1-1-1998

A methodology for the selection of a paradigm of reasoning under uncertainty in expert system development

Vivian Campbell
Edith Cowan University

A METHODOLOGY FOR THE SELECTION OF A PARADIGM OF REASONING UNDER UNCERTAINTY IN EXPERT SYSTEM DEVELOPMENT

by

Vivian Campbell B.App.Sc.

A Thesis submitted in Partial Fulfilment of the Requirements for the Award of

Master of Science(Computer Science)

at the School of Mathematics, Information Technology and Engineering,

Edith Cowan University

Name of Supervisor : Dr Chaiyaporn Chirathamjaree

Date of Submission: 20th April 1998

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

Abstract

The aim of this thesis is to develop a methodology for the selection of a paradigm of reasoning under uncertainty for the expert system developer. This is important since practical information on how to select a paradigm of reasoning under uncertainty is not generally available.

The thesis explores the role of uncertainty in an expert system and considers the process of reasoning under uncertainty. The possible sources of uncertainty are investigated and prove to be crucial to some aspects of the methodology.

A variety of Uncertainty Management Techniques (UMTs) are considered, including numeric, symbolic and hybrid methods. Considerably more information is found in the literature on numeric methods, than the latter two. Methods that have been proposed for comparing UMTs are studied and comparisons reported in the literature are summarised. Again this concentrates on numeric methods, since there is more literature available.

The requirements of a methodology for the selection of a UMT are considered. A manual approach to the selection process is developed. The possibility of extending the boundaries of knowledge stored in the expert system by including meta-data to describe the handling of uncertainty in an expert system is then considered. This is followed by suggestions taken from the literature for automating the process of selection.

Finally consideration is given to whether the objectives of the research have been met and recommendations are made for the next stage in researching a methodology for the selection of a paradigm of reasoning under uncertainty in expert system development.

DECLARATION

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or a diploma in any institution of higher education; and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signature.

Date 31/7/98

CONTENTS

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Chapter overview

Chapter One introduces the thesis, which is concerned with the development of a methodology for the selection of a paradigm of reasoning under uncertainty in expert system development. The chapter explains the research that will be covered. The significance of the study and the purpose of the study are described in the light of literature material. The research questions are posed and the organisation of the thesis is outlined.

1.2 Introduction to the thesis

This thesis is concerned with the investigation of paradigms of reasoning under uncertainty that have been applied to expert systems. It will consider in detail a number of Uncertainty Management Techniques (UMTs) and consider their application. Procedures to compare UMTs will be investigated and an attempt made to develop a methodology that can be used to select the appropriate UMT for a particular expert system development.

1.3 The significance of the study

The research topic outlined above, was suggested to me by a (then) lecturer in Computer Science at Edith Cowan University, Mr. Tim Roberts. He has some experience in the development of Diagnostic Expert Systems. He had discovered that although there were many theoretical papers on the subject of uncertainty in expert systems, practical information on how to select a paradigm of reasoning under uncertainty was not readily available.

Expert systems are designed to solve real-life problems. Such problems are often not straightforward enough to be dealt with by the use of applied predicate calculus -- as was hoped in the 1960s (Lucas & Van Der Gaag, 1991). Expert systems may be distinguished from classical decision theory systems by the importance of the "representation of knowledge in an explicit qualitative form rather than implicitly in an algorithmic form" (Fox, Clark, Glowinski, O'Neil, 1990).

The real-life situations tackled by expert systems are often typified by a degree of uncertainty. This may include imprecise or conflicting information. Since expert system applications are designed to deal with real-life problems at the level of the human expert, they must cope with uncertain information. However, uncertainty does not arise from a single source and may arise even in completely deterministic systems (Rothman, 1989). This concept is dealt with in more detail in section 2.5 Sources of uncertainty.

In the development of expert systems that deal with uncertain information, the selection of a paradigm of reasoning under uncertainty is critical (Hsu & Chu, 1989). If uncertainty is not properly dealt with, the expert system may have an illusion of precision (Kerr,1992).

The largest group of expert systems that must deal with uncertainty are referred to as diagnostic systems (Weichselberger & Pohlman, 1990). This type of system is most often connected with the medical field but examples are also present in many other fields including financial planning, accounting, geology, meteorology and the control of technical installations.

Techniques have been developed specifically for handling uncertainty in expert systems. For example Certainty Factors were developed for the early medical expert system MYCIN (Shortliffe & Buchanan, 1975). Others especially the probability based methods have evolved from long established mathematical techniques (Bhatnagar & Kanal, 1986). However, in the theory of the management of uncertainty there is often criticism of Certainty factors as being mathematically invalid. Is the criticism valid or is it enough for a technique to produce satisfactory results? If so, then how should the selection of a method be made and on what criteria should that selection be based?

1.4 The purpose of the study

Many different paradigms for reasoning under uncertainty in expert systems have been proposed in the last twenty-five years. Most of them have been numeric systems often based, however loosely, on mathematical probability. Some have argued that the uncertainty in expert systems cannot be combined into numeric values (Cohen, 1988). More recently there has been the development of hybrid systems that have numeric and non-numeric components (D'Ambrosio, 1988) (Cohen, 1985).

In recent years there has been an almost religious debate about which is the right system to use for reasoning under uncertainty. Some have claimed that the established mathematical probability methods must be used (Cheeseman, 1986) (Lindley, 1985), whilst others have claimed that new methods are required (Zadeh, 1986). Max Henrion, in the preface to Uncertainty in Artificial Intelligence 5 indicates that this debate is inappropriate (Henrion, 1990). He suggests that it is not possible to select one UMT over another by considering only the basic mechanics or mathematical soundness of the theory. There are other practical considerations that must also be made such as the reliability and complexity of calculations. In addition users must be able to understand the model in order to provide data that can be used with confidence.

Henrion suggests that the criterion for success of an approach is its effectiveness for application. "The marketplace for ideas, like more tangible goods, is ultimately ruled more by consumers than producers." (Henrion et al. 1990, p. v). This thesis will explore the experience gained in the marketplace and look for recommendations that can be made to "consumers".

Clark agrees that attempts to demonstrate that one particular UMT was the best for all situations, were unfortunate. He concurs with the aim of this study "to suggest the most appropriate paradigm for a particular situation" (Clark, 1990, p. 140).

Saffioti supports the argument that it is appropriate to identify the paradigm of reasoning under uncertainty that should be used for a given application (Saffioti, 1988). Fox suggests that the debate about the correct way of dealing with uncertainty is unfortunate. He submits rather that debate should focus on considering the strengths and weaknesses of alternate methods of representing uncertainty (Fox, 1986).

The objective of this study is to select from the many alternatives, the most appropriate paradigm for reasoning under uncertainty for a particular application. Ginsberg (1986) advises that comparing the theory of UMTs is difficult. "The true advantages of the various competing paradigms will only be apparent when these paradigms have been incorporated in full-scale systems." It is now true that many paradigms are in use, and this thesis will investigate them. However, there continue to be theoretical developments that cannot be ignored. Where possible these developments will be given some consideration.

1.5 Statement of research questions

In the early stages of expert system development, the selection of a paradigm for reasoning under uncertainty is important. However Hsu and Chu in their paper "Practical issues in designing knowledge-based expert systems" (Hsu & Chu, 1989) identify the representation of uncertainty as one area that is often neglected in the design of an expert system. This is unfortunate, since an Uncertainty Management Technique (UMT) provides the expert system with a means of assessing evidence and making credible inferences about hypotheses in an indefinite environment.

This project has two aims.

1. To define the criteria on which the selection of a paradigm of reasoning under uncertainty for an expert system should be made.

2. To consider which recent advances in the theory of reasoning under uncertainty are worthy of consideration for incorporation into expert system developments.

The first is the major aim and the thesis will be structured around this aim. The second is considered to be of secondary importance and will be considered alongside the first.

In an attempt to answer these questions, information will be gathered from two major sources.

1. The theory of reasoning under uncertainty. There is a great deal of material available in journals and books

2. Expert System applications. Detailed information on the success or failure of the particular UMT used is more difficult to obtain. There are a few useful studies that discuss the attributes of systems in relation to their reasoning under uncertainty.

1.6 Organisation of the thesis

Chapter 2 sets the scene for this study. It outlines the structure of expert systems and discusses how uncertainty may become a part of this structure. It discusses the role of uncertainty and explains the process of reasoning under uncertainty. The problem faced by the expert system developer of having to select a paradigm of reasoning under uncertainty is described, sources of uncertainty are outlined and the concept of validating expert systems is discussed.

Chapter 3 provides the detail of the major paradigms for reasoning under uncertainty in expert systems. Several numeric approaches are discussed at length, whilst the symbolic and hybrid approaches receive rather less detailed consideration. For each paradigm the advantages and disadvantages of the technique are discussed and in some cases the relationship to other UMTs is clarified.

Chapter 4 begins by considering structures by which comparisons of paradigms for reasoning under uncertainty may be made. A list of the requirements of a theory of uncertainty management due to Bonissone (Bonissone, 1987) is discussed. A comparison of techniques is then made in the light of these ideas.

Suggestions concerning how to select a specific UMT for a particular application are to be found in chapter 5. This includes a discussion of the process of making a decision and the concepts of the expert systems that should be given cognisance.

Chapter 6 concludes this study by considering whether the objectives have been met and making recommendations for the next step in the study of this topic.

Finally, a comprehensive bibliography is included at the end of the thesis.

Chapter 2: Expert Systems and reasoning under uncertainty

2.1 Chapter overview

This chapter explains the concept of an expert system. It considers the structure of expert systems and how uncertainty may be incorporated into them. The reasoning process is outlined, as are the required changes to reasoning when uncertainty is involved. The requirement for an expert system developer to consider uncertainty is discussed. Sources of uncertainty are investigated that will indeed prove important in selecting the appropriate paradigm of reasoning under uncertainty. Finally the chapter considers the process of validation of expert systems and explains why this validation process is necessarily different from that in a system that uses exact reasoning.

2.2 Expert systems

An expert system is designed to make judgments in a complex field. It is supposed to make judgments at least as well as a human expert. This goal can be approached from two different perspectives. The first is to concentrate solely on the results of the system, if it makes the same recommendations as the human expert then that is all that is required of the system. The second approach achieves the same results but places additional importance on the reasoning process that achieved those results. This process should be as close as is possible to the psychological reasoning process that is used by the human expert. It has been argued that backward chaining uses a process that is similar to the human expert when propositional data is used (Neopolitan, 1990).

Backward chaining is the process of reasoning from a conclusion to proving the facts that support that conclusion. Forward chaining is the reverse process of reasoning from the facts to the conclusions that the facts support. Both directions of reasoning have been used by the inference engine of expert systems. EMYCIN uses backward chaining whilst OPS5 and CLIPS use forward chaining (Giarratano & Riley, 1994). Some inference engines actually allow both types of reasoning. A backward chaining reasoning process is usually more convergent simply because irrelevant facts can be discarded immediately (Jackson, 1990).

The greater number of expert system applications may be defined as *classification* problems (Ignizio, 1991). Included here are the diagnostic systems that given a set of symptoms will attempt to diagnose the disease and also systems that consider the cause of machinery failure. Since this type of system is attempting to establish an hypothesis given the conclusion, backward chaining reasoning is preferable. Forward chaining should be selected for other types of expert systems, those that attempt to solve *construction* problems (Ignizio, 1991). This includes expert systems for prognosis, monitoring and control (Giarratano & Riley, 1994). For example XCON was to advise on suitable configurations of VAX computers (McDermott, 1982).

Expert systems generally work in a narrow domain. MYCIN, probably the world's best known expert system was intended to be used for diagnosis of infectious blood diseases (Shortliffe, 1975). Despite this, complete certain knowledge of that domain is the exception rather than the rule. Human experts very often must reach decisions with concepts that are unreliable, incomplete or inconsistent and expert systems must do the same. Velverde and Gehl emphasise that expert systems must be capable of managing uncertainty before they can thrive in their intended field and use their knowledge successfully (Valverde & Gehl, 1992).

2.3 Expert system structure

Expert systems are usually considered to have three main components, a knowledge base, inference engine and user interface. These components respectively represent, manipulate and communicate knowledge. In addition, most systems also contain explanation and trace facilities. The interfaces between these components are shown in Fig 2.1. Explanation facilities are primarily for the expert system user and will provide more information on such questions as "How did you reach that answer?", whereas trace facilities are for the knowledge engineer and will provide for stepping through the inferencing process (Lucas & Van der Gaag, 1991).

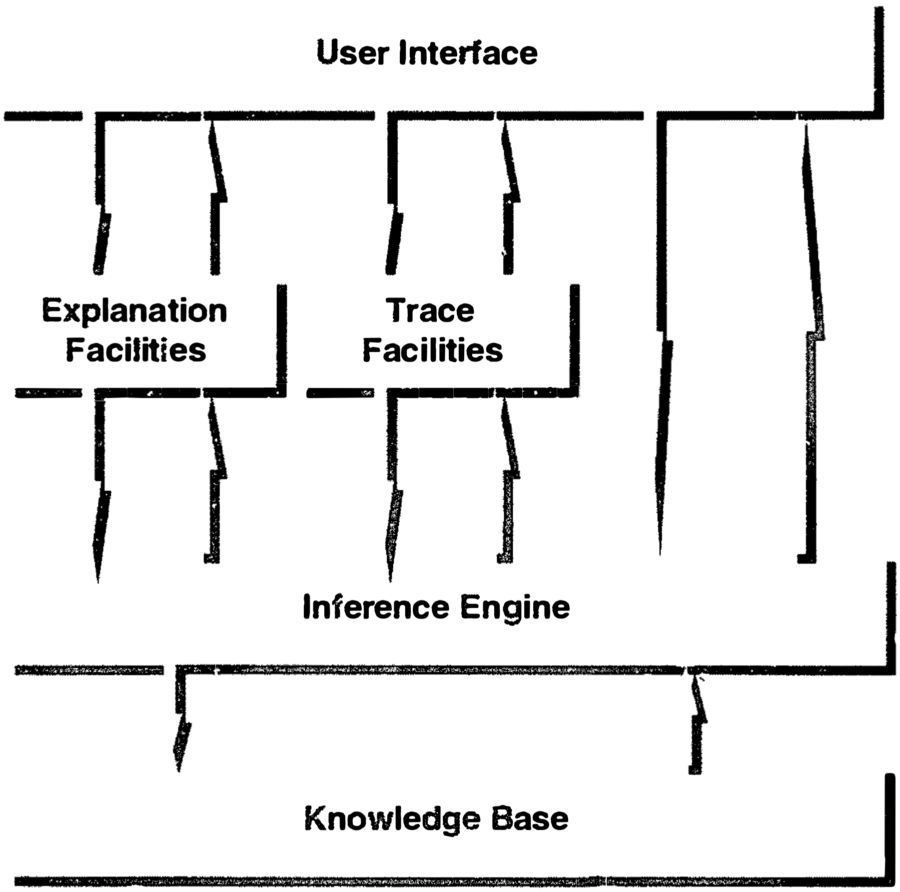


Figure 2.1 The Structure of an Expert System

How does the processing of uncertainty fit into the structure of an expert system? There are two possibilities. First from (Cortez-Rello and Golshani, 1990) is the separate approach that includes two additional modules in an expert system, the Belief base and Uncertainty module. They indicate that the Belief base communicates with the Uncertainty module which in turn communicates with the inference engine. An integrated approach would include aspects of uncertainty in all the components of the system. Since uncertainty must be represented, manipulated and communicated it will be a part of each of the three major components. Bonissone endorses the integrated approach

> In building expert-system architectures three distinct layers must be defined: representation, inference, and control layers. The treatment of uncertainty must address each of these layers. (Bonissone, 1987, p. 859)

Paul Cohen confirms that the integrated approach is important by emphasising that the management of uncertainty should not be an after thought that is an addition to a categorical inference system. He views uncertainty management as an "integral part of the problem solving process." (Cohen, 1989, p. 263)

2.4 The role of uncertainty in expert systems

Expert systems need to have the capability to infer from premises that are imprecise, incomplete or not totally reliable, just as human experts function in the same situation. The strict implication, "for all x, A(x) implies B(x)" is weakened by some degree, expressed as a scalar value, to "for most x, A(x) implies B(x)" (Bonissone, 1987). Less formally this statement has been phrased "the A's are B's" or "generally the A's are B's". A group of French researchers investigated the ways uncertainty was introduced into this statement in various formalisms. They used the title, Lea Sombe, from the French "les A sont B" (Lea Sombe, 1990). Their versions of this statement expressed using various paradigms of reasoning under uncertainty will be considered in the appropriate sections of Chapter 3.

The degree to which there is still a belief in the implication has been called a *degree of belief* (Shafer, 1976). The functions that propagate degrees of belief over inferences are called *combining functions*. Some systems propagate two degrees of belief, usually an upper and a lower bound, so indicating a range of values. Bonissone explains the process of reasoning under uncertainty:

> Facts must be aggregated to determine the degree to which the premise of a given rule has been satisfied, to verify the extent to which external constraints have been met, to propagate the amount of uncertainty through the triggering of a given rule, to summarize the findings provided be various rules or knowledge sources or experts, to detect possible inconsistencies among the various sources, and to rank different alternatives of different goals. (Bonissone, 1987, p. 854).

Bob Avanzato, an expert system developer, was looking for a suitable UMT to use in an Acoustic Signal Interpretation Expert System. He considered uncertainty to be an integral part of the expert system, and emphasised the importance of this part of the system. He felt that the UMT must be able to represent and reason with uncertainty, and should encompass all the facets of uncertainty in order to guarantee success in the evolution and installation of the expert system (Avanzato, 1991)(see Section 2.5 Sources of uncertainty).

Some have seen the numerical approaches to uncertainty as attempting to produce results with excessive precision. Most numeric UMTs require numerical values from the user as an accurate measure on a scalar or interval scale. They then perform complex calculations that produce seemingly precise results. It has been suggested that this apparent accuracy may not be justified given the difficulty of obtaining accurate initial figures (Bonissone, 1987).

2.5 Sources of uncertainty.

It would be convenient to package all forms of uncertainty into a single bundle and deal with this in a consistent manner throughout the expert system. Bonissone reminds us that uncertainty is not a single issue.

> the presence of uncertainty in reasoning systems is caused by a variety of sources: the reliability of the information, the inherent imprecision of the representation language in which the information is conveyed, the incompleteness of the information, and the aggregation or summarization of information from multiple sources. (Bonissone, 1987, p. 854)

All uncertainty involved in expert systems then. does not arise from a single source. In fact it may be inappropriate to package several different concepts together when considering expert systems or any reasoning system. Ng and Abramson (1990, p. 30) identify the same four sources of uncertainty as Bonissone but use different terminology. Each will be considered individually.

1. Lack of precision of knowledge/natural language.

Ambiguities may not be clarified during translation to a formal language. Thus it may be necessary to allow for the imperfect matching of facts with premises. Statements such as "the economy has a low inflation rate" are imprecise (Bhatnagar & Kanal, 1992).

2. Unreliable information.

May be due to:

1. Ill-defined domain concepts.

2. Inaccurate data possibly due to poor reliability of instruments used to make the observations.

3. Weak implications may occur because the system builder is not able to establish a concrete relationship between the antecedent and consequent.

This is perhaps the only true 'Uncertainty of knowledge' (Bhatnagar & Kanel. 1992).

3. Incomplete information.

Partial information results when the answers to questions are unknown. Approximate pattern matching is required here also. Bonissone (Bonissone, 1987) suggests that this type of uncertainty has often been modelled by non-numerical methods.

4. Disagreement amongst experts.

Conflicting information from a number of sources will result in conclusions that are suspect. Bonissone points out that when unconditional facts are combined three possible problems may appear:

- the single-valued certainty measure may be combined into an interval-value
- the combination of conflicting statements could generate a contradiction
- the rule of evidence may create an overestimate of the aggregated fact if a normalization is used to hide a contradiction. This was shown possible by Zadeh cited by Bonissone (Bonissone, 1987).

It has been shown that it is possible that a consensus can be reached by weighting each source, (depending on the expertise of the source) and thus calculating composite information. It would however be difficult to define the weights since this requires a weight for each expert and experts do not have uniform expertise across their domain (Ng & Abramson, 1990).

Graham (1991) suggests that it is important to consider two different kinds of uncertainty arising from:

1. Natural variation

   This includes such concepts as probability and possibility that can be handled by statistical and fuzzy methods respectively (in Graham's view)

2. Conceptual apprehension.

This includes ideas of vagueness, variations in belief, degrees of truth, etc.

It would appear then that only the first of Ng and Abramson's concepts is part of the first of Graham's. The second, third and fourth listed by Ng and Abramson however can all be included as Graham's second.

This thesis agrees with van der Lubbe and colleages that the type of uncertainty is crucial to the selection of an appropriate paradigm of reasoning under uncertainty for a given application (van der Lubbe, Backer & Krijgman, 1991).

2.6 Reasoning under Uncertainty

Exact reasoning involves the use of exact facts and exact conclusions follow. In a deductive argument, the conclusion must follow from the premises. When facts are uncertain there may be a great number of possible conclusions and the problem becomes selecting the best conclusion.

When reasoning under uncertainty, a conclusion may be arrived at with less than 100% certainty. A doctor may diagnose a certain treatment because it appears likely the patient has a disease. The treatment may be the correct decision without confirmation of the diagnosis if there are few side effects to the treatment and the cost (in time or money) of confirming the diagnosis is great.

It is not clear how human experts represent and reason under uncertainty. Some have argued that a form of "logic" is used, others that humans actually evaluate probabilities. At the other end of the spectrum are those that suggest that no explicit representations are used (Graham, 1991). Whatever the method, it is true that experts can make "useful and meaningful recommendations" even when faced with imprecise and uncertain information (Clarke, McLeish & Vyn, 1991).

If the presence of uncertainty is acknowledged and a method of approximate reasoning is to be included in an expert system then there remain two major problems that must be resolved:

- how to measure the degree of inexactness and calculate certainty factors for inexact situations,
- how to propagate uncertainty in making inferences and arrive at best conclusions in spite of some rules not being definite.

(Cortes-Rello and Golshani, 1990, p. 9)

These problems can be summarised as how to represent and reason with uncertainty. Chapter 3 considers various possible paradigms for reasoning under uncertainty. Each provides its own solution to these problems.

2.7 The problem for an expert system developer.

Some expert system developers have not considered including the handling of uncertainty in their systems. This has resulted in what Kerr referred to as an "illusion of precision" (Kerr, 1992). He reported that in some (scheduling) systems the lack of approximate reasoning can result in large numbers of calculations when minor data changes are made. So the lack of a method for dealing with uncertainty can result in major problems for the expert system.

The handling of uncertainty in expert systems is a complex task that has several possible solutions. There are no simple methods to provide an answer to the question "which UMT is appropriate for my expert system?". Expert systems are being asked to solve more challanging problems that involve many types of uncertainty and it is therefore becoming essential that the designers of expert systems are able to select a UMT that is appropriate (Avanzato, 1991).

Bonissone describes a change in view that has occurred in the process of looking for an appropriate method for dealing with uncertainty.

> The search for a normative uncertainty theory to be used in reasoning systems has long been a major driving force in our research community. ..... More recently, these controversies have subsided, and a slightly more tolerant view has emerged. Uncertainty tools have been divided into extensional and intentional approaches, according to their respective focus on computational efficiency or purer semantics[Pearl, 1988]. There has been an increased awareness of classes of problems requiring a

prescriptive rather than a normative approach to reasoning with uncertainty. (Bonissone, 1990, p. 237)

So it is no longer a matter of selecting *the method* of reasoning under uncertainty that is appropriate for expert systems. Rather a matter of selecting the method that is appropriate for *a particular* expert system (or set of similar systems).

This research aims to provide assistance in this selection. Chapter 4 compares paradigms of reasoning under uncertainty and Chapter 5 provides a methodology for the selection of an appropriate method.

2.8 Validation of expert systems -- its implication for UMT's.

Definitions of expert systems often include some mention of the notion that they can function at close to human expert levels (O'Keefe, Balci & Smith, 1987). They are usually expert only in a narrow domain, can produce recommendations, make enquiries to complete gaps in their knowledge and often explain how a conclusion has been reached (Graham, 1991). It is imperative that the expert system is able to supply accurate responses and perform in a manner that is dependable (Guida & Spampinato, 1989).

Quality assurance is a concept that has recently become important in almost every endeavor. Yet in the past, the quality of many systems was rarely tested rigorously to ascertain if a satisfactory level of performance was achieved. As with all other software, expert systems should undergo both validation and verification. Validation means substantiating that the system performs accurately, whilst verification is substantiating that a system has implemented its specification. Yet this terminology does not in itself define a clear methodology that will allow the quality of an expert systems to be assured.

It is likely to be a longer process to validate an expert system that uses reasoning under uncertainty than a system with a crisp reasoning process (Chang & Hall, 1992). In addition the importance of testing any system will depend on the nature of advice given by the system and whether anything critical is at stake if an incorrect decision is made by the system. A critical domain has been defined as one "where the occurrence of inappropriate or incorrect decision may cause damage" (Guida & Spampinato, 1989). It is easy to imaging damage occurring if incorrect decisions were made in many medical and industrial fields.

It is worthwhile considering what it is that is to be validated, especially in an expert system that reasons under uncertainty. Guida and Spampinato in their paper "Assuring adequacy of expert systems in critical application domains" distinguish two fundamental parts of the quality process (Guida & Spaminato, 1989). These are the external behaviour and the internal ontology. The former can be observed as the results of the system but the later deals with the structure (knowledge representation and reasoning algorithms) and content (knowledge base) of the system.

It could be argued for example that an expert system that produces satisfactory (valid) results should be acceptable; MYCIN has received support of this kind (Horvitz & Heckerman,1986). However it is a more widely held view that the internal ontology including the reasoning process itself should be validated, not simply the results it produces (O'Keefe, Balci & Smith, 1987) (Guida & Spampinato, 1989). Some suggest it would be unreasonable to extend the knowledge base or scale up to a larger application domain, a system that had a poor reasoning process. The implication of this argument to UMTs is that they must not only be shown to produce reasonable results but must support a valid reasoning process.

An order has been suggested to this validation process. The inference engine, knowledge acquisition facility and explanation facility should be validated first. This is because these parts of the system are the most procedural and therefore standard methods as used for more general computer systems could be used. This may be the easier part of the validation when compared to validating the knowledge base (Hollnagel, 1989). The performance of the system is heavily reliant on the structure and content of the knowledge base. Consequently the system must be validated continuously throughout the life of the system every time the knowledge base is updated.

The validation process requires some known or expected behaviours to validate against. It is important that these should not have been used in the development of the system. The opinion of an expert or group of experts should normally be used but in some problem domains it may be feasible to use known results instead (Chang & Hall, 1992). There may be a difficulty if the validation fails in determining whether the error is in the expert system or is with the test results themselves.

The validation of an imprecise or fuzzy expert system has an additional level of complexity. Not only must the correct recommendation be made by the system but it must be made with an appropriate strength. In a system using fuzzy logic -- "the fuzzy set defined by the conclusion must be within the acceptable bounds of its possible range" (Chang & Hall, 1992, p. 600).

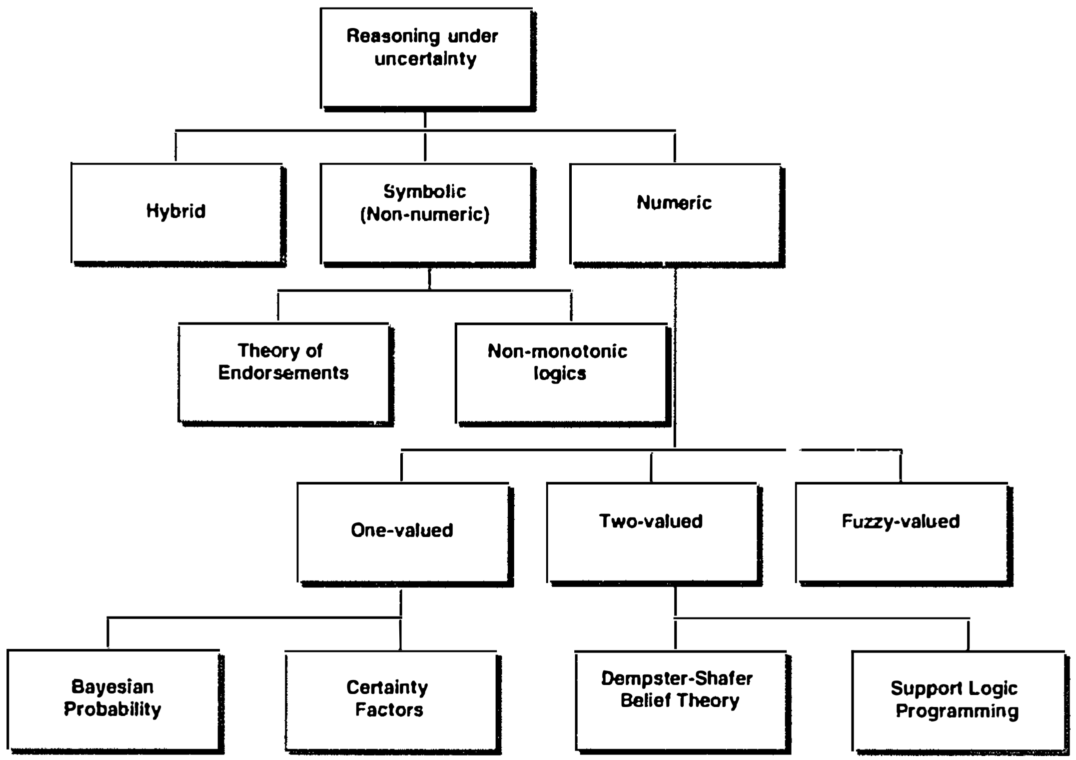
Chapter 3: The theory and practice of uncertainty management

3.1 Chapter overview

This chapter discusses the variety of methods that have been suggested for handling reasoning under uncertainty in expert systems. The UMTs presented here are divided into three main groups, numeric, symbolic and hybrid (which is a combination of the previous two).

In this chapter each technique is presented, advantages and disadvantages of the techniques are discussed and in some cases the relationship to other UMTs is clarified.

An objective comparison of paradigms for reasoning under uncertainty may be found in chapter 4, whilst suggestions as to how to select a specific UMT for a particular application are to be found in chapter 5. Figure 3.1 on the next page shows various classifications of UMTs that will be considered.



F ig 3.1 Paradigms of reasoning under uncertainty

3.2 Numeric approaches

3.2.1 Bayesian

Probability is the oldest and most widely used formalism for representing uncertainty. Shafer and Pearl (1990) explain that the concept of the degree of probability was used "in law and philosophy before mathematical probability was invented" (Shafer & Pearl, 1990). Scholars developed mathematical probability in the late 1600s and early 1700s. James Bernoulli's book *Art of Conjecture* was one of the first books on mathematical probability and from the title it can be seen that he intended the theory as a mechanism for plausible reasoning.

The frequentist view of probability did not emerge until the mid nineteenth century (Shafer & Pearl, 1990). The frequentist view leads to the most widely used description of Classical Probability Theory, that it is used for games involving the toss of a coin and the throw of a dice. From this point of view, probabilities are defined as the proportion in the long run, a frequency interpretation of probability.

The original account of probability is more useful for Expert Systems , this is to interpret probabilities as personal or subjective evaluations (Freund, 1972, p. 36). The probability of a proposition is a measure of a person's degree of belief in it, given the person's current level of information. A probability is the degree of belief in a particular proposition. (Cheeseman, 1986, p. 86). Hunter agrees with this interpretation

> ...for there are decision problems involving uncertainties that cannot plausibly be given a frequency interpretation, but which are really uncertainties about the truth of non-vague propositions.
>
> (Hunter, 1986, p. 209)

Zadeh, the inventor of Fuzzy Sets, (Zadeh, 1986) has the opinion that probability is not appropriately expressive for representing the many kinds of uncertainty that can be found in expert systems. He also believes that most probabilities are not known with "sufficient precision to be representable as real numbers". Zadeh prefers fuzzy terminology such as likely, unlikely and not very likely.

Hunter maintains that Zadeh's interpretation of probability, as being unable to represent vagueness, is incorrect. Hunter also distinguishes between static and dynamic views of uncertainty. He asserts that for a complete theory of uncertain reasoning, the static probability theory must be combined with a dynamic theory, (he suggests Maximum Entropy Theory) which is concerned with how one's degree of belief should change in the light of new evidence.

3.2.1.1 Probability - the basics

Let A be an event. Then S is the set of all possible events called the Sample Space. The probability of event A is denoted P(A). The probability measures must satisfy three given postulates. (Freund, 1972, p. 38)

1. P (A) >= 0 , for any subset A of S.

2. P(S) = 1.

3. If A_1, A_2, A_3, ..., is a sequence of disjoint subsets of S, then

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots$$

Further rules that can be derived include :

a) Probabilities cannot exceed one

b) The probabilities of A and not A, sum to 1.

Postulate 3 is the addition rule for mutually exclusive events. But if events are not mutually exclusive ,

eg. P(student 901087 will pass Intermediate Algebra) and

P(student 901087 will pass Advanced Algebra) then the appropriate rule for addition of probabilities is

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

3.2.1.2 Conditional probability

Any measure of probability is relative to the sample space, thus P(student 901087 will score top grade in Advanced Programming) may vary depending on whether S, the sample space includes students from one or all campuses. To clarify this P(A|S) is the conditional probability of event A relative to the sample space S. (Freund, 1972, p. 51).

When considering two events, A and B. The conditional probability of A given B, is the probability of event A occurring given that event B has occurred. This is defined as

$$P(A|B) = P(A \text{ and } B) \; / \; P(B)$$

or alternatively in the Bayes rule format.

$$P(A|B) = (\; P(B|A) * \; P(A)\;) \; / \; P(\bar{B})$$

Another description of conditional probability that is more directly useful in expert systems is that, the conditional probability of a hypothesis P(H|E) is the probability of the hypothesis in the light of the evidence E (Lee, Grize & Dehnad, 1987).

3.2.1.2.1 Bayes Theorem

Bayes Theorem allows for the calculation of the updated degree of belief in a hypothesis when new evidence becomes available. For a given hypothesis H_k, there is a prior probability that it is true, $P(H_k)$. In the light of new evidence our belief is altered to produce a posterior probability, $P(H_k|E)$ for the hypothesis H_k :

$$P(H_k|E) = \frac{P(H_k)P(E|H_k)}{P(E)}$$

Bayes rule can be used to infer the probability of a disease from the given symptoms, if one has knowledge of the probability of the symptoms given each disease, and the prior probability of each disease (Neapolitan, 1992).

Lee and Clark provide examples of the application of Bayes Theorem in the expert system domain (Lee et al, 1987, p. 18) (Clark, 1990, p.114).

Expert systems can reason through forward-chaining or backward-chaining processes. Bayes Theorem is appropriate for either type of reasoning. Thus if probablilities are more readily available to support reasoning in a certain direction, Bayes Theorem can be adapted to support that direction of argument (Valverde, 1992).

A particular version of Bayes Theorem is pertinent to the question of assessing a set of competing hypotheses in the light of a set of evidence (Valverde, 1992). There are dangers in progressively updating an assessment in the light of a new piece of evidence. It is vital that the interelationships between separate pieces of evidence be considered so that conflicts will become evident rather than be submerged. (Buxton, 1989)

3.2.1.2.2 Independent Events

Two events are independent if the occurrence of one has no effect on the occurrence of the other. Then by definition P(A|B) = P(A) and P(B|A) = P(B).

3.2.1.2.3 Probabilities in Expert Systems

When an expert system is based on Bayesian probability many probability values are required. These will be provided by the domain expert and will be both the estimates of the probability of hypotheses and also the probability of hypotheses in the light of evidence.

Valverde and Gehl (1992) report on an expert system used to determine an accurate diagnosis of the reason for Boiler tube failure in fossil fuel driven power plants. They implemented two systems, one using a Bayesian model and the other using Dempster-Shafer (see Section 3.2.3 Dempster-Shafer belief theory). They report that it was possible to acquire estimates of probabilities from actual data. The historical records of observed causes of failure in the boiler provided the required estimates of relative frequencies (Valverde &Gehl, 1992).

Where historical information is not available the probabilities required by an expert system must be provided by human experts. This will usually include all prior and conditional probabilities. The required amount of data grows exponentially with the number of hypotheses. This, together with the enormous amount of computational effort required when new evidence becomes available provides the reason that full probabilistic representations have not been popular in expert system development (Wise &Henrion, 1986).

Bayes Theorem must be adapted in the case that the evidence itself is uncertain. The changes required are due to the interlinkea nature of system, and the use of Bayes theorem to calculate posterior probabilities (Lee et al., 1987).

They also identify four 'significant drawbacks' in using Bayesian techniques in expert systems (Lee et al., 1987).

1. The subjective nature of the assignments of probability by a domain expert may lead to a set of probabilities that are internally inconsistent. This can be prevented but only through a lot of work on the part of the domain expert and the knowledge engineer.

2. The hypotheses that are used in Bayes theorem are assumed to be disjoint. This requirement may not be practical.

3. If disjoint sets of hypothesis cannot be achieved then the results achieved may not be valid.

4. A single change to the probability of an event requires the recalculation of many probabilities.

Although Bayesian inference is the most common strategy used in expert systems, there are some situations where it is inappropriate (Neopolitan, 1992). Such situations are those in which a probability cannot be assigned to all pertinent events. In this instance, other techniques such as Dempster-Shafer may be appropriate (see Section 3.2.3 Dempster-Shafer belief theory).

The next section considers two early and yet successful expert systems that were based on probability, but in different ways.

3.2.1.3 Some examples of probability based expert systems

In this section the thesis will consider two expert systems that used probability to handle uncertainty. The first expert system, Prospector was reported to be responsible for making a great deal of money. The second, Inferno was able to make recommendations even when provided with certain inconsistent information.

3.2.1.3.1 Prospector

This expert system was intended as an aid to geologists in their search for ore deposits and its fame is based on its success (Dan & Dudeck, 1992).

Prospector's designers intended that the system would provide answers that were "reasonably close approximations" to those that would result from the use of probability analysis (Yadrick et al, 1988, p. 81). The uncertainty handling mechanism is regarded as having a stronger theoretical foundation than that of MYCIN's certainty factors and therefore "it has not been reviewed in a critical way" (Dan & Dudeck, 1992). Prospector used an inference network to identify dependent probabilities. This was an early version of the Bayesian Belief Network (See section 3.2.1.4 Bayesian Belief Networks).

3.2.1.3.2 Inferno

Quilin's Inferno(1983) is a probabalistic inference system that solves some of the problems of earlier systems. Inferno can make inferences in a cyclic way, a desirable but unaccomplished feature in systems like Prospector. Inferno is able to use both forward and backward chaining in its reasoning process (Section 2.2 Expert Systems).

Another uncommon feature of Inferno is its ability to deal with inconsistent information. If information is inconsistent, Inferno can make this fact evident along with some alternative ways that the information could be made consistent. However the appropriateness of this feature has been questioned (Cheeseman 1985).

3.2.1.4 Bayesian Belief Networks

Bayesian Belief Networks in respect to expert systems have their origins in the inference networks of Prospector. They are based on probability theory and have been largely developed to their current form by Pearl. However the Bayesian Network is an annotated directed graph and was first used by the statistician Wright in 1921 for the analysis of crop failure (Heckerman, Mamdani & Wellman, 1995).

A belief network representation consists of two components, a qualitative directed acyclic graph (DAG) that demonstrated the existence of probabilistic dependence between variables and a quantitative set of conditional probability tables for the graph.

A belief network has to be sparse if it is to be comprehensible to the user and inference using the network is to be computationally tractable (Srinivas, Russell & Agogino, 1990). This is a technique which allows for the explicit representation of dependencies as well as independencies, thus allowing the experts to accurately represent their beliefs with respect to some domain. One of Heckerman's objections to Certainty factors was that they did not allow any such explicit representation and thus were ambiguous with respect to dependence or independence (Heckerman, 1986).

The belief network is seen as a realistic approach to building expert systems. It solves some of the short comings of the Certainty factor model whilst not requiring the huge volumes of data of classical probabilty (Heckermann & Shortliffe, 1992).The requirement that the relations between variables be specified by a conditional probability matrix, forces the knowledge engineer to consider the various combinations of variable values (Morowski, 1989).

Booker, Hota and Ramsey (1990) suggest that belief networks solve one of the commonly made mistakes of early systems - the idea that uncertain inferences are modular. They developed BaRT - a Bayesian Reasoning Tool for knowledge based systems, to "make state of the art techniques for uncertain reasoning available to researchers concerned with the classificatory problem solving" (Booker, Hota and Ramsey , 1990, p.280). Its designers claim that BaRT is efficient and practical for real applications.

3.2.1.5 Probability in practice

One of the most common criticisms of the use of probability theory in expert systems is that the theory is impractical to apply in realistic situations. Heckerman and others at Stanford converted Quick Medical Reference (QMR), one of the largest medical expert systems in existence, to a probabilistic framework (Heckermann, 1990). This was largely successful although several assumptions were made which may not always be present in other systems. One of the assumptions was that the variables under consideration (diseases and findings) were binary. A second that diseases are marginally independent is probably the most important in allowing them to limit the total quantity of data required and the least likely to apply in general.

It has been shown that if the requirement to provide an exponential quantity of data can be controlled, perhaps by some feature of the problem domain (as in QMR above), then probability can be usefully applied.

> An algorithm for computing the posterior probability of each disease given a set of observed findings, is presented. Although the time complexity is exponential in the number of positive findings, the algorithm is useful in practice because the number of observed positive findings is usually far less than the number of diseases under consideration. (Heckerman, 1990, p. 163)

There has been much discussion. about the reliability of the probability estimates elicited from human experts, since the earliest expert systems attempted to use numerical representations of uncertainty. This can be seen as less of a problem if these numerical estimates are considered a starting point that will be refined over time as the system is used. The ability of systems that use probabilistic representations to reason and produce reasonable results even with inaccurate numeric assessments can also be seen as a strength of the representation (Spiegelhalter, Franklin & Bull, 1990).

It may be that the amount of research carried out and the stability of the domain may affect the reliability of the probability estimates. Studies have shown that in certain domains probability assessments can be dependable although they may lean to more extreme values (Spiegelhalter, Franklin & Bull, 1990).

In this area, as in many others, the capabilities of modern computers are making it feasible to solve problems that were previously considered not practical. Programs that implement complex algorithms will execute within a reasonable time on today's computers where they could not have been considered practical 20 years ago (Heckermann and Shortliffe, 1992). Heckerman, Mamdani and Wellman report that small powerful computers and GUI interfaces have made Bayesian networks a more common choice for expert system applications in a number of different fields, including diagnosis, forecasting and manufacturing control (Heckerman, Mamdani & Wellman, 1995).

3.2.2 Certainty factors

One of the earliest and most widely used methods for reasoning under uncertainty in expert systems is the Certainty factors (CF) of MYCIN (Shortliffe & Buchanan, 1975), (Dan & Dudeck, 1992). EMYCIN (Empty MYCIN) is an expert system shell which made CF available for other expert system developments. Shortliffe and Buchanan developed CF in the mid-1970s specifically to be used with MYCIN, an expert system for the diagnosis and treatment of meningitis and bacteraemia.

> Our certainty factor model was developed in response to our desire to deal with uncertainty while attempting to keep knowledge modular and in rules. (Buchanan & Shortliffe, 1984, p. 56)

Certainty factors were introduced in the well-known expert system MYCIN and remain one of the most used uncertainty management paradigms. Certainty factors were devised because their creators felt good enough data did not exist to create a full statistical database for the medical application (Ng and Abramson, 1990). There are also indications that the artificial intelligence research community felt that full probability theory would prove too cumbersome (Heckermann & Shortliffe, 1992). At the time some probabilistic diagnostic medical systems used a 'simple Bayes' model. These included assumptions of mutually exclusive and exhaustive hypotheses and conditional independence. The assumptions were made for practical reasons. They made it possible to build diagnostic systems whereas without them the volume of probability estimates required and the complexity of calculations would have been restrictive. It is clear however, that the assumptions were unfaithful to the domain (Heckermann & Shortliffe, 1992).

3.2.2.1 The mechanics of Certainty factors

When Certainty factors are used, the knowledge base has the form:

If <evidence> then CF <hypothesis>.

CFs have values between -1 and 1 which represent the change in belief about a hypothesis given some evidence. Certainty factors were originally defined in terms of probability, a probability of 1 corresponding to a CF of 1 and a probability of 0 corresponding to a CF of -1. A CF of 0 represents the situation of using the prior probability. "Piece wise linear interpolation is used between these three points" (Wise and Henrion, 1986, p. 72).

An inference network of the connecting rules exists and they are combined using parallel and sequential combination as appropriate. These rules were devised by Shortliffe and Buchanan as approximations of related statistical techniques and showed that they satisfied certain intuitive properties. One such property is that parallel combination should be commutative (Heckermann, 1986). If all evidence and hypothesis in the knowledge base are simple propositions then only the serial and parallel combination rules are required.

3.2.2.1.1 Parallel Combination Function

$$CF_3 = \begin{cases} CF_1 + CF_2 - CF_1 CF_2 & CF_1 , CF_2 \geq 0 \\ CF_1 + CF_2 + CF_1 CF_2 & CF_1 , CF_2 \leq 0 \\ (CF_1 + CF_2)/(1 - \min[|CF_1|,|CF_2|] & otherwise \end{cases}$$

For example two pieces of evidence which support the same hypothesis result in a greater certainty factor. For $CF_1 = 0.8$ and $CF_2 = 0.9$ then

$CF_3 = 0.8 + 0.9 - (0.8)(0.9) = 0.98$

3.2.2.1.2 Serial Combination Function

The combination function is used to combine two rules where the hypothesis in the first rule is the evidence in the second rule.

$$CF = \begin{cases} CF_1 CF_2 & CF_1 > 0 \\ 0 & CF_1 \leq 0 \end{cases}$$

3.2.2.1.3 Combination of Rules with Conjunctions and Disjunctions

Suppose the knowledge base contains rules of the form:

R1: if A AND B then C, $CF_1 = 0.9$

R2: if X then A, $CF_2 = 0.7$

R3: if Y then B, $CF_3 = 0.9$

The new composite rule (R4) can be created and the Certainty factors combined with the following combination function:

R4: if X AND Y then C, CF_4

$CF_4 = CF_1 \min/\max (CF_2, CF_3)$

The minimum of the Certainty factors is used for combination in a conjunction and the maximum for a disjunction.

The parallel-combination function appeared in a different form in the earliest of the CF models. The terms Measure of belief (MB) and Measure of disbelief (MD) were used (for positive and negative CF) and the final CF was given as the difference between MD and MB.

Variations to the model have been made in implementations of it. For example in MYCIN Certainty factors of 0.2 or less are treated as if they were 0 (Heckermann & Shortliffe, 1992, p. 39). This meant that where there was very little probability (<0.2) of an hypothesis being used, it would be discarded thus avoiding pointless questions to the user of the expert system.

3.2.2.2 A critique of Certainty factors

The CF model was created for the domain of MYCIN and in blinded evaluations has been shown to provide recommendations for treatment equivalent to, or better than human experts (Heckermann & Shortliffe, 1992).

Two reasons have been suggested for the success of this method of modelling uncertainty. Certainty factors are relatively simple to implement when compared to other methods and the resulting modular knowledge base is helpful to the developer (Dan & Dudeck, 1992).

Heckermann has shown that the statistical definition of CF and the rules for combination show some gross inconsistencies. He suggests that the definition be abandoned in favour of one which works statistically (Heckermann, 1986).

Certainty factors are isomorphic to a subset of probability theory under an appropriate set of assumptions (Rothman, 1989). One of these assumptions is conditional independence of evidence given an hypothesis. This is a very strong assumption and is not the case in all rules in all expert systems. However, this system has been shown to be successful (Yu, Buchanan, Shortliffe et al., 1979). Buchanan and Shortliffe, the creators of Certainty factors wrote, "the motives were largely pragmatic, we justified the underlying assumptions by emphasizing the system's excellent performance"(Buchanan & Shortliffe, 1984).

Horvitz and Heckerman (1986) highlight a misuse of Certainty factors and provide examples of the problem in two well known expert systems . They suggest that the problem stems from the inability to distinguish between a change in belief and an absolute measure of belief.

> Positive certainty factors then, correspond to an increase in belief while negative certainty factors correspond to a decrease in belief. While certainty factors were intended to represent measure of belief update, they were elicited from experts as absolute beliefs. In particular certainty factors were elicited from experts with the phrase "On a scale of one to ten, how much certainty do you affix this conclusion?" (Horvitz and Heckerman, 1986, p. 146)

Certainty factors can produce some apparently illogical results. It is demonstrated that CF values can be the opposite of the conditional probabilities with the following example (Giarratano & Riley, 1994).

	$P(H_1)$	$= 0.8$	$P(H_2)$	$= 0.2$
	$P(H_1 \mid E)$	$= 0.9$	$P(H_2 \mid E)$	$= 0.8$
then	$CF(H_1, E)$	$= 0.5$	$CF(H_2, E)$	$= 0.75$

Since one purpose of CF is to rank hypotheses in terms of likely diagnosis, it is a contradiction for a disease to have a higher conditional probability (P(H|E) and yet have a lower certainty factor, CF(H,E). (Giarratano & Riley, 1994, p. 268).

Adams had reported the same problem using the same example several years earlier (Adams, 1985). It has been shown that these *contradictory* results are quite reasonable (Dan & Dudeck, 1992). The real problem may be that highlighted by Horvitz and Heckerman above, which is whether CF's measure absolute belief or a change in belief. So in the example above, the results are contradictory, if CF's are absolute measures of belief. However if CF's are measures of belief updating then the fact that $CF(H_1, E) < CF(H_2, E)$ results from $P(H_1) > P(H_2)$ and $P(H_1|E) > P(H_2|E)$ should not be surprising. It simply shows that the evidence E has provided for a greater increase in belief in hypothesis 1 than hypothesis 2.

The operational definition of CF is preferred by some researchers. They suggest that this is appropriate since CF's are elicited from domain experts as absolute beliefs, used by the inference engine as absolute beliefs and have results interpreted as if they were absolute beliefs. They suggest changes to the computations of the system to maintain consistency with this(Dan & Dudeck, 1992).

So Certainty factors were elicited without a clear operational definition. However MYCIN performs as well as experts in the field. This suggests that detailed considerations of uncertainty are not critical to the systems performance. Indeed it has been shown that performance does not change significantly when many of the certainty factors in the knowledge base were changed (Heckerman, 1986). Avanzato (Avanzato, 1991) agrees with Heckerman, he states that the CF model has been shown to be equivalent to probability theory with the additional assumption of statistical independence. Adams is cited by Avanzato as concluding that the success of MYCIN despite the theoretical difficulties is

> due to the fact that MYCIN uses short chains of reasoning and simple hypothesis. (Avanzato, 1991, p. 70)

However MYCIN's creators were interested first in getting a system that worked. They were not principally motivated with the mathematical correctness of their UMT but more especially concerned with designing a system that performed well in a particular medical area. Some of the problems may have since arisen by the use of Certainty factors in other domains that are unsuitable for its reasoning process (Horvitz & Heckerman, 1986).

To be more specific, there are features of MYCIN's problem domain that are unusual (Heckerman & Shortliffe, 1992). MYCIN's therapy recommendations are invariant to changes in the CF values, whereas the diagnostic assessments degrade more rapidly. However MYCIN is primarily a therapy advice system and the antibiotics recommended often cover several diagnostic assessments. Thus Heckerman and Shortliffe emphasise that

> "the CF model may be inadequate for diagnostic systems or in domains where appropriate recommendations of treatment are more sensitive to accurate diagnosis. Unfortunately, this point has been missed by many investigators who have built expert systems using CFs" (Heckerman & Shortliffe, 1992, p.36).

3.2.3 Dempster-Shafer belief theory

The belief theory was originally developed by Dempster (Dempster, 1967) and extended by Shafer (Shafer, 1976). It attempts to provide a measurable means of defining the concept of belief, which relates to "our conviction in the truth of some statement" (Valverde &Gehl, 1992). Dempster-Shafer (D-S) theory is a generalisation that was designed to take into account a short coming of probability theory, that is it cannot explicitly represent ignorance. D-S is also able to loosen the requirement for prior and conditional probabilities (Avanzato, 1990).

The essence of the Dempster-Shafer theory is that the language of belief functions is a generalisation of the Bayesian language (Shafer, 1986). Shafer states that a belief-function argument differs from a Bayesian argument in that the former involves a probability model for the 'evidence bearing on the question' whereas the latter involves a probability model for the 'answer to the question'. The belief-function generalisation makes it possible to use certain kinds of incomplete probability models.

So the use of belief-functions allows for the simplification and generalisation of Bayesian probability. Belief-functions concentrate on the evidence and also provide for an upper and lower level of probability. A wide margin between the upper and lower levels expresses explicitly a state of ignorance. This is not possible in classical probability where ignorance is typically represented by a probability of 0.5. This muddles the concept of ignorance with the actual probability of an event, which may indeed be 0.5.

Naturally Dempster-Shafer theory continues to develop as more experience is gained with its use. Yen cited by Avanzato points out that some of the problems with the system have been conquered.

> [YEN, 1989] describes an expert system, GERTIS (General Evidential Reasoning Tool for Intelligent Systems), which extends D-S theory to overcome several of the problems......this is accomplished in part by modifying Dempster's Rule to combine belief updates instead of absolute belief measures.(Avanzato, 1991, p. 70)

A full description of Dempster's Rule of Combination is found in the next section.

3.2.3.1 The mechanics of Dempster-Shafer

The Dempster-Shafer theory is a generalisation of probability theory with its roots in a theory of upper and lower probabilities (Fung & Chong, 1986).

The main difference between Dempster-Shafer Evidential reasoning (also called Belief Calculus) and standard probability theory is the relaxation of the constraint that the probability of an event and the probability of its negation must sum to one.

ie P(X) + P(not X) = 1 in probability theory.

P(X) + P(not X) <= 1 in Dempster-Shafer.

The central concept in this paradigm is that of the *frame of discernment* (*F*). This is similar to the sample space in probability. The elements of *F* are mutually exclusive and exhaustive and can be explained as the solutions to the question at hand (Valverde & Gelh, 1992). In the case of n possible outcomes there are 2^n possibilities, these are all possible subsets of the frame of discernment.

The basic probability function m defines a probability number for each subset of the frame of discernment.

This function satisfies two basic properties :

1. $m(\varnothing) = 0$. The probability number of a null event is 0.

2. The sum of the probabilities of all the other subsets is 1.

eg. If there are 2 suspects of a crime. Bill and Jim then

m(Bill) = .1 Strength of evidence that Bill is guilty

m(Jim) = .2 Strength of evidence that Jim is guilty

m({Bill, Jim}) = .7 Strength of evidence that the culprit is in the subset {Bill, Jim}

From the basic probability numbers, two other measures of a hypothesis can be derived, belief (Bel) and plausibility(Plaus). The belief interval for hypothesis(a), is then given by [Bel(a), Plaus(a)] and the difference Plaus(a) - Bel(a) represents the amount of uncertainty with respect to a (Cortez-Rello and Golshani, 1990, p. 13). Also since

Bel(a) <= Prob(a) <= Plaus(a)

the degree of belief and the degree of plausibility can be regarded as the lower and upper bound on the probability.

The degree of belief in a hypothesis (A a subset of *F*) is the combined sum of the basic probabilities of A and its proper subsets. m(A) is then a measure of belief assigned to A but Bel(A) is assigned to A and its subsets. Those basic probabilities in subsets that constitute Bel(A) are known as the *focal elements* of Bel (Valverde & Gehl, 1992).

The belief of a subset, measures the total belief which includes the belief in supersets.

eg Bel(Bill) = m(Bill) + m({Bill,Jim}). So Bel(Bill) = .8

The plausibility of a subset is (1 - Bel(not A)) e.g. Plaus(Bill) = 1 - m(not Jim) = .8

A frequent critisism of classical probability theory is in the case of little evidence for or against a hypothesis, the sum of the probabilities must still be one. This is not the case using belief functions where Bel(A) + Bel(not A) <= Bel(F) = 1

The probabilities are referred to as Measures of Belief and are combined according to Dempster's rule of combination (Spillman, 1989, p. 47-49). A composite belief function may be generated from two or more belief functions defined over the same Frame of Discernment. If Bel_1 and Bel_2 are two belief functions based on different evidence and $m_1(A)$, $m_2(B)$ and m(C) denote the basic probabilites for Bel_1, Bel_2 and Bel respectively then Dempsters rule of combination is defined as follows:

$$m(C) = \frac{\sum m_1(A_i)m_2(B_j)}{1 - \sum m_1(A_i)m_2(B_j)}$$

There are two perspectives of the Dempster-Shafer theory of Belief functions. The compatibility view and the probability allocation view.

> The compatibility view, interprets the theory of belief functions in terms of a mapping or a compatibility relation between two different but related sets of mutually exclusive propositions (Lingras & Wong, 1990, p. 468).

These two sets respectively provide an upper and lower value and together they define a belief interval. The Bayesian probability, which is considered to be the 'true' value is estimated to be contained within this interval. The second view is the probability allocation view.

> Another view of the theory constructs belief functions based on a body of evidence which is too vague to be described in terms of propositions. The belief functions in the second view are constructed by allocating a certain probability mass to not necessarily singleton sets of possible answers. (Lingras & Wong, 1990, p. 468)

3.2.3.2 Advantages of Dempster-Shafer

Dempster-Shafer theory is seen as a generalisation of probability theory (Fung & Chong, 1986). It explicitly represents measurement of a degree of belief (Valverde & Gehl, 1992), and allows for the explicit representation of ignorance (Spillman,1989). A common problem with the representation of uncertainty is that of effectively combining information from several sources (see Section 2.5 Sources of Uncertainty). Dempster-Shafer through its combining function explicitly provides a solution to this situation (Cortes-Rello & Golshani, 1990).

Along with the advantages of course come some disadvantages.

3.2.3.3 Disadvantages of Dempster-Shafer

A significant disadvantage of D-S theory is that the assumption of independence of evidence is not always realistic (Henkind & Harrison, 1988). This same disadvantage applies to Certainty factors and the early Bayesian Techniques. Bayesian Belief networks however, explicity express dependence by the arcs in the network.

Shafer, one of the creators of the theory points out that the combination rule is pragmatically rather than mathematically based. He indicates that there is no theoretical justification for the combination rule.

> "there is no conclusive a priori argument for Dempster's rule... the rule does seem to reflect the pooling of evidence" (Shafer, 1976, p.57)

The belief interval is defined by the values Bel (The degree of belief) and Plaus (the degree of plausibility). These values have been said to be estimates of the true probability (Lingras & Wong, 1990) but Neopolitan is not sure of their use.

> Bel and Plaus are nebulous entities. They are not probabilities of the event of interest nor the lower and upper probabilities therefore what meaning can we attach to them? (Neopolitan, 1992, p.73).

Zadeh in Giarratano and Riley illustrates a problem with Dempster-Shafer theory by the use of an example that produces unanticipated results (Giarratano & Riley, 1994). The example used is the belief by two doctors, A and B, in a patent's illness. The beliefs in the patient's problem are:

m_A (meningitis) = 0.99

m_A (brain tumor) = 0.01

m_B (concussion) = 0.99

m_B (brain tumor) = 0.01

The Dempster rule of combination gives a combined belief of 1 in the brain tumour. The problem arises in this instance because this is the only illness that is supported by both doctors.

Dempster-Shafer has also been considered to be computationally complex (Lee, 1987). There have been recent improvements to this by Xu and Kennes reported in their paper, "Steps towards an Efficient Implementation of Dempster-Shafer Theory" (Xu & Kennes, 1994).

3.2.3.4 Support Logic Programming

Support Logic Programming (SLP) was developed by J. Baldwin and co-workers (Dubois and Prade, 1990). At first sight this is simply one of the multiplicity of other techniques proposed for handling uncertainty. Careful consideration however reveals that the support pair is very similar to the belief interval [Belief, Plausibility] of the Dempster-Shafer Theory. Dubois and Prade(1990,p21) indicate that this model is in accordance with the theory of evidence- at least mathematically! It is unclear whether Support Logic Programming is a proposed improvement to Dempster-Shafer theory or simply an alternate form of implementation.

In SLP an uncertain statement is expressed as

$$A: [S_n, S_p]$$

where A is an atomic formula in first order logic.

S_n is the degree of necessary support for A

Sp is the degree of possible support for A.

The degree of possible support for A is interpreted as the fact that

1 - Sp is the support for not A.

eg. (0,0) A is certainly false

(1,1) A is certainly true

(0,1) It is unknown if A is true or false.

A voting model is considered when looking to combine information. This solves the problem of combining information from several sources (see Section 2.5 Sources of Uncertainty). It is done by considering the information to be from a number of different expert sources, they each have a vote, and the votes are considered to have equal influence. The proportion of the population voting yes to proposition A is denoted p(A). Baldwin extended the voting model to allow for don't know answers.

For two hypothesis A and B, information required is what proportion of voters support (A and B), (not A and B), (not A and not B) and (A and not B). This clearly depends on the voting behaviour of individuals and different combination rules are presented for the three cases of Independence (see Section 3.2.1.2.2 Independent Events), Mutual Dependence and Mutual Exclusion.

3.2.4 Possibility Theory (Fuzzy sets)

Fuzzy logic is one of the larger class of multi valued logics. They are named multi valued because they allow more values than the simple true and false of classical logic.

The difficulties of representing imprecise information in probability theory led to the development of Possibility Theory. This is an extension, by Zadeh, of his theory of Fuzzy Sets. Possibility Theory replaces the binary logic of probability with a multi valued logic. Lea Sombe (Lea Sombe, 1990) suggests that the logic statement "all A's are B's" should be expressed "the more one is A, the more one is B" in fuzzy logic. Fuzzy logic is able to represent and reason with such terms as *hot, dangerous, a little* and *very much* (Giarratano & Riley, 1994). Neopolitan (1992) explains that whereas probability theory allows us to attach a measure of how uncertain we are of the truth or falsity of a proposition, Fuzzy set theory "deals with propositions that have vague meaning" (Neopolitan, 1992, p.74). When a doctor says that an operation has a 90% chance of improving a patient's condition by 50%, then the 90% represents a probability while the 50% represents fuzzy set membership.

Shenoy has proposed a framework of VBS (Valuation Based Systems) for managing uncertainty in expert systems (Shenoy, 1992a). This framework is general enough to include many of the possible paradigms for managing uncertainty that have been proposed. In another article "Using possibility theory in expert systems", Shenoy shows how possibility theory can be fitted into the framework of VBS. In possibility theory the "basic representational unit is called the possibility function" (Shenoy, 1992). Projection and particularization are the main operations for manipulating possibility functions. Dubois and Prade cited in Shenoy (Shenoy, 1992) pointed out the correspondence between projection and marginalisation, an operation in VBS that corresponds to the coarsening of knowledge; and the correspondence between particularisation and combination, the VBS operations that is used for the aggregation of knowledge.

It has been argued that probability theory is all that is required to deal with uncertainty (Cheeseman, 1986) and therefore Fuzzy sets must simply be expressing a form of probability theory. However this example demonstrates that fuzzy set theory is able to express concepts not applicable to probability theory (a similar example appears in (Neopolitan, 1990)). Consider a cross bred animal, for example a sheep that has a pure bred Marino and a pure bred Dorset for its parents. It is neither a Marino nor a Dorset but has 50% membership in both sets, there is no probability involved.

It has been suggested that there are two ways of using fuzziness in expert systems:

> One method is to provide fuzzy truth values to rules and conditions in their premises,...,, The second approach is to handle uncertainty and imprecision with linguistic quantifiers and the use of fuzzy terms in the condition. eg. If the water level of the river is high and the water level of the river is rapidly rising then prepare to open the gates to the bypass canal. (Chang & Hall, 1992, p.598)

The second approach may be referred to as linguistic logic (Novak,1992). It is however always based on the first approach.

Miyoshi et al. (Miyoshi et al., 1992) have developed an expert system shell that incorporates two different kinds of uncertainty both based on fuzzy logic, a fuzzy production system and a fuzzy frame system. They report that they are working on expert systems in the fields of foreign exchange and image recognition.

3.2.4.1 The mechanics of Fuzzy sets

In (normal) set theory, membership of a set is a boolean value , that is either true(1) or false(0). A *characteristic function* is the established way of showing which objects are members of a set (Giarratano & Riley, 1994).

$$U_A(x) = \begin{cases} 1 & \text{if x is an element of set A} \\ 0 & \text{if x is not an element of set A} \end{cases}$$

An alternate definition is in terms of a *functional mapping*.

$$U_A(x) : x \to \{0,1\}$$

A Fuzzy set may be represented by a generalisation of the characteristic function that is called the *membership function* (Giarratano & Riley, 1994).

$$U_A(x) : x \to [0,1]$$

Although on the surface these two definitions appear very similar, the membership function is a real number between 0 and 1 that represents the *grade of membership* of the fuzzy set.

So in contrast to the crisp sets of standard set theory, Fuzzy set theory allows grades of membership. Imprecise terms such as "short man" can be represented by a Fuzzy set which has a value of 1 (conclusively is a member of the set of short men) for a height of 150cm and a value of 0 (definitely not in the set of short men) for a height of 180cm and is smooth and monotonic between these values (see Figure 3.2 An Example of a Fuzzy membership function, on the next page).

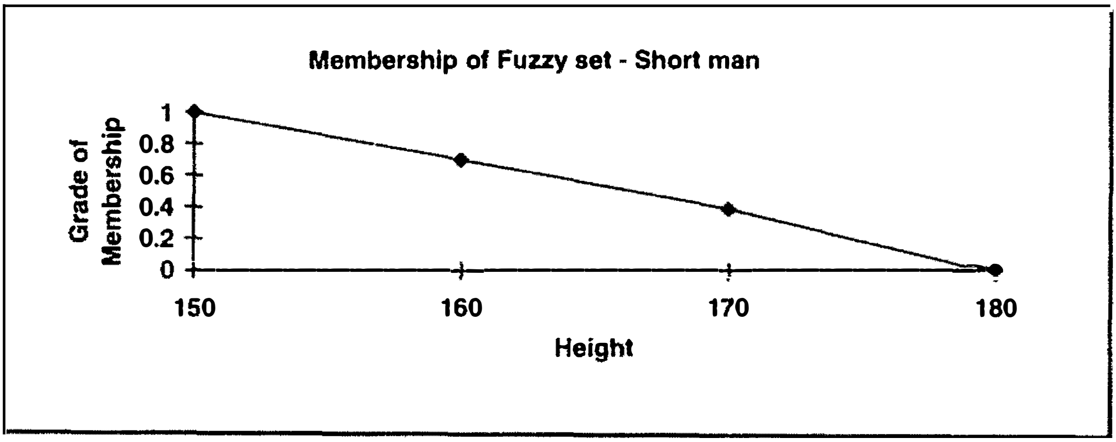


Figure 3.2 An example of a Fuzzy membership function

Fuzzy sets can be combined using the operations of intersection and union. The intersection operation is carried out by taking the minimum value of the two membership functions.

m(short and fat) = MIN(m(short), m(fat))

Whilst the union of two fuzzy sets is found by taking the maximum values.

m(short or fat) = MAX(m(short), m(fat))

The fuzzy set membership function can also be altered by the use of other linguistic terms. In Table 3.1(on the next page), m is a modifier, F is the modified membership function (Lee, Grize and Dehnad, 1987, p. 29), the descriptions are from a different source (Giarratano & Riley, 1994).

m	F	Description
Not	$1 - f(x)$	Negation
Very	$f^2(x)$	Intensification
More or Less	$f^{0.5}(x)$	Dilation

Table 3.1 Fuzzy Qualifiers

3.2.4.2 Approximate Reasoning

Approximate Reasoning was proposed by Zadeh as a mathematical method to model human reasoning with "vague notions present" (Novak, 1992). It contains two kinds of rules, Translation and Inference. The former are used to obtain fuzzy sets from natural language, the latter are to obtain conclusions from premises, that is to carry out the reasoning process.

Fuzzy "if-then" rules have the form:

If X is C_j then Y is S_j

where C_j and S_j are fuzzy sets over X and Y respectively. eg. If the road is quite wet then drive slowly. (Koczy & Hirota, 1992)

In their paper "A fast algorithm for Fuzzy Inference", Koczy and Hirota state that there have been various methods used for the Inference process over a knowledge base that contains fuzzy rules. They investigate two methods, "probably those two which are applied most frequently" and discover that one algorithm has a weak sensitivity in reasoning and has low computational complexity whilst the other has good reasoning but is complex computationally. They go on to suggest a "fast and sensitive" algorithm that itself contains some other minor problems (Koczy & Hirota, 1992).

3.2.4.3 A critique of Fuzzy sets

The main limitations of probability theory is that it is based on two valued logic. An event either occurs or it does not. Another limitation is that probabilities are real numbers. Important issues which Zadeh (Zadeh, 1985, p. 4) says can be dealt with by fuzzy logic and not probability are :

1. The fuzziness of antecedents or consequents

2. Partial match between the antecedents of a rule and a fact supplied by the user. - through the compositional rule of inference and interpolation.

3. The presence of fuzzy quantifiers in the antecedent and/or the consequent

> Fuzzy logic has been shown to be a successful representation for uncertainty in expert system design, some difficulties have been identified....elements of fuzzy set theory ignore some mutual exclusivity requirements and that some distributions and qualifier operations (eg. squaring for 'Very') are subjective in nature and may exhibit inaccuracies[Ng, Abramson, 1990]. (Avanzato, 1990, p.71)

Graham (1991) in his paper "Fuzzy logic in commercial expert systems" considers a number of expert systems and plant and machinery controllers that use fuzzy logic. He cites numerous examples mainly from the U.S.A. and Japan, many of which were, at the time of writing, in experimental form. Graham considered one of the most impressive applications of fuzzy logic to be the automatic train operations system developed by Hitachi for the Sendai municipal subway system. The system optimises fuel consumption and other aspects of system performance through the use of fuzzy rules, such as "if the speed is *far below* the limit then the power notch is selected" (Graham 1991).

Shiraishi (1989) suggest that fuzzy reasoning was useful in the development of an expert system for damage assessment.

1. By introducing the fuzzy set manipulation system into the expert system, it is possible to utilize the knowledge and rules which are expressed in terms of natural language.
2. Based on fuzzy reasoning, it is possible to reduce the number of rules necessary for deriving a meaningful conclusion. The reduction is very useful for building a practical expert system (Shiraishi, et al., 1989, p. 216).

It has been shown that as well as being applicable to rule-based expert systems, Fuzzy sets may also be applied to connectionist expert systems that is, those based on Neural networks (Cohen & Hudson,1992).

Some have claimed to have demonstrated that fuzzy logic has been incorrectly used in problems that are examples of uncertain inference (Cheeseman, 1986). Further claims have been made that fuzzy set theory can be subsumed by Bayesian probability. Others disagree and maintain that fuzzy set theory addresses a "fundamentally different class of problems from that of probability theory". This section concludes with a very useful example that aims to demonstrate this difference (Neopolitan, 1992, p.77).

Suppose we have the constraints:

C1 = "X should be close to 4"

C2 = "X should be close to 6"

and the goals

G1 = "X should be close to 5"

G2 = "X should be close to 3"

If we are restricted to the set of integers, these constraints and goals can be represented by the fuzzy sets in Table 3.2 If we take we obtain the fuzzy set D this is our decision. Since no X has full membership in D , we can define our optimal decision as being X that maximizes D. In this case that is X equal to 5. This problem has nothing to do with uncertainty, it is preferably called approximate. (Neopolitan, 1992, p.78)

	1	2	3	4	5	6	7	8	9	10
	0	0.1	0.4	0.8	1.0	0.7	0.4	0.2	0	0
	0.1	0.6	1.0	0.9	0.8	0.6	0.5	0.3	0	0
	0.3	0.6	0.9	1.0	0.8	0.7	0.5	0.3	0.2	0.1
	0.2	0.4	0.6	0.7	0.9	1.0	0.8	0.6	0.4	0.2
	0	0.1	0.4	0.7	0.8	0.6	0.4	0.2	0	0

Table 3.2 The fuzzy set membership in G1, G2, C1, C2 and D

3.3 Symbolic approaches

Symbolic approaches to handling uncertainty are also referred to as *non-numeric* or *qualitative* methods (Graham,1991). In addition the term *plausible reasoning* is often used. It has been defined as "reasoning that leads to uncertain conclusions because its methods are fallible or its premises are uncertain" (Shafer&Pearl,1990). Plausible reasoning has not developed a typical language because formalisations have been absorbed by probability theory. (This lack of typical language has caused difficulty when researching this topic for this thesis.)

Symbolic treatments of uncertainty are seen to have advantages and disadvantages in relation to numeric methods. Symbolic methods generally have "strong explanation capabilities" but that their fragility is in the combining of evidence (Avanzato. 1991, p. 71). This can be clearly seen in the ability of the theory of endorsements to provide its reasons for believing (or disbelieving) in an hypothesis (see Section 3.3.1 Theory of Endorsements).

Symbolic representations are also more suitable to handle uncertainty of particular types. The ability to reason with incomplete information has been identified as a strength but researchers have suggested that symbolic methods are unable to cope with imprecise information, "since they lack any measure to quantify confidence levels" (Bonissone & Decker, 1986, p. 218).

Other researchers use different terminology. Methods for handling uncertainty have been categorised as either quantitative or qualitative approaches (Graham, 1991). The qualitative methods are said to vary from those that hide uncertainty in linguistic terms to those that provide intricate methods using nonmonotonic logic or endorsements.

Sullivan and Cohen (1990) argue against the use of numbers to represent uncertainty. Their argument has been summarised to the following points :

1. Subjective degrees of belief do not behave as probabilities
2. Experts are uncomfortable in committing themselves to numbers
3. In some situations the accuracy has little effect on performance
4. Numbers tell us how much to believe, not why to believe (Sullivan & Cohen, 1990).

The theory of endorsements provides an answer to the concerns about the use of numbers to represent uncertainty by providing a clear alternative. Although, as will be explained in the next section numerical measures are not eliminated completely.

3.3.1 Theory of Endorsements

The main principle behind this uncertainty management system is to avoid the use of numbers to represent uncertainty. Cohen believes that where numbers are used to represent imprecise information, that they act as a summary of several different aspects of uncertainty, (see Section 2.5. Sources of Uncertainty above) and therefore information is lost.

The Theory of Endorsements explicitly records the reasons for believing or disbelieving a proposition. This method would appear to be closer to the actual method of reasoning used by human experts, who would simply endorse their belief in a statement by a list of reasons. Clark indicated that endorsements may be divided into five classes: rules, data, task, conclusion and resolution endorsements, however no detail on how the types compare is included (Clark, 1990).

Bonissone suggests that there are possible problems in the combinatorial explosion of information required.

> a set of rules is needed to propagate endorsements over inferences
>
> ..combination of endorsements in a premise, propagation of endorsements to a conclusion, and ranking of endorsements must be explicitly specified for each particular context (Bonissone, 1987, p. 859)

Although the process of endorsement is similar to the recording of justifications in truth maintenance systems (TMS), there is an important difference (de Kleer, 1984) (Cohen and Grinberg, 1988). TMS are discussed briefly in section 3.4.2. In the TMS the kind of support for a justification is irrelevant. However, endorsements consider the aspects of inferences that are relevant to reasoning about their certainty. Endorsements can be ranked. The user would have more confidence in an hypothesis with a higher ranked endorsement.

Clark suggests that the motivation for Cohen's Theory of Endorsements is the realisation that the composition of reasons to believe or disbelieve produce the level of certainty. (Clark, 1990).

The Theory of Endorsements developed by Cohen has been implemented in the Expert System shell appropriately named Solomon. The system is working in an ES to advise on portfolio investments (Bhatnager and Kanal, 1986, p. 14) and provides a natural approach to uncertainty although it still has difficulties to be overcome.

Bhatnager and Kanal (1986, p. 15) explain that

> When Solomon derives inferences using rules, all the endorsements carried by the antecedents are transferred to the consequent. The endorsements of the rule, the tasks, the data and the conclusion are all included in the endorsement of the consequent. Since such conclusions are used to prove other tasks, SOLOMON builds up huge bodies of endorsements for conclusions only after a few inferences.

There are still limitations to the theory of endorsements. Although combining evidence and ranking propositions are important in controlling inference, these operations are not readily available when using the theory of endorsements. A limited ranking of endorsements would be reasonable to consider when using combination rules but given a large number of endorsements it is not clear how combination could be performed. Cohen has not provided an answer to the question, "How do experts combine evidence?"

Cohen has pointed out that : "The model of endorsements does not preclude endorsements that include numerical measures such as degrees of belief" (Cohen, 1985, p. 53).

Grech and Sammut describe an expert system shell that was used to implement a system for the identification of radar emitters (Grech & Sammut, 1989). They suggest that the shell was developed for dynamic domains in which "the use of probabilities is highly questionable". As a result the shell uses a combination of an assumption-based truth maintenance system and a system of endorsements to enable it to reason under uncertainty. One of the features of the system which is important is that it "enables problem solving to occur incrementally as new information concerning the state of the world is acquired" (Grech and Sammut, 1989, p. 308).

3.3.2 Non-monotonic logics

There are three broad categories of Non-monotonic reasoning (Bonissone, Cyrluk, Goodwin and Stillman, 1990)

1. consistency - such as McDermott and Doyle's non-monotonic logic and Reiter's default logic.

2. minimization - circumscription, McCarthy(1980).

3. epistemology - autoepistemic logic. Moore(1983).

According to Cohen (1985) non-monotonic reasoning was first applied by Stallman and Sussman in 1977 in a system for electronic circuit analysis. Reiter (Reiter, 1987) describes non-monotonic reasoning as a "particular kind of plausible reasoning". He explains that most examples of such reasoning are of the kind: "*Normally, A Holds.*" This type of reasoning then is different to the predicate logics. Lea Sombe (1990) suggests that the logic statement "all the A's are B's" should be expressed "an A is a B, up to exceptions" in Reiter's default logic.

Traditional mathematic logic does not provide for reasoning with incomplete information as it is inherently monotonic. This means that whenever we have a relationship between a set of sentences (S) and a conclusion (c) such as S implies c then including new sentences in the antecedent will not change the conclusion. Reiter sums this up as "new information, preserves old conclusions." (Reiter. 1987).

As an example of default reasoning, suppose we know of a bird called Pengui and wish to know whether it is capable of flight. A non monotonic logic holds the following rules:

1. " if x is a bird and failing any evidence to the contrary then assume x can fly"

2. "if x is a penguin then x cannot fly"

3. "if x is an ostrich then x cannot fly"

If our knowledge of Pengui is incomplete but we know that she is a bird then we must assume from 1 that she can fly. If we later discover that Pengui is in fact a penguin then we must revise our assumption. It is quite clear that classical logic is inadequate to represent this type of logical mechanism because here adding information has changed the original conclusion.

Bonissone and co-workers (1990) suggest that non-monotonic logic allows a more natural form of reasoning, it mirrors more closely the manner that most people reason

> .. we are constantly making assumptions about the world and revising those assumptions as we obtain more information. Informally the common idea of non-monotonics is that we may want to be able to jump to conclusions. which might have to be retracted as new information about the world becomes available.
>
> (Bonissone, Cyrluk, Goodwin & Stillman, 1990, p. 69)

Non-monotonic logic does not manage without the use of numerical measures of uncertainty by magically transforming uncertainty to certainty. McDermott and Doyle (1980) explain

> the purpose of non-montonic inference rules is not to add certainty where there is none, but rather to guide the selection of tentatively held beliefs in the hope that fruitful investigations and good guesses will result. This means that one should not a priori expect non monotonic rules to derive valid conclusions independent of the non-monotonic rules. Rather one should expect to be led to a set of beliefs which, while perhaps eventually

> shown incorrect, will meantime coherently guide investigations. (McDermott & Doyle, 1980, p. 42)

So non-monotonic logics are useful in situations that are uncertain because of the lack of (incomplete) information but are not able to deal with probabilistic or fuzzy reasoning. Researchers have criticised non-monotonics for this inadequacy (Bonissone, 1987). Others however have elaborated on this aspect of default reasoning and demonstrated that non-monotonics perform a different type of reasoning under uncertainty and are therefore not in competition with the other methods of reasoning under uncertainty (Clark, 1990).

> At any point in time, propositions are considered to be true or false, but no degrees of credibility are permitted. So using a nonmonotonic logic it is not possible to deal comprehensively with partial information about an event. (Clark, 1990, p. 129)

Nonmonotonics were developed to deal with uncertainty resulting from incomplete not partial information.

Reiter (1987) suggests that there are two basic approaches to diagnostic reasoning. The experimental approach is dominant and uses rules of thumb, statistical intuition and past experiences of human experts.

> The second approach diagnosis from structure and behaviours, the only information at hand, is a description of some system together with an observation of that system's behaviour. If this observation conflicts with intended system behaviour then the diagnostic problem is to determine which components could by malfunctioning account for the discrepancy between observed and correct system behaviour. Since components can fail in various and often unpredictable ways, their normal or default behaviours should be described. These descriptions fit the pattern of plausible reasoning. (Reiter, 1987, p.638)

It has been noted that several distinct versions of Reiters default logic (DL) were suggested between 1988 and 1991. The main work published by Reiter on this topic was in 1980 (Reiter, 1980). This work is all entirely theoretical and appears not yet to have been applied to expert system development so it will not be further investigated here (Giordano & Martelli, 1994).

Marvin Cohen sums up non-monotonic logic as a "computationally efficient method for reasoning with incomplete information" (Cohen,1985). He also suggests that the features of non-monotonic reasoning make it particularly suitable for 'meta-reasoning', that is the process of controlling the application of the uncertainty calculus. This idea will be revisited in Chapter 5.

The method of Reasoned Assumptions is another form of non-monotonic logic. Uncertainty embedded in an implication is removed by listing all the exceptions to that rule. Like other non-monotonics Assumption-based systems can cope with the case of incomplete information, but they are inadequate to handle the case of imprecise information with reasoned assumptions (Bonissone, 1987).

It has been suggested that the essential difference between numeric and non-numeric approaches to uncertainty is that in numeric approaches each piece of evidence may be believed to only a partial extent whilst the reasoning may have a "high degree of confidence". This can be contrasted with non-numeric approaches where each piece of evidence is completely believed or disbelieved and confidence in the reasoning is based on the underlying assumptions (Bhatnagar & Kanal, 1986).

So in non-monotonic reasoning before inferencing can be performed assumptions have to be made (or defaults assigned). The results obtained may be later revised in the light of new evidence.

3.4 Hybrid approaches

Hybrid approaches to the handling of uncertainty use aspects of both numeric and symbolic reasoning. The aim is to take advantage of both methods of reasoning, and to combine these advantages into a single method of reasoning.

3.4.1 The Non-Monotonic Probabilist

Cohen (1985) developed the Non-Monotonic Probabilist (NMP), a hybrid approach to uncertainty, specifically for the field of image analysis. A domain that he describes as requiring an "explicit and valid quantitative model of uncertainty", and "a metastructure of qualitative reasoning", in which the conjectures of the model are reconsidered in the reasoning process. This method was introduced by considering the handling of conflict resolution in numeric and non-numeric paradigms. It is suggested that Bayesian methods (and all other numeric methods) actually expect divergence occasionally and because of this, the line of reasoning is similar to that where extreme measurements are expected to "cancel each other out". This perspective is quite different to the qualitative viewpoint where contrary evidence can only occur as a result of flawed knowledge, thus the response is to identify the mistake(s) in the argument and correct it (them). Cohen explains

> Pure probabilistic systems never learn anything new about their probabilistic beliefs and assumptions from the experience of applying them. Pure non-monotonic systems do learn, but they have an arbitrariness and an all-or-none quality about the new beliefs they acquire. Our argument, quite simply, is that both capabilities are needed, and that satisfactory systems will, in general require their combination" (Cohen, 1985, p. 3.18).

NMP is an expert system building tool that incorporates hybrid methods for reasoning under uncertainty. It uses Shaferian belief rather than Bayesian probability because of the possibility for the explicit representation of ignorance (Cohen, 1985).

3.4.2 Truth Maintenance Systems

Truth Maintenance Systems(TMS) are identified by de Kleer as having the problem of only considering one solution at a time (de Kleer, 1984). However Assumption-Based Truth Maintenance Systems (ATMS) allow "arbitrarily many contradictory solutions to coexist" (de Kleer, 1984, p. 81). D'Ambrosio discusses an hybrid approach to reasoning under uncertainty using ATMS:

> "the method relies on the propagation mechanisms in an ATMS to perform most evidence combination operations symbolically, and only substitutes numeric values when asked for the certainty of a proposition" (D'Ambrosio, 1989, p. 268).

Advantages of this technique include improved handling of dependent and partially independent evidence, rapid re-evaluation of propositional certainty values with different sets of assumption certainties, and the ability to obtain certainty values for a variety of different perspectives (partial solutions) with little computational effort (D'Ambrosio, 1989, p. 282).

Filman adds further weight to the argument, that this type of assumption-based reasoning is more similar to most human reasoning, than that of traditional logic.

> In general, reasoning is the process of deriving new knowledge from old. If the underlying knowledge never changes, if we never explore hypothetical spaces, and if our knowledge is free of internal contradictions, the accumulation of knowledge is straight forward: We just add the results of our reasoning to our pile of knowledge. Unfortunately, few problems are so simple. We usually find ourselves reasoning under a set of assumptions that may be withdrawn or changed. Often the entire reasoning process is focused on identifying preferred assumption sets. Ideally when the assumptions change we would like to withdraw those conclusions that are no longer valid, retaining those that are still true. This requires attaching to derived facts justifications or dependencies, that is, reasons for belief in

these facts (Filman, 1988, p.384).

So with this method of reasoning we may concentrate on the assumptions that drive the process of reasoning. Doyle's system uses the concept that certain assumptions are either believed or not believed. "A particular derivation would be valid, for example, if assumptions X and Y were in, but Z out" (Filman, 1988, p.384).

Assumption based Truth Maintenance Systems then provide for many contradictory solutions to be held and a natural way of reasoning.

Chapter 4: A comparison of uncertainty management techniques

4.1 Chapter overview

This chapter will attempt to compare uncertainty management techniques. The chapter is in four parts. The first introduces the concept of comparison and contains a warning for the expert system developer. The second part of the chapter considers three methods that have been suggested in the literature to perform the comparison. Each suggests features that are important in the comparison. The first (Wise & Henrion, 1986) considers the results of the expert system to be of paramount importance. The second (Cohen, 1985) is in effect a cost benefit analysis that suggests it is important to weigh up the Validity (Benefits) and Feasibility (Costs). The third method of comparison (Bonissone, 1987) is in the form a Desiderata for uncertainty management techniques. Each UMT is classified on whether or not it meets each of thirteen objectives.

The third part of the chapter considers comparisons of UMTs that have been reported in the literature. This section is dominated by numeric UMT's with little on Non-numeric and barely a mention of hybrid methods reflecting the amount of material available. Hybrid methods especially are in their infancy and therefore are rarely mentioned beyond the hope that they may provide for a better method for the future. The complexity of such implementation is a limiting factor for the present.

Finally the chapter considers some of the recent advances in the theory of reasoning under uncertainty.

4.2 An introduction to comparison and warning

Ginsberg advised that comparing Uncertainty Management Techniques was in the too hard basket. "The true advantages of the various competing paradigms will only be apparent when these paradigms have been incorporated in full-scale systems" (Ginsberg, 1986). Even then, a method of performing the comparison, or a scale along which the performance of the UMT's is to be measured may be difficult or inappropriate to find.

There appears to be a trend in the literature that identifies a shift in belief over the years from the mid-eighties to the early nineties. The start of the period is characterised by claims that certain UMT's are the one and only correct system; eg. Cheeseman (Cheeseman, 1986) argued in favour of probability and Zadeh (Zadeh, 1986) in favour of fuzzy set theory.

The following quote from Shafer was ahead of its time and is far more characteristic of the early nineties.

> I believe that in the next few years both Bayesian and belief-function designs will find their niches in the world of expert systems. Bayesian designs will predominate in systems that are repeatedly applied under conditions so constant that the picture of answers determined at random with known chances fits. Belief-function designs will be more successful in systems whose each use represents a relatively unusual conjunction of different small worlds of experience (Shafer, 1986, p. 135).

4.2.1 Warning to the expert system developer, your UMT may not be what it claims

Magill and Leech (1991) investigated two commercially available expert system tools that used respectively, Bayes' Theorem and Certainty Factors, for the handling of uncertainty. Their aim was to recommend the more appropriate tool for a particular task.

They discovered that the complex decision of which UMT was more appropriate for an individual ES development was further hampered by the fact that "the two specific tools do not follow strictly the theories on which they are based".

This matter is beyond the scope of this investigation since it was never intended to investigate particular implementations. It is included here as merely a warning to expert system developers that it is possible that the methodology selected may not be implemented in its purest form.

Returning to the comparison of UMTs, is this a matter of comparison of apples with oranges? If so, then when should the apple be selected for a particular application ahead of the orange.

4.3 Methods for comparing UMT's

Before beginning to compare UMTs, the manner in which they are to be compared should be considered. This section will consider three methods of comparison that have been suggested. The first suggests accuracy of results (Wise & Henrion, 1986), the second suggests a framework of features for comparison - it is in essence a cost benefit analysis (Cohen, 1985), and the third is a Desiderata - a list of requirements (Bonissone, 1987).

4.3.1 Comparison using results produced

Wise and Henrion in 1986 felt that it was important to test Uncertain Inference systems (UISs, usually elsewhere in this paper referred to as paradigms of reasoning under uncertainty or UMTs) in respect of the results they produced whilst acknowledging that other aspects were important.

> The main purpose of this paper is to present and try to justify a framework for testing the accuracy of UIS's results, ignoring for the moment issues of computational effort, clarity, or simplicity. ...we believe that clearer presentation of these fundamentals and examination of the methods against the full range of criteria, including the theoretical, pragmatic issues, as well as the experimental comparison of performance explored here, could shed some needed light. (Wise & Henrion, 1986, p.82)

Despite the drawbacks of this method of comparison they consider that information regarding the accuracy of outcomes from the expert system will assist the expert system developer in making a selection.

> Different people will have different weightings for these criteria, reflecting their different goals, and so it may never be appropriate to attempt definitive evaluation of the techniques. But in any case, better analytic and experimental evidence which compares the performance of UIS's in terms of their results, should help to provide system designers a more solid basis for choosing among them (Wise & Henrion, 1986, p.82).

It is true that the system designer does need to know that a certain UMT provides reasonable results, but it is not likely that this is to be the overriding selection criterion on every occasion. Accuracy of result is only one of several possible criteria for the selection of a paradigm. Other criteria, including the feasibility of implementing a particular paradigm are considered by the framework suggested by Marvin Cohen and outlined in the following section (Cohen, 1985).

4.3.2 A framework for evaluating paradigms

A framework for evaluating theories of uncertainty is presented by Marvin Cohen (Cohen, 1985, p. 2-4). He suggests that the framework:

- provides an opportunity to clarify our comprehension of the task.
- suggests ways in which models may be changed.
- possibly provides the structure on which to build new inference methods.

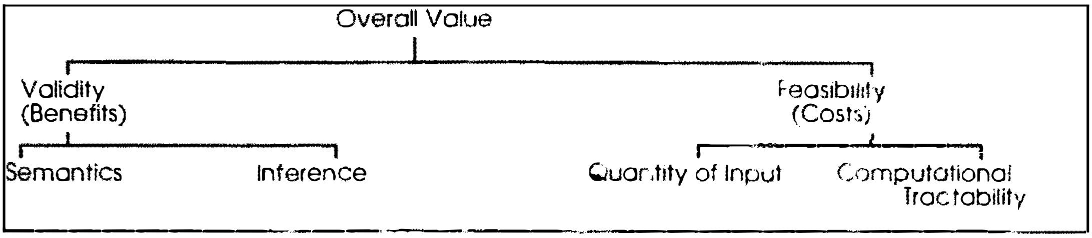


FIGURE 4.1: A Framework for Evaluating Theories of Uncertainty

This framework is illustrated in Figure 4.1. It provides for a number of features of the various paradigms to be evaluated without specifying which of the features is the most important. This will depend on the specific development being undertaken, especially the importance of the overall system and the available budget. A developer of a low budget system may not have the luxury of selecting validity as their most important criterion. They may make a different decision because of time constraints or limited equipment and be unable to consider more complex methods.

4.3.3 The requirements of a theory of uncertainty management

Bonissone (1987) presents a Desiderata for Uncertainty management techniques in a paper entitled Plausible reasoning. This consists of a set of thirteen objectives for an UMT. It concentrates on the theoretical aspects rather than the practical, twelve of the thirteen objectives are related to the *validity* rather than the *feasibility* of the UMT. Nine UMT's are evaluated in respect of whether or not they meet these objectives. Each of these objectives will be discussed in turn. Bonissone's results are presented in Table 4.1.

1. Combination rules should not be based on global assumptions of evidence independence.

Certainty factors are said to have this independence assumption. Heckerman (1986) objected to Certainty factors on the grounds that they did not allow explicit representation of dependence or independence. The early Bayesian methods also made this assumption.

2. The combination rules should not assume the exhaustiveness and exclusiveness of the hypotheses.

Given this assumption there could quite clearly be inaccuracies in a system that had not included all possible hypotheses if it were using probabilities of a variation. Since all probability is relative to a sample space (Freund, 1972).

3. There should be an explicit representation of the amount of evidence for and against each hypothesis.

Since it is "the amount of evidence" that is to be represented, then a numeric representation would be appropriate but cannot be combined into a single figure.

4. There should be an explicit representation of the reasons for and against each hypothesis.

Cohen's Theory of Endorsements provides the most explicit representation of the reasons to support of an hypothesis or not. This is one of Cohen's major arguments against numeric methods -- that they mask the reasons by simply combining them into a number.

5. The representation should allow the user to decide the uncertainty of any information at the available level of detail (i.e. allowing heterogeneous information granularity).

It would appear pointless to insist that the user provide numbers (uncertainty levels) that are not known. These must be only *guess-timates*. It would be better to allow information that is actually known even if less detail is therefore provided. This is the sole objectives to consider the feasibility of the system, specifically the quantity of input (see Figure 4.1: A Framework for Evaluating Theories of Uncertainty).

6. There should be explicit representation of consistency.

7. There should be an explicit representation of ignorance to allow noncommittal statements.

Proponents of Dempster-Shafer (that includes an upper and lower limit to allow for representation of ignorance) argue that this is one of the major limitations of probability theory.

8. There should be a clear distinction between a conflict in the information (violation of consistency) and ignorance about the information.

9. There should be a second order measure of uncertainty recording the uncertainty of the information as well as the uncertainty of the measure itself.

10. The representation must be, or appear to be, natural to the user to facilitate graceful interaction, natural to the expert to permit, elicitation of consistent weights or reasons, and the semantics of procedures for

propagating and summarising information must be clear.

Graham (Graham, 1991) suggests that since it is true that people are generally very bad at estimating probabilities then the Bayesian approach is not suitable for systems to be used by non-statisticians.

11. The syntax and semantics of the representation should be closed under the rules of combination.

12. Making pairwise comparisons of uncertainty should be feasible as these are required for decision making.

In general this type of comparison is possible with numeric values, but not symbolic. Proponents of symbolic methods argue that this comparison may be invalid.

13. The traceability of the aggregation and propagation of uncertainty through the reasoning process must be available to resolve conflicts of contradictions, to explain the support of conclusions, and to perform meta-reasoning for control.

This support for the reasoning process is available with symbolic methods and not with numeric methods.

With the final two requirements of his Desiderata, Bonissone has dismissed, in general, all numeric and symbolic methods. This leaves only the hybrid methods as options to be further considered when looking for a method of uncertainty representation that passes all his stipulations.

Table 4.1 of the next page summarizes how Bonissone sees various UMTs in relation to his Desiderata. He considers seven numeric and two non-numeric systems (Reasoned assumptions and The Theory of Endorsements).

Uncertainty Representation	1	2	3	4	5	6	7	8	9	10	11	12	13
Modified Bayesian	N	N	N	N	N	N	N	N	N	Y	Y	Y	N
Confirmation	N	Y	Y/N	N	N	N	Y	N	N	N	N	Y	N
Upper and lower probabilities	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N
Evidential reasoning	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N
Probability bounds	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N
Fuzzy necessity and possibility	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N
Evidence space	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N
Reasoned assumptions	Y	Y	N	Y	N	Y	N	N	N	Y	Y	N	Y
Endorsements	Y	Y	N	Y	N	Y	N	N	N	Y	Y	N	Y

Table 4.1 Bonissone's view of Uncertainty Representations

4.4 Comparisons reported in the literature

In this section the thesis will consider comparisons that have been made between various UMTs. There are three major sections here, reflecting firstly the overall comparison between numeric and non-numeric methods in general, then two more specific sections dealing with each of those in turn. The number of comparisons of numeric methods found in the literature far outweighs that of the non-numeric methods. This reflects the quantity of research in each area at this point in time.

4.4.1 Numeric versus non-numeric

Bonissone and colleagues argue in favour of the use of a numerical representation of uncertainty on the grounds that this provides a method that can be used in the inference engine.

> With numerical representations, it is possible to define a calculus that provides a mechanism for propagating uncertainty throughout the reasoning process. The use of aggregation operators provides summaries which can then be ranked to perform rational decisions. (Bonissone, Cyrluk, Goodwin & Stillman, 1990, p. 69)

They continue to suggest that models based on qualitative approaches are usually designed to handle the aspect of uncertainty derived from the incompleteness of the information. Doyle's method of Reasoned Assumptions (Doyle, 1983) and Reiter's Default reasoning (Reiter, 1980) are examples of these. With a few exceptions they are generally inadequate to handle the case of imprecise information, as they lack any measure to quantify confidence levels.

Graham (Graham, 1991) suggests that since it is true that people are generally very bad at estimating probabilities then the Bayesian approach is not suitable for systems to be used by non-statisticians.

However Bonissone in an earlier paper (Bonissone, 1987) had argued that numerical approaches to uncertainty required precision that an expert could simply not provide. He considered that the complex calculations may not be justified given the difficulty of obtaining accurate initial figures.

> it is clear that these models of uncertainty require an unrealistic level of precision that does not actually represent a real assessment of uncertainty. (Bonissone, 1987, p. 73)

One of the very interesting expert systems of recent times is Cyc. This enormous system is designed to capture common sense. Lenat and Guha (1990) published a "snapshot of research in progress" at the half way point in the ten year project. Certainty factors were initially used in the project but were not popular with the authors. They criticised CFs because of the "problem" that all numbers could be compared. This meant that unreasonable comparisons could be made between very similar CFs, numbers that really should not be compared. The Cyc project abandoned CFs in favour of a system having only five possible values, absolutely certain, currently believed true (but capable of being overridden), unknown, currently believed false (but capable of being overridden) and absolutely impossible. Lenat and Guha suggest that this method works well when there "isn't too much semantic knowledge" and fails generally when some knowledge is missing (Lenat & Guha, 1990, p. 307). Since no other detail of Cyc's uncertainty handling has been located, this method is not included in the section on symbolic approaches (Section 3.3 Symbolic approaches).

Henkind and Harrison surveyed four numeric UMTs and concluded that although each had its strong points they saw the common disadvantage that they "compute aggregate numbers but keep no record of divergent opinion" (Henkind & Harrison, 1988, p.713).

Bonissone suggests that non-numeric UMTs have deficiencies in their ability to "represent and summarize" measures of uncertainty (Bonissone, 1987, p. 860). Yet he also points to restrictions in some of the numeric representations of uncertainty.

> The numerical approaches tend to impose some restrictions on the type and structure of the information (e.g. mutual exclusiveness of hypotheses, conditional independence of evidence) (Bonissone, 1987, p.860).

Most numerical UMTs represent uncertainty as an exact quantity (scalar or interval) on a given scale. They direct the user or expert to provide an accurate and consistent numerical assessment of the uncertainty of both the facts and rules in the knowledge base. The results of these systems are produced by lengthy calculations guided by well-defined methods and appear to be equally accurate. However, given the difficulty in obtaining such numerical values from the user, "it is clear that these models of uncertainty require an unrealistic level of precision that does not actually represent a real assessment of the uncertainty" (Bonissone,1987, p.860).

4.4.2 Comparisons of numeric UMTs

Wise and Henrion cited by Ng and Abramson(1986, p. 44) compared the performance of different schemes using the same set of rules and data. They report Bayesian networks produced better results than both fuzzy sets and Certainty factors, which were on a par. Heckermann (1990, p. 283) reported that the Bayesian approach outperformed Dempster-Shafer and Certainty factors in a large scale system in the domain of lymph-node pathology. Unfortunately in neither case were the criteria for measurement clearly stated. Hence it remains indeterminate as to how one UMT outperformed the others.

	Bayesian Probability	**Dempster-Shafer**	**MYCIN's Certainty factors**	**Fuzzy Set Theory**
Theoretical Background	*Strong*	*Strong*	*Weak*	*Moderate*
Computational Complexity	*Low*	*Moderate*	*Low*	*Moderate*
Model Setup	*Moderate*	*Moderate*	*Low*	*Moderate*
Model Execution	*Low*	*Moderate*	*Low*	*Moderate*
Complexity of Theory	*Low*	*Moderate*	*Low*	*Moderate*
Ease of Application	*Easy*	*Difficult*	*Easy*	*Easy*

Table 4.2 Comparison of Theories (Lee et al., 1987, p. 35)

Table 4.2 shows a comparison of UMTs as developed by Lee et al. Some of the entries for Bayesian probability are surprising in the light of results reported elsewhere. The theoretical background of Bayesian methods are undoubtedly high but the computational complexity has been reported as exponential (NP hard). There are however a number of different ways of using Bayesian methods.

The ease of application is surprising when considering a comment from Shafer (Shafer, 1986), that has already been noted earlier in this section, that Bayesian design does not have the modularity of production rules. Certainty factors for example does have this modularity and has been claimed to be easier to use for this reason (Dan & Dudeck, 1992).

Ramsbottom and Adams report on a series of expert systems that were developed using an expert system shell specifically to compare three UMTs (Bayesian logic, Certainty factors and Fuzzy logic). They conclude that "the use of fuzzy logic functions allow easier expansion of the system and more accurately represent the nature of the uncertainty and vagueness associated with the analytical test performed" (Ramsbottom & Adams, 1993, p.53).

4.4.2.1 Probability Theory and its suitability for expert system development

Probability theory is where the concept of the management of uncertainty started. Zadeh (Zadeh, 1986) and Kosko (Kosko, 1992) have been among the most vocal critics of its use in expert systems, others consider that probabilities can be applied generally to any system requiring the handling of uncertainty. The major conceptual change necessary for applying probability theory to typical fuzzy situations is to interpret probabilities as a measure of belief in a relevant proposition rather than a long run frequency (Cheeseman , 1986).

Probability theory and its use in handling uncertainty may be mathematically sound but there are still difficulties with the volume of data required from experts regarding conditional probabilities.

It is clear that a Bayesian design does not have the modular character of expert system production rules. We are not free to add or remove probability judgments from a Bayesian design in the way that we are free to add or remove production rules from a production system. A Bayesian design specifies very rigidly just what probability judgment it requires. (Shafer, 1986)

It has been suggested that we are bound to apply probability theory if uncertainty is represented by real numbers and every relevant event may be allocated a real number (Neopolitan, 1992). When this is not possible then other techniques must be investigated.

Lindley cited in Neopolitan states:

> It was good to realize that workers in expert systems are beginning to understand that uncertainty statements must be combined according to the rules of probability. What is surprising is that they took so long to see this. The explanation presumably is that workers in new fields seem to think that everything is new and sometimes fail to recognize connections with older work (Neopolitan, 1992, p.69).

This is certainly not a universally held view. Lindley (1985) asserts that decision making under uncertainty consists of three steps:

1. Quantify uncertainties with probability values

2. Describe the results of all actions in terms of utility.

3. Select the action that will be the most useful.

Yet this is clearly not always possible. "Circumstances do not always permit quantification of uncertainties yet a decision may still be urgently required" (Fox et al., 1990). Lindley claims then that a decision cannot be made. However there are expert systems that are able to produce results under such circumstances. They do not however, use classical decision making.

Pearl (1988) cited in Neopolitan (Neopolitan, 1992) has demonstrated that his approach to probability (e-semantics) "can better handle many of the problems for which default logic and nonmonotonic logic were specifically designed".

Others have stated more explicitly the restricted applicability of the Bayesian technique (Magill and Leech, 1991). This is a summary of the problems that have been found:

1. Experts required to quantify uncertainty in a probabilistic manner based on long past experience and prohibitively large samples.

2. Two or more pieces of evidence in a rule are assumed independent.

3. The algebraic requirement is contravened by the intuitive beliefs of experts.

Problem 1 is based on a statement by Shortliffe and Buchanan (Shortliffe & Buchanan, 1975) in support of their work on Certainty factors and should not be considered current thinking. Problem 2 is also wrong when leveled at Bayesian Techniques in general. Since the independence assumption is completely optional, it may be made to simplify the situation but is open and acknowledged. (Early Bayesian Techniques often made this assumption!) It is strange to consider the independence assumption a problem of Bayesian Techniques in the light of evidence from Rothman that Certainty factors are isomorphic to a subset of probability theory under an appropriate set of assumptions (Rothman, 1989). One of these assumptions is conditional independence of evidence given an hypothesis. Let us consider the possible correlation between two events and the corresponding probabilities of the conjunction and disjunction.

1. Maximum correlation between two events is present when the less probable event occurs only when the more probable event occurs.

The conjunction $p(A\&B) = Min(p(A),p(B))$

The disjunction $p(A \text{ or } B) = Max(p(A), p(B))$

2. If two events are Independent then:

The conjunction $p(A\&B) = p(A)p(B)$

The disjunction $p(A \text{ or } B) = p(A) + p(B) - p(A)p(B)$

3. If minimum correlation applies then

The conjunction $p(A\&B) = Max(0,p(A)+p(B) - 1)$

The disjunction $p(A \text{ or } B) = Min(1, p(A) + p(B))$

It can be seen from this set of rules that the combination rules for disjunction and conjunction that are used in Fuzzy set theory have used the assumption of maximum correlation between the events. This may be correct or incorrect depending on the example.

4.4.2.2 Certainty factors -- the original UMT

Magill and Leech (1991) investigated two commercially available expert system tools that used respectively, Bayes' Theorem and Certainty factors, for the handling of uncertainty. Their aim was to recommend the more appropriate tool for a certain task. They discovered that the complex decision of which UMT was more appropriate for a particular ES development was further hampered by the fact that "the two specific tools do not follow strictly the theories on which they are based". They simply quote Shortliffe and Buchanan when looking at the applicability of MYCIN's Certainty factors to other applications. This thesis will follow their example.

"it is potentially applicable to any problem area in which real world knowledge must be combined with expertise before an informed opinion can be obtained to explain observations, or to suggest a course of action". (Shortliffe & Buchanan, 1975, p. 353)

Some of the difficulties with Certainty factors were considered in Chapter 3, Section 3.2.2.2 A critique of Certainty factors. Problems that have been identified by Giarratano and Riley (Giarratano & Riley, 1994), Heckerman (Heckerman, 1986) and others, that could severely limit other potential application areas, were discussed.

Wise and Henrion explain that in attempting to simplify standard probabilistic methods, UMTs such as Fuzzy sets and Certainty factors, are actually making unacknowledged assumptions about the relationship between propositions.

Any uncertain inference methods, by implication at least, makes certain assumptions about the unspecified parameters, particularly the correlation between propositions (Wise and Henrion, 1986).

Bonissone (1987) suggests that there are "numerous serious problems" with the use of CF's. These include the interpretation of the number, the supposition of independence of evidence and the impossibility of interpreting a CF of zero, which does not allow one to distinguish between lack of information and discordant information.

4.4.2.3 In favour of Dempster-Shafer

When compared to ad hoc techniques, Dempster-Shafer is considered by some to be more desirable because of its rigorous mathematical underpinning. (Cortes-Rello & Golshani, 1990). This is despite Shafer's comment that Dempster's rule of combination had a pragmatic rather than mathematical basis (Shafer, 1976). Compared with other probability based methods such as Bayesian, the Dempster-Shafer theory is more powerful since it can work with probability of sets of points instead of probability of just individual points. In addition it can handle contradictory evidence in a satisfactory manner. (Cortes-Rello & Golshani, 1990)

Neopolitan (1992) is a supporter of probabilistic techniques yet he states that there is still a place for Dempster-Shafer. He suggests it is unfortunate that this theory has been applied inappropriately in some expert systems because there are situations when Dempster-Shafer can provide meaningful results and Bayesian analysis cannot.

Lingras and Wong (Lingras & Wong, 1990, p. 468) discuss two views of belief functions. They suggest that the *compatibility view* allows the use of conditionalizations that are more usually confined to Bayesian theory. This is contrasted with *the probability allocation view* that should be used when the information available cannot be explicitly expresses in terms of propositions, but probability allocation based on the evidence is possible. Lingras and Wong explain that this is because Dempster's rule for combining belief functions is defined for independent bodies of evidence. If there are dependencies then they should be identified and Bayes rule of conditionalization used.

Cortes-Rello and Golshani (1990) selected the Dempster-Shafer method for an expert system in forecasting and marketing management because they felt it better handled the application, because:

1. the solution is not a single method, but a set of methods.

2. the problem is complex and the solution is based on subjective (and possibly contradictory) opinions of expert in forecasting techniques.

3. we can build for different levels of abstraction (for example, rules referring to an individual method, or rules referring to a 'class' of methods.

4. the concept of methods having strengths and weaknesses can be modelled using rules with confirming and 'deconfirming' beliefs over sets of hypotheses.

(Cortes-Rello & Golshani, 1990, p. 17)

4.4.2.4 Other Comparisons.

It is also interesting that although these techniques all claim to be different from the next several researchers have shown that some are (just) special cases of other techniques.

Cheeseman (Cheeseman, 1986) and Barclay in Rothman (Rothman, 1989) both liken

Certainty factors to probability methods. Heckerman (Heckerman, 1986) demonstrates a "clear relationship between certainty factors and probabilities," that he suggests, adds weight, to the idea that probability theory is sufficient for managing uncertainty.

Grosof has shown that the revised versions of MYCIN's Certainty factors are equivalent to a special case of Dempster-Shafer theory. (Grosof, 1986, p. 163)

Zadeh's method of combining fuzzy sets via the max and min functions has been criticized as "failing to describe the real world" (Jumarie, 1993). Jumarie suggests that other methods that are closer to subjective probability should be used when appropriate.

4.4.3 Comparisons of non-numeric UMTs

The research for this thesis has uncovered very few attempts to compare non-numeric UMTs. When mentioned in comparison it has mainly been in view of the discussion about non-numeric versus numeric techniques that was covered in Section 4.4.1.

One criticism of Cohen's theory of endorsements is due to Fox (Fox, 1986). He suggests that although Cohen's theory of endorsements is able to successfully preserve information about the sources of beliefs, it does not show how to deal with knowledge that must be revised in the light of new information.

4.5 Recent advances in the theory of reasoning under uncertainty

There are recent advances in the theory of reasoning under uncertainty that could be considered for incorporation into an expert system. Probably the most important developments are in the areas of Fuzzy Logic and Non-monotonic logics. Hybrid systems provide an area for useful investigation.

Fuzzy logics have recently been further investigated (Novak, 1992), (Graham, 1991) and their use is becoming accepted. Shiraishi describes Fuzzy logic as useful because it simplifies the process of building an expert system by reducing the number of rules required (Shiraishi, 1989). Further work that has been done on the efficiency of implementation algorithms will assist with the practicalities of Fuzzy logic in expert systems (Koczy & Hirota, 1992) and also Dempster-Shafer theory (Xu & Kennes, 1994). Fuzzy logic has also been shown to be useful in connectionist expert systems (Cohen & Hudson, 1992).

Default logics are said to mirror more closely the reasoning process of humans than do other forms of automated reasoning (Lea Sombe, 1990). Giordano and Martelli have summarised the work of Reiter and others in the area of default logics (Giordano & Martelli, 1994) and this method of reasoning is likely to become more widely used. It is perhaps going to be most useful in the future in hybrid systems that are today still in their infancy (Clark, 1990) (Bonissone et al. 1990).

Distributed expert systems and the problems of cooperation between expert systems is considered by Zhang (Zhang, 1992). This interesting work also considers transformations from one UMT to another. This is done by considering the UMTs as members of a mathematical structure known as a group. Isomorphic transformations between the UMTs are defined. This type of definition and the transformations may prove to be useful when designing expert systems that can implement several UMTs.

Chapter 5: A methodology for the selection of a paradigm of reasoning under uncertainty in expert system development.

5.1 Chapter overview

This chapter presents a methodology for the selection of a paradigm of reasoning under uncertainty in expert system development. It begins by considering the requirements of a methodology for the selection of an UMT. Several methodologies from the literature will be considered and shortcomings in each noted. The thesis will then present its own methodology to assist the expert system developer in selecting an appropriate paradigm of reasoning. The possibility of viewing the process of selection of a paradigm as a meta-problem is then considered. The final section of the chapter considers using this to incorporate the selection process into an expert system.

5.2 Some methodologies for the selection of an UMT.

This section will consider several methodologies for the selection of an UMT suggested in the literature and consider their appropriateness for the required task.

It has been suggested that the decision as to which is the appropriate UMT for an expert system development is a trade-off between complexity and precision (Bonissone & Decker, 1986). However consideration of the nature, reliability and characteristics of the data is also important. Clark agrees that the selection process is multidimensional.

> The most appropriate technique for a particular application will thus depend upon a number of factors, such as the nature of the domain, how much data, expertise and time is available to construct the appropriate representation, what level of accuracy is required, what functions the system is intended to support, the importance of meta-level capabilities and so on. (Clark, 1990, p. 140)

The thesis now considers three methods.

5.2.1 Saffioti's method - an outline

Saffioti suggests a method of selecting a paradigm of reasoning under uncertainty. This approach to the problem is very much simply an outline of the process and does little to assist the expert system developer with the information required to complete each of the steps. His method indicates that the comparison should be done in three phases (Saffioti, 1988, p. 93)

1. Select those techniques which are applicable to the problem (the problem fits its preconditions)
2. Verify the epistemological and computational adequacy of the selected techniques for the uncertainty at hand.
3. Weigh the remaining techniques and choose one: the general context should be taken into account

This is a method that requires a great deal of work on the part of the expert system developer. The first point, that very simply states "select those techniques that are applicable to the problem" is a very complex search of the literature in itself, unless some form of assistance is provided to summarize the various options. The second point requires yet more work when to the expert system developer, uncertainty is only one (albeit important) aspect of the system development that they have to consider. It is the objective of this thesis to simplify the process.

5.2.2 Lee's method - numeric UMTs only

Lee et al. provided a more reasonable approach in a table (reproduced as Table 5.1) that compares four aspects of four numeric UMTs (Lee et al, 1987, p.36).

	Bayesian Probability	**Dempster-Shafer**	**MYCIN's Certainty Factors**	**Fuzzy Set Theory**
Problem Definition	*Well-defined*	*Well-defined*	*Well/Ill-defined*	*Well/Ill-defined*
Computing Power Needed	*Small*	*Small-Large*	*Small*	*Small-Large*
Needed amount of training in Theory	*Little*	*Moderate*	*Little*	*Moderate*
Needed amount of training in application	*Little*	*Substantial*	*Little*	*Moderate*

Table 5.1 Guidelines of Selection

Unfortunately Lee et al. have only considered numeric approaches to handling uncertainty and have not justified the content of the table. There are further aspects that required consideration including the source of uncertainty and whether large amounts of historical imformation are available. These aspects will be included in the methodology for selection that is proposed in section 5.3, A Manual approach to the selection.

5.2.3 Kline and Dolins - guidelines and quotation

The scope of this methodology is broader than managing uncertainty. Kline and Dolins provide guidelines to selecting techniques for the total implementation of an expert system. In their book Designing Expert Systems - a guide to selecting implementation techniques, Kline and Dolins devised a number of guidelines that consider aspects of the problem and suggest recommended techniques to use in an expert system development (Kline & Dolins, 1989). An example guideline is presented here (Figure 5.1) since this is very similar to the kind of advice that this thesis was looking to give regarding the selection of an UMT.

> Will the expert system be solving a signal-interpretation problem?
> and
> Is it hard to distinguish true signals from noise (i.e, low S/N ratio)?
> or
> Is it easy to distinguish true signals from noise(i.e. high S/N ratio)?
>
> Low S/N ratio -> Model-Driven Reasoning
> Evidence: Weak, moderate
> High S/N ratio -> Data-Driven Reasoning
> Evidence: Moderate, powerful

Figure 5.1 An example guideline to selecting implementation techniques (Kline & Dolins, 1989).

Quotations are then used in supporting arguments for the design guidelines. Two advantages are given for this type of supporting evidence:

> 1. The quotations help to ensure that the design guidelines have some degree of support among expert system builders, as opposed to merely reflecting the personal biases of the authors of this book.
>
> 2. The quotations provide pointers to additional source of information on a particular issue. (Kline and Dolins, 1989, p. 6)

This method provides some easily accessible advice to the expert system developer. It is provided in a manner that requires careful consideration but does not require too much additional work beyond the development of the expert system itself. This thesis will use a similar technique. It will also use quotations to back up the manual approach to selection that is presented in the following section.

5.3 A manual approach to the selection

This section will present a manual approach to the selection of a paradigm of reasoning under uncertainty. It should be noted that whilst this thesis has attempted to cover a broad range of UMTs the methodology for selection concentrates on numeric methods. This is because Hybrid methods are still largely experimental and the Theory of Endorsements (see Section 3.3.1) is only considered suitable in situations where the reasoning chains are very short (Bhatnager & Kanel, 1986).

In selecting the order for the decision making process consideration was given to any overriding features that would clearly indicate the required form of UMT. It can be seen in Figure 5.2 A manual approach to the selection of an UMT that the first step is making far more clear cut decisions that the latter steps. It is also true that the early steps tend to be more important. This can be likened to a process of first sorting the goats from the sheep and then going on to sort the sheep into Merinos, Leicesters and Suffolks.

The manual approach that follows is in the form of a number of questions that the expert system developer should answer regarding a proposed ES development.

Questions should be considered in the order given and can be considered to form a decision tree, as is presented in Figure 5.2 A manual approach to the selection of a UMT. If the answer to any question is positive and this is confirmed by the guideline then that recommendation should be followed. A negative answer means the following step should be considered.

Step 1a: Considers the source of uncertainty

1a) Is uncertainty mainly in terms of incomplete data?

Guidelines 1a

If uncertainty is mainly in terms of incomplete data then default reasoning and non-monotonic logic is probably appropriate.

Support/Reasons 1a

The process of reasoning using non-monotonic logics is that :

- judgements are made using the available evidence by making assumptions
- assumptions are revised in the light of new evidence.

Non-monotonic logic is not suitable to deal with imprecise data and so another method will be required if this type of uncertainty is present. Refer also to Section 3.3.2. Non-monotonic logics. (McDermott & Doyle, 1980, p.42), (Clark, 1990, p. 129)

__Step 1b: Considers the source of uncertainty__

1b) Is uncertainty mainly in terms of imprecision of knowledge?

Guidelines 1b

b) If uncertainty is mainly in terms of imprecision of knowledge then fuzzy sets may be appropriate.

Support/Reasons 1b

Research has shown that fuzzy set theory is able to express concepts that are not applicable to probability theory (Zadeh, 1985)(Neopolitan, 1990). This may be the only calculus that has systematically addressed the issue of imprecision of statements (Bhatnagar & Kanal, 1992). More efficient algorithms have been developed for the inference process (Koczy & Hirota, 1992). When fuzzy set theory is applicable it may reduce the number of rules required (Shiraishi, 1989).

__Step 2: Historical Data__

Are large amounts of historical data available?

Guidelines 2

If large amounts of data are available then Bayesian methods are likely to be suitable (Valverde & Gehl, 1992, p. 23). There are further requirements if Bayesian methods are to be chosen, go now to step 3.

If not then go on to step 4.

Support/Reasons 2

It is important to consider the structure of the problem and the ease of obtaining numerical measures. If the situation is well developed and a full history of data is available then a system based on mathematical probabilities would be appropriate (Wise & Henrion, 1986) (Buxton, 1989) (Neopolitan, 1992). If the structure of the problem is less well defined then a more flexible approach is required.

Step 3: Conditional Independence

Is there Conditional Independence among cases?

Guidelines 3

Yes - then use Bayes' rule.

No - then Subjective Bayesian may be suitable but steps 4 and 5 should also be considered.

Support/Reasons 3

Bayes' rule assumes conditional independence (Kline and Dolins, 1989) if not then the number of conditional probabilities required becomes prohibitive. Subjective Bayesian Methods using networks reduce this requirement (Heckerman & Shortliffe, 1992) (Srinivas et al., 1990) (Buxton, 1989) (Heckerman, 1990).

Step 4: Representation of Ignorance

Is there an explicit representation of ignorance required?

Guidelines 4

Yes - then use Dempster-Shafer.

No - then go on to step 5.

Support/Reasons 4

Dempster-Shafer provides explicit representation of ignorance through the use of an upper and lower probability (Spillman,1989) (Avanzato, 1990)(Fung & Chong, 1986)(Cortez-Rello & Golshani, 1990)(Valverde & Gehl, 1992). The advantage is the ability to use incomplete probability models (Shafer, 1986. p.133).

Step 5: Difficulty assigning probability

Bayesian inference may be suitable but probability cannot be assigned to all pertinent events

Guidelines 5

Use Dempster-Shafer (Neopolitan, 1992).

Support/Reasons 5

Dempster-Shafer is suitable when a probability cannot be assigned to all events (Neopolitan, 1992) (Buxton, 1989).

Step 6

Is ease of implementation important?

Guidelines 6

IF YES then Certainty Factors were devised to be straight forward

IF NO then use Subjective Bayesian methods.

Support/Reasons 6

Certainty Factors may be used when a simple implementation is important (Dan & Dudeck, 1992) (Heckermann & Shortliffe, 1992). The modular knowledge base is helpful to the developer (Dan & Dudeck, 1992). Certainty Factors have been shown to work (Buchanan & Shortliffe, 1984) and expert systems that use CFs have performed equivalent to, or better than human experts (Heckermann & Shortliffe, 1992)(Yu et al., 1979).

Thus the process of selection has been described and it has resulted in a recommendation of an UMT to be chosen for an expert system.

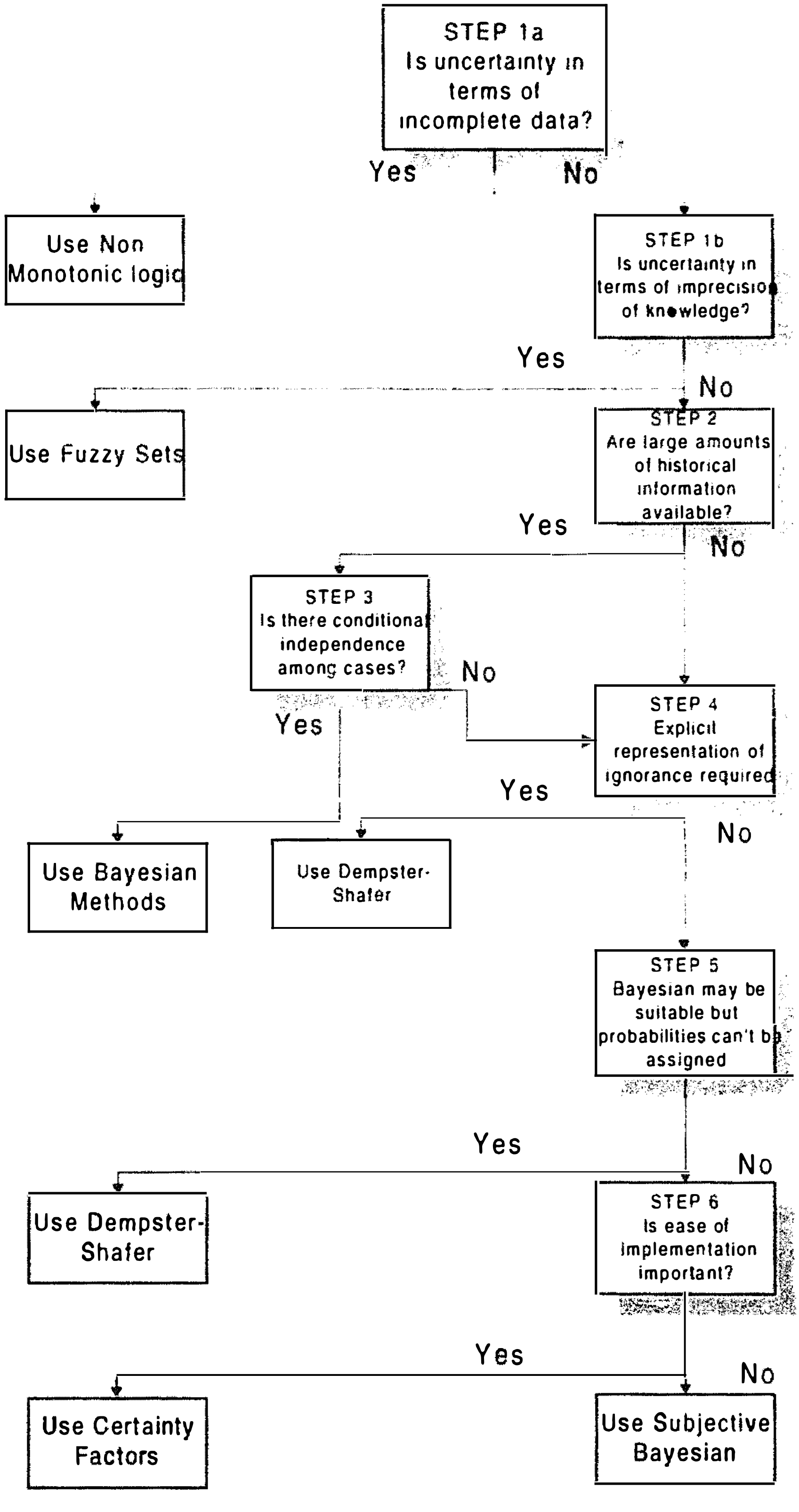


Figure 5.2 A manual approach to the selection of an UMT

5.4 The process of selection as a meta-problem

This section will consider the process of selecting an UMT as a meta-problem. It should be possible to extend the knowledge stored by the expert system to beyond that of just the problem domain. The expert system will then include meta-knowledge about the UMTs themselves. This will mean that during the development process, features of the expert system can be identified that will allow an appropriate paradigm of reasoning under uncertainty to be selected. It will then be put into place automatically by the expert system shell.

Researches have argued for an explicit representation of the methods used for uncertainty management.

> One of the most innovatory characteristics of AI is its concern with representing and using knowledge in as explicit a form as possible. This principle does not seem to have been applied to uncertainty, for which the implicit numerical methods have usually been the only possibility. (Saffioti, 1988, p. 86)

This suggests that assumptions about the validity of a certain UMT for a particular application would then become apparent. If this principle were to be applied it would mean storing meta-knowledge of the uncertainty representation and reasoning process.

The path for selecting an UMT is thus defined as a *meta-problem* which is the approach taken by Fox (Fox, 1986). It does not however provide guidelines as to the particular features of the UMT to consider.

Fox (Fox, 1986) presents a radical approach to the problem of reasoning under uncertainty. He puts forward three arguments for extending the framework of probability:

- explicit representation of several types of uncertainty, specifically possibility and plausibility, as well as probability
- the use of weak methods for uncertainty management in problems which are poorly defined
- symbolic representation of different uncertainty calculi and methods for choosing between them

(Fox, 1986, p. 447)

So the paradigms of reasoning themselves could become part of an extended knowledge base of an expert system. More appropriately this would be a separate meta-knowledge base that would be a part of the expert systems shell. This is more appropriate since this section of knowledge would be standard and appropriate for any expert system development and quite separate from the domain knowledge that is currently stored.

In the following section this idea is further considered.

5.5 Incorporating the process of selection in an expert system

If knowledge about UMTs can be abstracted then the process of selection of a particular UMT can be performed as a part of the automated process. That is -- within the expert system itself.

It has been suggested that the expert system shell should be structured to be able to help in the decision making process. To provide such an implementation a system would require a set of rules that would provide for the selection of a UMT, automating the process outlined in section 5.3 A manual approach to the selection. Also required would be explicit representation of the control processes to implement a number of selected calculi. The calculi provided could be selected by a trade-off between complexity and precision and the rules used to select them rely on a number of features including the nature of uncertainty, availablity of historical data and the importance of ease of implementation -- as described in section 5.3 (Bonissone and Decker, 1986).

Others have supported the suggestion that the uncertainty calculi themselves should be represented in the knowledge base. Fox (1986) demonstrated that the language of probability theory has a framework similar to other languages. It consists of many of the features usually associated with context-free grammars and BNF form (Louden, 1993). This includes a "vocabulary of terminal symbols, and a set of composition or transformation rules for generating sentences from elements of the vocabulary" (Fox, 1986).

> The terminal symbols of probability theory include the real numbers, operators (+,-,.. etc) and relations (= , >, < etc). The composition and transformation rules are the ordinary algebraic composition and manipulation rules, extended operators (for example sum, product) and specific revision rules (eg Bayes rule) (Fox, 1986, p. 455).

The probabilistic reasoning process is represented by the production rules. When carried out these rules will use a composed set of terminal symbols, to generate a new set of terminals.

The advantages of including the representation of UMTs explicitly are clear. Once the methods are included as options for an expert system development then the UMTs become alternatives to be used as required. This will not be useful until it is possible to explicitly represent the methodology by which the selection of a UMT will be performed. This was begun in section 5.3 A manual approach to the selection, but remains to be validated, refined and automated. Fox suggests that once this in done UMTs will be seen "not as rivals for all the honours, but as alternatives to be used as circumstances demand" (Fox, 1986, p. 455).

Figure 5.3 on the next page illustrates this idea with a fragment of an expert system for advising on the selection of an uncertainty calculus under development in PROPS 2.

> The first two rules generate the set of possible methods and the assumptions which must be tested in order to evaluate them. The second two rules generate the subset of plausible methods on the argument that their assumptions are satisfied. The last pair of rules considers the number of plausible candidates and recommends accordingly. If neither of these rules is satisfied a weak uncertainty calculus can be suggested. (Fox, 1986, p.455)

Including meta-knowledge about the method of reasoning under uncertainty would provide another level of flexibility that is not currently available. Before it can be done expert system shells will need to become more flexible so that the knowledge of UMTs can be translated into the reasoning process of the expert system itself.

> if uncertainty calculus is required
> and Method is a kind of uncertainty calulus
> and assumptions of Method are not violated
> then Method is a possible uncertainty calculus.
>
> if method is a possible uncertainty calculus
> and assumptions of Method includes Assumption
> then check Assumption is true
> and record that assumptions of Method are checked.
>
> ---
>
> if assumptions of Method includes Assumption
> and Assumption is false
> then assumption of Method are violated
>
> if Method is a possible uncertainty calculus
> and assumptions of Method are checked
> and assumptions of Method are not violated
> then plausible uncertainty methods includes Method.
>
> ---
>
> if plausible uncertainty methods includes Method
> and number of plausible uncertainty methods = 1
> then method is recommended.
>
> if number of plausible uncertainty methods > 1
> Then cost benefit analysis of methods is required.

Figure 5.3 Productions rules demonstrating meta-knowledge (Fox, 1986)

Chapter 6: Conclusion

6.1 Chapter overview

This chapter will conclude the thesis. The Bibliography is to follow.

It will be shown that the objectives of the project have been met. Consideration will be given to the applicability of the results. Suggestions will be made for the next stage of research in this area.

6.2 Have objectives been met?

To enable consideration of whether the objectives of the project have been met they are restated here (taken from Chapter 1)

The major aim of this project is:

1. To define the criteria on which the selection of a paradigm of reasoning under uncertainty for an expert system should be made.

A secondary aim is:

2. To consider what recent advances in the theory of reasoning under uncertainty are worthy of consideration for incorporation into expert system developments.

In an attempt to answer these questions, I have gathered information from two major sources.

1. The theory of reasoning under uncertainty. There is a great deal of material available in journals and books

A great deal of the literature on this subject relates to the formal methods involved. There are some parts of the field (for example Default and non-monotonic Logic) that are in the development stages and as far as I can ascertain are yet to be incorporated into expert systems.

2. Expert System applications. Detailed information on the success or failure of the particular UMT used is more difficult to obtain.

In chapter 5 a manual methodology for the selection of a paradigm of reasoning under uncertainty was developed. This defines clearly the criteria that should be applied during the selection process. This methodology should now undergo a process of validation and verification that is quite likely to require the methodology to be revised.

There are recent advances in the theory of reasoning under uncertainty that could be considered for incorporation into an expert system. The use of efficient algorithms for Fuzzy logic and the further use of default logic were considered in Section 4.5: Recent advances in the theory of reasoning under uncertainty.

6.3 Consideration of the applicability of the results

Throughout this research an attempt has been made to cover all types of expert system development. The main consideration has been to diagnostic systems but necessarily all sources of uncertainty were considered. This led to an attempt to research all major areas of numeric, symbolic and hybrid systems. The latter two areas have proved particularly difficult to address but perhaps also represent the main areas for future research. Especially combining the symbolic methods with numeric methods.

The manual approach to the selection of an UMT that has been developed (5.3 A manual approach to the selection) makes no recommendations to select an hybrid method. It may be that if an expert system developer is not able to clearly answer the steps of the method with a discrete answer then it may become apparent that more than one method for handling uncertainty should ideally be used in an expert system.

The manual method also makes no attempt to select the type of non-monotonic logic that would be most appropriate. This level of detail is beyond the scope of the thesis.

6.4 Recommendations and suggestions for the next step

The manual methodology for the selection of a paradigm of reasoning under uncertainty that has been developed should now undergo a process of validation and verification to ascertain its usefulness for the expert system developer.

It is clear that UMTs that are able to combine the ability to deal with different aspect of uncertainty in the same system need further investigation. Current expert systems may select the method that appears most appropriate but they are not able to cope with the full spectrum of uncertainty. Clark emphasises this

> However many domains of interest are composed of a mixture of quantitative and qualitative relations. So no UMT may be unequivocally appropriate. This raises the need to intelligently combine different UMTs and suggests that an important area of research is the use of both symbolic and quantitative representations of uncertainty in the same application. (Clark, 1990, p. 142)

It will be important to combine methods effectively since some UMTs are able to deal better with uncertainty arising from different sources.

Non-monotonic logics mirror more closely the reasoning used by humans and it may be that as expert systems become able to cover a broader knowledge base these methods of reasoning become more important. It would be worthwhile investigating their use in expert systems (Bonissone et al, 1990) .

6.5 Conclusions

In conclusion this chapter has shown that the objectives of the thesis have largely been met. This is clear since the methodology for selection has been presented.

The next stages in the research process are in three areas. The first is to validate and verify the methodology that has been developed in the thesis. The second is to implement the process of selection of an UMT as a portion of an expert system shell. The third is to further explore symbolic and especially hybrid methods of reasoning under uncertainty. Hybrid methods are those that will be able to reason with data that contains uncertainties of several types. Current examples include a numeric and symbolic components, in the future they may contain mutiple components both numeric and symbolic. For example Bayesian probability may deal with uncertainty that pertains to unreliable information, Fuzzy sets for uncertainty that originates from lack of precision and non-monotonic methods deal with incomplete information. It is clear that there is much work still to be done in this area.

A bibliography completes the thesis.

Bibliography

Adams, J. (1984). Probabilistic reasoning and certainty factors. In B. Buchanan and E. Shortcliffe (Eds.), Rule-based Expert Systems, (pp. 263-271). Reading, MA: Addison-Wesley.

Avanzato, Robert. (1991). Uncertainty management issues in the development of an acoustic signal interpretation expert system. Proceedings of the IEEE/ACM International Conference on Developing and Managing Expert System Programs, pp. 68-75.

Bhatnagar, R.K. and Kanal, L.N. (1986). Handling uncertain information: A review of numeric and non-numeric methods. In L. Kanal and J. Lemmer (Eds.), Uncertainty in Artificial Intelligence (pp. 3-26). Amsterdam: Elsevier Science Publishers B.V.

Bonissone, P.P. and Decker, K.S. (1986). Selecting uncertainty calculi and granularity. In L. Kanal and J. Lemmer (Ed.), Uncertainty in Artificial Intelligence (pp. 217-247). Amsterdam: Elsevier Science Publishers B.V.

Bonissone, P.P. (1987). Reasoning, Plausible. In Encyclopedia of Artificial Intelligence, (pp. 854-863).

Bonissone, P.P. (1990). Now that I have a good theory of uncertainty, what else do I need?. In M. Henrion, R. Schachter, L. Kanal, & J. Lemmer (Eds.), Uncertainty in Artificial Intelligence 5 (pp. 237-253). Amsterdam: North-Holland.

Bonissone, P.P., Cyrluk, D.A., Goodwin, J.W., Stillman, J. (1990). Uncertainty and Incompleteness: Breaking the symmetry of defeasible reasoning. ?. In M. Henrion, R. Schachter, L. Kanal, & J. Lemmer (Eds.), Uncertainty in Artificial Intelligence 5 (pp. 67-85). Amsterdam:North-Holland.

Booker, L., Hota, N., Ramsey, C. (1990). BaRT: A Bayesian Reasoning Tool for knowledge based systems?. In M. Henrion, R. Schachter, L. Kanal, & J. Lemmer (Eds.), Uncertainty in Artificial Intelligence 5 (pp. 271-282). Amsterdam: North-Holland.

Buchanan, B. and Shortcliffe, E. (1984). Rule Based Expert Systems. Addison-Wesley: Reading, Massachusetts.

Buxton, R. (1990). Modelling uncertainty in expert systems. International Journal Man-Machine Studies, 31, 415-476.

Chang, A. and Hall, L. (1992). The validation of fuzzy knowledge-based systems. In L. Zadeh and J. Kacprzyk (Eds.), Fuzzy Logic for the Management of Uncertainty (pp. 589-604). John Wiley & Sons, Inc.: New York.

Cheeseman, P. (1986). Probabilistic versus fuzzy reasoning. In L. Kanal and J. Lemmer (Eds.), Uncertainty in Artificial Intelligence (pp. 85-102). Amsterdam: North-Holland.

Clark, D. A. (1990). Numerical and symbolic approaches to uncertainty management in AI. Artificial Intelligence Review, 4, 109-146.

Clarke, N., McLeish, M. and Vyn, T. (1991). Using Certainty Factors and possibility theory methods in a tillage selection expert system. Expert Systems with Applications, 4, 53-62.

Cohen, M. (1985). Alternative Theories of Inference in Expert Systems for Image Analysis. Falls Church, Virginia : Decision Science Consortium.

Cohen, M. and Hudson, D. (1992). Management of uncertainty in knowledge-based medical systems. In L. Zadeh and J. Kacprzyk (Eds.). Fuzzy Logic for the Management of Uncertainty (pp. 553-568). New York : John Wiley & Sons, Inc.

Cohen, P. (1985). Heuristic Reasoning about Uncertainty : An Artificial Intelligence Approach. London: Pitman.

Cohen, P.R. and Grinberg, M.R. (1988). A Theory of Heuristic Reasoning About Uncertainty. London: Pitman.

Cohen, P.R. (1989). Steps toward programs that manage uncertainty. In L. Kanal, K. Levitt, J. Lemmer (Eds.), Uncertainty in Artificial Intelligence 3 (pp. 257-265). Amsterdam: North-Holland.

Cortes-Rello, E. and Golshani, F. (1990). Uncertain reasoning using the Dempster-Shafer method: an application in forecasting and marketing management. Expert Systems (U.K.), 7(1). 9-18.

D'Ambrosio, B. (1989). A Hybrid Approach to Reasoning Under Uncertainty. In L. Kanal, K. Levitt and J. Lemmer (Eds.), Uncertainty in Artificial Intelligence 3 (pp. 267-282). Amsterdam: North-Holland.

Dan, Q. and Dudeck, J. (1992). Certainty factor theory and its implementation in a medical expert system shell. Medical Inform, 17(2), 87-103.

De Kleer, J. (1984). Choices without backtracking. Proceedings of the National Conference on Artificial Intelligence (AAAI-84), Austin, Texas, 1984, (pp. 79-85).

Dempster, A. (1967). Upper and lower probabilities induced by a multi-valued mapping. Annals of Mathematical Statistics, 38, 325-339.

Dubois, D. and Prade, H. (1990). A Discussion of uncertainty handling in support logic programming. International Journal of Intelligent Systems, 5, 65-82.

Filman, R. (1988). Reasoning with worlds and truth maintenance in a knowledge-based programming environment. Communications of the ACM, 31 (4), 382 - 401.

Fox, J. (1986). Three Arguments for extending the framework of probability. In L. Kanal and J. Lemmer (Eds.), Uncertainty in Artificial Intelligence (pp. 447-458). Amsterdam: Elsevier Science Publishers B.V.

Fox, J., Clark, D., Glowinski, A. and O'Neil, M. (1990). Using predicate logic to integrate qualitative reasoning and classical decision theory. IEEE Transactions on Systems, Man and Cybernetics, 20(2), 347-357.

Freund, J. E. (1972). Mathematical Statistics. London: Prentice-Hall International.

Giordano, L. and Martelli, A. (1994). On cumulative default logics. Artificial Intelligence, 66, 161-179.

Ginsburg, M. (1986). Implementing Probabilistic Reasoning. In L. Kanal and J. Lemmer (Eds.) Uncertainty in Artificial Intelligence (pp. 331-338). Amsterdam: Elsevier Science Publishers B.V.

Giarratano, J. and Riley, G. (1994). Expert Systems: Principles and Programming. (2nd ed). Boston: PWS-Kent Publishing Company.

Graham, I. (1991). Fuzzy Logic in commercial expert systems - results and prospects. Fuzzy Sets and Systems, 40(3), 451-472.

Grech, A. and Sammut, C. (1989). Qualitative plausible reasoning and assumptions. In J.R. Quinlan, (Ed.) Applications of Expert Systems (pp. 290-309). Sydney: Addison-Wesley Publishing Company.

Grosof, B. (1986). Evidential reasoning as transformed probability. In L. Kanal and J. Lemmer (Eds.), Uncertainty in Artificial Intelligence (pp. 153-166). Amsterdam: Elsevier Science Publishers B.V.

Guida, G. and Spampinata, L. (1989). Assuring adequacy of expert systems in critical application domains: A constructive approach. In E. Hollnagel (Ed), (1989). The Reliability of Expert Systems (pp.134-167). Chichester: Ellis Horwoord.

Heckerman, D. (1990). A Tractable inference algorithm for diagnosing multiple diseases. In M. Henrion, R. Schachter, L. Kanal and J. Lemmer (Eds.), (1990). Uncertainty in Artificial Intelligence 5 (pp. 163-171). Amsterdam: Elsevier Science Publishers B.V.

Heckerman, D. (1990). An Empirical Comparison of Three Inference Methods. In R. Schachter, K. Levitt, L. Kanal and J. Lemmer (Eds), Uncertainty in Artificial Intelligence 4 (pp. 283-302). Amsterdam: Elsevier Science Publishers B.V.

Heckerman, D. (1986). Probabilistic Interpretations for Mycin's Certainty Factors. In L. Kanal and J. Lemmer (Eds.), Uncertainty in Artificial Intelligence (pp. 167-196). Amsterdam: Elsevier Science Publishers B.V.

Heckerman, D., Mamdani, E. and Wellman, M. (1995). Real-world applications of Bayesian Networks. Communications of the ACM. 38(3), 24-26.

Heckerman, D. and Shortcliffe, E. (1992). From Certainty Factors to Belief Networks. Artificial Intelligence in Medicine, 4(1), 35-52.

Henkind, S. and Harrison, M. (1988). An analysis of four uncertainty calculi. IEEE Transactions on Systems, Man and Cybernetics, 18(5), 700-714.

Henrion, M., Schacter, R.D., Kanel, L.N. and Lemmer, J. (Eds.). (1990). Uncertainty in Artificial Intelligence 5. Amsterdam: Elsevier Science Publishers B.V.

Hollnagel, E. (1989). Issues in reliability of expert systems. In E. Hollnagel (Ed.). Reliability of Expert Systems (pp. 168-209). Chichester: Ellis Horwood.

Hsu, K. and Chu, C. (1989). Practical issues in designing knowledge-based expert systems. Computers and Industrial Engineering, 17 (1-4), 61-66.

Hunter, D. (1986). Uncertain Reasoning using Maximum Entropy Inference. In L. Kanal and J. Lemmer (Eds.), Uncertainty in Artificial Intelligence (pp. 203-215). Amsterdam: Elsevier Science Publishers B.V.

Ignizio, J. (1991). Introduction to Expert Systems. New York: McGraw-Hill, Inc.

Isaac, S. and Michael, W. B. (1974). Handbook in research. San Diego: Knapp.

Jackson, P. (1990). Introduction to Expert Systems. Wokingham, England: Addison-Wesley.

Jumarie, G. (1993). Expert systems and Fuzzy Systems: A new approach via possibility-probability conversion. Kybernetes, 22(8), 21-36.

Kerr, R. (1992). Expert systems in production scheduling : Lessons from a failed implementation. Journal of Systems Software, 19, 123-130.

Kline, P. and Dolines, S. (1989). Designing Expert Systems : A Guide to Selecting Implementaion Techniques. New York: John Wiley & Sons.

Kobalicek, J., Gorig, J., Malik, K. and Dohnal, M. (1993). Examples of the application of fuzzy expert systems to rubber compounding. Plastics, Rubber and Composites Processing and Applications, 19 (3), 159-173.

Koczy, L. and Hirota, K. (1992). A fast algorithm for fuzzy inference by compact rules. L. Zadeh and J. Kacprzyk (Eds). Fuzzy Logic for the Management of Uncertainty (pp. 297-317). New York: John Wiley & Sons.

Kosko, B. (1992). Neural Networks and Fuzzy Systems. Englewood Cliffs, NJ: Prentice Hall.

Lea Sombe (1990). Reasoning under Incomplete Information in Artificial Intelligence. New York: John Wiley & Sons, Inc.

Lee, N., Grize, Y. and Dehnad, K. (1987). Quantitative Models for Reasoning under Uncertainty in Knowledge-Based Expert Systems. International Journal of Intelligent Systems, 2, 15-38.

Lenat, D. and Guha, R. (1990). Building Large Knowledge-based Systems. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.

Lindley, D. (1985). Making Decisions. London, U.K.: Wiley.

Lingras, P and Wong, S.K.M. (1990). Two perspectives of the Dempster-Shafer theory of belief functions. International Journal of Man-Machine Studies, 33, 467-487.

Liu, X. and Gammerman, A. (1986). On the Validity and Appplicability of the Inferno System. In M.A. Bramer (Ed.) Research and Development in Expert Systems III (pp. 47-56). Cambridge: Cambridge University.

Louden, K. (1993). Programming Languages: Principles and Practice. Boston, Massachusetts: PWS Publishing Co.

Lucas, P. and Van der Gaag, L. (1991). Principles of Expert Systems. Wokingham, England: Addison-Wesley Publishing Co.

Magill, W. and Leech, S. (1991). Uncertainty techniques in system software. Decision Support Systems, 7, 55-65.

McDermott, D. and Doyle, J. (1980). Nonmonotonic Logic. Artificial Intelligence, 13, 41-71.

McDermott, J. (1982). R1: A Rule-based configurer of computer systems. Artificial Intelligence, 19, 39-88.

Miyoshi, T., Fukami, S. and Koyama, H. (1992). Management of uncertainty in LIFE shell fuzzy frame system. Expert Systems with Applications, 5, 359-368.

Morawski, P. (1989). Understanding Bayesian Belief Networks. AI Expert , May 1989,(pp. 44-48).

Neopolitan, R. (1992). A survey of uncertain and approximate inference. In L. Zadeh and J. Kacprzyk (Eds) Fuzzy Logic for the Management of Uncertainty (pp. 55-82). New York : John Wiley & Sons, Inc.

Ng, K. and Abramson, B. (1990). Uncertainty Management in Expert Systems. IEEE Expert , 5(4), 29-48.

Novak, V. (1992). Fuzzy logic as a basis of approximate reasoning. In L. Zadeh and J. Kacprzyk (Eds.), Fuzzy Logic for the Management of Uncertainty, (pp. 247-264). New York: John Wiley & Sons, Inc.

O'Keefe, R., Balci, O. and Smith, E. (1987). Validating Expert System Performance. IEEE Expert, 2(12), 81-87.

Oliphant, J. and Blockley, D. (1989). Knowledge-based system: advisor on the selection of earth retaining structure. In B.H.V. Topping (Ed.). Artificial Intelligence Techniques and Applications for Civil and Structural Engineers. (pp. 253-262). Edinburgh: Civil-comp Press.

Ramsbottom, D. and Adams, M. (1993). Uncertainty within a commercial expert system shell for polymer analysis. Chemometrics and Intelligent Laboratory Systems, 19, 53-63.

Reiter, R. (1980). A logic for default reasoning. Artificial Intelligence. 13, 81-132.

Reiter, R. (1987). Nonmonotonic reasoning. In G. Shafer and J. Pearl (Eds). Uncertain Reasoning (pp. 637-656). San Mateo, California: Morgan Kaufman Publishers, Inc.

Rothman, Peter. (1989). Selecting an Uncertainty Management System. AI Expert. July 1989, (pp. 56-62).

Saffiotti, A. (1988). An AI view of the treatment of uncertainty. Knowledge Engineering Review, 2, 75-98.

Shafer, G. (1986). Probability Judgement in Artificial Intelligence. In L. Kanal and J. Lemmer (Eds). Uncertainty in Artificial Intelligence. Amsterdam: Elsevier Science Publishers B.V.

Shafer, G. (1976). A Mathematical Theory of Evidence. Princeton, NJ : Princeton University Press.

Shafer, G. and Pearl, J. (1990). Non-numerical approaches to plausible inference. In G. Shafer and J. Pearl (Eds) Uncertain Reasoning (pp. 625-628). Morgan Kaufman : San Mateo, California.

Shenoy, P. (1992). Using possibility theory in Expert Systems. Fuzzy Sets and Systems, 52, 129-142.

Shiraishi, N., Furata, H., Umano, M. and Kawakami, K. (1989). Knowledge-based expert system for damage assessment based on fuzzy reasoning. In Topping, B.H.V. (Ed). Artificial Intelligence Techniques and Applications for Civil and Structural Engineers (pp. 211-216). Edinburgh: Civil-comp Press.

Shortliffe . E. and Buchanan, B. (1975). A Model of Inexact Reasoning in Medicine. Mathematical Biosciences, 23 , 351-379.

Srinivas, S., Russell, S., and Agogino, A. (1990). Automated construction of sparse bayesian networks from unstructured probabilistic models and domain information. In M. Henrion, R. Schachter, L. Kanal and J. Lemmer (Eds). Uncertainty in Artificial Intelligence 5 (pp. 295-308). Amsterdam: Elsevier Science Publishers B.V.

Spillman, Richard. (1990). Managing Uncertainty with Belief Functions. AI Expert, May 1990, (pp.44-49).

Spiegelhalter, D., Franklin, R. and Bull, K. (1990). Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert system. In M. Henrion, R. Schachter, L. Kanal & J. Lemmer (Eds). Uncertainty in Artificial Intelligence 5 (pp.163-171). Amsterdam: Elsevier Science Publishers B.V.

Sullivan. M. and Cohen, P. (1990). An Endorsement-based plan recognition program. In G. Shafer and J. Pearl (Eds) Uncertain Reasoning (pp. 672-676). San Mateo, California: Morgan Kaufman.

Tonn, B. E. and Goeltz, R. T. (1990). Psychological validity of uncertainty combining rules in expert systems. Expert systems, 7(2).

van der Lubbe, J., Backer, E. and Krijgsman, W. (1991). Models for reasoning with multitype uncertainty in expert systems. In B. Bouchon-Meunier, R. Yager and L. Zadeh (Eds), Uncertainty in Knowledge Bases. 3rd International Conference on Information Processing and Management of Uncertainty. Berlin, Germany: Springer-Verlag.

Vaughan, D., Perrin, B. and Yadrick, R. (1990). Comparing Expert Systems built using different Uncertain Inference Systems. In M. Henrion, R.D. Schacter, L.N. Kanel & J.F. Lemmer (Eds), Uncertainty in Artificial Intelligence 5. Amsterdam: Elsevier Science Publishers B.V.

Valverde, L. and Gehl, S. (1992). Probabilistic Models for Uncertainty Management in Expert Systems. Journal of Power and Energy. 206(A1), 19-26.

Weichselberger, K. and Pohlmann, S. (1990). A Methodology for Uncertainty in Knowledge-Based Systems. Berlin: Springer-Verlag.

Xu, H. and Kennes, R. (1994). Steps towards an Efficient Implementation of Dempster-Shafer Theory. In R.R. Yager, M. Fedrizzi and J. Kacprzyk (Eds), Advances in the Dempster-Shafer Theory of Evidence (pp. 153-174) New York: Wiley.

Yadrick, R., Perrin, B. and Vaughan, D. (1988). Evaluation of Uncertain Inference Models I : Prospector. In J.F. Lemmer and L.N. Kanel (Eds), Uncertainty in Artificial Intelligence 2 (pp. 77-88) Amsterdam: Elsevier Science B.V.

Yu, V., Buchanan, B., Shortcliffe, E., Wraith, S., Davis, R., Scott, A. and Cohen, N. (1979). An evaluation of the performance of a computer-based consultant. Computer Programs in Biomedicine, 9, 95-102.

Zadeh, L. (1985). The role of Fuzzy Logic in the Management of Uncertainty in Expert Systems. In Gupta, Kandel, Bandler and Kiska (Eds), Approximate Reasoning in Expert Systems (pp. 132-166). Amsterdam: Elsevier Science.

Zadeh, L. (1986). Is Probability theory sufficient for dealing with uncertainty in AI: A Negative View. In L.N. Kanal and J.F. Lemmer (Eds). Uncertainty in Artificial Intelligence (pp. 103-116). Amsterdam: Elsevier Science.

Zhang, C. (1992). Cooperation under uncertainty in distributed expert systems. Artificial Intelligence, 56, 21-69.