

13

The Dempster-Shafer Theory of Evidence

Jean Gordon and Edward H. Shortliffe

The drawbacks of pure probabilistic methods and of the certainty factor model have led us in recent years to consider alternate approaches. Particularly appealing is the mathematical theory of evidence developed by Arthur Dempster. We are convinced it merits careful study and interpretation in the context of expert systems. This theory was first set forth by Dempster in the 1960s and subsequently extended by Glenn Shafer. In 1976, the year after the first description of CF's appeared, Shafer published *A Mathematical Theory of Evidence* (Shafer, 1976). Its relevance to the issues addressed in the CF model was not immediately recognized, but recently researchers have begun to investigate applications of the theory to expert systems (Barnett, 1981; Friedman, 1981; Garvey et al., 1981).

We believe that the advantage of the Dempster-Shafer theory over previous approaches is its ability to model the narrowing of the hypothesis set with the accumulation of evidence, a process that characterizes diagnostic reasoning in medicine and expert reasoning in general. An expert uses evidence that, instead of bearing on a single hypothesis in the original hypothesis set, often bears on a larger subset of this set. The functions and combining rule of the Dempster-Shafer theory are well suited to represent this type of evidence and its aggregation.

For example, in the search for the identity of an infecting organism, a smear showing gram-negative organisms narrows the hypothesis set of all possible organisms to a proper subset. This subset can also be thought of as a new hypothesis: the organism is one of the gram-negative organisms. However, this piece of evidence gives no information concerning the relative likelihoods of the organisms in the subset. Bayesians might assume equal priors and distribute the weight of this evidence equally among the gram-negative organisms, but, as Shafer points out, they would thus fail to distinguish between uncertainty, or lack of knowledge, and

equal certainty. Because he attributes belief to subsets, as well as to individual elements of the hypothesis set, we believe that Shafer more accurately reflects the evidence-gathering process.

A second distinct piece of evidence, such as morphology of the organism, narrows the original hypothesis set to a different subset. How does the Dempster-Shafer theory pool these two pieces of evidence? Each is represented by a belief function, and two belief functions are merged via a combination rule to yield a new function. The combination rule, like the Bayesian and CF combining functions, is independent of the order in which evidence is gathered and requires that the hypotheses under consideration be mutually exclusive and exhaustive. In fact, the Dempster-Shafer combination rule includes the Bayesian and CF functions as special cases.

Another consequence of the generality of the Dempster-Shafer belief functions is avoidance of the Bayesian restriction that commitment of belief to a hypothesis implies commitment of the remaining belief to its negation, i.e., that $P(h) = 1 - P(\neg h)$. The concept that, in many situations, evidence partially in favor of a hypothesis should not be construed as evidence partially against the same hypothesis (i.e., in favor of its negation) was one of the desiderata in the development of the CF model, as discussed in Chapter 11. As in the CF model, the beliefs in each hypothesis in the original set need not sum to 1 but may sum to a number less than or equal to 1; some of the belief can be allotted to subsets of the original hypothesis set.

Thus the Dempster-Shafer model includes many of the features of the CF model but is based on a firm mathematical foundation. This is a clear advantage over the *ad hoc* nature of CF's. In the next sections, we motivate the exposition of the theory with a medical example and then discuss the relevance of the theory to MYCIN.

13.1 Basics of the Dempster-Shafer Theory

13.1.1 A Simple Example of Medical Reasoning

Suppose a physician is considering a case of cholestatic jaundice for which there is a diagnostic hypothesis set of hepatitis (hep), cirrhosis (cirr), gallstone (gall) and pancreatic cancer (pan). There are, of course, more than four causes of jaundice, but we have simplified the example here for illustrative purposes. In the Dempster-Shafer theory, this set is called a *frame of discernment*, denoted Θ . As noted earlier, the hypotheses in Θ are assumed mutually exclusive and exhaustive.

One piece of evidence considered by the physician might support the diagnosis of intrahepatic cholestasis, which is defined for this example as

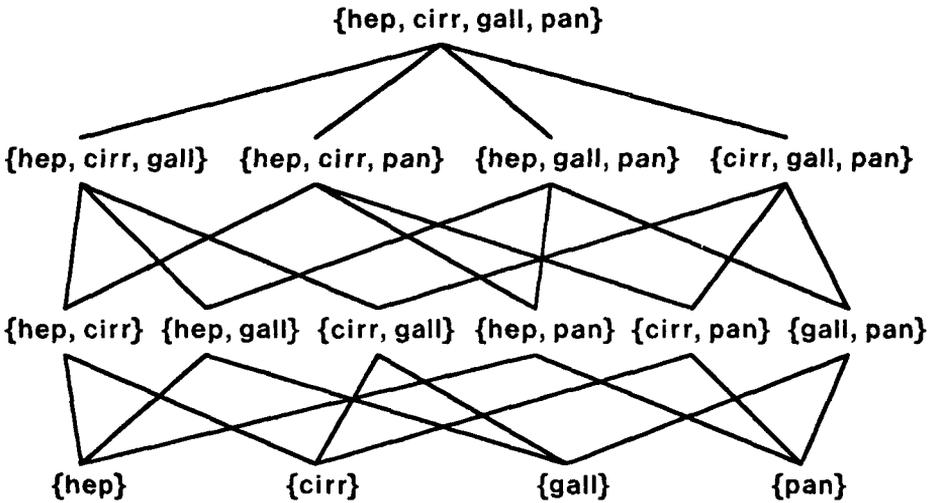


FIGURE 13-1 The subsets of the set of causes of cholestasis.

the two-element subset of Θ $\{\text{hep, cirr}\}$, also represented by the hypothesis HEP-OR-CIRR. Similarly, the hypothesis extrahepatic cholestasis corresponds to $\{\text{gall, pan}\}$. Evidence confirming intrahepatic cholestasis to some degree will cause the physician to allot a proportional amount of belief to that subset.

A new piece of evidence might help the physician exclude hepatitis to some degree. Evidence disconfirming hepatitis (HEP) is equivalent to evidence confirming the hypothesis NOT-HEP, which corresponds to the hypothesis CIRR-OR-GALL-OR-PAN or the subset $\{\text{cirr, gall, pan}\}$. Thus evidence disconfirming hepatitis to some degree will cause the physician to allot a proportional amount of belief to this three-element subset.

As illustrated above, a subset of hypotheses in Θ gives rise to a new hypothesis, which is equivalent to the disjunction of the hypotheses in the subset. Each hypothesis in Θ corresponds to a one-element subset (called a *singleton*). By considering all possible subsets of Θ , denoted 2^Θ , the set of hypotheses to which belief can be allotted is enlarged. Henceforth, we use the term *hypothesis* in this enlarged sense to denote any subset of the original hypotheses in Θ .

A pictorial representation of 2^Θ is given in Figure 13-1. Note that a set of size n has 2^n subsets. (The empty set, \emptyset , is one of these subsets, but corresponds to a hypothesis known to be false and is not shown in Figure 13-1.

In a given domain, only some subsets in 2^Θ will be of diagnostic interest. Evidence often bears on certain disease categories as well as on specific disease entities. In the case of cholestatic jaundice, evidence available to

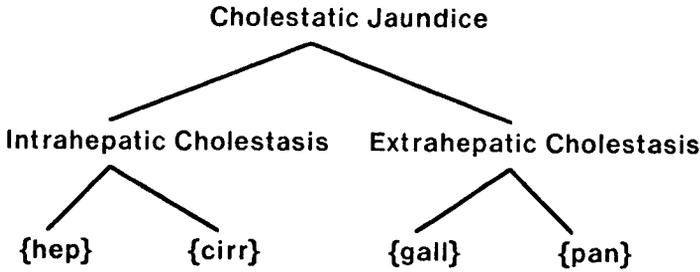


FIGURE 13-2 The subsets of clinical interest in cholestatic jaundice.

the physician tends to support either intrahepatic cholestasis, extrahepatic cholestasis, or the singleton hypotheses. The tree of Figure 13-1 can thus be pruned to that of Figure 13-2, which summarizes the hierarchical relations of clinical interest. In at least one medical artificial intelligence system, the causes of jaundice have been usefully structured in this way for the diagnostic task (Chandrasekharan et al., 1979).

13.1.2 Basic Probability Assignments

The Dempster-Shafer theory uses a number in the range $[0,1]$ to indicate belief in a hypothesis given a piece of evidence. This number is the degree to which the evidence supports the hypothesis. Recall that evidence against a hypothesis is regarded as evidence for the negation of the hypothesis. Thus, unlike the CF model, the Dempster-Shafer model avoids the use of negative numbers.

The impact of each distinct piece of evidence on the subsets of Θ is represented by a function called a *basic probability assignment* (bpa). A bpa is a generalization of the traditional probability density function; the latter assigns a number in the range $[0,1]$ to every singleton of Θ such that the numbers sum to 1. Using 2^Θ , the enlarged domain of all subsets of Θ , a bpa denoted m assigns a number in $[0,1]$ to every subset of Θ such that the numbers sum to 1. (By definition, the number 0 must be assigned to the empty set, since this set corresponds to a false hypothesis. It is false because the hypotheses in Θ are assumed exhaustive.) Thus m allows assignment of a quantity of belief to every element in the tree of Figure 13-1, not just to those elements on the bottom row, as is the case for a probability density function.

The quantity $m(A)$ is a measure of that portion of the total belief committed exactly to A , where A is an element of 2^Θ and the total belief is 1. This portion of belief cannot be further subdivided among the subsets of A and does not include portions of belief committed to subsets of A . Since

belief in a subset certainly entails belief in subsets containing that subset (i.e., nodes “higher” in the network of Figure 13-1), it would be useful to define a function that computes a total amount of belief in A . This quantity would include not only belief committed exactly to A but belief committed to all subsets of A . Such a function, called a *belief function*, is defined in the next section.

The quantity $m(\Theta)$ is a measure of that portion of the total belief that remains unassigned after commitment of belief to various proper subsets of Θ . For example, evidence favoring a single subset A need not say anything about belief in the other subsets. If $m(A) = s$ and m assigns no belief to other subsets of Θ , then $m(\Theta) = 1 - s$. Thus the remaining belief is assigned to Θ and not to the negation of the hypothesis (equivalent to A^c , the set-theoretic complement of A), as would be required in the Bayesian model.

Examples

Example 1. Suppose that there is no evidence concerning the specific diagnosis in a patient with known cholestatic jaundice. The bpa representing ignorance, called the vacuous bpa, assigns 1 to $\Theta = \{\text{hep, cirr, gall, pan}\}$ and 0 to every other subset of Θ . Bayesians might attempt to represent ignorance by a function assigning 0.25 to each singleton, assuming no prior information. As remarked before, such a function would imply more information given by the evidence than is truly the case.

Example 2. Suppose that the evidence supports, or confirms, the diagnosis of intrahepatic cholestasis to the degree 0.6, but does not support a choice between cirrhosis and hepatitis. The remaining belief, $1 - 0.6 = 0.4$, is assigned to Θ . The hypothesis corresponding to Θ is known to be true under the assumption of exhaustiveness. Bayesians would assign the remaining belief to extrahepatic cholestasis, the negation of intrahepatic cholestasis. Such an assignment would be an example of Paradox 1, discussed in Chapter 11. Thus $m(\{\text{hep, cirr}\}) = 0.6$, $m(\Theta) = m(\{\text{hep, cirr, gall, pan}\}) = 0.4$, and the value of m for every other subset of Θ is 0.

Example 3. Suppose that the evidence disconfirms the diagnosis of hepatitis to the degree 0.7. This is equivalent to confirming that of NOT-HEP to the degree 0.7. Thus $m(\{\text{cirr, gall, pan}\}) = 0.7$, $m(\Theta) = 0.3$, and the value of m for every other subset of Θ is 0.

Example 4. Suppose that the evidence confirms the diagnosis of hepatitis to the degree 0.8. Then $m(\{\text{hep}\}) = 0.8$, $m(\Theta) = 0.2$, and m is 0 elsewhere.

13.1.3 Belief Functions

A belief function, denoted Bel , corresponding to a specific bpa, m , assigns to every subset A of Θ the sum of the beliefs committed exactly to every subset of A by m . For example,

$$\begin{aligned} \text{Bel}(\{\text{hep, cirr, pan}\}) = & m(\{\text{hep, cirr, pan}\}) + m(\{\text{hep, cirr}\}) \\ & + m(\{\text{hep, pan}\}) + m(\{\text{cirr, pan}\}) \\ & + m(\{\text{hep}\}) + m(\{\text{cirr}\}) + m(\{\text{pan}\}) \end{aligned}$$

Thus, $\text{Bel}(A)$ is a measure of the total amount of belief in A and not of the amount committed precisely to A by the evidence giving rise to m .

Referring to Figure 13-1, Bel and m are equal for singletons, but $\text{Bel}(A)$, where A is any other subset of Θ , is the sum of the values of m for every subset in the subtree formed by using A as the root. $\text{Bel}(\Theta)$ is always equal to 1 since $\text{Bel}(\Theta)$ is the sum of the values of m for every subset of Θ . This sum must be 1 by definition of a bpa. Clearly, the total amount of belief in Θ should be equal to the total amount of belief, 1, since the singletons are exhaustive.

To illustrate, the belief function corresponding to the bpa of Example 2 is given by $\text{Bel}(\Theta) = 1$, $\text{Bel}(A) = 0.6$, where A is any proper subset of Θ containing $\{\text{hep, cirr}\}$, and the value of Bel for every other subset of Θ is 0.

13.1.4 Combination of Belief Functions

As discussed in Chapter 11, the evidence-gathering process in medical diagnosis requires a method for combining the support for a hypothesis, or for its negation, based on multiple, accumulated observations. The Dempster-Shafer model also recognizes this requirement and provides a formal proposal for its management. Given two belief functions, based on two observations, but with the same frame of discernment, Dempster's combination rule, shown below, computes a new belief function that represents the impact of the combined evidence.

Concerning the validity of this rule, Shafer (1976) writes that although he can provide "no conclusive *a priori* argument, . . . it does seem to reflect the pooling of evidence." In the special case of a frame of discernment containing two elements, Dempster's rule can be found in Johann Heinrich Lambert's book, *Neues Organon*, published in 1764. In another special case where the two bpa's give support to exactly one and the same hypothesis, the rule reduces to that found in the MYCIN CF model and in *Ars Conjectandi*, the work of the mathematician Jakob Bernoulli in 1713.

The Dempster combination rule differs from the MYCIN combining function in the pooling of evidence supporting mutually exclusive hypotheses. For example, evidence supporting hepatitis reduces belief in each

of the singleton hypotheses—CIRR, GALL, and PAN—and in any disjunction not containing HEP, e.g., CIRR-OR-GALL-OR-PAN, NOT-HEP, CIRR-OR-PAN, etc. As we discuss later, if the Dempster-Shafer model were adapted for use in MYCIN, each new piece of evidence would have a wider impact on other hypotheses than it does in the CF model. The Dempster combination rule also gives rise to a very different result regarding belief in a hypothesis when confirming and disconfirming evidence is pooled.

Let Bel_1 and Bel_2 and m_1 and m_2 denote two belief functions and their respective bpa's. Dempster's rule computes a new bpa, denoted $m_1 \oplus m_2$, which represents the combined effect of m_1 and m_2 . The corresponding belief function, denoted $\text{Bel}_1 \oplus \text{Bel}_2$, is then easily computed from $m_1 \oplus m_2$ by the definition of a belief function.

If we sum all products of the form $m_1(X)m_2(Y)$, where X and Y run over all subsets of Θ , the result is 1 by elementary algebra and the definition of a bpa:

$$\sum m_1(X)m_2(Y) = \sum m_1(X) \sum m_2(Y) = 1 \times 1 = 1 \quad (1)$$

The bpa representing the combination of m_1 and m_2 apportions this number 1, the total amount of belief, among the subsets of Θ by assigning $m_1(X)m_2(Y)$ to the intersection of X and Y . Note that there are typically several different subsets of Θ whose intersection equals that of X and Y . Thus, for every subset A of Θ , Dempster's rule defines $m_1 \oplus m_2(A)$ to be the sum of all products of the form $m_1(X)m_2(Y)$, where X and Y run over all subsets whose intersection is A . The commutativity of multiplication ensures that the rule yields the same value regardless of the order in which the functions are combined. This is an important property since evidence aggregation should be independent of the order of its gathering. The following two examples illustrate the combination rule.

Example 5. As in Examples 2 and 3, suppose that for a given patient one observation supports intrahepatic cholestasis to degree 0.6 (m_1) whereas another disconfirms hepatitis (i.e., confirms {cirr, gall, pan}) to degree 0.7 (m_2). Then our net belief based on both observations is given by $m_1 \oplus m_2$. For computational purposes, an "intersection tableau" with values of m_1 and m_2 along the rows and columns, respectively, is a helpful device. Only nonzero values of m_1 and m_2 need be considered, since if $m_1(X)$ and/or $m_2(Y)$ is 0, then the product $m_1(X)m_2(Y)$ contributes 0 to $m_1 \oplus m_2(A)$, where A is the intersection of X and Y . Entry i,j in the tableau is the intersection of the subsets in row i and column j . Clearly, some of these entries may be the same subset. The product of the bpa values is in parentheses next to the subset. The value of $m_1 \oplus m_2(A)$ is computed by summing all products in the tableau adjacent to A .

		m_2	
		{cirr, gall, pan} (0.7)	Θ (0.3)
	m_1	{hep, cirr} (0.6)	{cirr} (0.42)
			{hep, cirr} (0.18)
			{cirr, gall, pan} (0.28)
			Θ (0.12)

In this example, a subset appears only once in the tableau and $m_1 \oplus m_2$ is easily computed:

$$\begin{aligned}
 m_1 \oplus m_2(\{\text{cirr}\}) &= 0.42 \\
 m_1 \oplus m_2(\{\text{hep, cirr}\}) &= 0.18 \\
 m_1 \oplus m_2(\{\text{cirr, gall, pan}\}) &= 0.28 \\
 m_1 \oplus m_2(\Theta) &= 0.12 \\
 m_1 \oplus m_2 &\text{ is 0 for all other subsets of } \Theta
 \end{aligned}$$

Since $\text{Bel}_1 \oplus \text{Bel}_2$ is fairly complex, we give only a few sample values:

$$\begin{aligned}
 \text{Bel}_1 \oplus \text{Bel}_2(\{\text{hep, cirr}\}) &= m_1 \oplus m_2(\{\text{hep, cirr}\}) + m_1 \oplus m_2(\{\text{hep}\}) \\
 &\quad + m_1 \oplus m_2(\{\text{cirr}\}) \\
 &= 0.18 + 0 + 0.42 \\
 &= 0.60
 \end{aligned}$$

$$\begin{aligned}
 \text{Bel}_1 \oplus \text{Bel}_2(\{\text{cirr, gall, pan}\}) &= m_1 \oplus m_2(\{\text{cirr, gall, pan}\}) \\
 &\quad + m_1 \oplus m_2(\{\text{cirr, gall}\}) \\
 &\quad + m_1 \oplus m_2(\{\text{cirr, pan}\}) \\
 &\quad + m_1 \oplus m_2(\{\text{gall, pan}\}) + m_1 \oplus m_2(\{\text{cirr}\}) \\
 &\quad + m_1 \oplus m_2(\{\text{gall}\}) + m_1 \oplus m_2(\{\text{pan}\}) \\
 &= 0.28 + 0 + 0 + 0 + 0.42 + 0 + 0 \\
 &= 0.70
 \end{aligned}$$

$$\text{Bel}_1 \oplus \text{Bel}_2(\{\text{hep, cirr, pan}\}) = \text{Bel}_1 \oplus \text{Bel}_2(\{\text{hep, cirr}\}) = 0.60$$

since

$$m_1 \oplus m_2(\{\text{hep, cirr, pan}\}) = m_1 \oplus m_2(\{\text{hep, pan}\}) = m_1 \oplus m_2(\{\text{cirr, pan}\}) = 0$$

In this example, the reader should note that $m_1 \oplus m_2$ satisfies the definition of a bpa: $\sum m_1 \oplus m_2(X) = 1$, where X runs over all subsets of Θ and $m_1 \oplus m_2(\emptyset) = 0$. Equation (1) shows that the first condition in the definition is always fulfilled. However, the second condition is problematic in cases where the “intersection tableau” contains null entries. This situation did not occur in Example 5 because every two sets with nonzero bpa values always had at least one element in common. In general, nonzero products

of the form $m_1(X)m_2(Y)$ may be assigned when X and Y have an empty intersection.

Dempster deals with this problem by normalizing the assigned values so that $m_1 \oplus m_2(\emptyset) = 0$ and all values of the new bpa lie between 0 and 1. This is accomplished by defining κ as the sum of all nonzero values assigned to \emptyset in a given case ($\kappa = 0$ in Example 5). Dempster then assigns 0 to $m_1 \oplus m_2(\emptyset)$ and divides all other values of $m_1 \oplus m_2$ by $1 - \kappa$.¹

Example 6. Suppose now that, for the same patient as in Example 5, a third observation (m_3) confirms the diagnosis of hepatitis to the degree 0.8 (cf. Example 4). We now need to compute $m_3 \oplus m_4$, where $m_4 = m_1 + m_2$ of Example 5.

		$m_4 = m_1 \oplus m_2$			
		{cirr} (0.42)	{hep, cirr} (0.18)	{cirr, gall, pan} (0.28)	Θ (0.12)
m_3	{hep} (0.8)	\emptyset (0.336)	{hep} (0.144)	\emptyset (0.224)	{hep} (0.096)
	Θ (0.2)	{cirr} (0.084)	{hep, cirr} (0.036)	{cirr, gall, pan} (0.056)	Θ (0.024)

In this example, there are two null entries in the tableau, one assigned the value 0.336 and the other 0.224. Thus

$$\begin{aligned} \kappa &= 0.336 + 0.224 = 0.56 \text{ and } 1 - \kappa = 0.44 \\ m_3 \oplus m_4(\{\text{hep}\}) &= (0.144 + 0.096)/0.44 = 0.545 \\ m_3 \oplus m_4(\{\text{cirr}\}) &= 0.084/0.44 = 0.191 \\ m_3 \oplus m_4(\{\text{hep, cirr}\}) &= 0.036/0.44 = 0.082 \\ m_3 \oplus m_4(\{\text{cirr, gall, pan}\}) &= 0.056/0.44 = 0.127 \\ m_3 \oplus m_4(\Theta) &= 0.024/0.44 = 0.055 \\ m_3 \oplus m_4 &\text{ is 0 for all other subsets of } \Theta \end{aligned}$$

Note that $\sum m_3 \oplus m_4(X) = 1$, as is required by the definition of a bpa.

13.1.5 Belief Intervals

After all bpa's with the same frame of discernment have been combined and the belief function Bel defined by this new bpa has been computed, how should the information given by Bel be used? $\text{Bel}(A)$ gives the total

¹Note that the revised values will still sum to 1 and hence satisfy that condition in the definition of a bpa. If $a + b + c = 1$ then $(a + b)/(1 - c) = 1$ and $a/(1 - c) + b/(1 - c) = 1$.

amount of belief committed to the subset A after all evidence bearing on A has been pooled. However, the function Bel contains additional information about A , namely, $\text{Bel}(A^c)$, the extent to which the evidence supports the negation of A , i.e., A^c . The quantity $1 - \text{Bel}(A^c)$ expresses the plausibility of A , i.e., the extent to which the evidence allows one to fail to doubt A .

The information contained in Bel concerning a given subset A may be conveniently expressed by the interval

$$[\text{Bel}(A) \ 1 - \text{Bel}(A^c)]$$

It is not difficult to see that the right endpoint is always greater than the left: $1 - \text{Bel}(A^c) \geq \text{Bel}(A)$ or, equivalently, $\text{Bel}(A) + \text{Bel}(A^c) \leq 1$. Since $\text{Bel}(A)$ and $\text{Bel}(A^c)$ are the sum of all values of m for subsets of A and A^c , respectively, and since A and A^c have no subsets in common, $\text{Bel}(A) + \text{Bel}(A^c) \leq \sum m(X) = 1$ where X ranges over all subsets of Θ .

In the Bayesian situation, in which $\text{Bel}(A) + \text{Bel}(A^c) = 1$, the two endpoints of the belief interval are equal and the width of the interval $1 - \text{Bel}(A^c) - \text{Bel}(A)$ is 0. In the Dempster-Shafer model, however, the width is usually not 0 and is a measure of the belief that, although not committed to A , is also not committed to A^c . It is easily seen that the width is the sum of belief committed exactly to subsets of Θ that intersect A but that are not subsets of A . If A is a singleton, all such subsets are supersets of A , but this is not true for a nonsingleton A . To illustrate, let $A = \{\text{hep}\}$:

$$\begin{aligned} 1 - \text{Bel}(A^c) - \text{Bel}(A) &= 1 - \text{Bel}(\{\text{cirr}, \text{gall}, \text{pan}\}) - \text{Bel}(\{\text{hep}\}) \\ &= 1 - [m(\{\text{cirr}, \text{gall}, \text{pan}\}) + m(\{\text{cirr}, \text{gall}\}) \\ &\quad + m(\{\text{cirr}, \text{pan}\}) + m(\{\text{gall}, \text{pan}\}) + m(\{\text{cirr}\}) \\ &\quad + m(\{\text{gall}\}) + m(\{\text{pan}\})] - m(\{\text{hep}\}) \\ &= m(\{\text{hep}, \text{cirr}\}) + m(\{\text{hep}, \text{gall}\}) \\ &\quad + m(\{\text{hep}, \text{pan}\}) + m(\{\text{hep}, \text{cirr}, \text{gall}\}) \\ &\quad + m(\{\text{hep}, \text{cirr}, \text{pan}\}) \\ &\quad + m(\{\text{hep}, \text{gall}, \text{pan}\}) + m(\Theta) \end{aligned}$$

Belief committed to a superset of $\{\text{hep}\}$ might, on further refinement of the evidence, result in belief committed to $\{\text{hep}\}$. Thus the width of the belief interval is a measure of that portion of the total belief, 1, that could be added to that committed to $\{\text{hep}\}$ by a physician willing to ignore all but the disconfirming effects of the evidence.

The width of a belief interval can also be regarded as the amount of uncertainty with respect to a hypothesis, given the evidence. It is belief that is committed by the evidence to neither the hypothesis nor the negation of the hypothesis. The vacuous belief function results in width 1 for all belief intervals, and Bayesian functions result in width 0. Most evidence leads to belief functions with intervals of varying widths, where the widths are numbers between 0 and 1.

13.2 The Dempster-Shafer Theory and MYCIN

MYCIN is well suited for implementation of the Dempster-Shafer theory. First, mutual exclusivity of singletons in a frame of discernment is satisfied by the sets of hypotheses in MYCIN constituting the frames of discernment (single-valued parameters; see Chapter 5). This condition may be a stumbling block to the model's implementation in other expert systems where mutual exclusivity cannot be assumed. Second, the belief functions that represent evidence in MYCIN are of a particularly simple form and thus reduce the combination rule to an easily managed computational scheme. Third, the variables and functions already used to define CF's can be adapted and modified for belief function values. These features will now be discussed and illustrated with examples from MYCIN. It should be noted that we have not yet implemented the model in MYCIN.

13.2.1 Frames of Discernment in MYCIN

How should the frames of discernment in MYCIN be chosen? Shafer (1976, p. 36) points out:

It should not be thought that the possibilities that comprise Θ will be determined and meaningful independently of our knowledge. Quite to the contrary: Θ will acquire its meaning from what we know or think we know; the distinctions that it embodies will be embedded within the matrix of our language and its associated conceptual structures and will depend on those structures for whatever accuracy and meaningfulness they possess.

The "conceptual structures" in MYCIN are the associative triples found in the conclusions of the rules, which have the form (object attribute value).² Such a triple gives rise to a singleton hypothesis of the form "the attribute of object is value." A frame of discernment would then consist of all triples with the same object and attribute. Thus the number of triples, or hypotheses in Θ , will equal the number of possible values that the object may assume for the attribute in question. The theory requires that these values be mutually exclusive, as they are for single-valued parameters in MYCIN.

For example, one frame of discernment is generated by the set of all triples of the form (*Organism-1 Identity X*), where *X* ranges over all possible identities of organisms known to MYCIN—*Klebsiella*, *E. coli*, *Pseudomonas*, etc. Another frame is generated by replacing *Organism-1* with *Organism-2*. A third frame is the set of all triples of the form (*Organism-1 Morphology*

²Also referred to as (context parameter value); see Chapter 5.

X), where X ranges over all known morphologies—coccus, rod, bacillus, etc.³

Although it is true that a patient may be infected by more than one organism, organisms are represented as separate contexts in MYCIN (not as separate values of the same parameter). Thus MYCIN's representation scheme is particularly well suited to the mutual exclusivity demand of the Dempster-Shafer theory. Many other expert systems meet this demand less easily. Consider, for example, how the theory might be applicable in a system that gathers and pools evidence concerning the identity of a patient's disease. Then there is often the problem of multiple, coexistent diseases; i.e., the hypotheses in the frame of discernment may not be mutually exclusive. One way to overcome this difficulty is to choose Θ to be the set of all subsets of all possible diseases. The computational implications of this choice are harrowing, since if there are 600 possible diseases (the approximate scope of the INTERNIST knowledge base), then

$$|\Theta| = 2^{600} \quad \text{and} \quad |2^\Theta| = 2^{2^{600}}$$

However, since the evidence may actually focus on a small subset of 2^Θ , the computations need not be intractable. A second, more reasonable alternative would be to apply the Dempster-Shafer theory after partitioning the set of diseases into groups of mutually exclusive diseases and considering each group as a separate frame of discernment. The latter approach would be similar to that used in INTERNIST-1 (Miller et al., 1982), where scoring and comparison of hypotheses are undertaken only after a special partitioning algorithm has separated evoked hypotheses into subsets of mutually exclusive diagnoses.

13.2.2 Rules as Belief Functions

In the most general situation, a given piece of evidence supports many of the subsets of Θ , each to varying degrees. The simplest situation is that in which the evidence supports only one subset to a certain degree and the remaining belief is assigned to Θ . Because of the modular way in which knowledge is captured and encoded in MYCIN, this latter situation applies in the case of MYCIN rules.

If the premises confirm the conclusion of a rule with degree s , where s is above threshold value, then the rule's effect on belief in the subsets of

³The objection may be raised that in some cases all triples with the same object and attribute are not mutually exclusive. For example, both (Patient-1 Allergy Penicillin) and (Patient-1 Allergy Ampicillin) may be true. In MYCIN, however, these triples tend not to have partial degrees of belief associated with them; they are usually true-false propositions ascertained by simple questioning of the user by the system. Thus it is seldom necessary to combine evidence regarding these multi-valued parameters (see Chapter 5), and these hypotheses need not be treated by the Dempster-Shafer theory.

Θ can be represented by a bpa. This bpa assigns s to the singleton corresponding to the hypothesis in the conclusion of the rule, call it A , and assigns $1-s$ to Θ . In the language of MYCIN, the CF associated with this conclusion is s . If the premise disconfirms the conclusion with degree s , then the bpa assigns s to the subset corresponding to the negation of the conclusion, A^c , and assigns $1-s$ to Θ . The CF associated with this conclusion is $-s$. Thus, we are arguing that the CF's associated with rules in MYCIN and other EMYCIN systems can be viewed as bpa's in the Dempster-Shafer sense and need not be changed in order to implement and test the Dempster-Shafer model.

13.2.3 Types of Evidence Combination in MYCIN

The revised quantification scheme we propose for modeling inexact inference in MYCIN is the replacement of the previous CF combining function with the Dempster combination rule applied to belief functions arising from the triggering of domain rules. The combination of such functions is computationally simple, especially when compared to that of two general belief functions.

To illustrate, we consider a frame of discernment, Θ , consisting of all associative triples of the form (Organism-1 Identity X), where X ranges over all identities of organisms known to MYCIN. The triggering of two rules affecting belief in these triples can be categorized in one of the three following ways.

Category 1. Two rules are both confirming or both disconfirming of the same triple, or conclusion. For example, both rules confirm *Pseudomonas* (Pseu), one to degree 0.4 and the other to degree 0.7. The effect of triggering the rules is represented by bpa's m_1 and m_2 , where $m_1(\{\text{Pseu}\}) = 0.4$, $m_1(\Theta) = 0.6$, and $m_2(\{\text{Pseu}\}) = 0.7$, $m_2(\Theta) = 0.3$. The combined effect on belief is given by $m_1 \oplus m_2$, computed using the following tableau:

		m_2	
		{Pseu} (0.7)	Θ (0.3)
m_1	{Pseu} (0.4)	{Pseu} (0.28)	{Pseu} (0.12)
	Θ (0.6)	{Pseu} (0.42)	Θ (0.18)

Note that $\kappa=0$ in this example, so no normalization is required (i.e., $1-\kappa=1$).

$$m_1 \oplus m_2(\{\text{Pseu}\}) = 0.28 + 0.12 + 0.42 = 0.82$$

$$m_1 \oplus m_2(\Theta) = 0.18$$

Note that $m_1 \oplus m_2$ is a bpa that, like m_1 and m_2 , assigns some belief to a certain subset of Θ , {Pseu}, and the remaining belief to Θ . For two confirming rules, the subset is a singleton; for disconfirming rules, the subset is a set of size $n - 1$, where n is the size of Θ .

This category demonstrates that the original MYCIN CF combining function is a special case of the Dempster function (MYCIN would also combine 0.4 and 0.7 to get 0.82). From earlier definitions, it can easily be shown, using the Dempster-Shafer model to derive a new bpa corresponding to the combination of two CF's of the same sign, that

$$\begin{aligned}
 m_1 \oplus m_2(A) &= s_1 s_2 + s_1(1 - s_2) + s_2(1 - s_1) \text{ where } s_i = m_i(A), i = 1, 2 \\
 &= s_1 + s_2(1 - s_1) \\
 &= s_2 + s_1(1 - s_2) \\
 &= 1 - (1 - s_1)(1 - s_2) \\
 &= 1 - m_1 \oplus m_2(\Theta)
 \end{aligned}$$

Category 2. One rule is confirming and the other disconfirming of the same singleton hypothesis. For example, one rule confirms {Pseu} to degree 0.4, and the other disconfirms {Pseu} to degree 0.8. The effect of triggering these two rules is represented by bpa's m_1 and m_3 , where m_1 is defined in the example from Category 1 and $m_3(\{Pseu\}^c) = 0.8$, $m_3(\Theta) = 0.2$. The combined effect on belief is given by $m_1 \oplus m_3$.

		m_3	
		{Pseu} ^c (0.8)	Θ (0.2)
m_1	{Pseu} (0.4)	\emptyset (0.32)	{Pseu} (0.08)
	Θ (0.6)	{Pseu} ^c (0.48)	Θ (0.12)

Here $\kappa = 0.32$ and $1 - \kappa = 0.68$.

$$\begin{aligned}
 m_1 \oplus m_3(\{Pseu\}) &= 0.08/0.68 = 0.118 \\
 m_1 \oplus m_3(\{Pseu\}^c) &= 0.48/0.68 = 0.706 \\
 m_1 \oplus m_3(\Theta) &= 0.12/0.68 = 0.176 \\
 m_1 \oplus m_3 &\text{ is 0 for all other subsets of } \Theta
 \end{aligned}$$

Given m_1 above, the belief interval of {Pseu} is initially $[Bel_1(\{Pseu\}) 1 - Bel_1(\{Pseu\}^c)] = [0.4 1]$. After combination with m_3 , it becomes $[0.118 0.294]$. Similarly, given m_3 alone, the belief interval of {Pseu} is $[0 0.2]$. After combination with m_1 , it becomes $[0.118 0.294]$.

As is illustrated in this category of evidence aggregation, an essential aspect of the Dempster combination rule is the reducing effect of evidence

supporting a subset of Θ on belief in subsets disjoint from this subset. Thus evidence confirming $\{\text{Pseu}\}^c$ will reduce the effect of evidence confirming $\{\text{Pseu}\}$; in this case the degree of support for $\{\text{Pseu}\}$, 0.4, is reduced to 0.118. Conversely, evidence confirming $\{\text{Pseu}\}$ will reduce the effect of evidence confirming $\{\text{Pseu}\}^c$; 0.8 is reduced to 0.706. These two effects are reflected in the modification of the belief interval of $\{\text{Pseu}\}$ from $[0.4 \ 1]$ to $[0.118 \ 0.294]$, where $0.294 = 1 - \text{Bel}(\{\text{Pseu}\}^c) = 1 - 0.706$.

If $A = \{\text{Pseu}\}$, $s_1 = m_1(A)$, and $s_3 = m_3(A^c)$, we can examine this modification of belief quantitatively:

$$m_1 \oplus m_3(A) = s_1(1 - s_3)/(1 - s_1s_3) \text{ where } \kappa = s_1s_3$$

$$m_1 \oplus m_3(A^c) = s_3(1 - s_1)/(1 - s_1s_3)$$

$$m_1 \oplus m_3(\Theta) = (1 - s_1)(1 - s_3)/(1 - s_1s_3)$$

Thus s_1 is multiplied by the factor $(1 - s_3)/(1 - s_1s_3)$, and s_3 is multiplied by $(1 - s_1)/(1 - s_1s_3)$. Each of these factors is less than or equal to 1.⁴ Thus combination of confirming and disconfirming evidence reduces the support provided by each before combination.

Consider the application of the MYCIN CF combining function to this situation. If CF_p is the positive (confirming) CF for $\{\text{Pseu}\}$ and CF_n is the negative (disconfirming) CF:⁵

$$\begin{aligned} CF_{\text{COMBINE}}[CF_p, CF_n] &= (CF_p + CF_n)/(1 - \min\{|CF_p|, |CF_n|\}) \\ &= (s_1 - s_3)/(1 - \min\{s_1, s_3\}) \\ &= (0.4 - 0.8)/(1 - 0.4) \\ &= -0.667 \end{aligned}$$

When this CF is translated into the language of Dempster-Shafer, the result of the MYCIN combining function is belief in $\{\text{Pseu}\}$ and $\{\text{Pseu}\}^c$ to the degrees 0 and 0.667, respectively. The larger disconfirming evidence of 0.8 essentially negates the smaller confirming evidence of 0.4. The confirming evidence reduces the effect of the disconfirming from 0.8 to 0.667.

By examining CF_{COMBINE} , it is easily seen that its application to CF's of the opposite sign results in a CF whose sign is that of the CF of greater magnitude. Thus support for A and A^c is combined into reduced support for one or the other. In contrast, the Dempster function results in reduced support for *both* A and A^c . The Dempster function seems to us a more realistic reflection of the competing effects of conflicting pieces of evidence.

Looking more closely at the value of 0.667 computed by the MYCIN function, we observe that its magnitude is less than that of the correspond-

⁴ $s_1s_3 \leq s_i$ implies $1 - s_1s_3 \geq 1 - s_i$ implies $(1 - s_i)/(1 - s_1s_3) \leq 1$ for $i = 1, 3$.

⁵See Section 10.2 for a discussion of this modified version of the original CF combining function, which was defined and defended in Chapter 11.

ing value of 0.706 computed by the Dempster function. It can be shown that the MYCIN function always results in greater reductions. To summarize, if s_1 and s_3 represent support for A and A^c , respectively, with $s_1 \geq s_3$, and if s_1' and s_3' represent support after Dempster combination, then the MYCIN function results in support for only A , where this support is less than s_1' . Similarly, if $s_3 \geq s_1$, the MYCIN function results in support for only A^c , where the magnitude of this support is less than s_3' .

The difference in the two approaches is most evident in the case of aggregation of two pieces of evidence, one confirming A to degree s and the other disconfirming A to the same degree. MYCIN's function yields $CF=0$, whereas the Dempster rule yields belief of $s(1-s)/(1-s^2) = s/(1+s)$ in each of A and A^c . These results are clearly very different, and again the Dempster rule seems preferable on the grounds that the effect of confirming and disconfirming evidence of the same weight should be different from that of no evidence at all.

We now examine the effect on belief of combination of two pieces of evidence supporting mutually exclusive singleton hypotheses. The MYCIN combining function results in no effect and differs most significantly from the Dempster rule in this case.

Category 3. The rules involve different hypotheses in the same frame of discernment. For example, one rule confirms {Pseu} to degree 0.4, and the other disconfirms {Strep} to degree 0.7. The triggering of the second rule gives rise to m_4 defined by $m_4(\{\text{Strep}\}^c) = 0.7$, $m_4(\Theta) = 0.3$. The combined effect on belief is given by $m_1 \oplus m_4$.

		m_4	
		{Strep} ^c (0.7)	Θ (0.3)
m_1	{Pseu} (0.4)	{Pseu} (0.28)	{Pseu} (0.12)
	Θ (0.6)	{Strep} ^c (0.42)	Θ (0.18)

In this case, $\kappa = 0$.

$$m_1 \oplus m_4(\{\text{Pseu}\}) = 0.28 + 0.12 = 0.40$$

$$m_1 \oplus m_4(\{\text{Strep}\}^c) = 0.42$$

$$m_1 \oplus m_4(\Theta) = 0.18$$

$$m_1 \oplus m_4 \text{ is } 0 \text{ for all other subsets of } \Theta$$

$$\text{Bel}_1 \oplus \text{Bel}_4(\{\text{Pseu}\}) = 0.40$$

$$\begin{aligned} \text{Bel}_1 \oplus \text{Bel}_4(\{\text{Strep}\}^c) &= m_1 \oplus m_4(\{\text{Strep}\}^c) + m_1 \oplus m_4(\{\text{Pseu}\}) \\ &= 0.42 + 0.40 \\ &= 0.82 \end{aligned}$$

$$\text{Bel}_1 \oplus \text{Bel}_4(\{\text{Pseu}\}^c) = \text{Bel}_1 \oplus \text{Bel}_4(\{\text{Strep}\}) = 0$$

Before combination, the belief intervals for {Pseu} and {Strep}^c are [0.4 1] and [0.7 1], respectively. After combination, they are [0.4 1] and [0.82 1], respectively. Note that evidence confirming {Pseu} has also confirmed {Strep}^c, a superset of {Pseu}, but that evidence confirming {Strep}^c has had no effect on belief in {Pseu}, a subset of {Strep}^c.

13.2.4 Evidence Combination Scheme

We now propose an implementation in MYCIN of the Dempster-Shafer method, which minimizes computational complexity. Barnett (1981) claims that direct translation of the theory, without attention to the order in which the belief functions representing rules are combined, results in exponential increases in the time for computations. This is due to the need to enumerate all subsets or supersets of a given set. Barnett's scheme reduces the computations to linear time by combining the functions in a simplifying order. We outline his scheme adapted to MYCIN.

Step 1. For each triple (i.e., singleton hypothesis), combine all bpa's representing rules confirming that value of the parameter. If s_1, s_2, \dots, s_k represent different degrees of support derived from the triggering of k rules confirming a given singleton, then the combined support is

$$1 - (1 - s_1)(1 - s_2) \dots (1 - s_k)$$

(Refer to Category 1 combinations above if this is not obvious.) Similarly, for each singleton, combine all bpa's representing rules disconfirming that singleton. Thus all evidence confirming a singleton is pooled and represented by a bpa, and all evidence disconfirming the singleton (confirming the hypothesis corresponding to the set complement of the singleton) is pooled and represented by another bpa. We thus have $2n$ bpa's, where n is the size of Θ . These functions all have the same form as the original functions. This step is identical to the original approach for gathering confirming and disconfirming evidence into MB's and MD's, respectively.

Step 2. For each triple, combine the two bpa's computed in Step 1. Such a computation is a Category 2 combination and has been illustrated. We now have n bpa's, which are denoted $Evi_1, Evi_2, \dots, Evi_n$.

Step 3. Combine the bpa's computed in Step 2 in one computation, using formulae developed by Barnett (1981), to obtain a final belief function Bel. A belief interval for each singleton hypothesis can then be computed. The form of the required computation is shown here without proof. See Barnett (1981) for a complete derivation.

Let $\{i\}$ represent the i th of n singleton hypotheses in Θ and let

$$\text{Evi}_i(\{i\}) = p_i$$

$$\text{Evi}_i(\{i\}^c) = c_i$$

$$\text{Evi}_i(\Theta) = r_i$$

Since $p_i + c_i + r_i = 1$, $r_i = 1 - p_i - c_i$. Let $d_i = c_i + r_i$. Then it can be shown that the function Bel resulting from combination of $\text{Evi}_1, \dots, \text{Evi}_n$ is given by

$$\text{Bel}(\{i\}) = K[p_i \prod_{j \neq i} d_j + r_i \prod_{j \neq i} c_j]$$

For a subset A of Θ with $|A| > 1$,

$$\text{Bel}(A) = K([\prod_{\text{all } j} d_j] [\sum_{j \in A} p_j/d_j] + [\prod_{j \notin A} c_j] [\prod_{j \in A} d_j] - \prod_{\text{all } j} c_j)$$

where

$$K^{-1} = [\prod_{\text{all } j} d_j] [1 + \sum_{\text{all } j} p_j/d_j] - \prod_{\text{all } j} c_j$$

as long as $p_j \neq 1$ for all j .

An Example

The complex formulation for combining belief functions shown above is computationally straightforward for limited numbers of competing hypotheses such as are routinely encountered in medical domains. As we noted earlier, the INTERNIST program (Miller et al., 1982) partitions its extensive set of possible diagnoses into a limited subset of likely diseases that could be seen as the current frame of discernment. There are likely to be knowledge-based heuristics that can limit the search space in other domains and thereby make calculations of a composite belief function tenable.

Example 7. Consider, for example, the net effect of the following set of rules regarding the diagnosis of the infecting organism. Assume that all other rules failed and that the final conclusion about the beliefs in competing hypotheses will be based on the following successful rules:

- R1: disconfirms {Pseu} to the degree 0.6
- R2: disconfirms {Pseu} to the degree 0.2
- R3: confirms {Strep} to the degree 0.4
- R4: disconfirms {Staph} to the degree 0.8
- R5: confirms {Strep} to the degree 0.3
- R6: disconfirms {Pseu} to the degree 0.5

R7: confirms {Pseu} to the degree 0.3

R8: confirms {Staph} to the degree 0.7

Note, here, that $\Theta = \{\text{Staph}, \text{Strep}, \text{Pseu}\}$ and that for this example we are making the implicit assumption that the patient has an infection with one of these organisms.

Step 1. Considering first confirming and then disconfirming evidence for each organism, we obtain:

{Pseu} confirmed to the degree $s_1 = 0.3$, disconfirmed to the degree $s_1' = 1 - (1 - 0.6)(1 - 0.2)(1 - 0.5) = 0.84$

{Staph} confirmed to the degree $s_2 = 0.7$, disconfirmed to the degree $s_2' = 0.8$

{Strep} confirmed to the degree $s_3 = 1 - (1 - 0.4)(1 - 0.3) = 0.58$,
disconfirmed to the degree $s_3' = 0$

Step 2. Combining the confirming and disconfirming evidence for each organism, we obtain:

$$\text{Evi}_1(\{\text{Pseu}\}) = \frac{0.3(1 - 0.84)}{1 - (0.3)(0.84)} = 0.064 = p_1$$

$$\text{Evi}_1(\{\text{Pseu}\}^c) = \frac{0.84(1 - 0.3)}{1 - (0.3)(0.84)} = 0.786 = c_1$$

Thus $r_1 = 0.15$ and $d_1 = 0.786 + 0.15 = 0.936$.

$$\text{Evi}_2(\{\text{Staph}\}) = \frac{0.7(1 - 0.08)}{1 - (0.7)(0.8)} = 0.318 = p_2$$

$$\text{Evi}_2(\{\text{Staph}\}^c) = \frac{0.8(1 - 0.07)}{1 - (0.7)(0.8)} = 0.545 = c_2$$

Thus $r_2 = 0.137$ and $d_2 = 0.545 + 0.137 = 0.682$.

$$\text{Evi}_3(\{\text{Strep}\}) = 0.58 = p_3$$

$$\text{Evi}_3(\{\text{Strep}\}^c) = 0 = c_3$$

Thus $r_3 = 0.42$ and $d_3 = 0.42$.

Step 3. Assessing the effects of belief in the various organisms on each other, we obtain:

$$\begin{aligned}
 K^{-1} &= d_1d_2d_3(1 + p_1/d_1 + p_2/d_2 + p_3/d_3) - c_1c_2c_3 \\
 &= (0.936)(0.682)(0.42)(1 + 0.064/0.936 + 0.318/0.682 \\
 &\quad + 0.58/0.42) - (0.786)(0.545)(0) \\
 &= 0.268(1 + 0.068 + 0.466 + 1.38) \\
 &= 0.781 \\
 K &= 1.28
 \end{aligned}$$

$$\begin{aligned}
 \text{Bel}(\{\text{Pseu}\}) &= K(p_1d_2d_3 + r_1c_2c_3) \\
 &= 1.28((0.064)(0.682)(0.42) + (0.15)(0.545)0) \\
 &= 0.023
 \end{aligned}$$

$$\begin{aligned}
 \text{Bel}(\{\text{Staph}\}) &= K(p_2d_1d_3 + r_2c_1c_3) \\
 &= 1.28((0.318)(0.936)(0.42) + (1.137)(0.786)0) \\
 &= 0.160
 \end{aligned}$$

$$\begin{aligned}
 \text{Bel}(\{\text{Strep}\}) &= K(p_3d_1d_2 + r_3c_1c_2) \\
 &= 1.28((0.58)(0.936)(0.682) + (0.42)(0.786)(0.545))
 \end{aligned}$$

$$\begin{aligned}
 \text{Bel}(\{\text{Pseu}\}^c) &= K(d_1d_2d_3(p_2/d_2 + p_3/d_3) + c_1d_2d_3 - c_1c_2c_3) \\
 &= 1.28(0.268(0.466 + 1.381) + (0.786)(0.682)(0.42)) \\
 &= 0.922
 \end{aligned}$$

$$\begin{aligned}
 \text{Bel}(\{\text{Staph}\}^c) &= K(d_1d_2d_3(p_1/d_1 + p_3/d_3) + c_2d_1d_3 - 0) \\
 &= 1.28(0.268(0.068 + 1.381) + (0.545)(0.936)(0.42)) \\
 &= 0.771
 \end{aligned}$$

$$\begin{aligned}
 \text{Bel}(\{\text{Strep}\}^c) &= K(d_1d_2d_3(p_1/d_1 + p_2/d_2) + c_3d_1d_2 - 0) \\
 &= 1.28(0.268(0.068 + 0.466) + 0) \\
 &= 0.184
 \end{aligned}$$

The final belief intervals are therefore:

$$\text{Pseu: [0.023 0.078]} \quad \text{Staph: [0.160 0.229]} \quad \text{Strep: [0.704 0.816]}$$

13.3 Conclusion

The Dempster-Shafer theory is particularly appealing in its potential for handling evidence bearing on categories of diseases as well as on specific disease entities. It facilitates the aggregation of evidence gathered at varying levels of detail or specificity. Thus collaborating experts could specify rules that refer to semantic concepts at whatever level in the domain hierarchy is most natural and appropriate. They would not be limited to the most specific level—the singleton hypotheses of their frame of discernment—but would be free to use more unifying concepts.

In a system in which all evidence either confirms or disconfirms sin-

gleton hypotheses, the combination of evidence via the Dempster scheme is computationally simple if ordered appropriately. Due to its present rule format, MYCIN provides an excellent setting in which to implement the theory. Claims by others that MYCIN is ill-suited to this implementation due to failure to satisfy the mutual exclusivity requirement (Barnett, 1981) reflect a misunderstanding of the program's representation and control mechanisms. Multiple diseases are handled by instantiating each as a separate context; within a given context, the requirements of single-valued parameters maintain mutual exclusivity.

In retrospect, however, we recognize that the hierarchical relationships that exist in the MYCIN domain are not adequately represented. For example, evidence suggesting *Enterobacteriaceae* (a family of gram-negative rods) could have explicitly stated that relationship rather than depending on rules in which an observation supported a list of gram-negative organisms with varying CF's based more on guesswork than on solid data. The evidence really supported the higher-level concept, *Enterobacteriaceae*, and further breakdown may have been unrealistic. In actual practice, decisions about treatment are often made on the basis of high-level categories rather than specific organism identities (e.g., "I'm pretty sure this is an enteric organism, and would therefore treat with an aminoglycoside and a cephalosporin, but I have no idea which of the enteric organisms is causing the disease").

If the MYCIN knowledge base were restructured in a hierarchical fashion so as to allow reasoning about unifying high-level concepts as well as about the competing singleton hypotheses, then the computations of the Dempster-Shafer theory would increase exponentially in complexity. The challenge is therefore to make these computations tractable, either by a modification of the theory or by restricting the evidence domain in a reasonable way. Further work should be directed to this end.