Implementing Dempster's Rule for Hierarchical Evidence*

Glenn Shafer and Roger Logan
School of Business, University of Kansas, Lawrence, KS 66045, U.S.A.

Recommended by Richard Duda

ABSTRACT

This article gives an algorithm for the exact implementation of Dempster's rule in the case of hierarchical evidence. This algorithm is computationally efficient, and it makes the approximation suggested by Gordon and Shortliffe unnecessary. The algorithm itself is simple, but its derivation depends on a detailed understanding of the interaction of hierarchical evidence.

Introduction

Gordon and Shortliffe [4] propose an algorithm for approximating the results of Dempster's rule of combination for the case where the evidence being combined is evidence for and against hypotheses that can be arranged in a hierarchical or tree-like structure. This proposal is motivated by the computational complexity of Dempster's rule. In general, the amount of computation needed to implement the rule increases exponentially with the number of possible answers in a diagnostic problem. Gordon and Shortliffe's algorithm avoids this exponential explosion; the amount of computation it requires increases only linearly with the number of possible answers.

Gordon and Shortliffe's algorithm usually produces a good approximation. In the case of highly conflicting evidence, however, the approximation can be poor; an example is given in Section 2. Moreover, the algorithm does not give degrees of belief for all hypotheses (i.e., all subsets of the set of possible answers). It gives degrees of belief only for hypotheses in the tree.

In this article we show that it is not necessary to resort to Gordon and Shortliffe's approximation. We give an algorithm for exact implementation that is linear in its computational complexity. This algorithm works for slightly more general types of evidence than Gordon and Shortliffe's algorithm, and it

*This is a revision of K.U. School of Business Working Paper No. 174.

Artificial Intelligence 33 (1987) 271–298
0004-3702/87/$3.50

gives degrees of belief for more hypotheses. In particular, it gives plausibilities as well as degrees of belief for hypotheses in the tree.

Dempster's rule is part of the theory of belief functions, sometimes called the Dempster–Shafer theory in the artificial intelligence community. A basic reference for the elementary aspects of this theory is Shafer [9]. A more recent exposition and an extensive bibliography are included in Shafer [10]. Expositions that discuss the theory's relevance to artificial intelligence include Garvey, Lowrence, and Fischler [2], Gordon and Shortliffe [3], and Shafer [11].

In the next section, we provide a reasonably self-contained discussion of those mathematical aspects of the theory of belief functions that are relevant to the algorithm presented in this article. Readers will need to turn to the references just cited for further details of the theory and for information on its intuitive interpretation.

In Section 2, we review the problem posed by Gordon and Shortliffe and describe the approximation they propose. In Section 3 we derive some mathematical facts about the problem, and in Section 4 we use these facts to derive our algorithm. In Section 5, we discuss generalizations.

1. The Mathematics of Belief Functions

Suppose Θ denotes a set of possible answers to some question, and assume that one and only one of these answers can be correct. We call Θ a *frame of discernment*. A function Bel that assigns a degree of belief $\mathrm{Bel}(A)$ to every subset A of Θ is called a *belief function* if it satisfies certain mathematical conditions.

Those familiar with the usual mathematical theory of probability can understand the mathematical structure of belief functions by thinking about random sets. A function Bel defined for every subset A of Θ qualifies as a belief function if and only if there is a random non-empty subset S of Θ such that

$$\mathrm{Bel}(A) = \Pr[S \subseteq A]$$

for all A. (It should be emphasized that this interpretation in terms of a random subset S provides insight only into the mathematical structure of belief functions. It does not provide insight into the interpretation of $\mathrm{Bel}(A)$ as a degree of belief based on evidence. See Shafer [9, 10] for explanations of the evidential interpretation.)

The information in a belief function Bel can also be expressed in terms of the *plausibility function* Pl, given by

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\bar{A}) = \Pr[S \cap A \neq \emptyset]\,,$$

where $\bar{A}$ denotes the complement of A. In the evidential interpretation, $\mathrm{Pl}(A)$

is the plausibility of A in light of the evidence—a measure of the extent to which the evidence fails to refute A. To recover Bel from Pl, we use the relation $\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\bar{A})$. Notice that $\mathrm{Bel}(A) \leq \mathrm{Pl}(A)$ for every subset A of Θ. Both Bel and Pl are monotone: $\mathrm{Bel}(A) \leq \mathrm{Bel}(B)$ and $\mathrm{Pl}(A) \leq \mathrm{Pl}(B)$ whenever $A \subseteq B$.

In this article we assume that the frame of discernment Θ is finite. In this case the information in Bel or Pl is also contained in the *commonality function* Q, defined by

$$Q(A) = \Pr[S \supseteq A]$$

for every subset A of Θ. Indeed, it is shown in [9, Chapter 2] that

$$Q(A) = \sum \{(-1)^{|B|+1}\mathrm{Pl}(B) \mid \emptyset \neq B \subseteq A\} \tag{1}$$

and

$$\mathrm{Pl}(A) = \sum \{(-1)^{|B|+1}Q(B) \mid \emptyset \neq B \subseteq A\} \tag{2}$$

for every non-empty subset A of Θ, where $|B|$ denotes the number of elements in the set B. (Formulas (1) and (2) do not give values for $Q(\emptyset)$ or $\mathrm{Pl}(\emptyset)$, but we know that $Q(\emptyset) = 1$ and $\mathrm{Pl}(\emptyset) = 0$ for any belief function.)

1.1. Dempster's rule

Consider two random non-empty subsets S_1 and S_2. Suppose S_1 and S_2 are probabilistically independent—i.e.,

$$\Pr[S_1 = A_1 \text{ and } S_2 = A_2] = \Pr[S_1 = A_1]\Pr[S_2 = A_2]\,.$$

And suppose $\Pr[S_1 \cap S_2 \neq \emptyset] > 0$. Let S be a random non-empty subset that has the probability distribution of $S_1 \cap S_2$ conditional on $S_1 \cap S_2 \neq \emptyset$—i.e.,

$$\Pr[S = A] = \frac{\Pr[S_1 \cap S_2 = A]}{\Pr[S_1 \cap S_2 \neq \emptyset]} \tag{3}$$

for every non-empty subset A of Θ.

If Bel_1 and Bel_2 are the belief functions corresponding to S_1 and S_2, then we denote the belief function corresponding to S by $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$, and we call $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ the *orthogonal sum* of Bel_1 and Bel_2. The rule for forming $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ is called *Dempster's rule of combination*. This rule corresponds, in the evidential interpretation, to the combination or pooling of independent bodies of evidence. (If $\Pr[S_1 \cap S_2 \neq \emptyset] = 0$, then the two belief functions ontradict each other—i.e., there exists A such that $\mathrm{Bel}_1(A) = 1$ and $\mathrm{Bel}_2(\bar{A}) = 1$. It makes no sense to try to pool the evidence in this case.)

The formation of orthogonal sums by Dempster's rule corresponds to the nultiplication of commonality functions. Indeed, if the commonality functions or Bel_1, Bel_2, and $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ are denoted by Q_1, Q_2, and Q, respectively, hen

$$\begin{aligned} Q(A) &= \Pr[S \supseteq A] \\ &= K \Pr[S_1 \cap S_2 \supseteq A] = K \Pr[S_1 \supseteq A \text{ and } S_2 \supseteq A] \\ &= K \Pr[S_1 \supseteq A] \Pr[S_2 \supseteq A] = K\, Q_1(A) Q_2(A)\,, \end{aligned}$$

vhere K does not depend on A;

$$K^{-1} = \Pr[S_1 \cap S_2 \neq \emptyset]\,.$$

Ve can find K from Q_1 and Q_2 if we substitute $KQ_1(B)Q_2(B)$ for $Q(B)$ and Θ or A in (2). Since $\mathrm{Pl}(\Theta) = 1$, this gives

$$1 = \sum \{(-1)^{|B|+1} K Q_1(B) Q_2(B) \mid \emptyset \neq B \subseteq \Theta\}\,,$$

r

$$K^{-1} = \sum \{(-1)^{|B|+1} Q_1(B) Q_2(B) \mid \emptyset \neq B \subseteq \Theta\}\,. \tag{4}$$

We may summarize by saying that the multiplication of commonality funcions gives a recipe for computing the plausibility function Pl for $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$. irst we find the plausibility functions Pl_1 and Pl_2 using the relation

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

hen we find the commonality functions Q_i using the relation

$$Q_i(A) = \sum \{(-1)^{|B|+1}\, \mathrm{Pl}_i(B) \mid \emptyset \neq B \subseteq A\}\,. \tag{5}$$

hen we find Pl using the relation

$$\mathrm{Pl}(A) = K \sum \{(-1)^{|B|+1} Q_1(B) Q_2(B) \mid \emptyset \neq B \subseteq A\}\,, \tag{6}$$

vhere K is given by (4). This recipe generalizes to the case where we wish to ombine more than two belief functions; we merely put $Q_1(B) \cdots Q_n(B)$ in the lace of $Q_1(B)Q_2(B)$ in (4) and (6).

Unfortunately, this recipe is computationally forbidding if Θ contains a large umber of elements. The number of subsets of Θ increases exponentially with

the number elements of Θ, and the sum in (4), for example, involves a term for each of these subsets.

This computational complexity seems to be intrinsic to Dempster's rule. There does not seem to be any general way of implementing the rule that will always involve fewer computations than are involved in (4), (5), and (6). There are, however, special cases where alternative methods involving less computation are possible.

1.2. Focal elements, simple support functions, and dichotomous belief functions

A subset S of Θ is called a *focal element* of Bel if $\Pr[S = S]$ is positive.

The simplest belief function is the belief function whose only focal element is the whole frame Θ; in this case $\Pr[S = \Theta] = 1$. This belief function is called the *vacuous belief function*. It is obvious that if Bel is the vacuous belief function, then $\mathrm{Bel} \oplus \mathrm{Bel}' = \mathrm{Bel}'$ for any other belief function Bel′.

A belief function is called a *simple support function* if it has at most one focal element not equal to the whole frame Θ. If a simple support function does have a focal element not equal to Θ (i.e., if it is not vacuous), then this focal element is called the *focus* of the simple support function.

A belief function is called *dichotomous* with dichotomy $\{A, \bar{A}\}$ if it has no focal elements other than A, $\bar{A}$, and Θ.

In general, combination by Dempster's rule involves the intersection of focal elements. The focal elements for $\mathrm{Bel}_1 \oplus \cdots \oplus \mathrm{Bel}_n$ will consist of all non-empty intersections of the form $S_1 \cap \cdots \cap S_n$, where S_i is a focal element of Bel_i. Therefore, the orthogonal sum of simple support functions with a common focus will be another simple support function with that focus. Similarly, the orthogonal sum of dichotomous belief functions with a common dichotomy will be another dichotomous belief function with that dichotomy.

1.3. Bayesian belief functions

This theory of belief functions is a generalization of the more familiar Bayesian theory, which uses probability measures as expressions of subjective judgments and updates these measures by conditioning. A probability measure is a belief function, and conditioning is a special case of Dempster's rule.

Let us call a belief function a *Bayesian belief function* if it is a probability measure. A belief function is Bayesian if and only if its focal elements are all singletons. This is equivalent to saying that the corresponding random subset is always equal to a singleton. Since a singleton is contained in a subset A if and only if it has a non-empty intersection with A, a Bayesian belief function is equal to its plausibility function.

In the Bayesian theory, conditioning a belief function Bel_1 on knowledge that a subset B of Θ is true means changing one's degree of belief for each subset A from $\mathrm{Bel}_1(A)$ to

$$\frac{\mathrm{Bel}_1(A \cap B)}{\mathrm{Bel}_1(B)} . \tag{7}$$

In the theory of belief functions, on the other hand, knowledge that a subset B of Θ is true is represented by a belief function, say Bel_2, that has B as its only focal element. And the way to change Bel_1 to take this knowledge into account is to combine Bel_1 with Bel_2 by Dempster's rule. In order to see that this application of Dempster's rule gives the same result as (7), let us return to (3) for a moment.

It is clear that from (3) that if S_1 is always a singleton, then S is also always a singleton, so $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ will indeed be Bayesian. Moreover, if we substitute $\{s\}$ for A in (3) and bear in mind that S_1 is always a singleton and S_2 is always equal to B, then we obtain

$$\Pr[S = \{s\}] = \frac{\Pr[S_1 \cap B = \{s\}]}{\Pr[S_1 \cap B \neq \emptyset]} = \begin{cases} \dfrac{\mathrm{Bel}_1(\{s\})}{\mathrm{Bel}_1(B)}, & \text{if } s \in B, \\ 0, & \text{if } s \notin B. \end{cases}$$

Adding $\Pr[S = \{s\}]$ for all s in A, we obtain (7) for our degree of belief in A.

1.4. Partitions

One case where the computational complexity of Dempster's rule can be reduced is the case where the belief functions being combined are carried by a partition $\mathcal{P}$ of the frame Θ. In this case, $\mathcal{P}$, which has fewer elements than Θ, can in effect be used in the place of Θ when the computations (5), (6) and (4) are carried out.

A *partition* of a frame of discernment Θ is a set of disjoint non-empty subsets of Θ whose union equals Θ. Such a partition $\mathcal{P}$ can itself be regarded as a frame of discernment; it is the set of possible answers to the question, "which element of $\mathcal{P}$ contains the correct answer to the question corresponding to Θ?" If $\mathcal{P}_1$ and $\mathcal{P}_2$ are partitions of Θ and for every element P_1 in $\mathcal{P}_1$ there is an element P_2 in $\mathcal{P}_2$ such that $P_1 \subseteq P_2$, then we say that $\mathcal{P}_1$ is a *refinement* of $\mathcal{P}_2$.

Given a partition $\mathcal{P}$ of Θ, we denote by $\mathcal{P}^*$ the set consisting of all unions of elements of $\mathcal{P}$; $\mathcal{P}^*$ is a field of subsets of Θ.

We say that a belief function Bel over Θ is *carried by* $\mathcal{P}$ if the random subset S corresponding to Bel satisfies

$$\Pr[S \in \mathcal{P}^*] = 1 .$$

It is evident that if Bel_1 and Bel_2 are both carried by $\mathcal{P}$, then $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ will also be carried by $\mathcal{P}$; for if S_1 and S_2 are both in the field $\mathcal{P}^*$ with probability one, then $S_1 \cap S_2$ is as well.

For a given partition $\mathcal{P}$ of Θ and a given subset A of Θ there is a largest element of $\mathcal{P}^*$ contained in A, namely

$$A_{\mathcal{P}} = \bigcup \{P \mid P \in \mathcal{P}, P \subseteq A\} .$$

There is also a smallest element of $\mathcal{P}^*$ containing A, namely

$$A^{\mathcal{P}} = \bigcup \{P \mid P \in \mathcal{P}, P \cap A \neq \emptyset\} .$$

When Bel is carried by $\mathcal{P}$, its values for elements of $\mathcal{P}^*$ determine its values for the other subsets of Θ. Indeed, since $S \in \mathcal{P}^*$, $S \subseteq A$ if and only if $S \subseteq A_{\mathcal{P}}$, and so

$$\begin{aligned} \mathrm{Bel}(A) &= \Pr[S \subseteq A] = \Pr[S \subseteq A_{\mathcal{P}}] \\ &= \mathrm{Bel}(A_{\mathcal{P}}) = \max\{\mathrm{Bel}(B) \mid B \subseteq A, B \in \mathcal{P}^*\} . \end{aligned} \tag{8}$$

Similarly,

$$\mathrm{Pl}(A) = \mathrm{Pl}(A^{\mathcal{P}}) = \min\{\mathrm{Pl}(B) \mid B \supseteq A, B \in \mathcal{P}^*\} . \tag{9}$$

It turns out that when Bel is carried by $\mathcal{P}$ we can replace (1) and (2) by analogous formulas that only involve elements of $\mathcal{P}^*$:

$$Q(A) = \sum \{(-1)^{|B|^{\mathcal{P}}+1} \mathrm{Pl}(B) \mid B \in \mathcal{P}^*, \emptyset \neq B \subseteq A\} \tag{10}$$

and

$$\mathrm{Pl}(A) = \sum \{(-1)^{|B|^{\mathcal{P}}+1} Q(B) \mid B \in \mathcal{P}^*, \emptyset \neq B \subseteq A\} \tag{11}$$

for every non-empty element A of $\mathcal{P}^*$, where $|B|^{\mathcal{P}}$ denotes the number of elements of $\mathcal{P}$ contained in B. It follows that if Bel_1 and Bel_2 are both carried by $\mathcal{P}$, we can compute the plausibility function Pl for $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ by first computing

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

just for A in $\mathcal{P}^*$, then computing

$$Q_i(A) = \sum \{(-1)^{|B|^{\mathcal{P}}+1} \mathrm{Pl}_i(B) \mid B \in \mathcal{P}^*, \emptyset \neq B \subseteq A\} \tag{12}$$

just for A in $\mathcal{P}^*$, and then computing

$$\mathrm{Pl}(A) = K \sum \{(-1)^{|B|^{\mathcal{P}}+1} Q_1(B) Q_2(B) \mid B \in \mathcal{P}^*, \emptyset \neq B \subseteq A\} \tag{13}$$

or A in $\mathcal{P}^*$, where

$$K^{-1} = \sum \{(-1)^{|B|^{\mathcal{P}}+1} Q_1(B) Q_2(B) \mid B \in \mathcal{P}^*, \emptyset \neq B \subseteq A\} . \tag{14}$$

The values $\mathrm{Pl}(A)$ for A not in $\mathcal{P}^*$ can then be obtained, if they are desired, rom (9).

Why do (10) and (11) hold for elements of $\mathcal{P}^*$? The easiest way to see that hey do hold is to recognize that $\mathcal{P}^*$ is isomorphic to the set of all subsets of $\mathcal{P}$. \nd when we do this, we see that (10) and (11) are merely (1) and (2) with $\mathcal{P}$ n the place of Θ. When we use (12)–(14) we are treating our belief functions ıs if they were really belief functions on the simpler frame $\mathcal{P}$.

Formulas (13) and (14) generalize, of course, to the case where more than wo belief functions carried by $\mathcal{P}$ are combined. As before, we simply replace $Q_1(B)Q_2(B)$ by $Q_1(B) \cdots Q_n(B)$.

.5. Coarsenings

Given a random subset S and a partition $\mathcal{P}$, let $S^{\mathcal{P}}$ denote the random subset hat is always equal to $A^{\mathcal{P}}$ when S is equal to A. If Bel is the belief function orresponding to S, then let $\mathrm{Bel}_{\mathcal{P}}$ denote the belief function corresponding to $S^{\mathcal{P}}$. Since $S^{\mathcal{P}}$ is always in $\mathcal{P}^*$, Bel is carried by $\mathcal{P}$. Since $S^{\mathcal{P}} \subseteq A$ if and only if $S \subseteq A_{\mathcal{P}}$,

$$\mathrm{Bel}_{\mathcal{P}}(A) = \Pr[S^{\mathcal{P}} \subseteq A] = \Pr[S \subseteq A_{\mathcal{P}}] = \mathrm{Bel}(A_{\mathcal{P}}) .$$

This means in particular that $\mathrm{Bel}_{\mathcal{P}}(A) = \mathrm{Bel}(A)$ if $A \in \mathcal{P}^*$. Thus, $\mathrm{Bel}_{\mathcal{P}}$ is the ınique belief function that agrees with Bel on $\mathcal{P}^*$ and is carried by $\mathcal{P}$.

Suppose we want to combine two belief functions Bel_1 and Bel_2. And uppose we are tempted to do so using (12), (13), and (14), even though Bel_1 ınd Bel_2 are not carried by the partition $\mathcal{P}$. We know that we will not get the ight answer; we will get $\mathrm{Bel}_{1\mathcal{P}} \oplus \mathrm{Bel}_{2\mathcal{P}}$ instead of $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$. But suppose we ıre not interested in the whole belief function $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$. Suppose we are nterested only in the values of $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ on $\mathcal{M}^*$ for some partition $\mathcal{M}$. We vill get these values right if and only if

$$(\mathrm{Bel}_{1\mathcal{P}} \oplus \mathrm{Bel}_{2\mathcal{P}})_{\mathcal{M}} = (\mathrm{Bel}_1 \oplus \mathrm{Bel}_2)_{\mathcal{M}} .$$

This is equivalent to

$$(S_1^{\mathcal{P}} \cap S_2^{\mathcal{P}})^{\mathcal{M}} = (S_1 \cap S_2)^{\mathcal{M}} . \tag{15}$$

t is also equivalent to the condition that $M \cap P \neq \emptyset$, $S_1 \cap P \neq \emptyset$, and $S_2 \cap P \neq \emptyset$ ogether imply $S_1 \cap S_2 \cap M \neq \emptyset$ whenever $M \in \mathcal{M}$, $P \in \mathcal{P}$, S_1 is a focal element of Bel_1, and S_2 is a focal element of Bel_2. If this condition is satisfied, then we say that *$\mathcal{P}$ discerns the interaction* between Bel_1 and Bel_2 that is relevant to $\mathcal{M}$.

It is easy to see that if $\mathcal{P}$, $\mathcal{P}'$, $\mathcal{M}$, and $\mathcal{M}'$ are all partitions, $\mathcal{P}'$ is finer than $\mathcal{P}$, $\mathcal{M}'$ is coarser than $\mathcal{M}$, and $\mathcal{P}$ discerns the interaction relevant to $\mathcal{M}$, then $\mathcal{P}'$ discerns the interaction relevant to $\mathcal{M}'$.

We are most often interested in whether $\mathcal{P}$ discerns the interaction relevant to itself. In this case (15) becomes

$$S_1^{\mathcal{P}} \cap S_2^{\mathcal{P}} = (S_1 \cap S_2)^{\mathcal{P}} ,$$

and this is equivalent to the condition that $S_1 \cap P \neq \emptyset$ and $S_2 \cap P \neq \emptyset$ together imply $S_1 \cap S_2 \cap P \neq \emptyset$ whenever $P \in \mathcal{P}$, S_1 is a focal element of Bel_1, and S_2 is a focal element of Bel_2. Notice that if one of the pair Bel_1 and Bel_2 is carried by $\mathcal{P}$, then $\mathcal{P}$ will necessarily discern the interaction between Bel_1 and Bel_2 that is relevant to itself.

It might be thought that if $\mathcal{P}$ discerns the interaction relevant to itself and $\mathcal{P}'$ is finer than $\mathcal{P}$, then $\mathcal{P}'$ will also discern the interaction relevant to itself. But this is not necessarily true; $\mathcal{P}'$ will discern the interaction relevant to $\mathcal{P}$, but it may not discern the interaction relevant to $\mathcal{P}'$. Figure 1 illustrates this point. If our two belief functions are simple support functions with foci S_1 and S_2, respectively, then the partition $\{P_1, P_2 \cup P_3\}$ discerns the interaction relevant to itself, but the partition $\{P_1, P_2, P_3\}$ does not. Figure 2 illustrates the opposite situation; $\{P_1, P_2, P_3\}$ discerns the interaction relevant to itself, but $\{P_1, P_2 \cup P_3\}$ does not.

The preceding discussion generalizes readily to the case where we have more than two belief functions. For example, $\mathcal{P}$ discerns the interaction among $\mathrm{Bel}_1, \ldots, \mathrm{Bel}_n$ that is relevant to itself if and only if

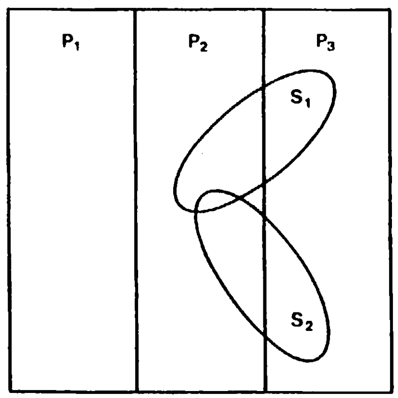

FIG. 1. A partition that does not discern the interaction relevant to itself.

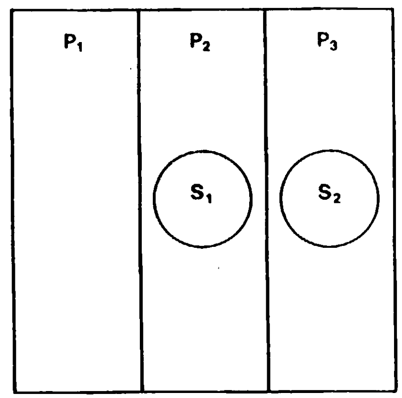

FIG. 2. A partition that does discern the interaction relevant to itself.

$$S_1^{\mathcal{P}} \cap \cdots \cap S_n^{\mathcal{P}} = (S_1 \cap \cdots \cap S_n)^{\mathcal{P}} ,$$

and this is equivalent to the condition that $S_i \cap P \neq \emptyset$ for $i = 1, \ldots, n$ implies $S_1 \cap \cdots \cap S_n \cap P \neq \emptyset$ whenever $P \in \mathcal{P}$ and S_i is a focal element of Bel_i. Notice that if $\mathcal{P}$ discerns the interaction among $\mathrm{Bel}_1, \ldots, \mathrm{Bel}_n$ that is relevant to itself and $\mathrm{Bel}_{n+1}, \ldots, \mathrm{Bel}_{n+m}$ are carried by $\mathcal{P}$, then $\mathcal{P}$ discerns the interaction among $\mathrm{Bel}_1, \ldots, \mathrm{Bel}_{n+m}$ that is relevant to itself.

1.6. Barnett's technique

Barnett [1] has shown that Dempster's rule can be implemented in a number of computations that increases only linearly with the number of elements in Θ if the belief functions being combined are all simple support functions focused on singletons or their complements. Here we will explain Barnett's technique in terms of the commonality function.

Recall that a simple support function focused on S is a belief function whose only focal elements are S and Θ. If θ is an element of Θ, Bel_1 is a simple support function focused on the singleton $\{\theta\}$, and Bel_2 is a simple support function focused on its complement $\overline{\{\theta\}}$, then $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ is easy to calculate; it is dichotomous with the dichotomy $\{\{\theta\}, \overline{\{\theta\}}\}$. In describing Barnett's technique we may, therefore, assume that we begin with dichotomous belief functions of this form. In fact, we may assume, without loss of generality, that we have such a dichotomous belief function, Bel_θ say, for every element θ of Θ; our task is to combine the Bel_θ. (If $\mathrm{Bel}_\theta(\{\theta\}) = \mathrm{Bel}_\theta(\overline{\{\theta\}}) = 0$, then Bel_θ is vacuous, and its presence in the combination makes no difference.)

For brevity, we denote $\mathrm{Bel}_\theta(\{\theta\})$ and $\mathrm{Bel}_\theta(\overline{\{\theta\}})$ by θ^+ and θ^- respectively. In order to avoid trivialities, we assume that both θ^+ and θ^- are less than one. Then the commonality function for Bel_θ is given by

$$Q_\theta(B) = \begin{cases} 1 - \theta^- , & \text{if } B = \{\theta\} , \\ 1 - \theta^+ , & \text{if } \theta \not\in B , \\ 1 - \theta^- - \theta^+ , & \text{if } \theta \in B \text{ and } |B| > 1 \end{cases}$$

for all non-empty subsets B of Θ, and

$$\prod_{\theta \in \Theta} Q_\theta(B) = \begin{cases} (1 - \theta_0^-) \prod\limits_{\theta \neq \theta_0} (1 - \theta^+) , & \text{if } B = \{\theta_0\} \\ \prod\limits_{\theta \in B} (1 - \theta^- - \theta^+) \prod\limits_{\theta \notin B} (1 - \theta^+) , & \text{if } |B| > 1 \end{cases}$$

$$= \begin{cases} \prod\limits_{\theta \in \Theta} (1 - \theta^+)(1 - \theta_0^-)/(1 - \theta_0^+) , & \text{if } B = \{\theta_0\} , \\ \prod\limits_{\theta \in \Theta} (1 - \theta^+) \prod\limits_{\theta \in B} (1 - \theta^- - \theta^+)/(1 - \theta^+) , & \text{if } |B| > 1 . \end{cases}$$

So implementation of the generalization of (6) involves calculating

$$\sum_{\substack{B \subseteq A \\ B \neq \emptyset}} (-1)^{|B|+1} \prod_{\theta \in \Theta} Q_\theta(B)$$

$$= \prod_{\theta \in \Theta} (1 - \theta^+) \left[\sum_{\theta \in A} \frac{1 - \theta^-}{1 - \theta^+} - \sum_{\substack{B \subseteq A \\ |B| > 1}} (-1)^{|B|} \prod_{\theta \in B} \frac{1 - \theta^- - \theta^+}{1 - \theta^+} \right]$$

$$= \prod_{\theta \in \Theta} (1 - \theta^+)$$

$$\times \left[\sum_{\theta \in A} \frac{1 - \theta^-}{1 - \theta^+} + 1 - \sum_{\theta \in A} \frac{1 - \theta^- - \theta^+}{1 - \theta^+} - \sum_{B \subseteq A} (-1)^{|B|} \prod_{\theta \in B} \frac{1 - \theta^- - \theta^+}{1 - \theta^+} \right]$$

$$= \prod_{\theta \in \Theta} (1 - \theta^+) \left[1 + \sum_{\theta \in A} \frac{\theta^+}{1 - \theta^+} - \prod_{\theta \in A} \left(1 - \frac{1 - \theta^- - \theta^+}{1 - \theta^+} \right) \right]$$

$$= \prod_{\theta \in \Theta} (1 - \theta^+) \left[1 + \sum_{\theta \in A} \frac{\theta^+}{1 - \theta^+} - \prod_{\theta \in A} \frac{\theta^-}{1 - \theta^+} \right] . \tag{16}$$

The next to last equality is the crucial step; it reduces the summation over subsets of A to a product over elements of A, which can be implemented in linear time.

Substituting (16) in the generalizations of (4) and (6) and omitting the common factor $\Pi_{\theta \in \Theta}(1 - \theta^+)$, we obtain

$$\mathrm{Pl}(A) = K \left(1 + \sum_{\theta \in A} \frac{\theta^+}{1 - \theta^+} - \prod_{\theta \in A} \frac{\theta^-}{1 - \theta^+} \right) , \tag{17}$$

where

$$K^{-1} = 1 + \sum_{\theta \in \Theta} \frac{\theta^+}{1 - \theta^+} - \prod_{\theta \in \Theta} \frac{\theta^-}{1 - \theta^+} . \tag{18}$$

The statement that (17) and (18) allow the implementation of Dempster's rule in linear time should be interpreted with caution. It is true that the number of computations required by (18) increases only linearly with the number of elements in Θ, and the same is true of any particular instance of (17). If, however, we wish to compute the whole belief function Bel, then we need to calculate $\mathrm{Pl}(A)$ for every subset A of Θ, and the number of such subsets increases exponentially with the size of Θ. In some problems this will cause no difficulty, for we will be able to identify a priori a few subsets A of Θ

s the only ones for which we need to know Bel(A) or Pl(A). But in other roblems we may be interested simply in finding the smallest subsets A that ave high values of Bel(A), and if it is not feasible to calculate and look at Bel(A) for all A, then some search strategy may be needed.

If $\theta^+ + \theta^- = 1$ for all elements θ in Θ, then it is easy to locate the subsets A f Θ that have the highest values of Bel(A). Indeed, in this situation Bel is a Bayesian belief function; Bel(A) = Pl(A) for all subsets A, and (17) and (18) ecome

$$\mathrm{Bel}(A) = \sum_{\theta \in A} f(\theta)\,, \tag{19}$$

vhere

$$f(\theta) = \frac{\theta^+}{1-\theta^+} \Big/ \sum_{\theta' \in \Theta} \frac{\theta'^+}{1-\theta'^+}\,. \tag{20}$$

n this case, to locate subsets A with high values of Bel(A) we need only order he elements of Θ from largest to smallest in the value of $f(\theta)$, and consider ubsets obtained by taking initial sequences from this list.

In general $\theta^+ + \theta^-$ will not equal one; in fact, $\theta^+ + \theta^-$ can approach one nly as the weights of evidence for and against θ become infinitely large (see [9, Chapter 9]). However, when there is a substantial amount of evidence both for nd against most of the θ, (19) and (20) may be nearly enough correct to help s identify subsets for which (17) should be computed.

Barnett's technique applies, of course, not only to the case where we begin vith simple support functions for and against singletons but also to the case vhere we begin with simple support functions for and against elements of some oarser partition $\mathcal{P}$. Indeed, if Pl(A) is the plausibility function for the belief unction $\oplus\{\mathrm{Bel}_P \mid P \in \mathcal{P}\}$, where Bel_P is dichotomous with dichotomy $\{P, \bar{P}\}$, and we write P^+ for $\mathrm{Bel}_P(P)$ and P^- for $\mathrm{Bel}_P(\bar{P})$, then

$$\mathrm{Pl}(A) = K\Bigg(1 + \sum_{\substack{P \subseteq A \\ P \in \mathcal{P}}} \frac{P^+}{1-P^+} - \prod_{\substack{P \subseteq A \\ P \in \mathcal{P}}} \frac{P^-}{1-P^+}\Bigg) \tag{21}$$

or every element A of $\mathcal{P}^*$, where

$$K^{-1} = 1 + \sum_{P \in \mathcal{P}} \frac{P^+}{1-P^+} - \prod_{P \in \mathcal{P}} \frac{P^-}{1-P^+}\,. \tag{22}$$

2. Gordon and Shortliffe's Problem

Gordon and Shortliffe [3, 4] discussed the problem of implementing Dempster's rule in the case where one begins with simple support functions focused for or against subsets of Θ that can be arranged hierarchically in a tree. They concluded that it is not feasible to compute Dempster's rule in such cases, and they proposed a simplification of the rule that can be computed easily.

Figure 3 shows a tree of the kind Gordon and Shortliffe considered. This tree represents the frame $\Theta = \{a, b, c, d, e, f\}$. We have labeled each node of the tree with a capital letter, which we will use to name both the node and the subset of Θ to which it corresponds. The terminal nodes of the tree correspond to singleton subsets; $A = \{a\}$, $F = \{f\}$, etc. Each nonterminal node corresponds to the union of the terminal nodes below it; $G = \{a, b, c\}$, $H = \{d, e\}$, and $I = \{a, b, c, f\}$. Notice that most subsets of Θ are not represented in the tree; there is no node, for example, that corresponds to the subset $\{d, f\}$.

In Gordon and Shortliffe's example, the elements of Θ are possible diseases, so that higher nodes in the tree correspond to classes of diseases. They suggested that diagnostic evidence tends either to support or refute particular diseases or natural classes of diseases that appear in the tree. Thus, they posed the problem of combining simple support functions focused on nodes of the tree and on the complements of these nodes.

Gordon and Shortliffe found that it is not difficult to combine simple support functions focused on nodes of the tree, because the intersection of two subsets corresponding to nodes will either be empty (because neither node lies below the other) or else equal to one of the two subsets (the one lying below the other). Combining negative evidence leads to computational difficulties, how-

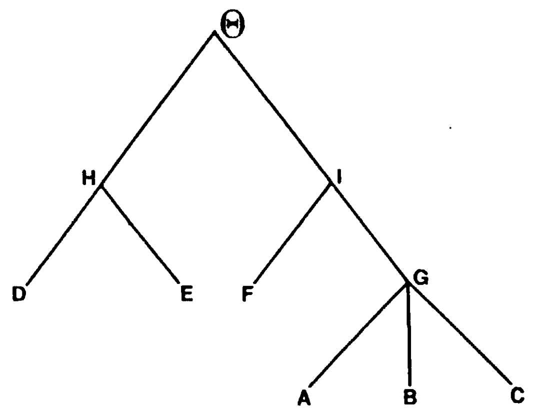

FIG. 3. A tree of diseases.

ever, because the intersection of the complements of nodes may fail to correspond to a node or its complement. The intersection of $\bar{E}$ and $\bar{G}$ in Fig. 3, for example, results in the subset $\{d, f\}$, and neither this subset nor its complement is represented by a node in the tree.

Gordon and Shortliffe suggested the following procedure. First we combine all the simple functions focused on nodes of the tree by Dempster's rule. Then we successively bring into the combination the simple support functions focused on the complements, working down the tree. But when we bring in one of the simple support functions focused on a complement, we modify Dempster's rule by replacing each intersection of focal elements by the smallest subset in the tree that contains it. The final result depends, in general, on the order in which the simple support functions focused on complements are brought in, but Gordon and Shortliffe conjectured that if we bring these simple support functions in as we work down the tree, then the result will approximate the result that we would get using Dempster's rule correctly.

We have found that Gordon and Shortliffe's approximation is usually very good when the degrees of support for the simple support functions are drawn at random from a uniform distribution. It is easy to construct examples, however, where the approximation is poor. Consider the tree in Fig. 4, and suppose that we have three items of evidence. One of these indicates fairly strongly that a patient's disease is in I, while the other two indicate very strongly that it is not f and not g. More precisely, we have three simple support functions to combine:

$$\mathrm{Bel}_1 \text{ focused on } I, \text{ with } \mathrm{Bel}_1(I) = 0.8\,,$$
$$\mathrm{Bel}_2 \text{ focused on } \bar{F}, \text{ with } \mathrm{Bel}_2(\bar{F}) = 0.99\,,$$
$$\mathrm{Bel}_3 \text{ focused on } \bar{G}, \text{ with } \mathrm{Bel}_3(\bar{G}) = 0.99\,.$$

Combining these by Dempster's rule, we obtain a belief function $\mathrm{Bel} = \mathrm{Bel}_1 \oplus \mathrm{Bel}_2 \oplus \mathrm{Bel}_3$, with $\mathrm{Bel}(H) \approx 0.91$, corresponding to the judgment that

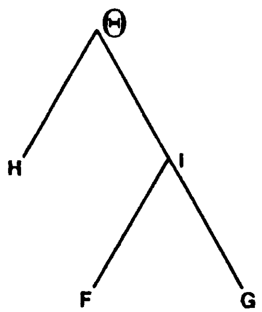

FIG. 4. A smaller tree of diseases.

the positive evidence for I represented by Bel_1 is overwhelmed by the negative evidence represented by Bel_2 and Bel_3. If, however, we combine using Gordon and Shortliffe's procedure, then we obtain, $\mathrm{Bel}(H) = 0$.

Another shortcoming of Gordon and Shortliffe's procedure is that it assigns degrees of belief only to the subsets of Θ that correspond to nodes in the tree. It does not assign degrees of belief to the complements of these nodes. Thus it does not allow us to assign plausibilities to the nodes. (Recall that the plausibility of A, $\mathrm{Pl}(A)$, is equal to $1 - \mathrm{Bel}(\bar{A})$.) Nor, for example, does it assign a degree of belief to the subset $\{d, f\}$ in Fig. 3. Since $\{d, f\}$ is not a natural class of diseases, it may be rare for evidence to support this class without supporting either d or f alone. But such a situation is conceivable; it would arise, for example, if one item of evidence weighed strongly against E and another weighed strongly against G. If this did happen, we would want it to come to our attention, so that we would know to look for further evidence that might help us decide which of these two diseases the patient really has.

Gordon and Shortliffe used the term "hierarchical hypothesis space" to emphasize that they were interested only in hypotheses corresponding to nodes of a tree. Since we think it is appropriate to be interested in degrees of belief for a broader class of hypotheses, we use instead the term "hierarchical evidence." This term reflects the assumption that the evidence bears directly only on hypotheses in the tree, but it leaves open the possibility that we might want to calculate degrees of belief for other hypotheses as well.

3. The Interaction of Hierarchical Evidence

In this section we derive some mathematical facts about the interaction of hierarchical evidence. In the next section we show how these facts enable us to implement Dempster's rule efficiently.

Here, as in the preceding section, we assume that we are working with a finite tree such as the one in Fig. 3. We denote by $\mathscr{A}$ the collection of all the nodes below Θ—i.e., all the nodes except Θ itself. If B is directly below A, we say that B is A's *daughter* and A is B's mother. In order to avoid trivialities, we assume that every node that is not a terminal node has more than one daughter. We call a set of nodes that consists of all the daughters of a given nonterminal node a *sib*. We denote by $\mathscr{S}_A$ the sib consisting of the daughters of A.

We suppose that for each node A in $\mathscr{A}$ we have one simple support function focused on A and another focused on the complement $\bar{A}$. Here, as in our discussion of Barnett's technique, we begin by combining these two simple support functions. Then for each node A in $\mathscr{A}$ we have a single dichotomous belief function Bel_A with the dichotomy $\{A, \bar{A}\}$. We assume that $\mathrm{Bel}_A(A)$ and $\mathrm{Bel}_A(\bar{A})$ are both strictly less than one, but we allow either or both to be zero.

For any node A in the tree, we denote by $\mathrm{Bel}_A^{\downarrow}$ the orthogonal sum of Bel_B

for all nodes B that are strictly below A. In Fig. 3, for example, $\mathrm{Bel}_H^{\downarrow} = \mathrm{Bel}_D \oplus \mathrm{Bel}_E$, and

$$\mathrm{Bel}_I^{\downarrow} = \mathrm{Bel}_F \oplus \mathrm{Bel}_G \oplus \mathrm{Bel}_G^{\downarrow} = \mathrm{Bel}_F \oplus \mathrm{Bel}_G \oplus \mathrm{Bel}_A \oplus \mathrm{Bel}_B \oplus \mathrm{Bel}_C .$$

If A is a terminal node, then $\mathrm{Bel}_A^{\downarrow}$ is vacuous. Our purpose, of course, is to calculate values of $\mathrm{Bel}_\Theta^{\downarrow} = \oplus \{\mathrm{Bel}_A \mid A \in \mathscr{A}\}$.

For each node A in $\mathscr{A}$, we denote by $\mathrm{Bel}_A^{\diamond}$ the orthogonal sum of Bel_B for all nodes B in $\mathscr{A}$ that are neither below A nor equal to A. Thus

$$\mathrm{Bel}_\Theta^{\downarrow} = \mathrm{Bel}_A^{\downarrow} \oplus \mathrm{Bel}_A \oplus \mathrm{Bel}_A^{\diamond} . \tag{23}$$

Lemma 3.1. *Suppose $\mathscr{P}$ is a partition of Θ, and $P \in \mathscr{A} \cap \mathscr{P}$. Then* $(\mathrm{Bel}_P^{\downarrow})_{\mathscr{P}} = (\mathrm{Bel}_P^{\downarrow})_{\{P,\bar{P}\}}$.

Proof. The belief function Bel_A has only A, $\bar{A}$, and Θ as focal elements. If $A \subseteq P$, then each of these focal elements either contains $\bar{P}$ or else is contained in P. A focal element S of $\mathrm{Bel}_P^{\downarrow}$ is obtained by intersecting such focal elements and hence must also either contain $\bar{P}$ or else be contained in P. If S contains $\bar{P}$ but is not equal to $\bar{P}$, then $S^{\mathscr{P}} = S^{\{P,\bar{P}\}} = \Theta$. If S is equal to $\bar{P}$, then $S^{\mathscr{P}} = S^{\{P,\bar{P}\}} = \bar{P}$. If S is contained in P, then $S^{\mathscr{P}} = S^{\{P,\bar{P}\}} = P$. In any case, $S^{\mathscr{P}} = S^{\{P,\bar{P}\}}$. □

Lemma 3.2. *Suppose $\mathscr{P}$ is a partition of Θ, $A \in \mathscr{A}$, and $\bar{A} \in \mathscr{P}$. Then* $(\mathrm{Bel}_A^{\diamond})_{\mathscr{P}} = (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}$.

Proof. Again, Bel_B has only B, $\bar{B}$, and Θ as focal elements. If $B \in \mathscr{A}$ and $B \not\subseteq A$, then B is either disjoint from A or else contains A, and hence each focal element of Bel_B either contains A or else is contained in $\bar{A}$. Any focal element S of $\mathrm{Bel}_A^{\diamond}$ is the intersection of such focal elements and hence must also contain A or else be contained in $\bar{A}$. If S is equal to A, then $S^{\mathscr{P}} = S^{\{A,\bar{A}\}} = S$. If S contains A but is not equal to A, then $S^{\mathscr{P}} = S^{\{A,\bar{A}\}} = \Theta$. If A is contained in $\bar{A}$, then $S^{\mathscr{P}} = S^{\{A,\bar{A}\}} = \bar{A}$. In any case, $S^{\mathscr{P}} = S^{\{A,\bar{A}\}}$. □

Lemma 3.3. *Suppose $\mathscr{P}$ is a partition of Θ. Then $\mathscr{P}$ discerns the interaction relevant to itself among the belief functions in* $\{\mathrm{Bel}_P^{\downarrow} \mid P \in \mathscr{A} \cap \mathscr{P}\}$.

Proof. Suppose $\mathscr{A} \cap \mathscr{P} = \{P_1, \ldots, P_n\}$, and let S_i be a focal element of Bel_{P_i} for $i = 1, 2, \ldots, n$. Fix an element P of $\mathscr{P}$, and suppose $S_i \cap P \neq \emptyset$ for $i = 1, \ldots, n$. We must show that $S_1 \cap \cdots \cap S_n \cap P \neq \emptyset$.

By the proof of Lemma 3.1, S_i either contains $\bar{P}_i$ or else is contained in P_i.

Since $\mathscr{P}$ is a partition, P_i and P are either disjoint or equal. If they are disjoint, then since $S_i \cap P \neq \emptyset$, S_i cannot be contained in P_i; instead it must contain $\bar{P}_i$, and hence it must contain P.

At most one of the P_i can equal P. If none equal P, then all the S_i contain P, and hence $S_1 \cap \cdots \cap S_n \cap P = P$. If one, say P_j, does equal P, then

$$S_1 \cap \cdots \cap S_n \cap P = S_j \cap \Big[\bigcap_{i \neq j} (S_i \cap P) \Big] = S_j \cap P .$$

In either case, $S_1 \cap \cdots \cap S_n \cap P \neq \emptyset$. □

Since the partition $\mathscr{P}$ carries Bel_P for each $P \in \mathscr{A} \cap \mathscr{P}$, we can strengthen Lemma 3.3 to the statement that $\mathscr{P}$ discerns the interaction relevant to itself among the belief functions in

$$\{\mathrm{Bel}_P \mid P \in \mathscr{A} \cap \mathscr{P}\} \cup \{\mathrm{Bel}_P^{\downarrow} \mid P \in \mathscr{A} \cap \mathscr{P}\} .$$

Consider, for example, the partition $\mathscr{S}_A \cup \{\bar{A}\}$, where A is a nonterminal node in $\mathscr{A}$. This partition discerns the interaction relevant to itself among

$$\{\mathrm{Bel}_B \mid B \in \mathscr{S}_A\} \cup \{\mathrm{Bel}_B^{\downarrow} \mid B \in \mathscr{S}_A\} .$$

Since $\mathrm{Bel}_A^{\downarrow}$ is the orthogonal sum of these belief functions, it follows that

$$(\mathrm{Bel}_A^{\downarrow})_{\mathscr{S}_A \cup \{\bar{A}\}} = \oplus \{(\mathrm{Bel}_B)_{\mathscr{S}_A \cup \{\bar{A}\}} \oplus (\mathrm{Bel}_B^{\downarrow})_{\mathscr{S}_A \cup \{\bar{A}\}} \mid B \in \mathscr{S}_A\} .$$

This can be written more simply as

$$(\mathrm{Bel}_A^{\downarrow})_{\mathscr{S}_A \cup \{\bar{A}\}} = \oplus \{\mathrm{Bel}_B \oplus (\mathrm{Bel}_B^{\downarrow})_{\{B,\bar{B}\}} \mid B \in \mathscr{S}_A\} ; \tag{24}$$

$$(\mathrm{Bel}_B)_{\mathscr{S}_A \cup \{\bar{A}\}} = \mathrm{Bel}_B$$

because Bel_B is carried by $\mathscr{S}_A \cup (\bar{A}\}$, and

$$(\mathrm{Bel}_B^{\downarrow})_{\mathscr{S}_A \cup \{\bar{A}\}} = (\mathrm{Bel}_B^{\downarrow})_{\{B,\bar{B}\}}$$

by Lemma 3.1. It should be borne in mind that if the element B of $\mathscr{S}_A$ is a terminal node, then $\mathrm{Bel}_B^{\downarrow}$ is vacuous, and the orthogonal sum $\mathrm{Bel}_B \oplus (\mathrm{Bel}_B^{\downarrow})_{\{B,\bar{B}\}}$ reduces to Bel_B.

The reasoning of the preceding paragraph applies to the case where A is the topmost node Θ, except that in this case the partition is simply $\mathscr{S}_\Theta$, not $\mathscr{S}_\Theta \cup \{\bar{\Theta}\}$. So

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\mathscr{S}_\Theta} = \bigoplus \{\mathrm{Bel}_B \oplus (\mathrm{Bel}_B^{\downarrow})_{\{B,\bar{B}\}} \mid B \in \mathscr{S}_\Theta\}\,. \tag{25}$$

Formulas (24) and (25) tell us that in order to find for A and her immediate daughters the degrees of belief resulting from all the evidence bearing on nodes below A, it is sufficient to consider each daughter separately. We find the degrees of belief for and against each daughter resulting from evidence bearing directly on it and on nodes below it, and then we combine the results for the different daughters.

Lemma 3.4. *Suppose A is a nonterminal element of $\mathscr{A}$. Then the partition $\mathscr{S}_A \cup \{\bar{A}\}$ discerns the interaction relevant to itself between* $\mathrm{Bel}_A^{\downarrow}$ *and* $\mathrm{Bel}_A^{\diamond}$.

Proof. Suppose S_1 is a focal element of $\mathrm{Bel}_A^{\downarrow}$, and S_2 is a focal element of $\mathrm{Bel}_A^{\diamond}$. Then S_1 either contains $\bar{A}$ or is contained in A, while S_2 either contains A or is contained in $\bar{A}$. Table 1 lists the four possibilities and shows what can happen when $S_1 \cap S_2$ is intersected with an element P of $\mathscr{S}_A \cup (\bar{A}\}$. Inspection of the table shows that if $S_1 \cap P \neq \emptyset$ and $S_2 \cap P \neq \emptyset$, then $S_1 \cap S_2 \cap P \neq \emptyset$. This establishes that $\mathscr{S}_A \cup (\bar{A}\}$ discerns the interaction relevant to itself between $\mathrm{Bel}_A^{\downarrow}$ and $\mathrm{Bel}_A^{\diamond}$. □

Since Bel_A is carried by $\mathscr{S}_A \cup \{\bar{A}\}$, Lemma 3.4 can be strengthened to the statement that $\mathscr{S}_A \cup \{\bar{A}\}$ discerns the interaction relevant to itself among $\mathrm{Bel}_A^{\downarrow}$, Bel_A, and $\mathrm{Bel}_A^{\diamond}$. So from (23) we can obtain

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\mathscr{S}_A\cup\{\bar{A}\}} = (\mathrm{Bel}_A^{\downarrow})_{\mathscr{S}_A\cup\{\bar{A}\}} \oplus (\mathrm{Bel}_A)_{\mathscr{S}_A\cup\{\bar{A}\}} \oplus (\mathrm{Bel}_A^{\diamond})_{\mathscr{S}_A\cup\{\bar{A}\}}\,. \tag{26}$$

Since Bel_A is carried by $\{A, \bar{A}\}$, and since

$$(\mathrm{Bel}_A^{\diamond})_{\mathscr{S}_A\cup\{\bar{A}\}} = (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}$$

TABLE 1. Verifying the discernment

<table>
<tr><th></th><th>$P = \bar{A}$</th><th>$P \in S_A$</th></tr>
<tr><td>$S_1 \supseteq \bar{A},\ \ S_2 \supseteq A$</td><td>$S_1 \cap S_2 \cap P = S_2 \cap P$</td><td rowspan="2">$S_1 \cap S_2 \cap P = S_1 \cap P$</td></tr>
<tr><td>$S_1 \subseteq A,\ \ S_2 \supseteq A$</td><td rowspan="2">$S_1 \cap S_2 \cap P = S_1 \cap P = \emptyset$</td></tr>
<tr><td>$S_1 \subseteq A,\ \ S_2 \subseteq \bar{A}$</td><td rowspan="2">$S_1 \cap S_2 \cap P = S_2 \cap P = \emptyset$</td></tr>
<tr><td>$S_1 \supseteq \bar{A},\ \ S_2 \subseteq \bar{A}$</td><td>$S_1 \cap S_2 \cap P = S_2$</td></tr>
</table>

by Lemma 3.2, (26) reduces to

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\mathscr{S}_A\cup\{\bar{A}\}} = (\mathrm{Bel}_A^{\downarrow})_{\mathscr{S}_A\cup\{\bar{A}\}} \oplus \mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}\,. \tag{27}$$

This formula tells us that evidence from above A and down other branches affects our degrees of belief about the daughters of A only inasmuch as it affects our degrees of belief for and against A itself.

In the next section, we will have occasion to use two consequences of (27):

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\{A,\bar{A}\}} = (\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}} \oplus \mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}} \tag{28}$$

and

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\{B,A-B,\bar{A}\}} = (\mathrm{Bel}_A^{\downarrow})_{\{B,A-B,\bar{A}\}} \oplus \mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}} \tag{29}$$

for every B in $\mathscr{S}_A$. These formulas follow from (27) because both the partition $\{A, \bar{A}\}$ and the partition $\{B, A-B, \bar{A}\}$ carry the belief function $\mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}$. Whenever a partition carries a belief function, it discerns the interaction relevant to itself between that belief function and any other belief function.

4. Implementing Dempster's Rule

We now present our algorithm for calculating $\mathrm{Bel}_\Theta^{\downarrow}(A)$ for A in $\mathscr{A}$. We first present the algorithm in general terms and explain how it is justified by the results of the preceding section. We then give detailed formulas for the actual calculations. We conclude with a complexity analysis and a comparison of the complexity with that of Gordon and Shortliffe's algorithm.

The algorithm can be broken down into three states. In the first stage we begin with sibs of terminal nodes, combine the belief functions attached to them to find degrees of belief for and against their mothers, then do the same for the mothers' mothers, and so on, until we have a dichotomous belief function attached to each daughter of Θ to obtain the values of $\mathrm{Bel}_\Theta^{\downarrow}$ for these daughters. In the third stage we use information stored as we moved up the tree to move back down, calculating further values of $\mathrm{Bel}_\Theta^{\downarrow}$ as we go.

4.1. First stage

Recall that we begin with a dichotomous belief function Bel_A attached to each node A of $\mathscr{A}$.

Choose a sib of terminal nodes, and let A denote its mother. According to (24),

$$(\mathrm{Bel}_A^{\downarrow})_{\mathcal{S}_A \cup \{\bar{A}\}} = \oplus \{\mathrm{Bel}_B \mid B \in \mathcal{S}_A\}. \tag{30}$$

Since Bel_B is dichotomous with dichotomy $\{B, \bar{B}\}$, and since B is an atom of the partition $\mathcal{S}_A \cup \{\bar{A}\}$, Barnett's technique can be used to calculate values of the orthogonal sum in this formula. We use it to calculate $\mathrm{Bel}_A^{\downarrow}(A)$ and $\mathrm{Bel}_A^{\downarrow}(\bar{A})$ — i.e., to find $(\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$.

We now compute $\mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$. This is easy, since both $(\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$ and Bel_A are dichotomous with dichotomy $\{A, \bar{A}\}$. We discard Bel_A and store in its place both $(\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$ and $\mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$. This means that we store four numbers at A: $\mathrm{Bel}_A^{\downarrow}(A)$, $\mathrm{Bel}_A^{\downarrow}(\bar{A})$, $(\mathrm{Bel}_A \oplus \mathrm{Bel}_A^{\downarrow})(A)$, and $(\mathrm{Bel}_A \oplus \mathrm{Bel}_A^{\downarrow})(\bar{A})$.

After we have completed the procedure of the two preceding paragraphs for every sib of terminal nodes, we temporarily prune these terminal nodes from the tree, as it were, so that the mothers of the original sibs of terminal nodes are not themselves terminal nodes. We then repeat the procedure with the sibs of terminal nodes we now see, except that instead of (30), we now use (24),

$$(\mathrm{Bel}_A^{\downarrow})_{\mathcal{S}_A \cup \{\bar{A}\}} = \oplus \{\mathrm{Bel}_B \oplus (\mathrm{Bel}_B^{\downarrow})_{\{B,\bar{B}\}} \mid B \in \mathcal{S}_A\}$$

to calculate $\mathrm{Bel}_A^{\downarrow}(A)$ and $\mathrm{Bel}_A^{\downarrow}(\bar{A})$ for the mother A of what are now terminal sibs. (Of course, we really used (24) in the first round, too. When we wrote Bel_B instead of $\mathrm{Bel}_B \oplus (\mathrm{Bel}_B^{\downarrow})_{\{B,\bar{B}\}}$ in (30) above, we were just taking advantage of the fact that $\mathrm{Bel}_B^{\downarrow}$ is vacuous when B is terminal.)

We continue this process until we have reached the daughters of the topmost node Θ. We then have $(\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$ and $\mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$ stored at every node A in $\mathcal{A}$.

4.2. Second stage

Recall (25),

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\mathcal{S}_\Theta} = \oplus \{\mathrm{Bel}_B \oplus (\mathrm{Bel}_B^{\downarrow})_{\{B,\bar{B}\}} \mid B \in \mathcal{S}_\Theta\}.$$

We apply Barnett's technique to this formula to calculate $\mathrm{Bel}_\Theta^{\downarrow}(A)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{A})$ for each A in $\mathcal{S}_\Theta$. Knowing these two numbers amounts to knowing $(\mathrm{Bel}_\Theta^{\downarrow})_{\{A,\bar{A}\}}$. We store them at A, along side the four numbers already there.

4.3. Third stage

Now consider a particular daughter A of Θ. We want to calculate $\mathrm{Bel}_\Theta^{\downarrow}(B)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{B})$ for each daughter B of A. We can do this using (24), (28), and (29).

Consider first (28):

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\{A,\bar{A}\}} = (\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}} \oplus \mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}.$$

All the belief functions in this formula are dichotomous with dichotomy $\{A, \bar{A}\}$, and $(\mathrm{Bel}_\Theta^{\downarrow})_{\{A,\bar{A}\}}$ and $(\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$ are stored at A. So we can easily find $\mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}$ by division.

Now consider (24) again. We have already applied Barnett's technique to this formula to calculate $\mathrm{Bel}_A^{\downarrow}(A)$ and $\mathrm{Bel}_A^{\downarrow}(\bar{A})$. We now apply it again to calculate $\mathrm{Bel}_A^{\downarrow}(B)$, $\mathrm{Bel}_A^{\downarrow}(\bar{B})$, $\mathrm{Bel}_A^{\downarrow}(A - B)$, and $\mathrm{Bel}_A^{\downarrow}(B \cup \bar{A})$ for each B in $\mathcal{S}_A$. This gives us the belief function $(\mathrm{Bel}_A^{\downarrow})_{\{B, A-B, \bar{A}\}}$. (Actually, as we shall see in the next section, we do not need to calculate $\mathrm{Bel}_A^{\downarrow}(A - B)$.)

Now consider (29):

$$(\mathrm{Bel}_\Theta^{\downarrow})_{\{B, A-B, \bar{A}\}} = (\mathrm{Bel}_A^{\downarrow})_{\{B, A-B, \bar{A}\}} \oplus \mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}.$$

We have just found $\mathrm{Bel}_A \oplus (\mathrm{Bel}^{\diamond})_{\{A,\bar{A}\}}$ and $(\mathrm{Bel}_A^{\downarrow})_{\{B, A-B, \bar{A}\}}$. So we can use (29) to calculate $\mathrm{Bel}_\Theta^{\downarrow}(B)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{B})$. (Barnett's technique cannot be used here, since $(\mathrm{Bel}_A^{\downarrow})_{\{B, A-B, \bar{A}\}}$ is not dichotomous. But since the partition we are working with is only a trichotomy, a brute force application of Dempster's rule involves little computation.)

We have just seen how to go from $\mathrm{Bel}_\Theta^{\downarrow}(A)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{A})$ to $\mathrm{Bel}_\Theta^{\downarrow}(B)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{B})$ for the daughters B of A. This process can be repeated for the daughters of each B, and so on, until we have calculated $\mathrm{Bel}_\Theta^{\downarrow}(C)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{C})$ for every node C in the tree.

Usually, of course, we will not be interested in $\mathrm{Bel}_\Theta^{\downarrow}(C)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{C})$ for every node C in the tree. Once we have seen that $\mathrm{Bel}_\Theta^{\downarrow}(B)$ is very small, we know that $\mathrm{Bel}_\Theta^{\downarrow}(C)$ will be at least as small for every descendant C of B, and so we may not want to go to the trouble of finding these values. We may decide to look at descendants of B only if $\mathrm{Bel}_\Theta^{\downarrow}(B)$ is greater than 0.5, say. Since two disjoint sets cannot both have degree of belief greater than 0.5, this decision will result in our moving down the tree along just one path, which may stop before reaching a terminal node.

4.4. Details of the algorithm

The numerical calculations that our algorithm requires can be described by formulas, and we can group these formulas into six subroutines.

The following notation will allow us to write these formulas concisely. For each node A in $\mathcal{A}$, we set

$$
\begin{aligned}
A_0^+ &= \mathrm{Bel}_A(A)\,, & A_0^- &= \mathrm{Bel}_A(\bar{A})\,,\\
A_\downarrow^+ &= \mathrm{Bel}_A^{\downarrow}(A)\,, & A_\downarrow^- &= \mathrm{Bel}_A^{\downarrow}(\bar{A})\,,\\
A^+ &= (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^{\downarrow})(A)\,, & A^- &= (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^{\downarrow})(\bar{A})\,,\\
A_\diamond^+ &= (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^{\diamond}(A)\,, & A_\diamond^- &= (\mathrm{Bel}_A \oplus \mathrm{Bel}_A^{\diamond})(\bar{A})\,,\\
A_\Theta^+ &= \mathrm{Bel}_\Theta^{\downarrow}(A)\,, & A_\Theta^- &= \mathrm{Bel}_\Theta^{\downarrow}(\bar{A})\,.
\end{aligned}
$$

(If A is a terminal node, then $\mathrm{Bel}_A^{\downarrow}$ is vacuous, and therefore $A_\downarrow^+ = A_\downarrow^- = 0$, $A^+ = A_0^+$, and $A^- = A_0^-$.) For each node B other than Θ and its daughters, we set

$$
B_A^+ = \mathrm{Bel}_A^{\downarrow}(B)\,, \qquad B_A^- = \mathrm{Bel}_A^{\downarrow}(B)\,,
$$
$$
B_A^* = \mathrm{Bel}_A^{\downarrow}(B \cup \bar{A})\,,
$$

where A is B's mother.

Recall that the first stage of our algorithm begins with the computation of $(\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$ for mothers of sibs of terminal nodes. Subroutine 1 specifies how this is done. This is followed by the calculation of $\mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\downarrow})_{\{A,A\}}$, by Subroutine 2. After these operations have been completed for every node A whose daughters are all terminal nodes, we pretend to prune all these terminal nodes from the tree, and we repeat the process with the new sibs of terminal nodes, and so on. Each round uses Subroutine 1 followed by Subroutine 2. We continue until we have calculated $(\mathrm{Bel}_A^{\downarrow})_{\{A,\bar{A}\}}$ for the daughters A of Θ.

At the second stage we apply Barnett's technique to (25) to find $\mathrm{Bel}_\Theta^{\downarrow}(A)$ for $A \in \mathscr{S}_\Theta$. This is Subroutine 3.

In the second stage, we go back down the tree. When we go from A to its daughters, we first find $\mathrm{Bel}_A \oplus (\mathrm{Bel}_A^{\diamond})_{\{A,\bar{A}\}}$ using (28); this is Subroutine 4. Then we return to formula (24) and calculate $\mathrm{Bel}_A^{\downarrow}(B)$, $\mathrm{Bel}_A^{\downarrow}(\bar{B})$ and $\mathrm{Bel}_A^{\downarrow}(B \cup \bar{A})$ for each B in $\mathscr{S}_A$; this is Subroutine 5. Finally, we use (29) to calculate $\mathrm{Bel}_\Theta^{\downarrow}(B)$ and $\mathrm{Bel}_\Theta^{\downarrow}(\bar{B})$ for each B in $\mathscr{S}_A$; this is Subroutine 6. (Alternatively, to minimize storage, we may execute Subroutines 5 and 6 for a particular B in $\mathscr{S}_A$, then for another, and so on.)

In summary, we repeatedly cycle through Subroutines 4, 5 and 6 as we move up the tree, we execute Subroutine 3 once at the top of the tree, and then we repeatedly cycle through Subroutines 4, 5 and 6 as we move back down.

Subroutine 1. Calculating $A_\downarrow^+$ and $A_\downarrow^-$ from B^+ and B^- for B in $\mathscr{S}_A$;

$$
A_\downarrow^+ = 1 - K\,, \qquad A_\downarrow^- = K \prod_{B \in \mathscr{S}_A} B^-/(1 - B^+)\,,
$$

where

$$
K^{-1} = 1 + \sum_{B \in \mathscr{S}_A} B^+/(1 - B^+)\,.
$$

Subroutine 2. Calculating A^+ and A^- from A_0^+, A_0^-, $A_\downarrow^+$, and $A_\downarrow^-$:

$$
A^+ = 1 - K(1 - A_0^+)(1 - A_\downarrow^+)\,, \qquad A^- = 1 - K(1 - A_0^-)(1 - A_\downarrow^-)\,,
$$

where

$$
K^{-1} = 1 - A_0^+ A_\downarrow^- - A_0^- A_\downarrow^+\,.
$$

Subroutine 3. Calculating A_Θ^+ and A_Θ^- for A in $\mathscr{S}_\Theta$ from A^+ and A^- for A in $\mathscr{S}_\Theta$:

$$
\begin{aligned}
A_\Theta^+ &= 1 - K\Big(1 + \sum_{\substack{B \in \mathscr{S}_\Theta \\ B \neq A}} B^+/(1 - B^+)\Big) - \prod_{\substack{B \in \mathscr{S}_\Theta \\ B \neq A}} B^-/(1 - B^+)\,,\\
A_\Theta^- &= 1 - K(1 - A^-)/(1 - A^+)\,,
\end{aligned}
$$

where

$$
K^{-1} = 1 + \sum_{B \in \mathscr{S}_\Theta} B^+/(1 - B^+) - \prod_{B \in \mathscr{S}_\Theta} B^-/(1 - B^+)\,.
$$

Subroutine 4. Calculating $A_\diamond^+$ and $A_\diamond^-$ from A_Θ^+, A_Θ^-, $A_\downarrow^+$, and $A_\downarrow^-$;

$$
A_\diamond^+ = 1 - K(1 - A_\Theta^+)/(1 - A_\downarrow^+)\,, \qquad A_\diamond^- = 1 - K(1 - A_\Theta^-)/(1 - A_\downarrow)\,,
$$

where

$$
K^{-1} = \frac{1 - A_\Theta^+}{1 - A_\downarrow^+} + \frac{1 - A_\Theta^-}{1 - A_\downarrow^-} - \frac{1 - A_\Theta^+ - A_\Theta^-}{1 - A_\downarrow^+ - A_\downarrow^-}\,.
$$

Subroutine 5. Calculating B_A^+, B_A^-, and B_A^* from C^+ and C^- for C in $\mathscr{S}_A$:

$$
\begin{aligned}
B_A^+ &= 1 - K\Big(1 + \sum_{\substack{C \in \mathscr{S}_A \\ C \neq B}} C^+/(1 - C^+)\Big)\,,\\
B_A^- &= 1 - K(1 - B^-)/(1 - B^+)\,,\\
B_A^* &= 1 - K\Big(1 + \sum_{\substack{C \in \mathscr{S}_A \\ C \neq B}} C^+/(1 - C^+) - \prod_{\substack{C \in \mathscr{S}_A \\ C \neq B}} C^-/(1 - C^+)\Big)\,,
\end{aligned}
$$

where

$$K^{-1} = 1 + \sum_{C \in \mathcal{S}_A} C^+/(1 - C^+)\,.$$

Subroutine 6. Calculating B_Θ^+ and B_Θ^- from $A_\downarrow^+$, $A_\downarrow^-$, $A_\diamond^+$, $A_\diamond^-$, B_A^+, B_A^- and B_A^*, where B is a daughter of A:

$$B_\Theta^+ = K(A_\diamond^+(B_A^* - A_\downarrow^-) + (1 - A_\diamond^+ - A_\diamond^-)B_A^+)\,,$$
$$B_\Theta^- = 1 - K(1 - A_\diamond^-)(1 - B_A^-)\,,$$

where

$$K^{-1} = 1 - A_\downarrow^+ A_\diamond^- - A_\downarrow^- A_\diamond^+\,.$$

4.5. Miscellaneous comments

(1) The constant K in Subroutine 5 is the same as the constant K in Subroutine 1. Recognition of this fact will save computation on the way back down the tree, since we store $\mathrm{Bel}_A^\downarrow(A) = 1 - K$ on our way up the tree. It is probably most efficient, in fact, to store K^{-1} or

$$K^{-1} - 1 = \sum_{B \in \mathcal{S}_A} B^+/(1 - B^+)$$

instead of $\mathrm{Bel}_A^\downarrow(A)$.

(2) Each sum or product in Subroutine 5 differs from the corresponding product in Subroutine 1 only by the omission of a single term or factor. So if we save the sums and products from Subroutine 1, we can obtain those in Subroutine 5 by subtraction and division. This may be advantageous when the sib sizes are large.

(3) In our description of the procedure for moving up the tree, we specified that $\mathrm{Bel}_A^\downarrow(A)$ and $\mathrm{Bel}_A^\downarrow(\bar{A})$ should be calculated first for those A whose daughters are all terminal, then for those A whose daughters are either terminal or else have only terminal daughters, and so on. In fact, however, we have more freedom of choice than this. In order to calculate $\mathrm{Bel}_A^\downarrow(A)$ and $\mathrm{Bel}_A^\downarrow(\bar{A})$ it is necessary only that these quantities should already have been calculated for each nonterminal daughter of A.

(4) We could move up the tree faster if we were to calculate $(\mathrm{Bel}_A^\downarrow)_{\{A,\bar{A}\}}$ for disjoint A in parallel. A similar opportunity for parallelism occurs when we move back down the tree, provided we want to move down all the branches.

(5) We set out to calculate only $\mathrm{Bel}_\Theta^\downarrow(A)$ for all A in $\mathcal{A}$. As it turned out, we also calculated $\mathrm{Bel}_\Theta^\downarrow(\bar{A})$, since this was necessary for calculating the values of $\mathrm{Bel}_\Theta^\downarrow$ for A's daughters. (This means we can calculate the plausibility of A, $\mathrm{Pl}_\Theta^\downarrow(A) = 1 - \mathrm{Bel}_\Theta^\downarrow(\bar{A})$.) A glance at (24) and (27) makes it clear that we can also calculate $\mathrm{Bel}_\Theta^\downarrow(B)$ for any B that is in the field $(\mathcal{S}_A \cup \{\bar{A}\})'$ for some node A. In general, however, there will remain many subsets B of Θ for which our method is not helpful. It does, not, for example, help us calculate $\mathrm{Bel}_\Theta^\downarrow(\{d, f\})$ in Fig. 3.

(6) We have used Barnett's technique in Subroutines 1, 3, and 5. (Subroutine 2 can also be regarded as an application of Barnett's technique, but there is really no distinction between Barnett's technique and brute-force calculation of an orthogonal sum when we are working with a single dichotomy.) However, we have used this technique only on the partitions $\mathcal{S}_A \cup \{\bar{A}\}$. If the sibs $\mathcal{S}_A$ are all relatively small — if, say, no sib contains more than three or four daughters — then those calculations would be manageable even without Barnett's technique. Thus, the efficiency of our algorithm is mainly due not to Barnett's technique but to the fact that we are able to break the overall computation down into local computations.

4.6. Complexity analysis

It is clear from our description of the algorithm that the amount of arithmetic involving a particular node does not depend on the size of the tree. It depends only on the number of the node's daughters, and it increases linearly with the number of daughters. (Subroutine 1, for example has a product with a factor for each daughter and a sum with a term for each daughter.) It follows that the computational complexity of the algorithm is linear in the number of nodes in the tree.

We can make a closer complexity analysis if we assume that the number of daughters in a sib (the branching factor) is constant throughout the tree. Let f denote the branching factor. Let n denote the number of sibs or, equivalently, the number of nonterminal nodes. Then we can expect $a + bf$ arithmetic operations for each sib, or $n(a + bf)$ altogether, where a and b are positive constants.

This formula clarifies the role of Barnett's technique in our algorithm. Barnett's technique is responsible for the linearity with respect to the sib size f, while the localization of the computation is responsible for the linearity with respect to the number of sibs, n. If we did not use Barnett's technique, the computational complexity would be exponential in f but stll proportional to n. In place of $n(a + bf)$, we would have $n\exp(a + bf)$.

Instead of talking about the number of arithmetic operations per sib, we might wish to talk about the number per node. Since there are $nf + 1$ nodes altogether, this is

$$\frac{n(a + bf)}{nf + 1} \approx \frac{n(a + bf)}{nf} = \frac{a}{f} + b\,. \tag{31}$$

Alternatively, we might wish to talk about the number of operations per terminal node, since the number of terminal nodes is the size of our frame. Since there are $nf - n + 1$ terminal nodes, the number of operations per terminal node is

$$\frac{n(a + bf)}{nf - n + 1} \approx \frac{n(a + bf)}{n(f - 1)} = \frac{a}{f - 1} + \frac{bf}{f - 1}\,. \tag{32}$$

Both (31) and (32) are greatest for binary trees ($f = 2$) and tend towards b as f increases. (There is no paradox here. When f is large, most nodes are terminal nodes.)

The formula '$a + bf$ for the number of operations per sib can be verified empirically. We have verified it using a LISP implementation in a variety of trees, with f ranging up to 5 and n ranging up to 30,000. The fit was excellent, with 99.8% of the variance explained. The least squares estimates were $a = 158$ and $b = 44$. (Strictly speaking, the counts on which these estimates were based are counts of arguments in operations rather than counts of operations. Thus an addition of k terms counts as k, and the division of one number by another counts as 2.)

As we mentioned in Section 4.3, it is often possible to save computation by moving down only some of the branches in the third stage. Since Subroutine 5 involves the greatest computation, the savings can be substantial.

4.7. Comparison with Gordon and Shortliffe's algorithm

Gordon and Shortliffe [4] do not give details for the implementation of their algorithms. We have found, however, that it can also be implemented in linear time. The particular implementation we have used is analogous to the implementation of our own algorithm: it involves movements up and down the tree. We have found that this implementation of Gordon and Shortliffe's algorithm is comparable in complexity to our algorithm. In all the trees we checked it required fewer arithmetic operations than our algorithm, but never fewer than half as many.

The details of our implementation of Gordon and Shortliffe's algorithm are nearly as complicated as the details of our algorithm, and it is possible that a more efficient implementation might be found.

5. Generalizations

In this article, we have retained Gordon and Shortliffe's assumption that the belief functions being combined are simple support functions focused on nodes or their complements. The essence of our computational scheme can be retained, however, whenever each belief function is carried by a sib (more precisely, by a partition $\mathscr{S}_A \cup \{\bar{A}\}$ for some node A). Under this more general assumption, Barnett's technique is no longer available, and the amount of

arithmetic is exponential in the sib size, but it remains proportional to the number of sibs. An interesting special case occurs when each belief function is conditionally Bayesian—i.e., when the belief function Bel_A carried by $\mathscr{S}_A \cup \{\bar{A}\}$ satisfies

$$\mathrm{Bel}_A(B \mid A) + \mathrm{Bel}_A(\bar{B} \mid A) = 1$$

and

$$\mathrm{Bel}_A(B \mid \bar{A}) + \mathrm{Bel}_A(\bar{B} \mid \bar{A}) = 1$$

for every element B of the field $(\mathscr{S}_A \cup \{\bar{A}\})^*$. In this case, the result of combining all the belief functions is Bayesian, and the computations can be simplified; the amount of arithmetic is again linear rather than exponential in the sib size. This case has been studied by Pearl [7].

A further generalization is to replace diagnostic trees with general trees of partitions or variables. We need only a "Markov" property: a given node in the tree should discern the interaction among the belief functions on the different branches of tree separated by the node. The problem of propagating belief functions in such Markov trees is discussed by Shenoy and Shafer [14] and by Shafer, Shenoy, and Mellouli [13]. The Bayesian special case is discussed by Pearl [8].

The generalization to networks of variables has been studied by Kong [5]; see also Mellouli, Shafer, and Shenoy [6]. The last chapter of Kong [5] is of particular interest; it shows how the algorithm of this article can be generalized, without loss of computational efficiency, to the case where a patient may have more than one disease.

ACKNOWLEDGMENT

Research for this article has been partially supported by NSF grants IST-8405210 and IST-8610293 and by ONR grant N00014-85-K-0490. The article has been improved by the thoughtful criticism of a referee. Chuck Hand and Khaled Mellouli assisted with the complexity analysis in Section 4.6.

REFERENCES

1. Barnett, J.A., Computational methods for a mathematical theory of evidence, in: *Proceedings IJCAI-81*, Vancouver, BC (1981) 868–875.
2. Garvey, T.D., Lowrance, J.D. and Fischler, M.A., An inference technique for integrating knowledge from disparate sources, in: *Proceedings IJCAI-81*, Vancouver, BC (1981) 319–325.
3. Gordon, J. and Shortliffe, E.H., The Dempster–Shafer theory of evidence, in: B.G. Buchanan and E.H Shortliffe (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley, Reading, MA, 1985) 272–292.
4. Gordon, J. and Shortliffe, E.H., A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence* **26** (1985) 323–357.
5. Kong, A., Multivariate belief functions and graphical models, Doctoral Dissertation, Department of Statistics, Harvard University, Cambridge, MA, 1986.

6. Mellouli, K., Shafer, G. and Shenoy, P., Qualitative Markov networks, in: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Paris (1986) 31–35.
7. Pearl, J., On evidential reasoning in a hierarchy of hypotheses, *Artificial Intelligence* **28** (1986) 9–15.
8. Pearl, J., Fusion, propagation, and structuring in belief networks, *Artificial Intelligence* **29** (1986) 241–288.
9. Shafer, G., *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, 1976.
10. Shafer, G., Belief functions and possibility measures, in: J.C. Bizdek (Ed.), *The Analysis of Fuzzy Information* **2** (CRC Press, 1987).
11. Shafer, G., Probability judgment in artificial intelligence and expert systems, *Stat. Sci.* **2** (1987) 3–16.
12. Shafer, G., Hierarchical evidence, in: *Proceedings Second Conference on Artificial Intelligence Applications*, Miami, FL (1985) 16–21.
13. Shafer, G., Shenoy, P. and Mellouli, K., Propagating belief functions in qualitative Markov trees, Working Paper No. 190, School of Business, University of Kansas, Lawrence, KS, 1987.
14. Shenoy, P. and Shafer, G., Propagating belief functions with local computations, *IEEE Expert* **1**(3) (1986) 43–52.

Received August 1985; revised version received March 1987