reasonable.

								$(1,1,1,2)$		$(1,1,1,3)$	
m	a	b	c	$a \Rightarrow b$	$a \Rightarrow c$	$\neg b \wedge \neg c \Rightarrow a$	$\neg a \wedge \neg b \Rightarrow c$	Z^+	ME	Z^+	ME
m_1	0	0	0	-	-	f	f	2	3	3	3
m_2	0	0	1	-	-	-	v	0	0	0	0
m_3	0	1	0	-	-	-	-	0	0	0	0
m_4	0	1	1	-	-	-	-	0	0	0	0
m_5	1	0	0	f	f	v	-	1	2	1	2
m_6	1	0	1	f	v	-	-	1	1	1	1
m_7	1	1	0	v	f	-	-	1	1	1	1
m_8	1	1	1	v	v	-	-	0	0	0	0

Table 6.1: The Z^+- and ME-rankings for two different strength assignments.

It is interesting to note that, although the idea that lies behind it is quite simple, ME has a much more sophisticated behaviour since it can capture more subtle differences in models than the rather crude minimisation underlying system Z^+. The ME-ranks of defaults are also better indicators of their significance to the ME-ranking than the Z^+-ranks are to the Z^+-ranking. These findings suggest that the ME approach is not only better motivated semantically than Z^+, but also that it is far more expressive, although neither system subsumes the other.

6.3 Dynamic behaviour

All systems compared in this chapter are based on the same framework. That is, a set of defaults is used to induce a ranked ordering of models corresponding to a rational consequence relation which determines which further defaults are entailed.

In the same way that reasonable properties have been proposed for nonmonotonic consequence relations, the theory of belief change has postulated reasonable behaviours for systems of nonmonotonic reasoning themselves, that is, how do the outputs of a reasoning system change as the inputs change—a process of belief revision (Gärdenfors 1988). While it is difficult to prescribe the behaviour of a system when radical new beliefs are to be incorporated—what might be called a paradigm shift in beliefs—it seems reasonable to require that there ought to be some continuity in beliefs when the new beliefs to be incorporated are things which had previously been expected or derivable. For nonmonotonic consequence relations this was formalised as two rules: cautious monotonicity and rational monotonicity. Cautious monotonicity (CM) requires that when one learns a fact which was

previously a defeasible belief, one should not retract any other defeasible beliefs. Rational monotonicity is slightly stronger and requires that when one learns a fact which does not contradict a previously held defeasible belief, one should not retract any other defeasible beliefs.

This section examines the three systems, Z^+, ME and LEX, to see whether they satisfy the property of CM at the higher level when the new beliefs to be incorporated are defaults themselves. Firstly, though, it is necessary to formalise exactly what is meant by "learning a previously held defeasible belief"; whereas the rules of system P work well when applied to beliefs which are propositional formulæ, their meaning is less clear when applied to beliefs which are defaults. Since the inputs to these systems are sets of defaults, adding new beliefs means adding defaults to the original set and, for both Z^+ and ME, this means they must be given appropriate strengths.

The behaviour of Z^+ was established in the previous section. Adding a default which was previously Z^+-entailed with the same strength as the degree to which it was entailed does not affect the Z^+-ranking. The Z^+-ranking therefore corresponds to the rational closure of Δ^+ using variable strength defaults, in the same way that system Z (Pearl 1990) corresponds to the rational closure of Δ using standard defaults (Lehmann & Magidor 1992). This behaviour is extremely straightforward and system Z^+ clearly satisfies CM.

The behaviour of ME, however, is somewhat more complex. If a default is ME-entailed to some degree and is subsequently added to that degree, the resultant ME-ranking is highly likely to be non-robust, implying that other ME-rankings exist. Since one of these ME-rankings will be that in which the added default has an ME-rank of zero, its addition is effectively ignored and the ranking is identical with the original. The fact that this unchanged ranking is bound to exist clearly demonstrates that the ME approach can satisfy CM. However, there may be other ME-rankings, each of which is also perfectly reasonable given the data. In these other rankings it will be another default which is redundant and the one which is added will have a non-zero ME-rank. These alternative ME-rankings can quite possibly lead to old beliefs being retracted leading to a failure of CM. This behaviour is very interesting since what appears to be happening is that learning something one previously anticipated may "explain away" other parts of the knowledge base, a characteristic which is considered useful in evidence-based reasoning (Pearl 1988). So ME can both satisfy CM and fail to satisfy it, depending on which ME-ranking is considered preferable. This ambiguity may not be entirely satisfactory for those

who insist that CM is desirable, but it does offer a justification for the opposing view that it need not necessarily always apply. In some ways this strengthens the argument for the acceptability of the ME approach since it lends support to both points of view and, perhaps, explains why there are proponents both for and against satisfying CM.

The remainder of this section examines the behaviour of LEX when new defaults are added. First, theorem 6.3.1 shows how the Z-partition changes when a LEX-entailed default is added.

Theorem 6.3.1 (Dynamics of Z-partition) *Consider a set of defaults, Δ, with Z-partition $\Delta_0 \cup \ldots \cup \Delta_n$. Let r be a default LEX-entailed by Δ such that the Z-rank of its minimal verifying model is k. Then (1) the Z-partition of $\Delta' = \{r\} \cup \Delta$ is such that $\Delta'_i = \Delta_i$ for $i < k$, (2) $r \in \Delta'_k$ and (3) for all $r' \in \Delta_{j \geq k}$, either $r' \in \Delta'_j$ or $r' \in \Delta'_{j+1}$.*

Proof. All confirming models for the defaults in $\Delta_0 \cup \ldots \cup \Delta_{k-1}$ neither verify nor falsify r by the conditions of the theorem, hence the first k partition-sets in the new Z-partition will be the same, that is, for $i < k$, $\Delta'_i = \Delta_i$, as required.

Now if v_r is a minimum verifying model of r, it is also a confirming model for r with respect to $\{r\} \cup \Delta_k \cup \ldots \cup \Delta_n$, since it may falsify defaults in $\Delta_{i<k}$ but not in higher sets. Thus $r \in \Delta'_k$, as required.

Finally, consider $v_{r'}$, a verifying model for some default $r' \in \Delta_k$ which previously confirmed r' with respect to $\Delta_k \cup \ldots \cup \Delta_n$. If $v_{r'}$ satisfies r then it is also a confirming model of r' with respect to $\{r\} \cup \Delta_k \cup \ldots \cup \Delta_n$, so $r' \in \Delta'_k$. Otherwise r' does not tolerate $\{r\} \cup \Delta_k \cup \ldots \cup \Delta_n$. Therefore separate Δ_k into those defaults which tolerate $\{r\} \cup \Delta_k \cup \ldots \cup \Delta_n$, say Δ_{T_k}, and those which do not, say $\Delta_{\neg T_k}$. Then $\Delta'_k = \{r\} \cup \Delta_{T_k}$ and it remains to partition $\Delta_{\neg T_k} \cup \Delta_{k+1} \ldots \cup \Delta_n$. Clearly all defaults in $\Delta_{\neg T_k}$ tolerate $\Delta_{\neg T_k} \cup \Delta_{k+1} \ldots \cup \Delta_n$ since they did previously and so $\Delta_{\neg T_k} \subset \Delta'_{k+1}$. Separate Δ_{k+1} into those defaults which tolerate $\Delta_{\neg T_k} \cup \Delta_{k+1} \ldots \cup \Delta_n$, say $\Delta_{T_{k+1}}$, and those which do not, say $\Delta_{\neg T_{k+1}}$. Then $\Delta'_{k+1} = \Delta_{\neg T_k} \cup \Delta_{T_{k+1}}$ and it remains to partition $\Delta_{\neg T_{k+1}} \cup \Delta_{k+2} \ldots \cup \Delta_n$. Proceeding in this way, the Z-partition of Δ' is formed such that for any default, $r' \in \Delta_{j \geq k}$, it holds that either $r' \in \Delta'_j$ or $r' \in \Delta'_{j+1}$, as required. □

Thus the default is incorporated into a particular set and the members of that and higher sets experience a "ripple" effect and may be shunted up by one position; indeed, a new partition set may be created.

This has unfortunate effects for the LEX-ordering since it implies that some defaults—the ones which get shunted up—are made semantically stronger. Thus adding a derived default leads to significant changes in the LEX-ordering and some

old beliefs are sure to be retracted, causing the failure of CM. This may seem a little confusing since in section 6.1 it was claimed that LEX is a form of ME-entailment and ME can satisfy CM but LEX always fails to; that is, some defaults which belong to the consequence relation corresponding to the canonical ME-ranking of the original set may not appear in that corresponding to the enlarged set. This confusion is easily resolved, however, when one realises that it is not possible to add a default under the LEX system without changing the effective strengths of both itself and other defaults. Thus it is impossible to add a default with the same "strength" with which it was LEX-entailed. For this reason, LEX cannot satisfy CM[5].

This section has shown that system Z^+ satisfies CM, ME both satisfies it and allows for alternative interpretations of the enlarged set (i.e., a choice of which defaults may be redundant), and LEX is bound to falsify it.

6.4 Summary

This chapter has seen a comparison between the ME approach and two other systems of default reasoning. The results are twofold. Firstly, connections between the systems have been identified; for LEX this turns out to be a direct connection at the semantic level, whereas for system Z^+ it is only a superficial, mechanical connection. The second result concerns the behaviour of all systems when new beliefs are added. While Z^+ and ME behave reasonably, LEX cannot incorporate new beliefs without changes which result in previous beliefs being retracted.

[5]This point was originally noted by Lehmann (1995).

Chapter 7

Constructing and testing default knowledge bases

This chapter describes how the ME approach may be used to assist translation of default knowledge into a set of defaults, or knowledge base (KB), and gives a guide for using a software implementation of a default reasoning system, called DRS, which is available to create and test default knowledge bases[1]. A brief discussion on the complexity limitations of the implementation is also given.

7.1 Creating a default knowledge base

Using the ME approach to default reasoning necessarily involves coming to some quite specific conclusions, mainly because the result is an ME-ranking which 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 (Geffner 1992, Bacchus *et al.* 1996). 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 KB designers since any significant deviations from the conclusions obtained using ME imply that extra, or different, constraints exist. This use of ME has been suggested by Jaynes for finding physical constraints (Jaynes 1979), but is equally valid when applied to the more abstract problem of eliciting default knowledge.

There have already been several examples of this process in the preceding chapters. For example, Reiter and Criscuolo rejected the ME-consequence that "typically high school dropouts are employed" (see page 80) but admitted that

[1]The program is available at: `http://www2.elec.qmw.ac.uk/~rach/drs.html`.

they wished to remain "agnostic" on this point. This amounts to putting an extra constraint on the problem, which was not represented in the original default set.

Similarly, Touretzky *et al.* were reluctant to concede that Marine chaplains were not beer drinkers under any circumstances (see page 89) but added that this is because Marines may well be much heavier drinkers than men in general (Touretzky, Horty, & Thomason 1987). Again, this extra knowledge was not present in the original problem but by adding it as a new default, it is possible to obtain the ME-consequence that Marine chaplains may be beer drinkers after all.

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 KB designer 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 KB from some background information illustrates how the ME approach can be put to work in practice. The example is taken from Brewka (1989):

> *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 indicates that it is. There may be situations in which "unless" means only "if [something] happens to be the case as well". So the KB designer must be aware of those of his intuitions which relate to the semantics of everyday language and need to be represented explicitly. Given this point, the following set seems to represent the information in a reasonable way:

$$\Delta = \{\texttt{True} \Rightarrow m, s \Rightarrow \neg m, c \Rightarrow s, c \Rightarrow m, v \Rightarrow \neg m\}$$

with the symbols standing for m meeting, s sick, c cold, and v vacation, and the strengths of all defaults being equal, initially.

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 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 KB designer must make and this is where using the ME approach can help.

In fact, given Δ, the default $c \wedge v \Rightarrow \neg m$ is an ME-consequence; but it is not an uncontroversial 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 $c \Rightarrow m$.

It is therefore up to the KB designer to decide whether the default set is sufficient as it stands, so that altering the strengths, or adding extra defaults to the KB, might lead to a different result, 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 KB designer, 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 examined for the benchmark examples, this will only produce the "correct" answers if the KB designer has correctly encoded his semantic intuitions about the propositions. The ME approach can assist the KB designer in clarifying his intuitions because it treats all defaults equally, giving unbiased conclusions enabling the designer to determine both the nature and extent of his own biases.

7.2 How to use DRS

DRS is a program which implements all the default reasoning systems described in chapters 3 and 4. The system can be used for querying the examples of default knowledge bases given in chapter 5 and elsewhere. This section can be treated as the user manual for the program and describes how it can be used to create and test default knowledge bases.

Figure 7.1 shows the graphical user interface for DRS. The user can either select a pre-set benchmark default set, e.g., Penguins, or input his own under the option User-defined. The defaults are displayed in the left-hand panel which is editable. A default is made up of two propositional formulæ connected by a default connective, `=>`. An optional integer strength attribute may be attached to the default connective by enclosing it in square brackets, e.g., `a =>[3] b`. A formula is either a simple proposition, or a complex expression made using simple propositions connected using the logical connectives, `~`, `^`, `v`, `->`, and brackets, '`)`' and

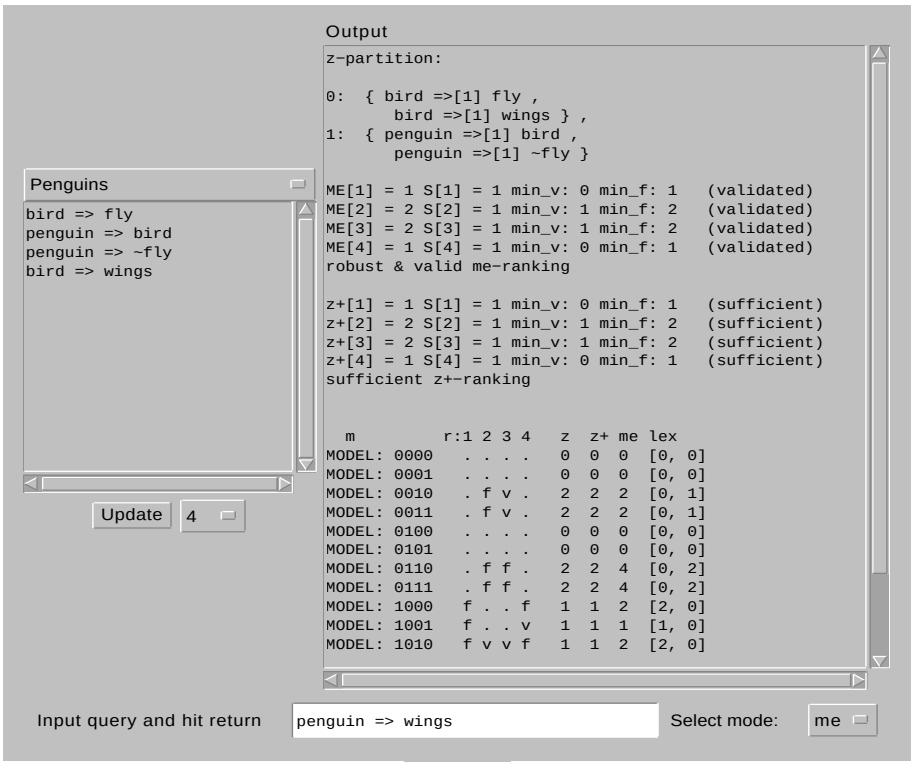

Figure 7.1: User interface of DRS.

'(', according to the syntactic rules of classical logic. A simple proposition is either a single letter or a string of letters (excluding the letter 'v' but including '_'). Examples of valid default syntax are:

```
bird => fly
bears => like_honey ^ inhabit_woods
chaplain =>[3] ~beer_drinker
```

The following are not valid:

`bird => [3] fly`	*space between* `=>` *and* `[3]`
`bears => live_in_the_woods`	*proposition contains the letter* `v`
`chaplains => v ~beer_drinker`	*logical syntax error*

When the program is first loaded, the default knowledge base (KB) is set to the benchmark default set Penguins (see section 5.3). The KB can be updated at any time by clicking the Update button. If the KB is parsed successfully, the output appears in the Output panel on the right-hand side of the screen. The output is composed of the Z-partition, the ME-ranks for each default along with a description of whether the computed ME-ranking is robust or contains redundancy, the Z^+-ranks for each default along with whether the Z^+-ranking is sufficient or not, and,

finally, a grid which displays all models of the underlying language and whether they verify or falsify any default plus their Z, Z^+ and ME ranks, and their LEX-tuples. Since the Output panel is of limited size, the grid is only displayed when the language has fewer than 7 propositional atoms. The number of atoms can be changed by adjusting the numeric choice button next to the Update button. If, for some reason, the KB could not be parsed, an error message appears in the Output panel. The most common reasons for failure are that the vocabulary bounds have been exceeded (the numeric choice button needs to be adjusted), or that one of the defaults contains a syntax error.

Once the KB has been parsed and all ranks have been computed, it is ready to be queried. The user selects which mode of entailment is required by adjusting the Select mode choice button to the bottom right-hand side of the screen. The options are:

P	P-entailment
Z	Z-entailment
LEX	LEX-entailment
ME	ME-entailment
Z^+	Z^+-entailment

For the last two mode options, the user can adjust the strengths of defaults by optionally adding `[x]` after the default symbol `=>`, where `x` is a positive integer.

To query the KB, the user inputs a default in the Input query panel and presses return. The result of the query pops up in a new window. For example, the answer to the query given in the figure would be `"penguin => wings is me-entailed to degree 1"`. The degree of entailment is given only in modes ME and Z^+. If the query contains a syntax error, or if the vocabulary bounds are exceeded, an error message will appear in the results window. The user must dismiss the results window by clicking the OK button before the KB can deal with another query.

The user may add, delete or edit defaults and their strengths by editing the text in the left-hand panel and then clicking the Update button. To change the KB to another benchmark example, or to the (empty) User-defined option, the user adjusts the choice button above the panel containing the defaults.

7.3 Complexity of DRS

DRS was designed and implemented to enable the benchmark examples to be analysed in greater depth than is feasible manually, and to aid in the development and

Mode	Parsing	Querying
P	D^3M^2	D^3M^2
Z	D^3M^2	M^2
LEX	D^3M^2	M^2
ME	D^3M^2	M^2
Z$^+$	D^3M^2	M^2

Table 7.1: Differences in complexity for DRS implementation.

testing of the ME algorithm. Since the implementation is semantical, i.e., is based on enumerating the models of the language, it works only for small KBs of less than 20 defaults, and for small propositional languages of less than 16 propositional atoms.

For this implementation, there is little difference in the relative complexity of the systems. If the number of defaults is D and the number of models of the underlying language is M, then the complexity, in terms of primitive operations, is given in table 7.1. As the table shows, parsing the KB is the most computationally expensive task for the program. Creating the Z-partition takes D^3M^2 operations, and this is necessary for testing the consistency of the KB and before the Z-ranks and LEX-tuples can be computed. Computing the Z$^+$-ranks and ME-ranks is of the same order of complexity as creating the Z-partition. Testing for entailment is relatively less complex for all the ranked-based systems: the ranks are computed at parse time and queries take just M^2 operations. Testing for P-entailment, however, requires another consistency test and therefore takes D^3M^2 operations, as before. The differences in the complexity of parsing the KB and of querying it are easily seen when working with larger languages. For example, with 13 propositional atoms and 4 defaults, parsing the KB takes approximately 60 seconds while queries can be answered in just one second.

Most of the default systems described in this thesis suffer from severe intractibility problems, a common problem for systems of nonmonotonic reasoning (Ben-Eliyahu-Zohary & Palopoli 1997). However, this implementation is naive, being semantically based, and represents the worst case scenario. Systems P, Z and Z$^+$ can all be implemented using propositional satisfiability tests rather than enumeration of models so that efficient implementations are feasible (though none are known to this author). For these systems, algorithms exist which are polynomial in the number of propositional satisfiability tests required (Pearl 1990,

Goldszmidt & Pearl 1996), so that restricting defaults to be Horn clauses can lead to realistic implementations (Dowling & Gallier 1984). Since both LEX- and ME-entailment require the enumeration of models of the underlying language, practical implementations are currently a long way off, although recent advances in inconsistency handling offer interesting possibilities for LEX-entailment (Grégoire 1999).

Chapter 8

Conclusion

This thesis has argued that using the ε-semantics for defaults, augmented to allow priorities between defaults to be represented, and extended using the principle of maximum entropy, provides the most acceptable rational consequence relation admissible with respect to a set of defaults. In view of this, the ME approach provides a general theory of default reasoning which can be used as the basis for understanding default inference and patterns of commonsense reasoning. This final chapter summarises the arguments supporting the thesis and looks at future uses of the ME approach.

8.1 Review of thesis

The main argument in support of the thesis is that selecting a distribution by maximising entropy is the only consistent method of inference under uncertainty (Paris & Vencovská 1990). This is because it maximises the uncertainty contained in a distribution and hence leads to the one containing the least bias or which is the least committed while still implying the original information. By using this method to obtain more information from a distribution, one can be confident that one is not making any unnecessary assumptions which may subsequently turn out to be false. In fact, if the inferences do turn out to be incorrect it almost certainly implies that some unspecified factors or constraints exist and this method can be used to establish what these might be.

This thesis has used the ε-semantics for defaults which sanctions inferences accepted as core behaviour for any reasonable nonmonotonic system (Geffner 1992, Pearl 1990). The ε-semantics, and its translation into ranking functions, is ideally suited to the application of maximum entropy, since it equates defaults with constraints on probability distributions. However, one must be careful to be explicit about how these constraints correspond to defaults, since this will greatly affect the

conclusions obtained. The original application of ME to the ε-semantics by Goldszmidt (1992) was inadequate in certain respects exactly because the constraints he used to define the meaning of a set of defaults were too inflexible. His definition led to just one ME-ranking for each set of defaults and, moreover, his algorithm could not compute this in all cases. He also did not address, nor interpret, the problem of multiple solutions to the ME-ranking constraint equations. Thus the ME-ranking which was derived by Goldszmidt is not identical with that proposed in this thesis, although the results coincide for the cases in which his approach was successful. By reassessing the assumptions on which the approach is based, this revised application of ME to default reasoning has produced a system of default reasoning which allows far greater expressiveness in representing default knowledge, mainly through the use of variable strengths, as well as providing an algorithm which will compute the solution in all meaningful cases.

By using the ε-semantics, one is effectively examining what conclusions can be drawn from one's default knowledge by taking the assumptions it represents to the extreme. The extension using ME requires only that one also commit oneself to the relative orders of magnitude with which defaults hold. However, this is no great extra commitment, since making only order of magnitude judgements means that all defaults can be treated equally favourably by assuming them all to have the same strength. In this way, minor differences in default strengths, i.e., differences within the same order of magnitude, need not affect default conclusions. One of the successes of the ME approach is that, making only these order of magnitude commitments for the strengths of defaults, usually results in a unique order of magnitude description, i.e., ranking function, over all possible worlds and hence a unique rational consequence relation. Although in some cases it turns out that there are multiple solutions for a given strength assignment, this is relatively rare. The cause is a truly ambiguous set of defaults for which there are several candidates for redundancy; not only will this rarely occur in practice but it is also unlikely that a knowledge base designer would require such a situation to be represented.

Despite the simplicity of the assumption made to produce this extension to the ε-semantics—basically a principle of indifference—it leads to a consequence relation which satisfies all the requirements which have been postulated as necessary for a general theory of default reasoning, as well as some more abstract meta-properties of reasoning systems such as cautious monotonicity. More interestingly, this has been accomplished with no reference to those requirements in its design,

with the exception of the introduction of variable strengths to represent priorities between defaults. In contrast, practically every known system which has been designed to perform default reasoning including logical extensions like default logic and circumscription, as well as consequence relations like lexicographic entailment and system Z^+, fails to satisfy one or other of these requirements. Perhaps this is an indication of the inconsistency of making arbitrary design decisions which may satisfy one requirement at the expense of violating another. The fact that under the ME approach all the intuitive solutions to the benchmark examples can be reproduced suggests that the default conclusions obtained from ME-rankings might be considered as benchmarks themselves, that is, *perhaps what underlies default intuitions—i.e., common sense—is exactly some form of* ME*-inference*. This point of view has recently been put forward by Paris (1998).

In comparison with other systems which produce rational consequence relations, it has been seen that the ME approach subsumes the only other system to satisfy the general requirements of default reasoning (up to and including exceptional inheritance). It was also shown to be semantically more reasonable than system Z^+, both because it takes account of all non-redundant information and is therefore far more expressive, and because it assigns more meaningful values to defaults themselves in terms of their ranks. As for the more controversial meta-level properties of belief revision, the ME approach provides some insight. While system Z^+ satisfies cautious monotonicity, almost trivially, and the LEX system fails to satisfy it, the ME approach not only allows for the possibility of incorporating defaults into one's knowledge base without any change taking place, but also allows for the possibility that this new knowledge may affect the relevance of the defaults used to derive it. This means that some new defaults may "explain away" old ones making them redundant. This situation corresponds to strength assignments which lead to multiple ME-rankings indicating ambiguity in the default knowledge caused by the abstraction to orders of magnitude. It also goes some way to explaining the controversy surrounding the acceptance of any blanket postulates for belief revision, and justifies a skeptical position.

Taken as a whole, this method of default reasoning appears to be extremely successful, since it passes all benchmark tests and behaves reasonably at the meta-level. Given the "inevitability of maximum entropy" (Paris & Vencovská 1990), one may conjecture that it also ought to satisfy any future requirements. Furthermore, the ME approach has a sound justification, being based on probability theory and a principle of indifference. This concludes the argument in support of the thesis.

8.2 Future uses of the ME approach

This section looks at theoretical and practical problems that need resolving for the ME approach and how it could be used in future research.

From a theoretical point of view, the ME approach is almost complete in that the ME algorithm can be used to compute an ME-ranking for any set of variable strength defaults. There are two main theoretical problems. The first involves the analytic assumption which is used to derive the main equations which constrain the ME-ranking. In particular, is it the case that certain strength assignments over convergence functions uniquely determine the asymptotic abstraction to the corresponding ME-distribution? While it is clear that the derivation is valid in one direction—from real convergence functions to the ME-ranking—it is unclear if solutions to these equations are the only abstraction to potential ME-distributions. This will only be established by a more detailed mathematical analysis, far beyond the intended scope of this thesis. The second problem involves isolating the cases when multiple solutions occur, i.e., adjusting the robustness condition so that it was both necessary and sufficient. The current condition is inadequate for two reasons. Firstly, there are obvious cases when a unique solution exists but the robustness condition fails. For example, whenever a redundant default is present, that is, one with an insufficient strength, the condition will clearly fail, but it is still very likely that the ME-ranking produced is unique. This case may be easily resolved by applying the robustness condition to only those defaults which are active. Secondly, there are cases, for instance the example given on page 74, which have multiple solutions for the ME-ranks over defaults but the ME-ranking itself is unique. Thus the main remaining theoretical problem with the ME approach is to characterise multiple and unique solutions. However, such cases are rare and, from the point of view of the knowledge engineer, unlikely to be useful models of default knowledge, so the impetus to resolve these problems is rather lacking.

From a practical perspective, the challenges are both very hard and potentially much more rewarding. The complexity of most model-based approaches is highly discouraging and it is very likely that severe restrictions, or interesting approximations, will be required to make this approach at all feasible. In fact, one restriction, i.e., insisting on a language of fewer than 15 atoms, has already resulted in a successful implementation as detailed in the penultimate chapter. However, to be of realistic practical use, much larger languages will be necessary. One application could be to use defaults as condition-action rules and allow the ME approach to resolve what actions should be taken in exceptional situations. There may be do-

mains in which the possible states are too large to be able to prescribe behaviour exactly, but small enough to be manageable using limited propositional languages. Attempting to use the ME approach on a real problem of this kind would certainly provide new and interesting avenues of research.

The example given in chapter 7 indicates that even an apparently simple problem needs to be encoded carefully as slight differences in representing the background knowledge can lead to different solutions. This suggests a use for the ME approach in default knowledge base design. Encoding knowledge in the form of defaults and applying the ME approach may help to elicit hidden information from the user which perhaps he was unaware of using. By using the ME-solution for a set of defaults, one can test whether the defaults which one expects to entail some conclusion actually do, and whether any other conclusions which were unexpected are in fact entailed. This may help one to decide exactly which defaults should be contained in one's knowledge base, and whether they require different strengths. Indeed, it seems very important to be able to distinguish between semantic and structural intuitions and using ME can help to sort out these differences.

Finally, the ME approach can be used as a benchmark itself from which to compare and assess other related default systems as already performed in chapter 6. While it is relatively easy to compare the ME-ranking with other systems which produce rational consequence relations, it may also be possible to compare it with preferential relations. One could compare, for example, what were called "uncontroversial" ME-consequences, that is, default conclusions which hold for all strength assignments with the conclusions which are obtained from preferential systems. Some systems, while not producing rational consequence relations, are still very closely connected to them, for example, Geffner's conditional entailment (Geffner 1992) and Benferhat *et al.*'s LCD consequence (Benferhat, Saffioti, & Smets 1995). Any conclusions which are not uncontroversial ME-consequences imply that the system is in some way incorporating additional assumptions, but it is not always obvious what the effect of these might be. Such comparisons may reveal the underlying assumptions or implications of these systems which are often difficult to establish by looking at the systems in isolation.

Bibliography

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

Antoniou, G. 1997. *Nonmonotonic Reasoning*. Cambridge, MA: MIT Press.

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

Beach, L. R., and Braun, G. P. 1994. Laboratory studies of subjective probability: a status report. In Wright, G., and Ayton, P., eds., *Subjective Probability*, 107–127. Wiley.

Ben-Eliyahu-Zohary, R., and Palopoli, L. 1997. Reasoning with minimal models: efficient algorithms and applications. *Artificial Intelligence* 96:421–449.

Ben-Eliyahu, R. 1990. *NP-complete problems in optimal horn clause satisfiability*. Technical report R-158, Cognitive Systems Laboratory, UCLA, Los Angeles.

Benferhat, S., Cayrol, C., Dubois, D., Lang, J., and Prade, H. 1993. Inconsistency management and prioritzed syntax-based entailment. In Bajcsy, R., ed., *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 640–645. Morgan Kaufmann.

Benferhat, S., Saffioti, A., and Smets, P. 1995. Belief functions and default reasoning. In *Uncertainty in Artificial Intelligence*, 19–26.

Bourne, R. A., and Parsons, S. 1998. Propagating probabilities in System P. In *Proceedings of the 11th Florida Artificial Intelligence Research Symposium*, 440–445.

Bourne, R. A., and Parsons, S. 1999a. An algorithm for computing the maximum entropy ranking for variable strength defaults. In *Proceedings of the 6th Bar-Ilan Symposium on the Foundations of Artificial Intelligence*.

Bourne, R. A., and Parsons, S. 1999b. Connecting lexicographic with maximum entropy entailment. In Hunter, A., and Parsons, S., eds., *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (LNAI 1638)*, 80–91. Springer.

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

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

Brewka, G. 1994. Reasoning about priorities in default logic. In *AAAI*, 358–363.

Buck, B., and Macaulay, V. A., eds. 1991. *Maximum Entropy in Action*. Oxford: Clarendon Press.

Cheeseman, P. 1983. A method for computing generalized bayesian probability values for expert systems. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 198–202.

Clark, L. 1978. Negation as failure. In Gallaire, H., and Minker, J., eds., *Logics and Data Bases*, 293–322. New York: Plenum.

Darwiche, A., and Ginsberg, M. 1992. A symbolic generalization of probability theory. In *AAAI*, 622–627.

Darwiche, A., and Goldszmidt, M. 1994. On the relation between kappa calculus and probabilistic reasoning. In *Uncertainty in Artificial Intelligence*, 145–153.

Darwiche, A., and Pearl, J. 1997. On the logic of iterated belief revision. *Artificial Intelligence* 89:1–29.

Darwiche, A. 1992. *A symbolic generalization of probability theory*. PhD thesis, Stanford University, CA.

Darwiche, A. 1994. *CNETS: A computational environment for generalized causal networks*. Technical memorandum, Rockwell International, Palo Alto Laboratory.

De Finetti, B. 1974. *Theory of probability : a critical introductory treatment*. Wiley.

Delgrande, J. P., Schaub, T., and Jackson, W. K. 1994. Alternative approaches to default logic. *Artificial Intelligence* 70:167–237.

Dowling, W., and Gallier, J. 1984. Linear-time algorithms for testing the satisfiability of propositional horn formulæ. *Journal of Logic Programming* 3:267–284.

Doyle, J. 1990. Methodological simplicity in expert system construction: The case for judgments and reasoned assumptions. In Shafer, G., and Pearl, J., eds., *Readings in Uncertainty Reasoning*, 689–693. San Mateo, CA: Morgan Kaufmann.

Dubois, D., and Prade, H. 1988. *Possibility Theory: an Approach to Computerized Processing of Uncertainty*. NY: Plenum Press.

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

Gärdenfors, P. 1988. *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.

Garside, G. R., and Rhodes, P. C. 1996. Computing marginal probabilities in causal multiway trees given incomplete information. *Knowledge-Based Systems* 9:315–327.

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

Geffner, H. 1996. A formal framework for causal modeling and argumentation. In *International Conference on Formal and Applied Practical Reasoning*, 208–222.

Ginsberg, M. 1986. Counterfactuals. *Artificial Intelligence* 30:35–79.

Goldszmidt, M., and Pearl, J. 1990. On the relation between rational closure and system Z. In *Third International Workshop on Nonmonotonic Reasoning*, 130–140. South Lake Tahoe: (UCLA Technical Report R-139).

Goldszmidt, M., and Pearl, J. 1992. Ranked-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. In *Principles of Knowledge Representation and Reasoning*, 661–672.

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

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

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

Grégoire, E. 1999. Handling inconsistency efficiently in the incremental construction of stratified belief bases. In Hunter, A., and Parsons, S., eds., *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (LNAI 1638)*, 168–178. Springer.

Grove, A. J., , Halpern, J. Y., and Koller, D. 1994. Random worlds and maximum entropy. *Journal of Artificial Intelligence Research* 2:33–88.

Halpern, J. Y., and Koller, D. 1995. Representation dependence in probabilistic inference. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1853–1860.

Henrion, M., Provan, G., Del Favero, B., and Saunders, G. 1994. An experimental comparison of numerical and qualitative probabilistic reasoning. In *Uncertainty in Artificial Intelligence*, 319–326.

Holmes, D. E., and Rhodes, P. C. 1998. Reasoning with incomplete information in a multivalued multiway causal tree using the maximum entropy formalism. *International Journal of Intelligent Systems* 13:841–858.

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

Hunter, D. 1989. Causality and maximum entropy updating. *International Journal of Approximate Reasoning* 3:87–114.

Jaeger, M. 1996. Representation independence of nonmonotonic inference relations. In *Proceedings of the Sixth International Conference on the Principles of Knowledge Representation and Reasoning*.

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

Jeffrey, R. C. 1965. *The logic of decision*. Chicago: University Press.

Katsuno, H., and Mendelzon, A. O. 1991. Propositional knowledge base revision and minimal change. *Artificial Intelligence* 52:263–294.

Kern-Isberner, G. 1997. A logically sound method for uncertain reasoning with quantified conditionals. In Gabbay, D. M., Kruse, R., Nonnengart, A., and Ohlbach, H. J., eds., *Qualitative and Quantitative Practical Reasoning (LNAI 1244)*, 365–379. Berlin: Springer.

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

Kowalski, R. A., and Sergot, M. J. 1986. A logic-based calculus of events. *New Generation Computing* 4:67–95.

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

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

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

Lifschitz, V. 1987. Pointwise circumscription. In Ginsberg, M., ed., *Readings in nonmonotonic reasoning*, 179–193. San Mateo: Morgan Kaufmann.

Lifschitz, V. 1988. Benchmark problems for formal non-monotonic reasoning, version 2.00. In Reinfrank, M., de Kleer, J., Ginsberg, M. L., and Sandewall, E., eds., *Non-monotonic reasoning (LNAI 346)*, 202–219. Berlin: Springer.

Lukasiewicz, T., and Kern-Isberner, G. 1999. Probabilistic logic programming under maximum entropy. In Hunter, A., and Parsons, S., eds., *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (LNAI 1638)*, 279–292. Springer.

Lukaszewicz, W. 1988. Considerations on default logic: an alternative approach. *Computational Intelligence* 4:1–16.

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

Makinson, D. 1988. General theory of cumulative inference. In Reinfrank, M., de Kleer, J., Ginsberg, M. L., and Sandewall, E., eds., *Non-monotonic reasoning (LNAI 346)*, 1–18. Berlin: Springer.

McCarthy, J. 1968. Situations, actions and causal laws. In Minksy, M., ed., *Semantic Information Processing*, 410–417. Cambridge, MA: MIT Press.

McCarthy, J. 1980. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence* 13:27–39.

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

Mikitiuk, A. 1996. Semi-representability of default theories in rational default logic. In *Lecture Notes in Artificial Intelligence 1126*, 192–207.

Neapolitan, R. E. 1990. *Probabilistic reasoning in expert systems : theory and algorithms*. New York, NY: Wiley.

Neufeld, E. 1991. Notes on "a clash of intuitions". *Artificial Intelligence* 48:225–240.

Neumann, I. 1996. Graded inheritance nets for knowledge representation. In *International Conference on Formal and Applied Practical Reasoning*, 436–448.

Nute, D. 1980. *Topics in conditional logic*. Dordrecht, Netherlands: Reidel.

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

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

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

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

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

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

Pradhan, M., Henrion, M., Provan, G., Del Favero, B., and Huang, K. 1996. The sensitivity of belief networks to imprecise probabilities: An experimental investigation. *Artificial Intelligence* 85:363–397.

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

Reiter, R. 1978. On closed-world data bases. In Gallaire, H., and Minker, J., eds., *Logics and Data Bases*, 55–76. New York: Plenum.

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

Rhodes, P. C., and Garside, G. R. 1995. Use of maximum-entropy method as a methodology for probabilistic reasoning. *Knowledge-Based Systems* 8:249–258.

Rhodes, P. C., and Garside, G. R. 1998. Computing marginal probabilities in causal inverted binary trees given incomplete information. *Knowledge-Based Systems* 10:213–224.

Sandewall, E. 1986. Nonmonotonic inference rules for multiple inheritance with exceptions. *Proceedings of the IEEE* 74:1345–1353.

Shafer, G. 1976. *A mathematical theory of evidence*. New Jersey: Princeton University Press.

Shannon, C. E., and Weaver, W. 1949. *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

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

Shoham, Y. 1988. *Reasoning about change: Time and causation from the standpoint of Artificial Intelligence*. Cambridge, MA: MIT Press.

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

Shore, J. E. 1986. Relative entropy, probabilistic inference, and ai. In *Uncertainty in Artificial Intelligence*, 211–215.

Smets, P. 1988. Belief functions. In Smets, P., Mamdani, E. H., Dubois, D., and Prade, H., eds., *Non-Standard Logics for Automated Reasoning*, 253–286. New York: Academic Press.

Spohn, W. 1988. Ordinal conditional functions: a dynamic theory of epistemic states. In Harper, W. L., and Skyrms, B., eds., *Causation in Decision, Belief Change, and Statistics, II*, 105–134. Kluwer Academic Publishers.

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

Spohn, W. 1998. How to understand the foundations of empirical belief in a coherentist way. In *Proceedings of the Aristotelian Society, New Series*, volume 98, 23–40.

Stalnaker, R. 1975. A theory of conditionals. In Sosa, E., ed., *Causation and Conditionals*, 165–179. Oxford University Press.

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

Touretzky, D. S. 1984. Implicit ordering of defaults in inheritance systems. In *AAAI*, 322–325.

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

Tversky, A., and Kahneman, D. 1981. The framing of decisions and the psychology of choice. *Science* 211:453–458.

Weydert, E. 1995. Defaults and infinitesimals: defeasible inference in nonarchimedean entropy maximization. In *Uncertainty in Artificial Intelligence*, 540–547.

Weydert, E. 1996. System J - Revision Entailment. Default reasoning through ranking measure updates. In *International Conference on Formal and Applied Practical Reasoning*, 637–649.

Weydert, E. 1998. 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*.