CHAPTER 7

BELIEF FUNCTIONS

INTRODUCTION

by
Glenn Shafer

The theory of belief functions provides a non-Bayesian way of using mathematical probability to quantify subjective judgements. Whereas a Bayesian assesses probabilities directly for the answer to a question of interest, a belief-function user assesses probabilities for related questions and then considers the implications of these probabilities for the question of interest.

Though antecedents for belief functions can be found in the seventeenth and eighteenth centuries, the theory in its present form is due to the work of A.P. Dempster in the 1960s and my own work in the 1970s. For this reason, it is sometimes called the "Dempster-Shafer theory." It came to the attention of workers in artificial intelligence in the 1980s, in part because of its resemblance to the less systematic calculus of certainty factors developed for MYCIN by Edward Shortliffe and Bruce Buchanan (see Chapter 5).

My 1976 book, *A Mathematical Theory of Evidence*, remains the most comprehensive single reference for the mathematical theory of belief functions, but it has been followed by a large literature on interpretation, application, and computation. This chapter presents a selection from that literature.

The first two articles in the chapter represent recent thinking by Dempster, myself, and our co-workers. The article by Rajendra Srivastava and me introduces the belief-function formalism and compares it with the Bayesian formalism in the context of an important application, financial auditing. The article by Dempster and Augustine Kong illustrates how belief functions can be used in belief networks.

The third and fourth articles, the article by Jean Gordon and Edward Shortliffe, and the article by Judea Pearl, provide alternative introductions to belief functions. Gordon and Shortliffe present the theory in the context of a medical application, and they compare it with the framework and rules of combination of MYCIN. Pearl's article takes a more critical look at belief functions.

The fifth article, by Prakash Shenoy and me, extends to belief functions the methods of local computation in belief networks that are presented for Bayesian networks in Chapter 6. Shenoy and I show that these methods apply to a broad class of formalisms, including the Bayesian formalism, belief functions, Wolfgang Spohn's natural conditional functions, and any other formalism that has operations of combination and marginalization that satisfy certain simple axioms.

The last two articles describe consultation systems that help users represent and combine belief functions. The article by John Lowrance, Thomas Garvey, and Thomas Strat describes an early version of the Gister system, which helps users build up belief networks graphically. This article is significant not only because of Gister itself, but also because Lowrance and his co-workers at SRI International were influential in bringing belief functions into artificial intelligence. The article by Debra Zarley, Yen-Teh Hsia, and me describes DELIEF, a system developed at the University of Kansas. This system goes beyond the early versions of Gister by using the mathematics of belief networks to carry out the combination automatically.

This introduction surveys some of the issues in the belief-function literature and provides references for further reading. For a fuller survey of current issues, see Shafer (1990).

1. The Basics of Belief-Function Theory

The theory of belief functions is based on two ideas: the idea of obtaining degrees of belief for one question from subjective probabilities for a related question, and Dempster's rule for combining such degrees of belief when they are based on independent items of evidence.

These ideas are illustrated by example in several of the articles in this chapter. The simplest, perhaps, is the example of testimony discussed by Srivastava and me. We can derive degrees of belief for statements made by witnesses from subjective probabilities for the reliability of these witnesses.

Degrees of belief obtained in this way differ from probabilities in that they may fail to add to 100%. Suppose, for example, that Betty tells me a tree limb fell on my car. My subjective probability that Betty is reliable is 90%; my subjective probability that she is unreliable is 10%. Since they are probabilities, these numbers add to 100%. But Betty's statement, which must be true if she is reliable, is not necessarily false if she is unreliable. From her testimony alone, I can justify a 90% degree of belief that a limb fell on my car, but only a 0% (not 10%) degree of belief that no limb fell on my car. (This 0% does not mean that I am sure that no limb fell on my car, as a 0% probability would; it merely means that Betty's testimony gives me no reason to believe that no limb fell on my car.) The 90% and the 0%, which do not add to 100%, together constitute a "belief function."

In this example, we are dealing with a question that has only two answers (Did a limb fall on my car? Yes or no.). Belief functions can also be derived for questions for which there are more than two answers. In this case, we will have a degree of belief for each answer and for each set of answers. If the number of answers (or the size of the "frame") is large, the belief function may be very complex.

Dempster's rule is based on the standard idea of probabilistic independence, applied to the questions for which we have subjective probabilities. I can use the rule to combine evidence from two witnesses if I consider the first witness's reliability subjectively independent (before I take account of what the witnesses say) of the second's reliability. (This means that finding out whether one witness is reliable would not change my subjective probability for whether the other is reliable.) The rule uses this subjective independence to determine joint probabilities for various possibilities as to which of the two are reliable.

Though it begins with an initial judgment of independence, Dempster's rule goes beyond this independence. After using independence to compute joint probabilities for who is reliable, I must check whether some possibilities are ruled out by what the witnesses say. (If Betty says that a tree limb fell on my car, and Sally says nothing fell on my car, then they cannot both be reliable.) If so, I renormalize the probabilities of the remaining possibilities so they add to one. This is an example of probabilistic conditioning, and it may destroy the initial independence. (After I notice that Betty and Sally have contradicted each other, their reliabilities are no longer subjectively independent for me. Now finding out that one is reliable would tell me that the other is not.) Then I determine what each possibility for the reliabilities implies about the truth of what the witnesses said, and I use the renormalized probabilities to get new degrees of belief.

The net effect of Dempster's rule is that concordant items of evidence reinforce each other (two independent witnesses for a limb falling on my car make me believe it more than either alone), conflicting items of evidence erode each other, and a chain of reasoning is weaker than its weakest link. Section 2.3 of the article by Srivastava and me illustrates all these points by example.

Bayesian probability measures qualify as belief functions; they represent, in effect, the special case where the unreliable witness always lies. Another special case is that of categorical, completely non-probabilistic information. This is the case where we are 100% confident in the reliability of a witness or other evidence, yet this evidence does not completely answer our question. I might have conclusive evidence, for example, that I lost my wallet in one of three places, without any clue as to which one. This calls for a belief function that assigns a degree of belief of 100% to the three places as a set, but a degree of belief of 0% to each of the three individually.

As Dempster and Kong emphasize in their article in this chapter, the ability to represent both probabilistic and categorical information makes belief functions a bridge between Bayesian and categorical reasoning. Bayesian conditioning itself can be understood in terms of this bridge. Conditioning a Bayesian probability measure on given information is equivalent to combining it, by Dempster's rule, with a categorical belief function representing that information (Shafer 1976, pp. 66-67).

2. Belief Functions Do Not Express Lower Bounds on True but Unknown Probabilities

Mathematically, the degrees of belief given by a single belief function can be related to lower bounds on probabilities, but conceptually they must be sharply distinguished from such lower bounds. If we make up numbers by thinking of them as lower bounds on true probabilities, and we then combine these numbers by Dempster's rule, we are likely to obtain erroneous and misleading results.

It is easy to see the temptation to interpret belief-function degrees of belief as lower bounds on unknown true probabilities. Consider again my 90% belief that a limb fell on my car, and my 0% belief that no limb fell on my car. These degrees of belief were derived from my 90% and 10%

subjective probabilities for Betty being reliable or unreliable. Suppose these subjective probabilities were based on my knowledge of the frequency with which witnesses like Betty are reliable. Then I might think that the 10% of witnesses like Betty who are not reliable make true statements a definite (though unknown) proportion of the time and false statements the rest of the time. Were this the case, I could think in terms of a large population of statements made by witnesses like Betty. In this population, 90% of the statements would be true statements by reliable witnesses, x% would be true statements by unreliable witnesses, and (10-x)% would be false statements by unreliable witnesses, where x is an unknown number between 0 and 10. The total chance of getting a true statement from this population would be (90+x)%, and the total chance of getting a false statement would be (10-x)%. My degrees of belief of 90% and 0% are lower bounds on these chances; since x is anything between 0 and 10, 90% is the lower bound for (90+x)%, and 0% is the lower bound for (10-x)%.

As this example suggests, there is a sense in which a *single* belief function can always be interpreted as a consistent system of probability bounds. It is always possible to find a probability distribution such that each probability is greater than the corresponding degree of belief given by the belief function.

The fallaciousness of the probability-bound interpretation of belief functions becomes clear, however, when we consider two or more belief functions addressing the same question but representing different and possibly conflicting items of evidence. The disagreements that such belief functions represent are not disagreements about the values of true probabilities. When Betty says a limb fell on my car, and Sally says nothing fell on my car, they are disagreeing about whether something fell on my car, not about the true probability of something having fallen on my car.

Were we to insist on a probability-bound interpretation of belief functions, then we would only be interested in groups of belief functions whose degrees of belief, when interpreted as probability bounds, can be satisfied simultaneously. When belief functions are given their proper interpretation, however, it is of no particular significance whether there exist probabilities that simultaneously satisfy the bounds defined by a whole group of belief functions. Consider two cases that might arise when we use belief functions to represent contradictory evidence from Betty and Sally:

> Case 1. Before hearing their testimony, we think highly of the reliability of both Betty and Sally. We represent Betty's evidence by a belief function that gives a 95% degree of belief to a limb having fallen on my car, and we represent Sally's evidence by a belief function that gives a 95% degree of belief to nothing having fallen on my car. In this case, the two belief functions are contradictory as probability bounds; if the true probability of a limb having fallen on my car is greater than 95%, then the true probability of nothing having fallen on my car cannot also be greater than 95%.
>
> Case 2. Before hearing their testimony, we think that both Betty and Sally are fairly unreliable. So in both belief functions, we assign a 35% degree of belief rather than a 95% degree of belief. In this case, the two belief functions define consistent probability bounds; the true probability of a limb having fallen on my car and of nothing having fallen on my car can both be greater than 35%.

From the belief-function point of view, there is no conceptual difference between these two cases. In both cases, we can combine the two belief functions by Dempster's rule. In both cases, there is conflict in the evidence being combined, and normalization is required.

It can be shown that if no renormalization is required in the combination of a group of belief functions by Dempster's rule, then there do exist consistent probabilities that simultaneously bound all the belief functions being combined as well as the belief function that results from the combination. We cannot count on this, however, when renormalization is required. Consequently, authors who favor a probability-bound interpretation of belief functions are uncomfortable with renormalization (see, e.g., Zadeh 1986).

Probability bounds provide the basis for yet another mathematical theory of evidence, which I have called the "theory of lower probability" (Shafer 1981). In this theory, an analogy is drawn between actual evidence and knowledge of bounds on unknown true probabilities for the question of interest. This theory is not always useful, because unknown true probabilities exist only if a population and sampling scheme are well defined. An unknown true probability for the truth of Betty's statement, for example, exists only if a the population of true and false statements of witnesses like Betty is well-defined. In a problem where a reference population for the question of interest is well-defined, the theory of lower probability may be more useful than the theory of belief functions. But in other problems belief functions may be more useful.

As Shafer and Tversky explain in our article in Chapter 2, both Bayesian and belief-function arguments involve analogies to sampling situations. But belief-function analogies are less complete than Bayesian analogies. They are useful when it is reasonable to evaluate certain evidence (e.g., Betty's reputation) using a sampling analogy, but this evidence will not support extending the analogy to all potentially relevant issues (what Betty would say were my good impression of her erroneous, how often limbs fall from that tree, etc.).

There has been some confusion about the original relation between belief functions and probability bounds, because some of Dempster's early articles hinted at a probability-bound interpretation. But Dempster's "upper and lower probabilities" were not derived by bounding unknown true probabilities. My 1976 book, in which the term "belief function" was introduced, explicitly disavowed any probability-bound interpretation (p. ix). This disavowal was elaborated at length in Shafer (1981) and seconded by Dempster (1982).

3. General Metaphors and Canonical Examples

The metaphor of the witness who may or may not be reliable can serve as a standard of comparison, or canonical example, for judging the strength of other evidence. We can assess given evidence by saying that it is comparable in strength to the evidence of a witness who has a certain chance of being reliable.

A witness testifying to a specific proposition leads to a relatively simple belief function—one that gives a specified degree of belief to that proposition and its consequences, and zero degree of belief to all other propositions. Arbitrarily complex belief functions can be built up by combining such simple belief functions (Shafer 1976, p. 200), but in some cases we may want to produce complex belief functions more directly, in order to represent evidence that conveys a complex or mixed message but cannot be broken down into independent components. This requires more complex metaphors or canonical examples.

Two distinct general metaphors have been suggested. Shafer (1981) suggests the metaphor of a randomly coded message. Pearl (1988) suggests the metaphor of random switches.

Shafer's randomly coded messages are explained in detail in the article by Shafer and Tversky in Chapter 2. In this metaphor, we have probabilities for which of several codes was used to encode a message. We do not yet know what the message says, but we know it is true. We have this message in hand in its coded form, and we will try to decode it using each code, but the probabilities are judgments we make before this decoding. When we do decode using the different codes, we sometimes get nonsense, and we sometimes get a comprehensible statement. It seems sensible, in this situation, to condition our probabilities for the codes by eliminating the ones with which we get nonsense. The conditioned probability for each remaining code can then be associated with the statement we get by decoding using that code. These statements may be related in various ways; some may be inconsistent with each other, and some may be stronger than others. Thus we obtain the complexity of an arbitrary belief function.

In this metaphor, the independence of two belief functions means that two different people independently choose codes with which to send two possibly different (though both true) messages. Our uncertainties about the codes in the two cases remain independent unless possible codes imply contradictory messages. If s_1 is a possible code for the first person, and s_2 is a possible code for the second person, and the first message as decoded by s_1 contradicts the second message as decoded by s_2, then it cannot be true that these were the two codes used. We eliminate such pairs of codes and renormalize the probabilities of the remaining possible pairs. The probability of each pair is then associated with the conjunction of the two implied messages. This is Dempster's rule.

The metaphor can be presented in a way that forestalls the interpretation of belief-function degrees of belief in terms of bounds on probabilities. There is no probability model for the choice of the true message sent. The probabilities are only for the choice of codes. We might visualize these probabilities in terms of a repetition of the choice of codes, but since the true message can vary arbitrarily over this population of repetitions, the idea of this population does not lead to the idea of a true unknown probability for the true message or for the true answer to the question of interest. It leads only to an argument about what the true message says or implies—an argument whose strength can be represented in terms of the derived "degrees of belief."

Pearl's metaphor of random switches is explained in detail in his article in this chapter. In this metaphor, a switch oscillates randomly among propositions about the question of interest. Our probabilities are probabilities for the position of the switch, but when the switch points to a certain proposition, this indicates that the proposition is to be adopted as an assumption or axiom for further reasoning. A given proposition about the question of interest may be implied by several of these possible axioms, and its total degree of belief will be the total probability that an axiom implying it is adopted—the total probability, to speak more briefly, that it is proven.

Pearl's metaphor seems well fitted for computer science, since it mixes the language of electrical engineering with that of symbolic logic. I find it difficult, however, to construe the metaphor in a way that completely avoids interpretation in terms of bounds on true probabilities. The random-code metaphor allows us to interpret the probabilities for a related question in terms of a population of repetitions completely unconnected with the question of interest. But since the switch positions in Pearl's metaphor are defined in terms of axioms about the question of interest, it seems to me that each repetition of the random selection of switch positions will generally constrain the true answer to the question of interest. Thus the probabilities do bear directly on the question of interest, and this leads to the objections to renormalization I discussed in the preceding section. In Pearl's view, however, his metaphor is compatible with renormalization (section 5.2 of his article in this section).

4. Sorting Evidence into Independent Items

Dempster's rule should be used to combine belief functions that represent independent items of evidence. But when are items of evidence independent? How can we tell? These are probably the questions asked most frequently about belief functions.

The independence required by Dempster's rule is simply probabilistic independence, applied to the questions for which we have probabilities, rather than directly to the question of interest. In the metaphor of the randomly coded messages, this means that the codes are selected independently. In the more specialized metaphor of independent witnesses, it means that the witnesses (or at least their current properties as witnesses) are selected independently from well-defined populations.

Whether two items of evidence are independent in a real problem is a subjective judgment, in the belief-function as in the Bayesian approach. There is no objective test.

In practice, our task is to sort out the uncertainties in our evidence. When items of evidence are not subjectively independent, we can generally identify what uncertainties they have in common, thus arriving at a larger collection of items of evidence that are subjectively independent. Typically, this maneuver has a cost—it forces us to refine, or make more detailed, the frame over which our belief functions are defined.

We can illustrate this by adapting an example from Pearl's "Bayes Decision Methods," in Chapter 6. Suppose my neighbor Mr. Watson calls me at my office to say he has heard my burglar alarm. In order to assess this testimony in belief-function terms, I assess probabilities for the frame

S_1 = {Watson is reliable, Watson is not reliable}.

Here Watson being reliable means he is honest and he can tell whether it is my burglar alarm he is hearing. I can use these probabilities to get degrees of belief for the frame

T = {My alarm sounded, My alarm did not sound}.

Putting a probability of 90%, say, on Watson being reliable, I get a 90% degree of belief that my burglar alarm sounded, and a 0% degree of belief that my burglar alarm did not sound.

I now call another neighbor, Mrs. Gibbons, who verifies that my alarm sounded. I can assess her testimony in the same way, by assessing probabilities for the frame

S_2 = {Gibbons is reliable, Gibbons is not reliable}.

Suppose I also put a probability of 95% on Gibbons being reliable, so that I again obtain a 95% degree of belief that my burglar alarm sounded, and a 0% degree of belief that it did not sound.

Were I to combine these two belief functions by Dempster's rule, I would obtain an overall degree of belief of 99.5% that my burglar alarm sounded. This is inappropriate, however, for the two items of evidence involve a common uncertainty—whether there might have been some other noise similar to my burglar alarm.

In order to deal with this problem, I must pull my skepticism about the possibility of a similar noise out of my assessment of Watson's and Gibbons' reliability, and identify my grounds for this skepticism as a separate item of evidence. So I now have three items of evidence—my evidence for Watson's honesty (I say honesty now instead of reliability, since I am not including here the judgment that there are no other potential noises in the neighborhood that Watson might confuse with my burglar alarm), my evidence for Gibbons' honesty, and my evidence that there are no potential noises in the neighborhood that sound like my burglar alarm.

These three items of evidence are now independent, but their combination involves more than the frame T. In its place, we need the frame U = $\{u_1,u_2,u_3\}$, where

u_1 = My alarm sounded,
u_2 = There was a similar noise,
u_3 = There was no noise.

(Let us exclude, for simplicity of exposition, the possibility that there were two noises, my alarm and also a similar noise.) My first two items of evidence (my evidence for Watson's and Gibbons' honesty) both provide a high degree of belief in $\{u_1,u_2\}$, while the third item (my evidence against the existence of other noise sources) provides a high degree of belief in $\{u_1,u_3\}$. Combining the three by Dempster's rule produces a high degree of belief in $\{u_1\}$.

A Bayesian approach to this problem would be somewhat different, but it too would involve refining the frame T to U or something similar. In the Bayesian case, we would ask whether the events "Watson says he heard a burglar alarm" and "Gibbons says she heard a burglar alarm" are subjectively independent. They are not unconditionally independent, but they are independent conditional on a specification of what noise actually occurred. I can exploit this conditional independence in assessing my subjective probabilities, but in order to do so, I must bring the possibility of other noises into the frame.

In the belief-function approach, one talks not about conditional independence of propositions, but rather about the overlapping and interaction of evidence. For further explanation and more examples, see Shafer (1976, Chapter 8), Shafer (1981), Shafer (1984), Shafer (1987) and Srivastava, Shenoy, and Shafer (1989).

5. Frequency Thinking

A common but dangerous temptation is to use Dempster's rule to combine opinions that are really fragments of information about a single probability distribution. This is usually inappropriate and can give misleading results.

Suppose we are concerned with a bird Tweety. We want to know whether Tweety flies and whether Tweety is a penguin. We decide to make judgments about this by thinking of Tweety as randomly selected from a certain population of birds. We have guesses about the proportion of birds in this population that fly and the proportion that are penguins. Should these guesses be represented as belief functions over a set of statements about Tweety, and then combined by Dempster's rule? No. Both guesses bear on the particular bird only through their accuracy as guesses about the population. This means that they have in common the uncertainty involved in choosing Tweety at random from the population. Depending on how we obtained the guesses, they may also have other uncertainties in common.

Like every problem of dependence, we can deal with this problem within the belief-function approach by sorting out the uncertainties and properly refining our frame. In this case, we must bring the possible values for the population frequencies into the frame. We can then formalize the connection between these frequencies and Tweety as one of our items of evidence. We must also identify our sources of evidence about the frequencies, sort our their uncertainties, and use them to assess belief functions about what these frequencies are.

This brings us into the difficult realm of statistical inference. Belief functions provide only one of many approaches to statistical inference, and even the possible belief-function techniques are varied and complex (Dempster 1966, 1967a,b, 1968a,b, 1969; Chapter 11 of Shafer 1976; Shafer 1982a,b). No statistical approach is likely to be of much value unless considerable frequency information is available.

An alternative is the Bayesian approach, in which (according to the constructive interpretation) we compare our hunches about Tweety to fairly precise knowledge of frequencies in a population. One of the messages of Pearl's article on probabilistic semantics in Chapter 9, with which I agree, is that many of the intuitions discussed in the literature on non-monotonic logic are best represented in this way, in spite of the authors' protestations that their intuitions are not probabilistic.

In section 7 of his article in this chapter, Pearl gives a number of examples in which Dempster's rule gives inappropriate results. I believe the intuitions in most of these examples are based on frequency thinking, and I agree that Dempster's rule is not appropriate for such problems. If there is sufficient evidence on which to base detailed probability judgments in analogy to frequency knowledge, then the Bayesian approach will be far more useful. In some other cases, a lower probability approach will be useful. Belief functions are not likely to be useful unless the focus can be shifted from frequency thinking to the assessment of specific items of evidence.

6. Computation

It has often been pointed out that the computation involved in Dempster's rule can be prohibitive. As the number of possibilities in the frame T grows, the computation required to combine arbitrary belief functions on T grows exponentially.

Barnett (1981) was the first to address this problem; he showed how to compute Dempster's rule in linear time in the special case where each belief function supports only a single element of T. Unfortunately, this condition is not often met. Even if it is met in an initial formulation of a problem, it may fail to be met in a more careful formulation. A more careful formulation may require us to split elements of T. (In the example in Section 4 above, we split “My alarm did not sound” into two possibilities, “There was a similar noise” and “There was no noise.”) Thus what appears as a single element at one stage may appear as a set of elements at a later stage.

More recent work on computation has focused on exploiting patterns of interaction in evidence to reduce computation on very large frames to computation on many smaller frames. This is completely analogous to the Bayesian exploitation of conditional independence (see the articles in this chapter by Dempster and Kong and by Shenoy and myself). The smaller frames involve variables that are closely tied to one another, either by definition or by evidence. The results of “local” computations involving one small cluster of variables can affect beliefs about distant variables, but only after being “propagated” to them through other clusters. The clusters together form a belief network.

In both the belief-function and Bayesian cases, belief networks are more than merely computational tools. They are also conceptual tools, tools that we use in sorting out our evidence. In the Bayesian case, they provide a graphical representation of the conditional independence structure that is needed to make probability judgments manageable. In the belief-function case, they provide a representation of the sorting into independent uncertainties discussed in section 4 above.

Work on belief-function networks was initiated by the work of Gordon and Shortliffe (1985) on the problem of combining evidence by Dempster's rule when different items of evidence are relevant to different levels of specificity in a hierarchy of diseases, which they mention at the end of their article in this chapter. Gordon and Shortliffe's method for approximating Dempster's rule in this case was strengthened by Shafer and Logan (1987) to an efficient method of computing the exact results of the rule. Pearl discusses the analogous Bayesian approach in his article, “On evidential reasoning in a hierarchy of hypotheses,” in Chapter 6.

The belief network defined by a hierarchy is basically a tree, and the computations are simplest in this case. The article by Shenoy and Shafer in this chapter is concerned primarily with propagation in trees. The problem of belief networks that are not trees was addressed by Kong (1986), who developed methods similar to those later used for the Bayesian case by Lauritzen and Spiegelhalter (Chapter 6). As Dempster and Kong explain in their article in this chapter, these methods can be thought of as representing the network as a “tree of cliques” and then applying propagation to this tree.

7. Implementation

Belief functions have been implemented in a number of expert systems, and it is beyond the scope of this introduction to evaluate or even list all these implementations. It should be pointed out, however, that the implementations that are discussed in this chapter are designed for interactive use. Neither Gister nor DELIEF is a fully automatic system for evidential reasoning. Instead, they are systems that help human users build and evaluate belief networks. They require the user to make the judgments of independence that justify the network and to provide the numerical judgments of support based on each item of evidence.

Interactive systems seem appropriate to belief functions. As we saw in sections 2 and 3 of this introduction, belief functions are appropriate to situations where the direct application of probability is not useful—situations where different populations of repetitions, real or imagined, justify probability judgments for different items of evidence, and these populations bear on the question of interest and interact with each other in ways unique to each application. In the example of section 2, the relation between the question for which we have probabilities (whether Betty is reliable) and the question of interest (whether a limb fell on my car) is defined by the particular testimony given by Betty. So in general, the belief network and the numerical judgments must be constructed anew for case.

In contrast, some probabilistic systems—such as the HUGIN system for medical diagnosis described in Chapter 5—apply the same conditional independence structure and, for the most part, the same numerical judgments (or frequencies) to each new case. In this case, we are relating the entire structure of the evidence in each case to the same population of repetitions, and this is a process we can hope to automate more fully.

8. Pearl's Critique

The article by Judea Pearl in this chapter provides a comprehensive critique of the belief-function formalism from a sympathetic but ultimately skeptical Bayesian. In the end, Pearl prefers probability-bound approaches. I conclude this section with a brief response to Pearl's critique.

One of the purposes of Pearl's article is to introduce his random switch metaphor for belief functions. As I indicated in section 2, I do not find this metaphor helpful in explaining Dempster's rule. I also doubt that calling the degrees of belief given by belief functions "probabilities of necessity" will be helpful in delimiting the applications of the theory. By calling attention to the necessity of logical relations, this phrase distracts from the contingent nature of the relation between questions that is typical of belief-function applications.

The propositions

A = Betty is reliable

and

B = Betty says a limb fell on my car,

properly translated into a formal language, will make the proposition

C = A limb fell on my car

logically necessary. So if we think of B as fixed, we can say that A's probability contributes to C's probability of being necessary. But it is best not to think of B as fixed. It is fixed for the particular belief-function calculation, but as I emphasized in the preceding section, it is contingent relative to the population that defines the probabilities for A, and we must bear this contingency in mind if we are to avoid the fallacies that Pearl exposes in section 7 of his article.

Pearl's examples of illegitimate and misleading uses of Dempster's rule are extremely valuable. They are not mere straw men, for they correspond to uses of Dempster's rule that have been suggested and implemented. Pearl's discussion should clarify the general principles governing the use of Dempster's rule that I discussed in sections 4 and 5 above.

Though Pearl and I agree on specific examples, I do not agree with his claim that the Bayesian formalism is more useful than the belief-function formalism because it is concerned directly with probabilities for questions of interest (this is my understanding of the remark, in section 5.2 of his article, that we should be concerned with the probability of truth). While I do want to know about the probability of the question that interests me when I think there is such a probability, I do not always think there is. In order to define the probability that a person has a disease, we must either specify a population or else specify other evidence and draw an analogy between that evidence and the situation where we draw a person at random from a population. If we succeed in this, then we may speak of the probability that the person has the disease; if we do not succeed, then there simply is no probability.

Both the Bayesian and the belief-function formalisms are tools for the assessment of evidence. Both draw analogies between actual evidence and certain idealized canonical examples. The Bayesian analogy is more familiar, but this does not always guarantee its success.

References

Barnett, J.A. (1981). Computational methods for a mathematical theory of evidence. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence,* Vancouver, pp. 868-875.

Dempster, A.P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics* **37** 355-374.

Dempster, A.P. (1967a). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* **38** 325-339.

Dempster, A.P. (1967b). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* **54** 515-528.

Dempster, A.P. (1968a). Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics* **39** 957-966.

Dempster, A.P. (1968b). A generalization of Bayesian inference (with discussion). *Journal of the Royal Statistical Society, Series B* **30** 205-247.

Dempster, A.P. (1969). Upper and lower probability inferences for families of hypotheses with monotone density ratios. *Annals of Mathematical Statistics* **40** 953-969.

Dempster, A.P. (1982). Comment. *Journal of the American Statistical Association* **77** 339-341.

Gordon, Jean, and Edward H. Shortliffe (1985). A method for managing evidential reasoning in hierarchical hypothesis spaces. *Artificial Intelligence* **26** 323-358.

Kong, Augustine (1986). *Multivariate Belief Functions and Graphical Models*. Doctoral dissertation. Department of Statistics. Harvard University.

Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, California.

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

Shafer, Glenn (1981). Constructive probability. *Synthese* **48** 1-60.

Shafer, Glenn (1982a). Lindley's paradox (with discussion). *Journal of the American Statistical Association* **77** 325-351.

Shafer, Glenn (1982b). Belief functions and parametric models (with discussion). *Journal of the Royal Statistical Society, Series B* **44** 322-352.

Shafer, Glenn (1984). The problem of dependent evidence. Working paper no. 164, K.U. School of Business.

Shafer, Glenn (1987). Belief functions and possibility measures. *The Analysis of Fuzzy Information, Vol. 1: Mathematics and Logic*, pp. 51-84. James C. Bezdek, ed., CRC Press.

Shafer, Glenn (1990). Perspectives on the theory and practice of belief functions. To appear in the *International Journal of Approximate Reasoning*.

Shafer, Glenn, and Roger Logan (1987). Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence* **33** 271-298.

Srivastava, Rajendra P., Prakash P. Shenoy, and Glenn Shafer (1989). Audit risk: A belief-function approach. Working Paper No. 212, School of Business, University of Kansas.

Zadeh, Lotfi A. (1986). A simple view of the Dempster-Shafer theory of evidence and its implications for the rule of combination. *AI Magazine* **7:2** 85-90.

Articles for Chapter 7

1. "The Bayesian and Belief-Function Formalisms; A General Perspective for Auditing," by Glenn Shafer and Rajendra Srivastava. To appear in *Auditing: A Journal of Practice and Theory*, 1990.
2. "Uncertain Evidence and Artificial Analysis," by A.P. Dempster and Augustine Kong. (1988) *Journal of Statistical Planning and Inference* **20** 355-368.
3. "The Dempster-Shafer Theory of Evidence," by Jean Gordon and Edward H. Shortliffe. (1984) In *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, edited by B.G. Buchanan and E.H. Shortliffe. Addison-Wesley.
4. Pearl, J. (1988) Bayesian and Belief-Function Formalisms for Evidential Reasoning: A Conceptual Analysis. *Proceedings, 5th Israeli Symposium on Artificial Intelligence,* Tel Aviv, December 1988.
5. "Axioms for Probability and Belief-Function Propagation," by Prakash P. Shenoy and Glenn Shafer. To appear in *Uncertainty in Artificial Intelligence*, Vol. 4, edited by Ross Schachter, Tod S. Levitt, John F. Lemmer, and Laveen N. Kanal, Machine Intelligence and Pattern Recognition Series, North-Holland, 1990.
6. Lowrance, John D. Thomas D. Garvey, and Thomas M. Strat (1989). A Framework for Evidential-Reasoning Systems. *AAAI-86*, pp. 896-903.
7. Zarley Debra, Yen-Teh Hsia, and Glenn Shafer (1988). Evidential Reasoning using DELIEF. *AAAI-88; Proceedings of the Seventh National Conference on Artificial Intelligence*, St. Paul, Minnesota, pp. 205-209.