

Rochester Institute of Technology

## RIT Digital Institutional Repository

---

### Theses

---

1990

## Uncertainty reasoning and representation: A Comparison of several alternative approaches

Barbara S. Smith

Follow this and additional works at: <https://repository.rit.edu/theses>

---

### Recommended Citation

Smith, Barbara S., "Uncertainty reasoning and representation: A Comparison of several alternative approaches" (1990). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact [repository@rit.edu](mailto:repository@rit.edu).

Rochester Institute of Technology  
Department of Computer Science

Uncertainty Reasoning and Representation:  
A Comparison of Several Alternative Approaches

by  
Barbara S. Smith

A Thesis, submitted to  
The Faculty of the Department of Computer Science  
in partial fulfillment of the requirements for the degree of  
Master of Science in Computer Science

Approved by: \_\_\_\_\_ 8/10/90  
Professor John A. Biles Date

\_\_\_\_\_  
Professor Feredoun Kazemian Date

\_\_\_\_\_  
Professor Peter G. Anderson Date

Uncertainty Reasoning and Representation:  
A Comparison of Several Alternative Approaches

I, Barbara S. Smith, hereby grant permission to the Wallace Memorial Library of RIT to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Date: August 13, 1990

## ABSTRACT

Much of the research done in Artificial Intelligence involves investigating and developing methods of incorporating uncertainty reasoning and representation into expert systems. Several methods have been proposed and attempted for handling uncertainty in problem solving situations. The theories range from numerical approaches based on strict probabilistic reasoning to non-numeric approaches based on logical reasoning. This study investigates a number of these approaches including Bayesian Probability, Mycin Certainty Factors, Dempster-Shafer Theory of Evidence, Fuzzy Set Theory, Possibility Theory and non-monotonic logic. Each of these theories and their underlying formalisms are explored by means of examples. The discussion concentrates on a comparison of the different approaches, noting the type of uncertainty that they best represent.

### CR Categories

- I.2.1      Applications and Expert Systems  
            Medicine and Science
- I.2.4      Artificial Intelligence  
            Knowledge Representation  
            Formalisms and Methods

Keywords : Uncertainty, approximate reasoning, Dempster-Shafer Theory of Evidence, Fuzzy set theory

## Table of Contents

1. Introduction.....	3
2. The Nature of Uncertainty.....	6
3. Probability Theory.....	10
3.1 Background: What are probabilities?.....	10
3.2 Bayesian Probability Theory.....	14
3.2.1 Conditional Probabilities.....	15
3.2.2 Bayes' Theorem.....	17
3.2.3 Statistical Independence Assumptions.....	19
3.2.4 Ignorance.....	21
3.3 Mycin - An Early Expert System.....	22
3.3.1 Mycin Certainty Factors.....	23
3.3.2 Combining evidence in Mycin.....	26
3.3.3 Certainty Factors - Examples.....	28
3.3.4 Ignorance.....	31
3.4 Dempster-Shafer Mathematical Theory of Evidence.....	33
3.4.1 Representation of Uncertain Information.....	33
3.4.2 Belief, Commonality, Plausibility - An Example...	38
3.4.3 Combining Evidence from Multiple sources.....	40
3.4.4 Demster's Rule of Combination - Examples.....	42
3.4.5 Ignorance.....	48
3.5 Comparison of Probability-based approaches.....	50
3.5.1 Representation of Belief.....	50
3.5.2 Ignorance.....	52
3.5.3 Prior Probability - Problem of Subjectivity....	53

3.5.4 Assumptions: Statistical Independence and Mutual Exclusivity.....	54
3.5.5 Inferencing methods.....	56
4. Fuzzy Set Theory.....	58
4.1 Definition of Fuzzy Sets.....	59
4.2 Fuzzy Logic.....	63
4.3 Possibility Theory.....	65
4.4 Ignorance.....	67
4.5 Comparison of Possibility Theory and Probability.....	68
5. Non-monotonic Logic.....	74
5.1 Basic Characteristics.....	75
5.2 Modal Logic.....	76
5.3 Doyle's Truth Maintenance System.....	77
6. Summary and Conclusion.....	78
REFERENCES.....	83
ANNOTATED BIBLIOGRAPHY.....	85

## Chapter One

### Introduction

Every day people are faced with the dilemma of making decisions and solving problems in uncertain environments. These domains can be characterized in several ways. The information obtained could only be partial, in that answers to several questions are not known. The available information also could be approximate, in that all the required answers are known but they are not totally accurate. The information obtained from multiple sources also could be conflicting or inconsistent. If the available information is uncertain, then the problem can be solved only with uncertainty.

Expert systems created to support people in decision making must be designed with the ability to cope with these uncertain domains. The goal in expert system design is to combine 'expert' knowledge of a domain with 'expert' ability to reason and conceptualize about a domain. Several methods have been proposed and attempted for handling uncertainty in problem solving situations. The theories range from numerical approaches based on strict probabilistic reasoning to totally non-numeric approaches based on logical reasoning or specifically, first order predicate calculus. Each of these theories is supported by different assumptions and/or

interpretations of other existing theories. Therefore, the uncertainty best handled by each is quite different.

It is the primary objective of this thesis to investigate six different approaches that have been proposed by AI researchers for problem solving and decision making in uncertain domains. The theories that will be reviewed are: Bayesian Probability Theory, Mycin Certainty Factors, Dempster-Shafer Theory of Evidence, Fuzzy Set Theory, Possibility Theory and non-monotonic logic. The discussion of each theory will focus primarily on four general areas. The first area is concerned with how each theory represents the uncertain information or evidence. Secondly, the methods used for combining uncertain information from different sources will be examined along, with the underlying assumptions that are required. How each theory handles some special cases of uncertainty such as ignorance or conflicting information then will be discussed. Finally we will look at how inferences can be made using the information from each theory.

Chapter Two will describe different types of uncertainty and how they arise in problem solving situations. An overview of how people process information for decision making and task resolution also will be included. In Chapter Three the Bayesian Probability Theory, Mycin Certainty Factors and the Dempster-Shafer Theory of Evidence will be presented as three very distinct yet related interpretations of probability. Chapter Four will discuss Fuzzy Set theory and Zadeh's Possibility Theory.



Chapter Five will be concerned with non-monotonic logic as a non-numeric approach to reasoning with uncertainty. In Chapter Six the strong and weak aspects of each approach will be summarized.

## Chapter Two

### The Nature of Uncertainty

The complex domains within which people make decisions are very diverse and dynamic. Though each situation is often unique in many ways, the underlying decision process usually can be characterized by three primary steps [Sage, 1987]. The first step is the formulation of the decision problem. In this step the desired objectives and necessary requirements are identified, and all the potentially acceptable alternative solutions are generated. The second step involves the analysis of all these alternative solutions to evaluate the impacts and costs of the alternatives. In the last step an interpretation of the analysis done in step two is made and the most appropriate alternative for implementation or further evaluation is selected. In order to actually solve problems, there is a need for a tremendous amount of iteration back to earlier steps when either more information is needed or new information has an effect on a previously selected alternative.

Information is without doubt the most important building block of this decision process model. The first step in understanding human reasoning is to characterize and understand the information used by people for problem

resolution. In order to understand how information processing is done, particular attention must be paid to how the information is acquired, how the information is represented, and how it then is used to make inferences.

The situation within which people are forced to make decisions is often very complex in that several alternative solutions are possible. In some cases, the best alternative can be chosen quickly and with little difficulty. More commonly, however, the choice is not very clear cut due to the nature of the available information. The evidence or information that is available is often characterized as being partial, inaccurate, inconsistent, vague or otherwise ill-suited for a positive judgement to be made. Therefore, people are forced to make decisions based on uncertain information.

There are a number of reasons that evidential information could be partial or incomplete. It could be due to the fact that the tools required to obtain all the required information are not readily available. This could be a result of the high cost of obtaining the proper tools or the fact that the actual technology required to obtain the information has not been developed adequately. Often, when a time constraint is placed on a decision process, all the information cannot be collected quickly enough, forcing a decision to be made based on partial information.

The sources of our information, whether they be an electronic sensor or another person, are inherently imperfect. Therefore, the information that we obtain from

these sources could in fact be inaccurate. In order to solve the majority of the problems we face, we rely on consulting with multiple sources. The information from these sources is then combined to support a decision. It is possible that the information obtained from one source could be either inconsistent or actually conflicting with the information from another source. In order to continue the problem solving process, these errors and conflicts must be recognized. Once an error has been recognized, additional tests can be made to resolve the discrepancy, or more data can be collected in an effort to outweigh the false information.

People do not always approach decision making in a formal, rational manner. Their decision making processes are strongly biased by how they perceive the current 'state' of the world surrounding them. Unfortunately, this perceptual information describing the 'state' of the world is not easily expressed in simple truths and falsities. As this information is translated to a formal representation scheme for expert system development, it becomes less accurate, more inconsistent with its original meaning, or at times it is lost all together.

It is not only important to evaluate the available information but also to recognize what information remains unknown. By understanding what information remains unknown, a person may choose to take specific actions to obtain the evidence required to eliminate this ignorance.

To properly describe the decision making process, it is imperative to understand the difference between a partial belief and total ignorance. Ignorance is characterized as a situation where none of the relevant information required to support a decision is available. In other words, there is no supporting evidence present that will bias a choice between any of the alternative solutions. On the other hand, a situation representing partial belief exists when not all the information required to make a completely certain decision is available, though a subset of the necessary information is available. Based on this subset of information, a person will not be able to make a certain choice between alternative solutions. She will, however, be able to limit the number of alternatives that are feasible. This difference between ignorance and partial belief will be discussed further as it relates to the formalisms described in the following chapters.

## Chapter Three

### Probability Theory

Probability theory has played an active role in decision making in AI and expert systems for over 10 years [Nutter, 1987]. At the same time, the role of probability theory in reasoning with uncertainty continues to be a widely debated subject. A fact in favor of probability theory is that it provides a proven mathematical formalism for dealing with uncertainty. The opponents to probability, however, are quick to point out that the requirement for numbers or evidence that is not generally available is a severe shortcoming. At the same time, the evidence that is available is often discounted on the basis that it is subjective. A source of contention with Bayesian Theory is the strong independence assumptions required in order to make this theory computationally feasible. It is also argued [Zadeh, 1986] that the inadequacy of probability theory stems from its lack of expressiveness, especially for describing 'fuzzy' events.

#### 3.1 Background : What are probabilities?

Numerical probability theory has been interpreted in many different ways. Each of these different

interpretations or theories has produced its own specific 'language' to be used to express probability judgement [Shafer, 1986].

Traditionally, probabilities have been defined in terms of frequency ratios. A set of all possible hypotheses or outcomes, generally known as the sample space, is defined first. Each element in the sample space is assigned a weight ranging from zero to one. This weight defines the likelihood of the occurrence of the element resulting from a statistical experiment. That is to say, given a sufficiently large number of samples or tests, the likelihood of an element (A) is defined as the proportion of the number of outcomes that correspond to element (A) to that of all elements that fall in this set. A weight of zero represents that the given hypothesis is false, whereas a weight of one represents that it is true. An event is defined as a subset of the sample space. The probability of an event (E) is then defined as the sum of all the weights assigned to elements in E.

In AI, the frequency interpretation of probability has limited application. First of all, the amount of data required to derive a probability would not be available in most cases. Secondly, the data that is available is often very subjective with no strong statistical foundation. It should be noted, however, that in areas that are well studied and for which large amounts of statistical data are available, this approach offers a good method for reasoning.

In AI, due to the nature of the majority of domains,

large amounts of statistical data is generally unavailable. As a result of these data limitations, the Bayesian approach has adopted the definition of a probability as a subjective measure of belief. In this approach, the probability of an event, given certain evidence, is a subjective measure of the degree a person believes the event will occur. The 'chance' that an event will occur is attached directly to the event.

A generalization of the Bayesian Theory is the theory of belief functions or the Dempster-Shafer Theory. In this theory, the probability is a subjective measure of belief as in the Bayesian Theory. The Dempster-Shafer Theory differs from the Bayesian approach in that the belief is the measure of the degree a person believes that the evidence proves the hypothesis to be true, not whether the hypothesis itself is true. The 'chance' is attached to whether the evidence supports the hypothesis.

Mycin certainty factors were developed as an early model for inexact reasoning for use in medical diagnosis. In Mycin, the probability is a subjective measure of a physician's belief that a hypothesis is supported by a given observation. The theoretical basis for certainty factors is in the interpretation of probability known as confirmation [Shortliffe, 1976]. Confirmation of a hypothesis does not indicate that a hypothesis is true or proven, it merely indicates that there is evidence to support the hypothesis. As in the Dempster-Shafer Theory, the 'chance' is attached



to whether the evidence supports the hypothesis.

All three of these approaches use the standard calculus of probability. Their differences lie in the fact that they apply this calculus in different ways. The next section will present the Bayesian Theory of Probability in considerable detail paying particular attention to the features and assumptions that are presently subject to much debate. In section 3.3, the model of inexact reasoning implemented in the expert system, Mycin, will be presented. The Dempster-Shafer theory will be discussed further in section 3.4.

### 3.2 Bayesian Probability Theory

The Bayesian Theory of probability is based on the work of Thomas Bayes (1702 - 1761) [Shafer, 1986]. In this approach, a probability is a subjective measure of certainty. The probability that hypothesis (H) will occur represents the degree to which a person believes it will occur, given the current evidence. The 'chance' that a specific hypothesis is true or false is represented by a probability value in the range of  $[0,1]$ . A hypothesis assigned a probability of one is believed to be totally true, whereas a probability value of zero would imply absolutely false. All alternative hypotheses in a given sample space are assigned probability values such that their sum would be one.

Many early expert systems used Bayesian Theory as a foundation for their designs for dealing with uncertainty and this approach does, in fact, provide a good mathematical formalism for dealing with uncertainty [Clancy, Shortliffe, 1984]. Unfortunately, due to several inherent limitations of Bayes Theorem, these early systems were not implemented with true Bayesian mechanisms but instead with ad hoc schemes using heuristic methods. A discussion of the mathematical model of the Bayesian Theory will be presented in the following sections and the assumptions and inherent

limitations of this model will be highlighted.

### 3.2.1 Conditional Probabilities

Bayesian probability is based on the idea of conditional probabilities, which are useful in situations where the information concerning an event is incomplete or otherwise uncertain. A conditional probability is the probability of a hypothesis (H) occurring based on the fact that event (E) has already occurred. More formally, a conditional probability is represented as:

$$P(H|E) = N$$

which is read as the probability of hypothesis (H) given event (E). The probability is set to a value (N) between zero and one, which represents our belief that the hypothesis is true. Consider the following example.

$$P(\text{It is raining}|\text{It is lightning})$$

This represents the likelihood that it is raining given the fact that we know that it is lightning. If it was our belief that it always rains when there is lightning, we would set  $P(\text{It is raining}|\text{It is lightning})$  to one.

An unconditional probability, in contrast, is the probability of an hypothesis before any other information or evidence is known. In discussions of Bayesian Theory these are most often referred to as prior probabilities, because they are set prior to any evidence. A prior probability is

represented as  $P(H) = N$ , where  $N$  is a number between zero and one representing the degree of belief that a hypothesis is true before any other information is known.

In order to build a formal definition of conditional probabilities, let us return to the idea of frequency ratios [Lukacs, 1972]. Given a statistical experiment where  $n$  trials were made and events  $(X)$  and  $(Y)$  were observed. The relative frequency of event  $(X)$  is:

$$RF(X) = N_X / n \quad ( 3.1 )$$

where  $N$  represents the number of observations of event  $(X)$ . Likewise, the relative frequency of the event  $(X \cap Y)$  representing that both events were observed is

$$RF(X \cap Y) = N_{X \cap Y} / n \quad ( 3.2 )$$

What we are really interested in is the relative frequency of event  $Y$  given event  $X$ . In other words, we need to know the number of outcomes that correspond to event  $(Y)$  with respect to a reduced sample space. This reduced sample space is composed of all hypotheses from the original set in which event  $(X)$  is true. The relative frequency of  $Y$  given  $X$  is represented as

$$RF(Y|X) = \frac{N_{X \cap Y}}{N_X}$$

by substitution from 3.1 and 3.2 we get

$$RF(Y|X) = \frac{RF(X \cap Y)}{RF(X)} \quad ( 3.3 )$$

Given that the number of trials of the experiment is large, equation 3.3 suggests the following definition.

Let  $(\Omega, S, P)$  be a probability space and  $X$  and  $Y$  be two events. Suppose  $P(X) > 0$  then :

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad ( 3.4 )$$

is called the conditional probability of  $Y$  given event  $X$  has occurred. Now let  $X$  and  $Y$  be two events and suppose that  $P(X) > 0$  and  $P(Y) > 0$ . From 3.4 we see that

$$P(X \cap Y) = P(X) P(Y|X)$$

$$P(X \cap Y) = P(Y) P(X|Y)$$

By eliminating  $P(X \cap Y)$  we obtain

$$P(Y|X) = \frac{P(Y) P(X|Y)}{P(X)} \quad ( 3.5 )$$

### 3.2.2 Bayes' Theorem

Bayes' Theorem specifies a way that conditional probabilities can be calculated. From equation 3.5 we see that

$$P(H|E) = \frac{P(H) P(E|H)}{P(E)} \quad ( 3.6 )$$

Where  $P(H)$  and  $P(E)$  represent prior probabilities that hypothesis  $(H)$  or event  $(E)$  occur, respectively.  $P(E|H)$  represents the conditional probability of the event  $(E)$

given hypothesis (H). The left side of the equation represents the posterior probability that (H) occurs after the evidence is known. Equation 3.6 is generally known as Bayes' formula. Bayes' Theorem can be derived from this formula by applying the total probability rule to the denominator. The Total Probability Rule states:

Let A be a set of mutually exclusive and exhaustive events, let B be an arbitrary event,

$$P(B) = \sum P(A_j) P(B|A_j)$$

Therefore, applying this rule to equation 3.6 we arrive at Bayes' Theorem:

Assume a set of mutually exclusive and mutually exhaustive events or hypotheses (H) in sample space (S),

$$P(H_i|E) = \frac{P(H_i) P(E|H_i)}{\sum_j P(H_j) P(E|H_j)}$$

The Total Probability rule introduces the requirement that all hypotheses in the sample space are mutually exclusive. This means that at any given time only one hypothesis in the sample space can occur. As a basis for the remainder of our discussion, Bayes' formula ( Equation 3.5 ) will be used.

Most decision making environments are not generally characterized by a single piece of evidence. Instead, decisions are made based on a collection of information and evidence. As new information is gathered, beliefs must be

revised to select this new evidence. Bayes' formula can be extended to accommodate multiple pieces of evidence as follows:

$$P(H|E_1...E_n) = \frac{P(H) (E_1...E_n|H)}{P(E_1...E_n)} \quad ( 3.7 )$$

Unfortunately as the amount of evidence increases, the number of probabilities required also increases, making this solution computationally impractical. One way to simplify the above equation to a more workable form is to apply the concept of statistical independence.

### 3.2.3 Statistical Independence Assumptions

The preceding sections have been concerned primarily with probabilities of events based on the occurrence of other events or conditional probabilities. Conditional probabilities are useful in describing situations where evidence has a direct affect on our belief that another situation or hypothesis is either true or false. At the other extreme, however, we need to characterize a situation where the occurrence of an event has no bearing on our belief in a particular hypothesis. In this case, event (E) and hypothesis (H) are totally independent.

If it is thought that events X and Y are independent of each other, it follows that the probability of Y given event X is merely equal to the probability of Y alone. This is represented as:

$$P(Y|X) = P(Y)$$

By substituting this into equation 3.4 we see that two events are independent if:

$$P(X \cap Y) = P(X) P(Y)$$

This can be extended to include all relevant events:

$$P(X_1 \cap \dots X_n) = P(X_1) \dots P(X_n) \quad ( 3.8 )$$

This describes events that are independent of each other in relation to the world. In order to incorporate the idea of conditional probabilities, equation 3.8 can be extended further to describe events that are independent in relation to a subset of the world in which a particular hypothesis is true. This can be represented as follows:

$$P(X \& Y|H) = P(X|H) P(Y|H)$$

By adopting these independence assumptions, Bayes' formula can be written as:

$$P(H|E_1 \& E_2) = \frac{P(H) P(E_1|H) P(E_2|H)}{P(E_1) P(E_2)} \quad ( 3.9 )$$

In this form, the probabilities required are reduced to those required in the case of one event [Charniak, 1984]. As new evidence is gathered, the hypothesis must be re-evaluated based on the new evidence. From 3.9, it is apparent that as new evidence is received the probability of the hypothesis can be updated easily.



#### 3.2.4 Ignorance

The Bayesian Probability Theory does not offer a way to explicitly represent total ignorance. In order to simulate the total lack of knowledge, the maximum entropy assumption is used. The maximum entropy assumption assumes that all events are independent and then distributes the uncertainty as evenly as possible over all events [Cheeseman, 1985]. As a result each event is assigned a probability value representing the least amount of commitment. For example, in the case of two independent events A and B the probability would be distributed as follows:

$$P(A) = P(B) = .5$$

This sets up a neutral background that any new changes can then be compared against. The validity of these assumptions will be further discussed in section 3.5.3.

### 3.3 MYCIN - An Early Expert System

Mycin medical diagnosis system was developed as part of the Stanford Heuristic Programming Project by Edward Shortliffe and in collaboration with the Infectious Disease Group at Stanford Medical School in 1974 [Michie, 1982]. Mycin was developed to aid physicians in the diagnosis and treatment of meningitis and blood infections. One of the key challenges faced in Mycin was how to account for the uncertainty that is inherent in clinical decision making. [Buchanan, Shortliffe, 1984].

Mycin is based on an interactive dialogue with a physician, during which the physician is asked a number of questions concerning the condition of their patient. A backward chaining inference method is then used to backtrack through an and/or tree to determine the disease and the appropriate treatment. The benefit of backward chaining is that the groups of questions asked by the system are focused towards a particular hypothesis [Michie, 1982]. This was a critical feature attributing to the acceptance of Mycin by the physicians.

To handle the uncertainty involved in medical diagnosis, Shortliffe and Buchanan developed and implemented a reasoning model in Mycin which allows a physician to express varying degrees of belief in facts and hypotheses. This degree of belief is characterized as a probabilistic

weight called a certainty factor. During an interactive session with Mycin, a physician could express his degree of certainty of an event by entering a number between 1 and 10 which would then be automatically converted to probability values [Shortliffe, 1976].

Their goal in developing this reasoning model was to avoid strict Bayesian probability and its inherent assumptions and restrictions. It has been shown, however, that a substantial part of this model can be derived from probability theory and in some cases is equivalent [Adams, 1984]. The definition and formal notation of certainty factors and the functions used to combine them will be presented in the following sections.

### 3.3.1 Mycin's Certainty Factors

Mycin certainty factors are based on two independent units of measure: belief and disbelief. The formal representation of belief is

$$MB[h,e] = x$$

which is read as the measure of increased belief in the hypothesis,  $h$ , given the evidence,  $e$ , is equal to  $x$ . Similarly, the formal notation of disbelief is

$$MD[h,e] = x$$

representing the measure of increased disbelief in the hypothesis,  $h$ , given the evidence,  $e$ . The requirement for two separate and independent measures is due to an interpretation of confirmation theory stating that  $C[h,e] =$

1 - C[not h,e] [Shortliffe, 1976]. Simply, this states that a piece of evidence that supports a particular hypothesis does not necessarily affect the support of the negation of that hypothesis. In support of this, many experts have expressed that though they believe in a hypothesis to a particular degree, x, they are uncomfortable stating that they believe in the negation of the hypothesis to a degree, 1-x. Even though some of the restrictions inherent in probability theory can be avoided by using two independent measures of belief, certainty factors retain a strong foundation in Bayesian probability theory.

These measures of belief (MB) and disbelief (MD) can be represented in terms of Bayesian probability as follows:

$$MB[h,e] = \frac{P(h|e) - P(h)}{1 - P(h)} \quad ( 3.9 )$$

$$MD[h,e] = \frac{P(h) - P(h|e)}{P(h)} \quad ( 3.10 )$$

where P(h) represents the subjective belief in the hypothesis prior to any evidence and P(h|e) represents the conditional probability given the known evidence. In subjective probability theory, all alternative hypotheses must be assigned beliefs that sum to one. Therefore, 1-P(h) represents the disbelief in the hypothesis, h, and equation 3.9 represents the proportionate decrease in disbelief given the evidence [Buchanan,Shortliffe, 1984]. If a piece of

evidence increases our belief in the hypothesis,  $P(h|e)$  will be greater than  $P(h)$  and the value of MB will increase representing the growth of our belief. On the other hand, if a piece of evidence decreases our belief in a hypothesis,  $P(h|e)$  would be less than  $P(h)$  and the value of MD would increase representing our increased disbelief.

The certainty factor combines these two independent measures of belief and disbelief into a single measure as follows:

$$CF[h,e] = MB[h,e] - MD[h,e]$$

Substituting the probability ratios, equations 3.9 and 3.10, into this formula shows that both the prior probability,  $P(h)$ , and the conditional probability,  $P(h|e)$ , are combined into the single measure of  $CF[h,e]$ . Because of this, it has been suggested that the certainty factor may be a more naturally intuitive term for an expert to express [Buchanan, Shortliffe, 1984]. The reasoning model implemented in Mycin assumes that the values of MB and MD received from the expert are adequate estimates of the values that would be calculated if the necessary probabilities were known.

In Mycin, certainty factors are applied to both rules and facts that contain some level of uncertainty. By definition, MB and MD are restricted to values in the range of  $[0,1]$ . Therefore, CF is limited to values in the range of  $[-1,1]$ . There are a number of special cases that define

the properties of these measures. In the case where  $MB[h,e] = 1$  and  $MD[h,e] = 0$ , the expert is absolutely certain of the hypothesis and  $CF[h,e] = 1$ . On the other hand, when  $MB[h,e] = 0$  and  $MD[h,e] = 1$ ,  $CF[h,e] = -1$  representing absolute disbelief in the hypothesis or absolute certainty in the negation of the hypothesis. In the situation where  $MB[h,e] = MD[h,e] = 1$ , there is conflicting or contradictory evidence and  $CF[h,e]$  is undefined.

### 3.3.2 Combining Evidence in Mycin

Our belief in a hypothesis or conclusion is usually based on numerous pieces of information. In Mycin, as new evidence is obtained the values of MB and MD are adjusted to reflect the joint effect of all the evidence on our belief. It is important to remember that since MB and MD are independent measures, the combination of favoring and disfavoring evidence is done independently.

The following functions define how incrementally acquired evidence is combined in Mycin. It is important to point out that an underlying assumption of these functions is that of independence. That is to say, it is assumed that all the evidence is independent from each other [Adams, 1984]. As new evidence is obtained that increases our belief in a hypothesis, MB is adjusted as follows:

$$MB[h,s_1 \& s_2] = MB[h,s_1] + MB[h,s_2](1 - MB[h,s_1]) \quad (3.11)$$

In the case that  $MD[h,s_1 \& s_2] = 1$ , by definition MB will

equal 0. On the other hand, to reflect disfavoring evidence, MD will be adjusted as follows:

$$MD[h, s_1 \& s_2] = MD[h, s_1] + MD[h, s_2](1 - MD[h, s_1])$$

Simply stated, as new evidence is acquired the values of MB and MD are increased proportionally to the belief or disbelief already present.

In Mycin, functions were also defined to allow the description of our belief or disbelief in multiple hypotheses given a particular fact or observation. Functions are defined for both the Boolean 'and' and 'or' operations. Our belief and disbelief in the conjunction of two different hypotheses are defined as follows:

$$MB[h_1 \& h_2, e] = \min (MB[h_1, e], MB[h_2, e])$$

$$MD[h_1 \& h_2, e] = \max (MD[h_1, e], MD[h_2, e])$$

Our belief and disbelief in the disjunction or the boolean 'or' of two hypotheses are represented by the following equations.

$$MB[h_1 \text{ or } h_2, e] = \max (MB[h_1, e], MB[h_2, e])$$

$$MD[h_1 \text{ or } h_2, e] = \min (MD[h_1, e], MD[h_2, e])$$

In viewing these functions, there are some underlying assumptions concerning the relationship of the hypotheses. For example, consider the case where hypothesis,  $h_1$ , and  $h_2$  are mutually exclusive. In this case, the conjunction of the hypotheses would equal zero regardless of our belief in either hypothesis. The disjunction of the hypotheses, on

the other hand, would be equal to a number larger than our belief in either hypothesis separately [Adams, 1984].

In our decision making, uncertainty is not only apparent in our beliefs of a particular hypothesis but also in our beliefs of evidence or observations. In order to account for this, certainty factors are applied to both facts and rules in Mycin. To insure that our actual belief in the supporting evidence is properly reflected in the measure of our belief or disbelief of the hypothesis, Shortliffe and Buchanan defined the following functions which incorporate the actual strength of the evidence.

$$MB[h,s] = MB'[h,s] * \max ( 0, CF[s,e] )$$

$$MD[h,s] = MD'[h,s] * \max ( 0, CF[s,e] )$$

In these functions  $MB'[h,s]$  and  $MD'[h,s]$  represent the measure of belief and disbelief in the hypothesis,  $h$ , given that we are totally certain of the given evidence,  $s$ . In Mycin, these represent the decision rules that are acquired from the experts [Buchanan, Shortliffe, 1984].  $CF[s,e]$  represents the certainty factor describing our actual belief in the evidence,  $s$ , based on the prior evidence,  $e$ . The examples in the following sections will illustrate how these functions are actually used in Mycin.

### 3.3.3 Certainty Factors - Examples

Below is an example of an actual decision rule that



could be found in the expert system Mycin [Waterman, 1986].

IF 1) The strain of the organism is grampos, and  
2) The morphology of the organism is coccus, and  
3) The growth conformation of the organism is chains

THEN

There is suggestive evidence (0.7) that the identity of the organism is streptococcus.

In more generic terms, this rule takes the form of

IF X & Y & Z

THEN H with certainty of 0.7.

In mathematical notation this rule would be represented as

$$CF[h, x \& y \& z] = 0.7$$

As decision rules are acquired from experts, certainty factors are assigned to these rules based on the absolute belief in the truth of the supporting evidence. In this example, our expert has expressed a relatively high level of belief that the hypothesis is true given that the three underlying facts are known to be certain. Therefore, a certainty factor of 0.7 has been assigned to this rule.

Consider now, that the expert's actual belief in the supporting evidence is represented by the following certainty factors.

X with certainty 0.5  
Y with certainty 0.7  
Z with certainty 0.3

Our experts prior belief in the hypothesis is given as

$$CF'[h, x \& y \& z] = 0.7.$$

By applying the functions defined in section 3.3.2 to incorporate the actual strength of evidence, we can calculate the experts actual belief in the hypothesis.

$$CF[h, x \& y \& z] = CF'[h, x \& y \& z] * \max(0, CF[x \& y \& z])$$

$$CF[h, x \& y \& z] = 0.7 * \max(0, CF[x \& y \& z])$$

Now by applying the functions describing the conjunction of hypotheses we arrive at

$$CF[h, x \& y \& z] = 0.7 * \max(0, \min(0.5, 0.7, 0.3))$$

$$CF[h, x \& y \& z] = 0.7 * 0.3$$

$$CF[h, x \& y \& z] = 0.21$$

The expert expressed a relatively low level of certainty of the supporting evidence and consequently the value of certainty assigned to the hypothesis has been adjusted downward. It is worth noting that if an expert has no knowledge concerning the truth or falsity of a piece of evidence he would assign a value of zero certainty to the evidence. Missing information is simply disregarded from the rule. For example,  $MB[h, s_1 \& s_2] = MB[h, s_1]$  in the case that the information concerning  $s_2$  is missing.

It is also possible that a given hypothesis is the result of several different decision rules. In this case, as new evidence is obtained to substantiate rules, the actual belief in the hypothesis must be recalculated as a combination of all the relevant rules. For example, consider these two rules:

RULE 1 :        IF X & Y  
                 THEN H with certainty 0.6

EVIDENCE: X with certainty 0.5  
          Y with certainty 0.7

RULE 2 :        IF Z  
                 THEN H with certainty 0.8

EVIDENCE: Z with certainty 0.5

First we must calculate the actual belief in the hypothesis for each rule based on the present belief in the supporting evidence.

RULE 1:  $CF[H, X \& Y] = 0.6 * \max ( 0, \min(0.5, 0.7) )$   
 $CF[H, X \& Y] = 0.3$

RULE 2:  $CF[H, Z] = 0.8 * \max ( 0, 0.5 )$   
 $CF[H, Z] = 0.4$

The certainty factors from both rules must now be combined by applying equation 3.11.

$CF[H, (X \& Y) \& Z] = CF[H, X \& Y] + CF[H, Z] ( 1 - CF[H, X \& Y] )$   
 $CF[H, (X \& Y) \& Z] = 0.3 + 0.4(0.7)$   
 $CF[H, (X \& Y) \& Z] = 0.58$

#### 3.3.4 Ignorance

Mycin certainty factors do not offer an unambiguous representation of ignorance or total lack of knowledge. In the event that an expert has no evidence or knowledge

concerning a particular hypothesis,  $h$ , his belief and disbelief would be equal to 0. In this case,  $CF[h,e]=0$  would indicate that the evidence is independent of the hypothesis and neither confirms or disconfirms it [Shortliffe, 1984].

Unfortunately, a certainty factor of zero may also arise due to equal non-zero values of MB and MD. In this case, the evidence that has been observed both confirms and disconfirms the hypothesis by equal amounts resulting in a net belief of zero.

### 3.4 Dempster-Shafer Mathematical Theory of Evidence

The Dempster-Shafer Mathematical Theory of evidence was first developed by Arthur Dempster in the 1960's. Glenn Shafer further extended the theory and published "A Mathematical Theory of Evidence" in 1976. The Dempster-Shafer Theory uses the same standard calculus of probability as the Bayesian Theory, although it applies this calculus in a much different way. The Dempster-Shafer Theory deals with weights of evidence and numerical degrees of support based on evidence rather than on Bayesian probabilities.

This theory has several characteristics that make it an attractive approach to approximate reasoning. First of all, the Dempster-Shafer theory has the ability to narrow the hypothesis set with the accumulation of evidence. This can be achieved because the evidence accumulated by the expert does not bear on a single hypothesis in the hypothesis set but instead bears on a larger subset of this set. The actual order in which evidence is gathered will not affect the solution. This theory allows an explicit distinction between lack of knowledge and certainty.

#### 3.4.1 Representation of Uncertain Information

The representation of evidence in the Dempster-Shafer Theory begins with the frame of discernment (  $\Theta$  ) [Shafer,

1976]. The frame of discernment is the set of all possible hypotheses or events in a domain. All possibilities in a given frame of discernment are mutually exclusive and exhaustive. The set of hypotheses or events to which belief can be assigned is represented by  $2^\Theta$ , corresponding to all possible subsets of  $\Theta$ . Figure 3.4.1 represents the set of all possible subsets over the frame of discernment - [Spruce, Fir, Pine].

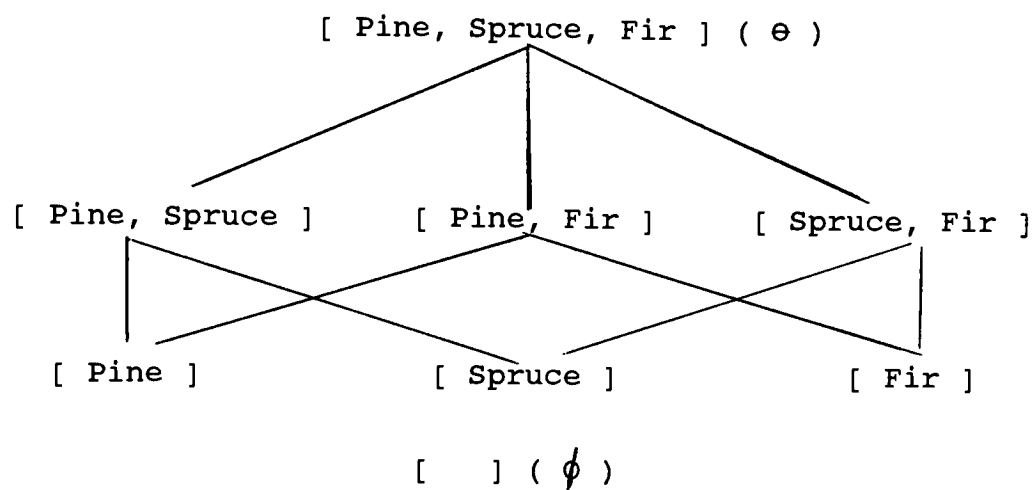
Given a piece of evidence, the Dempster-Shafer theory indicates a belief in the evidence by assigning a number in the range of  $[0,1]$ . A numerical function used to represent our exact belief in a proposition is defined as the basic probability assignment. The basic probability assignment is defined as follows:

If  $\Theta$  is a frame of discernment, then the function  $m:2^\Theta \rightarrow [0,1]$  is called the basic probability function whenever

- 1)  $m(\emptyset) = 0$
- 2)  $m(\Theta) = 1$

This implies that no belief ought to be committed to the empty set and one's total belief is equal to one. Each subset of  $\Theta$  is assigned a value of  $m$  such that all the numbers will sum to 1. The probability assignment  $m(A)$ , represents the belief committed exactly to the proposition  $A$ . This belief cannot be subdivided by the subsets of  $A$ ; it represents only the belief committed to  $A$  alone.

The total belief that is committed to a hypothesis is



The subsets of the set of coniferous trees ( $2^\Theta$ )

Figure 3.4.1

represented by the belief function. The total belief in a given hypothesis (A) is the sum of all the exact beliefs in the hypotheses that imply A and the exact belief in A itself.

$$\text{BEL}(A) = \sum_{B \subseteq A} m(B)$$

To obtain the total belief in the hypothesis A, one must add to  $m(A)$  the quantities  $m(B)$  for all proper subsets B of A.

The basic probability assignment can then be recovered from the belief function:

$$m(A) = \sum_{\substack{B \subseteq A \\ \text{for all } A \subseteq \Theta}} (-1)^{|A-B|} \text{Bel}(B)$$

It follows that for a given belief function there is only one basic probability assignment, and that for a given probability assignment there is only one belief function. Therefore, the same information is being conveyed by the belief function and the basic probability function.

Corresponding to belief functions, there are three other functions that are useful in characterizing evidence. The Doubt function expresses the degree to which the evidence refutes a hypothesis A. The Doubt function is represented as follows:

$$\text{DOU}(A) = \text{BEL}(\bar{A})$$

Our doubt of A is equal to our belief in the negation of A.

The Plausibility function, on the other hand, expresses



the extent of which one fails to doubt or refute the hypothesis A. This function is expressed as follows:

$$\begin{aligned} PL(A) &= 1 - BEL(\bar{A}) \\ PL(A) &= 1 - DOU(A) \end{aligned}$$

$PL(A)$ , also called the upper probability function, expresses the extent to which one finds the hypothesis plausible. In terms of the basic probability assignments, the plausibility function can be stated as follows:

$$\begin{aligned} Pl(A) &= \sum_{B \subset \Theta} m(B) - \sum_{B \subset \bar{A}} m(B) \\ Pl(A) &= \sum_{B \cap A \neq \emptyset} m(B) \end{aligned}$$

Another function that is useful when dealing with belief functions is the commonality function. The commonality function is defined as follows:

$$Q(A) = \sum_{\substack{B \subset \Theta \\ A \subset B}} m(B)$$

The commonality  $Q(A)$ , is equal to the sum of all subsets that have A as their subset. It is interesting to note that  $Q(\emptyset) = 1$ .

The belief function now can be represented in terms of commonality  $Q(A)$ .

$$Bel(A) = \sum_{B \subset \bar{A}} (-1)^{|B|} Q(B)$$

assignment now can be recovered. The plausibility and commonality functions are also related as follows:

$$PL(A) = \sum_{B \subset A} (-1)^{|B|+1} Q(B)$$

Therefore, it follows that the belief, plausibility, commonality and basic probability assignment are in one to one correspondence and represent the same information.

For each body of evidence, these functions define an "evidential interval" [Garvey, Lowrance, Wesley, 1984] within which our belief about a hypothesis must lie. The belief function represents the lower bound of this interval, and the plausibility function represents the upper bound. When our knowledge of a hypothesis (A) is certain and precise, the BEL(A) and PL(A) are equal to one. In this special case, this theory reduces to general Bayesian probability.

### 3.4.2 Belief, Commonality and Plausibility - An Example

In an attempt to clarify the definitions in the previous section, consider again the scenario in Figure 3.4.1. Based on evidence that is gathered, the variable (X) could take on one of three values: Spruce, Pine or Fir. The number of subsets to which we will assign belief is represented by  $2^\theta$ . Based on the available evidence, we assign a basic probability assignment to all members of  $\theta$ , such that they sum to one. The probability assignments for each set are shown in Figure 3.4.2. Figure 3.4.2 also shows the calculated values for the belief, plausibility and

Set	m	Bel	Pl	Q
-----				
Pine	0.2	0.2	0.8	0.8
Spruce	0.1	0.1	0.4	0.4
Fir	0.1	0.1	0.5	0.5
Pine, Spruce	0.2	0.5	0.9	0.3
Pine, Fir	0.3	0.6	0.9	0.4
Spruce, Fir	0.0	0.2	0.8	0.1
Pine, Spruce, Fir	0.1	1.0	1.0	0.1
$\phi$	0.0	0.0	0.0	1.0

Calculations for the belief, plausibility, and commonality for a given body of evidence pertaining to the subsets of the set of coniferous trees.

Figure 3.4.2

commonality functions, based on these probability assignments. As shown, there is no direct evidence for [Spruce,Fir] though the belief function is greater than zero. This is due to the non-zero values for the basic probability assignments (m) pertaining to the subsets [Spruce] and [Fir].

The plausibility of a given set is represented by the summation of the probability assignments for all sets that do intersect with that set. In the case of the set [Spruce,Fir], all sets with the exception of [Pine] intersect with it. Therefore, the Plausibility value is equal to 0.8.

Commonality of the set [Spruce,Fir] is equal to the sum of the probability assignments for all sets that have [Spruce,Fir] as a subset. In this case, the only values to be summed are those for [Spruce,Fir] and [Spruce,Fir,Pine]. The resulting value for Q is 0.1.

### 3.4.3 Combining Evidence from Multiple Sources

In order to accommodate information obtained from multiple sources, Dempster-Shafer has developed a method that provides the facility for the combination of evidence. There are generally two scenarios where evidence must be combined. The first scenario is when two or more facts in a given body of evidence need to be combined to obtain their joint effect, and the second scenario is when two entirely different bodies of evidence need to be combined. Given a

specified  $\theta$ , and a given body of evidence, we have already shown how to combine elementary facts. We will now concentrate on combining evidence that supports or negates a hypothesis from two totally different bodies of evidence. The Dempster-Shafer model provides a formal mechanism to combine different bodies of evidence.

Given two belief functions based on two pieces of evidence, Dempster's Rule of Combination computes a new belief function that represents the impact of the combined evidence. The belief function to be combined must be over the same frame of discernment. The order in which evidence is combined is not important since the combination rule is both commutative and associative. Let  $m_1$  and  $m_2$  be basic probability assignments on the same frame of discernment  $\theta$ . Let  $m = m_1 \oplus m_2$ , defining their orthogonal sum. The new probability assignment, as a result of combining  $m_1$  and  $m_2$ , is defined as:

$$m(A) = k \left( \sum_{X \cap Y = A} m_1(X) * m_2(Y) \right)$$

Where  $X$  and  $Y$  are all subsets whose intersection is  $A$ .  $m(A)$  is the new basic probability assignment as a result of combining  $m_1$  and  $m_2$ .

In order to intuitively understand this combination rule, it can be represented in a geometric model [Barnett, 1981]. Let the basic probability assignments for each body of evidence be depicted as portions of a line segment of length one. Now Consider a unit square with two sides, one

side representing  $m_1$  and the other side representing  $m_2$ . Figure 3.4.3 shows the two line segments  $m_1$  and  $m_2$  orthogonally combined to form a square. The intersection of a vertical strip  $m_1(A)$  and a horizontal strip  $m_2(A)$  results in mass  $m_1(A_j) * m_2(A_j)$  associated with the combination of  $A \cap B$ . There may be more than one rectangle in the square committed for a particular subset; therefore, we must sum all values.

For rectangles where there is no intersection,  $a \cap b = \emptyset$ , a commitment is made to the empty set. Dempster states, however, that  $m(\emptyset) = 0$ . The assigned values are therefore, normalized so the  $m_1 \oplus m_2 (\emptyset) = 0$ , and all new values are between 0 and 1. This is accomplished by setting  $m_1 \oplus m_2 (\emptyset) = 0$  and multiplying each other value by  $K$ , where  $K$  is equal to  $1/(1-k)$ . The value of  $k$  is defined as the sum of all non-zero values assigned to the empty set. The weight of conflict between two sources is defined as  $\log(K)$ .

#### 3.4.4 Dempster's Rule of Combination - An Example

Let us look at some examples of Dempster's rule of combination. First consider a particularly interesting example due to Zadeh (Figure 3.4.4). There are two sources of evidence,  $m_1$  and  $m_2$ .  $M_1$  shows a large amount of certainty in the possibility of hypothesis  $a$  and also in the impossibility of hypothesis  $c$ .  $M_2$ , on the other hand, is very certain in the possibility of hypothesis  $c$  and the impossibility of hypothesis  $a$ . The result shows hypothesis

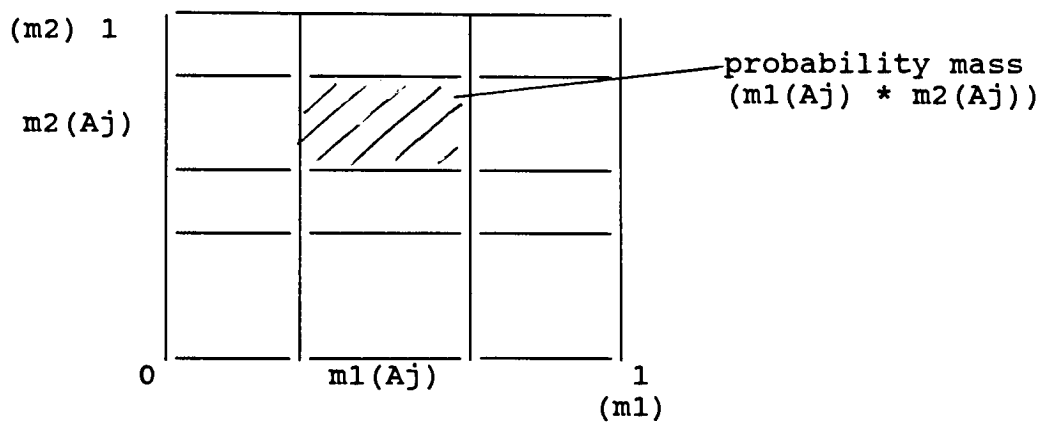


Figure 3.4.3

Set	m1	m2	Orthogonal sum m1 + m2	Normalized orthogonal sum (m1 + m2)/(1-k)
a	0.9	0.0	0.0	0.0
b	0.1	0.1	0.01	1.0
c	0.0	0.9	0.0	0.0

	a (0.0)	b (0.1) <sup>(m2)</sup>	c(0.9)
(0.9) a (m1)	0	0.09 $\phi$	0.81 $\phi$
(0.1) b	0	0.01 b	0.09 $\phi$
(0.0) c	0	0	0

$$k = \sum_{x \cap y = \phi} m1(X) * m2(Y) = 0.99$$

$$\log(K) = \log(1 / (1-k)) = \text{weight of conflict} = 2$$

Figure 3.4.4



b as the only possible answer even though both  $m_1$  and  $m_2$  considered it to be highly improbable which is shown by the probability assignment of 0.1. Shafer has provided an answer for this rather counter-intuitive result. Each source has distributed the entire unit weight among the hypothesis a,b,and c, which implies that the source is entirely reliable. In other words, neither source has any degree of uncertainty or ignorance. Due to the large discrepancy between the beliefs of the two sources, it seems that the reliability of each source is slightly questionable. If the sources are not considered to be fully reliable, the result is not so unexpected. It is interesting to note that the resultant probability assignments do not reflect the conflict seen in the probability of the sources.

Another interesting example of Dempster's rule for combination is shown in Figure 3.4.5. Again there are two sources of evidence  $m_1$  and  $m_2$ . Both sources have committed the same amount of belief to hypotheses a, b and c respectively. This can be seen in that the two sources have identical probability assignments for each hypothesis. The result shows that the dominant hypothesis c, is reinforced. This shows that the rule of combination reinforces the agreements between sources and discards the conflicts. This example implies that as more evidence is collected, we can narrow in on a particular hypothesis.

Figure 3.4.6 illustrates a combination of beliefs over the frame of discernment [Spruce, Pine, Fir]. This

Set	m1	m2	Orthogonal sum m1 + m2	Normalized orthogonal sum (m1 + m2)/(1-k)
a	0.3	0.3	0.09	0.26
b	0.3	0.3	0.09	0.26
c	0.4	0.4	0.16	0.48

	a (0.3)	b (0.3)	c (0.4)
(0.3) a	0.09 a	0.09 $\phi$	0.12 $\phi$
(0.3) b	0.09 $\phi$	0.09 b	0.12 $\phi$
(0.4) c	0.12 $\phi$	0.12 $\phi$	0.16 c

$$k = \sum_{X \cap Y = \phi} m1(X) * m2(Y) = 0.66$$

$$\log(K) = \log(1 / (1-k)) = \text{weight of conflict}$$

= 47

Figure 3.4.5

Set	m1	m2	Orthogonal sum m1 + m2	Normalized orthogonal sum (m1 + m2)/(1-k)
Pine (P)	0.2	0.0	0.17	0.24
Spruce (S)	0.1	0.3	0.22	0.315
Fir (F)	0.1	0.1	0.14	0.20
Pine, Spruce (PS)	0.2	0.3	0.11	0.16
Pine, Fir (PF)	0.3	0.0	0.03	0.04
Spruce, Fir (SF)	0.0	0.2	0.02	0.03
Pine, Spruce, Fir (PSF)	0.1	0.1	0.01	0.015

	P	S	F	(m2) PS	PF	SF	PSF
(.2) P	0	.06 $\phi$	.02 $\phi$	.06 P	0	.04 $\phi$	.02 P
(.1) S	0	.03 S	.01 $\phi$	.03 S	0	.02 S	.01 S
(.1) F	0	.03 $\phi$	.01 F	.03 $\phi$	0	.02 F	.01 F
(.2) PS	0	.06 S	.02 $\phi$	.06 PS	0	.04 S	.02 PS
(.3) PF	0	.09 $\phi$	.03 F	.09 P	0	.06 F	.03 PF
(.0) SF	0	0	0	0	0	0	0
(.1) PSF	0	.03 S	.01 F	.03 PS	0	.02 SF	.01 PSF

$$k = \sum_{X \cap Y = \phi} m1(X) * m2(Y) = 0.3$$

$$\log(K) = \log(1 / (1-k)) = \text{weight of conflict} \\ = 0.154$$

Figure 3.4.6

completes our discussion of the coniferous tree example.

### 3.4.5 Ignorance

The Dempster-Shafer theory of evidence allows for an explicit distinction between a total lack of knowledge or ignorance and a high level of disbelief. When a person is ignorant concerning a hypothesis, he has no belief either in the hypothesis or in its negation. In terms of the Dempster-Shafer Theory ignorance is defined as follows:

$$\text{Ignorance} = \text{PL}(A) - \text{BEL}(A)$$

Consider the case of determining the sex of a person. The frame of discernment consists of male and female:

$$\theta = [\text{Male}, \text{Female}]$$

The set of all possible subsets is

$$2^\theta = [0, \emptyset, \text{Male}, \text{Female}]$$

$\text{Bel}(\text{Male})$  represents the degree of belief that the person is male and likewise  $\text{Bel}(\text{Female})$  represents the belief that the person is female. In the case, that very little evidence was available that specifically supported the belief that the person was either male or female,  $m(\text{Male})$  and  $m(\text{Female})$  would be set very low. It would follow that  $\text{Bel}(\text{Male})$  and  $\text{Bel}(\text{Female})$  also would be very low. In the case that no evidence at all was available,  $\text{Bel}(\text{Male})$  and  $\text{Bel}(\text{Female})$  would be zero and the  $\text{Bel}(\theta)$  would be one. This situation, called the vacuous belief function, represents total ignorance. Generally stated, the vacuous belief function is

obtained by setting:

$$m(\theta) = 1, m(A) = 0, \text{ for all } A \neq \theta$$

$$\text{Bel}(\theta) = 1, \text{Bel}(A) = 0, \text{ for all } A \neq \theta$$

### 3.5 Comparison of Bayesian, Mycin and Dempster-Shafer Approaches

In an attempt to describe rational decision making, all three of these approaches to uncertainty provide a mathematical formalism primarily based in probability theory. A 'rational' decision is defined as a decision that abides by the requirements and limitations of the underlying functions and data structures that define each specific theory. People, however, do not always approach decision making in a truly rational manner. An 'actual' decision is more of a psychological concept that describes how people truly make decisions in situations of uncertainty [Carnap, 1971]. All three of these theories require the expert to adhere to their specific functional requirements therefore maintaining a rational approach to problem solving. However, the actual level of restriction imposed by these requirements varies among the three approaches. The Dempster-Shafer theory, by offering a more naturally intuitive way to express uncertainty, appears to be the least restrictive.

#### 3.5.1 Representation of Belief

In the Bayesian probability theory, all alternative hypotheses are assigned probability values so that they sum to one. The expert is forced to divide his belief among

each individual hypothesis in the sample space. This is a relatively easy task when we are sure of our beliefs in all of the propositions. In the case that we are not confident of our belief in a proposition, however, we might find it very difficult or uncomfortable to assign a number to our belief. In the event that an expert is either doubtful of his belief or in fact ignorant concerning a given hypothesis, he is still required to assign a probability value to the hypothesis when using the Bayesian approach. This often leads to an overcommitment of his actual belief.

In Mycin, this restriction is avoided by independently measuring belief and disbelief. In this case, the expert is no longer forced to assign probabilities to all alternative hypotheses regardless of his confidence in his belief. However, the expert is restricted to assigning point probabilities to individual hypothesis so that the upper limit of their sum is one.

In the situation where our knowledge is incomplete, the Dempster-Shafer theory offers a good alternative approach. In the Dempster-Shafer theory, belief is assigned to all possible subsets of the hypothesis space or the frame of discernment so that the total belief will sum to one. This provides the expert more alternative ways to express his beliefs, no longer restricting him to supporting only individual hypotheses. In this approach, the expert is no longer forced to commit to a belief that he is not confident in, therefore providing a better representation of his true beliefs. This advantage of the Dempster-Shafer theory is

most evident in the case of total ignorance.

### 3.5.2 Ignorance

Ignorance is characterized as a situation where there is no relevant information or knowledge available to support a decision. In this case, the Dempster-Shafer surpasses both Bayesian probability and Mycin by offering a way to explicitly represent ignorance. In the Dempster-Shafer approach, as described in section 3.4.5, when a person is ignorant concerning a hypothesis he can assign all his belief to the frame of discernment ( $\Theta$ ) which represents the entire set of possible hypotheses. He is not forced to assign his belief to any individual hypothesis, therefore allowing a more comfortable and natural way to express his lack of knowledge without overcommitting his belief.

In Bayesian probability, Cheeseman argues that by applying the maximum entropy assumptions a neutral background is established representing the least amount of commitment. Opponents argue that this assumes more than is truly known by assuming that all events are independent. Though this argument is justified, statistical independence is an assumption found in all three approaches and is further discussed in section 3.5.4. Even though theoretically, maximum entropy represents the least amount of commitment, it does not offer the expert a clear and unambiguous way to express his lack of knowledge. A person is still required to assign a point probability to individual hypotheses,



therefore forcing a statement of belief that exceeds his actual belief.

In Mycin, ignorance is represented by assigning a certainty factor equal to zero to a hypothesis. Given that a person has no information concerning a hypothesis, he would be naturally more comfortable assigning a value of zero to his belief in contrast to a non-zero probability representing maximum entropy. Even though this approach provides a more intuitive way to express ignorance than Bayesian probability it is still a very ambiguous representation. Although both Bayesian probability and Mycin offer reasonable methods to simulate ignorance, neither approach offers a representation as explicit as the Dempster-Shafer theory.

### 3.5.3 Prior Probability - Problem of Subjectivity

All three of these approaches rely on the integrity of the subjective beliefs of experts. They differ in that Bayesian probability also requires prior probabilities or beliefs. Prior probabilities are considered to be more subjective than conditional probabilities and are often disregarded as being valid. The reason that prior probabilities are considered to be highly subjective is because prior to any evidence being known the 'state' of the world of each individual expert may be quite different and therefore their stated prior probabilities may vary tremendously. In Mycin, though the theoretical basis of MB

and MD included prior probabilities experts are not specifically asked to estimate their prior beliefs. In practice, experts are asked to estimate the certainty factors, which combine both prior probability ( $P(H)$ ) and conditional probability ( $P(H|E)$ ) into one value. These questions concerning the validity or quality of prior probabilities are not in themselves sufficient reason to abandon probability theory. Expert systems can be designed in such ways that they can improve and refine our initial estimates as new evidence is gathered [Nutter, 1987].

Many still maintain that the strict statistical concept of probability is the only legitimate interpretation. However, in the majority of cases the required statistical data to support this approach is simply not available. Therefore, the idea of a subjective or personal probability must be applied to support decision making and problem solving.

#### 3.5.4 Statistical Independence & Mutual Exclusivity

Statistical independence and mutual exclusivity are assumptions that can be found in all three theories discussed here. In most situations, these assumptions predicate more information than is actually known and in many cases, especially those involving medical diagnosis, are actually known to be invalid. The repercussions of these assumptions on the outcome of these approaches is not well documented which is due in part to the limited number

of actual implementations.

A number of studies have been conducted on Mycin to verify the integrity of the program results. One such study was designed to compare the certainty factor computed from known probability data (  $P(H)$  and  $P(H|E)$  are known ) and the certainty factor computed using the combining functions and known values for MB and MD. The outcome of this study showed that the largest discrepancies between the two CF values were due to interrelated evidence and longer reasoning chains [Buchanan, Shortliffe, 1984]. However, their overall impact of these assumptions in Mycin appears to be rather small due to the predominant use of relatively short reasoning chains. [Adams, 1984].

A number of solutions have been offered to help accomodate the restrictions of independence and mutual exclusivity. One such solutions suggests that all non-independent evidence be grouped together into a single piece of evidence or a single rule. It is important to note that though this may help solve the problem theoretically it makes the actual implementation much more unmanageable by increasing the difficulty in obtaining rules. There also have been a number of heuristic techniques proposed to avoid the restrictions of mutual exclusivity. A good example can be seen in the expert system Caduceus. [Charniak, McDermott, 1984].

### 3.5.5 Inferencing Methods

A key function of expert systems is the ability to make decisions and draw inferences from the available information. It is here that the Dempster-Shafer Theory falls short by not offering any effective means for decision making. In the Dempster-Shafer approach, ignorance and uncertainty are represented directly in belief functions and are carried through the combination process [Barnett, 1981]]. The BEL and PL functions can be calculated for each statement defining the bounds of the evidential interval. There is presently no accepted mechanism for decision making based on these bounds. Possibly the value for Bel or Pl could be used solely for decision making. However, which of these values should be used is unknown [Barnett, 1981].

The Bayesian approach differs in that it masks our ignorance in the prior probabilities. The probabilities calculated by the updating mechanism can be used in making decisions that will minimize the expected loss [Thompson, 1985]. The reasoning methods used in Mycin were implemented on this basis in conjunction with a deductive style of reasoning. There are a number of drawbacks to deductive reasoning which have limited the success of Mycin. In most cases, deductive reasoning does not represent how people actually make decisions. People tend to incorporate a more inductive-style of reasoning based on their personal intuition and perception of the 'state' of their situation. Inductive reasoning usually results in a solution that best

represents or is most typical of the data or evidence. Deductive reasoning, also has been shown to have severe shortcomings with default reasoning or typicality. Typicality-based uncertainty, which is involved in generalizations such as "Birds Fly", centers on whether an individual has a property that is typical of other things of its kind [Nutter, 1987].

## Chapter Four

### Fuzzy Set Theory

The theory of fuzzy sets was first introduced by Lotfi Zadeh in 1965, due to his strong interest in the analysis of complex systems [Kandel, 1982]. Since that time, the development of fuzzy set theory has grown significantly. Possibility Theory was later introduced by Zadeh in 1978 as a development of fuzzy sets. Unlike probability-based approaches, Possibility Theory offers a formal method of representing uncertainty due to vague concepts or hypotheses where distinct boundaries or definitions do not exist. Fuzzy set theory is particularly well suited for natural language applications where vague linguistic variables such as tall, short, young etc. are prevalent. Fuzzy set theory is a generalization of abstract set theory, in that all the definitions and theorems that apply to fuzzy sets also hold true for non-fuzzy sets [Kandel, 1982].

The basis of fuzzy sets and fuzzy logic will be presented in sections 4.1 and 4.2, respectively. Possibility Theory will be examined in section 4.3 as a model for representing uncertainty. A discussion of how ignorance can be represented in possibility theory also will be included. A comparison of both Possibility and Probability Theory will be discussed in the final section.

#### 4.1 Definition of Fuzzy Sets

Since a fuzzy set is a generalization of a crisp set, it is first necessary to understand the concept of a crisp set. In a crisp set, each element of the universal set is assigned a value of either 1 or 0 indicating whether it is contained within or is strictly outside of the specified set. In describing a crisp set, all elements of the universal set are separated into two groups, members and non-members. A sharp distinctive boundary delineates the members from the non-members. Formally, in a crisp set each element is defined by a discrimination function:

$$u_A(X) = \begin{cases} 1 & \text{if and only if } X \in A \\ 0 & \text{if and only if } X \notin A \end{cases}$$

An example of a crisp set is the set of all integer temperature values in the range [50F, 100F]. Given the universal set  $X = \{0, 30, 60, 90, 120\}$ , each element is assigned a value of 1 or 0 indicating its membership in the set (T), as seen in figure 4.1.

Elements	(T)
0	0
30	0
60	1
90	1
120	0

Figure 4.1

It can be determined with certainty that each element is either contained within the set or is strictly outside of the set. A sharp distinctive boundary delineates members and non-members. In a fuzzy set, this sharp boundary is replaced by a gradual 'fuzzy' boundary.

In defining a fuzzy set, each element of the universal set is assigned a value in a specified range representing the membership grade of each element in respect to the specified set. This is formally represented by the following membership function:

Let  $X$  = Universal set,

$\mu_A : X \rightarrow [0,1]$

Where  $[0,1]$  denotes the interval of real numbers from 0 to 1 [Klir,Folger, 1988].

As in a crisp set, if an element of the universal set is clearly contained within the specified set it is assigned a value of 1. Likewise, if an element is strictly outside of the set it is assigned a value of 0. A fuzzy set differs from a crisp set in that it also allows the assignment of values in the range of 0 to 1 representing an elements position in the vague boundary between being either wholly within or outside of the set. Clearly, a crisp set represents the special case of a fuzzy set when no uncertainty or vagueness exists concerning the membership of the elements in the specified set. In this case, all elements of the universal set would be assigned values of 0



or 1, and no elements would exist in the 'fuzzy' boundary.

To help illustrate the concepts of a fuzzy set, consider the following two fuzzy sets: COLD (set of cold temperatures) and HOT (set of hot temperatures). Given the universal set,  $X = \{0, 30, 40, 50, 60, 70, 80, 100\}$ , each element is assigned a value in the range of 0 to 1 representing its membership in the particular fuzzy set (Figure 4.2).

Elements (Temperatures, F)	COLD	HOT
0	1	0
30	1	0
40	0.8	0.1
50	0.5	0.3
60	0.3	0.4
70	0.1	0.7
80	0	1
100	0	1

FIGURE 4.2

In the fuzzy set HOT, the elements  $\{40, 50, 60, 70\}$  are in the boundary area, indicating that there is some level of uncertainty or vagueness concerning their membership in the fuzzy set. This uncertainty is due to the vagueness inherent in the linguistic concept of hot. The amount of uncertainty or vagueness is indicated by the size of the fuzzy boundary. It is important to note that the membership grades assigned to each element are subjective in nature, representing the degree that a person believes that an element belongs in the particular fuzzy set.

The support of a fuzzy set is equal to the crisp set of all the elements that have non-zero values. From Figure

4.2, the support of the fuzzy set, HOT, is

$$\text{SUPP}(\text{HOT}) = \{40, 50, 60, 70, 80, 100\}$$

A fuzzy set is considered to be normalized if at least one element of the set is assigned the highest possible membership grade. The height of a fuzzy set is equal to the highest membership grade assigned to any of its elements. Both fuzzy sets, HOT and COLD, are normalized with a height of 1.

Fuzzy modifiers such as very, really, usually, etc., can be applied to linguistic variables to create new fuzzy sets or subsets. Generally speaking, the subsets resulting from fuzzy modifiers are more concise and, therefore, less uncertain. Fuzzy modifiers are often referred to in the literature as linguistic hedges [Kandel, 1982]. As an illustration, if the modifier 'very' is applied to the fuzzy set, HOT, a new fuzzy subset, VERYHOT, is created. As shown in Figure 4.3, the support of the subset, VERYHOT, is half

Elements (Temperatures, F)	HOT	VERYHOT
0	0	0
30	0	0
40	0.1	0
50	0.3	0
60	0.4	0
70	0.7	0.4
80	1	0.8
100	1	1

FIGURE 4.3

of the support for the subset, HOT, indicating that the new

fuzzy set is a more concise set. The number of elements found in the fuzzy boundary has also decreased to include only two elements indicating that the new fuzzy subset is less uncertain or vague.

## 4.2 Fuzzy Logic

In classical or two-valued logic, every hypothesis is assigned a value of either true (1) or false (0). In other words, each hypothesis is considered to be very clear and precise, either being completely true or completely false. It is the goal of fuzzy logic to provide a method to do approximate reasoning with imprecise or fuzzy propositions [Klir, Folger, 1988]. In fuzzy logic, each hypothesis is assigned a value in the range of  $[0,1]$  representing the degree that the hypothesis is either true or false. It is important to note, that in the case that no uncertainty exists, fuzzy logic can be reduced to two-valued logic.

As a basis for fuzzy logic, the functions of union, intersection and complement have been defined in fuzzy set theory as follows:

Let A and B be fuzzy subsets of X

1) Complement of a fuzzy subset

$$u_{\bar{A}}(X) = 1 - u_A(X)$$

2) Intersection of two fuzzy subsets

$$u_{A \cap B}(X) = \min[u_A(X), u_B(X)]$$

### 3) Union of two fuzzy subsets

$$u_{A \cup B}(X) = \max[u_A(X), u_B(X)]$$

The complement, intersection and union functions are illustrated in Figure 4.4, as they pertain to the fuzzy subsets, HOT and VERYHOT.

TEMP	HOT	VERYHOT	HOT U VERYHOT	HOT $\cap$ VERYHOT	HOT
0	0	0	0	0	1
30	0	0	0	0	1
40	0.1	0	0.1	0	0.9
50	0.3	0	0.3	0	0.7
60	0.4	0	0.4	0	0.6
70	0.7	0.4	0.7	0.4	0.3
80	1	0.8	1	0.8	0

FIGURE 4.4

In fuzzy logic, an imprecise or fuzzy proposition can consist of fuzzy predicates (hot, cold), fuzzy modifiers (very, usually) and also fuzzy truth values (very true, fairly false) [Klir, Folger, 1988]. For example, consider the fuzzy proposition 'Yesterday was a cold day'. The truth value of this proposition depends on the membership grade of the actual temperature recorded yesterday in the fuzzy subset, COLD, and also on the strength of the truth being claimed. Consider for example the following two claims:

'Yesterday was a cold day is very true'

'Yesterday was a cold day is fairly true'

Each of these claims will result in different truth values

based on the fuzzy set representing the appropriate truth claim expressed.

Fuzzy truth values can be represented by their own fuzzy subsets as described below. Given a universal set,  $U$ , of real numbers in the range,  $[0,1]$ , and assuming that truth can be represented by these numerical values, varying degrees of membership can be assigned to the linguistic variables, true and false. Consider for example that absolutely true is represented numerically by the value of 1 and absolutely false is represented by the value of 0. Within this unit interval, as the membership grade of true increases towards 1 the membership grade for false would decrease toward 0. Fuzzy modifiers such as very or fairly now can be applied to these fuzzy sets to create new fuzzy subsets as shown in Figure 4.5.

U	True	Fairly True	False	Very False
0	0	0	1	1
0.2	0.2	0.3	0.8	0.8
0.4	0.4	0.5	0.6	0.5
0.6	0.6	0.7	0.4	0.2
0.8	0.8	1	0.2	0
1	1	1	0	0

Figure 4.5

### 4.3 Possibility Theory

Possibility Theory was initially proposed by Zadeh in 1978 as a development of fuzzy set theory. The main purpose behind Possibility Theory was to provide a good formalism to

deal with the inherent fuzziness found in much of the information that is used to make decisions. Possibility theory is centered around the concept of a possibility distribution.

A possibility distribution, denoted as  $\Pi_X$ , is defined as a fuzzy constraint on the values that may be assigned to X. Zadeh defines the possibility distribution in terms of a fuzzy restriction as follows:

Let F be a fuzzy subset and U be the Universal set. Let X be a variable taking values in U, and let F act as a fuzzy restriction.

Then the proposition "X is F" translates into

$$R(X) = F.$$

Associated with this proposition is a possibility distribution which is equal to R(X)

$$\Pi_X = R(X)$$

This possibility distribution function of X is defined to be numerically equal to the membership function of F.

$$\Pi_X = \mu_F$$

As an example, consider the proposition 'Yesterday was a cold day'. In this example the fuzzy set, COLD, restricts the values that can be assigned to yesterday's temperature. In formal terms, the above proposition can be written as

$$R(\text{Temperature}(\text{Yesterday})) = \text{Cold}$$

In order to relate this idea of a fuzzy restriction to a possibility distribution, consider the temperature, 40F,

whose degree of membership in the fuzzy set, COLD, is 0.8. The degree of membership, 0.8, is interpreted as a degree of compatibility of 40F with the concept of cold. Now by combining this with the proposition that 'Yesterday was a cold day' we can state that the degree of possibility that yesterday was 40F is 0.8 given the fact that yesterday was a cold day. The compatibility, 0.8, of the value 40F with the concept cold becomes the possibility of that temperature given that the proposition is true.

The possibility distribution,  $\Pi_X(u)$ , is numerically equal to the membership function and therefore may be assigned values in the range of  $[0,1]$ . The possibility distribution for the fuzzy set, COLD, is  $\Pi_X = 1/0 + 1/30 + 0.8/40 + 0.5/50 + 0.3/60 + 0.1/70 + 0/80 + 0/90 + 0/100$  where each term is read as the possibility that X is 40 given that X is cold is equal to 0.8. Although for a given set, a possibility distribution may appear to be very similar to a probability distribution, there lies a number of fundamental differences between the two which will be investigated in the section 4.5.

#### 4.4 Ignorance

In Possibility Theory, total ignorance is represented by a possibility distribution consisting of all 1's.

$$\Pi_X(u) = \{1,1,1,\dots,1\}$$

Generally speaking, the less specific the evidence or

information, the larger the possibility distribution representing the high degree of uncertainty. Hence, total ignorance is represented by the 'largest' possibility distribution, consisting of all 1's [Klir, Folger, 1988]. On the other hand, the situation where we have perfect evidence and consequently no uncertainty, would be represented by the following possibility distribution.

$$\Pi_X(u) = \{1, 0, 0, 0 \dots 0\}$$

It seems naturally intuitive to state that in the absence of any relevant evidence, all propositions are possible. Unfortunately, in possibility theory we are still forced to address all propositions as singletons as we were in probability theory.

#### 4.5 Comparison of Possibility and Probability

Both Probability and Possibility Theory offer methods to represent uncertainty that is found in situations involving problem solving and decision making. However, the type of uncertainty that is best handled by these two theories is quite different. Probabilities are well suited to represent the uncertainty inherent in our belief of a particular hypothesis, given the evidence and information that is currently available. This uncertainty lies in the ambiguity associated with being faced with many well defined alternative hypotheses or solutions. On the other hand, possibilities best represent the uncertainty that results



from the use of vague or indistinct concepts or linguistic terms.

As an iterative part of the decision process, a person must choose between the available alternative hypotheses, based on their belief that a particular hypothesis is true. In this type of environment, the alternative solutions are clear and well defined, the uncertainty lies with our belief of which alternative to choose. This type of uncertainty, most prevalent in diagnostic and problem solving environments, is best handled by a probabilistic approach. For instance, suppose the problem that we are faced with is to determine a person's age. Our alternative solutions, in this example, can be viewed as well defined crisp sets such as the set of all people 16 years old. The information we have concerning this person's age is limited to the following two observations; the person drives a car and the person is currently attending highschool. Given this information we may assign our probabilistic beliefs as follows:

Age	15	16	17	18	19	20
	.05	.3	.25	.25	.1	.05

In this example, our 'best guess' is 16 and consequently is assigned the highest value. The uncertainty is inherent in the information or lack of information that we have available to us to support our decision. On the other hand, the alternative solutions are very clear and certain. If we

were given perfect evidence, we could predict with certainty the person's age. Possibilities represent the uncertainty due to vagueness found in situations where distinct definitions do not exist. Possibility theory concentrates on the actual meaning of the information rather than the measure of it [Klir, Folger, 1987]. This type of uncertainty is prevalent in natural language applications. For example, consider the task of determining whether a person is young, given their age. In this example, the problem is to determine how this person's age relates to the vague concept of young. Here the uncertainty lies in the subjective interpretation of the concept young.

The difference between probabilities and possibilities can be best illustrated by the following example given by Zadeh [Zadeh, 1977]. Consider the statement "Hans ate X eggs for breakfast" with X taking values in  $U = \{1, 2, 3, 4, \dots\}$ . A possibility distribution,  $P(u)$ , may be associated with X by interpreting  $P(u)$  to be the degree of ease with which Hans can eat u eggs. A probability distribution,  $Pr(u)$ , can also be associated with X by interpreting  $Pr(u)$  as the probability of Hans's eating u eggs for breakfast. The possibility and probability distributions associated with X may look as follows:

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

From this example, we can observe that a high degree of possibility does not imply a high degree of probability, nor does a low degree of probability imply a low degree of possibility. However, Zadeh did propose that there does exist a weak heuristic link between probabilities and possibilities which he called the 'possibility/probability consistency principle'. This principle is based on the observation that as the possibility of an event decreases so does the probability of the event. In other words, if an event is impossible it is also improbable. A further understanding of the differences between probability and possibility can be gained by comparing the concepts of the probability and the possibility measure. A probability measure represents the chance that a hypothesis will occur, based on certain evidence. A possibility measure represents the degree that a hypothesis is feasible. Probability measures are represented by the sum of the probabilities of the events in the distribution space, whereas possibility measures are represented by the maximum value of the possibility distribution. A possibility measure is similar to the plausibility measure of the Dempster-Shafer Theory. In fact, it has been proposed that possibility measures are a special subset of plausibility measures [Klir,Folger, 1988].

With the introduction of possibility theory began a debate concerning the validity of the new theory and its applicability. Specifically, Cheeseman claims that a probabilistic approach may be used to describe the

uncertainty due to 'fuzzy' concepts, and therefore eliminating the need for a new theory [Cheeseman, 1986]. Cheeseman argues that a possibility can be represented as the probability that "it is possible that event X occurs" instead of the proposition "event X occurs". One problem with this interpretation is that in the first proposition our event space does not include all possible values of X, whereas in the later case it does [Bhatnagar, Kanai, 1986]. In support of his arguments, Cheeseman further identifies the underlying dependency assumptions for the max/min rules found in fuzzy logic as severe shortcomings. For example consider the following rule for the intersection of two fuzzy sets,

$$u_{A \cap B}(X) = \min[u_A(X), u_B(X)]$$

In the case that A and B are mutually exclusive, this rule will state their possibility as the minimum of the two and it should instead be stated as zero. However, Probability Theory also suffers from certain assumptions. In fact, Cheeseman defends the underlying independence assumptions found in probability by stating that in order to make decisions with uncertain information one is forced to make assumptions. In the defense of Possibility Theory, Zadeh contends that the vagueness surrounding a particular concept is not equal to the uncertainty about its truth value in a particular context. Although fuzzy set theory lacks much of the empirical verification found in probability theory, it

offers a good method to reason with a type of uncertainty that is inherent in natural language.

## Chapter 5

### Non-Monotonic Logic

The preceding sections have dealt with various numeric approaches to uncertainty. Each of these approaches were based on unique mathematical formalisms based on their own set of requirements and assumptions. One requirement shared by each of these methods, is that the expert must express his uncertainty in terms of specific numbers. This has been viewed as a severe shortcoming due to the fact that people often find it difficult to translate their beliefs into hard numbers. Consequently, many have argued that non-numeric approaches should be further investigated as more realistic methods for modeling uncertainty.

Non-monotonic logic offers a good non-numeric approach, and as such the type of uncertainty that it best represents is quite different from the uncertainty represented by the previous numerical approaches. Some of the better known non-monotonic logics include McCarthy's circumscription, Reiter's default logic and McDermott and Doyle's non-monotonic logic [Shoham, 1987]. Each of these logical systems offer unique approaches each worthy of consideration. For the purposes of this thesis, however, the discussions in the following sections will be limited to the non-monotonic logic developed by McDermott and Doyle.

## 5.1 Basic Characteristics

The ability to revise our beliefs based on the most current information is essential in modeling environments that are changing or uncertain. In order to truly reflect the underlying uncertainty of a hypothesis, incomplete and contradictory information also must be considered. Classical or traditional logic is monotonic in nature and, therefore, does not allow the revision of previous conclusions as new information is obtained. Consequently, traditional logic cannot be effectively applied to dynamic problem solving situations having incomplete or uncertain information.

Non-monotonic logic, on the other hand, allows previous conclusions to be invalidated or revised based on inconsistencies that arise from new information. Non-monotonic logic is based on first order predicate calculus and requires that each variable is restricted to a single value. Consequently, in the case that the evidence is either contradictory or missing, assumptions must be made up front to resolve any conflicts, and a single value must be assigned. It should be noted that due to this restriction there is presently no way to represent ignorance with this method. In the numerical approaches discussed in the previous chapters, hypotheses could be partially believed. In the case of non-monotonic logic, each hypothesis is either

believed or disbelieved based on the initial assumptions that are applied. The uncertainty is masked in the suppositions made and is not carried through the decision making process. Therefore, the level of confidence we have in our decision strongly depends on the nature of these underlying assumptions.

## 5.2 Modal Logic

The non-monotonic logic introduced by McDermott and Doyle incorporates modal operators to represent the idea of "consistency". In this logic, classical logic is extended by adding modality,  $M$ , as follows:

Infer  $(Mp)$  from the inability to infer  $(NOT\ p)$

Here  $Mp$  is the modal operator representing "consistency" and can be read as "p is consistent with the theory" [Maselli, 1985]. The modal operators allow a contingent conclusion to be drawn based on the fact that there is no evidence to the contrary. As new information is received, the previous conclusions may be withdrawn. As an example consider the following.

- 1)  $BIRD \wedge M(CAN-FLY) \rightarrow CAN-FLY$
- 2)  $PENGUIN \rightarrow BIRD \wedge NOT(CAN-FLY)$
- 3)  $BIRD$

In this theory we can prove

- 4)  $CAN-FLY$



the premises of 1,2 and 3. Now if the following new information is received

#### 5) PENGUIN

we find that our previous conclusion (4) is now inconsistent and must be re-evaluated. The modal operator changes its meaning based on the context of the situation.

### 5.3 Doyle's Truth Maintenance System

Doyle's Truth Maintenance System (TMS) provides a framework for updating and revising models as new information is obtained. The TMS is a program that maintains the knowledge base of a reasoner in a tree structure. In this tree structure each node represents a hypothesis and the children of each node represent the various justifications and assumptions. In turn, each child node may contain hypotheses and justifications. If the system receives new input that is inconsistent with the hypothesis or theorem, TMS revises the necessary assumptions in the tree in order to restore the system to consistency. TMS also will inform the user of all the changes that were made.

TMS has been able to effectively demonstrate a non-monotonic logic system. The decision reached by TMS, however, does not explicitly reflect the amount of underlying uncertainty in the assumptions. In large expert systems, with longer inferencing chains, the system response for truth maintenance systems may degrade significantly.

## Chapter 6

### Summary and Conclusion

Probability theory has played an important role in modeling uncertainty in problem solving environments for many years. This can be seen not only in the Bayesian Probability Theory but also as the basis for a number of subsequent theories such as Mycin certainty factors and the Dempster-Shafer Theory of Evidence. In all three of these theories, probabilities are defined as a subjective measure of belief, in contrast to the traditional frequency ratio approach.

In Bayesian probability, the probability or 'chance' that an event will occur, represented by a value in the range  $[0,1]$ , is attached directly to the event. In general, Bayesian probability has experienced limited success due to the large volume of data required and the resulting computational complexity when applied in large expert systems. Its limited success is also due in part to its lack of expressiveness in the representation of partial beliefs, especially for ignorance and 'fuzzy' events.

Mycin represented a tremendous breakthrough by offering one of the first workable implementations of a reasoning model that incorporates uncertainty. In Mycin, a probabilistic weight called a certainty factor is used to

represent one's degree of belief. The theoretical basis for certainty factors lies in confirmation theory.

Consequently, the value assigned to a certainty factor represents the degree of belief that the available evidence supports the hypothesis. One of the goals in developing this reasoning model was to avoid strict Bayesian probability and its inherent assumptions and restrictions. It has been shown, however, that a substantial part of this model can be derived from probability theory, and consequently it suffers from some of the same assumptions, such as statistical independence. Much of Mycin's success has been attributed to its more simplified model, in comparison to Bayesian probability. Its short inferencing chains have minimized the effects of the independence and mutual exclusivity assumptions.

The Dempster-Shafer Theory of Evidence uses the same calculus of probability as the Bayesian theory, however, it applies it in a much different way. This approach is based on numerical degrees of support based on evidence. In other words, the 'chance' is attached to whether the evidence supports the hypothesis. One notable advantage of this theory is that the evidence accumulated by the expert does not bear on a single hypothesis but instead on all possible subsets of the hypothesis set. This provides a truer representation of the expert's beliefs and also allows an explicit representation of ignorance. Drawbacks of the Dempster-Shafer Theory of evidence include the lack of an effective inferencing method and difficulties in defining

frames of discernment in complex domains.

Probability-based theories are effective for expressing partial beliefs in evidence and hypotheses where the uncertainty lies in the context of the hypotheses. These approaches would be best applied to domains that are well defined and where the data is available. It is important to realize that no mathematical model can be expected to give reasonable results unless all of the underlying restrictions and assumptions are followed.

Fuzzy set and Possibility theories, developed by Zadeh, offer formal methods of representing uncertainty due to vague concepts and consequently are very well suited for natural language applications. Possibility theory represents the uncertainty due to unclear or imprecise definitions of the actual meaning of information. In this theory, degrees of membership are used to relate hypotheses to vague or 'fuzzy' concepts. In turn, fuzzy logic offers the functional methodology to do approximate reasoning with fuzzy propositions. A possibility measure represents the feasibility of a hypothesis, whereas a probability measure represents the chance that a hypothesis will occur. Possibility suffers from its own set of restrictions and assumptions such as the underlying dependency restrictions found in the max/min rules. Many have been skeptical of Possibility Theory since it presently lacks much of the empirical testing found in probability theory.

Non-monotonic logic offers a non-numeric approach to

handling uncertainty and, therefore, avoids the difficult translations of beliefs into hard numbers. McDermott and Doyle's non-monotonic logic allows conclusions to be revised as new information is received by incorporating modal operators. These modal operators are used to represent the "consistency" in the model, thereby allowing us to jump to conclusions in the absence of evidence to the contrary. Non-monotonic logic is effective in modeling default reasoning in environments with missing or incomplete information.

In conclusion, there is not one global theory that can provide a panacea for all situations of uncertainty. Probability-based approaches are best in situations where our uncertainty lies in the evidential context of a hypothesis. The Dempster-Shafer Theory appears to be the most expressive approach, especially in describing ignorance. Fuzzy set theory should be used when our uncertainty is due to the vagueness of the information. Non-monotonic logic is applicable in generalizations based on typicality allowing a method for default reasoning.

In order to model human reasoning completely in expert systems, we would need to combine all these modes of uncertainty into one reasoning system. In the majority of domains people must make decisions by incorporating all types of uncertainties including partial evidential information, vagueness and generalizations. In order to ensure that our information is not misrepresented, systems need to be developed that allow experts to represent all

their different types of uncertainty with the most appropriate method. The difficulty lies in successfully combining the numbers from these different methods into rules allowing good inferences to be made. Unfortunately, these types of systems though ideal are not currently developed. The best alternative at present is to choose the best theory for the particular domain, being aware of the limitations and trade-offs that are made.

It is important to remember that not all of the difficulty in developing reasoning systems that handle uncertainty is with the representation of the information and inferencing methods. Many of the limitations are due to the difficulty in acquiring all the information describing the full state of the expert's world or domain. Consequently, the knowledge base that we begin with is usually incomplete. All in all the realm of human reasoning is extremely diverse and complex, and no one theory will have the global key to the answer.

## REFERENCES

- [Adams, 1984] Adams, J. Barclay, "Probabilistic Reasoning and Certainty Factors" in Rule Based Expert Systems, The Mycin Experiments of the Stanford Heuristic Programming Project, edited by Buchanan and Shortliffe, Addison-Wesley Publishing Co., 00 263-271, 1984.
- [Barnett, 1981] Barnett, Jeffrey A., "Computational Methods for a Mathematical Theory of Evidence", Seventh International Joint Conference on Artificial Intelligence, pp. 868-875, 1981.
- [Buchanan, Shortliffe, 1984] Buchanan, Bruce G. and Shortliffe, Edward H. , "A Model of Inexact Reasoning in Medicine" in Rule Based Expert Systems, The Mycin Experiments of the Stanford Heuristic Programming Project, edited by Buchanan and Shortliffe, Addison-Wesley Publishing Co., pp 233-263, 1984.
- [Carnap, 1971] Carnap, Rudolf and Jeffrey, Richard C., Studies in Inductive Logic and Probability, University of California Press, pp. 7-19, 1971.
- [Charniak, 1984] Charniak, Eugene, and McDermott, Drew, Introduction to Artificial Intelligence, Addison-Wesley Publishing Co., pp. 457,482, 1984.
- [Cheeseman, 1985] Cheeseman, Peter, "In Defense of Probability," International Joint Conference on Artificial Intelligence, pp. 1002-1009, 1985.
- [Cheeseman, 1986] Cheeseman, Peter, "Probabilistic versus Fuzzy Reasoning" in Uncertainty in Artificial Intelligence edited by Kanal and Lemmer, Elsevier Science Publishers (North Holland) pp. 85-101, 1986.
- [Clancy, Shortliffe, 1984] Clancy, William J., and Shortliffe, Edward H., Readings in Medical Artificial Intelligence, Addison-Wesley Publishing Co., 1984.
- [Garvey, Lowrance, Wesley, 1984] Garvey, Thomas D., Lowrance, John D. and Wesley, Leonard P., "Reasoning about Control: An Evidential Approach", SRI International, Technical Note 324, pp1-40, July 1984.
- [Kandel, 1982] Kandel, Abraham, Fuzzy Techniques in Pattern Recognition, John Wiley and Sons Inc., pp. 23-91, 1982.
- [Klir, Folger, 1988] Klir, George J., and Folger, Tina A., Fuzzy Set, Uncertainty and Information, Prentice-Hall, pp. 107-131, 1988.
- [Lukacs, 1972] Lukacs, Eugene, Probability and Mathematical Statistics, Academic Press, 1972.

[Maselli, 1985] Maselli, A. Raul, "Basic Characteristics to Achieve Common Sense Reasoning", March 1985.

[Michie, 1982] Michie, Donald, Introductory Readings in Expert Systems, Gordon and Breach Science Publishers, pp. 3-25, 1982.

[Nutter, 1987] Nutter, Jane T., "Uncertainty and Probability", International Joint Conference on Artificial Intelligence, pp. 373-378, August 1987.

[Shafer, 1976] Shafer, Glenn, A Mathematical Theory of Evidence, Princeton University Press, 1976.

[Shafer, 1986] Shafer, Glenn, "Probability Judgement in Artificial Intelligence," in Uncertainty in Artificial Intelligence, edited by Kanal and Lemmer, Elsevier Science Publishers (North Holland), pp. 127-135, 1986.

[Shoham, 1987] Shoham, Yoav, "Nonmonotonic Logics: Meaning and Utility", International Joint Conference on Artificial Intelligence, August 1987.

[Shortliffe, 1976] Shortliffe, Edward Hance, Computer-Based Medical Consultations : Mycin, American Wlsevier Publishing Co. Inc., pp 159-194, 1976.

[Thompson, 1985] Thompson, Terence R., "Parallel Formulation of Evidential-Reasoning Theories," International Joint Conference on Artificial Intelligence, pp. 321-327, 1985.

[Waterman, 1986] Waterman, Donald A., A Guide to Expert Systems, Addison-Wesley Publishing Co., pp 22-40, 1986.

[Zadeh, 1977] Zadeh, L. A., "Fuzzy Sets as a Basis for a Theory of Possibility" in Fuzzy Sets and Applications: Selected Papers by L. A. Zadeh, edited by Yager, Ovchinnikov, Tong and Nguyen, pp. 193-218, 1987.

[Zadeh, 1986] Zadeh, Lofti A., "Is Probability Theory Sufficient for dealing with Uncertainty in AI: A Negative View," in Uncertainty in Artificial Intelligence edited by Kanal and Lemmer, Elsevier Science Publishers (North Holland), pp. 103-116, 1986.



## ANNOTATED BIBLIOGRAPHY

Adams, J. Barclay, "Probabilistic Reasoning and Certainty Factors" in Rule Based Expert Systems - The Mycin Experiments of the Stanford Heuristic Programming Project, edited by Buchanan and Shortliffe, Addison-Wesley Publishing Co., pp 263-271, 1984.

Adams demonstrates that a substantial part of the Mycin model is based on and at time equivalent to probability theory. Consequently he contends the Mycin model assumes statistical independence and is subjected to the inherent limitations.

Barnett, Jeffrey A., "Computational Methods for a Mathematical Theory of Evidence," Seventh International Joint Conference on Artificial Intelligence, pp 868-876, 1981.

Barnett presents the mathematical basis for Dempster-Shafer Theory. He compares the Dempster-Shafer Theory with Bayesian statistics. He presents all the current drawbacks for Dempster-Shafer Theory. He then presents a computational theory that reduces the calculation time from exponential to linear.

Bhatnagar, Raj K., and Kanal, Laveen N., "Handling Uncertain Information: A Review of Numeric and Non-numeric Methods," in Uncertainty in Artificial Intelligence, edited by Kanal and Lemmer, Elsevier Science Publishers (North Holland), pp. 3-25, 1986.

They briefly compare and contrast several methods of reasoning with uncertainty including: Bayesian statistics, Dempster-Shafer theory, Possibility Theory, Theory of Endorsements and Non-monotonic logic. In their comparison, they focus on three areas: How uncertain information is represented, how information is combined and how inferences are made.

Buchanan, Bruce G. and Shortliffe, Edward H., "A Model of Inexact Reasoning in Medicine" in Rule Based Expert

Systems - The Mycin Experiments of the Stanford Heuristic Programming Project, edited by Buchanan and Shortliffe, Addison-Wesley Publishing Cp, pp 233-262, 1984.

Buchanan and Shortliffe present the model they developed for approximate reasoning in medicine using certainty factors in Mycin. A brief discussion of Bayesian probability is included as a basis of the decision model. They full describe the model, including definitions, notations and terminology of all the functions. A general discussion of how this model is then implemented in Mycin is included.

Carnap, Rudolf and Jeffrey, Richard C., Studies in Inductive Logic and Probability, University of California Press, pp 7-19, 1971.

A quick review of inductive logic and rational decision making. Statistical probability and personal (subjective) probability are discussed as two different theories of probability. Carnap states the personal concept of probability must be used in decision theory. Personal probability is divided into two version: actual degree of belief and rational degree of belief.

Charniak, Eugene, "The Bayesian Basis of Common Sense Medical Diagnosis," Proceedings of the National Conference on Artificial Intelligence, pp. 70-73, 1983.

Charniak argues that the statistical independence assumptions imposed on Bayes Theorem to make it computationally feasible are not necessary. Thus Bayesian statistics offers a realistic basis for medical diagnosis. Eventhough, Bayesian statistics doesn't work for multiple disease case, it does work well with various heuristic solutions.

Charniak, Eugene, and McDermott, Drew, Introduction to Artificial Intelligence, Addison-Wesley Publishing Co., pp 457-482, 1984.

A good discussion of the use of statistics in plausible reasoning. Bayes Theorem is defined and its limitations are discussed. Includes a discussion of Mycin and Caduceus and how Bayesian statistics and heuristic rules were combined to overcome the limitations of Bayesian statistics.

Cheeseman, Peter, "In Defense of Probability," International Joint Conference on Artificial Intelligence, pp 1002-1009, 1985.

Cheeseman argues that probability is the only scheme needed for reasoning about uncertainty. He defines a probability as a measure of belief rather than a frequency ratio. On this basis he refutes most claims presently held against probability.

Cheeseman, Peter, "Probabilistic versus Fuzzy Reasoning," in Uncertainty in Artificial Intelligence edited by Kanal and Lemmer, Elsevier Science Publishers (North Holland), pp. 85-102, 1986.

Cheeseman demonstrates how probability theory can be used to solve "fuzzy" problems. He maintains the view that probabilities are a measure of belief in a proposition. He presents a detailed example using this view of probability to solve a fuzzy problem. He concludes that in many cases the fuzzy set approach is identical to probability and therefore the fuzzy approach is not necessary.

Clancy, William J., and Shortliffe, Edward H., Readings in Medical Artificial Intelligence, Addison-Wesley Publishing Co., 1984.

A survey of artificial intelligence work in medicine from 1971 to 1981. A collection of works describing much of the work done in this period. Reviews of Internist and Puff are included.

Garvey, Thomas D., Lowrance, John D., and Wesley, Leonard P., "Reasoning About Control: A Evidential Approach." SRI International, Technical Note 324, pp. 1-40, July 1984.

Present a model for reasoning with evidential or uncertain information. The computational model is a dependence graph model of evidential support. The inference engine used is one developed by Lowrance to extrapolate the mass distributions to the remaining dependent propositions.

Garvey, Thomas D., and Lowrance, John D., "Evidential Reasoning: A Developing Concept," IEEE Cybernetics and Society, pp. 6-9, October 1982.

They discuss their method for evidential reasoning in knowledge based systems. Their work is based on the Dempster-Shafer approach. They contrast this to Bayesian methods.

Heckerman, David E., and Horvitz, Eric J., "On the expressiveness of Rule-Based Systems for reasoning with Uncertainty," AAAI proceedings, pp. 121-126, July 1987.

They argue that "certain" beliefs can be represented by rule based systems whereas "uncertain" beliefs can not. This is due to the fact that uncertain beliefs are not modular and the rule based approach cannot express dependencies. They present a modification to rule based systems that will accommodate dependencies and therefore, uncertain beliefs.

Hoenkamp, Edward, "An Analysis of Psychological Experiments on Non-monotonic Reasoning," International Joint Conference on Artificial Intelligence, pp. 115-117, August 1987.

This paper is based on psychological experiments that show that initial inferences made by a person may not change even though the original basis for the inferences is discredited.

Kahneman, D., Slovic, P., and Tversky, A., Judgement Under Uncertainty: Heuristics and Biases, Cambridge University Press, 1985.

This article describes three heuristics that people use when making judgements with uncertainty. It is shown that people using these heuristics often exhibit errors and biases in their judgements.

Lukacs, Eugene, Probability and Mathematical Statistics, Academic Press, 1972.

Probability and Statistics fundamentals.

McCord, James R. and Moroney, Richard, Introduction to Probability, Macmillian Publishing Company Inc., New York, 1964.

Probability and Statistics fundamentals.

Myers, Raymond H., and Walpole, Ronald E., Probability and Statistics for Engineers and Scientists, Macmillian Publishing Company, Inc., 1972.

Probability and Statistics fundamentals.

Nutter, Jane T., "Uncertainty and Probability," International Joint Conference on Artificial Intelligence, pp. 373-378, August 1987.

Nutter argues that there is definitely a place for probability theory in reasoning with uncertainty, though it cannot provide the answer to all problems of uncertainty. Where typicality, vagueness and ignorance are involved, other approaches to uncertainty must be incorporated.

Sage, Andrew P., "On the Management of Information Imperfections in Knowledge-Based Systems," in Uncertainty in Knowledge-Based Systems, edited by Bouchon and Yager, pp3-29, 1987.

This research was concerned with designing knowledge base systems that can deal effectively with imperfect knowledge. It encompassed all phases of design: acquisition, representation, and inferencing with imperfect knowledge. Emphasis on how humans process information for problem solving.

Shafer, Glenn, A Mathematical Theory of Evidence, Princeton University Press, 1976.

Shafer presents a full description of the mathematical theory of evidence which is the basis for the Dempster-Shafer theory. Detailed derivations of all formulas that are the basis of the theory.

Shafer, Glenn, "Hierarchical Evidence," IEEE Second Conference on Artificial Intelligence Applications, pp.

16-21, December 1985.

Shafer reviews the mathematics of the Dempster-Shafer theory. He proposes an algorithm for using this theory in the case of hierarchical evidence.

Shafer, Glenn, "Probability Judgement in Artificial Intelligence," in Uncertainty in Artificial Intelligence, edited by Kanal and Lemmer, Elsevier Science Publishers (North Holland), pp. 127-135, 1986.

Shafer presents two theories of probability judgement: Bayesian Theory and the Theory of Belief functions. Three examples are presented detailing the differences between the Bayesian approach and the belief function of the Dempster-Shafer approach.

Shortliffe, Edward Hance, Computer-Based Medical Consultations : Mycin, American Elsevier Publishing Co., Inc., pp 159-194, 1976.

Shortliffe presents the model for inexact reasoning in medicine that he jointly developed with Buchanan for use in the expert system Mycin. Much of the article is excerpts from Buchanan, Shortliffe, 1984.

Thompson, Terence R., "Parallel Formulation of Evidential-Reasoning Theories," International Joint Conference on Artificial Intelligence, pp. 321-327, 1985.

Thompson discusses five approaches to evidential reasoning: Classical Bayes, Convex Bayes, Dempster-Shafer, Kyburg and Possibility. He reviews each approach according to four categories: Background elements, Observation reports, Updating mechanism and Decision mechanism. He identifies key assumptions, similarities and differences of the approaches.

Turner, Raymond, Logics for Artificial Intelligence, Ellis Horwood Limited, pp. 101-114, 1984.

Turner describes the basics for fuzzy set theory and fuzzy set logic. A general discussion of the applications of this theory in artificial intelligence and expert systems.

Zadeh, Lofti A., "Is Probability Theory Sufficient for dealing with Uncertainty in AI: A Negative View," in Uncertainty in Artificial Intelligence edited by Kanal and Lemmer, Elsevier Science Publishers, (North Holland), pp. 103-116, 1986.

Zadeh refutes the idea that classic probability can properly treat any kind of uncertainty, especially for fuzzy events and fuzzy probabilities. He details a number of examples where fuzzy methods instead of probabilistic methods would be required.

Zadeh, Lofti, A., Fuzzy Sets and Applications: Selected papers by L.A. Zadeh, edited by Yager, Ovchinnikov, Tong and Hguyen, John Wiley and Sons Inc., 1987.

A collections of works by Zadeh. Included is a good detaled study of Fuzzy set theory and decision making in a fuzzy environment. Also looks at the role of fuzzy logic in expert systems.