Analysis of Fuzzy Information
Volume I: Mathematics and Logic
James C. Bezdek, ed.
CRC Press, 1987

Chapter 3

BELIEF FUNCTIONS AND POSSIBILITY MEASURES

Glenn Shafer

ABSTRACT

This article studies the connections between the theories of belief functions and possibility measures. It shows how both use compatibility relations linking different frames of discernment. It explains the canonical examples on which belief functions are based and challenges students of possibility measures to provide alternative canonical examples to justify those rules for possibility measures that diverge from the rules for belief functions.

TABLE OF CONTENTS

I. INTRODUCTION

There are close formal connections between the theory of belief functions, developed by Dempster[1-6] and Shafer,[7-19] and the theory of possibility measures, developed by Zadeh.[20-22] Associated with every belief function is another function called its plausibility function, and a possibility measure is a special kind of plausibility function, a consonant plausibility function. Moreover, the two theories agree in their treatment of marginalization and extension. However, how much do the two theories have in common conceptually?

This article reviews the conceptual basis of the theory of belief functions and studies the rules for consonant plausibility functions. It then examines the points of agreement and disagreement between these rules and the rules Zadeh and others have suggested for possibility measures. It asks whether there is a conceptual basis for the theory of possibility measures that can account for the divergences.

The theory of belief functions and the theory of possibility measures are both languages for subjective probability judgment; they both provide a vocabulary within which a person can make judgments about the strength of inconclusive evidence. One way to explain the conceptual basis of such a language is to describe the language's canonical examples — the idealized examples to which the language compares real problems.[23] As we see in Section V below, the language of belief functions compares real problems to canonical examples in which the meaning of evidence, though uncertain, depends on known chances. Possibility measures, on the other hand, are said to gauge imprecision as well as uncertainty. Can we formulate canonical examples that elucidate the nature of this imprecision? If we use canonical examples in which the meaning of a statement is more or less precise with known probabilities, then we are working with a subset of the canonical examples for belief functions, and in this case possibility measures should obey all the rules for plausibility functions. To justify different rules for possibility measures, we need different canonical examples.

The article begins with a study of frames of discernment and compatibility relations, general concepts that provide a setting for both belief functions and possibility measures. We then study Sugeno measures, the most general set functions that can be used to represent subjective probability judgments within this framework. We then briefly review probability measures, emphasizing those aspects that are generalized or imitated by belief functions and possibility measures. After reviewing belief functions at length, we then study possibility measures — their relation to consonant plausibility functions and the problem of finding canonical examples for them. The article concludes with an examination of two examples that Zadeh has used to criticize the normalization of belief functions.

II. FRAMES OF DISCERNMENT

A set is called a *frame of discernment*, or simply a *frame*, when its elements are interpreted as possible answers to a certain question, and we know or believe that exactly one of these answers is correct. A frame is an epistemic or doxic object; its elements are possible relative to our knowledge and opinion.

Consider two frames S and T. We say that an element s of S is *compatible* with an element t of T if it is possible, relative to our knowledge and opinion, that s is the answer to the question considered by S *and* t is the answer to the question considered by T. It is worth repeating that both compatibility and possibility, as the words are used here, are relative to our knowledge and opinion. If we learn more or change our opinion, our frames themselves may change (some previously possible answers may be ruled out) and even if they do not change, the compatibility relation between them may change (some previously possible combinations of answers may be ruled out). More radical changes are also possible.

Given two frames S and T and an element s of S, there must be some element t of T that is compatible with s. Otherwise, consideration of T would tell us that s is impossible, and it is part of our idea of a frame that all its elements are possible relative to our knowledge. Similarly, given three frames, S, T, and U, and given compatible elements s of S and u of U, there must exist an element t of T such that t is compatible both with s and with u. If we write sCt for "s is compatible with t", then these two principles can be written as follows:

Principle I: If s is in S, then there exists t in T such that sCt.

Principle II: If s is in S, u is in U, and sCu, then there exists t in T such that sCt and tCu.

We also take it for granted, of course, that compatibility is symmetric: sCt if and only if tCs.

If sCt for all s in S and all t in T, we say that S and T are *independent*; otherwise they are *dependent*. If for each s in S there is only one t in T such that sCt, we say that T is a *coarsening* of S and that S is a *refinement* of T.

A relation between two sets is a subset of their Cartesian product — the subset consisting of all pairs for which the relation holds. Let us denote by C(S,T) the subset of $S \times T$ consisting of all pairs (s,t) such that sCt. By Principle I, we must have $C(S,T) \cap (\{s\} \times T) \neq \emptyset$ for all s in S and $C(S,T) \cap (S \times \{t\}) \neq \emptyset$ for all t in T. In general, let us call any subset C of SxT such that $C \cap (\{s\} \times T) \neq \emptyset$ for all s in S and $C \cap (S \times \{t\}) \neq \emptyset$ for all t in T a *compatibility relation* between S and T.

The actual compatibility relation C(S,T) between two frames S and T can be thought of as a third frame that is a refinement of both S and T. Indeed, the elements of C(S,T) are the possible answers to the conjunction of the questions asked by S and T. The frame C(S,T) is the *minimal* refinement of S and T; any other frame that refines both S and T also refines C(S,T). (Note: When I call C(S,T) the *actual* compatibility relation between S and T, I mean only that it is the compatibility relation that actually represents our knowledge and opinion; I do not mean that it tells which elements of S and T are compatible in some objective sense, independent of our knowledge and opinion.)

Another way to represent a compatibility relation mathematically is through a *multivalued mapping* from one of the frames to the power set of the other.[2] A compatibility relation between S and T can be represented by the multivalued mapping $G{:}S \rightarrow 2^T$ given by $G(s) = \{t \text{ in } T | sCt\}$. (It can also be represented by the multivalued mapping $H{:}T \rightarrow 2^S$ given by $H(t) = \{s \text{ in } S | sCt\}$.) In general, a multivalued mapping $G{:}S \rightarrow 2^T$ represents a compatibility relation if and only if $G(s) \neq \emptyset$ for all s in S and $\cup\{G(s) | s \text{ in } S\} = T$.

A. Mediating and Discerning Interaction

Given three frames S, T, and U, set

$$C_T(S,U) = \{(s,u) | \text{for some t in T, sCt, and tCu}\} \tag{1}$$

It follows from Principle II above that $C_T(S,U)$ must contain C(S,U). However, the two may fail to be equal. If they are equal, then we say that *the interaction between S and U is mediated by T*. This means that the compatibility relation between S and U can be deduced from the compatibility relations between S and T and between T and U. It follows from Principle II that the compatibility relation between S and U is mediated by T whenever either S or U is a coarsening of T.

We will find it useful later to express Equation 1 in terms of multivalued mappings: if $G_1:S \rightarrow 2^T$, $G_2:T \rightarrow 2^U$, and $G:S \rightarrow 2^U$ are multivalued mappings describing C(S,T), C(T,U), and $C_T(S,U)$, respectively, then

$$G(s) = \cup\{G_2(t)|t \text{ in } G_1(s)\} \qquad (2)$$

for all s in S.

It may be helpful to give an example where T mediates the relation between S and U even though neither S nor U are coarsenings of T. Suppose Tom and Mary know that Early Bird Airways has only one flight a day from San Francisco to Kansas City. However, they are uncertain about the time of departure. One schedule, which Tom got from a friend yesterday, says that the flight departs at 8:00 a.m. Another, which Mary just found in their hotel's rack of travel literature, says that it departs at 7:00 a.m. Suppose Tom and Mary form three frames: S, which considers whether the hotel's travel literature is kept scrupulously up to date; T, which considers which, if either, of the schedules is current; and U, which considers the time of the flight. We can list the elements of these frames as follows:

S = {kept up to date, not kept up to date},
T = {Tom's schedule current, Mary's current, neither current},
U = {00:00, 00:05, 00:10, . . . , 23:45, 23:50, 23:55}.

For brevity, let us also denote the elements of S and T by s_1, s_2, t_1, t_2, t_3, respectively. It is evident that s_1 is compatible only with t_2 in T, while s_2 is compatible with all three elements of T. Similarly, t_1 is compatible only with 8:00 in U, t_2 is compatible only with 7:00 in U, and t_3 is compatible with any element of U. We can see, moreover, that T mediates the relation between S and U. So s_1 is compatible only with 7:00 in U, while s_2 is compatible with all 288 elements of U.

Thus far we have considered only pairwise compatibility of elements of different frames. However, we can also ask about the compatibility of larger collections of answers to different questions. Given, for example, elements s, t, and u of three frames S, T, and U, we can ask whether these elements are compatible as a triplet: is it possible that s is the correct answer to the question considered by S, t is the correct answer to the question considered by T, *and* u is the correct answer to the question considered by U? Compatibility as a triplet can also be defined in terms of the notation we have already introduced; s, t, and u are compatible as a triplet if (s,u) is in C(S,U) and ((s,u),t) is in C(C(S,U),T). In general, the pairwise compatibility of s,t, and u does not imply their compatibility as a triplet: we can have sCt, tCu, and sCu without having (s,u)Ct. We can, however, strengthen Principle II to the following.

Principle III: If s is in S, u is in U, and sCu, then there exists t in T such that s, t, and u are compatible as a triplet.

It follows from Principle III that if one of the three frames S, T, and U is a refinement of at least one of the others, then any pairwise compatible elements s, t, and u are also compatible as a triplet. If, for example, U is a refinement of T, then there is only one element of T that is compatible with a given element u of U, and so if sCu and tCu, then t must be the element of T that satisfies Principle III.

In general, the possibility that s, t, and u can be pairwise compatible without being compatible as a triplet can always be eliminated by making one of the frames finer, and usually it is not necessary to make it so fine that it is a refinement of one of the others. By putting more detail into t, say, we can assure that whenever it is ruled out by s and u together, it is also ruled out by at least one of them separately. This suggests the following terminology: if three frames are such that three pairwise compatible elements are always compatible as a triplet, then we say that each frame *discerns the relevant interaction* between the other two.

Let us say that *T strongly mediates the interaction between S and U* whenever T both mediates the interaction between S and U and also discerns the relevent interaction between them. This means that from sCt and tCu it follows that s, t, and u are compatible as a triplet. In the example concerning Early Bird Airways, we would probably say that T strongly mediates the interaction between S and U.

B. Constructing Frames and Compatibility Relations

When probability is studied in the context of formal logic, it is assumed that probabilities are attached to sentences or propositions in a formal language. In this context, a frame of discernment is a list of sentences. If it is assumed that the language is rich enough to express all our knowledge that is revelant to the question considered by the frame, then it can further be assumed that the sentences in this list are detailed enough that our knowledge that only one can be correct is reflected in the syntax of the sentences: they are mutually contradictory.

In practice, however, a frame of discernment is constructed for a particular problem in the context of our knowledge of that problem, using a plethora of assumptions and judgments, many of which are both implicit and provisional. The elements of the frame are described linguistically, but in a language that is, to some extent, made for the occasion. The meaning of each element is only partly communicated by its linguistic form; it is further specified by the assertion that one and only one of the elements is correct and by discussion of how we might learn which is correct.

The compatibility relations between frames are also constructed, and this construction further refines the meaning of the elements in each frame. One kind of judgment that can help in the construction of a compatibility relation is the judgment that one frame mediates the interaction between two others. In the example concerning Early Bird Airways, we made such a judgment in order to construct C(S,U) on the basis of our construction of C(S,T) and C(T,U). In this example, and probably in most other examples where we make the judgment that one frame mediates the interaction between two other frames, we are also implicitly making the judgment that it strongly mediates this interaction, and thus we are constructing not only a compatibility relation between the two frames but also a compatibility relation among all three.

III. SUGENO MEASURES

Suppose S is a frame, and suppose Bel is a real-valued function on 2^S, the set of all subsets of S, satisfying

$$\text{(i) } Bel(\emptyset) = 0,$$
$$\text{(ii) } Bel(S) = 1,$$
$$\text{(iii) } Bel(A) \leq Bel(B) \text{ whenever } A \subseteq B \subseteq S$$

Then we may call Bel a *Sugeno measure*. A Sugeno measure Bel is of interest when Bel(A) can be interpreted as our degree of belief that A contains the correct answer to the question considered by S. Conditions (i) and (ii) can be regarded as conventions; they indicate that zero means no belief, while one means total belief. Condition (iii) has more content; it is based on the idea that when A ⊆ B, the correct answer can be in A only if it is also in B, so that belief in the former should also count as belief in the latter. (Note: The set functions called *Sugeno measures* here have been called *fuzzy measures* by Sugeno[24] and other authors.[25,26] Such measures are not, however, related to ordinary measures in the way fuzzy sets are related to ordinary sets.)

Degrees of belief must, of course, be constructed; they are judgments based on evidence. The degree of belief Bel(A) marks a judgment about how strongly the evidence we are considering points to the correct answer being in A. When we talk about degrees of belief being defined for all subsets of the frame S, we are assuming that this judgmental work is feasible. We are assuming that S is small enough or easily enough manipulated, or that the evidence is simple enough, that it is practical for us to consider the relation of the evidence to all the subsets of S. We are also assuming, when we state condition (iii) for Sugeno measures, that our way of considering these subsets makes the containment relations between them manifest. It is of interest to relax these assumptions and to consider situations where we construct degrees of belief for only some subsets of a frame, but we will not do so in this article.

What does it mean when Bel(A) is greater than zero? Various interpretations are conceivable, but in this article we will suppose that a positive value of Bel(A) represents a judgment that our evidence gives us positive reason to believe A. Lack of evidence against A does not suffice. We distinguish, therefore, between lack of belief and disbelief. A degree of belief of zero in A indicates lack of belief in A, resulting from lack of evidence favoring A. It does not necessarily indicate disbelief in A, such as might result from evidence against A. It does not, that is to say, necessarily indicate any belief in $\bar{A}$, the negation of A.

The simplest kind of Sugeno measure is one that assigns degree of belief zero to every proper subset of a frame S. (It must, of course, assign degree of belief one to the frame itself; this is condition (ii).) Such a *vacuous* Sugeno measure is appropriate for representing evidence that tells us nothing at all about the question considered by S.

A. Extending Sugeno Measures

Suppose we have constructed two frames S and T and a compatibility relation between them, and suppose we have also constructed degrees of belief for both frames, giving us Sugeno measures Bel_S on 2^S and Bel_T on 2^T. Then the same logic that demands condition (iii) also demands that

$$Bel_T(B) \geqslant Bel_S(A) \tag{3}$$

whenever A and B are such that the correct answer to the question considered by S can be in A only if the correct answer to the question considered by T is in B — i.e., whenever A and B are such that

$$s \text{ in } A,\ t \text{ in } T, \text{ and } sCt \Longrightarrow t \text{ in } B. \tag{4}$$

For fixed B, the demand is that Equation 3 should be satisfied for all A satisfying Equation 4; given (iii), this is equivalent to the demand that

$$Bel_T(B) \geqslant Bel_S(\{s \text{ in } S | t \text{ in } T \text{ and } sCt \Longrightarrow t \text{ in } B\}),$$

or

$$Bel_T(B) \geqslant Bel_S(\{s \text{ in } S | G(s) \subseteq B\}), \tag{5}$$

where G is the multivalued mapping from S to 2^T.

Now suppose we have constructed frames S and T, a compatibility relation between them, and degrees of belief for S, forming a Sugeno measure Bel_S, but we have not yet constructed degrees of belief for T, and suppose we make the judgment that the evidence we are considering bears on the question considered by T only indirectly, through its relevance to the question considered by S. (This is analogous to the judgment that S mediates the interaction between T and another frame; here we may say that S mediates the interaction between T and the evidence.) Then it seems reasonable that we should accord to each subset B of T just that degree of belief required by our degrees of belief on S and the compatibility relation between S and T — i.e., we should define Bel_T on 2^T by Equation 5, with $=$ in the place of $\geq$:

$$Bel_T(B) = Bel_S(\{s \text{ in } S | G(s) \subseteq B\}). \quad (6)$$

We may call the function Bel_T defined by Equation 6 the *extension* of Bel_S to T. If T is a coarsening of S, then the extention of Bel_S to T is also called the *marginal* of Bel_S over T. The extension of a Sugeno measure is always a Sugeno measure.

B. Impossible Possibilities

In Section II, I insisted on the idea that a frame is a list of answers to a question, all of which are possible relative to our knowledge and opinion. This means that whenever we learn something that rules out some of these possible answers, we must formally change the frame. It also means that we will not be interested in a Sugeno measure that assigns degree of belief one to a proper subset of a frame, since such a Sugeno measure would indicate complete belief that the answers not in the proper subset are impossible.

It is sometimes convenient, however, to relax this attitude. If we do relax it, we will not need to change our frames so often; instead, we can just change our Sugeno measures. If, for example, we begin with the vacuous Sugeno measure over S and we then acquire new evidence that tells us that the correct answer to the question considered by S is in a proper subset S_0, but tells us nothing more than this, then we need not change our frame from S to S_0; instead we can simply change our Sugeno measure to the Sugeno measure Bel over S that is given by

$$Bel(A) = \begin{cases} 0 & \text{if } S_0 \not\subseteq A \\ 1 & \text{if } S_0 \subseteq A \end{cases}$$

for every subset A of S. This Sugeno measure may be called the *binary Sugeno measure* focused on S_0.

C. Probability Languages Based on Sugeno Measures

We have looked at some very simple Sugeno measures, and we have considered how a Sugeno measure on one frame might be extended to a Sugeno measure on another. However, how can we construct more complicated Sugeno measures? How can we construct a Sugeno measure that gives degrees of belief that are less than one, but greater than zero?

There are two important problems here. First, there is the problem of giving a meaning to the scale. What is meant by a degree of belief of one third, say? More generally, what is meant by the adoption of a whole Sugeno measure? Secondly, there is the problem of locating real evidence on the scale. How do we decide what degree of belief is appropriate for a given subset of a frame in light of given evidence?

In order to give meaning to the scale, we need canonical examples, examples in which we agree that given numerical degrees of belief, and hence a given Sugeno measure, are appropriate, and in order to locate real evidence on the scale, we need ways of relating real evidence to the kinds of evidence and knowledge that are assumed in the canonical examples. When we specify a set of appropriate canonical examples together with methods for matching real evidence to the scale of Sugeno measures determined by these canonical examples, we have, in effect, formulated a particular theory or language for subjective probability judgment.[12,23]

A language for subjective probability judgment formulated in this way typically does not use the whole set of all Sugeno measures. It does not, that is to say, have a canonical example corresponding to every Sugeno measure, and different languages usually use different subsets of the Sugeno measures. It is therefore convenient, though sometimes misleading, to label different languages with the names of the different kinds of of Sugeno measures they use.

The most familiar language of subjective probability judgment is the Bayesian language, which is reviewed in the next section. This language uses *probability measures,* Sugeno measures that have the structure of frequencies. It compares problems to canonical examples in which the correct answer to a question is determined by chance, with the chances completely known.

A Sugeno measure obtained by extending a probability measure is called a *belief function.* As we explain in Section V below, the language of belief functions compares problems to canonical examples in which the meaning and reliability of the evidence depend on known chances.

A Sugeno measure obtained by taking lower bounds over a class of probability measures is called a *lower probability measure*. The language of lower probabilities uses canonical examples in which the answer to a question is determined by chance, with the chances only partially known. Lower probabilities have been studied by Smith,[27] Good,[28] Williams,[29] Walley and Fine,[30] and others. The set of belief functions is actually a subset of the set of lower probability measures, but the two languages use different canonical examples and hence use different rules to manipulate these functions; this point is discussed in detail by Shafer[12] and is touched on in Section V. We will not study the language of lower probabilities in detail in this article.

Are there yet other languages for subjective probability judgment that use Sugeno measures? Other classes of Sugeno measures have been studied, but so far this study does not seem to have led to other languages for probability judgment. Several authors, for example, have studied a class of Sugeno measures called the g_λ measures,[25,26,31] and Dubois[32] has called attention to the class of Sugeno measures obtained by extension of the g_λ measures. However, there is no language of g_λ measures; no one has given canonical examples that would enable us to use these measures in probability judgment. Fine and students[33,34] have recently investigated a class of Sugeno measures that is broader than the class of lower probability measures, but the intention of this work seems to be to generalize probability as a model for objective phenomena rather than to provide another language for subjective judgment.

Possibility measures constitute another subset of the Sugeno measures. However, again, there is not yet a language of possibility measures; we do not yet have appropriate canonical examples. This point will be discussed further in Section VI.

IV. PROBABILITY MEASURES

A Sugeno measure Bel over a frame S is called a *probability measure* if in addition to conditions (i) to (iii) it also satisfies

$$\text{(iv) } \mathrm{Bel}(A \cup B) = \mathrm{Bel}(A) + \mathrm{Bel}(B)$$

for all disjoint subsets A and B of S. Usually the letter P is used instead of Bel to denote a probability measure. When P is a probability measure and S is its frame, the pair (S,P) is called a *probability space*.

Probability measures are important because we are often able to construct them from observed or conjectured frequencies. Sometimes our evidence about the question considered by S consists just of knowledge of the answers to similar questions, and in these cases we may construct a probability measure P over S by setting P(A) equal to the proportion of all answers that are in A. In other cases, our evidence may allow us to make reasonable conjectures about such frequencies. In yet other cases, we may judge that our evidence is similar in strength to the knowledge of such frequencies. In this section we will review some facts about probability measures and some ideas about subjective probability judgment using probability measures that are relevant to our later discussion of belief functions and possibility measures.

A. Canonical Examples

The mathematical picture of chance, which originated in the 17th century study of games of chance, is well known in our culture. It is a picture in which different events happen with different chances. These chances are objective frequencies, and since they are known to us we can also call them our degrees of belief. They define fair betting rates — rates at which we can repeatedly offer to bet and expect to break even in the long run.

In games of chance, and also in modern particle physics, we can find examples that fit the mathematical picture of chance perfectly. When we construct a probability measure to represent more ordinary sorts of evidence, we may say that we are comparing our ordinary evidence to those special, or canonical, examples from gambling and physics. The comparison may not be perfect, but it still provides an argument, sometimes a convincing one.[23] If we make a number of probability judgments by giving numerical probabilities that seem to match in strength the evidence we have for several different assertions and we then combine these numerical probabilities using the rules that follow from the picture of chance, then we may be able to construct a probability measure that yields convincing conclusions that were not obvious at the outset.

When we used observed or conjectured frequencies to construct a probability measure over a frame S, we are locating our evidence on a scale of canonical examples in which the correct answer to the question considered by S is determined by chance. We may also say that in doing so we are using the *Bayesian* language of probability judgment, for this name has come to be associated with subjective probability judgments that directly match the picture of chance.[17]

B. Countable Additivity

It may be appropriate to offer some technical remarks to excuse the lack of any mention of countable additivity in our definition of probability measure. In the mathematical study of probability, it is usually assumed that a probability measure is defined only on an algebra of subsets of a frame S instead of on 2^S, the set of all subsets of S.[35] (An algebra of subsets of a set S is a collection of subsets of S that contains $\emptyset$ and S and is closed under complementation and finite unions and intersections. An algebra that is closed under countable unions and intersections is called a σ-algebra.) This assumption is sound from a constructive viewpoint, since it may not be feasible, especially in the case of complicated infinite sets, to specify values of a probability measure for all subsets.

A probability measure P that is defined on a σ-algebra of subsets of S is said to be *countably additive* if

$$P(\cup A_i) = \Sigma P(A_i)$$

for every sequence $A_1, A_2, \ldots$ of disjoint subsets in the σ-algebra. This condition is equivalent to the requirement that P be *continuous*, in the sense that

$$P(\cap A_i) = \inf P(A_i) \tag{7}$$

for every decreasing sequence $A_1 \supseteq A_2 \supseteq \ldots$ of elements of the σ-algebra.[36] The requirement of countable additivity is associated with the definition of probability measures on σ-algebras because many important countably additive probability measures cannot be extended to countably additive probability measures defined for all subsets of their frames.

Some advocates of the Bayesian language of probability judgment, e.g., de Finetti,[37] have argued against countable additivity and for the assumption that probabilities are at least potentially defined for all subsets of one's frame. I do not agree with this argument, but since the issue of countable additivity is not relevant to the questions studied in this article and since it is tiresome to repeatedly refer to algebras on which probability measures are defined, I will often write as if all probability measures are defined for all subsets of their frames. When convenient, I will even assume that the frames are finite.

C. Product Probability Measures

Given two probability spaces (S_1,P_1) and (S_2,P_2), we may form a product probability space $(S_1 \times S_2, P)$, where P, the *product* of P_1 and P_2, is characterized by the relation

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2) \tag{8}$$

for all subsets A_1 of S_1 and A_2 of S_2. (If S_1 and S_2 are finite, P_1 is defined for all subsets of S_1, and P_2 is defined on all subsets of S_2, then Equation 8 uniquely characterizes a probability measure defined for all subsets of $S_1 \times S_2$. If P_1 and P_2 are defined only on algebras of subsets, then Equation 8 uniquely characterizes a probability measure on the product algebra.)

The use of product probability measures in Bayesian probability judgment involves two judgments of independence: (a) a judgment that S_1 and S_2 are independent, so that $S_1 \times S_2$ is a frame, and (b) a judgment that the evidence on which the probability measure P_1 is based is independent of the evidence on which the probability measure P_2 is based, so that the probabilities given by $P_1 \times P_2$ may be used as degrees of belief on the frame $S_1 \times S_2$. After judging that our evidence about the question considered by S_1 is similar in import to knowledge that the answer has been drawn at random from a certain frequency distribution and after making a similar judgment about the question considered by S_2, we further judge that our evidence about the one question is so unrelated to our evidence about the other that answering both is similar to making these two drawings independently.

D. Conditional Probability

The 17th century theory of games of chance, from which the mathematical theory of probability derives, considered games that unfold step by step. The rules of such a game determine a *protocol* — a specification, at each step, of what may happen next. Since each player keeps abreast of events, his expectations and probabilities about the eventual outcome of the game change as the game unfolds in accordance with this protocol. Since the game is a closed system, these expectations and probabilities depend only on what has happened so far in the game. For mathematical treatments of the idea of a protocol, see Shafer.[38-40]

Suppose the frame U considers the question of what the eventual outcome of the game will be; each element u of U completely describes one possible way the game might unfold. Let P denote the initial probability measure on U. Suppose that at a certain point in the game, only the outcomes in the subset U_0 of U remain possible. This subset amounts to a specification of everything that has happened so far in the game and therefore, according to the preceding paragraph, fully determines how the probabilities given by P have changed. So the phrase "the probability for a subset A of U at the point when only the outcomes in U_0 remain possible" labels a unique number. Let us denote this probability by $P(A|U_0)$. Using the frequency or betting interpretation of the probabilities involved, we can demonstrate that[40]

$$P(A \cap U_0) = P(U_0)P(A|U_0). \tag{9}$$

The demonstration depends on the existence of the protocol and requires that U_0 be specified by the protocol as one of the possibilities for the set of still possible outcomes at some point in the game. We may call a subset U_0 of U an *exact event* if the protocol specifies U_0 as one of these possibilities; in general not all subsets of U will be exact events. When U_0 is an exact event, so that $P(A|U_0)$ is well defined, this number may be called the *conditional probability* of a given U_0. Formula 9 is called the *rule of compound probability*.

Suppose $U = S \times T$, and suppose the protocol specifies, for each s in S, a subset B(s) of T such that (1) $\{s\} \times B(s)$ is an exact event and (2) if B is any other subset of T such that $\{s\} \times B$ is an exact event, then $B = B(s)$. (This means that at the point in the unfolding of the game where we first find out that s is the answer to the question considered by S, B(s) will constitute our knowledge about the question considered by T.) Notice that for each s in S, $\{s\} \times \overline{B(s)}$ is impossible, so that $P(\{s\} \times B(s)) = P(\{s\} \times T) = P_S(\{s\})$. For each s in S and each subset A of T, let us denote $P(S \times A|\{s\} \times B(s))$ by $P_{T|s}(A)$. The set function $P_{T|s}$ defined in this way is a probability measure over T; it is called the *conditional* over T of P given s.

If t is in B(s), then substitution of $S \times \{t\}$ for A and $\{s\} \times B(s)$ for U_0 in Equation 9 yields

$$P(\{(s,t)\}) = P_S(\{s\})P_{T|s}(\{t\}). \tag{10}$$

This formula also holds if t is not in B(s), for then both sides are zero. If we suppose that $S \times T$ is finite, then P is completely determined by its values on the points $\{(s,t)\}$. So in this case at least, P can be completely reconstructed from knowledge of its marginal P_S and its conditionals $P_{T|s}$.

It follows from Equation 9 that

$$P(A|U_0) = \frac{P(A \cap U_0)}{P(U_0)} \tag{11}$$

This formula can serve as the symbolic expression of a rule. This rule, *the Bayesian rule of conditioning*, says that when new knowledge or evidence tells us that the correct answer to the question considered by U is in the subset U_0, we should change our probability for another subset A from P(A) to the new probability given by this quotient. Strictly speaking, this rule is valid only when the receipt of the information U_0 is itself a part of our probability model — i.e., was foreseen by the protocol. In practice, however, the rule is used more broadly than this in subjective probability judgment. This broader use is reasonable, but it must be emphasized that the calculation is only an argument — an argument that is imperfect because it involves a comparison of the actual evidential situation with a situation where U_0 was specified by a protocol.

V. BELIEF FUNCTIONS

A Sugeno measure Bel on 2^T is called a *belief function* over T if it satisfies

$$\text{(v) } Bel(B_1 \cup \ldots \cup B_n) \geq \sum_i Bel(B_i) - \sum_{i,j} Bel(B_i \cap B_j) + \sum_{i,j,k} Bel(B_i \cap B_j \cap B_k) - + \ldots$$

for any finite collection $B_1, \ldots, B_n$ of subsets of T. (It is also possible to consider belief functions defined only for some subsets of a frame.[11])

It can be shown[11,41] that a function Bel_T on 2^T is a belief function if and only if there exists a frame S, a compatibility relation between S and T, and a probability measure Bel_S on 2^S that is related to Bel_T by Equation 6. In other words, belief functions are the Sugeno measures that can be obtained by extending probability measures.

Since a probability measure is, in a trivial sense, an extension of itself, all probability measures are belief functions. It can also be shown that vacuous Sugeno measures and, more generally, binary Sugeno measures are belief functions.

Given a belief function Bel on 2^S, we may define another real-valued function Pl on 2^S by

$$Pl(A) = 1 - Bel(\bar{A}).$$

This function is called the *plausibility function* associated with Bel. A belief function can be recovered from its plausibility function: $Bel(A) = 1 - Pl(\bar{A})$. We always have $Bel(A) \leq Pl(A)$, but Bel and Pl are equal only if Bel is a probability measure. While Bel(A) is our judgment of the extent to which the evidence we are considering supports A, Pl(A) is our judgment of the extent to which this evidence falls short of refuting A. If we judge that the evidence provides neither support nor refutation for A, we may set $Bel(A) = 0$ and $Pl(A) = 1$.

(*Note:* Mathematically, a plausibility function is a Sugeno measure. We do not, however, interpret plausibility functions in the way that we proposed to interpret Sugeno measures in Section III above. A positive value of Pl(A) does not necessarily indicate a judgment that there is evidence favoring A.)

Any extension of a belief function is also a belief function. To see this, note that if Bel_T is the extension of Bel_S from S to T and Bel_U is the extension of Bel_T from T to U, then Bel_U is the extension of Bel_S from S to U. This can be verified using Equation 6; if the compatibility relations C(S,T), C(T,U), and $C_T(S,U)$ are represented by multivalued mappings G_1, G_2, and G, as in Equation 2 above, then

$$\begin{aligned} Bel_U(C) &= Bel_T(\{t \text{ in } T | G_2(t) \subseteq C\}) \\ &= Bel_S(\{s \text{ in } S | G_1(s) \subseteq \{t \text{ in } T | G_2(t) \subseteq C\}\}) \\ &= Bel_S(\{s \text{ in } S | \cup\{G_2(t) | t \text{ in } G_1(s)\} \subseteq C\}) \\ &= Bel_S(\{s \text{ in } S | G(s) \subseteq C\}) \end{aligned}$$

for any subset C of U (compare Section 4 of Dubois and Prade[42]). If Bel_T is a belief function, then we may choose S and Bel_S so that Bel_S is a probability measure; this means that Bel_U is also an extension of a probability measure and hence also a belief function.

We say that one belief function Bel_1 over a frame T is *stronger* than another belief function Bel_2 over the same frame if $Bel_1(B) \geq Bel_2(B)$ for all subsets B of T. This is equivalent to the requirement that $Pl_1(B) \leq Pl_2(B)$ for all subsets B of T.

A. Canonical Examples

If we construct a probability measure using observed or conjectured frequencies and then extend this probability measure through a compatibility relation, we obtain a belief function. However, when is such extension appropriate?

In order to deepen our understanding, we need canonical examples — stories about situations where we have a probability measure P over a frame S and a compatibility relation between S and another frame T and where we can make the judgment that S mediates the interaction between T and the evidence on which P is based, thereby justifying the degree of belief

$$\mathrm{Bel}_T(B) = P\{s \text{ in } S|G(s) \subseteq B\} \tag{12}$$

for each subset B of T.

Here are three stories in which the extension of a probability measure to a belief function seems reasonable. As we shall see, the second and third of these stories leads to the language that I have called the "language of belief functions". The first story provides a basis for some of the rules of that language, including the rule for forming product belief functions, but it does not provide a basis for Dempster's rule of combination.

1. Consider two frames S and T and a probability measure P over S. Suppose the answers to the questions considered by S and T are both initially undetermined. However, we know that s, the answer to the question considered by S, is to be determined by a random selection from S, this selection being governed by the probability distribution P. We also know that this determination of s partially determines t, the answer to the question considered by T; it forces t to be in a certain subset of T, say G(s). The further determination of t is to be accomplished in some way that is not known to us. We are assuming that our only evidence about s is our knowledge of P and our knowledge that s is selected at random according to P. It is reasonable, therefore, to adopt the probabilities given by P as our degrees of belief about s. We are also assuming that our evidence tells us nothing further about t — i.e., that S mediates the interaction between this evidence and T. We therefore adopt the belief function over T given by Equation 12. Here the probability measure P over S is causally as well as epistemically relevant to T; the answer to the question considered by S is the partial cause of the answer to the question considered by T.
2. Wierzchon's story:[31] Consider two frames S^0 and T^0 and a probability measure P^0 over S^0. Initially the answer to the question considered by T^0 is already determined, but the answer to the question considered by S^0 is not. Each element s of S^0 is actually a mapping from T^0 to another set Q. It is arranged that s is to be selected at random from S^0, this selection being governed by the probability distribution P^0. We will not be told the element s that is selected, but we will be told the value of s(t), where t is the correct answer to the question considered by T^0. Suppose this value turns out to be q. Then we know that s could have been selected and t can be the correct answer to the question considered by T^0 only if s(t) = q. So we may replace the frames S^0 and T^0 by S and T, respectively, where

$$S = \{s \text{ in } S^0|\text{for some } t \text{ in } T^0\ s(t) = q\},$$

and

$$T = \{t \text{ in } T^0|\text{for some } s \text{ in } S^0\ s(t) = q\}.$$

Our knowledge that the mapping s actually selected is in the subset S of S^0 justifies changing our degrees of belief about this answer by conditioning the probability measure P^0 on S; we may denote the resulting probability measure over S by P. Our knowledge also establishes a compatibility relation between the frames S and T:

$$C(S,T) = \{(s,t)|s(t) = q\}.$$

It is clear that our evidence about S bears on T only through this compatibility relation — i.e., S mediates the interaction between this evidence and T. We can make this story more concrete by thinking of q as the result of measuring t; s describes the random influences that determine how the measurement will err. There is a causal element in this story; t is one of the causes of s. However, the reverse relationship, from s to t, is not causal. Here, in contrast to the preceding example, the probability measure P over S has purely epistemic relevance to T.

3. Shafer's story:[12] Consider again frames S^0 and T^0 and a probability measure P^0 over S^0. However, now suppose that each element s of S^0 maps subsets of T^0 to elements of a set Q. Again, t, the correct answer to the question considered by T^0, has already been determined. Moreover, someone has singled out a subset M of T^0 that contains t. An element s of S^0 will be selected at random, the selection being governed by P^0. We will not be told t, M, or s, but we will be told the value of s(M). Suppose this value turns out to be q. Then we know that the correct t and the s that was selected must be such that s maps some set containing t to q. So we may replace the frames S^0 and T^0 by

$$S = \{s \text{ in } S^0|s(M) = q \text{ for some subset } M \text{ of } T^0\},$$

and

$$T = \cup\{M|s(M) = q \text{ for some } s \text{ in } S^0\}.$$

We may change our degrees of belief for S to those given by P, the result of conditioning P^0 on S, and we may use the compatibility relation

$$C(S,T) = \{(s,t)|s(M) = q \text{ for some } M \text{ containing } t\}$$

Again, it is clear that the evidence represented by P bears on T only through this compatibility relation — i.e., S mediates the interaction between this evidence and T. We can make this story more concrete by thinking of M as a message that is conveyed to us in coded form. The code s is selected at random. After we receive the encoded version q, we try to decode it using all the possible codes. The results may rule out some of the codes in S^0, but they may leave a number of possibilities for which code s was used, what plaintext message M was sent, and which answer t is correct. Our knowledge about how s was selected bears on the identity of M and t only through what it tells us about which s was selected.

In each of these stories, the knowledge on which the probabilities for s are based provides evidence about t only through the compatibility relation between S and T. The value of s has a causal bearing on the value of t in the first story, but a purely epistemic bearing in the second and third stories. In the first story, Equation 12 is interpreted as the probability that the selection of s will force t into B, but in the other two stories it is interpreted as the probability that the evidence (the measurement or the encoded message) means or implies that t is in B.

(*Note:* The second and third stories seem very similar. How are they related? The simplest way to relate them is to make explicit, in the third story, a frame U^0 that considers the question: Which subset M of T^0 has been singled out? The frames S^0 and U^0 are related in the third story just as the frames S^0 and T^0 are related in the second story. However, in the third story there is a further step, the extension of the belief function over $U = \{M|s(M) = q \text{ for some } s \text{ in } S^0\}$ to a belief function over T, which is the union of the sets in U. The compatibility relation between U and T is given by saying that MCt if and only if t is an element of M.)

Both the language of belief functions and the Bayesian language use observed or conjectured frequencies to construct probability judgments. However, whereas the Bayesian language uses canonical examples in which the correct answer to the question we are considering is directly determined by chance, the language of belief functions uses canonical examples in which only the meaning of the evidence bearing on this question depends on chance.

One consequence of this difference is the fact that it may be impossible to interpret the numerical degrees of belief given by a belief function over a frame T as hypothetical frequencies. This may be true even if these degrees of belief are derived from degrees of belief over a frame S that can be interpreted as hypothetical frequencies. The reason for this is that the compatibility relation between the frames S and T, since it is constructed for the particular problem, may not be constant over the set of hypothetical repetitions that we must consider in order to think of the degrees of belief on S as frequencies.

We can see this point more clearly if we look at the possibility of a hypothetical frequency interpretation in each of our three stories or canonical examples for belief functions. Such an interpretation may be possible in the first story, but not in the second two.

In the first story we can construct a set of hypothetical repetitions by imagining that elements of S are selected repeatedly in accordance with the chances given by P. Following each selection from S a selection from T is also made; if s is selected from S, then an element t is selected from G(s) according to some rules that are unknown to us. The degree of belief $Bel_T(B)$ given by Equation 12 has a frequency interpretation in terms of these hypothetical repetitions; we can say that the element of T selected will be in B with at least this frequency.

In the second and the third stories we begin with a probability measure not on the frame S, but on a frame S^0. We may imagine repeated selections of an element s from S^0, but it is not clear how the other elements of the story are to be varied as this selection is repeated. In the first story, the determination of s partially determined t, but here the selection of s is not supposed to affect t. It is also not supposed, in the third story, to affect the selection of M. This suggests that we imagine repeated selections of s with t (or both t and M) held constant, but doing so will not produce anything interesting; when t is held constant the frequency with which it will be in a fixed subset B of T will be either zero or one. It should also be noted that the conditioning of P^0 on S does not inherit a frequency interpretation from the imagined repeated selections of s from S^0. As I emphasized in Section IV.D above, the conditioning of probabilities can be given a frequency justification only if the process by which the information conditioned on is obtained is itself built into the probability model. Here the selection of t (or of t and M) and hence the process by which S^0 and T^0 are reduced to S and T are not part of the probability model.

Some readers will feel that the lack of a thorough frequency interpretation for our second and third stories for belief functions makes these stories unconvincing. I believe, however, that this attitude depends on a too uncritical acceptance of the relevance of imagined repetitions. To the extent that we recognize that the choice of a set of hypothetical repetitions is largely arbitrary, we will be more interested in this choice as a device for calibrating the strength of our evidence rather than as a basis for claiming objective validity for our judgments. I have already mentioned that the causal canonical example for belief functions (the first of our three stories) does not provide a basis for Dempster's rule of combination, while the purely epistemic canonical examples (the second and third stories) do provide such a basis (see Section V.C). This fact, together with the fact that only the first example leads to a frequency or betting interpretation for the degrees of belief given by a belief function, has led some authors[43-45] to claim that Dempster's rule is incorrect. It is more appropriate, however, simply to say that there is no hypothetical frequency or betting interpretation for final degrees of belief in the language of belief functions.

It is also enlightening to note that the first of our three canonical examples, the causal one, can be interpreted as a canonical example for the language of lower probabilities (see Section III.C above). Indeed, if we imagine that the further determination of t within G(s) is effected according to some unknown probabilities, then our knowledge of P over S amounts to partial knowledge of a probability measure that is used to determine the answer to the question considered by T. It follows from this observation that belief functions, *mathematically*, are lower probability functions. As we have already remarked, however, the *language* of belief functions, since it uses our purely epistemic canonical examples, is not a subset of the language of lower probabilities.

(*Note:* It may be useful to sketch the relation between the preceding statement of canonical examples for belief functions and the approaches to belief functions taken in earlier expositions. Dempster[2] introduced belief functions (or "lower probabilities", as he called them then) in terms of multivalued mappings. He did not give canonical examples to show how the multivalued mappings were to be interpreted, and many of his readers seem to have interpreted them in the causal sense spelled out in our first story above. Shafer,[8] in an attempt to avoid the confusion arising from this causal interpretation, left aside the multivalued mapping altogether and treated belief functions in purely axiomatic terms. This axiomatic approach did escape from the causal interpretation, but it did not solve the problem of providing an alternative interpretation. Shafer[12] introduced the canonical examples involving randomly coded messages in order to provide such an alternative interpretation. The present exposition returns to the level of abstraction of the multivalued mapping, but emphasizes the epistemic interpretation of "compatibility".)

The canonical example of the randomly coded message is very general, but it is also distant from our everyday experience, and it therefore seems very abstract. So it may be useful to conclude this section with a discussion of a simple but more concrete example, the example of testimony.

Suppose that a witness testifies that the answer to the question considered by the frame T is in the subset T_0. Suppose that we have no evidence about this question other than this testimony (or at least that we want to leave our other evidence aside for the moment) and that our only evidence about the reliability of the witness comes from experience that tells us that witnesses such as this one are usually reliable (say 80% of the time), are rarely deceitful (say 2% of the time), but fairly often do not know what they are talking about (the rest, or 18%, of the time).

The frequencies 80, 2, and 18% clearly do not bear directly on the frame T. Instead they bear directly on a frame S that considers the reliability of the witness: $S = \{a,b,c\}$, where a is the possibility that the witness is reliable, b is the possibility he is deceitful, and c is the possibility he does not know what he is talking about. These frequencies amount to a probability measure P over S: $P(\{a\}) = 0.8$, $P(\{b\}) = 0.02$, and $P(\{c\}) = 0.18$.

The compatibility relation between S and T is simple. We have aCt if and only if t is in T_0, bCt if and only if t is not in T_0, and cCt for all t in T. When we extend P to T using this compatibility relation, we get the belief function Bel_T given by

$$Bel_T(B) = \begin{cases} 0.80 & \text{if } T_0 \subseteq B \neq T \\ 0.02 & \text{if } \overline{T}_0 \subseteq B \neq T \\ 0.00 & \text{if } T_0 \not\subseteq B \text{ and } \overline{T}_0 \not\subseteq B \\ 1.00 & \text{if } B = T \end{cases}$$

In particular, $Bel_T(T_0) = 0.8$ and $Bel_0(\overline{T}_0) = 0.02$.

B. Product Belief Functions

Consider now two belief functions Bel_1 and Bel_2, defined over frames T_1 and T_2, respectively. It can be shown[8,11] that there is a unique belief function Bel over $T_1 \times T_2$ that satisfies both

$$Bel(A_1 x A_2) = Bel_1(A_1)Bel_2(A_2) \tag{13}$$

and

$$Pl(A_1 x A_2) = Pl_1(A_1)Pl_2(A_2) \tag{14}$$

for all subsets A_1 of T_1 and A_2 of T_2. (Again, Bel is uniquely determined for all subsets of $T_1 \times T_2$ if T_1 and T_2 are finite and Bel_1 and Bel_2 are defined for all subsets, and in general Bel is uniquely determined on the product of the algebras on which the two belief functions are defined.) This belief function is called the *product* of Bel_1 and Bel_2. It can be constructed as follows. For $i = 1$ and 2, let S_i be a frame and let P_i be a probability measure on S_i, and suppose there is a compatibility relation between S_i and T_i such that Bel_i is the extension of P_i to T_i. Define a compatibility relation between the Cartesian products $S_1 \times S_2$ and $T_1 \times T_2$ by saying that

$$(s_1,s_2)C(t_1,t_2) \text{ if and only if both } s_1Ct_1 \text{ and } s_2Ct_2 \tag{15}$$

Bel is the extension of $P_1 \times P_2$ to $T_1 \times T_2$ using this compatibility relation.[10]

What judgments are involved when we adopt the product belief function Bel? In addition to (a) and (b), listed in Section IV.C above, we appear to need the following:

(c) A judgment that T_1 and T_2 are independent, so that $T_1 \times T_2$ is a frame.
(d) A judgment that T_i mediates the interaction between S_i and $T_1 \times T_2$, for $i = 1,2$.
(e) A judgment that $T_1 \times T_2$ discerns the relevant interaction between S_1 and S_2.
(f) A judgment that $S_1 \times S_2$ mediates the interaction between the evidence on which $P_1 \times P_2$ is based and the frame $T_1 \times T_2$.

Judgments (d) and (e) assure that Equation 15 is the compatibility relation between $S_1 \times S_2$ and $T_1 \times T_2$. Judgment (f) then authorizes the extension of $P_1 \times P_2$ to $T_1 \times T_2$ using this compatibility relation.

It is instructive to run through these judgments for the construction of product belief functions in the context of each of our three stories, or canonical examples, for belief functions. In the first story, we would be willing to make these judgments if s_1 and s_2 were independently selected at random and partially but independently determined t_1 and t_2. In the second story, we would make them if we had independent measurements of independent quantities. In the third story, we would make them if we had independent and independently coded messages about independent questions. The judgments make sense for all three stories. Thus, both causal and purely epistemic conceptions can serve as a basis for forming product belief functions.

It should be mentioned that the rule for forming product belief functions can be described easily in terms of multivalued mappings. If Bel_1 is represented by the multivalued mapping $G_1:S_1 \rightarrow 2^{T_1}$ and Bel_2 is represented by the multivalued mapping $G_2:S_2 \rightarrow 2^{T_2}$, then their product can be represented by the multivalued mapping $G:S_1 \times S_2 \rightarrow 2^{T_1} \times 2^{T_2}$ given by $G(s_1,s_2) = G_1(s_1) \times G_2(s_2)$.

C. Dempster's Rules of Conditioning and Combination

Up to this point, our stories about constructing Bel on 2^T from P on 2^S have included, at least implicitly, the idea that we have no evidence bearing on T except for the evidence on which P is based. This is not realistic in general, but often whatever other evidence we do have is conceptually distinct, and it may be possible to represent it separately, either by a separate belief function or by some other means. This raises the problem of combining representations of different bodies of evidence to obtain probability judgments based on the total evidence.

The simplest case is one in which one body of evidence about T is well represented by a compatibility relation between T and a probability space (S,P) and by the resulting belief function Bel, while another body of evidence bears on T only inasmuch as it tells us with certainty that the correct answer is in a proper subset T_0 of T. In this case, the second body of evidence is adequately taken into account by reducing T to T_0. However, this change has repercussions for the compatibility relation with (S,P) and possibly for (S,P) itself. Indeed, by telling us that the correct answer to the question considered by T is in T_0, this second body of evidence also tells us that the correct answer to the question considered by S is in

$$S_0 = \{s \text{ in } S | sCt \text{ for some } t \text{ in } T_0\}$$

This suggests that we should change our degrees of belief for the answer to the question considered by S by conditioning P on S_0, thus obtaining a probability measure, say P_0, over S_0. If the reduction of T and S to T_0 and S_0 does not affect the compatibility of the elements of T and S that remain, then the compatibility relation between S_0 and T_0 is $(S_0 \times T_0) \cap C(S,T)$. So we can form a new belief function Bel′ over T_0 by extending P_0 from S_0 to T_0. This belief function is related to Bel by

$$Bel'(B) = \frac{Bel(B \cup (T - T_0)) - Bel(T - T_0)}{1 - Bel(T - T_0)}, \tag{16}$$

and the associated plausibility function, say Pl′, is related to the plausibility function Pl associated with Bel by

$$Pl'(B) = \frac{Pl(B \cap T_0)}{Pl(T_0)}. \tag{17}$$

The belief function Bel′ is called the result of *conditioning* Bel on T_0, and the process by which Bel′ is defined is called *Dempster's rule of conditioning*. It is clear from Equation 17 that this rule is a generalization of the Bayesian rule of conditioning for probability measures.

The conditional belief function Bel′ can be regarded either as a belief function over the reduced frame T_0 or as a belief function over the original frame T. In the latter case, we have, of course, some possibilities in the frame that are declared impossible by the belief function (see Section III.B above). In either case, Equations 16 and 17 give the values of Bel′ and Pl′ for all subsets of the frame.

Conditioning is readily understood in terms of the purely epistemic stories for belief functions. Indeed, conditioning to reduce T^0 to T is already built into those stories, and so further conditioning introduces no novelty. Conditioning does introduce a new element into the causal story, however. This new element is significant because it eliminates the possibility of a hypothetical frequency or betting interpretation of the degrees of belief given by the belief function.

Another important case arises when two bodies of evidence are both represented by belief functions, say Bel_1 and Bel_2, over a frame T. Suppose Bel_1 is the extension to T of a probability measure P_1 over a frame S_1 and Bel_2 is the extension to T of a probability measure P_2 over a frame S_2. Then we can construct a single belief function from these two by the following device. First, we think of Bel_1 and Bel_2 as being defined over distinct copies of T, and we form their product, which is a belief function over the Cartesian product $T \times T$. Then we remember, as it were, that the two copies of T consider the same question, and we take this into account by conditioning the product belief function on the diagonal of $T \times T$, which can be identified with T itself. The resulting belief function is called the *orthogonal sum* of Bel_1 and Bel_2 and the rule for forming it is called *Dempster's rule of combination*.

Dempster's rule can also be described in terms of the multivalued mappings, say $G_1{:}S_1 \rightarrow 2^T$ and $G_2{:}S_2 \rightarrow 2^T$, representing Bel_1 and Bel_2. We form a third multivalued mapping G from $S_1 \times S_2$ to 2^T by setting

$$G(s_1,s_2) = G_1(s_1) \cap G_2(s_2);$$

we restrict this multivalued mapping to the subset of $S_1 \times S_2$ consisting of all elements that it does not map to the empty set; and we condition the product probability measure $P_1 \times P_2$ on this subset.

It is evident from this description in terms of multivalued mappings that Dempster's rule does not make any sense if we use the causal story as our canonical example for belief functions. The problem is that the interpretation of G_1 and G_2 both involve the partial determination of t, and it does not make sense, in general, to partially determine t twice. At least we cannot determine that t is in $G_1(s_1)$ and then determine that it is in $G_2(s_2)$ if these two sets are disjoint. The rule directs us to eliminate such pairs (s_1,s_2) by conditioning. Such conditioning could be interpreted as a direction to repeat the joint selection of s_1 and s_2 if such an incompatible pair is obtained, but it is difficult to make sense of this repetition in the context of the combination of evidence.

Dempster's rule does make sense, however, in the context of the purely epistemic canonical examples. In the story of the randomly coded messages, for example, we are simply dealing with two independent and independently coded messages that deal with the same question. Dempster's rule amounts simply to treating the two messages as a joint message.[12]

What judgments are involved when we adopt the degrees of belief given by the orthogonal sum of Bel_1 and Bel_2? We can answer this question by looking at the six judgments that are involved in forming a product belief function. Of these six, we can drop (c) and (d) when we drop the temporary pretence that the two copies of T are concerned with distinct questions. And we must change (e) and (f) to refer to T instead $T_1 \times T_2$. This results in the following list:

(a′) A judgment that S_1 and S_2 are initially independent, so that $S_1 \times S_2$ is a frame.

(b′) A judgment that the evidence on which the probability measure P_1 is based is independent of the evidence on which the probability measure P_2 is based, so that the probabilities given by $P_1 \times P_2$ may be used as degrees of belief on the frame $S_1 \times S_2$.
(c′) A judgment that T discerns the relevant interaction between S_1 and S_2.
(d′) A judgment that $S_1 \times S_2$ mediates the interaction between the evidence on which $P_1 \times P_2$ is based and the frame T.

For a simple concrete example of Dempster's rule, let us consider the situation where two independent witnesses both testify to the same fact. Suppose, indeed, that they both testify that the answer to the question considered by T is in the subset T_0, and suppose, for simplicity, that our experience tells us that witnesses like these two are never deceitful; witnesses like the first are reliable 80% of the time and do not know what they are talking about 20% of the time, while witnesses like the second are reliable 90% of the time and do not know what they are talking about 10% of the time.

We may set $S_1 = \{a_1,c_1\}$ and $S_2 = \{a_2,c_2\}$, where a_i is the possibility that witness i is reliable, and c_i is the possibility that witness i does not know what he is talking about. We have $P_1(\{a_1\})=0.8$, $P_1(\{c_1\})=0.2$, $P_2(\{a_2\})=0.9$, and $P_2(\{c_2\})=0.1$. Both a_1 and a_2 are compatible only with t in T_0, while c_1 and c_2 are compatible with all t in T. So $Bel_1(T_0)=0.8$, while $Bel_2(T_0)=0.9$.

Judgments (a′) to (d′) are straightforward. Judgments (a′) and (b′) say that the two witnesses are independent; they were selected independently from their respective populations of similar witnesses, and, if they behave differently from case to case, then they are making their choices independently. It is this assumption of independence that is most likely to be questioned. Judgment (c′) says that the concurrence of the witnesses does not rule out any of the elements of T besides the ones in $\overline{T}_0$, the ones that can be ruled out by considering the testimony of the two witnesses separately. Judgment (d′) says that the experience upon which we based our assessment of the witnesses' reliability does not tell us anything more directly about T.

The calculations for Dempster's rule are also straightforward. The element (c_1,c_2) of $S_1 \times S_2$ is compatible with all t in T, while the other three elements are compatible only with t in T_0. So

$$\begin{aligned} Bel(T_0) &= P(\{(a_1,a_2), (a_1,c_2), (c_1,a_2)\}) \\ &= 0.72 + 0.08 + 0.18 \\ &= 0.98. \end{aligned}$$

Thus, the concurrence of the two witnesses warrants a much higher degree of belief in T_0 than was warranted by the testimony of either witness alone.

This simple rule for the combination of testimony is actually quite old. It was formulated by Hooper[46] in his 1689 tract against the infallibility of the Pope. It was rediscovered in the 1960s by Ekelöf.[47] There are now many other numerical examples of Dempster's rule in the literature.[4,5,14-19,48-51]

D. Conditioning and Products as Special Cases

We have just described Dempster's rule of combination as a composition of the rule for forming product belief functions and Dempster's rule of conditioning. It is also true, however, that these latter rules can be thought of as special cases of the rule of combination.

It is natural to think of Dempster's rule of conditioning as a special case of the rule of combination, because evidence that affects a frame T only by telling us that the correct answer to the question it considers is in a subset T_0 is naturally represented by a binary belief function focused on T_0, and combining this binary belief function with the belief function Bel over T results in the belief function Bel′ over T whose values are given by Equation 16, our formula for conditioning (see page 67 of Shafer[8]). The rule for product belief functions can also be described in terms of Dempster's rule of combination: given Bel_1 over T_1 and Bel_2 over T_2, extend them both to belief functions over $T_1 \times T_2$, and then form their orthogonal sum.[10]

E. Marginals and Conditionals

As we recalled in Section IV.D, a probability measure over $S \times T$ is uniquely determined by its marginal over S and its conditionals over T given the various values of s. This is not true in general for belief functions. Thus, we cannot expect to reconstruct a belief function from knowledge of a marginal and corresponding conditionals. We can, however, construct a belief function on a frame $S \times T$ by combining two belief functions, one of which represents evidence about S and the other of which represents evidence that relates T to S.

Let us call a belief function Bel over a joint frame $S \times T$ *relational* if its marginals for both S and T are vacuous. The evidence represented by such a belief function does not tell us anything directly about S or about T, but it may create some relation between the two. (Indeed, if the relational belief function is binary, then it amounts to a compatibility relation.) Suppose that in addition to a relational belief function Bel over $S \times T$, we also have a belief function Bel_0 over S, and suppose these two belief functions are based on distinct bodies of evidence. The belief function Bel_0 can be extended to a belief function over $S \times T$, so we actually have two belief functions over $S \times T$. We can combine these by Dempster's rule provided we accept the judgments listed in Section V.C above.

This idea of combining a relational belief function with a marginal belief function does generalize the idea of constructing a joint probability measure from a marginal and conditionals. This is because for any set of conditionals over T given the various values of s, there exists a (nonunique) relational belief function over $S \times T$ that has these conditionals, and combining this belief function with the marginal probability measure over S produces the joint probability measure over $S \times T$. For details and examples of relational belief functions, see Shafer.[16]

F. Condensability

The condition of continuity can be applied directly to belief functions in the form in which it is stated for probability measures in Section IV.B above (Equation 7). It turns out, however, that a much stronger condition is of interest in the language of belief functions — the condition that

$$\mathrm{Bel}(\cap\{A|A \text{ in } R\}) = \inf\{\mathrm{Bel}(A)|A \text{ in } R\}$$

for every collection R of subsets of S that has the property that if A and B are in R, then there is an element of R that is contained in $A \cap B$. If Bel satisfies this condition, we say that Bel is *condensable*.[7,11,52]

It is easy to show that a belief function Bel is condensable if and only if the plausibility it assigns to a subset A is always the supremum of the plausibilities it assigns to finite subsets of A. This is a very strong condition; it cannot be satisfied, for example, by a probability measure that is not discrete. It is satisfied, however, by all belief functions over infinite frames that were studied by Dempster.[2,4-6] Dempster's rule of combination can be defined in such a way that it preserves condensability.[10]

G. Consonant Belief Functions

Consider again a belief function Bel that is the extension to a frame T of a probability measure P over a frame S. And suppose the multivalued mapping G from S to 2^T is such that the subsets G(s) of T are nested — for each pair s_1 and s_2 of elements of S, $G(s_1)$ either is contained in or else contains $G(s_2)$. Then the belief function Bel is called *consonant*. A belief function Bel is consonant if and only if

$$Bel(A \cap B) = \min\{Bel(A), Bel(B)\} \tag{18}$$

for every pair of subsets A and B of T (see page 220 of Shafer[8]).

A consonant belief function represents evidence that points in a single direction, but leaves us uncertain about how far to go in that direction. Such evidence may mean that the correct answer is in one fairly broad subset of T, or it may mean that the correct answer is in a more specific (smaller) subset, etc.

Condition 18 is equivalent to the condition that the plausibility function Pl associated with Bel satisfy

$$Pl(A \cup B) = \max\{Pl(A), Pl(B)\}.$$

If the belief function Bel is condensable, then this condition is in turn equivalent to the condition that

$$Pl(A) = \sup\{Pl(\{t\}) | t \text{ in } A\}. \tag{19}$$

In the following, I will assume that the consonant belief functions we are studying are condensable, and I will take Equation 19 as the working definition of consonance.

Given a consonant belief function Bel over a frame T, I will write f(t) for Pl({t}), and I will call f the *contour function* for Bel (see page 221 of Shafer[8]). Since Pl(T) = 1, the contour function must satisfy

$$\sup\{f(t)\} | t \text{ in } T\} = 1. \tag{20}$$

This is the only restriction on contour functions; any function f that maps T to the interval [0,1] and satisfies Equation 20 is the contour function for some consonant belief function.

Consonant belief functions are much simpler than many other belief functions. One way to see this is to note that all the information in a consonant belief function is actually contained in its contour function, which is a point function rather than a set function and hence much less complicated than a set function might be.

A consonant belief function is stronger than any other belief function that assigns the same plausibilities to singletons; if Bel_1 is consonant, and $Pl_1(\{t\}) = Pl_2(\{t\})$ for all t, then

$$Pl_1(A) = \sup\{Pl_2(\{t\}) | t \text{ in } A\} \leq Pl_2(A)$$

for all A, and hence $Bel_1(A) \geq Bel_2(A)$ for all A. One consonant belief function is stronger than another if and only if its contour function is smaller; if Bel_1 and Bel_2 are consonant, then $Bel_1(A) \geq Bel_2(A)$ for all A if and only if $f_1(t) \leq f_2(t)$ for all t.

Any extension of a consonant belief function is itself consonant. To see this, we need only note that the G(s) given by Equation 2 are nested whenever the $G_1(s)$ are.

Extension can be expressed in terms of contour functions. If Bel_U is a belief function over U and Bel_V is its extension to V, then

$$f_V(v) = \sup\{f_U(u) | uCv\}. \tag{21}$$

We may note two special cases of Equation 21. First, if f is the contour function for a consonant belief function over $S \times T$, then the contour function for the marginal over S is given by

$$f_S(s) = \sup\{f(s,t)|t \text{ in } T\}. \tag{22}$$

Second, if f is the contour function for a consonant belief function over S, then the contour function for the extension to $S \times T$ is given by

$$f_{S \times T}(s,t) = f(s). \tag{23}$$

Conditioning preserves consonance; conditioning on T_0 amounts to intersecting all the sets G(s) with T_0, and nested sets remain nested when they are *all* intersected with a fixed set. In terms of the contour function, conditioning on T_0 means replacing f by the contour function f′ on T_0, where

$$f'(t) = \frac{f(t)}{\sup\{f(t')|t' \text{ in } T_0\}.} \tag{24}$$

(If we wish to think of the conditional belief function Bel′ as a belief function defined over the original frame T, then Equation 24 should be used only for t in T_0; $f'(t) = 0$ for t not in T_0.)

When f is the contour function for a consonant belief function over a product frame $S \times T$ and we condition on an element s of S, we obtain the consonant belief function over T with contour function

$$f_{T|s}(t) = \frac{f(s,t)}{\sup\{f(s,t')|t' \text{ in } T\}.} \tag{25}$$

From Equations 22 and 25, we obtain

$$f(s,t) = f_S(s)f_{T|s}(t). \tag{26}$$

This formula tells us that there is one and only one consonant belief function over $S \times T$ having f_S as its marginal over S and the $f_{T|s}$ as its conditionals over T given s. (Compare Section IV.D above.)

The product of two consonant belief functions is not usually consonant. This can easily be seen by considering multivalued mappings for the two belief functions. If these are denoted by G_1 and G_2, and if $G_1(s_1) \subseteq G_1(s'_1)$ and $G_2(s_2) \subseteq G_2(s'_2)$, but these sets are all nonempty and the inclusions are proper, then neither $G_1(s_1) \times G_2(s'_2)$ nor $G_1(s'_1) \times G_2(s_2)$ contains the other.

Though the belief-function product does not preserve consonance, there are simple methods of constructing a consonant belief function over a product frame from consonant belief functions over the separate frames. There are, that is to say, simple methods of constructing a consonant belief function Bel over $S \times T$ that has given consonant belief functions Bel_S and Bel_T as marginals. Here are two such methods.

1. We can ask for a consonant belief function Bel over $S \times T$ that satisfies

$$Pl(A \times B) = \min\{Pl_S(A), Pl_T(B)\} \tag{27}$$

for all subsets A of S and B of T. If we substitute {s} for A and {t} for B in Equation 27, we obtain

$$f(s,t) = \min\{f_S(s), f_T(t)\}, \tag{28}$$

which is a contour function. It is easily seen that the consonant belief function with this contour function does satisfy Equation 27 for all A and B. So there is a unique consonant belief function satisfying Equation 27. It is clear from Equation 27 that this belief function has Bel_S and Bel_T as its marginals. Since the contour function f for any consonant belief function over $S \times T$ that does have these marginals must satisfy $f(s,t) \leq f_S(s)$ and $f(s,t) \leq f_T(t)$, or

$$f(s,t) \leq \min\{f_S(s), f_T(t)\},$$

we see that the consonant belief function with the contour function given by Equation 28 is the weakest consonant belief function having these marginals.

2. We can ask for a consonant belief function Bel on $S \times T$ that satisfies

$$Pl(A \times B) = Pl_S(A)Pl_T(B) \tag{29}$$

for all subsets A of S and B of T. (This is the same as Condition 14 above — one of the two conditions that characterize the belief function product.) If we substitute {s} for A and {t} for B in Equation 29, we obtain

$$f(s,t) = f_S(s)f_T(t), \tag{30}$$

which is a contour function. It is easily seen that the consonant belief function with this contour function does satisfy Equation 29 for all A and B. So there is a unique consonant belief function satisfying Equation 29. It is clear from Equation 29 that this belief function has Bel_S and Bel_T as its marginals. We also see, from Equations 16 and 29, that the marginal degrees of belief for S are preserved when this belief function is conditioned on any subset of T, and viceversa. Finally, by comparing Equations 26 and 30, we see that this belief function can be characterized as the unique consonant belief function that has Bel_S as its marginal over S and Bel_T as its conditional over T given any s.

In order to use either of these methods in constructive probability judgment, we must first of all make the judgment that our evidence is consonant with respect to the frame $S \times T$. If we are then able to construct marginals Bel_S and Bel_T, but feel that we can make no probability judgments beyond these, then we can argue for completing our job by the first method. (The idea of adopting the weakest belief function consistent with given judgments is already familiar to us, after all, from the idea of extension. When we extend a belief function from one frame to another, we are adopting on the second frame the weakest degrees of belief that are consistent with our degrees of belief on the first frame.) If on the other hand, we can make the additional judgment that Bel_T should also give conditional degrees of belief given any value of s, then no further argument is needed; we obtain our consonant belief function by the second method.

Since these two ways of constructing belief functions over a product frame can be thought of as alternatives to the rule for forming product belief functions, we might ask whether they can be used to formulate alternatives to Dempster's rule for combining belief functions. We arrived at Dempster's rule by treating two belief functions over a frame S as if they were over independent frames, forming their product over $S \times S$, and then conditioning on the diagonal to obtain a new belief function over S. Given two consonant belief functions over a frame S, we might similarly use Equations 28 or 30 to construct a consonant belief function over $S \times S$ and then condition on the diagonal to obtain a new consonant belief function over S. If the two consonant belief functions being combined in this way have contour functions f_1 and f_2, respectively, then the contour function for the new consonant belief function will be given by

$$f(s) = \frac{\min\{f_1(s), f_2(s)\}}{\sup\{\min\{f_1(s'), f_2(s')\} | s' \text{ in } S\}} \tag{31}$$

in the case of Equation 28 and by

$$f(s) = \frac{f_1(s) f_2(s)}{\sup\{f_1(s') f_2(s') | s' \text{ in } S\}} \tag{32}$$

in the Equation case of 31. Let us call Equation 31 *Zadeh's rule of combination*, and let us call Equation 32 the *likelihood rule of combination*. (I will explain the motivation for these names below.)

What interpretation can be given to these rules of combination? Can they be interpreted as rules for pooling evidence? What kind of canonical examples might provide an interpretation for them?

It should be said first of all that neither rule seems to be interpretable in terms of the canonical examples for belief functions given in section A above. Thus, we can expect that any interpretation we can provide for them will lead us out of the language of belief functions.

It should also be noted that Zadeh's rule does not pool evidence in the sense in which Dempster's rule pools evidence. We can see this by comparing the behavior of the rules when they are used to combine a pair of identical belief functions. When we use Dempster's rule to combine a belief function Bel with itself, we obtain, in general, a different belief function, one that favors even more strongly the subsets of the frame most favored by Bel. This fits the idea of pooling independent bodies of evidence; if two independent bodies of evidence both support A, then they together give more support to A than either alone does. Zadeh's rule, on the other hand, is idempotent; when it is used to combine a consonant belief function with itself, it yields that same consonant belief function.

The likelihood rule, on the other hand, is more closely related to Dempster's rule. In fact, it gives the same answer as Dempster's rule when it is used to combine a consonant belief function with itself. One way to provide a basis for the likelihood rule, and the reason for its name, can be seen if we imagine that the frame S is the parameter space for a statistical model and that the belief functions being combined are derived from independent statistical observations in the way suggested in Chapter 11 of Shafer.[8] In this case, the contour functions are normalized likelihood functions, in the sense in which this term is used in statistics, and Equation 32 is simply the usual rule for combining independent statistical observations. If we use Equation 32 as a general rule for combining consonant belief functions, then we may say that we are comparing ordinary sorts of evidence with parametric statistical models; these are our canonical examples. For related ideas about using statistical examples as canonical examples for belief functions, see Krantz and Miyamoto.[53]

VI. POSSIBILITY MEASURES

The idea of a *possibility measure* was introduced by Zadeh,[20] who developed the idea from his earlier work on fuzzy sets. Several authors[54-56] have pointed out that formally, at least, a possibility measure is very nearly the same as a consonant plausibility function. Moreover, many of the rules Zadeh has formulated for possibility measures agree with the rules for consonant plausibility functions. Since the extent of this agreement can be obscured by differences in vocabulary and notation, I will review Zadeh's definitions using the vocabulary of frames and the notation of Section V.G above. After this review I will discuss the problem of finding canonical examples for possibility measures; I have not been able to formulate canonical examples to explain the rules for possibility measures that differ from those for consonant plausibility functions. Then I will turn to two examples that Zadeh has presented to criticize the language of belief functions.

It should be mentioned, for the sake of perspective, that mathematical structures for probability judgment that are similar to possibility measures have also been studied by Shackle,[57] Cohen,[58,59] and Levi.[60,61] I will not discuss the work of these authors in this article, but it is possible that some of this work might suggest canonical examples for a language for probability judgment.

A. Formalism

Instead of writing about a frame that lists possible answers to a certain question, Zadeh writes about a universe of discourse that lists possible values for a certain variable. However, if we translate his formulation into the terminology of frames, we may say that he begins with a frame S and a function $f{:}S \rightarrow [0,1]$ that he calls a *possibility distribution function*. The number f(s) is supposed to indicate the extent to which it is possible that s is the answer to the question considered by S. If $f(s) = 1$, then s is a fully possible answer; if $f(s) = 0$, then s is an impossible answer. Zadeh calls the function Pl on 2^S defined by

$$Pl(A) = \sup\{f(s) | s \text{ in } A\}$$

the *possibility measure* based on f. The number Pl(A) is supposed to indicate the extent to which it is possible that the answer to the question considered by S is in A. (*Note:* A function f that maps a set S into the interval [0.1] is also sometimes called a *fuzzy subset* of S; it generalizes the idea of the characteristic function of a subset of S. Zadeh was led to his interest in possibility measures through his interest in fuzzy sets.)

These definitions seem to agree with those in Section V.G; a possibility distribution function seems to be the same as a contour function, and a possibility measure seems to be the same as a consonant plausibility function. There is, however, one difference. In Section V.G, I required that the supremum of the values of f should be one, so that $Pl(S) = 1$. Zadeh does not insist on this requirement; he allows $Pl(S) < 1$.

Given a possibility measure Pl over a product frame $S \times T$, with possibility distribution function f, Zadeh[20] calls the possibility measure over S that has the contour function

$$f_S(s) = \sup\{f(s,t) | t \text{ in } T\} \tag{33}$$

the *marginal* of Pl for S. Given a possibility measure Pl over S, with possibility distribution function f, he calls the possibility measure over $S \times T$ that has the contour function

$$f_{S \times T}(s,t) = f(s) \tag{34}$$

the *cylindrical extension* of Pl. These definitions are again familiar from our study of belief functions; Equation 33 is the same as Equation 21, and Equation 34 is the same as Equation 23. Both are special cases of the general rule for extending belief functions.

As we saw in Section V.G, the specification of a possibility measure Pl_S over S and a possibility measure Pl_T over T, together with the condition that

$$Pl(A \times B) = \min\{Pl_S(A), Pl_T(B)\}$$

for all subsets A of S and B of T, completely determine a possibility measure Pl over $S \times T$. This possibility measure has Pl_S and Pl_T as its marginals, and it can also be defined by the relation

$$f(s,t) = \min\{f_S(s), f_T(t)\}$$

for all s in S and all t in T. When Pl is determined by its marginals in this way, Zadeh calls S and T *noninteractive* with respect to Pl. He also calls Pl the *Cartesian product* of Pl_S and Pl_T. In order to avoid confusion with the belief-function product, let us instead call it the *noninteractive product*. As we saw in Section V.G, the noninteractive product Pl of Pl_S and Pl_T is the weakest consonant plausibility function with the marginals Pl_S and Pl_T.

Zadeh formulates a rule for conditioning possibility measures that differs from the rule of conditioning for consonant plausibility functions only in not requiring normalization. His rule says simply that the result of conditioning a possibility measure Pl on a subset S_0 of its frame S is the possibility measure Pl′ given by

$$Pl'(A) = Pl(A \cap S_0). \tag{35}$$

(We can think of Pl′ as a possibility measure over S or over the reduced frame S_0; in either case Equation 35 holds for all subsets A of the frame.) In terms of the possibility distribution functions, Equation 35 may be written

$$f'(s) = f(s) \tag{36}$$

for all s in S_0. (If we wish to regard Pl as a possibility measure over S, then we must add that $f'(s) = 0$ for all s not in S.) If Pl is a possibility measure over a product frame $S \times T$, then conditioning on an element s of S results in the possibility measure $Pl_{T|s}$ over T given by

$$Pl_{T|s}(B) = Pl(\{s\} \times B).$$

In terms of the possibility distribution functions, this may be written

$$f_{T|s}(t) = f(s,t). \tag{37}$$

Notice that Equations 36 and 37 are the same as Equations 24 and 25, except that they omit the denominators that appear on the right-hand sides of Equations 24 and 25. One way to explain these denominators is to say that they normalize the new contour functions so that they have suprema equal to one. Since Zadeh does not insist that the suprema of possibility distribution functions should equal one, he does not hold that such normalization is always appropriate. "In some applications," he writes, "it may be appropriate to normalize . . . " (Zadeh,[20] page 18).

Suppose we have two possibility measures Pl_1 and Pl_2 over the same frame S. Then we may define another possibility measure Pl over S by

$$f(s) = \min\{f_1(s), f_2(s)\} \tag{38}$$

Zadeh calls Pl the result of *particularizing* Pl_1 by Pl_2, and he regards particularization as a kind of pooling of information.

Particularization is a generalization of conditioning. To see this, note that the information that the answer to the question considered by S is in the subset S_0 can be represented by the possibility measure Pl_2 over S that has the possibility distribution function

$$f_2(s) = \begin{cases} 1 & \text{if s is in } S_0 \\ 0 & \text{if s is not in } S_0 \end{cases}$$

(This representation accords with the representation of such information in the theory of belief functions, for this possibility measure is the plausibility function associated with the binary Sugeno measure focused on S_0.) If we particularize another possibility measure Pl by Pl_2, then according to Equation 38 we will obtain a possibility measure with the possibility distribution function

$$f'(s) = \begin{cases} f(s) & \text{if s is in } S_0 \\ 0 & \text{if s is not in } S_0, \end{cases}$$

where f is the possibility distribution function for Pl. And f′ is the same as the conditional possibility distribution function given by Equation 36.

Comparison of Equation 38 with Equation 31 shows that the rule of combination that I called "Zadeh's rule" in the preceding section can be thought of as particularization followed by normalization. Since Zadeh is wary of normalization, it may be somewhat misleading to put his name on this rule of combination. Doing so acknowledges, however, the importance of Zadeh's work in drawing attention to the minimum operation.

Not all the authors who have taken up Zadeh's ideas about possibility measures have been as wary of normalization as Zadeh. Dubois and Prade,[42] for example, have added to Zadeh's definition of possibility measure the requirement that the possibility for the whole frame should be one. By insisting on normalization, these authors have brought Zadeh's theory of possibility closer to the language of belief functions. At the same time, however, they have been interested in Zadeh's rule of combination and other rules of composition and combination that seem to lie outside of the language of belief functions.

I mentioned in Section V.G that the rules given there for constructing a consonant belief function over a product frame from given marginals (Equations 28 and 30) are just two of many possibilities. In fact, there is a distinct rule for every triangular norm. A *triangular norm*[26,62] is a real-valued function T on the unit square $[0,1] \times [0,1]$ that satisfies

(i) $T(0,0) = 0$ and $T(a,1) = T(1,a) = a$,
(ii) $T(a,b) \leq T(c,d)$ whenever $a \leq c$ and $b \leq d$,
(iii) $T(a,b) = T(b,a)$,
(iv) $T(a,T(b,c)) = T(T(a,b),c)$.

Any such function T maps the unit square to [0,1] and gives, through the formula

$$f(s,t) = T(f_S(s), f_T(t)) \quad (39)$$

a way of constructing a consonant belief function on $S \times T$ with given marginals on S and T.

Dubois, Smets, and Yager have all pointed out, in conversation, that any triangular norm also leads to a commutative and associative rule for combining possibility measures over a given frame. We simply apply to Equation 39 the device we have already applied to the special cases in Equations 28 and 30; given two possibility measures Pl_1 and Pl_2 over S, we form a possibility measure over $S \times S$ by

$$f(s_1, s_2) = T(f_1(s_1), f_2(s_2)),$$

and then we condition on the diagonal to obtain a possibility measure over S. Other rules for possibility measures have been proposed by Nguyen,[63] Hisdal,[64] and others. For a review, see Dubois and Prade.[65]

B. Interpretation

What is the meaning of a particular numerical value, say one third, for the "degree of possibility" Pl(A)? How might we assign such a value to a particular subset A of a frame of discernment? Also, how can we justify rules for manipulating such quantities?

We can begin to answer these questions only after we have specified a scale of canonical examples for possibility measures. Once we have such a scale, we can say that the assignment of a given numerical degree of possibility to a proposition means that our knowledge and evidence about the proposition is like the knowledge and evidence about an analogous proposition in one of the canonical examples, and the presence of such a scale transforms the problem of assigning numerical judgments into the richer and more concrete problem of matching our knowledge and evidence to a point on the scale. Finally, the canonical examples should elucidate the rules for manipulating the numbers; these rules should be intuitively reasonable, at least, in the context of these examples.

What, then, are the canonical examples for possibility measures? As we have seen, many of the rules that have been proposed for possibility measures can be accounted for by the canonical examples we proposed for belief functions in Section V.A, and the likelihood rule of combination can be accounted for by using statistical models as canonical examples. However, neither of these scales of canonical examples provide a basis for nonnormalized possibility measures, for the plethora of rules of combination resulting from different triangular norms, or for many of the other rules that various authors have proposed for possibility measures.

Zadeh has emphasized that possibility measures should take into account imprecision of the kind we find in natural language as well as uncertainty of the kind we can have about precisely posed questions. Can we find canonical examples for possibility measures by looking at the imprecision of natural languages? Can we, that is to say, tell stories in which this imprecision is guaged by a natural numerical scale?

One approach, of course, would be to bring in the usual scale of frequency or probability. We can imagine, that is to say, that a natural-language expression might have a sequence of progressively less precise possible meanings, each with a given known chance. "Bob is very tall", for example, might have a chance 0.1 of meaning that he is at least 6 ft. 6 in. tall, a chance 0.3 of meaning only that he is at least 6 ft. 3 in. tall, and a chance 0.6 of meaning only that he is at least 6 ft. tall. However, here we are back in the domain of belief functions: canonical examples where a message has different meanings with known chances.

Zadeh has suggested that "degree of possibility" should sometimes be interpreted as "degree of ease", either physical or figurative. The possibility that Hans ate a certain number of eggs may be interpreted, for example, as the degree of ease with which he can eat that many eggs.[20] This suggestion is a step towards a canonical example; it invokes a picture of "elastic" physical constraints. However, it does not tell us how to fill in this picture so as to give a meaning to the numerical scale, and without this, "degree of ease" remains a vague concept, not yet differentiated from the "probability" of the Bayesian language or the "degree of belief" of the language of belief functions. After all, James Bernoulli, who was the first to base a theory of subjective probability judgment on the mathematical picture of chance, described the probability of something as the degree of ease with which it happens.[66,67]

I have not been able to formulate canonical examples for possibility measures that lead to a language for probability judgment that diverges from the language of belief functions. I believe, however, that the formulation of such examples is necessary. This, rather than greater mathematical elegance or yet more alternative rules, is what is needed to make the idea of a possibility measure prosper.

C. Zadeh's Criticism of Normalization

Zadeh[68] has given two examples to illustrate why he thinks the normalization used in the language of belief functions is misguided. Here is one of the examples (it is also discussed more abstractly in Zadeh.[69]

> Suppose that a patient, P, is examined by two doctors, A and B. A's diagnosis is that P has either meningitis, with probability 0.99, or brain tumor, with probability 0.01. B agrees with A that the probability of brain tumor is 0.01, but believes that it is the probability of concussion rather than meningitis that is 0.99. Applying the Dempster rule to this situation leads to the conclusion that the belief that P has brain tumor is 1.0 — a conclusion that is clearly counterintuitive because both A and B agree that it is highly unlikely that P has a brain tumor. What is even more disconcerting is that the same conclusion, (i.e., Bel(brain tumor) = 1) would obtain regardless of the probabilities associated with the other possible diagnoses.

Abstractly, this example involves two belief functions Bel_1 and Bel_2 over a frame of discernment S with three elements:

$$S = \{\text{meningitis, tumor, concussion}\}$$

The belief functions Bel_1 and Bel_2 both happen to be probability measures, but this is not essential to the example. The essential point is that $Bel_1(\{\text{meningitis,tumor}\}) = 1$ and $Bel_2(\{\text{tumor,concussion}\}) = 1$. The first belief function rules out a concussion, and the second rules out meningitis, so taken together they leave a tumor as the only possibility.

I am puzzled by Zadeh's rejection of this reasoning. If the evidence on which the first belief function is based does rule out a concussion, and the evidence on which the second is based does rule out meningitis, and if we accept the initial assumption, embodied in our frame S, that meningitis, tumor, and concussion are the only possibilities, then it is a matter of logic, not merely probability, that the patient must have a tumor. As Sherlock Holmes put it, when you have eliminated the impossible, whatever remains, however improbable, must be the truth.

In practice, evidence seldom rules out anything with absolute certainty, and it may not be appropriate to take Bel_1 and Bel_2 at face value. Instead, we may want to adjust them to allow every element of S some plausibility, however small. One way to do this is to *discount*. Discounting a belief function means just what one might guess; when Bel is discounted at the rate a, the degree of belief Bel(A) is reduced to $(1 - a)$Bel(A) for every proper subset A of Bel's frame (Shafer,[8] page 252). The degree of belief we will obtain for a tumor if we combine Bel_1 and Bel_2 after discounting will depend, of course, on the discount rates — i.e., on our judgment of the reliabilities of the two doctors. For a study of the interaction between discounting and combination, see Shafer.[15]

It is also true that combination by Dempster's rule is probably not appropriate in this example. The two doctors are probably not looking at independent items of evidence. There is even a hint that what one takes as evidence for meningitis the other takes as evidence for a concussion. A better analysis of the problem would take a closer look at the structure of the evidence and would attempt to sort out just where the uncertainties lie, which are independent of which, and which are best assessed by which doctor.

Finally, I would like to raise a question about a verbal detail. Zadeh writes, "B agrees with A that the probability of brain tumor is 0.01, but believes that it is the probability of concussion rather than meningitis that is 0.99". What is meant by the reference to a belief about a probability? Such language is appropriate in the context of statistical models that posit objective probabilities, but it can be misleading in the context of subjective probability judgment. A degree of belief Bel(A) based on given evidence is a judgment in which we may have more or less confidence, but it is not a belief about *the* probability of A. Degrees of belief $Bel_1(A)$ and $Bel_2(A)$ that are based on independent bodies of evidence are answering different questions. One is telling about the support given to A by the one body of evidence; the other is telling about the support given by the other. They are not, just because they happen to be numerically equal, agreeing about a single well-defined number called the probability of A.

Here is Zadeh's second example.

> Country X believes that a submarine, S, belonging to Country Y is hiding in X's territorial waters. The Minister of Defense of X summons a group of experts, $E_1, \ldots E_n$, and asks each one to indicate the possible locations of S. Assume that the possible locations specified by the experts $E_1, \ldots E_m$, $m \leq n$, are $L_1 \ldots L_m$, respectively, where L_i, $i = 1 \ldots, m$, is a subset of the territorial waters; the remaining experts, $E_{m-1} \ldots, E_n$, assert that there is no submarine in the territorial waters, or, equivalently, that $L_{m-1} = \ldots = L_n = \emptyset$, where $\emptyset$ is the empty set.

Zadeh asserts that the basic idea of the "Dempster-Shafer theory" is that we should construct a belief function from this information by setting the degree of belief that the submarine is in any given subset A of the territorial waters equal to

$$Bel(A) = \frac{\Sigma \{w_i \mid E_i \subseteq A, E_i \neq \emptyset\}}{\Sigma \{w_i \mid E_i \neq \emptyset\}} \qquad (40)$$

where $w_1, \ldots, w_n$ are positive weights adding to one. The presence of the denominator that normalizes Equation 40 means, according to Zadeh, that we are disregarding the opinion of those experts who believe there is no submarine in the territorial waters.

Mathematically, the set function defined by Equation 40 is indeed a belief function, but its use to give degrees of belief about the location of the submarine would make sense in the language of belief functions only if:

1. We somehow knew that exactly one of the experts is reliable.
2. We assessed our evidence about which is reliable by assigning the probability w_i to E_i being reliable.
3. We took at face value the initial assumption that there is a submarine in territorial waters.

It is Assumption 3 that tells us to disregard the opinion of those experts who disagree. If we do not want to disregard these opinions, presumably we will not make Assumption 3; instead we will formulate a frame that allows for the possibility that no submarine is present and use the contrary evidence to justify a mere probability judgment.

Here, as in the story about medical diagnosis, a better analysis would take account of the details of the evidence the experts are using. It seems unlikely that we would have any *a priori* reason to think that one of the experts is completely reliable while all the others are unreliable. However, we might learn something by studying the evidence and pooling the expertise of the experts in order to assess the various uncertainties in that evidence.

ACKNOWLEDGMENTS

This article is based on a talk I gave at the First International Conference on Fuzzy Information Processing, Hawaii, July 22 to 26, 1984. James C. Bezdek invited me to speak at this conference, and Lotfi Zadeh and the rest of the "fuzzy" community gave me a very gracious reception. I would like to thank Didier Dubois for correcting an error I made in the talk, and I would like to thank Dubois, Hung T. Nguyen, Steve Hillmer, and Anca Ralescu for their comments on an earlier draft. Research for the article has been partially supported by National Science Foundation Grant IST-8405210.

LIST OF SYMBOLS

$\subseteq$	Contained in
$\not\subseteq$	Not contained in
$\cup$	Union
$\cap$	Intersection
Σ	Sigma (for summation)

REFERENCES

1. **Dempster, A. P.**, New methods for reasoning towards posterior distributions based on sample data, *Ann. Math. Stat.*, 37, 355, 1966.
2. **Dempster, A. P.**, Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Stat.*, 38, 325, 1967.
3. **Dempster, A. P.**, Upper and lower probability inferences based on a sample from a finite univariate population, *Biometrika*, 54, 515, 1967.
4. **Dempster, A. P.**, Upper and lower probabilities generated by a random closed interval, *Ann. Math. Stat.*, 39, 957, 1968.
5. **Dempster, A. P.**, A generalization of Bayesian inference (with discussion), *J. R. Stat. Soc. Ser. B*, 30, 205, 1968.
6. **Dempster, A. P.**, Upper and lower probability inferences for families of hypotheses with monotone density rations, *Ann. Math. Stat.*, 40, 953, 1969.
7. **Shafer, G.**, Allocations of Probability, Ph.D. thesis, Princeton University, Princeton, N. J., 1973.
8. **Shafer, G.**, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N. J., 1976.
9. **Shafer, G.**, A theory of statistical evidence, in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. 2, Harper, W. L. and Hooker, C. A., Eds., Reidel, Dordrecht, 1976, 365.
10. **Shafer, G.**, Dempster's Rule of Combination, Tech. Rep. University of Kansas, Lawrence, 1978.
11. **Shafer, G.**, Allocations of probability, *Ann. Probability*, 7, 827, 1979.
12. **Shafer, G.**, Constructive probability, *Synthese*, 48, 1, 1981.
13. **Shafer, G.**, Jeffrey's rule of conditioning, *Philos. Sci.*, 48, 337, 1981.
14. **Shafer, G.**, Two theories of probability, in *PSA 1978*, Vol. 2, Asquith, P. D. and Hacking, I., Eds., Philosophy of Science Association, East Lansing, Mich., 1981, 441.
15. **Shafer, G.**, Lindley's paradox (with discussion), *J. Am. Stat. Assoc.*, 77, 325, 1982.
16. **Shafer, G.**, Belief functions and parametric models, *J. R. Stat. Soc. Ser. B*, 44, 322, 1982.
17. **Shafer, G.**, The Combination of Evidence, Working Paper No. 161, School of Business, University of Kansas, Lawrence, 1984.
18. **Shafer, G.**, The Problem of Dependent Evidence, Working Paper No. 164, School of Business, University of Kansas, Lawrence, 1984.
19. **Shafer, G. and Breipohl, A. M.**, Reliability described by belief functions, in *Proc. 1979 Annual Reliability and Maintainability Symp.*, IEEE, Piscataway, N. J., 1979, 23.
20. **Zadeh, L.**, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets Syst.*, 1, 3, 1978.
21. **Zadeh, L.**, Test-score semantics for natural languages and meaning representation via PRUF, in *Empirical Semantics*, Rieger, B. B., Ed., Brockmeyer, Bochum, 1981, 281.

22. **Zadeh, L.,** The role of fuzzy logic in the management of uncertainty in expert systems, *Fuzzy Sets Syst.*, 11, 199, 1981.
23. **Shafer, G. and Tversky, A.,** Languages and designs for probability judgment, *Cognitive Sci.*, 9, 309, 1985.
24. **Sugeno, M.,** Fuzzy measures and fuzzy integrals: a survey, in *Fuzzy Automata and Decision Processes*, Gupta, M. M., Saridis, G. N., and Gaines, B. R., Eds., North-Holland, Amsterdam, 1977.
25. **Banon, G.,** Distinction between several subsets of fuzzy measures, *Fuzzy Sets Syst.*, 5, 219, 1981.
26. **Dubois, D. and Prade, H.,** A class of fuzzy measures based on triangular norms, *Int. J. Gen. Syst.*, 8, 43, 1982.
27. **Smith, C. A. B.,** Consistency in statistical inference and decision, *J. R. Stat. Soc. Ser. B*, 23, 1, 1961.
28. **Good, I. J.,** The measure of a non-measurable set, in *Logic, Methodology, and Philosophy of Science*, Nagel, E., Suppes, P., and Tarski, A., Eds., Stanford University Press, Stanford, Calif., 1962, 319.
29. **Williams, P. M.,** Indeterminate probabilities, in *Formal Methods in the Methodology of Empirical Sciences*, Przelecki, M., Szaniawski, K., and Wojciki, R., Eds., Kluwer Academic, Hingham, Mass., 1976, 229.
30. **Walley, P. and Fine, T. L.,** Towards a frequentist theory of upper and lower probability, *Ann. Stat.*, 10, 321, 1979.
31. **Wierzchón, S. T.,** An Inference Rule Based on Sugeno Measure, Institute of Computer Science, Polish Academy of Sciences, Warsaw, 1984.
32. **Dubois, D.,** personal communication, 1984.
33. **Kumar, A. and Fine, T. L.,** Stationary lower probabilities and unstable averages, *Z. Wahrscheinlichkeitstheorie und verwandte Gebiete*, in press.
34. **Papamarcou, A. and Fine, T. L.,** A Note on Undominated Lower Probabilities, School of Electrical Engineering, Cornell University, Ithaca, N.Y., 1984.
35. **Feller, W.,** *An Introduction to Probability Theory and Its Applications*, Vol. 2, John Wiley & Sons, New York, 1966.
36. **Kolmogorov, A.,** *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Basel, 1933.
37. **de Finetti, B.,** *Theory of Probability*, Vol. 1, John Wiley & Sons, New York, 1974.
38. **Shafer, G.,** Bayes's two arguments for the rule of conditioning, *Ann. Stat.*, 10, 1075, 1982.
39. **Shafer, G.,** A subjective interpretation of conditional probability, *J. Philos. Logic*, 12, 453, 1983.
40. **Shafer, G.,** Conditional probability, *Int. Stat. Rev.*, 53, 261, 1985.
41. **Choquet, G.,** Theory of capacities, *Ann. Inst. Fourier Grenoble*, 5, 131, 1953.
42. **Dubois, D. and Prade, H.,** Evidence measures based on fuzzy information, *Automatica*, in press.
43. **Freeling, A. N. S.,** Belief and Decision Aiding, Doctoral dissertation, St. John's College, Cambridge, U. K., 1983.
44. **Freeling, A. N. S. and Sahlin, N.-E.,** Combining evidence, in *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, Gardenfors, P., Hansson, B., and Sahlin, N.-E., Eds., C. W. K. Gleerups, Lund, Sweden, 1983, 58.
45. **Walley, P.,** Coherent Lower (and Upper) Probabilities, Statistics Res. Rep. 22, University of Warwick, Coventry, England, 1981.
46. **Hooper, G.,** A fair and methodical discussion of the first and great controversy between the Church of England and the Church of Rome concerning the infallible guide, in *Works*, Vol. 2, Oxford, 1855, 40.
47. **Ekelöf, P. O.,** My thoughts on evidentiary value, in *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, Gardenfors, P., Hansson, B., and Sahlin, N.-E., Eds., C. W. K. Gleerups, Lund, Swden, 1983, 9.
48. **Smets, P.,** Un Modele Mathematico-Statistique Simulant le Processus du Diagnostic Medical, These d'agregation, Universite Libre de Bruxelles, Belgium, 1978.
49. **Garvey, T. D., Lowrance, J. D., and Fischler, M. A.,** An inference technique for integrating knowledge from disparate sources, in *Proc. 7th Int. Joint Conf. on Artificial Intelligence*, Kaufmann, William, Los Altos, Calif., 1981, 319.
50. **Lowrance, J. D.,** Dependency-Graph Models of Evidential Support, Ph.D. thesis, University of Massachusetts, Amherst, 1982.
51. **Gordon, J. and Shortliffe, E. H.,** The Dempster-Shafer theory of evidence, in *Rule-Based Expert Systems*, Buchanan, B. G. and Shortliffe, E. H., Eds., Addison-Wesley, Reading, 1984, 272.
52. **Nguyen, H. T.,** On random sets and belief functions, *J. Math. Anal. Appl.*, 65, 531, 1978.
53. **Krantz, D. H. and Miyamoto, J.,** Priors and likelihood ratios as evidence, *J. Am. Stat. Assoc.*, 78, 418, 1983.
54. **Smets, P.,** The degree of belief in a fuzzy event, *Inf. Sci.*, 25, 1, 1981.
55. **Prade, H.,** On the link between Dempster's rule of combination and fuzzy set intersection, *Busefal*, 8, 60, 1981.
56. **Yager, R. R.,** Entropy and specificity in a mathematical theory of evidence, *Int. J. Gen. Syst.*, 9, 249, 1983.

57. **Shackle, G. L. S.**, *Decision, Order, and Time*, Cambridge, 1969.
58. **Cohen, L. J.**, *The Implications of Induction*, Methuen, 1970.
59. **Cohen, L. J.**, *The Probable and the Provable*, Oxford, 1977.
60. **Levi, I.**, On potential surprise, *Ratio*, 8, 107, 1966.
61. **Levi, I.**, Potential surprise: its role in inference and decision-making, in *Applications of Inductive Logic*, Cohen, L. J. and Hesse, M., Eds., Oxford, 1980, 1.
62. **Menger, K.**, Statistical metrics, *Proc. Natl. Acad. Sci. U.S.A.*, 28, 535, 1942.
63. **Nguyen, H. T.**, On conditional possibility distributions, *Fuzzy Sets Syst.*, 1, 229, 1978.
64. **Hisdal, E.**, Conditional possibilities, independence, and noninteraction, *Fuzzy Sets Syst.*, 1, 283, 1978.
65. **Dubois, D. and Prade, H.**, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.
66. **Bernoulli, J.**, *Ars Conjectandi*, Basel, 1713.
67. **Shafer, G.**, Non-additive probabilities in the work of Bernoulli and Lambert, *Arch. History Exact Sci.*, 19, 309, 1978.
68. **Zadeh, L.**, Review of *A Mathematical Theory of Evidence*, by Glenn Shafer, *The AI Magazine*, 5, No. 3, 81, 1984.
69. **Zadeh, L.**, On the Validity of Dempster's Rule of Combination of Evidence, ERL Memorandum M79/24, University of California, Berkeley, 1979.