MULTI-PLAYER BELIEF CALCULI:

MODELS AND APPLICATIONS

by

Shimon Schocken
Information Systems Department
Stern School of Business
New York University

Robert A. Hummel
Courant Institute of Mathematical Sciences
New York University

February 4, 1993

Center for Research on Information Systems
Information Systems Department
Stern School of Business
New York University

Working Paper Series

STERN IS-93-2

Multi-Player Belief Calculi: Models and Applications

Shimon Schocken [1]

Robert A. Hummel [2]

February 4, 1993

In developing methods for dealing with uncertainty in reasoning systems, it is important to consider the needs of the target applications. In particular, when the source of inferential uncertainty can be tracked to distributions of expert opinions, there might be different ways to model the representation and combination of these opinions. In this paper we present the notion of multi-player belief calculi – a framework that takes into consideration not only the 'regular' type of evidential uncertainty, but also the diversity of expert opinions when the evidence is held fixed. Using several applied examples, we show how the basic framework can be naturally extended to support different application needs and different sets of assumptions about the nature of the inference process.

[1] : Department of Information Systems, Management Education Center, Stern School of Business, New York University, 44 W. 4th Street, New York, NY 10003.
Email: `sschocken@stern.nyu.edu`

[2] : Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012. Email: `hummel@.cs.nyu.edu`

1 Introduction

The area of research that is generally referred to as "uncertainty in artificial intelligence" consists of many different methods for reasoning under uncertainty, and the reader is referred to Pearl [9] for a comprehensive review. The methods that are in common use include the Dempster-Shafer theory of evidence (Shafer, [12]), fuzzy set theory (Zadeh, [16]), and Bayesian networks (Pearl, [9]). There is now an entire subfield devoted to the topic, with compiled workshop proceedings [8]. Yet in spite of the great theoretical progress that was made during the last decade, a common criticism of the existing methods is that they are inapplicable to a wide range of real problems. Indeed, theories developed independent of applications tend to lack extensibility and miss important features of standard reasoning. In order to begin to address this problem, we argue in this paper that methods of uncertainty reasoning should be application-driven and sufficiently flexible to accommodate the needs of different problems and inference contexts.

In many methods, degrees of belief are encoded using probabilities. For example, Bayesian networks make use of probabilistic combinations to maintain degrees of uncertainty in propositions (Pearl, [10]). However, it is a well-observed fact that belief and probability are often two separate notions. For example, it is possible to be very certain (or alternatively, very uncertain) about a probabilistic statement. Early expert systems like MYCIN and Prospector represented propositions with pairs of numbers – one number relating to a probability and the other to a degree of certainty – and employed different methods for combining them [1], [3]. Subsequently, most expert systems use some form of an uncertainty reasoning calculus to deal with non-categorical inference rules and uncertain data.

Depending on whether probabilities are interpreted as subjective quantities or as frequentist measures of occurance rates, different methods for interpreting and combining them, result. While some of the methods use pairs of number in place of point probabilities, others use intervals of probabilities (Dempster, [2], Good, [4]). Intervals are equivalent to the "pair of numbers" representation, in the sense that there is a central value and an interval width. However, the concept of an interval leads naturally to the idea that a combination formula can be developed by tracking all possible pairings of values within intervals. One way to view this formulation is to distinguish between two sample spaces. The probabilities might be defined based on occurrences over one sample space, whereas the intervals might represent different estimates of the probability, as computed by or elicited from different agents indexed over a second and separate sample space. This duality is the foundation behind the multi-player belief formulation that is described in this paper.

Perhaps the most well-studied method for representing and combining belief intervals is the Dempster-Shafer theory of evidence, in which the end-points of the intervals are represented through belief and plausibility functions, together with a particular calculus (Shafer, [12]). As we treat in greater detail below, the theory of evidence can be viewed as replacing intervals of probabilities with collections of lists of possibilities (we take a *possibility* to be a subset of hypotheses, propositions, or simply *labels*). This gives rise to a multi-player interpretation, in which different players specify different possibilities and a calculus is used to pool and combine their opinions into a joint opinion.

In the standard Dempster-Shafer model, each 'player' designates a single nonempty subset of labels. Several authors proposed extensions to this basic setting. In Yen's GERTIS system [15], each 'player' designates a collection of disjoint nonempty subsets together with a probability distribution over the collection of subsets. A calculus is then developed using a modified Dempster's formula. Further, a hierarchy of labels is incorporated to allow players to refine the precision with which they designate subcollections of labels. Recently, Tzeng has introduced a mathematical model of uncertain information [14]. The model is related to the Dempster-Shafer calculus, in that there are "messages" that map to subsets of labels, but is considerably more general. Indeed, Tzeng shows how the Bayesian calculus and the Dempster-Shafer calculus fit into the scheme, and demonstrates a range of other possibilities that can be derived from the same model. Each message in Tzeng's formulation may have a weight, and for a given piece of evidence, there is an associated collection of distinct messages, each giving a single subcollection of labels. The collection of possible evidence codes are then given a prior probability distribution, and computations may be made in a Bayesian fashion conditioned on certain codes. The modeling that leads to structure in Tzeng's model occurs when the variety of available messages, codes, and sets of experts, are restricted.

The fact that there are alternative and more general models for handling uncertainty in reasoning systems suggests that any single framework is likely to be inadequate for some applications. Accordingly, in this paper we begin with a simple framework for reasoning under uncertainty that can be easily extended in many different ways. Next, we discuss how particular application needs and assumption sets can be addressed by modifications and extensions of the basic framework.

2 Multi-Player Belief Calculi

In a recent article, Hummel & Manevitz [6] noted that evidential reasoning is characterized by two types of uncertainty: 'intrinsic' and 'extrinsic.' The source of *intrinsic* uncertainty is pluralism, or diversity of opinions: when two or more experts are presented with the same piece of evidence, the conclusions that they draw may be different. In contrast, *extrinsic* uncertainty occurs because different pieces of evidence lead to different conclusions. When two or more pieces of evidence are considered simultaneously, the joint conclusion that they yield can either contradict, amplify, or be quite different than, any one of the individual conclusions. In what follows, evidential reasoning models that take both types of uncertainty into consideration are called *multi-player belief calculi.*

The mathematical foundations of multi-player belief calculi were articulated by Hummel & Landy [5], and the reader is referred there for a detailed analysis. In this section we wish to describe the basic idea by example, without getting into formal notation. Let A, B, and C be three exhaustive and mutually exclusive hypotheses, exactly one of which is known to be true. Suppose that two groups of experts are called upon to assess the likelihoods of these hypotheses in light of two different bodies of evidence, denoted e_1 and e_2. Specifically, all the experts in each group are presented with the same body of evidence and are then asked to state individually the hypotheses that the evidence renders as *unlikely*, in their opinion. Two restrictions are placed on the expert responses. First, the experts must rule out unlikely hypotheses categorically (this restriction will be lifted later). Second, the experts are not allowed to rule out *all* the hypotheses. Thus, each expert opinion can be represented as a binary vector in which 0 codes that the respective hypothesis has been ruled out by the expert and 1 otherwise. For example, the opinion $(1,0,1)$ codes that B was ruled out by the expert, implying the expert's belief that the truth lies in the subset $\{A,C\}$. Since the expert's response must imply at least one likely hypothesis, the opinion (0,0,0) is disallowed, and the case of insufficient reason, or total ignorance, is modeled by the opinion (1,1,1). This particular opinion is consistent with an expert who asserts that the available evidence is not sufficiently useful to rule out *any* hypothesis.

Figure 1 depicts a hypothetical scenario in which the sources of evidence e_1 and e_2 characterize two groups consisting of five and seven experts, respectively. All the experts express their beliefs in a common set of hypotheses – {A,B,C} – and their specific opinions are recorded in the two left-most tables in the figure. As the tables indicate, in the first group two experts say that e_1 rules out C and three experts say that e_1 rules out B. Similarly, in the second group two experts say that e_2 rules out A and C, four experts say that e_2 rules out A, and one expert says that e_2 rules out A and B. How can we summarize these distributions of opinions into a single statement about the joint impact of the body of evidence $\{e_1, e_2\}$ on the relative likelihoods of the three hypotheses?

Technically speaking, one can think of a variety of such multi-player belief calculi. The calculi will differ in terms of expert matching rules and belief combination operators, but their overall goal will be the same: reducing a multitude of possibly conflicting and redundant opinions into some plausible summary. However, we believe that the only calculi which are worth researching are those that have (i) a solid normative justification, and (ii) a practical face validity. Figure 1 illustrates the operation of one such calculus, which we call the *Boolean set-product* model. Later in the paper we will use this simple model as a point of departure towards developing more sophisticated calculi that can be customized for different applications.

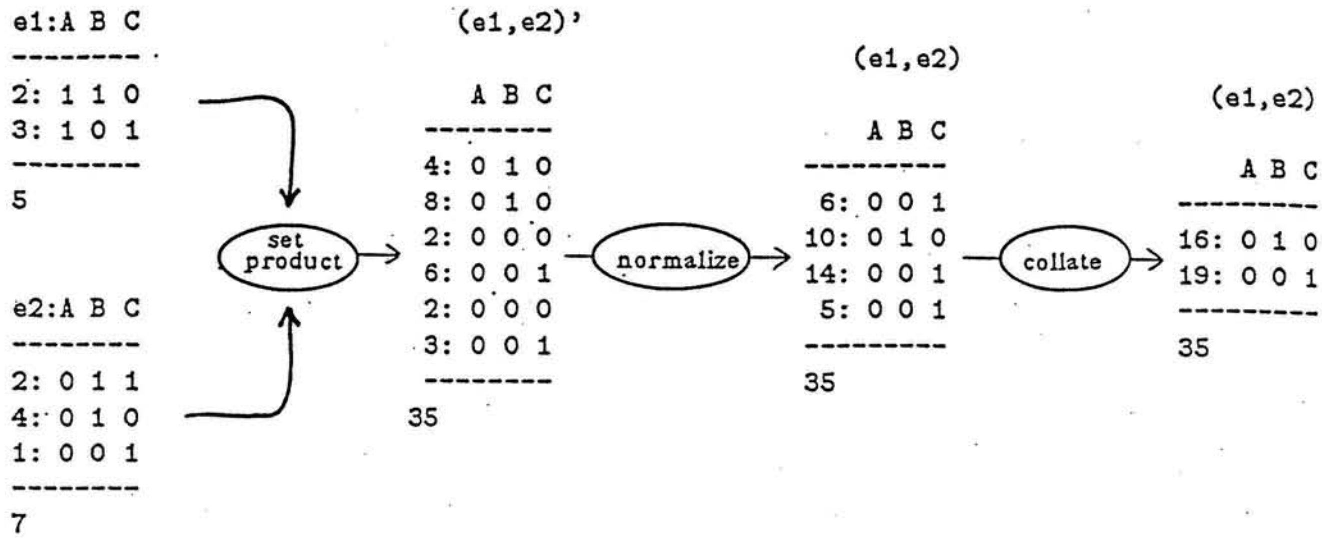

Figure 1

The Boolean set-product calculus implements what may be termed a cartesian consensus operation. First, one constructs all the committees of two that can be formed by matching each expert from the first group with each expert from the second group. In the above example, there are $5 \times 7 = 35$ such pairs of experts. The opinion of each committee is then taken to be the conjunction of the Boolean opinions of its members. For example, all the committees whose members had the individual opinions (1,1,0) and (0,1,1) yield the joint opinion (0,1,0) (first tuple in the third table). In the special case of figure 1, four committees ended up producing this particular opinion. Note that the Boolean conjunction operator rules out the hypotheses that were ruled out by *both* experts. This combination scheme would not work if the individual experts were instructed to select, rather than eliminate, hypotheses. This subtle point will be taken up later in the paper.

As it turns out, the Boolean conjunction of valid opinions may well generate an invalid opinion, which happens whenever the operator produces vectors of the form (0,0,0). This is an anomaly, since null opinions are disallowed at the axiomatic level. One way to resolve the problem is to disregard all the committees that produced null opinions and distribute their 'votes' evenly among all the committees that produced non-null opinions. In figure 1, this operation yields the table $(e1, e2)$. Finally, since the only thing that distinguishes one committee from another (in this particular model) is the joint opinion, the committees that have identical opinions are collated, leading to the right most table in figure 1.

The Boolean set-product calculus is of special interest to us because of its unique relationship to the Dempster-Shafer model. First, we recall that Dempster's rule computes a pooled mass function $m = m_1 \oplus m_2 : 2^\theta \rightarrow [0,1]$ as follows:

$$m'(X) = \sum_{A_i \cap B_j = X} m_1(A_i) \cdot m_2(B_j), \tag{1}$$

$$m(X) = \begin{cases} \frac{1}{1-m'(\emptyset)} \cdot m'(X) & X \neq \emptyset \\ 0 & X = \emptyset \end{cases} \tag{2}$$

As was shown by Hummel & Landy (1988), the Boolean set-product calculus and the standard Dempster-Shafer model are isomorphic. The remainder of this section demonstrates this important relationship in the context of figure 1. First, the frame of discernment is the hypotheses set $\theta = \{A, B, C\}$, exactly one of which is assumed to be true. Each expert rules out a subset of hypotheses, thereby retaining a subset $X \subseteq \theta$ of likely hypotheses which is said to be possible. For example, the opinion $(1, 0, 1)$ implies that $\{A, C\}$ is possible in the view of that expert. Hence, if we let X range over 2^{θ}, then the mass $m(X)$, belief $\text{Bel}(X)$, and plausibility $\text{Pl}(X)$ functions can be defined as (i) the fraction of experts who implied that the subset X is possible; (ii) the fraction of experts who ruled out all the hypotheses outside X; and (iii) the fraction of experts whose possible subsets intersect X. It is easy tho show that these fractional mappings preserve all the mathematical properties of $m(\cdot)$, $\text{Bel}(\cdot)$, and $\text{Pl}(\cdot)$ by construction.

To illustrate, consider the specific scenario depicted in figure 1. The two left-most tables induce two mass functions $m_1, m_2 : 2^{\{A,B,C\}} \rightarrow [0, 1]$ as follows: $m_1(\{A, B\}) = 2/5$, $m_1(\{A, C\}) = 3/5$, and $m_1(X) = 0$ for all other subsets of θ, and $m_2(\{B, C\}) = 2/7$, $m_2(\{B\}) = 4/7, m_2(\{C\}) = 1/7$, and $m_2(X) = 0$ for all other subsets of θ. Similarly, the right-most table – the outcome of the Boolean set product operation on the above two tables – induces the mass $m(\{B\}) = 16/35$, $m(\{C\}) = 19/35$, and $m(X) = 0$ for all other subsets of θ. The relationship of the Boolean set-product calculus to Dempster's rule is illustrated by the fact that had we applied the latter to the two functions m_1 and m_2 using eqn. (1) and (2), we would have obtained precisely the same function as m. That is, if we denote the mass function that a set of opinions T induces by m_T, the set-product/binary-conjunction operation as $\otimes$, and Dempster's rule as $\oplus$, we have the following isomorphism: $m_{T_1 \otimes T_2} = m_{T_1} \oplus m_{T_2}$.

The multi-player interpretation and the isomorphism imply that the Dempster-Shafer model is essentially a mechanism for representing and combining statistics of collections of opinions rather than tracking the collections themselves. However, different applications may require different ways to summarize and process collections of expert opinions. This view means that the Dempster-Shafer model can be seen as one instance in a parametric family of multi-player belief calculi.

3 Medical Diagnosis

Two common practices in medical diagnosis are: (i) referrals, and (ii) second opinions. A referral occurs when a general physician directs a patient to a specialist; a second opinion occurs when a patient consults independently with two doctors of the same specialty. This section illustrates how multi-player belief calculi can be used to model a multi-referral/multi-opinions diagnostic process.

As is normally done in diagnostic and fault-detection applications, we take the hypotheses set to be an exhaustive and mutually-exclusive set of potential diseases. If the initial set is not exhaustive, it can be made so by adding the hypothesis "none-of-the-above." If it is originally not mutually-exclusive, it can be augmented with all the plausible subsets of diseases that may co-occur in the same patient. The experts correspond to doctors who are grouped in different specialties, e.g. general physicians, neurologists, cardiologists, etc., all examining the same patient. Each doctor issues an independent prognosis that is essentially a list of possible diseases. The different bodies of evidence (e.g. e_1 and e_2 in figure 1) model the different training, diagnostic procedures, and information sets that characterize different specialists. For example, a hand surgeon who examines a patient with a severe pain in the wrist joint will ask different questions and administer different tests compared to a neurologist who examines the very same patient.

Since medicine is an inexact science, it is entirely possible that the prognoses of the same patient will be inconsistent. Hence, if all the doctors were considered equally qualified, it would be reasonable to try to (i) gauge the degrees of agreement/disagreement among the various specialists; and (ii) fuse the individual opinions into a joint prognosis that takes all the doctors into consideration. This pooling mechanism could be implemented in the context of a consultation system that makes use of previous prognoses of virtual patients that share common symptomatic profiles. There are at least two examples that could greatly benefit from such a consultation system: health management organizations (HMO)'s, and emergency rooms.

In a typical HMO, general physicians are routinely asked to examine numerous patients under time pressure, refer them to specialists, and then manage the overall information gathering and consultation process. Recent reports about the health care crisis in the United States indicate that the referral process is extremely wasteful, to the extent that insurance companies have begun to rebut the medical judgement of the referring doctors (New York Times, 1/20/93). In particular, it has been observed that general physicians often make unnecessary referrals in order to protect themselves from malpractice suits and compensate for lack of diagnostic experience. Under such circumstances, an inexperienced doctor who examines a new patient could clearly benefit from a consultation system that advises him/her on how different specialists have judged similar cases in the past. This information can be gathered from historical medical records, and then processed and summarized by a multi-player belief calculus.

However, there are several reasons why the simple Boolean set-product calculus described in the previous section is inappropriate for such an application. First, the prognoses that doctors make in practice are often non-categorical and non-committal, and thus the requirement that all opinions be Boolean is unrealistic. Second, the calculus yields a measure of average opinion which takes into account neither the number of doctors in each specialty group, nor the variance of their individual opinions. Needless to say, both parameters are critically important in real consultations with multiple doctors. Third, the opinions of all the doctors are weighted equally, and there is no provision for amplifying or discounting certain individual opinions, as is normally done in practice. Fourth, the bodies of evidence that characterize the different specialist groups have common features, and thus they cannot be considered independent.

Focusing on the latter point, note that in the absence of other assumptions, the probability distribution over a collection of hypotheses, conditioned on two different sources of evidence, can have an arbitrarily complex functional relationship to the probabilities that are conditioned on each source of evidence individually. Similarly, the level of uncertainty for each hypothesis, given two sources of evidence, may be arbitrarily related to the conditional probabilities and uncertainties for those hypotheses individually. To make order of the chaos, one must either have a functional model for the relationships, or one must make certain simplifying assumptions. In general, selection of the appropriate assumptions should be driven by the applications. Thus, different versions of multi-player calculi that address these and other application-specific concerns can be developed, yielding more realistic approaches to uncertainty reasoning.

4 Bibliographical indexing

Bibliographical indexing models concern the construction of data structures that enable rapid content-based access to collections of documents. Given a document, on the one hand, and a *keyword lexicon*, on the other, the goal of the indexing model is to select a subset of keywords that 'best' describes the document to its prospective users. Since some keywords are more relevant to the document than others, a numeric scale is often used to express the strength of association between the document and the selected keywords. The result is an *index vector*, consisting of pairs of keywords and their respective relevance weights.

Like medical diagnosis, bibliographical indexing is an inexact science. The relevance of a document to a keyword is a subjective relation which may be based on an aggregation of several indexing opinions. Specifically, each document has many *classifiers*, or discerning characteristics, that determine its relevance. For example, the *title* of a document can suggest one index, whereas the *abstract* can suggest another. Other aspects of the document, obtained through lexical, linguistic, and citations, analyses will yield additional indexing opinions that must be taken into consideration. Hence, even if the individual relevance opinions were forced to be binary, their aggregation would probably induce a continuous index. In addition, the indexing opinions may or may not be automatic. In the latter case, they would be elicited from human catalogers who inject yet another level of uncertainty to the indexing process. That is, when two catalogers are given access to the same classifier as background information, they may well supply two different indexing opinions. In fact, empirical studies have indicated that this kind of indexing uncertainty prevails even among well-trained catalogers (Jacoby & Slamecka [7], Stevens [13]).

In a recent paper, we explored how a multi-player belief calculus can be used to model and manage an indexing process that involves multiple relevance opinions (Schocken & Hummel, [11]). In this application, the frame of discernment ($\{A, B, C\}$ in figure 1) is taken to be a set of descriptive keywords, or a lexicon. The experts correspond to human catalogers who are asked to express relevance opinions regarding a given document. In the simplest model, each cataloger supplies a list of relevant keywords. Next, a multi-player relevance calculus is used to combine the individual relevance opinions into a joint set of committee opinions. The mass function that the (normalized) committee opinions induce is then taken to be the joint relevance vector associated with the document in question.

There are several reasons why the catalogers can be placed in different categories (corresponding to tables e_1 and e_2 in figure 1), and why the joint index may be more informative than any one of the individual indexes. From a cognitive standpoint, instead of asking each cataloger to read and index the entire document, it may better to expose different catalogers to different classifying criteria — e.g. the document's title, abstract, introduction and conclusion sections, etc. – and then combine their individual indexing opinions using a formal model. This approach can reduce information overload, debias classification errors, and generate new relevance opinions. Of course, the question of whether a pooled approach will yield better indexing decisions than individuals should be tested empirically.

As data communications and networking technologies continues to develop, retrieval systems like `WAIS` will connect masses of people to on-line libraries, opening new possibilities for interactive indexing. In an interactive indexing scenario, relevance opinions are dynamically elicited from library searchers and then used to refine (or even create) index vectors. The relevance opinions of the searchers can be obtained explicitly, or through keyword extraction algorithms that match their stated queries and the documents that they eventually find useful. In such a setting, the different groups of catalogers can correspond to the different domains, or queries, in the context of which the relevance opinions were made. Such an indexing scheme could be particularly effective in multi-disciplinary areas of interest. For example, it is quite possible that computer scientists, physicists, and neurologists will use different terms to describe and access the same neural networks paper. We believe that multi-player belief calculi, along with lexical analysis methods, could be used to join these different indexing opinions into a composite index that takes into consideration the information needs of a diverse group of library patrons.

5 Discussion

The example applications that were presented above lead us to two key observations: (1) individual players should be allowed to express probabilistic opinions, and the probabilities can be regarded as estimates or subjective quantities; (2) players can be grouped into classes and hierarchies based on the information that they bring to bear in arriving at their assessments. The variation in the resulting distribution of opinions can be attributed to the evidence that is presented, and to the specializations of the agents giving the opinions.

In the Hummel & Manevitz work [6], uncertainty calculi satisfying desiderata (1) are developed. In these calculi, multiple probabilistic opinions are expressed (actually, logarithmic opinions), and the degree of uncertainty is encoded in the spread of the opinions. These calculi also permit aggregation of the opinions into classes, satisfying portions of condition (2). The classes are then combined using a Bayesian combination of all tuples of opinions, potentially using non-independent conditioning and varying weights on the individual opinions. The degree of certainty in the resulting belief representation is naturally encoded in the resulting distribution.

On the other hand, the calculi given by Hummel & Manevitz only deal with 'intrinsic' uncertainty, so that uncertainty increases as constituent groups of opinions spread, but not necessarily as they diverge. In our applications, we can see how uncertainty should take into account divergence of opinions that might result from different classes of players, but that the formulas should reflect the fact that only certain divergences of opinions lead to uncertainty, whereas other divergences are explainable and might even lead to greater certainty. Further, we see that the dependencies among different classes should be modeled accurately, and that methods for uncertainty management need to be adaptive.

The purpose of this paper was to survey the theory of multi-player belief calculi and to explore its potential use in certain canonical applications. Our next step in this research program is to fully specify several calculi, implement them in reasoning systems, and then test their effectiveness in controlled experiments. While we have yet to build these systems, we believe that the multi-player paradigm, and the grouping of the players in ways suggested by the applications, will provide important and useful constraints to the development of practical methods for reasoning under uncertainty.

References

[1] B.G. Buchanan and E.H. Shortliffe. Uncertainty and evidential support. In B.G. Buchanan and E.H. Shortliffe, editors, *Rule-Based Expert Systems*, pages 217–219, Addison-Wesley, 1984.

[2] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals Mathematics Statistics*, 38:325–339, 1967.

[3] R.O. Duda, P.E. Hart, and N.J. Nilsson. *Development of a Computer-Based Consultant for Mineral Exploration.* Technical Report, SRI, 1977. International Projects 5821 and 6415.

[4] I. J. Good. The measure of a non-measurable set. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology, and Philosophy of Science*, pages 319–329, Stanford University Press, 1962.

[5] R.A. Hummel and M.S. Landy. A statistical viewpoint on the theory of evidence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(2):235–247, 1988.

[6] R.A. Hummel and L. Manevitz. *Foundations of Reasoning With Uncertainty.* Technical Report, Courant Institute of Mathematical Sciences, New York University, 1992.

[7] J. Jacoby and V. Slamecka. Indexing consistency under minimal conditions. Bethesda, MD: Documentation, Inc., 1962.

[8] L.N. Kanal and J.F. Lemmer. Uncertainty in artificial intelligence. North Holland, 1986. Subsequent volumes in 1987,1988,1989,1990,1991, and 1992.

[9] J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan-Kauffman, 1988.

[11] S. Schocken and R.A. Hummel. *On The Use of the Dempster-Shafer Model in Information Indexing and Retrieval Applications.* Technical Report, Information Systems Department, Stern School of Business, New York University, 1992. STERN IS-92-27.

[12] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[13] M.E. Stevens. Automatic indexing: a state of the art report. Washington, DC: U.S. Government Printing Office, 1965.

[14] C. H. Tzeng. A mathematical formulation of uncertain information. *Annals of Mathematics and Artificial Intelligence*, 4:69–88, 1991.

[15] J. Yen. Gertis: a Dempster-Shafer approach to diagnosing hierarchical hypotheses. *Communications of the ACM*, 32(5):573–585, 1989.

[16] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.