A Statistical Viewpoint on the Theory of Evidence

ROBERT A. HUMMEL, MEMBER, IEEE, AND MICHAEL S. LANDY

Abstract—**We describe a viewpoint on the Dempster/Shafer "theory of evidence," and provide an interpretation which regards the combination formulas as statistics of the opinions of "experts." This is done by introducing spaces with binary operations that are simpler to interpret or simpler to implement than the standard combination formula, and showing that these spaces can be mapped homomorphically onto the Dempster/Shafer theory of evidence space. The experts in the space of "opinions of experts" combine information in a Bayesian fashion. We present alternative spaces for the combination of evidence suggested by this viewpoint.**

Index Terms—**Bayes rule, Bayesian combination, evidence, evidential reasoning, expert systems, theory of evidence.**

I. Introduction

MANY problems in artificial intelligence call for assessments of degrees of belief in propositions based on evidence gathered from disparate sources. It is often claimed that probabilistic analysis of propositions is at variance with intuitive notions of belief [7], [17], [19]. Various methods have been introduced to reconcile the discrepancies, but no single technique has settled the issue on both theoretical and pragmatic grounds.

A. Theory of Evidence

One method for attempting to modify the probabilistic analysis of propositions is the Dempster/Shafer "theory of evidence." This theory is derived from notions of upper and lower probabilities, as developed by Dempster in [5]. The idea that intervals instead of probability values can be used to model degrees of belief had been suggested and investigated by earlier researchers [9], [13], [17], [31], but Dempster's work defines the upper and lower points of the intervals in terms of statistics on set-valued functions defined over a measure space. The result is a collection of intervals defined for subsets of a fixed labeling set, and a combination formula for combining collections of intervals.

Alternative theories based on notions of upper and lower probabilities were also pursued [13], [33], and can be formally related to the updating formulas used in the Dempster/Shafer theory [19], but are really a separate formulation.

Manuscript received December 13, 1985; revised June 12, 1987. This work was supported by the Office of Naval Research under Grant N00014-85-K-0077 and by the National Science Foundation under Grant DCR-8403300.

R. A. Hummel is with the Courant Institute of Mathematical Sciences, New York University, New York, NY 10012.

M. S. Landy is with the Department of Psychology, New York University, New York, NY 10003.

IEEE Log Number 8718601.

Dempster explained in greater detail how the statistical notion from his earlier work could be used to assess beliefs on propositions in [6]. In [4], Dempster gave examples of the use of upper and lower probabilities in terms of finite populations with discrete univariant observable characteristics, in correspondence with algebraic structure to be discussed later in this paper. The topic was taken up by Shafer [26], [27], and led to publication of a monograph on the "theory of evidence" [28]. All of these works after [4] emphasize the values assigned to subsets of propositions (the "beliefs") and the combination formulas, and deemphasize the connection to the statistical foundations based on the set-valued functions on a measure space.

The Dempster/Shafer theory of evidence has sparked considerable debate among statisticians and "knowledge engineers." The theory has been criticized and debated in terms of its behavior and applicability, e.g., [6], [21], [24], [33] (commentaries following). Some of the questions have been answered by Shafer [29], [30], but discussion of the theoretical underpinnings continues, e.g., [7], [18], [19]. A related, but distinct theory of lower probabilities is frequently discussed as another alternative for uncertain reasoning [13], [33]. An excellent study by Kyburg [19] relates the Dempster/Shafer theory to a lower probability framework where beliefs are viewed as extrema of opinions of experts. These viewpoints have similarities to the one developed here, but differ in the interpretation of belief values.

Recently, there has been increased interest in the use of the Dempster/Shafer theory of evidence in expert systems [11], [14]. Most of the recent attempts to map the theory to real applications and practical methods, such as described in [2], [8], [10], [15], [32], are based on the "constructive probability" techniques described by Shafer [29], and disregard the statistical theoretical foundations from which the theory was derived. The constructive theory is based on a notion of fitting particular problems to scales of canonical examples. In the case of belief functions, the cornerstone of the Dempster/Shafer theory, Shafer offers a set of examples of "coded messages" being sent by a random process, and a set of measures on belief functions to assist in fitting parameters of the "coded message" example to instances of subjective notions of belief. While the "coded message" interpretation is an essentially statistical viewpoint and isomorphic to the algebraic spaces discussed here and implicit in Dempster's work, the proposed fitting scheme attempts to apply alternate interpretations to the combination formula based on subjective similarities.

In this paper, we present a viewpoint on the Dempster/Shafer theory of evidence that regards the theory as statistics of opinions of "experts." We relate the evidence-combination formulas to statistics of experts who perform Bayesian updating in pairs. In particular, we show that the Dempster rule of combination, rather than extending Bayesian formulas for combining probabilities, contains nothing more than Bayes' formula applied to Boolean assertions, but tracks multiple opinions as opposed to a single probabilistic assessment. Finally, we suggest a related formulation that leads to simpler formulas and fewer variables. In this formulation, as in the Dempster combination formula, the essential idea is that we track the statistics of the opinions of a class of opinions. However, in our new formulation, the opinions are allowed to be probabilistic, as opposed to the Boolean opinions that are implicit in the Dempster formula.

The authors' interest in the Dempster/Shafer theory of evidence derives from a study of a large class of iterative knowledge aggregation methods [20]. These methods, which include relaxation labeling [16], stochastic relaxation [12], neural models [1], and other "connectionist networks," always attempt to find a true labeling by updating a state as evidence is accumulated. In the theory of evidence, as in many other models, the true labeling is one of a finite number of possibilities, but the state is a collection of numbers describing an element in a continuous domain. In the Shafer formulation, the state of the system is described by a distribution over the set of all subsets of the possible labels. That is, each subset A of labels has assigned to it a number representing a kind of probability that the subset of possible labels is precisely A. Implicit in this model is the notion that an incremental piece of evidence carries a certain amount of weight or confidence, and distinguishes a subset of possibilities. Evidence may point to a single inference among the set of labels or may point to a subset of the alternatives (see, e.g., [23]). As evidence is gained, belief values are updated according to a combination formula. The combination formula is commutative and associative, so a succession of incremental changes can be combined into a single state that can be regarded as a nonprimitive updating element.

Most of the other iterative models for combining evidence represent the degree of support for a label by a single number, although there may be additional numbers in a state vector corresponding to "hidden units." For the state of belief in the formulation discussed above, there are numbers for every subset of labels. Thus, if there are n labels, a state has (roughly) 2^n values. That is, there are many additional degrees of freedom. Further, not all iterative models have associative combination formulas. Commutativity is even more problematic since there is often a distinction between the current state of belief and the form of representation of incremental evidence. The Dempster/Shafer formulation is somewhat special in that evidence is represented by a second state of belief to be combined, on an equal basis, with a current state of belief.

B. Theory of Belief Functions

In this section, we amplify on the distinction between the viewpoint established in the remainder of this paper and the theory of belief functions, as used in the Dempster/Shafer theory of evidence.

The canonical examples from which belief functions are to be constructed are based on "coded messages" $c_1, \cdots, c_n$ which form the values of a random process with prior probabilities $p_1, \cdots, p_n$ [24]. Each message c_i has an associated subset A_i of labels, and carries the message that the true label is among A_i. The masses representing the current state are simply the probabilities (with respect to this random process) of receiving a message associated with a subset A. The *belief* in a subset A is the probability that a message points to a subset of A.

The coded-message formulation corresponds exactly with our space of Boolean opinions of experts (Section III-A). Moreover, the combination of coded messages and the combination of elements in the space of Boolean opinions coincide. Specifically, given a random process of messages $c_1, \cdots, c_n$ with priors $p_1, \cdots, p_n$, and another process of messages $c'_1, \cdots, c'_m$ with priors $p'_1, \cdots, p'_m$, then in combination, a pair of codes is chosen independently (c_i, c'_j), thus with prior probability $p_i p'_j$, and the associated message is that the truth lies in $A_i \cap A'_j$. There c_i carries the message A_i, and c'_j carries messages A'_j. It is our point, in introducing the spaces of experts, that the requisite independence includes not only the choice of messages, but also an assumption that the message is formed by the intersection of the subsets designated by the constituent messages. As opposed to being tautological, this intersection involves a conditional independence assumption, a point that we emphasize by treating the formulation as algebraic structures, and by considering the space of probabilistic opinions of experts.

In a sense, our space of Boolean opinions of experts can be thought of as an alternative set of canonical examples with which to construct states of belief to analogous real situations. Necessarily, these examples will be isomorphic to any other set of canonical examples, and only the language used to describe the same algebraic space varies. However, there is additional richness in the various classes of canonical examples since many distinct examples might correspond to an identical state of belief. By "backing up" to the richness of the space of probabilistic opinions of experts, we are better able to interpret the foundations of the Dempster rule of combination, and to suggest the alternative formulation that is presented in the second part of this paper (see Fig. 2).

When the theory of belief functions is actually applied to evidential reasoning situations with uncertain evidence, the belief function is typically regarded as a variant on a probability measure over the set of labels [25]. An important difference is that the belief function is not an additive measure. Nonetheless, the belief on a particular label is identified, in some subjective way, with a probability for that label, except that degrees of uncertainty are allowed to withhold "mass" to nonsingleton subsets. In the commentaries to Shafer's presentation of the theory of belief functions and example applications before the Royal Statistical Society [24], several discussants commented on the need for a closer connection between the canonical examples and the interpretation of belief values. Prof. Barnard, for example, states that "the connections between the logical structure of the . . . example and the story of the uncertain codes is not at all clear." Prof. Williams desires "a deeper justification of the method and a further treatment of 'unrelated bodies of evidence,'" while Prof. Krantz states simply that "comparison of evidence to a probabilistically coded message seems strained." Prof. Fine summarizes the problem by stating that "the coded message interpretation is ignored when actually constructing belief functions, calling into question the relevance of the canonical scales."

We believe that the viewpoint expounded here, and the analytic treatment of algebraic spaces embodying the combination formula for belief functions substantially answers these calls for elucidation of the meaning of belief functions. At the very minimum, our spaces provide canonical examples where belief values can be regarded as percentages of sets of experts stating that possible labels are restricted to within a specified subset. We believe, however, that the viewpoint reduces the need for subjective balancing between a given probabilistic situation and a "coded message" interpretation, and instead provides a way in which belief values can be estimated by, for example, sampling techniques. The crucial point (and presumably essential to the notion of uncertainty) is that uncertainty is measured over a different sample space from the labeling situation; in our parlance, the separate sample space is a set of experts. Further, the viewpoint that evidence can be represented by collections of opinions or the statistics on a collection of opinions leads, fairly naturally, to alternate representations from the space of belief states used in the Dempster/Shafer formulation. Given the fundamental simplicity of the parameterized statistics space that we introduce in Section V, we believe that the viewpoint yields structures for evidential reasoning that might well be applicable when neither Bayesian probabilistic reasoning nor theories of belief functions are suitable.

Belief functions are generally viewed as extensions of probability measures over the set of labels. When all masses occur on singleton subsets, then the belief function is an additive measure, and a combination of such elements yields a formula equivalent to Bayes' formula with conditional independence. Since more general belief functions are allowed, the Dempster combination formula is regarded, from this viewpoint, as an extension of Bayes' formula.

From the point of view of statistics of opinions of experts, as developed here, the Dempster combination formula is explained by Bayesian updating on Boolean opinions in all cases. The special-case Bayes' formula is explained as follows. When masses are concentrated on singletons, then each expert is naming a single label. Suppose that the percentage of experts naming a particular label is the same as the actual probability for that label given the information available to the experts. This is an ergodicity assumption since chances are being compared over two distinct sample spaces—the set of experts and the space of labeling situations. Then the independent sampling of a pair of experts from each of two such collections of experts mimics the independent probabilistic assessment of conditioning on multiple hypotheses.

To what extent can the various viewpoints coexist? As alternative scales of canonical examples, there is no conflict between opinions of experts and coded messages. However, the viewpoint that regards the masses and beliefs as probabilities of Boolean random variables defined on a sample space of experts, distinct from the sample space of labeling situations, seems to give additional intuitive insight, as stated by Prof. Kingman in the same commentaries to [24]. Further, as we emphasize here, this viewpoint is isomorphic to the structures for combining evidence, modulo the terminology. But in order to reconcile a view of beliefs as probabilities over sets of experts with a view of beliefs as extensions of probability measures over labels, some kind of ergodicity assumption is needed to relate distributions over the different spaces. It may well be that such assumptions can be formulated to give a deeper theoretical basis for the application of canonical examples to probabilistic situations with uncertainties. An advantage would be that judgments of the applicability of the formulation could be based on the validity of the assumption as opposed to the quality of empirical results. However, we do not pursue such a plan here, preferring to view uncertainty as a measure of concurrence of multiple opinions.

C. Objectives

This paper has three main points. First, we formulate the space of belief states as an algebraic structure, pointing out in the process that the normalization term in the Dempster rule of combination is essentially irrelevant. Our reason for treating these much-debated and motivated concepts in terms of mathematical structures such as semigroups and monoids is to follow Dempster's early admonition to avoid becoming "sidetracked into doctrinaire questions concerning whether probabilities are frequencies, or personal degrees of belief, or betting probabilities, etc." [4]. Having formulated the Dempster/Shafer theory of evidence as a simple algebraic structure, we can discuss interpretations in terms of their isomorphic relationship to the theory.

We then describe spaces that we call probabilistic and Boolean opinions of experts. Our intent is to survey the foundations of the Dempster/Shafer theory in a manner more accessible than the original Dempster works, and in a way that makes clear the relationship to Bayesian analysis. The key point here is that rather than extending Bayes' formula, the combination method is simply applying Bayes' formula to sets of Boolean opinions, updating on product sets of those opinions. The idea of a class of opinions, rather than a single probabilistic current opinion, occurs in the theory of lower probabilities [13], and is the theme of a unifying treatment of evidential reasoning in [22]. In the theory of evidence, the opinions are Boolean valued, giving lists of possible labels, and the state of the system is described by the statistics of these opinions. In, for example, a medical diagnosis application, the range of opinions might be held by different doctors, and the opinions themselves consist of a list of possible pathologies. The important distinction between measuring statistics over the set of doctors and over the set of patients forms the basis for measuring degrees of uncertainty.

Finally, we use the viewpoint established by these spaces or canonical examples to introduce the main original contribution of this paper. We use the space of probabilistic opinions of experts to define spaces that we call parameterized statistics of opinions. The idea and use of these spaces to tasks of evidence are fundamentally simple: a probabilistic opinion is maintained and updated, as in Bayesian analysis with conditional independence, and a concurrent measure of uncertainty is maintained in terms of a multivariate Gaussian distribution in log-probability space. Once again, we have the idea of a spread of opinions, but founded on notions of Bayes' theorem for updating, and with the connections to the Dempster/Shafer theory made clear.

II. The Rule of Combination and Normalization

The set of possible outcomes or labelings will be denoted in this paper by Λ. This set is the "frame of discernment," and in other works has been denoted, variously, by Ω, Θ, or $\mathcal{S}$. For convenience, we will assume that Λ is a finite set with n elements, although the framework could easily be extended to continuous label sets. More importantly, we will assume that Λ represents a set of states that are mutually exclusive and exhaustive. If Λ is not initially exhaustive, it can easily be made so by including an additional label denoting "none of the above." If Λ is not mutually exclusive, it can be made so by replacement with its power set (i.e., the set of all subsets), so that each subset represents the occurrence of exactly that subset of labels, excluding all other labels. Of course, replacing Λ by its power set is perilous in that it will greatly expand the cardinality of the label set. For practical applications, the implementer is more likely to want to replace Λ by the set of all plausible subsets describing a valid configuration.

An element (or state of belief) in the theory of evidence is represented by a probability distribution over the power set of Λ, $\boldsymbol{P}(\Lambda)$. That is, a state m is

$$m: \boldsymbol{P}(\Lambda) \rightarrow [0, 1]$$
$$\sum_{A \subseteq \Lambda} m(A) = 1.$$

There is an additional proviso that is typically applied, namely, that every state m satisfies

$$m(\varnothing) = 0.$$

Section III-B introduces a plausible interpretation for the quantities comprising a state.

A state is updated by combination with new evidence or information which is presented in the form of another state. Thus, given a current state m_1 and another state m_2, a combination of the two states is defined to yield a state $m_1 \oplus m_2$ given by

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B)\, m_2(C)}{1 - \sum_{B \cap C = \varnothing} m_1(B)\, m_2(C)} \qquad \text{if } A \neq \varnothing$$

and

$$(m_1 \oplus m_2)(\varnothing) = 0. \tag{1a}$$

This is the so-called "Dempster Rule of Combination." Note that the resulting function m is a probability mass due to the normalization factor, and that $(m_1 \oplus m_2)(\varnothing) = 0$ by definition.

The problem with this definition is that the denominator in (1a) might be zero, so that $(m_1 \oplus m_2)(A)$ is undefined. That is, there exist pairs m_1 and m_2 such that the combination of m_1 and m_2 is not defined. This, of course, is not a very satisfactory situation for a binary operation on a space. The solution which is frequently taken is to avoid combining such elements. An alternative is to add an additional element m_0 to the space:

$$m_0(A) = 0 \qquad \text{for } A \neq \varnothing$$
$$m_0(\varnothing) = 1.$$

Note that this additional element does not satisfy the condition $m(\varnothing) = 0$. Then define, as a special case,

$$m_1 \oplus m_2 = m_0 \qquad \text{if } \sum_{B \cap C = \varnothing} m_1(B)\, m_2(C) = 1. \tag{1b}$$

The binary operation is then defined for all pairs m_1, m_2. The special element m_0 is an absorbent state in the sense that $m_0 \oplus m = m \oplus m_0 = m_0$ for all states m.

This space has an identity element. The identity state m_I represents complete ignorance in that combination with it yields no change (i.e., $m_I \oplus m = m \oplus m_I = m$ for all states m). This state places full mass on the subset which is all of Λ:

$$m_I(\Lambda) = 1$$
$$m_I(A) = 0 \qquad \text{for } A \neq \Lambda.$$

Definition 1: We define $(\mathfrak{M}, \oplus)$, the *space of belief states*, by

$$\mathfrak{M} = \left\{ m: \boldsymbol{P}(\Lambda) \rightarrow \mathbf{R}^+ \cup \{0\} \,\middle|\, \sum_{A \subseteq \Lambda} m(A) = 1,\ m(\varnothing) = 0 \right\} \cup \{m_0\},$$

and define $\oplus$ by (1a) when the denominator in (1a) is nonzero, and by (1b) otherwise. ■

The set $\mathfrak{M}$, together with the combination operation $\oplus$, constitutes a *monoid* since the binary operation is closed and associative, and there is an identity element.[1] In fact, the binary operation is commutative, so we can say that the space is an Abelian monoid.

Still, because of the normalization and the special case in the definition of $\oplus$, the monoid $\mathfrak{M}$ is both ugly and cumbersome. It makes better sense to dispense with the normalization. We have the following.

Definition 2: We define $(\mathfrak{M}', \oplus')$, the *space of unnormalized belief states*, by

$$\mathfrak{M}' = \left\{ m: P(\Lambda) \to R^+ \cup \{0\} \,\middle|\, \sum_{A \subseteq \Lambda} m(A) = 1 \right\}$$

without the additional proviso, and set

$$(m_1 \oplus' m_2)(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \;\; \forall A \subseteq \Lambda \tag{2}$$

for all pairs $m_1, m_2 \in \mathfrak{M}'$. ■

One can verify that $m_1 \oplus' m_2 \in \mathfrak{M}'$, and that $\oplus'$ is associative and commutative. Further, the same element m_I defined above is also in $\mathfrak{M}'$, and it is an identity. Thus, $\mathfrak{M}'$ is also an Abelian monoid. Clearly, $\mathfrak{M}'$ is a more attractive monoid than $\mathfrak{M}$.

We define a transformation V mapping $\mathfrak{M}'$ to $\mathfrak{M}$ by the formulas

$$(Vm)(A) = \frac{m(A)}{1 - m(\varnothing)}, \; A \neq \varnothing$$

$$(Vm)(\varnothing) = 0 \tag{3}$$

if $m(\varnothing) \neq 1$ and

$$Vm = m_0$$

otherwise.

A computation shows that V preserves the binary operation, i.e.,

$$V(m_1 \oplus' m_2) = V(m_1) \oplus V(m_2).$$

Thus, V is a *homomorphism*.[2] Further, V is *onto*, since for $m \in \mathfrak{M}$, the same m is in $\mathfrak{M}'$, and $Vm = m$. The algebraic terminology is that V is an *epimorphism* of monoids, a fact that we record in the following.

Lemma 1: V maps homomorphically from $(\mathfrak{M}', \oplus')$ onto $(\mathfrak{M}, \oplus)$. ■

A "representation" is a term that refers to a map that is an epimorphism of structures. Intuitively, such a map is important because it allows us to consider combination in the space formed by the range of the map as combinations of preimage elements. Lemma 1 will eventually form a small part of a representation to be defined in the next section. In the case in point, however, if it is required to combine elements in $\mathfrak{M}$, one can perform the combinations in $\mathfrak{M}'$, and project to $\mathfrak{M}$ by V after all of the combinations are completed. Since combinations in $\mathfrak{M}'$ are much cleaner, this is a potentially useful observation. In terms of the Dempster/Shafer theory of evidence, this result says that the normalization in the combination formula is essentially irrelevant, and that combining can be handled by (2). Specifically, given a sequence of states in $\mathfrak{M}$ to be combined, say $m_1, m_2, \cdots, m_k$, we can regard these states as elements in $\mathfrak{M}'$. Since each m_i satisfies $m_i(\varnothing) = 0$, then each satisfy $Vm_i = m_i$. Thus, $V(m_1 \oplus' m_2 \oplus' \cdots \oplus' m_k) = Vm_1 \oplus \cdots \oplus Vm_k = m_1 \oplus \cdots \oplus m_k$, which says that it suffices to compute the combinations using $\oplus'$ (2), and then project by V (3). Of course, the final projection is necessary only if we absolutely insist on a result in $\mathfrak{M}$. If any more combining is to be done or if we are reasonably broad minded, intermediate results can be interpreted directly as elements in $\mathfrak{M}'$.

III. Spaces of Opinions of Experts

In this section, we introduce two new spaces, based on the opinions of sample spaces of experts, and discuss the evaluation of statistics of experts' opinions. Finally, we interpret the combination rules in these spaces as being a form of Bayesian updating. In the following section, we will show that these spaces also map homomorphically onto the space of belief states.

A. Opinions of Experts

We consider a set $\mathcal{E}$ of "experts," together with a map μ giving a weight or strength for each expert. It is convenient to think of $\mathcal{E}$ as a large but finite set, although the essential restriction is that $\mathcal{E}$ should be a measure space. Each expert $\omega \in \mathcal{E}$ maintains a list of possible labels: Dempster uses the notation $\Gamma(\omega)$ for this subset, i.e., $\Gamma(\omega) \subseteq \Lambda$. Here we will assume that each expert ω has more than just a subset of possibilities $\Gamma(\omega)$, but also a *probabilistic opinion* p_ω defined on Λ satisfying

$$p_\omega(\lambda) \geq 0, \qquad \forall \lambda \in \Lambda$$

$$p_\omega(\lambda) > 0 \qquad \text{iff } \lambda \in \Gamma(\omega)$$

and

$$\left(\sum_{\lambda \in \Lambda} p_\omega(\lambda) = 1 \;\; \text{or} \;\; p_\omega(\lambda) = 0 \, \forall \lambda \right), \qquad \forall \omega \in \mathcal{E}.$$

As suggested by the notation, $p_\omega(\lambda)$ represents expert ω's assessment of the probability of occurrence of the label λ. If an expert ω believes that a label λ is possible, i.e., $\lambda \in \Gamma(\omega)$, then the associated probability estimate $p_\omega(\lambda)$ will be nonzero. Conversely, if ω thinks that λ is impossible ($\lambda \notin \Gamma(\omega)$), then $p_\omega(\lambda) = 0$. We also include the possibility that expert ω has no opinion, which is indicated by the special element $p_\omega \equiv 0$. This state is included in order to ensure that the binary operation, to be defined

[1]A structure with a closed associative binary operation is sometimes call a *semigroup*, so that the space in question is an abelian semigroup with an identity.

[2]Strictly speaking, this merely shows that V is a homomorphism of semigroups; it is not hard to show that V maps the identity to the identity, which it must since it is onto, and thus it is also a homomorphism of monoids.

later, is closed. We denote the collection of maps $\{ p_\omega | \omega \in \mathcal{E} \}$ by P.

It will turn out that the central point in the theory of evidence is that the $p_\omega(\lambda)$ data are used only in terms of test for zero. Specifically, we set

$$x_\omega(\lambda) = \begin{cases} 1 & \text{if } p_\omega(\lambda) > 0 \\ 0 & \text{if } p_\omega(\lambda) = 0. \end{cases} \tag{4}$$

Note that x_ω is the characteristic function of the set $\Gamma(\omega)$ over Λ, i.e., $x_\omega(\lambda) = 1$ iff $\lambda \in \Gamma(\omega)$. The collection of all x_ω's will be denoted by X, and will be called the *Boolean opinions* of the experts $\mathcal{E}$.

If we regard the space of experts $\mathcal{E}$ as a sample space, then each $x_\omega(\lambda)$ can be regarded as a sample of a random (Boolean) variable $x(\lambda)$. In a similar way, the $p_\omega(\lambda)$'s are also samples of random variables $p(\lambda)$. The state of the system will be defined by statistics on the set of random variables $\{x(\lambda)\}_{\lambda \in \Lambda}$. These statistics are measured over the space of experts. If all experts have the same opinion, then the state should describe that set of possibilities, and the fact that there is a unanimity of opinion. If there is a divergence of opinions, the state should record the fact.

To compute statistics, we view $\mathcal{E}$ as a sample space with prior weights given by μ. We extend μ to a measure on $\mathcal{E}$, completely determined by the weights of the individual experts $\mu(\{\omega\})$ for $\omega \in \mathcal{E}$. (We are assuming that $\mathcal{E}$ is finite.) That is,

$$\mu(\mathfrak{F}) = \sum_{\omega \in \mathfrak{F}} \mu(\{\omega\}).$$

If all experts have equal weights, then μ is equivalent to a counting measure, and statistics are then measured in terms of percentages of experts. For minor technical reasons (explained in Section IV), we allow weights on the experts, so that statistics on the $x(\lambda)$'s are in terms of weighted percentages.

We are now ready to introduce the spaces which we will term "opinions of experts." The central point is that the set of labels Λ is fixed, but that the set of experts $\mathcal{E}$ can be different for distinct elements in these spaces. For the first space, we also use a fixed set of positive constants κ_λ, one for each label that will eventually be set to the prior probability for the label λ.

Definition 3: Let $K = \{\kappa_\lambda\}$ be a set of positive constants indexed over the label set Λ. The *space of probabilistic opinions of experts* $(\mathfrak{N}, K, \otimes)$ is defined by

$$\mathfrak{N} = \Big\{ (\mathcal{E}, \mu, P) \,\Big|\, \#\mathcal{E} < \infty, \mu \text{ is a measure on } \mathcal{E},$$
$$P = \{p_\omega\}_{\omega \in \mathcal{E}}, p_\omega : \Lambda \to [0, 1] \forall \omega,$$
$$\text{and } \forall \omega, \sum_{\lambda \in \Lambda} p_\omega(\lambda) = 1 \text{ or } p_\omega \equiv 0 \Big\}.$$

As noted earlier, the requirement that $\#\mathcal{E} < \infty$ is for clarity of presentation; Dempster defines the space $\mathfrak{N}$ in a more general setting.

We define a binary operation on $\mathfrak{N}$ as follows. Given $(\mathcal{E}_1, \mu_1, P_1)$ and $(\mathcal{E}_2, \mu_2, P_2)$ elements in $\mathfrak{N}$, define

$$(\mathcal{E}, \mu, P) = (\mathcal{E}_1, \mu_1, P_1) \otimes (\mathcal{E}_2, \mu_2, P_2)$$

by

$$\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2 = \{(\omega_1, \omega_2) | \omega_1 \in \mathcal{E}_1, \omega_2 \in \mathcal{E}_2\},$$
$$\mu(\{(\omega_1, \omega_2)\}) = \mu_1(\{\omega_1\}) \cdot \mu_2(\{\omega_2\})$$

and

$$P = \{p_{(\omega_1, \omega_2)}\}_{(\omega_1, \omega_2) \in \mathcal{E}}$$

$$p_{(\omega_1, \omega_2)}(\lambda) = \frac{p_{\omega_1}^{(1)}(\lambda)\, p_{\omega_2}^{(2)}(\lambda) [\kappa_\lambda]^{-1}}{\sum_{\lambda'} p_{\omega_1}^{(1)}(\lambda')\, p_{\omega_2}^{(2)}(\lambda') [\kappa_{\lambda'}]^{-1}}$$

provided the denominator is nonzero, and

$$p_{(\omega_1, \omega_2)} \equiv 0$$

otherwise. Here, $P_i = \{p_{\omega_i}^{(i)}\}_{\omega_i \in \mathcal{E}_i}$ for $i = 1, 2$, and the κ_λ's are a fixed set of positive constants defined for $\lambda \in \Lambda$. ■

To interpret this combining operation, consider two sets of experts $\mathcal{E}_1$ and $\mathcal{E}_2$, with each set of experts expressing opinions in the form of P_1 and P_2. We form a new set of experts, which is simply the set of all committees of two, consisting of one expert from $\mathcal{E}_1$ and another from $\mathcal{E}_2$. In each of the committees, the members confer to determine a consensus opinion. In Section III-C, we will see how to interpret the formulas as Bayesian combination (where κ_λ is the prior probability on λ). And in the following section, we will show that this space maps homomorphically onto the belief spaces. Finally, if as in Dempster [5], we only regard the opinions of these experts in terms of a test for zero (i.e., disregarding the strength of nonzero opinions), we arrive at yet another space. A depiction of the combination of two Boolean opinions is shown in Fig. 1.

Definition 4: The *space of Boolean opinions of experts* $(\mathfrak{N}', \odot)$ is defined similarly:

$$\mathfrak{N}' = \Big\{ (\mathcal{E}, \mu, X) \,\Big|\, \#\mathcal{E} < \infty, \mu \text{ is a measure on } \mathcal{E},$$
$$X = \{X_\omega\}_{\omega \in \mathcal{E}}, x_\omega : \Lambda \to \{0, 1\} \quad \forall \omega \Big\}.$$

If $(\mathcal{E}_1, \mu_1, X_1)$ and $(\mathcal{E}_2, \mu_2, X_2)$ are elements in $\mathfrak{N}'$, define their product

$$(\mathcal{E}, \mu, X) = (\mathcal{E}_1, \mu_2, X_1) \odot (\mathcal{E}_2, \mu_2, X_2)$$

by

$$\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2 = \{(\omega_1, \omega_2) | \omega_1 \in \mathcal{E}_1, \omega_2 \in \mathcal{E}_2\}$$
$$\mu(\{(\omega_1, \omega_2)\}) = \mu_1(\{\omega_1\}) \cdot \mu_2(\{\omega_2\}),$$

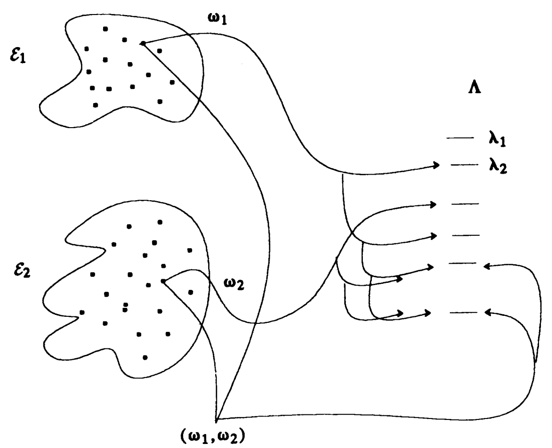


Fig. 1. A depiction of the combination of two Boolean opinions of two experts, as is present in combinations in $\mathfrak{M}'$, yielding a consensus opinion by the element in the product set of experts formed by the committee of two.

and

$$X = \{x_{(\omega_1, \omega_2)}\}_{(\omega_1, \omega_2) \in \mathcal{E}}$$

$$x_{(\omega_1, \omega_2)}(\lambda) = x^{(1)}_{\omega_1}(\lambda) \cdot x^{(2)}_{\omega_2}(\lambda),$$

where $X_i = \{x^{(i)}_{\omega_i} \mid \omega_i \in \mathcal{E}_i\}$ for $i = 1, 2$. ∎

B. Statistics of Experts

For a given subset $A \subseteq \Lambda$, the characteristic function χ_A is defined by

$$\chi_A(\lambda) = \begin{cases} 0 & \text{if } \lambda \notin A \\ 1 & \text{if } \lambda \in A. \end{cases}$$

Equality of two functions defined on Λ means, of course, that the two functions agree for all $\lambda \in \Lambda$. That is, $x_\omega = \chi_A$ means

$$x_\omega(\lambda) = \chi_A(\lambda) \qquad \forall \lambda \in \Lambda,$$

which is the same thing as saying $\Gamma(\omega) = A$.

Given a space of experts $\mathcal{E}$ and the Boolean opinions X, we define

$$\tilde{m}(A) = \frac{\mu\{\omega \in \mathcal{E} \mid x_\omega = \chi_A\}}{\mu\{\mathcal{E}\}} \tag{5}$$

for every subset $A \subseteq \Lambda$. It is possible to view the values as probabilities on the random variables $\{x(\lambda)\}$. We endow the elements of $\mathcal{E}$ with the prior probabilities $\mu(\{\omega\})/\mu(\mathcal{E})$, and say that the probability of an event involving a combination of the random variables $x(\lambda)$'s over the sample space $\mathcal{E}$ is the probability that the event is true for a particular sample where the sample is chosen at random from $\mathcal{E}$ with the sampling distribution given by the prior probabilities. This is equivalent to saying

$$\operatorname*{Prob}_{\mathcal{E}}(\text{event}) = \frac{\mu(\{\omega \in \mathcal{E} \mid \text{event is true for } \omega\})}{\mu\{\mathcal{E}\}}.$$

With this convention, we see that

$$\tilde{m}(A) = \operatorname*{Prob}_{\mathcal{E}}(x_\omega = \chi_A).$$

In fact, all of the priors and joint statistics of the $x(\lambda)$'s are determined by the full collection of $\tilde{m}(A)$ values. For example,

$$\operatorname{Prob}(x(\lambda_0) = 1) = \sum_{\{A \mid \lambda_0 \in A\}} \tilde{m}(A)$$

and

$$\operatorname{Prob}(x(\lambda_0) = 1 \text{ and } x(\lambda_1) = 1) = \sum_{\{A \mid \lambda_0, \lambda_1 \in A\}} \tilde{m}(A).$$

Further, the full set of values $\tilde{m}(A)$ for $A \subseteq \Lambda$ defines an element $\tilde{m} \in \mathfrak{M}'$. To see this, it suffices to check that $\Sigma\, \tilde{m}(A) = 1$, which amounts to observing that for every ω, $x_\omega = \chi_A$ for some $A \subseteq \Lambda$.

Recalling the definition of V (3), we may also consider the numbers $(V\tilde{m})(A)$. These values can also be interpreted as probabilities, provided we define probability in a way which ignores experts who give no possibilities, and provided there are some experts who give some possibilities (i.e., $\tilde{m}(\varnothing) \neq 1$). Then for $A \neq \varnothing$,

$$m(A) = (V\tilde{m})(A) = \frac{\tilde{m}(A)}{1 - \tilde{m}(\varnothing)}$$

is the probability that a randomly chosen expert ω will state that the subset of possibilities is precisely A conditioned on the requirement that the expert gives at least one possibility.

Under the assumptions that $A \neq \varnothing$, $\tilde{m}(\varnothing) \neq 1$, and that probability is measured over the set of experts expressing an opinion $\mathcal{E}' = \{\omega \mid x_\omega \not\equiv 0\}$, many of the quantities in the theory of evidence can be interpreted in terms of familiar statistics on the $x(\lambda)$'s. For example, the belief on a set A,

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

is simply the joint probability

$$\operatorname{Bel}(A) = \operatorname*{Prob}_{\mathcal{E}'}(x(\lambda) = 0 \qquad \text{for } \lambda \notin A).$$

Note that the prior probabilities on the experts in $\mathcal{E}'$ are given by $\mu(\{\omega\})/\mu(\mathcal{E}')$. The denominator in these priors is nonzero due to the assumption that $\tilde{m}(\varnothing) \neq 1$.

In a similar way, plausibility values

$$\operatorname{Pl}(A) = \sum_{B \cap A \neq \varnothing} m(B) = 1 - \operatorname{Bel}(\bar{A})$$

can be interpreted as disjunctive probabilities

$$\operatorname{Pl}(A) = \operatorname*{Prob}_{\mathcal{E}'}(x(\lambda) = 1 \qquad \text{for some } \lambda \in A).$$

The beliefs and plausibilities are the lower and upper probabilities as defined by Dempster. The commonality values

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

are joint probabilities:

$$Q(A) = \operatorname*{Prob}_{\mathcal{E}'}(x(\lambda) = 1 \qquad \text{for } \lambda \in A).$$

To recapitulate, we have defined a mapping from P values to X values, and then transformations from X to $\tilde{m}$ and m values. The resulting element m, which contains statistics on the X variables, is an element in the space of belief states $\mathfrak{M}$ of the Dempster/Shafer theory of evidence (Section II).

C. Bayesian Interpretation

We now interpret the manner in which pairs of experts achieve a consensus opinion. We will show that the combination formulas given for $\mathfrak{N}$ and $\mathfrak{N}'$ are consistent with a Bayesian interpretation. Our treatment is standard.

We first consider the combination of $(\mathcal{E}_1, \mu_1, P_1)$ and $(\mathcal{E}_2, \mu_2, P_2)$ in $\mathfrak{N}$. We assume that the experts in $\mathcal{E}_j$ have available to them information s_j. Note that all experts in a given set of experts share the same information. The information s_j consists of Boolean predicates constituting evidence about the labeling situation. For example, in a medical diagnosis application, s_j might consist of a statements about the presence or absence of a set of symptoms. Each set of experts $\mathcal{E}_j$ deals with a different set of symptoms.

In general, the information s_j is the result of a set of tests having Boolean outcomes. We could write $s_j = f_j(\sigma)$ where f_j represents the tests and σ is the current situation which is an element in some sample space of labeling problems $\sigma \in \Sigma$. Assuming Σ is also a measure space, there are prior probabilities on the information coefficients:

$$\text{Prob}(s_j) = \underset{\sigma\in\Sigma}{\text{Prob}}\left(f_j(\sigma) = s_j\right).$$

There are also prior probabilities on the true label $\lambda(\sigma)$ for labeling situation σ, given by

$$\text{Prob}(\lambda) = \underset{\sigma\in\Sigma}{\text{Prob}}\left(\lambda(\sigma) = \lambda\right).$$

Note that these probabilities are not measured over the space of experts $\mathcal{E}$, but instead are measured over the collection of instances Σ of the labeling problem. For example, in a medical diagnosis domain, Σ might represent the set of all patients.

For $j = 1, 2$, we will suppose that $p_{\omega j}^{(j)}(\lambda)$ represents expert ω_j's estimate of

$$\text{Prob}(\lambda \mid s_j),$$

the probability (over Σ) that $\lambda(\sigma) = \lambda$ conditioned on $f_j(\sigma) = s_j$. The "expert" (ω_1, ω_2) should then estimate $\text{Prob}(\lambda \mid s_1, s_2)$, which is the probability that $\lambda(\sigma) = \lambda$ given that $f_1(\sigma) = s_1$ and $f_2(\sigma) = s_2$, thus combining the two bodies of evidence seen by the two experts in that committee. This committee proceeds as follows.

Bayes' formula implies that

$$\begin{aligned}\text{Prob}(\lambda \mid s_1, s_2) &\\ &= \frac{\text{Prob}(\lambda) \cdot \text{Prob}(s_1, s_2 \mid \lambda)}{\text{Prob}(s_1, s_2)}\\ &= \frac{\text{Prob}(\lambda) \cdot \text{Prob}(s_1 \mid \lambda) \cdot \text{Prob}(s_2 \mid s_1, \lambda)}{\text{Prob}(s_1, s_2)}.\end{aligned}$$

Applying Bayes' formula to $\text{Prob}(s_1 \mid \lambda)$, this becomes

$$\frac{\text{Prob}(s_1)}{\text{Prob}(s_1, s_2)} \cdot \text{Prob}(\lambda \mid s_1) \cdot \text{Prob}(s_2 \mid s_1, \lambda). \tag{6}$$

At this point, we assume that

$$\text{Prob}(s_2 \mid s_1, \lambda) = \text{Prob}(s_2 \mid \lambda). \tag{7}$$

Using this assumption, we obtain by combining (6) and (7) and applying Bayes' formula to $\text{Prob}(s_2 \mid \lambda)$,

$$\begin{aligned}\text{Prob}(\lambda \mid s_1, s_2) &\\ &= c(s_1, s_2) \cdot \frac{\text{Prob}(\lambda \mid s_1) \cdot \text{Prob}(\lambda \mid s_2)}{\text{Prob}(\lambda)}\end{aligned} \tag{8}$$

where $c(s_1, s_2)$ is a constant independent of λ. Using (8), expert (ω_1, ω_2) estimates that

$$p_{(\omega_1,\omega_2)}(\lambda) = c(s_1, s_2) \cdot \frac{p_{\omega_1}^{(1)}(\lambda)\, p_{\omega_2}^{(2)}(\lambda)}{\kappa_\lambda} \tag{9}$$

based on the independence assumption (7) where $\kappa_\lambda = \text{Prob}(\lambda)$. Since the left-hand side of this equation should sum to 1 over λ, we have that

$$c(s_1, s_2) = \frac{1}{\sum_{\lambda'} p_{\omega_1}^{(1)}(\lambda')\, p_{\omega_2}^{(2)}(\lambda')[\kappa_{\lambda'}]^{-1}} \tag{10}$$

unless, of course, this denominator is zero, in which case we resort to setting $p_{(\omega_1,\omega_2)} \equiv 0$. Combining (9) and (10) gives the combination formula given in Definition 3. Thus, we have shown that combination in $\mathfrak{N}$ is a form of Bayesian updating of pairs of experts, based on an independence assumption.

To interpret the combination formula of $\mathfrak{N}'$ in a Bayesian fashion, a weaker independence assumption suffices. The combination formula can be restated as

$$x_{(\omega_1,\omega_2)}(\lambda) = 0 \quad \text{iff } x_{\omega_1}^{(1)}(\lambda) = 0 \text{ or } x_{\omega_2}^{(2)}(\lambda) = 0.$$

Using Bayes' formula and assuming that all prior probabilities are nonzero, it suffices to show that

$$\begin{aligned}\text{Prob}(s_1, s_2 \mid \lambda) = 0 \quad &\text{iff Prob}(s_1 \mid \lambda) = 0\\ &\text{or Prob}(s_2 \mid \lambda) = 0.\end{aligned}$$

The "if" part follows since

$$\begin{aligned}\text{Prob}(s_1, s_2 \mid \lambda) &= \text{Prob}(s_1 \mid \lambda) \cdot \text{Prob}(s_2 \mid s_1, \lambda)\\ &= \text{Prob}(s_2 \mid \lambda) \cdot \text{Prob}(s_1 \mid s_2, \lambda).\end{aligned}$$

The "only if" part becomes our independence assumption, and is equivalent to

$$\begin{aligned}\text{Prob}(s_1 \mid \lambda) > 0 \text{ and } \text{Prob}(s_2 \mid \lambda) > 0&\\ \text{implies Prob}(s_1, s_2 \mid \lambda) > 0.&\end{aligned} \tag{11}$$

This assumption is implied by our earlier hypothesis (7). However, assumption (11) is more defensible, and is actually all that is needed to regard updating in the space of "Boolean opinions of experts" $\mathfrak{N}'$ as Bayesian. Since the Dempster/Shafer theory deals only with the Boolean opinions, (11) is the required independence assumption.

IV. Equivalence with the Dempster/Shafer Rule of Combination

At this point, we have four spaces with binary operations, namely, $(\mathfrak{N}, \otimes)$, $(\mathfrak{N}', \circ)$, $(\mathfrak{M}', \oplus')$, and $(\mathfrak{M}, \oplus)$. We will now show that these four spaces are closely related. It is not hard to show that the binary operation is, in all four cases, commutative and associative, and that each space has an identity element, so that these spaces are Abelian monoids. We also have the following.

Definition 5: The map T

$$T: \mathfrak{N} \rightarrow \mathfrak{N}'$$

with $(\mathcal{E}, \mu, X) = T(\mathcal{E}, \mu, P)$ is given by (4), i.e., $x_\omega(\lambda) = 1$ iff $p_\omega(\lambda) > 0$, and $x_\omega(\lambda) = 0$ otherwise. ■

There is another mapping U, given by the following.

Definition 6:

$$U: \mathfrak{N}' \rightarrow \mathfrak{M}'$$

with $\tilde{m} = U(\mathcal{E}, \mu, X)$ given by (5), i.e.,

$$\tilde{m}(A) = \mu(\{\omega \in \mathcal{E} \mid x_\omega = \chi_A\})/\mu(\{\mathcal{E}\}). \quad ■$$

We will show that T and U preserve the binary operations. More formally, we show that T and U are homomorphisms of monoids.

Lemma 2: T is a homomorphism from $\mathfrak{N}$ onto $\mathfrak{N}'$.

Proof: It is a simple matter to verify that

$$T(\mathcal{E}_1, P_1) \circ T(\mathcal{E}_2, P_2) = T((\mathcal{E}_1, P_1) \otimes (\mathcal{E}_2, P_2)).$$

The essential point, it turns out, is that since the probabilistic opinions are all nonnegative:

$$p_{\omega_1}^{(1)}(\lambda) \cdot p_{\omega_2}^{(2)}(\lambda) > 0 \quad \text{iff } p_{\omega_1}^{(1)}(\lambda) > 0 \text{ and } p_{\omega_2}^{(2)}(\lambda) > 0.$$

T is easily seen to be onto. ■

Lemma 3: U is a homomorphism of $\mathfrak{N}'$ onto $\mathfrak{M}'$.

Proof: Consider $(\mathcal{E}, \mu, X) = (\mathcal{E}_1, \mu_1, X_1) \circ (\mathcal{E}_2, \mu_2, X_2)$. For each $\omega_1 \in \mathcal{E}_1$ and $\omega_2 \in \mathcal{E}_2$, the corresponding $x_{\omega_1}^{(1)}$ and $x_{\omega_2}^{(2)}$ are characteristic functions of subsets of Λ, say χ_B and χ_C, respectively. It is clear that

$$x_{\omega_1}^{(1)} \cdot x_{\omega_2}^{(2)} = \chi_A \qquad \text{iff } B \cap C = A.$$

Thus,

$$x_{(\omega_1, \omega_2)} = \chi_A \qquad \text{iff } x_{\omega_1}^{(1)} = \chi_B \text{ and } x_{\omega_2}^{(2)} = \chi_C \text{ where } B \cap C = A.$$

So

$$\{(\omega_1, \omega_2) \in \mathcal{E} \mid x_{(\omega_1, \omega_2)} = \chi_A\} = \bigcup_{B \cap C = A} \{\omega_1 \mid x_{\omega_1}^{(1)} = \chi_B\} \times \{\omega_2 \mid x_{\omega_2}^{(2)} = \chi_C\}.$$

Since this is a disjoint union, using properties of measures, this gives

$$\mu\{(\omega_1, \omega_2) \in \mathcal{E} \mid x_{(\omega_1, \omega_2)} = \chi_A\} = \sum_{B \cap C = A} \mu_1\{\omega_1 \in \mathcal{E}_1 \mid x_{\omega_1}^{(1)} = \chi_B\} \cdot \mu_2\{\omega_2 \in \mathcal{E}_2 \mid x_{\omega_2}^{(2)} = \chi_C\}.$$

We can divide both sides of this equation by $\mu\{\mathcal{E}\} = \mu_1\{\mathcal{E}_1\} \cdot \mu_2\{\mathcal{E}_2\}$ to obtain

$$\tilde{m}(A) = \sum_{B \cap C = A} \tilde{m}_1(B)\, \tilde{m}_2(C)$$

where $\tilde{m} = U(\mathcal{E}, \mu, X)$, and $\tilde{m}_i = U(\mathcal{E}_i, \mu_i, X_i)$, $i = 1, 2$. Thus,

$$U((\mathcal{E}_1, \mu_1, X_1) \circ (\mathcal{E}_2, \mu_2, X_2)) = U(\mathcal{E}_1, \mu_1, X_1) \otimes' U(\mathcal{E}_2, \mu_2, X_2),$$

which is to say that U is a homomorphism.

Finally, we show that U is onto. Recall that there are n elements in Λ, and so there are 2^n different subsets of Λ. For a given mass distribution $\tilde{m} \in \mathfrak{M}'$, consider a set of 2^n experts $\mathcal{E}$, with each expert $\omega \in \mathcal{E}$ giving a distinct subset $\Gamma(\omega) \subseteq \Lambda$ as the set of possibilities. If we give expert ω the weight $\mu\{\omega\} = \tilde{m}(\Gamma(\omega))$ and set $x_\omega = \chi_{\Gamma(\omega)}$, then it is easy to see that $\tilde{m} = U(\mathcal{E}, \mu, X)$. ■

In the immediately preceding proof that U is onto, we assigned weights to experts. This is the only place were we absolutely require the existence of differential weights on experts. However, if we content ourselves to spaces $\mathfrak{M}'$ and $\mathfrak{M}$ containing only rational values for the mass distribution functions (as, for example, is the case in any computer implementation), then the weights can be eliminated and replaced by counting measure.

Recall from Section II that the map $V: \mathfrak{M}' \rightarrow \mathfrak{M}$ is also a homomorphism. So we can compose the homomorphisms $T: \mathfrak{N} \rightarrow \mathfrak{N}'$ with $U: \mathfrak{N}' \rightarrow \mathfrak{M}'$ with $V: \mathfrak{M}' \rightarrow \mathfrak{M}$ to obtain the following obvious theorem.

Theorem: The map $V \circ U \circ T: \mathfrak{N} \rightarrow \mathfrak{M}$ is a homomorphism of monoids mapping onto the space of belief states $(\mathfrak{M}, \oplus)$. ■

This theorem provides the justification for the viewpoint that the theory of evidence space $\mathfrak{M}$ represents the space $\mathfrak{N}$ via the representation $V \circ U \circ T$. The proof follows from the lemmas; since each of the component maps in this representation is an onto homomorphism, the composition also maps homomorphically onto the entire theory of evidence space.

The significance of this result is that we can regard combinations of elements in the theory of evidence as combinations of elements in the space of opinions of experts. For if $m_1, \cdots, m_k$ are elements in $\mathfrak{M}$ which are to be combined under $\oplus$, we can find respective preimages in $\mathfrak{N}$ under the map $V \circ U \circ T$, and then combine those elements using the operation $\otimes$ in the space of opinions of experts $\mathfrak{N}$. After all combinations in $\mathfrak{N}$ are completed, we project back to M by $V \circ U \circ T$; the result will be the same as if we had combined the elements in $\mathfrak{M}$. The only advantage to this procedure is that combinations in $\mathfrak{N}$ are conceptually simpler: we can regard the combination as Bayesian updatings on the product space of experts.

V. An Alternative Method for Combining Evidence

With the viewpoint that the theory of evidence is really simply statistics of opinions of experts, we can make certain remarks on the limitations of the theory.

1) There is no use of probabilities or degrees of confidence. Although the belief values seem to give weighted results, at the base of the theory, experts only say whether a condition is possible or not. In particular, the theory makes no distinction between an expert's opinion that a label is likely or that it is remotely possible.

2) Pairs of experts combine opinions in a Bayesian fashion with independence assumptions of the sources of evidence. In particular, dependencies in the sources of information are not taken into account.

3) Combinations take place over the product space of experts. It might be more reasonable to have a single set of experts modifying their opinions as new information comes in, instead of forming the set of all committees of mixed pairs.

Both the second and third limitations come about due to the desire to have a combination formula which factors through to the statistics of the experts and is application-independent. The need for the second limitation, the independence assumption on the sources of evidence, is well known (see, e.g., [29]). Without incorporating much more complicated models of judgments under multiple sources of knowledge, we can hardly expect anything better.

The first objection, however, suggests an alternate formulation which makes use of the probabilistic assessments of the experts. Basically, the idea is to keep track of the density distributions of the opinions in probability space. Of course, complete representation of the distribution would amount to recording the full set of opinions p_ω for all ω. Instead, it is more reasonable to approximate the distribution by some parameterization, and update the distribution parameters by combination formulas.

We present a formulation based on normal distributions of logarithms of updating coefficients. Other formulations are possible. In marked contrast to the Dempster/Shafer formulation, we assume that all opinions of all experts are nonzero for every label. That is, instead of converting opinions into Boolean statements by test for zero, we will assume that all the values are nonzero, and model the distribution of their strengths.

A simple rewrite of (8) of Section III-C yields

$$\mathrm{Prob}(\lambda | s_1, s_2) = c(s_1, s_2) \cdot \mathrm{Prob}(\lambda) \cdot \frac{\mathrm{Prob}(\lambda | s_1)}{\mathrm{Prob}(\lambda)} \cdot \frac{\mathrm{Prob}(\lambda | s_2)}{\mathrm{Prob}(\lambda)}.$$

This equation depends on an independence assumption, (7). We can iterate this equation to obtain a formula for $\mathrm{Prob}(\lambda | s_1, \cdots, s_k)$. In this iteration process, s_1 and s_2 successively take the place of $s_1 \wedge \cdots \wedge s_i$ and s_{i+1}, respectively, as i increases from 1 to $k - 1$. Accordingly, we require a sequence of independence assumptions, which will take the form

$$\mathrm{Prob}(s_i | s_1 \wedge \cdots \wedge s_{i-1}, \lambda) = \mathrm{Prob}(s_i | \lambda)$$

for $i = 1, \cdots, k$. Under these assumptions, we obtain

$$\mathrm{Prob}(\lambda | s_1, \cdots, s_k) = c(s_1, \cdots, s_k) \cdot \mathrm{Prob}(\lambda) \cdot \prod_{i=1}^{k} \frac{\mathrm{Prob}(\lambda | s_i)}{\mathrm{Prob}(\lambda)}.$$

In a manner similar to [3], set

$$L(\lambda | s_i) = \log \left[\frac{\mathrm{Prob}(\lambda | s_i)}{\mathrm{Prob}(\lambda)} \right].$$

(Note, incidentally, that these values are not the so-called "log-likelihood ratios"; in particular, the $L(\lambda | s_i)$'s can be both positive and negative.) We then obtain

$$\log[\mathrm{Prob}(\lambda | s_1, \cdots, s_k)] = c + \log[\mathrm{Prob}(\lambda)] + \sum_{i=1}^{k} L(\lambda | s_i)$$

where c is a constant independent of λ (but not of $s_1, \cdots, s_k$).

The consequence of this formula is that if the independence assumptions hold, and if $\mathrm{Prob}(\lambda)$ and $L(\lambda | s_i)$ are known for all λ and i, then the values $\mathrm{Prob}(\lambda | s_1, \cdots, s_k)$ can be calculated from

$$\mathrm{Prob}(\lambda | s_1, \cdots, s_k) = \frac{\mathrm{Prob}(\lambda) \cdot \exp\left[\sum_{i=1}^{k} L(\lambda | s_i)\right]}{\sum_{\lambda'} \mathrm{Prob}(\lambda') \exp\left[\sum_{i=1}^{k} L(\lambda' | s_i)\right]}. \tag{12}$$

Accordingly, we introduce a space which we term "logarithmic opinions of experts." For convenience, we will assume that experts have equal weights. An element in this space will consist of a set of experts $\mathcal{E}_i$ and a collection of opinions $Y_i = \{ y_\omega^{(i)} \}_{\omega \in \mathcal{E}_i}$. Each $y_\omega^{(i)}$ is a map, and the component (λ) represents expert ω's estimate of $L(\lambda | s_i)$:

$$y_\omega^{(i)}: \Lambda \to \mathrm{R}, \; y_\omega^{(i)}(\lambda) \approx L(\lambda | s_i)$$

Note that the experts in $\mathcal{E}_i$ all have knowledge of the information s_i, and that the estimated logarithmic coefficients $L(\lambda | s_i)$ can be positive or negative. In fact, since the experts do not necessarily have precise knowledge of the value of $\mathrm{Prob}(\lambda)$, but instead provide estimates of logs of ratios, the estimates can lie in an unbounded range.

In analogy with our map to a statistical space (Section III-B), we can define a space which might be termed the "parameterized statistics of logarithmic opinions of experts." Elements in this space will consist of pairs $(\bar{u}, C)$ where $\bar{u}$ is in R^n and C is a symmetric $n \times n$ matrix. We next describe how to project from the space of logarithmic opinions to the space of parameterized statistics.

Let us suppose that for a set of experts $\mathcal{E}$ and for $\Lambda = \{ \lambda_1, \cdots, \lambda_n \}$, the n-vectors composed of the logarithmic opinions $\bar{y}_\omega \in \mathrm{R}^n$, $\bar{y}_\omega = (y_\omega(\lambda_1), \cdots, y_\omega(\lambda_n))$ are approximately (multi-) normally distributed. Thus, we model the distribution of the random vector $\bar{y} = (y(\lambda_1), \cdots, y(\lambda_n))$ by the density function

$$m(\bar{y}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det C}} \cdot \exp\left((\bar{y} - \bar{u})^T C^{-1} (\bar{y} - \bar{u})\right), \qquad \bar{y} \in \mathrm{R}^n$$

where $\bar{u} \in \mathrm{R}^n$ is the mean of the distribution, and C is the

$n \times n$ covariance matrix. That is, in terms of the expectation operator $E\{\cdot\}$ on random variables over the sample space $\mathcal{E}$,

$$\bar{u} = (u_1, \cdots, u_n)$$
$$u_i = E\{y(\lambda_i)\},$$

and for $C = (c_{ij})$,

$$c_{ij} = E\{(y(\lambda_i) - u_i) \cdot (y(\lambda_j) - u_j)\}.$$

These measurements of the statistics of the $y(\lambda)$'s can be made regardless of the true distributions. The accuracy of the model depends on the degree to which the multinormal distribution assumption is valid.

Next we discuss combination formulas in both spaces. Suppose $(\mathcal{E}_i, Y_i)$, $i = 1, 2$ are two elements in the space of logarithmic opinions, each describing a sample space of experts together with opinions. Since according to (12) the logarithmic opinions add, we define the combination of the two elements by $(\mathcal{E}, Y)$ where

$$\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$$
$$Y = \{y_{(\omega_1,\omega_2)}\}_{(\omega_1,\omega_2)\in\mathcal{E}}$$
$$y_{(\omega_1,\omega_2)}(\lambda) = y^{(1)}_{\omega_1}(\lambda) + y^{(2)}_{\omega_2}(\lambda).$$

To consider combinations in the space of statistics, let $m_i(\bar{y})$ be the density function over R^n for the random vector $\bar{y}^{(i)}$ over the sample space $\mathcal{E}_i$, $i = 1, 2$. Assume that each m_i is a multinormal distribution, associated with a mean vector $\bar{u}^{(i)}$ and a covariance $C^{(i)}$. In order that the projection to the space of statistics be a homomorphism, the definition of combination in the space of statistics should respect the true statistics of the combined opinions. The density function $m(\bar{y})$ for the combination $\bar{y}_{(\omega_1,\omega_2)}$, $(\omega_1, \omega_2) \in \mathcal{E}$ is given by

$$m(\bar{y}) = \int_{\mathrm{R}^n} m_1(\bar{y}')\, m_2(\bar{y} - \bar{y}')\, d\bar{y}'.$$

This is the point where we use the fact that the logarithmic opinions add under combination.

Projecting to the space of statistics, we discover the advantage of modeling the distributions by normal functions. Namely, since the convolution of a Gaussian by a Gaussian is once again a Gaussian, we define the combination formula

$$(\bar{u}^{(1)}, C^{(1)}) \oplus (\bar{u}^{(2)}, C^{(2)}) = (\bar{u}^{(1)} + \bar{u}^{(2)}, C^{(1)} + C^{(2)}).$$

That is, since m_1 and m_2 are multinormal distributions, their convolution is also multinormal with mean and covariance which are the sums of the contributing means and covariances. (This result is easily proven using Fourier transforms.) An extension to the case where $\mathcal{E}_1$ and $\mathcal{E}_2$ have nonequal total weights is straightforward.

Having defined combination in the space of statistics, one must show that the transformation from the space of opinions to the space of statistics is a homomorphism, even when the logarithmic opinions are not truly normally distributed. This is easily done since the means and covariances of the sum of two random vectors are the sums of the means and covariances of the two random vectors.

To interpret a state $(\bar{u}, C)$ in the space of parameterized statistics, we must remember the origin of the logarithmic-opinion values. Specifically, after k updating iterations combining information s_1 through s_k, the updated vector $\bar{y} = (y_1, \cdots, y_n) \in \mathrm{R}^n$ is an estimate of the sum of the logarithmic coefficients:

$$y_j \approx \sum_{i=1}^{k} L(\lambda | s_j).$$

According to (12), the *a posteriori* probabilities can then be calculated from this estimate (provided the prior $\mathrm{Prob}(\lambda)$'s are known). In particular, the *a posteriori* probability of a label λ_j is high if the corresponding coefficient $y_j + \log[\mathrm{Prob}(\lambda_j)]$ is large in comparison to the other components $y_j + \log[\mathrm{Prob}(\lambda_j)]$.

Since the state $(\bar{u}, C)$ represents a multinormal distribution in the log-updating space, we can transform this distribution to a density function for *a posteriori* probabilities. Basically, a label will have a high probability if $u_j + \log[\mathrm{Prob}(\lambda_j)]$ is relatively large. However, the components of $\bar{u}$ represent the center of the distribution (before bias by the priors). The spread of the distribution is given by the covariance matrix, which can be thought of as defining an ellipsoid in R^n centered at $\bar{u}$. The exact equation of the ellipse can be written implicitly as

$$(\bar{y} - \bar{u})^T C^{-1} (\bar{y} - \bar{u}) = 1.$$

This ellipse describes a "one sigma" variation in the distribution, representing a region of uncertainty of the logarithmic opinions; the distribution to two standard deviations lies in a similar but enlarged ellipse. The eigenvalues of C give the squared lengths of the semi-major axes of the ellipse, and are accordingly proportional to degrees of confidence. The eigenvectors give the directions in which the eigenvalues measure their uncertainty. Bias by the prior probabilities simply adds a fixed vector, with components $\log[\mathrm{Prob}(\lambda_j)]$, to the ellipse, thereby translating the distribution. We seek an axis j such that the components y_j of the vectors y lying in the translated ellipse are relatively much larger than other components of vectors in the ellipse. In this case, the preponderant evidence is for label λ_j.

For example, in a three-label case, we might have priors of approximately (0.01, 0.19, 0.8), and evidence with the following means and covariances in log-probability space of

$$u_1 = (1, 0, -0.01)$$

$$C_1 = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.001 \end{pmatrix}$$

and

$$u_2 = (0.4, -0.1, -0.2)$$

$$C_2 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0.05 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}.$$

Then adding means and covariances, and using (12) to reinterpret in terms of probabilities, we come up with a current estimated probability distribution (0.64, 0.08, 0.28), but with a large uncertainty region. For example, within a one-sigma displacement from the mean opinion, we have the distribution (0.13, 0.18, 0.69). We conclude that the evidence tends to indicate that label 1 is probable, but there is considerable uncertainty.

Clearly, the combination formula is extremely simple. Its greatest advantage over the Dempster/Shafer theory of evidence is that only $O(n^2)$ values are required to describe a state, as opposed to the 2^n values used for a mass distribution in $\mathfrak{M}$. The simplicity and reduction in numbers of parameters has been purchased at the expense of an assumption about the kinds of distributions that can be expected. However, the same assumption allows us to track probabilistic opinions (or actually, the logarithms), instead of converting all opinions into Boolean statements about possibilities.

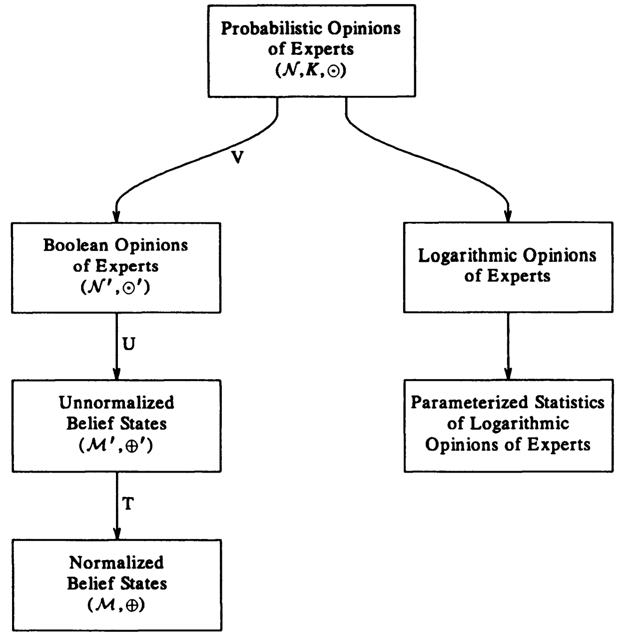


Fig. 2. Names of spaces and maps between them. Each box contains the name of space, and each arrow is a homomorphism that maps onto the next space, thereby defining a representation. Note that the left branch gives the spaces involved in the interpretation of the Dempster/Shafer theory of evidence, whereas the right branch is the alternative method for combining evidence presented in Section V.

VI. Conclusions

We have shown how the theory of evidence may be viewed as a representation of a space of opinions of experts where opinions are combined in a Bayesian fashion over the product space of experts. (Refer to Fig. 2.) By "representation," we mean something very specific—namely, that there is a homomorphism mapping from the space of opinions of experts onto the Dempster/Shafer theory of evidence space. This map fails to be an isomorphism (which would imply equivalence of the spaces) only insofar as it is many-to-one. That is, for each state in the theory of evidence, there is a collection of elements in the space of opinions of experts which all map to the single state. In this way, the state in the theory of evidence represents the corresponding collection of elements. In fact, what this collection of elements has in common is that the statistics of the opinions of the experts defined by the element are similar in terms of the way statistics are measured by the map U.

Furthermore, combination in the space of opinions of experts, as defined in Section III, leads to combination in the theory of evidence space. This allows us to implement combination in a somewhat simpler manner since the formulas for combination without the normalization are simpler than the more standard formulas, and also permits us to view combination in the theory of evidence space as the tracking of statistics of opinions of experts as they combine information in a pairwise Bayesian fashion over the product space of experts. Applying a Bayesian interpretation to the updating of the opinions of experts also makes clear the implicit independence assumptions which must exist in order to combine evidence in the prescribed manner.

From this viewpoint, we can see how the Dempster/Shafer theory of evidence accomplishes its goals. Degrees of support for a proposition, belief, and plausibilities are all measured in terms of joint and disjunctive probabilities over a set of experts who are naming possible labels given current information. The problem of ambiguous knowledge versus uncertain knowledge, which is frequently described in terms of "withholding belief," can be viewed as two different distributions of opinions. In particular, ambiguous knowledge can be seen as observing high densities of opinions on particular disjoint subsets, whereas uncertain knowledge corresponds to unanimity of opinions where the agreed-upon opinion gives many possibilities. Finally, instead of performing Bayesian updating, a *set* of values is updated in a Bayesian fashion over the product space, which results in non-Bayesian formulas over the space of labels.

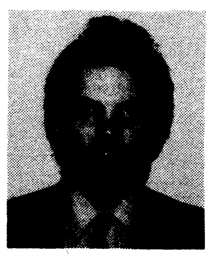
In meeting each of these goals, the theory of evidence invokes compromises that we might wish to change. For example, in order to track statistics, it is necessary to model the distribution of opinions. If these opinions are probabilistic assignments over the set of labels, then the distribution function will be too complicated to retain precisely. The Dempster/Shafer theory of evidence solves this problem by simplifying the opinions to Boolean decisions, so that each expert's opinion lies in a space having 2^n elements. In this way, the full set of statistics can be specified using 2^n values. We have suggested an alternate method, which retains the probability values in the opinions without converting them into Boolean decisions, and requires only $O(n^2)$ values to model the distribution, but fails to retain full information about the distribution. Instead, our method attempts to approximate the distribution of opinions with a Gaussian function.

Acknowledgment

We thank O. Faugeras for an introduction to the topic and T. Levitt for much assistance. We appreciate the helpful comments given by G. Reynolds, D. Strahman, and J.-C. Falmagne.

References

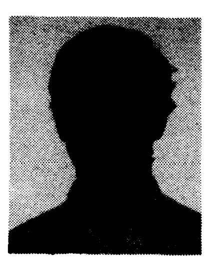
[1] J. A. Anderson, "Distinctive features, categorical perception, and probability learning: Some applications of a neural model," *Psychol. Rev.*, vol. 84, pp. 413-451, 1977.
[2] J. A. Barnett, "Computational methods for a mathematical theory of evidence," in *Proc. 7th Int. Joint Conf. Artificial Intell.*, 1981, pp. 868-875.
[3] E. Charniak, "The Bayesian basis of common sense medical diagnosis," *Proc. AAAI*, pp. 70-73, 1983.
[4] A. P. Dempster, "Upper and lower probability inferences based on a sample from a finite univariate population," *Biometrika*, vol. 54, pp. 515-528, 1967.
[5] —, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, vol. 38, pp. 325-339, 1967.
[6] —, "A generalization of Bayesian inference," *J. Roy. Statist. Soc.*, ser. B, vol. 30, pp. 205-247, 1968.
[7] J. C. Falmagne, "A random utility model for a belief function," *Synthese*, vol. 57, pp. 35-48, 1983.
[8] O. D. Faugeras, "Relaxation labeling and evidence gathering," in *Proc. 6th Int. Conf. Pattern Recognition*, IEEE Comput. Soc., Oct. 1982, pp. 405-412.
[9] P. C. Fishburn, *Decision and Value Theory*. New York: Wiley, 1964.
[10] L. Friedman, "Extended plausible inference," in *Proc. 7th Int. Joint Conf. Artificial Intell.*, 1981, pp. 487-495.
[11] T. D. Garvey, J. D. Lowrance, and M. A. Fischler, "An inference technique for integrating knowledge from disparate sources," in *Proc. 7th Int. Joint Conf. Artificial Intell.*, 1981, pp. 319-325.
[12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721-741, 1984.
[13] I. J. Good, "The measure of a non-measurable set," in *Logic, Methodology, and Philosophy of Science*, E. Nagel, P. Suppes, and A. Tarski, Eds. Stanford, CA: Stanford Univ. Press, 1962, pp. 319-329.
[14] J. Gordon and E. H. Shortliffe, "The Dempster-Shafer theory of evidence," in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, B. G. Buchanan and E. H. Shortliffe, Eds. Reading, MA: Addison-Wesley, 1984.
[15] —, "A method of managing evidential reasoning in a hierarchical hypothesis space," Dep. Comput. Sci., Stanford Univ., Stanford, Ca, Tech. Rep., 1984.
[16] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 267-287, May 1983.
[17] B. O. Koopman, "The axioms and algebra of intuitive probability," *Ann. Math.*, vol. 41, pp. 269-292, 1940. See also "The bases of probability," *Bull. Amer. Math. Soc.*, vol. 46, pp. 763-774, 1940.
[18] D. H. Krantz and J. Miyamoto, "Priors and likelihood ratios as evidence," *J. Amer. Statist. Ass.*, vol. 78, pp. 418-423, 1983.
[19] H. E. Kyburg, Jr., "Bayesian and non-Bayesian evidential updating," Dep. Comput. Sci., Univ. Rochester, Rochester, NY, Tech. Rep. 139, July 1985.
[20] M. S. Landy and R. A. Hummel, "A brief survey of knowledge aggregation methods," in *Proc. Int. Conf. Pattern Recognition*, Oct. 1986, pp. 248-252.
[21] D. V. Lindley, "Scoring rules and the inevitability of probability," *Int. Statist. Rev.*, vol. 50, pp. 1-26, 1982.
[22] H. Prade, "A computational approach to approximate and plausible reasoning with applications to expert systems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 260-283, 1985.
[23] G. Reynolds, L. Wesley, D. Strahman, and N. Lehrer, "Converting feature values to evidence," Dep. Comput. Inform. Sci., Univ. Massachusetts, Amherst, Tech. Rep., Nov. 1985, in preparation.
[24] G. Shafer, "Belief functions and parametric models," *J. Roy. Statist. Soc.*, vol. B44, pp. 322-352, 1982 (includes commentaries).
[25] —, "Probability judgement in artificial intelligence," in *Uncertainty in Artificial Intelligence*, L. Kanal and J. Lemmer, Ed. Amsterdam: Elsevier Science, North-Holland, 1986, pp. 127-135.
[26] —, "Allocations of probability," Ph.D. dissertation, Princeton Univ., Princeton, NJ, 1973. Available from University Microfilms, Ann Arbor, MI.
[27] —, "A theory of statistical evidence," in *Foundations and Philosophy of Statistical Theories in the Physical Sciences, Vol II*, W. L. Harper and C. A. Hooker, Ed. Reidel, 1975.
[28] —, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[29] —, "Constructive probability," *Synthese*, vol. 48, pp. 1-60, 1981.
[30] —, "Lindley's paradox," *J. Amer. Statist. Ass.*, vol. 77, pp. 325-351, 1982 (includes commentaries).
[31] C. A. B. Smith, "Personal probability and statistical analysis," *J. Roy. Statist. Soc.*, ser. A, vol. 128, pp. 469-499, 1965 (with discussion). See also "Personal probability and statistical analysis," *J. Roy. Statist. Soc.*, ser. B, vol. 23, pp. 1-25.
[32] T. M. Strat, "Continuous belief functions for evidential reasoning," in *Proc. Nat. Conf. Artificial Intell.*, 1984, pp. 308-313.
[33] P. M. Williams, "On a new theory of epistemic probability," *Brit. J. Phil. Sci.*, vol. 29, pp. 375-387, 1978.



Robert A. Hummel (M'82) received the B.A. degree in mathematics from the University of Chicago, Chicago, IL, and the Ph.D. degree in mathematics from the University of Minnesota, Minneapolis.

He was associated with the University of Maryland Computer Vision Laboratory while in school, and has also worked at Stanford Research Institute, International; Honeywell's Systems and Research Center; and Martin Marietta Aerospace, Orlando, FL. From 1980 to 1982, he was a Postdoctoral Fellow in mathematics at New York University's Courant Institute of Mathematical Sciences. Since then, he has been a faculty member in the Department of Computer Science at the Courant Institute. In mathematics, his research interests include variational methods for the study of partial differential equations. His main field of interest is computer vision, especially issues in the representation and inference of information from images.

Dr. Hummel is a member of Phi Beta Kappa and the Association for Computing Machinery.



Michael S. Landy received the B.S. degree in electrical engineering and computer science from Columbia University, New York, NY, in 1974, and the M.S. and Ph.D. degrees in computer and communication sciences from the University of Michigan, Ann Arbor, in 1977 and 1981, respectively.

From 1981 to 1984 he was a Research Scientist at the Human Information Processing Laboratory, Department of Psychology, New York University, New York, NY. He is currently an Assistant Professor in the Department of Psychology at NYU. His research interests include visual psychophysics and perception, applications of image processing to psychology, and mathematical psychology.

Dr. Landy is a member of the Association for Research on Vision and Ophthalmology, the Optical Society of America, and the Pattern Recognition Society.