A Review of Uncertainty Handling Formalisms

Simon Parsons[1] and Anthony Hunter[2]

[1] Department of Electronic Engineering, Queen Mary and Westfield College, University of London, London, E1 4NS, United Kingdom.
[2] Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, United Kingdom.

Abstract. Many different formal techniques, both numerical and symbolic, have been developed over the past two decades for dealing with incomplete and uncertain information. In this paper we review some of the most important of these formalisms, describing how they work, and in what ways they differ from one another. We also consider heterogeneous approaches which incorporate two or more approximate reasoning mechanisms within a single reasoning system. These have been proposed to address limitations in the use of individual formalisms.

1 Introduction

Practical AI systems are constrained to deal with imperfect knowledge, and are thus said to reason approximately under conditions of ignorance. Attempts to deal with ignorance, [45, 91] for example, often attempt to form general taxonomies relating different types and causes of ignorance such as uncertainty, incompleteness, dissonance, ambiguity, and confusion. A taxonomy, taken from Smithson [91], that is perhaps typical, is given in Figure 1. The importance of such taxonomies is not so much that they accurately characterise the nature of ignorance that those who build practical AI systems have to deal with—they are far too open to debate for that—but more that they allow distinctions to be drawn between different types of ignorance. This has motivated the development of a multitude of diverse formalisms each intended to capture a particular nuance of ignorance, each nuance being a particular leaf in Smithson's taxonomy tree. The most important distinction is that made between what Smithson calls uncertainty and absence, though this may be confused by a tendency in the literature to refer to "absence" as "incompleteness". Uncertainty is generally considered to be a subjective measure of the certainty of something and is thus modelled using a numerical value, typically between 0 and 1 with 0 denoting falsity and 1 denoting truth. Absence is the occurrence of missing facts, and is usually dealt with by essentially logical methods. The wide acceptance of the suggestion that uncertainty and absence are essentially different, and must therefore be handled by different techniques has lead to a schism in approximate reasoning between the "symbolic camp" who use logical methods to deal with absence and the "numerical camp" who use quantitative measures to deal with uncertainty.

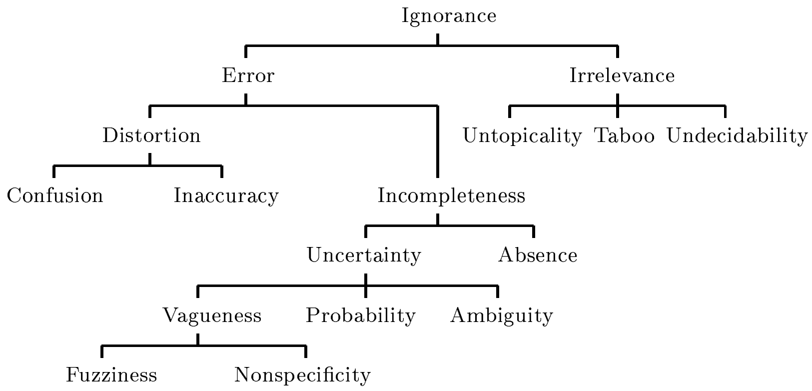

Fig. 1. Smithson's taxonomy of different types of ignorance.

This void between symbolic and numerical techniques, which remained unaddressed for many years as researchers concentrated on the finer technical details of their particular formalism, can be seen as symptomatic of the way in which research into approximate reasoning has been pursued. For many years researchers indulged in ideological slanging matches of almost religious fervour in which the formalism that they championed was compared with its "competitors" and found to exhibit superior performance. Examples of this behaviour abound, particularly notable are [10, 12, 42, 66, 88, 105]. It is only recently that a more moderate *eclectic* view has emerged, [13, 32, 48, 78] for example, which acknowledges that all formalisms are useful for the solution of different problems. A general realisation of the strength of this eclectic position has motivated research both into ways in which different formalisms may be used in combination to solve interesting practical problems, and into establishing the formal differences and similarities between different systems.

In this paper we discuss some of the best established and widely used formalisms from both the symbolic and numerical sides of the great divide. We suggest reasons for the introduction of the more novel techniques, and sketch in the technical differences between the approaches. With this background well established, we then consider work on bringing techniques together.

2 Numerical Approaches

Over the last two decades, numerous formal and informal systems have been introduced for reasoning under conditions of ignorance and uncertainty. The first uncertainty management technique to be introduced was probability theory. This was not only developed many years before the first computer, but was also used in computer decision aids before the advent of Artificial Intelligence as a discipline. Arthur Dempster generalised Bayes' theorem in 1967 [17, 18], though his work remained confined to the field of statistics until Glenn Shafer reformulated the theory and published it as "A Mathematical Theory of Evidence" in 1976 [82]. This body of work, often referred to as Dempster-Shafer theory, has several interpretations including the transferable belief model [87, 90]. Another much studied approach is possibility theory [26, 103] which grew out of work on fuzzy sets [102]. There are numerous other numerical techniques for dealing with uncertainty often developed from pragmatic considerations. These include certainty factors [86], probabilistic logic [63], and belief intervals [21] to name but a few.

2.1 Overview

The methods that we shall consider in the following sections are the main formal theories introduced to handle uncertainty—probability theory, possibility theory, and evidence theory. For theories that have traditionally been seen as rivals, one might expect that they would appear radically different, but this is not so. Indeed, they are remarkably similar, differing largely in subtleties of meaning or application, though this is not entirely surprising since they are intended to do much the same thing.

The basic problem is how to weigh up the degree to which several uncertain events are believed to occur so that the most believed may be unambiguously identified. The basis on which the "belief" is assigned is a contentious issue, though all the theories that we shall consider assume allocation by an assignment function that distributes belief to possible events under consideration. Belief may be distributed on the basis of statistical information [81, 92], physical possibility [103], or purely subjective assessment [12] by an expert or otherwise. The belief assigned is a number between 0 and 1, with 0 being the belief assigned to a fact that is known to be false, and 1 the belief assigned to a fact known to be definitely true. The infinite number of degrees of belief between the limits represent various shades of uncertainty. Now, some formalisms restrict the amount of belief that may be assigned. Both probability theory and evidence theory, which is after all derived from probability theory, limit the total belief that may be assigned by a particular distribution function by constraining the sum of all the beliefs to be 1. This may be interpreted as meaning that one particular observer cannot believe in a set of uncertain events more than she would have believed in a particular event of total certainty. There is no such restriction on a possibility distribution, since one may conceive of several alternative events that are perfectly possible, and so have a possibility of 1. Probability theory, unlike the other theories, also introduces a restriction on the belief that may be applied to a hypothesis based on the belief assigned to its negation. If we have an event A, then

$$\Pr(A) = 1 - \Pr(\neg A)$$

Given the result of a belief distribution, we are interested in how the assigned beliefs may be manipulated. Given our belief in two events, what is our belief in either of them occurring (our belief in their union), and what is our belief that both will occur (our belief in their intersection)? More importantly perhaps, especially for artificial intelligence applications where we often wish to assess the combined belief that results from several different pieces of information, we are interested in combining the effects of two or more belief distributions over the same set of hypotheses. Each distribution will, in general, assign different beliefs to a given hypothesis, and we require some means of assessing a final belief that takes account of all the different assignations. The way in which this is done is based upon the interpretation that the theory gives to the belief it assigns, and thus it is not surprising that each theory should "pool the evidence" in a different way.

2.2 Probability theory

Probability theory has existed in one form or another for several hundred years. During this time various alternative formulations have been introduced, and it is now difficult to say where the definitive account may be found. This is in contrast to the other methods described in this paper where the descriptions are drawn from the original paper on the subject. The introduction presented here is drawn from the discussion of probability theory in Lindley's excellent book "Making Decisions" [53]. Lindley asserts that probability theory is built on three axioms or laws that define the behaviour of a probability measure, which may be used as an estimate of the degree to which an uncertain event is likely to occur. The measure may be assessed by reference to a standard, such as the likelihood of drawing a black ball out of an urn containing five black balls and ten red balls. The first law of probability theory is the *convexity* law which states that the probability measure for an event A given information H is such that:

$$0 \leq \Pr(A|H) \leq 1$$

The second law is the *addition* law, which relates the probabilities of two events to the probability of their union. For two exclusive events A and B, that is two events that cannot both occur, we have:

$$\Pr(A \cup B|H) = \Pr(A|H) + \Pr(B|H)$$

which is commonly written

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

without explicit reference to the information H, since the information is the same in all cases. If the events are not exclusive we have, instead:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Furthermore, the sum of the probabilities of a set of mutually exclusive and *exhaustive* events, the latter meaning that they are the only possible events that may occur, are constrained to sum to 1 so that:

$$\Pr(A) + \Pr(\neg A) = 1$$

or, more generally for a set of n such events A_i:

$$\sum_{i=1,\ldots,n} \Pr(A_i) = 1$$

The final law is the multiplication law, which gives us the probability of two events occurring together; the probability of the intersection of A and B:

$$\Pr(A \cap B|H) = \Pr(A|H). \Pr(B|A \cap H)$$

Again this may be written as

$$\Pr(A \cap B) = \Pr(A). \Pr(B|A)$$

without explicit reference to H. Note that $\Pr(A \cap B)$ is often written as $\Pr(A, B)$. The probability measure $\Pr(B|A)$ is the *conditional* probability of B given A, the probability that B will occur, given that A is known to have occurred. From these laws we can derive two further results which are crucial from the point of view of artificial intelligence. The first of these is Jeffrey's rule:

$$\Pr(A) = \sum_{i=1,\ldots,n} \Pr(A|B_i) \Pr(B_i)$$

The second is Bayes' theorem, named after an eighteenth century non-conformist English clergyman. This states that:

$$\Pr(A|B) = \frac{\Pr(B|A). \Pr(A)}{\Pr(B)}$$

and thus gives a means of computing one conditional probability relating two events from another conditional probability.

Under the assumption that the events in which we are interested are mutually exclusive and exhaustive, and following some manipulation, we can obtain a version of Bayes' rule [14] that is suitable for assessing the probability of a hypothesis h_i that is a member of the set $h_1, \ldots, h_n$ given a set of pieces of evidence $e_1, \ldots, e_m$, a set of probabilities of occurrence of the hypotheses $\Pr(h_1), \ldots, \Pr(h_n)$, and a set of conditional probabilities for each piece of evidence given each hypothesis $\Pr(e_1|h_1)$, $\Pr(e_1|h_2), \ldots, \Pr(e_m|h_n)$:

$$\Pr(h_i|e_1, e_2, \ldots, e_m) = \frac{\Pr(e_1|h_i) \Pr(e_2|h_i) \ldots \Pr(e_m|h_i) \Pr(h_i)}{\sum_{j=1,\ldots,n} \Pr(h_j) \Pr(e_1|h_j) \Pr(e_2|h_j) \ldots \Pr(e_m|h_j)}$$

This may be used, say, to reason about the likelihood of a particular disease (h_i), from a set of possible diseases $\{h_1, \ldots, h_n\}$, given a set of recorded symptoms $\{e_1, ..., e_m\}$.

There have been several adaptations of probability theory within the literature of artificial intelligence including the odds-likelihood formulation used by Prospector [28], and the cautious approach adopted by Inferno [74]. Another is

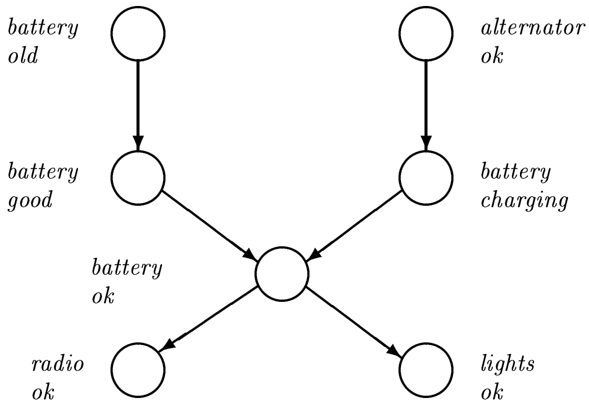

Fig. 2. Part of a probabilistic network for diagnosing faults in a car.

the use of probability theory by Tawfik and Neufeld [93] in their chapter in this volume where they consider the probability of failure of components over time and use this to guide diagnosis. Nilsson [63] provided an interesting variation with his probabilistic logic, an attempt to combine propositional calculus with a numerical uncertainty handling formalism by associating probability measures with logical sentences. Perhaps the most important feature of the formalism is it handles incompletely specified probabilistic models by computing the allowed bounds on the derived consequents.

An increasing important approach to using probability theory in computing is probabilistic networks, also called Bayesian networks or causal networks [39, 67, 68]. By augmenting the use of conditional probabilities with extra structural information, they can be used to represent and reason more efficiently with probabilistic information. In particular they incorporate assumptions about which propositions are independent of other propositions, thereby decreasing the dimensionality and number of conditional probability statements, and simplifying the computations. Essentially, probabilistic networks are a set of nodes with directed arcs (arrows) providing connections between nodes. Every node is connected to another node, but each node is not necessarily connected to every other node. Each node denotes a random variable, which is a variable that can be instantiated with an element from the sample space for the variable. They are used to model situations in which causality, or influence is prevalent, but in which we only have a partial understanding, hence the need to model probabilistically.

As an example, consider the network in Figure 2 which is part of a probabilistic network for diagnosing faults in a car (this example is drawn from [41]). This captures the fact that the age of the battery (the node *battery_old*) has an influence on whether or not the battery is good (*battery_good*), and that whether or not the alternator is good (*alternator_ok*) has an effect on whether or not the battery is charging (*battery_charging*), together the quality of the battery and whether or not the battery is charging affect whether the battery is working (*bat-*

tery_ok), and this has an effect on the radio (*radio_ok*) and the lights (*lights_ok*). All the random variables in this example are either "true" or "false" so that the random variable *battery_old*, can be instantiated with the event *battery_old* meaning that the battery is old, or the event $\neg$*battery_old* meaning that the battery is not old.

Each of the links in the network is quantified by giving the relevant conditional probabilities, which in this case will include:

$$
\begin{aligned}
&\Pr(\mathit{battery\_good}|\mathit{battery\_old}) && = 0.1\\
&\Pr(\mathit{battery\_good}|\neg\mathit{battery\_old}) && = 0.8\\
\\
&\Pr(\mathit{battery\_ok}|\mathit{battery\_good}, \mathit{battery\ charging}) && = 0.9\\
&\Pr(\mathit{battery\_ok}|\mathit{battery\_good}, \neg\mathit{battery\ charging}) && = 0.2\\
&\Pr(\mathit{battery\_ok}|\neg\mathit{battery\_good}, \mathit{battery\ charging}) && = 0.6\\
&\Pr(\mathit{battery\_ok}|\neg\mathit{battery\ good}, \neg\mathit{battery\_charging}) && = 0.05
\end{aligned}
$$

Note that the conditional probabilities reflect the direction of the arrows. Both, broadly speaking, capture a notion of causality (which is why probabilistic networks are also known as "causal networks")—if the battery is old it causes the battery to be less likely to be good, and it is therefore easier to assess $\Pr(\mathit{battery\_good}|\mathit{battery\_old})$ than $\Pr(\mathit{battery\_old}|\mathit{battery\_good})$ though the two probabilities may each be computed from the other using Bayes' theorem.

Now, given the network and the prior probabilities of the battery being old, $\Pr(\mathit{battery\_old})$, and the alternator being ok, $\Pr(\mathit{alternator\_ok})$, it is possible to compute the probability of each state of each random variable in the network (for instance $\Pr(\mathit{battery\_good}) = 0.58$ if $\Pr(\mathit{battery\_old}) = 0.4$) by simple application of Jeffrey's rule. It is also possible to take account of evidence that, for instance, the radio is not ok (which means that $\Pr(\neg\mathit{radio\_ok}) = 1$) and to use Bayes' theorem to revise the probabilities.

Much attention has been given to the problem of propagating probabilities through probabilistic networks efficiently. Pearl [69] provides a comprehensive introduction to the use of probabilistic causal networks, along with an efficient scheme for the propagation of probabilities in singly-connected networks[1] between every that is based on autonomous message passing. Another network-based method that has received wide attention is that of Lauritzen and Spiegelhalter [52] which has been used as the basis of the expert system shell HUGIN [1], and the paper in this volume by Magni *et al.* [56] makes use of a graphical representation similar to that discussed above.

2.3 Evidence theory

Evidence theory is the term commonly used to refer to the body of work carried out by Arthur Dempster [17, 18] and Glenn Shafer [82] to remedy some of what

[1] Singly-connected networks are those in which for every pair of nodes there is at most one path along arcs which joins them. When assessing connectedness, arcs may be traversed both directions, but any arc may only be traversed once.

they saw as the limitations of probability theory, in particular [19] disposing with the "completeness" axiom of probability theory [42]. The theory deals with the so-called *frame of discernment*, the set of base elements $\Theta = \{\theta_1, ..., \theta_n\}$ in which we are interested, and its power set 2^Θ, which is the set of all subsets of the interesting elements. The basis of the measure of uncertainty is a *probability mass function* $m(\cdot)$ that assigns zero mass to the empty set, $m(\emptyset) = 0$, and a value in $[0, 1]$ to each element of 2^Θ, the total mass distributed being 1 so that:

$$\sum_{A \subseteq \Theta} m(A) = 1$$

Since we deal with all possible subsets of the set of all base propositions, rather than the propositions themselves as in probability theory, we can apportion the probability mass exactly as we wish, ignoring assignments to those levels of detail that we know nothing about. This allows us to model ignorance, $m(\Theta)$ being the probability mass we are unable, through lack of knowledge, to assign to any particular subset of Θ. We can define our belief in a subset A of the set of all propositions as the sum of all the probability masses that support its constituents:

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B)$$

and the plausibility of A may be defined as the probability mass not supporting $\neg A$:

$$\mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

which may also be written as:

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\neg A)$$

The interval $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$ can be considered to be a measure of our ignorance about A, and can vary from zero when we have the same degree of belief in A as would be generated by probability theory, to 1 when A has belief 0 and plausibility 1. This means that no mass is assigned to A or any of its subsets, but equally no mass is assigned to $\neg A$.

Evidence is combined by Dempster's rule of combination. This computes the probability mass assigned to $C \subset \Theta$ from the probability mass assigned to A and B where both A and B are also subsets of Θ. If the distribution function assigning probability mass to A is $m_1(\cdot)$ and the function distributing probability mass to B is $m_2(\cdot)$, then the mass assigned to C is defined by:

$$m_{12}(C) = \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A) m_2(B)}$$

where the division *normalises* the new distribution by re-assigning any probability mass which is assigned to the empty set, $\emptyset$, by the combination. To clarify

	$\{Toyota, GM, Chrysler\}$ 0.8	Θ 0.2
$\{Nissan, Toyota\}$ 0.4	$\{Toyota\}$ 0.32	$\{Nissan, Toyota\}$ 0.08
Θ 0.6	$\{Toyota, GM, Chrysler\}$ 0.48	Θ 0.12

Table 1. Applying Dempster's rule.

what is going on here we will consider a simple example of the use of Dempster's rule in combining evidence.

Consider a world [14] with only four car manufacturers, Nissan, Toyota, GM and Chrysler, all trying to break into a new car market. We are interested in who will dominate the market. There are four singleton hypotheses corresponding to the assertions that each of the four manufacturers will dominate the market. Consider the case in which there are two mass functions m_1 and m_2 stemming from the opinions of two independent experts. Now, m_1 assigns 0.4 to $\{Nissan, Toyota\}$, the hypothesis that Japanese manufacturers dominate, and the remaining 0.6 to the set $\{Nissan, Toyota, GM, Chrysler\}$ modelling ignorance about the behaviour of American manufacturers. Similarly, m_2 assigns 0.8 to the set $\{Toyota, GM, Chrysler\}$ and 0.2 to Θ, and Dempster's rule of combination assigns the product of the two belief masses to the intersection of the sets to which they are assigned. Table 1 explains the calculation.

The masses after combination are as follows:

$$\begin{aligned} m_{12}(\{Toyota\}) &= 0.32 \\ m_{12}(\{Nissan, Toyota\}) &= 0.08 \\ m_{12}(\{Toyota, GM, Chrysler\}) &= 0.48 \\ m_{12}(\Theta) &= 0.12 \end{aligned}$$

The belief that a Japanese manufacturer will dominate is computed from the sum of the belief masses of all the subsets of the hypothesis. Thus:

$$\begin{aligned} Bel_{12}(\{Nissan, Toyota\}) &= m_{12}(\{Toyota\}) + m_{12}(\{Nissan, Toyota\}) \\ &\quad + m_{12}(\{Nissan\}) \\ &= 0.32 + 0.08 + 0 \\ &= 0.4 \end{aligned}$$

For this simple example, no normalisation is required.

The problems of the computational complexity of Dempster's rule have been discussed by several authors. Barnett [2] showed that the apparent exponential time requirement of the theory could be reduced to simple polynomial time if the theory was applied to single hypotheses, rather than sets of hypotheses, and the evidence combined in an orderly fashion. Gordon and Shortliffe [37] extended Barnett's approach to compute approximate beliefs in a space of hierarchically organised sets of hypotheses in linear time. This approach was then subsumed by that of Shafer and Logan [84], who provided an exact algorithm for hierarchically organised sets of hypotheses that is also linear in time whilst being slightly more general than that of Gordon and Shortliffe. More recently, Shenoy and Shafer[85] have introduced a method for the efficient propagation of belief functions in networks by means of local computations, and Nic Wilson [100] has proposed a method in which the explicit use of Dempster's rule of combination is avoided. This permits an exact calculation of belief to be performed in worse than polynomial but better than exponential time even when the hypotheses are not hierarchically structured. Wilson has also proposed an approximate calculation, based on a Monte-Carlo simulation, which gives results that are arbitrarily close to the exact solution, and which can be performed in linear time. More recent advances are explored in [61, 101].

The application of evidence theory is the subject of three papers in this volume. Lalmas [50] uses it as a means of deciding which document to retrieve, van Dam [96] uses it to control a radio communication system, and Duncan Wilson [98] considers how to apply it to the classification of faults in automated inspection.

2.4 Possibility theory

A formal theory of possibilities, based on the notion of a fuzzy set [102], was first introduced by Zadeh [103]. However, the concept of using the notion of possibilities as an alternative to probabilities was mooted much earlier. The economist G. L. S. Shackle [81], unhappy with the use of subjective probability for handling uncertainty, proposed an alternative formalism. This formalism was the calculus of *potential surprise* where uncertainty about an event is characterised by a subjective measure of the degree to which the observer in question would be surprised by its occurrence. Potential surprise is clearly linked to the intuitive notion of possibility. If an event is entirely possible, then there is no surprise attached to its occurrence. If an event is wholly impossible, or is believed to be so, then if it occurs it will be accompanied by the maximum degree of surprise. In this section we present a simple overview of Zadeh's theory, demonstrating its similarity to and differences from probability theory, and briefly discuss how possibility theory may be combined with logic.

Firstly we need the concept of a *fuzzy set* [102]. A fuzzy set is a set whose membership is not absolute, but a matter of degree, such as the set of tall people. A fuzzy set F is characterised by a membership function μ_F which specifies the degree to which each object in the universe U is a member of F. One way of considering F is as a fuzzy restriction on X, a variable which takes values in U, in that it acts as an elastic constraint upon the values that may be assigned to X. The assignment of a value u to X has the form

$$X = u : \mu_F(u)$$

where $\mu_F(u)$ is the degree to which the constraint F is satisfied when u is assigned to X. To denote the fact that F is a fuzzy restriction on X we write:

$$R(X) = F$$

Now, the proposition "X is F", which translates into "$R(X) = F$", associates a possibility distribution Π_X with X and this distribution is taken to be equal to $R(X)$:

$$\Pi_X = R(X)$$

Along with this we have a possibility distribution function π_X which is defined to be equal to the membership function of F:

$$\pi_X = \mu_F$$

Thus $\pi_X(u)$, the possibility that $X = u$, is taken to be equal to $\mu_F(u)$. As an example, let U be the set of positive integers, and F be the fuzzy set of small integers. This set is described by the following set of pairs each of the form $(u, \mu_F(u))$:

$$F = \{(1,1), (2,1), (3,0.8), (4,0.6), (5,0.4), (6,0.2)\}$$

Given this, the proposition "X is a small integer" associates the possibility distribution Π_X with X where Π_X is written as a set of pairs $(u, \pi_X(u))$:

$$\Pi_X = \{(1,1), (2,1), (3,0.8), (4,0.6), (5,0.4), (6,0.2)\}$$

Thus, the possibility that X takes the value 3, given that X is a small integer, is 0.8. We can use possibility distributions to define possibility measures. If A is fuzzy subset of U, then the possibility measure $\Pi(A)$ of A is defined by

$$\begin{aligned} Poss(X \text{ is } A) &= \Pi(A) \\ &= \sup_{u \in U} \min\left(\mu_A(u), \pi_X(u)\right) \end{aligned}$$

When A is a strict subset of U, this reduces to:

$$\Pi(A) = \sup_{u \in U} \pi_X(u)$$

Possibility measures clarify the comparison between possibility and probability theory. We can establish that:

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)) \tag{1}$$
$$\Pi(A \cap B) = \min(\Pi(A), \Pi(B)) \tag{2}$$

which contrast with the corresponding results for probability theory[2]. Zadeh stresses the fact that possibility and probability are different concepts with the

[2] The use of maximum and minimum is not compulsory. For further discussion of this point, see [26].

example of Hans' breakfast. Consider the statement "Hans ate X eggs for breakfast" with $X \in \{1, \ldots, 8\}$. We can associate both a possibility distribution (based on our view of the ease with which Hans can eat eggs) and a probability distribution (based on our observations of Hans at breakfast) with X, giving something of the form:

u	1	2	3	4	5	6	7	8
$\Pi_X(u)$	1	1	1	1	0.8	0.6	0.4	0.2
$\Pr_X(u)$	0.1	0.8	0.1	0	0	0	0	0

So that, while it is perfectly possible that Hans can eat three eggs for breakfast, he is unlikely to do so. There is a heuristic connection between possibility and probability, since if some thing is impossible, it is likely to be improbable, but (as the previous example shows) a high degree of possibility does not imply a high degree of probability, nor does a low degree of probability reflect a low degree of possibility. Dubois and Prade [25] point out that a weak theoretical connection exists since for all A,

$$N(A) \leq \Pr(A) \leq \Pi(A)$$

where $N(A)$ is the necessity of A, defined by:

$$N(A) = 1 - \Pi(\neg A)$$

It is possible to extend these ideas to possibility distributions that depend on more than one attribute, and marginal possibility distributions. Explanations of these concepts will be omitted in the interests of saving space (but see [103]), but it should be noted that the kind of graphical structures discussed above in relation to probability theory can be adapted for use with possibility theory as well [35].

Possibility has been applied to reasoning with vague statements [27, 104]. For example, suppose we have the following statement.

If the clothes are dirty then wash them in hot water

Both the concepts "dirty" and "hot" are vague or fuzzy in this context. For a given collection of clothes, we are interested in using this general statement to determine whether to wash them in hot water. In other words, we wish to determine whether for some fuzzy value for dirty, we should derive the instruction to wash the clothes in hot water.

Now, in classical logic we would perform this kind of reasoning using *modus ponens*, a rule for reasoning which formalises the argument that if α is *true* and $\alpha \supset \beta$ is *true*, then β is *true*. For reasoning with fuzzy statements such as the one above about dirty clothes, we need to develop a notion of modus ponens which can handle fuzzy concepts. Generalized modus ponens is such a development [57]. For example, suppose the clothes are "not very dirty", then "not very dirty" does not directly match with "dirty". We need to adapt the statement to allow the data "not very dirty" to apply. This means changing the consequence in some way, perhaps to "warm water". Since "dirty clothes" and "hot water" can be modelled by fuzzy sets, the manipulations can be done on the fuzzy sets. For this, we represent propositions as:

$$X \text{ is } A$$

So for example, "clothes are dirty" is a a proposition, where X is "clothes", and A is "dirty". Generalized modus ponens is then of the following form.

$$\begin{array}{l} X \text{ is } A^* \\ \text{If } X \text{ is } A, \text{ then } Y \text{ is } B \\ \hline Y \text{ is } B^* \end{array}$$

Here, B^* is calculated from the possibility distribution of A^*, and of A given B. The possibility distribution for B provides an upper bound on the possibility distribution for B^*. This calculation decreases the possibility that Y is B^* is *true*, the further A^* is from A. This combination of possibility theory and logic into *possibilistic logic* has been investigated at length by Dubois and Prade [23, 24]. Possibilistic logic is one of the techniques explored by Bigham in his paper in this volume [6], and both the contributions of Ramalho [75] and Saffiotti [79] make use of fuzzy inference of the kind discussed above, while Bosc *et al.* [7] consider the application of fuzzy techniques to databases.

2.5 Other approaches

There are a number of other numerical techniques which, although we do not have space to consider them in any detail, are worth mentioning for their particular historical or theoretical interest. Certainty factors [86], perhaps because of their simplicity and intuitive appeal, have been widely used to handle uncertainty. The certainty factor approach assigns a numerical weight, the certainty factor, to the consequent of every

If ⟨*evidence*⟩ then ⟨*hypothesis*⟩

rule in a rule-based system. The value of the certainty factor, which lies in the interval [-1, 1], is assessed by the domain expert from the degree, between 0 and 1, to which a given piece of evidence causes her belief and disbelief in the hypothesis to be increased. The certainty factor is then the difference of the degree of belief, MB and the degree of disbelief, MD:

$$CF = MB - MD$$

The certainty factors of rules fired during inference are then combined to give an overall certainty for the support given to a particular hypothesis by the known evidence. Recently several people have challenged the validity of the certainty factor model. For instance, Heckerman [40] has shown that the original definition of the model is flawed since the belief in a hypothesis given two pieces of evidence will depend upon the order in which the effect of the pieces of evidence is computed.

Smets has adapted evidence theory as introduced by Dempster and Shafer in two important ways [87,90]. The first was to relax the assumption that all hypotheses have been identified before the evidence is considered. Instead Smets makes an *open-world* assumption that the frame of discernment does not necessarily contain an exhaustive set of hypotheses. Under this assumption there is no normalisation in Dempster's rule of combination since the mass pertaining to the empty set is taken to indicate belief in a hypothesis outside the frame of discernment. The open world assumption requires a modification of the definitions given earlier for the calculation of belief and plausibility from probability mass distributions (which are just called "mass functions" by Smets). Belief is defined as:

$$\mathrm{Bel}(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B)$$

while plausibility is defined as:

$$\begin{aligned}\mathrm{Pl}(A) &= \sum_{B \cap A \neq \emptyset} m(B) \\ &= 1 - (m(\emptyset) + \mathrm{Bel}(\neg A))\end{aligned}$$

Dempster's rule of combination becomes:

$$m_{12}(C) = \sum_{A \cap B = C} m_1(A) m_2(B)$$

Smets' other innovation was to introduce an alternative interpretation of the theory of evidence called the transferable belief model. The transferable belief model rejects any suggestion that the numbers manipulated by the theory are probabilities. Instead they are taken to be pure expressions of belief suitable for reasoning at an abstract *credal* level and are transformed into probabilities at the *pignistic* level when decisions are necessary. In the model, the basic belief mass $m(A)$ in any subset A of a frame of discernment Θ is the amount of belief supporting A, that, due to ignorance, does not support any strict subset of A. If we have new evidence that excludes some of our original hypotheses, and so points to the truth being in $\Theta' \subset \Theta$, then the basic belief mass $m(A)$ now supports $A \subset \Theta'$. Thus the belief originally attributed to A is transfered to that part of A not eliminated by the new evidence, thus giving the system its name.

Another interesting proposal is due to Driankov [21]. In Driankov's system, we have degrees of belief and plausibility, related, as in the original theory of evidence, by:

$$\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\neg A)$$

However the system also allows contradictory beliefs so that it is possible that:

$$\mathrm{Bel}(A) + \mathrm{Bel}(\neg A) > 1$$

These ideas lead to the definition of a calculus of belief intervals, where a belief interval for A is $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$, in which combination is carried out by a family of general functions called triangular norms and conorms [80], and explicit reasoning about the degree to which a proposition is believed and disbelieved is possible.

2.6 Limitations of numerical techniques

As one might expect, none of the systems mentioned in preceding sections is perfect, and there are a number of problems common to all numerical formalisms. The first is perhaps the simplest. When Cohen [14] criticises possibility theory saying:

> "relatively little has been made of the idea of fuzzy sets and possibility theory ... (this) may be because the idea does not improve on any of the difficult methodological problems that beset probability theory, such as the assessment of prior probabilities"

he is restating an argument that has been made time and again, perhaps most tersely by Cheeseman [11] who asked:

> "where are all the numbers coming from?"

Obtaining the "numbers", be they probabilities, possibilities, or mass distributions does seem to be a major problem. Clearly, without good numerical assessments sophisticated computational mechanisms are of little value. It is also true that there are domains in which it is not possible to obtain the kind of strong statistical data necessary to apply probability theory in its "frequentist" interpretation, where the probability of an event is the value to which the ratio of occurrences to non-occurrences converges after a large number of trials. This has been used by many (see for example [32]) to argue against the use of probability theory for dealing with uncertainty. However, the *personalist* and *necessarian* [83] schools of probabilists argue that probabilities may always be obtained, either from rational human reasoning, or because they exist as a measure of the degree to which sets of propositions confirm one another. It seems, then, as though there is no clear cut winner in this argument; the moral appears to be:

> "if you can obtain the numbers to your satisfaction, then use them."

As a final word, it is worth mentioning that it has also been convincingly argued in several places (see for example [9, 71]) that even if the numbers are available, they make little difference to the business of weighing up the evidence. This, however, is a different argument altogether, and we will say no more about it.

A second problem stems from the use of numbers; the interpretation of the results of applying a numerical formalism given the notorious irrationality that human beings exhibit when dealing with numbers [95]. All the techniques generate results as numerical values. These values, however, have been generated in different ways, and thus measure different things, although they are just numbers and may be compared and contrasted by the uninitiated as though they represented the same thing. Indeed, to interpret them correctly, it is perhaps necessary to label them with the type of belief that they measure to prevent a probability of 0.5 being compared unfavourably with a possibility of 0.8. In addition, there is the problem of ranking different solutions. A common argument for including numbers is in order to choose the best of several courses of action that must be differentiated between using uncertain knowledge, and of course numerical results can be used to do this. However, using the ordinal value of the results alone to do this can obscure important information concealed in the ratio of the results; namely how close the second largest value is to the largest. If they are close, but separated by a wide margin from the third, then, rather than choosing the first, it might be profitable to review the criteria upon which the assessment was made in the hope that some telling difference between the alternatives might be found.

Finally, there are the problems associated with computational expense. The massive amount of time needed to apply the full formal methods to realistically large problems was one of the main reasons that such *ad hoc* methods as certainty factors were introduced. Whilst, as outlined in earlier sections, there have been several recent attempts to find computationally efficient methods of calculating the results of applying probability and evidence theories in particular situations, the general problem of inefficiency remains.

3 Symbolic approaches

Nonmonotonic logics were introduced in order to allow programs to deal with incompleteness by exhibiting "commonsense" reasoning, thus avoiding the need to state every possible exception to a general rule. Two key approaches are Reiter's default logic [76] and McCarthy's circumscription [60]. In this section we start with a discussion of the limitations of first order logic as a basis for practical reasoning systems, introducing the notions of retraction, monotonicity and defeasibility. Then we consider the family of default logic in more detail—as it is probably the most developed approach for non-monotonic reasoning. In the subsequent two sections, we consider the logic-based approaches of argumentation and truth maintenance systems.

3.1 Overview

The common motivation behind all of the systems of nonmonotonic logic that we will discuss below is the attempt to devise sound formal mechanisms for reasoning that overcome the limitations of first order logic. At first sight, first order logic seems to be a panacea for all the problems of knowledge representation and deduction for AI systems. This is unfortunately a naive impression, and there are many problems that beset the use of classical logic, especially when attempting to model the kind of "commonsense" reasoning which human beings excel at. Israel [44] credits Minsky with being the first to consider the matter, pointing out that there are two particular properties of first order logic that are at odds with commonsense human behaviour. The first results in the so called *qualification* problem. Say, to take the classic example, we are interested in building a system that reasons about animals and their athletic abilities. One of the facts that we want to encode is the fact that generally birds can fly. Unfortunately, there is no "generally" quantifier in first order logic, so we must approximate this by asserting that all birds fly:

$$\forall x, \mathit{bird}(x) \supset \mathit{flies}(x) \tag{3}$$

This seems fine until we recall that penguins don't fly, and so we have to augment the rule. This may be done in several ways, we will choose to write:

$$\forall x, \mathit{bird}(x) \wedge \neg \mathit{penguin}(x) \supset \mathit{flies}(x) \tag{4}$$

However, this formulation becomes problematic when we want to reason about ostriches, kiwis, and birds whose feet have been set in concrete. For any general rule of this kind, we can think up an arbitrarily large number of exceptions, and it is the provision of a compact means of handling all of these exceptions that is the qualification problem. The second troublesome property is that of monotonicity. In first order logic there is no mechanism for retracting inferences once they have been made, or facts once they have been added to the database. If a sentence S is a logical consequence of a theory A then it is still a consequence of any theory that includes A, such as the theory $A \cup \phi$. This is true even if we have $\phi = \neg S$, though in this latter case adding ϕ leads to inconsistency (since we can derive both S and its negation). Monotonicity is particularly troublesome when, in attempting to solve the qualification problem, we allow systems to make "guesses" about the state of the world which are used in the absence of more detailed information. For instance consider making the assumption that a particular bird, Joe, flies when nothing is known to the contrary. In a monotonic system, when it is learnt that Joe has been nailed to his perch there is no means of retracting the inference that Joe can fly. To solve such problems researchers turned to nonmonotonic logical systems that allow for plausible inferences to be made to defeat the qualification problem, and then allow those inferences to be withdrawn if their falsity becomes apparent.

There are three main ways in which a solution to these problems have been attempted; closed world reasoning, prototypical reasoning, and reasoning about beliefs. These methods may be summarised as follows. Closed world reasoning makes the assumption that all relevant positive knowledge has been explicitly stated. Working on this assumption, systems are permitted to deduce any negative facts that they desire in order to reason about the state of the world. Thus a system reasoning about connecting flights which has no knowledge of a flight between London and Ankara is allowed to deduce that there is no such flight, and is only allowed to postulate the existence of a flight joining London and Paris if such a flight is explicitly recorded in its database. Prototypical, or default, reasoning proceeds from rules relating to typical individuals of some class to make plausible assumptions about particular individuals. If and when specific information about the individual that contravenes the plausible deduction comes to light, the assumptions are retracted. Our example about flying is of this kind. We know that birds fly in general, so that when we hear of a bird Opus we assume that he can fly. Later we learn that Opus is a penguin, and knowing that penguins don't fly allows us to retract our assumption that Opus is capable of flying. Finally, reasoning about beliefs allows a system to make sound deductions based on what it believes to be true. Assuming rationality, the system is allowed to logically deduce facts from what it knows and what it believes to be true, and what it believes to be false, which is everything that it does not believe to be true. Thus a system reasoning about its siblings can deduce that it is an only child because if it wasn't, it would know about a brother or sister.

In addition to the distinction between closed world reasoning, prototypical reasoning, and reasoning about beliefs, there is another distinction between systems of nonmonotonic logic which it is worth making. This is between *brave* systems and *cautious* systems (also known as *credulous* and *skeptical* systems, respectively). Brave systems are those which are prepared to accept any conclusion which they can hypothesise. As a result they typically suffer from being able to derive two contradictory conclusions, both of which they deem to be acceptable but are unable to choose between. Cautious systems on the other hand are only prepared to accept conclusions which cannot be contradicted. As a result if they can hypothesise both ϕ and $\neg\phi$, they conclude neither, even though one must be true.

3.2 Default logic

Default logic, introduced by Reiter in [76], models prototypical reasoning by allowing special inference rules, known as default rules, to be added to a standard first order logic. These rules differ from first order rules of the form:

$$\forall A(x) \supset B(x)$$

in that they include an explicit consistency check that prevents the rule being applied in inappropriate situations and allow the expression of rules such as:

$$\frac{Bird(x) : Flies(x)}{Flies(x)}$$

which is read as "if x is a bird, and it is consistent to believe that x flies, then conclude that x flies". Default rules can be considered as meta-rules that tell us how to complete first order theories that are incompletely specified. Now, a default theory (W, D), is a set of first order axioms W, and a set of default rules D of the form:

$$\frac{\alpha(\tilde{x}) : \beta_1(\tilde{x}) \ldots \beta_m(\tilde{x})}{\gamma(\tilde{x})}$$

Where $\alpha(\tilde{x})$, $\beta_i(\tilde{x})$ and $\gamma(\tilde{x})$ are all formulae whose free variables are among those in $\tilde{x} = x_1, \ldots, x_n$. $\alpha(\tilde{x})$ is termed the precondition or prerequisite, the $\beta_i(\tilde{x})$ are known as the gating facts or justifications, and $\gamma(\tilde{x})$ is called the consequent. Given a set of default rules D and a first order theory W, it is possible to define an extension of the default theory as the closure of W plus a maximal consistent set of consequences of D. It is possible to distinguish several classes of such default rules, some of which have attractive properties such as always having extensions. Chief among these are those with a single justification $\beta(\tilde{x})$ which divide into normal defaults, the set of defaults such that $\beta(\tilde{x}) = \gamma(\tilde{x})$, and semi-normal defaults where $\beta(\tilde{x}) = \gamma(\tilde{x}) \wedge \omega(\tilde{x})$ for some $\omega(\tilde{x})$.

An extension E of a default theory is a minimal set of beliefs that contain W are deductively closed, and maximally consistent with the rules in D. Thus E is an extension for (W, D) if $\Gamma(E) = E$ where for any set of sentences S, $\Gamma(S)$ is a minimal set such that:

$$\begin{aligned} W &\subseteq \Gamma(S) \\ Th(\Gamma(S)) &= \Gamma(S) \end{aligned}$$

where $Th(T)$ is the deductive closure of T, and if D contains:

$$\frac{\alpha(\tilde{x}) : \beta_1(\tilde{x}) \ldots \beta_m(\tilde{x})}{\gamma(\tilde{x})}$$

and both $\alpha(\tilde{x}) \in \Gamma(S)$ and $\neg\beta_i(\tilde{x}) \notin \Gamma(S)$ for all i, then it is the case that $\gamma(x) \in \Gamma(S)$.

Reiter proved some interesting results for normal default theories that include no free variables. Firstly every closed normal default theory has an extension, so something can always be conjectured about such a theory. Secondly, if a closed normal default has two extensions, then the union of these are inconsistent, so that multiple extensions are only generated if the default rules have inconsistent consequents. Finally, Reiter showed that closed normal default theories are *semi-monotonic*. This means that if we have two default theories where the sets of default rules of one are a subset of the default rules of the other, then an extension of the theory with the smaller set of defaults will be a subset of an extension of the other. Thus adding default rules to a theory does not cause its extensions to need revision, instead new default inferences are simply added to the existing extensions (they may of course cause new extensions to arise). There are also some more general results, applicable to all closed default theories, the most important of which are that if such a theory (D, W) has an inconsistent extension, then it is its only extension, and it is inconsistent because W is inconsistent. Thus default rules alone do not generate inconsistent extensions.

Many authors have worked on default logic in the years since it was first introduced. One those whose work is worth considering is Łukaszewicz, who proposed two important extensions to the original formulation. The first of these [54] takes the form of translations between different types of default, in particular to replace the general default:

$$\frac{\alpha(\tilde{x}) : \beta(\tilde{x})}{\gamma(\tilde{x})}$$

by the semi-normal default:

$$\frac{\alpha(\tilde{x}) : \beta(\tilde{x}) \wedge \gamma(\tilde{x})}{\gamma(\tilde{x})}$$

and to replace the semi-normal default:

$$\frac{\alpha(\tilde{x}) : \beta(\tilde{x}) \wedge \gamma(\tilde{x})}{\gamma(\tilde{x})}$$

by the normal default:

$$\frac{\alpha(\tilde{x}) : \beta(\tilde{x}) \wedge \gamma(\tilde{x})}{\beta(\tilde{x}) \wedge \gamma(\tilde{x})}$$

The first is non-controversial, but the second, despite being applicable for a large range of practically occurring defaults, has some rather alarming exceptions [30]. By using both translations sequentially, we can replace the eminently sensible:

$$\frac{\mathit{has\_motive}(x) : \mathit{guilty}(x)}{\mathit{suspect}(x)}$$

by the rather unreasonable:

$$\frac{\mathit{has\_motive}(x) : \mathit{suspect}(x) \wedge \mathit{guilty}(x)}{\mathit{suspect}(x) \wedge \mathit{guilty}(x)}$$

In a further paper, Łukaszewicz [55] generalises default logic, providing an alternative formalisation of an extension, and proving that semi-normal default theories are guaranteed such extensions. He also shows that semi-normal default theories are semi-monotonic, that is monotonic with respect to default rules.

Despite the maturity of the theoretical work on default logic, there are as yet few applications, partly because there has been less attention paid to providing prospective application builders with useful tools for using default logic than has been paid to providing tools for using approaches such as probability. However this situation is beginning to change. This volume includes a paper by Nicolas and Schaub [62] which describes a system on which to build default logic applications, while Brazier *et al.* [8] have applied default logic to a problem from ecology.

3.3 Argumentation

Argumentation is the process by which arguments are constructed and compared. Following Toulmin [94], an argument can be structured so that from facts a qualified claim (a conclusion) can be argued (inferred) if and only if:

1. there is some warrant (some further assumptions) that can be used with the facts to logically derive the claim, and
2. there is no other argument that would act as a rebuttal of the claim (a counter-argument).

Argumentation can be further developed with the notion of an undercutting argument, which is an argument that acts as a rebuttal for one of the assumptions of an argument.

An argument can be modelled by a pair (Φ, α), where Φ is a set of formulae, and α is a formula derived as a conclusion from the assumptions Φ. These assumptions are also known as the grounds of the argument. For an argument (Φ, α), a rebutting argument is an argument $(\Psi, \neg\alpha)$, and an undercutting argument is an argument $(\Pi, \neg\gamma)$, where $\gamma \in \Phi$. For a set of arguments $\{\Phi_1, ..., \Phi_n\}$, let Δ denote the union of the set of assumptions, ie. $\Delta = \Phi_1 \cup .. \cup \Phi_n$. Often in argumentation Δ will be inconsistent, and it may incorporate more than one minimally inconsistent subset[3]. Now, we can identify some arguments as safer than others according to the nature of the arguments and counter-arguments (both rebutting arguments and undercutting arguments). For example, an argument with no counter arguments is safer than an argument with counterarguments. As a result, we can rank conclusions on the basis of how safe the arguments for it are. As an example, suppose all maximally consistent subsets of Δ imply ϕ, and so all arguments for ϕ are relatively safe, yet a more preferred conclusion is a formula that follows from the intersection of the maximally consistent subsets of Δ. This approach to argumentation has been developed in [3, 29]. A number of other approaches to argumentation, including [70, 72, 97], focus on default reasoning by incorporating default connectives (which can be used to build up default statements similar to the default rules in default logic) into their languages together with associated machinery.

Argumentation can also be used to handle uncertain information by extending the pair (Φ, α) to a triple (α, ϕ, δ) in which δ is a measure of the degree to which α is believed to be true on the basis of Φ. In this way, argumentation can be used as a framework which can capture a number of different formalisms for handling uncertainty, with different formalisms entailing different meanings for δ (often called the "sign" of the argument) and different ways of handling the signs. This approach is described in more detail in this volume [33], and elsewhere [46], and its historical development is charted in [65]. It also forms the basis for one of the applications case studies in this book [47].

3.4 Truth maintenance systems

When reasoning with inconsistent information, questions of belief in assumptions and belief in conclusions arise. These questions include [58]:

Inferences from beliefs. How do new beliefs follow from existing beliefs?

Default beliefs. How do we record that a belief depends on the absence of other beliefs?

Dependency recording. How do we record that one belief depends on another belief?

[3] A minimally inconsistent set is a set of propositions which is inconsistent in the sense that $p \wedge \neg p$ can be derived from it for some p, and which is such that the removal of any one proposition from it will mean that the resulting set is not inconsistent.

Disbelief propagation. How do we withdraw belief in the consequences of a proposition that is disbelieved?

Revision of beliefs. How do we change beliefs in order to remove a contradiction?

These kinds of question led to approaches for truth maintenance[4]. A truth maintenance system (TMS) records information about each inference that is generated from a set of assumptions. The two main types of truth maintenance system are the justification-based truth maintenance system (JTMS) [20] and the assumption-based truth maintenance system (ATMS) [15]. A JTMS records a single set of consistent facts and all the inference which may be proved form them. When an inconsistency is detected some external system (which may be the user) is invoked to resolve the inconsistency and the JTMS then retracts the necessary inferences. In its simplest form, an ATMS maintains all the consistent subsets of the set of known facts and all the inferences which may be drawn from each. Inconsistency is handled by creating new consistent subsets and identifying which inferences may be made from them. Both types of system make it possible to identify consistent sets of beliefs and so make it possible to isolate inconsistency and avoid trivialization.

Truth maintenance can be considered to be concerned with lemma storage for non-monotonic reasoning. Thinking in terms of default logic, a JTMS can be considered to be a means of establishing a single extension and an ATMS as a means of establishing all the possible extensions. In a JTMS the discovery of a new fact which contradicts something in the existing extension will prompt the revisions necessary to establish a single new extension (if any exists). In an ATMS the introduction of a piece of contradicting information will generate a new set of extensions (if such extensions exist). The question of computational viability is then dependent upon the balance between on inferencing (consistency-checking and theorem proving) versus storage requirements (consistent subsets of data and inferential interdependencies). The aim of a TMS is to find the most parsimonious choice. A number of different implementations are given in [31], and a particular approach to assumption-based reasoning is described by Haenni [38]. In addition, more sophisticated truth maintenance systems will emerge from advanced theoretical frameworks such as that described by Benferhat and Garcia [4].

The notion of arguments discussed above provides useful concepts for formalising truth maintenance: For each explicit argument (Φ, α) there is classical proof of α from Φ so addressing the question of inferences from beliefs, and for the belief α, α is dependent on Φ so addressing the question of dependency recording. Let us assume that (Φ, α) follows from some assumptions Δ. To disbelieve some contradictory inferences from Δ requires a minimally inconsistent subset, Γ of Δ to be removed. Furthermore, Γ needs to be removed from the assumptions of all the argument, so all arguments (Ψ, β) become $(\Psi - \Gamma, \beta)$. This of course may involve withdrawing some arguments since the revised assumptions

[4] Now often referred to as "reason maintenance".

no longer imply the conclusion. In this way it is possible to address the questions of disbelief propagation and the revision of beliefs.

Truth maintenance systems have proved to be of particular interest for incorporation in diagnostic systems. Given some set of observations, such as symptoms, diagnosis involves determining the cause of those observation by selecting an appropriate consistent set of hypotheses from which the observations can be logically derived. So diagnosis can be viewed as constructing an argument. Furthermore, since the diagnostic process can take place over time, new observations can be obtained that can be inconsistent with the current diagnosis, so forcing the need for revision of beliefs. Diagnosis can therefore benefit from appropriate truth maintenance.

4 Combining approaches

Most of the research into uncertainty handling formalisms which has been mentioned so far has dealt with the use of single formalisms in isolation. However, if one accepts, as we do, that the eclectic position outlined in Section 1 is correct, then the following argument may be made. If different formalisms are good for representing different aspects of ignorance, then it follows that there are some problems which require the modelling of aspects of ignorance which are best covered by two or more different formalisms. Thus there is merit in investigating both the use of several formalisms in combination, and on determining the differences between different formalisms, and there is a growing body of work on this subject (though it should be noted that not all researchers working on such matters would explicitly acknowledge the validity of the eclectic position).

Possibly the most interesting strand of this kind of work is that which combines essentially logical techniques with numerical measures. This is commonly done by using a logical technique to establish a set of possible hypotheses from a larger initial set of exhaustive hypotheses, and then using a numerical techniques to rank the plausible set. Typical of such systems are those of Provan [73], Bigham [5] and Laskey and Lehner [51]. In all three of these systems, the semantic equivalence of the ATMS [15] and the Dempster-Shafer method, proved by Provan, is exploited ensuring that no information is lost in the initial round of inference. Bigham's system is particularly interesting in that it includes an extension of the clause based approach of McAllester's logic-based truth maintenance system (LTMS) [59] as a symbolic inference engine, and also permits beliefs based on possibility theory to be propagated. A similar system is de Kleer and Williams' [16] GDE for fault diagnosis. In GDE all inference directed at discovering the fault is carried out by symbolic methods, with probabilities invoked, not to determine the most likely of several solutions in a static analysis, but to suggest the next measurement to be taken by the user of the system. This measurement leads to new information which, when entered, leads to further symbolic computation. Thus the numerical computation sparks off another round of symbolic inference, and the cycle continues until the fault is found.

In contrast to these ATMS-based approaches, van Dam [96] uses a JTMS in combination with Dempster-Shafer theory.

It is also possible to use possibility measures with an assumption-based truth maintenance system instead of belief functions or probabilities. This is exactly the course followed by Dubois, Lang and Prade in their possibilistic ATMS [22]. A possibilistic ATMS is an ATMS in which both assumptions and justifications may be associated with a possibility weight, and, since the propagation of the weights is carried out for every clause in the ATMS, there is no separation of the management of uncertainty from the usual functionality of an ATMS. Bigham [6] has extended this work by adapting the possibilistic ATMS to take account of temporal information. Furthermore, the possibilistic ATMS allows inconsistent knowledge bases to be revised using the principles of epistemic entrenchment [34].

Another set of interesting developments which bridge the gap between symbolic and numerical techniques is the discovery of relationships between default logic and evidence theory. Wilson [99] considers the similarities between belief functions and default logic. He shows that, despite their initial dissimilarities they are, in fact, closely related. Indeed, in Łukaszewicz's [55] modification of default logic, the extensions of general closed default theories correspond to the sets of formulae whose beliefs, calculated by the theory of evidence, tend to 1 when the reliability of the sources of evidence tend to 1. The existence of a strong relationship between default logic and the theory of evidence is borne out by Smets and Hsia [89] who demonstrate how to represent normal defaults (both with and without prerequisites) using the transferable belief model. Both of these papers can be seen as an extension of the work of Rich [77] and Ginsberg [36], who considered ways of applying numerical certainty measures to logical inference rules.

It is also possible to use argumentation to combine symbolic and numerical reasoning. For instance, Fox and Krause [49] discuss a simple inference mechanism, based on argumentation, which is suitable for joint symbolic and numerical reasoning. Nonmonotonic reasoning about Tweety's ability to fly is handled in the following way. The result of applying the default rule that "typically birds fly" is marked as supported by a "possible" argument, thus explicitly recording the fact that the conclusion need not be true. A certain inference of the form that Tweety doesn't fly because she is an ostrich is supported by a true argument. When two facts are in conflict, reasoning that a default fact is a less powerful argument than a true one resolves the situation. Similarly, numerical techniques generate arguments quantified by numerical degrees of belief, which can be compared to order hypotheses. However, this method is more than just a fancy method for quantifying propositions. The quantifier also allows the reasoning mechanism to refer to the grounds of the argument, identifying why the argument was generated. This provides the vital connection between the degree of belief and the underlying uncertainty that is missing from most methods of approximate reasoning. From the grounds, we can establish the reasons for the uncertainty, and the nature of the uncertainty, and reasoning about this allows us to proceed when we would otherwise be held up by the incomparability of the degrees of belief with which the propositions we are dealing with are quantified.

5 Summary

In this introduction to uncertainty formalisms we have only been able to briefly cover some of the many uncertainty formalisms which have been proposed over the years[5]. However, despite this diversity, we strongly believe that no single approach is appropriate for all uncertainty handling problems. Furthermore, for some uncertainty handling problems, we believe that a mixture of approaches is required. While this statement is still controversial in some quarters, there seems to be a growing realisation that the position it represents has some merit, and so there are clear arguments for the development of a range of uncertainty formalisms. In particular, there is still more work to be done in developing the range of uncertainty formalisms and in learning more about how to use them effectively in a wider range of uncertainty problems.

References

1. S. K. Andersen, F. V. Jensen, K. G. Olesen, and F. Jensen. HUGIN—a shell for building Bayesian belief universes for expert systems. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 783–791, San Mateo, CA, 1989. Morgan Kaufmann.
2. J. A. Barnett. Computational methods for a mathematical theory of evidence. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 868–875, Los Altos, CA, 1981. William Kaufmann.
3. S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proceedings of the 9th Uncertainty in Artificial Intelligence*, pages 411–419. Morgan Kaufmann, 1993.
4. S. Benferhat and L. Garcia. A local handling of inconsistent knowledge and default bases. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
5. J. Bigham. Computing beliefs according to Dempster-Shafer and possibilistic logic. In *Proceedings of the 3rd International Conference on Information Processing and Management of Uncertainty, Paris*, pages 59–61, 1990.
6. J. Bigham. Exploiting uncertain and temporal information in correlation. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
7. P. Bosc, L. Lietard, and H. Prade. An ordinal approach to the processing of fuzzy queries with flexible quantifiers. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
8. F. Brazier, J. Engelfreit, and J. Truer. Analysis of multi-interpretable ecological monitoring information. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.

[5] For a more detailed survey, see [43, 48, 64].

9. T. Chard. Qualitative probability versus quantitative probability in clinical diagnosis: a study using a computer simulation. *Medical Decision Making*, 11:38–41, 1991.
10. P. Cheeseman. Probabilistic vs. fuzzy reasoning. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 85–102. Elsevier Science Publishers, Amsterdam, The Netherlands, 1986.
11. P. Cheeseman. Discussion of the paper by Lauritzen and Spiegelhalter. *Journal of the Royal Statistical Society, B*, 50:203, 1988.
12. P. Cheeseman. An inquiry into computer understanding. *Computational Intelligence*, 4:58–142, 1988.
13. D. A. Clark. Numerical and symbolic approaches to uncertainty management in AI. *Artificial Intelligence Review*, 4:109–146, 1990.
14. P. R. Cohen. *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach*. Pitman, London, UK, 1985.
15. J. de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.
16. J. de Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.
17. A. P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
18. A. P. Dempster. A generalisation of Bayesian inference (with discussion). *Journal of the Royal Statistical Society B*, 30:205–232, 1968.
19. A. P. Dempster. Comments on 'An inquiry into computer understanding' by Peter Cheeseman. *Computational Intelligence*, 4:72–73, 1988.
20. J. Doyle. A truth maintenance system. *Artificial Intelligence*, 12:231–272, 1979.
21. D. Driankov. A calculus for belief-intervals representation of uncertainty. In B. Bouchon-Meunier and R. R. Yager, editors, *Uncertainty in Knowledge-Based Systems*, pages 205–216. Springer-Verlag, Berlin, Germany, 1986.
22. D. Dubois, J. Lang, and H. Prade. A possibilistic assumption-based truth maintenance system with uncertain justifications, and its application to belief revision. In J. P. Martins and M. Reinfrank, editors, *Truth Maintenance Systems*, pages 87–106, Berlin, Germany, 1990. Springer Verlag.
23. D. Dubois, J. Lang, and H. Prade. Fuzzy sets in approximate reasoning, Part 2: Logical approaches. *Fuzzy Sets and Systems*, 40:203–244, 1991.
24. D. Dubois and H. Prade. Necessity measures and the resolution principle. *IEEE Transactions on Systems, Man and Cybernetics*, 17:474–478, 1987.
25. D. Dubois and H. Prade. Modelling uncertainty and inductive inference: a survey of recent non-additive probability systems. *Acta Psychologica*, 68:53–78, 1988.
26. D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, NY, 1988.
27. D. Dubois and H. Prade. Processing of imprecision and uncertainty in expert system reasoning models. In C. Ernst, editor, *Management Expert Systems*, pages 67–88. Addison Wesley, 1988.
28. R. O. Duda, P. E. Hart, and N. J. Nilsson. Subjective Bayesian methods for a rule-based inference system. In *Proceedings of the National Computer Conference*, pages 1075–1082, 1976.
29. M. Elvang-Gøransson and A. Hunter. Argumentative logics: Reasoning from classically inconsistent information. *Data and Knowledge Engineering Journal*, 16:125–145, 1995.
30. D. W. Etherington. *Reasoning with Incomplete Information*. Pitman, London, UK, 1988.

31. K. D. Forbus and J. de Kleer. *Building Problem Solvers.* MIT Press, Cambridge, MA, 1993.
32. J. Fox. Three arguments for extending the framework of probability. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 447–458. Elsevier Science Publishers, Amsterdam, The Netherlands, 1986.
33. J. Fox and S. Parsons. Arguing about beliefs and actions. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume).* Springer Verlag, Berlin, 1998.
34. P. Gärdenfors. *Knowledge in flux: Modelling the Dynamics of Epistemic States.* MIT Press, Cambridge, MA, 1988.
35. J. Gebhardt and R. Kruse. Background to and perspectives on possibilistic graphical models. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume).* Springer Verlag, Berlin, 1998.
36. M. Ginsberg. Non-monotonic reasoning using Dempster's rule. In *Proceedings of the 4th National Conference on Artificial Intelligence*, pages 112–119, Los Altos, CA, 1984. William Kaufmann.
37. J. Gordon and E. H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985.
38. R. Haenni. Modelling uncertainty in propositional assumption-based systems. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume).* Springer Verlag, Berlin, 1998.
39. D. Heckerman and M. Wellman. Bayesian networks. *Communications of the ACM*, 38:27–30, 1995.
40. D. E. Heckerman. Probability interpretation for MYCIN's certainty factors. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 167–196. Elsevier Science Publishers, Amsterdam, The Netherlands, 1986.
41. M. Henrion, G. Provan, B. Del Favero, and G. Sanders. An experimental comparison of numerical and qualitative probabilistic reasoning. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, pages 319–326, San Francisco, CA, 1994. Morgan Kaufmann.
42. E. J. Horvitz, D. E. Heckerman, and C. P. Langlotz. A framework for comparing alternative formalisms for plausible reasoning. In *Proceedings of the 5th National Conference on Artificial Intelligence*, pages 210–214, Los Altos, CA, 1986. Morgan Kaufmann.
43. A. Hunter. *Uncertainty in Information Systems.* McGraw-Hill, London, UK, 1996.
44. D. J. Israel. Some remarks on the place of logic in knowledge representation. In N. Cercone and G. McCalla, editors, *The Knowledge Frontier: Essays in the Representation of Knowledge*, pages 80–91. Springer Verlag, New York, NY, 1987.
45. G. J. Klir. Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like? *Fuzzy Sets and Systems*, 24:141–160, 1987.
46. P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11:113–131, 1995.
47. P. Krause, J. Fox, P. Judson, and M. Patel. Qualitative risk assessment fulfills a need. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume).* Springer Verlag, Berlin, 1998.
48. P. J. Krause and D. A. Clark. *Representing Uncertain Knowledge: An Artificial Intelligence Approach.* Intellect, Oxford, UK, 1993.
49. P. J. Krause and J. Fox. Combining symbolic and numerical methods for reasoning under uncertainty. In D. J. Hand, editor, *AI and Computer Power; The Impact on Statistics*, pages 99–114. Chapman and Hall, London, UK, 1994.

50. M. Lalmas. Modelling information retrieval with Dempster-Shafer's theory of evidence. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
51. K. B. Laskey and P. E. Lehner. Assumptions, beliefs and probabilities. *Artificial Intelligence*, 32:65–77, 1990.
52. S. L. Lauritzen and D. J. Spiegelhalter. Local computations on graphical structures, and their application to expert systems. *Journal of the Royal Statistical Society, B*, 50:157–224, 1988.
53. D. V. Lindley. *Making Decisions*. John Wiley & Sons, Chichester, UK, 1975.
54. W. Łukaszewicz. Two results on default logic. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 459–461, Los Altos, CA, 1985. Morgan Kaufmann.
55. W. Łukaszewicz. Considerations on default logic: an alternative approach. *Computational Intelligence*, 4:1–16, 1988.
56. P. Magni, R. Bellazi, and F. Locatelli. Using uncertainty management techniques in medical therapy planning: a decision theoretic approach. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
57. P. Magrez and Ph. Smets. Fuzzy modus ponens: A new model suitable for applications in knowledge-based systems. *International Journal of Intelligent Systems*, 4:181–200, 1975.
58. J. Martins and S. Shapiro. A model of belief revision. *Artificial Intelligence*, 35:25–79, 1988.
59. D. A. McAllester. An outlook on truth maintenance. AI Memo 551, AI Laboratory, MIT, 1980.
60. J. McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.
61. S. Moral and N. Wilson. Importance sampling Monte-Carlo algorithms for the calculation of Dempster-Shafer belief. In *Proceedings of the 6th International Conference on Information Processing and the Management of Uncertainty*, pages 1337–1344, 1996.
62. P. Nicolas and T. Schaub. The XRay system: an implementation platform for local query-answering in default logics. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
63. N. J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.
64. S. Parsons. *Qualitative approaches to reasoning under uncertainty*. MIT Press, Cambridge, MA, 1998.
65. S. Parsons and J. Fox. Argumentation and decision making: a position paper. In *Formal and Applied Practical Reasoning*, pages 705–709, Berlin, Germany, 1996. Springer Verlag.
66. J. Pearl. How to do with probabilities what people say you can't. Technical Report CSD-850031, Cognitive Systems Laboratory, Computer Science Department UCLA, 1985.
67. J. Pearl. Fusion, propagation and structuring belief networks. *Artificial Intelligence*, 29:241–288, 1986.
68. J. Pearl. Bayesian decision methods. In *Encyclopedia of Artificial Intelligence*, pages 48–56. John Wiley, 1987.
69. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

70. G. Pinkas and R. Loui. Reasoning from inconsistency: A taxonomy of principles for resolving conflict. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*. Morgan Kaufmann, 1992.
71. M. Pradhan, M. Henrion, G. Provan, B. Del Favero, and K. Huang. The sensitivity of belief networks to imprecise probabilities: an experiemntal investigation. *Artificial Intelligence*, 85:363–397, 1996.
72. H. Prakken. An argumentation framework for default reasoning. *Annals of Mathematics and Artificial Intelligence*, 9, 1993.
73. G. M. Provan. Solving diagnostic problems using extended assumption-based truth maintenance systems: foundations. Technical Report 88-10, Department of Computer Science, University of British Columbia, 1988.
74. J. R. Quinlan. INFERNO: a cautious approach to uncertain inference. *Computer Journal*, 26:255–269, 1983.
75. M. Ramalho. Uncertainty measures associated with fuzzy rules for connection admission control in ATM networks. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
76. R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.
77. E. Rich. Default reasoning as likelihood reasoning. In *Proceedings of the 3rd National Conference on Artificial Intelligence*, pages 348–351, Los Altos, CA, 1983. William Kaufmann.
78. A. Saffiotti. An AI view of the treatment of uncertainty. *The Knowledge Engineering Review*, 2:75–97, 1987.
79. A. Saffiotti. Handling uncertainty in control of autonomous robots. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume)*. Springer Verlag, Berlin, 1998.
80. B. Schweitzer and A. Sklar. Associative functions and abstract semigroups. *Publicationes Mathematicae Debrecen*, 10:69–81, 1963.
81. G. L. S. Shackle. *Decision, order and time in human affairs*. Cambridge University Press, Cambridge, UK, 1961.
82. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.
83. G. Shafer. Comments on 'An inquiry into computer understanding' by Peter Cheeseman. *Computational Intelligence*, 4:121–124, 1988.
84. G. Shafer and R. Logan. Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence*, 33:271–298, 1987.
85. P. P. Shenoy and G. Shafer. Axioms for probability and belief function propagation. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 169–198. North-Holland, Amsterdam, The Netherlands, 1990.
86. E. H. Shortliffe. *Computer-Based Medical Consultations: MYCIN*. Elsevier, New York, NY, 1976.
87. Ph. Smets. Belief functions. In Ph. Smets, E. H. Mamdani, D. Dubois, and H. Prade, editors, *Non-Standard Logics for Automated Reasoning*, pages 253–275. Academic Press, London, UK, 1988.
88. Ph. Smets. Belief functions versus probability functions. In B Bouchon-Meunier, L. Saitta, and R. R. Yager, editors, *Uncertainty and Intelligent Systems*, pages 17–24. Springer Verlag, Berlin, Germany, 1988.
89. Ph. Smets and Y-T. Hsia. Default reasoning and the transferable belief model. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 495–504. Elsevier Science Publishers, Amsterdam, The Netherlands, 1991.

90. Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.
91. M. Smithson. *Ignorance and Uncertainty: Emerging Paradigms.* Springer Verlag, New York, NY, 1989.
92. L. E. Sucar, D. F. Gillies, and D. A. Gillies. Objective probabilities in expert systems. *Artificial Intelligence*, 61:187–208, 1993.
93. A. Tawfik and E. Neufeld. Model-based diagnosis: a probabilistic extension. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume).* Springer Verlag, Berlin, 1998.
94. S. Toulmin. *The uses of argument.* Cambridge University Press, Cambridge, UK., 1957.
95. A. Tversky and D. Kahneman. Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.
96. K. van Dam. Using uncertainty techniques in radio communication systems. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume).* Springer Verlag, Berlin, 1998.
97. G. Vreeswijk. Abstract argumentation systems. In M de Glas and D Gabbay, editors, *Proceedings of the First World Conference on Fundamentals of Artificial Intelligence.* Angkor, 1991.
98. D. Wilson, A. Greig, John Gilby, and Robert Smith. Some problems in trying to implement uncertainty techniques in automated inspection. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms (this volume).* Springer Verlag, Berlin, 1998.
99. N. Wilson. Rules, belief functions, and default logic. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, pages 443–449, Mountain View, CA, 1990. Association for Uncertainty in AI.
100. N. Wilson. *Some theoretical aspects of the Dempster-Shafer theory.* PhD thesis, Oxford Polytechnic, 1992.
101. N. Wilson and S. Moral. Fast Markov chain algorithms for calculating Dempster-Shafer belief. In *Proceedings of the 12th European Conference on Artificial Intelligence*, pages 672–676, Chichester, UK, 1996. John Wiley & Sons.
102. L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
103. L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:1–28, 1978.
104. L. A. Zadeh. A theory of approximate reasoning. In J. Hayes, D. Michie, and L. Mikulich, editors, *Machine Intelligence 9*, pages 149–194. Ellis Horwood, 1979.
105. L. A. Zadeh. Is probability theory sufficient for dealing with uncertainty in AI? a negative view. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 103–116. Elsevier Science Publishers, Amsterdam, The Netherlands, 1986.