# Reviews of Books

**A Mathematical Theory of Evidence.** GLENN SHAFER. Princeton University Press, Princeton, NJ, 1976.

The seminal work of Glenn Shafer—which is based on an earlier work of Arthur Dempster—was published at a time when the theory of expert systems was in its infancy and there was little interest within the AI community in issues relating to probabilistic or evidential reasoning.

Recognition of the relevance of the Dempster-Shafer theory to the management of uncertainty in expert systems was slow in coming. Today, it is the center of considerable attention within AI due in large measure to (a) the emergence of expert systems as one of the most significant areas of activity in knowledge engineering, and (b) the important extensions, applications and implementations of Shafer's theory made by John Lowrance at SRI International, Jeff Barnett at USC/ISI, and Ted Shortliffe and Jean Gordon at Stanford University.

What are the basic ideas behind the Dempster-Shafer theory? In what ways is it relevant to expert systems? What are its potentialities and limitations? My review of Shafer's book will be more of an attempt to provide some answers to these and related questions than a chapter-by-chapter analysis of its contents.

To understand Shafer's theory, it is best to start with a careful reading of Dempster's original paper (Dempster, 1967) which provides the basis for it.

Stated in somewhat simplified terms, the central problem considered by Dempster is the following. If $y$ is a function of $x$, say $y = f(x)$, and $x$ has a specified probability distribution, then an elementary result in probability theory yields the probability distribution of $y$ as a function of $f$ and the probability distribution of $x$. But what if $f$ is a relation or, equivalently, a set-valued function, which implies that to a given value of $x$ corresponds a set of values of $y$? Suppose that $x$ and $y$ take values in $U$ and $V$, respectively, and $A$ is a specified subset of $V$—which is referred to as the *frame of discernment* in Shafer's book. Then the question is: What is the probability that $y$ is in $A$? If $f$ is a point function, the answer is a real-valued probability. But $f$ is a set-valued function, the answer is not unique and all that can be asserted is that the probability in question lies between two bounds which are the lower and upper probabilities, $P_*(A)$ and $P^*(A)$, respectively. The lower and upper probabilities associated with the proposition "$y$ is in $A$" correspond to what in Shafer's theory are called the degree of belief, $\mathrm{Bel}(A)$, and the degree of plausibility, $P1(A)$. Do $P_*(A)$ and $P^*(A)$ capture our intuitive perception of the meaning of belief and plausibility? This is a semantic question that, objectively, does not have an obvious answer. In my opinion, however, the answer is in the negative.

What makes Shafer's exposition much harder to understand than Dempster's is that Dempster, starting with the concept of a set-valued mapping, derives a number of properties of $P_*(A)$ and $P^*(A)$. Shafer, on the other hand, uses the belief function, $\mathrm{Bel}(A)$, as the point of departure. This is more elegant mathematically, but has the effect of obscuring the motivation of the postulated properties of the belief function, one of which is that of super-additivity, *i.e.*, $\mathrm{Bel}(A \cup B) \geq \mathrm{Bel}(A) + \mathrm{Bel}(B) - \mathrm{Bel}(A \cap B)$. Actually, the Dempster-Shafer theory is a natural, important and useful extension of classical probability theory in which the probabilities may be assigned not just to points in the frame of discernment—which is assumed to be a finite set—but, more generally, to subsets of $V$. Viewed in this perspective, the Dempster-Shafer theory is closely related to the theory of random sets. Another related viewpoint which is developed in Zadeh (1979a) involves the notion of information granularity—a notion which is partly probabilistic and partly possibilistic in nature.

The basic ideas underlying the Dempster-Shafer theory are actually quite simple and can readily be understood through a concrete example. Specifically, assume that 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.

Now suppose that the question raised by the Minister of Defense is: Is $S$ in a specified subset, $A$, of the territorial waters? In this regard, there are two cases to consider: (i) $E_i \subset A$, which implies that, in the view of $E_i$, it is *certain*, *i.e.*, *necessarily true*, that $S$ is in $A$; and (ii) $E_i \cap A \neq 0$, *i.e.*, $E_i$ intersects $A$, in which case it is possible that $S$ is in $A$. Note that (i) implies (ii).

Furthermore, suppose that the Minister of Defense aggregates his experts opinion by averaging. Thus, if $k$ out of $n$ experts vote (i), the *average certainty* (or *necessity*) is $k/n$; and if 1 $(1 \geq k)$ experts vote (ii), the *average possibility* is $1/n$. Finally, if the opinion of those experts who believe that there is no submarine in the territorial waters is disregarded, the average certainty and possibility will be $k/m$ and $1/m$, respectively. The disregarding of those experts whose $L_i$ is the empty set corresponds to what is called *normalization* in the Dempster-Shafer theory. As we shall see at a later point, *normalization* can lead to highly counterintuitive results, for it suppresses an important aspect of the experts opinion.

In a more general setting, the opinion of each expert may be weighted, so that the vote of $E_i$ might be multiplied by a number $w_i$, $o \leq w_i \leq 1$, with $w_1 + \ldots + w_n = 1$. In this case, the *average normalized certainty*, $P_*(A)$, and the *average normalized possibility*, $P^*(A)$, may be expressed as

$$P_*(A) = \frac{1}{K}\left(\sum_j w_j\right), E_j \subset A$$

and

$$P^*(A) = \frac{1}{K}\left(\sum_j w_j\right), (E_j \cap A \neq \emptyset$$

where the normalization factor $K$ is given by

$$K = 1 - \sum_j w_j, E_j \cap A = \emptyset.$$

The average normalized certainty and the average normalized possibility are, respectively, the *belief* and *plausibility functions* of Shafer's theory, while the weights $w_1, \ldots, w_n$ are the *basic probability functions* of Shafer's theory, while the weights $w_1, \ldots, w_n$ are the basic probability numbers. (Note that it follows at once from (1) and (2) that $P_*(A) + P^*(\text{not}A) = 1$. This, in a nutshell, is the basic idea underlying the Dempster-Shafer theory.

An important observation is in order at this juncture. If the Minister of Defense had asked the question: What is the probability, $P(A)$, that $S$ is in $A$, the answer would be (after normalization) $P_*(A) \leq P(A) \leq P^*(A)$, where $P_*(A)$ and $P^*(A)$ are the degrees of belief and plausibility associated with $A$. How tight are these bounds, in general?

An approximate answer to this question is provided by what might be called the *principle of information granularity.* Thus, if each $L_i$ is regarded as granule whose "size" is proportional to $W_i$, then the tightness of the bounds bears an inverse relationship to the "size" of $A$, *i.e.*, $P_*(A)$. Thus, if $A$ is regarded as a granule, then the granularity of the $L_i$ must be small compared to that of $A$. This statement may be interpreted as the disposition: "Granularity of data should be finer than the granularity of questions," with the understanding that a disposition is a proposition which is preponderantly, but not necessarily always, true.

Dealing with granular data has an undesirable side effect which is a significant shortcoming of the Dempster-Shafer theory. However, this shortcoming may be overcome by fuzzifying the $L_i$, as is done in Zadeh (1979b).

Specifically, the problem—which, for convenience reference may be labeled the *problem of crisp containment*—is the following. suppose that some particular $L_i$, say $L_a$, violates the containment relation, $L_a \subset A$, by a very small margin. but, no matter how small the margin, so long as the relation is crisp, $E_a$ will have to vote "possibly in $A$" rather than "certainly in $A$." in legal terms, this may be expressed as the dictum "reject any evidence which is not completely certain." the result of such a strict interpretation of containment may be unacceptably loose bounds on $P(A)$ when the size of the $L_i$ is not small compared to that of $A$. However, as shown in Zadeh (1979b), the seriousness of this problem may be reduced by replacing the crisp relation of containment $L_a \subset A$ with a graded relation in which the transition from containment to non-containment is gradual rather than abrupt.

A basic result in the Dempster-Shafer theory which is of direct relevance to the management of uncertainty in expert systems is the so-called Dempster's rule of combination. To understand how this rule works, let us return to the submarine example and assume that there are two groups of experts $E_1, \ldots, E_n$ and $F_1, \ldots, F_r$, with respective weights $w_1, \ldots, w_n$ and $v_1, \ldots, v_r$, and possible locations $L_1, \ldots, L_n$ and $M_1, \ldots, M_r$. Pairing each expert in the first group with an expert in the second group leads to the collection of all possible intersections of the $L_i$ with the $M_j$, $L_i \cap M_j$, $\imath = 1, \ldots, n$, $j = 1, \ldots, r$. Then, according to the Dempster rule of combination of evidence, this combined collection may be treated as a single collection in which the weight associated with $L_i \cap M_j$ is the product $w_{ij} = w_i w_j$. In probabilistic terms, this implies that the bodies of evidence represented by the two groups of experts are independent.

As is pointed out in Zadeh (1979a), the Dempster rule of combination of evidence may lead to counterintuitive conclusions as a result of the application of normalization. The reason for this, as was pointed out earlier, is that normalization throws out the opinion of those experts who assert that the object under consideration does not exist. As a concrete illustration of its effect on the Dempster rule of combination of evidence, consider the following situation. 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. This example and other easily constructed examples call into question the validity of Dempster's rule of combination when it involves a normalization of belief and plausibility.

What is the relevance of the Dempster-Shafer theory to the management of uncertainty in expert systems? The pioneering work of Ted Shortliffe and Bruce Buchanan on MYCIN, and that of Richard Duda and Peter Hart on PROSPECTOR, has made it clear that classical probability techniques are not directly applicable to the derivation of rules of combination of evidence in a rigorous framework, largely because in a typical expert system there are many gaps in the knowledge of the conditional probabilities which are needed for updating the probabilities of hypotheses. Thus, implicitly or explicitly, it becomes necessary to rely on the assumption of independence or, more or less equivalently, on the maximum entropy principle, to fill the gaps with synthetic information. The Dempster-Shafer theory and, more generally, the theory of possibility (Zadeh, 1979b) suggest an alternative approach in which the incompleteness of information in the knowledge base propagates to the conclusion and results in an interval-valued or, more generally, a possibilistic answer. It may be argued that this, in principle, is a more realistic approach because it addresses, rather than finesses, the problem of incomplete information in the knowledge base.

On the other hand, the Dempster-Shafer theory provides a basis—at least at present—for only a small subset of the rules of combination which are needed for inferencing in expert systems. In particular, the theory does not address the issue of chaining, nor does it come to grips with the fuzziness of probabilities and certainty factors.

Thus, although the theory is certainly a step in the right direction, for it provides a framework for dealing with granular data, it does require a great deal of further development to become a broadly useful tool for the management of uncertainty in expert systems.

Although not easy to understand, Shafer's book contains a wealth of significant results, and is a must for anyone who wants to do serious research on problems relating to the rules of combination of evidence in expert systems. Indeed, there is no doubt that, in the years to come, the Dempster-Shafer theory and its extensions will become an integral part of the theory of such systems and will certainly occupy an important place in knowledge engineering and related fields.

### References


Dempster, A. P. (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statistics*, 38:325-339.

Zadeh, L. A. (1979a) Fuzzy sets and information granularity. In Gupta, M., Ragade, R. and Yager, R., (Eds.), *Advances in Fuzzy Set Theory and Applications* Amsterdam: North-Holland, 3-18.

Zadeh, L. A. (1979b) On the validity of Dempster's rule of combination of evidence. *ERL Memorandum M79/24*, University of California, Berkeley, CA.



**Lotfi A. Zadeh**
Computer Science Division
University of California, Berkeley