# Propagating Belief Functions in Qualitative Markov Trees

**Glenn Shafer, Prakash P. Shenoy, and Khaled Mellouli**
*School of Business, University of Kansas*

## ABSTRACT

*This article is concerned with the computational aspects of combining evidence within the theory of belief functions. It shows that by taking advantage of logical or categorical relations among the questions we consider, we can sometimes avoid the computational complexity associated with brute-force application of Dempster's rule.*

*The mathematical setting for this article is the lattice of partitions of a fixed overall frame of discernment. Different questions are represented by different partitions of this frame, and the categorical relations among these questions are represented by relations of qualitative conditional independence or dependence among the partitions. Qualitative conditional independence is a categorical rather than a probabilistic concept, but it is analogous to conditional independence for random variables.*

*We show that efficient implementation of Dempster's rule is possible if the questions or partitions for which we have evidence are arranged in a qualitative Markov tree—a tree in which separations indicate relations of qualitative conditional independence. In this case, Dempster's rule can be implemented by propagating belief functions through the tree.*

KEYWORDS: ***Bayesian propagation, belief functions, causal trees, Dempster-Shafer theory, Dempster's rule, combination of evidence, qualitative independence, partitions, qualitative Markov trees, diagnostic trees, probability, propagation of evidence***

Research for this article has been partially supported by NSF grants IST-8405210 and IST-8610293, by ONR grant N00014-85-K-0490 and by a Research Opportunities in Auditing grant from the Peat, Marwick, Mitchell Foundation. The authors have profited from conversations and correspondence with Paul Cohen, Arthur Dempster, Augustine Kong, and Judea Pearl. They have also profited from comments by anonymous referees.

*Address correspondence to Glenn Shafer, School of Business, University of Kansas, Lawrence, KS 66045-2003.*

International Journal of Approximate Reasoning 1987; 1:349–400

52 Vanderbilt Ave., New York, NY 10017 0888-613X/87/$3.50

## INTRODUCTION

A practical problem may involve a number of related questions. The evidence bearing on the problem may consist of a number of independent arguments and other items. Typically, each item of evidence bears directly on only a few of the questions. It is natural in this situation to assess the force of each item of evidence separately, obtaining in each case probability judgments about the questions on which that evidence directly bears. But since all the questions are related, it is then necessary to combine the probability judgments.

In this article, we study this task of combination within the theory of belief functions (Shafer [1]), with an emphasis on computational problems. We assume that the items of evidence are independent and that each item has been used to construct a belief function for the questions on which it bears. The theory of belief functions tells us that these belief functions, since they are independent, can be combined by Dempster's rule. Unfortunately, a brute-force application of Dempster's rule may be computationally infeasible in a problem involving many questions. We show how to take advantage of logical or categorical relations among the questions to reduce the computation.

We show that if the different questions are arranged in a tree, the structure of which captures the categorical relations among the questions, then the combination of the belief functions can be accomplished step by step, with each step involving only a small number of directly related questions. The task of combining the belief functions reduces, in effect, to the task of propagating them through the tree.

The propagation scheme we present here generalizes the scheme for diagnostic trees studied by Shafer and Logan [2] and the slightly more general scheme for hierarchical evidence studied by Shafer [3]. It also generalizes Pearl's scheme for propagating the effects of new data in Bayesian causal trees [4–8]. (Recall that the theory of belief functions is a generalization of the Bayesian theory.) We believe this general scheme will prove useful in expert systems, where a compromise is often needed between the modularity of production rules and the ideal structure of Bayesian models. Belief function models offer one such compromise, and propagation in trees makes this compromise feasible in a variety of problems.

Our exposition is relatively abstract. We treat related questions as related partitions of a single frame of discernment. We call this the *partitive* formalism, and we contrast it with the *multivariate* formalism, which uses variables rather than partitions. We use the partitive formalism because we feel its abstractness makes for mathematical clarity. The multivariate formalism is closer to aplications, but we only sketch it in this article. For a more thorough development, see Kong [9]. For more on the relation between the two formalisms, see Mellouli [10].

In the next section, we give a mathematical treatment of partitions, emphasizing qualitative conditional independence. This is a logical or categorical concept rather than a probabilistic concept, but it is analogous to conditional independence for random variables. In the third section we study qualitative Markov trees of partitions, which are analogous to Markov trees of random variables. Following that we review the mathematical theory of belief functions, with an emphasis on partitions. In the final section we use the background provided by the preceding three sections to give a concise description of our general propagation scheme.

Our purpose in this article is to provide a general mathematical understanding that unifies a number of practical approaches. This general level of understanding is somewhat distant from actual implementation. When working at the level of implementation, we would expect to replace the general ideas and terminology used here with more concrete ideas and more specific terminology selected to fit the nature of the particular problem. In some cases, we might use diagnostic or fault trees; in other cases we might use variables in causal trees; and so on.

One question only touched on here is how to deal with networks of variables or partitions when these networks are not trees. In the Bayesian case, this has been studied by Pearl [6] and by Lauritzen and Spiegelhalter [11]. The computational problems are more severe in the general belief-function case than in the Bayesian case, but in both cases it seems to be necessary to collapse the networks to trees so that the general method we give here can be applied. Kong [9] has initiated the study of how to do this efficiently in the general belief-function case. See also Mellouli [10].

## PARTITIONS

We begin this section with a discussion of how to formalize relations among questions. We distinguish between the multivariate formalism, which begins with a number of relatively simple questions or variables, and the partitive formalism, which begins with a set of alternative comprehensive statements and treats variables as functions that are defined on and hence partition this set.

After this discussion, we turn to the mathematics of the partitive formalism. We review some basic facts about partitions, and we study qualitative independence for partitions.

In the section on belief functions, we will use the notation and concepts of the partitive formalism to explain what is involved in coarsening, extending, and projecting belief functions. These operations are important in both the partitive and multivariate formalisms, but we have found that we can explain them most succinctly using the partitive formalism. This is only a manner of exposition, of course. Once the operations are understood, they can be used freely in the multivariate formalism.

## The Multivariate and Partitive Formalisms

Suppose $\Theta$ is a finite set of possible answers to some question, and suppose we know that one and only one of these answers can be correct. In this case we call $\Theta$ a *frame of discernment*, or simply a *frame*. Adoption of a frame involves an assumption that can usually be challenged—the assumption that exactly one of the elements of the frame is the right answer to the question.

The act of adopting a frame for a question formalizes a *variable*. The elements of the frame are the possible values of the variable. For example, when we adopt the set {red, white, yellow} as our frame for the question "What color rose is Bill wearing today?," we formalize the variable "the color rose Bill is wearing today," with possible values, red, white, and yellow.

Questions, related or unrelated, can be conjoined. We can conjoin the question of what color rose Bill is wearing with the question of what color shirt he is wearing, obtaining the joint question, "What color rose and what color shirt is Bill wearing today?" The frame we adopt for such a joint question should, of course, be consistent with the frame we adopt for the individual questions. If we adopt the frame $\Theta_1 = \{\text{red, white, yellow}\}$ for the question about the rose, and the frame $\Theta_2 = \{\text{white, blue}\}$ for the question about the shirt, then we might adopt the Cartesian product

$$\Theta_1 \times \Theta_2 = \{(\text{red, white}), (\text{red,blue}), (\text{white,white}), (\text{white,blue}), (\text{yellow,white}), (\text{yellow,blue})\}$$

for the joint question. Alternatively, we might rule out the possibility of Bill's wearing either a red rose on a blue shirt or a white rose on a white shirt and adopt the smaller frame

$$\Theta = \{(\text{red,white}), (\text{white,blue}), (\text{yellow,white}), (\text{yellow,blue})\}$$

which is a subset of the Cartesian product $\Theta_1 \times \Theta_2$.

The *multivariate formalism* uses Cartesian products freely. In order to study together variables $X_1, \ldots, X_n$, which are individually formalized by frames $\Theta_1, \ldots, \Theta_n$, the formalism introduces a joint variable $(X_1, \ldots, X_n)$ and adopts as the frame for this joint variable either the Cartesian product $\Theta_1 \times \cdots \times \Theta_n$ or a subset of it.

The multivariate formalism has a flexibility that is useful in applied statistics and other practical work. It allows us to formalize a problem step by step, introducing new questions and corresponding variables as the need arises.

When we are concerned with general theory rather than applications, however, the proliferation of frames that accompanies the flexibility of the multivariate formalism can seem cumbersome. Consider, for example, the case of a variable that is a function of variables we have already introduced. In practical work, it is often useful to introduce such a variable explicitly, giving it its own name and frame, even though knowledge of its value would add nothing to knowledge

of the values of the previous variables, and even though its explicit introduction complicates our notation. Suppose, for example, that we have a variable $X_1$ = "the color rose Bill is wearing," with the frame $\Theta_1$ = {red, white, yellow}. For practical reasons, we may want to consider explicitly both $X_1$ and the variable $X_2$ = "whether Bill is wearing a red rose," with the frame $\Theta_3$ = {yes, no}. This shifts our attention from the simple frame $\Theta_1$ to the Cartesian product $\Theta_1 \times \Theta_3$ or its subset {(red, yes), (white, no), (yellow, no)}. This shift may be useful, but from an abstract point of view it is a complication that adds nothing in meaning; the new alternatives {(red, yes), (white, no), (yellow, no)} have the same meaning as the original alternatives {red, white, yellow}.

The *partitive formalism,* which we use in this article, avoids this proliferation of notation. It is based on the assumption that we have a fixed overall frame $\Theta$ that is detailed enough to take account of all the questions we want to consider in a particular problem. Since we begin with this frame rather than with a question, we think of it as a set of statements rather than as a set of answers. Adopting the frame means assuming that exactly one of these statements is true. If, for example, we are thinking about Bill, we might begin with a frame $\Theta$ consisting of four statements: $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, where

$\theta_1$ = Bill is wearing a red rose on a white shirt

$\theta_2$ = Bill is wearing a white rose on a blue shirt

$\theta_3$ = Bill is wearing a yellow rose on a white shirt

$\theta_4$ = Bill is wearing a yellow rose on a blue shirt

In this formalism, a variable is a function or mapping defined on the frame $\Theta$. For example, the variable $X_1$, "the color rose Bill is wearing," is a mapping that maps $\theta_1$ to "red," $\theta_2$ to "white," and $\theta_3$ and $\theta_4$ to "yellow."

On reflection, we see that the concept of a variable is not essential in the partitive formalism. What is essential is not the values to which the variable maps but how it partitions the frame on which it is defined. Our variable $X_1$ partitions the frame $\Theta$ into three subsets, the singletons $\{\theta_1\}$ and $\{\theta_2\}$ and the pair $\{\theta_3,\theta_4\}$. Once we know that this is the partition corresponding to $X_1$, we can tell from the $\theta_i$ themselves that $X_1$ is telling us what color rose Bill is wearing.

The partitive formalism concentrates on the frame $\Theta$ and on partitions of $\Theta$. The frame $\Theta$ itself corresponds to a very detailed question, and each partition corresponds to a less detailed question.

Advanced mathematical treatments of probability theory (e.g., Breiman [12]) generally use the partitive formalism, for the same reason we are using it. But these works are largely concerned with infinite sets and continuous probability measures, and hence they emphasize fields (also called algebras) of subsets rather than partitions. A *field* of subsets of a set $\Theta$ is a set of subsets that contains both $\Theta$ and the empty set $\varnothing$, contains $A$'s complement $A^c$ whenever it contains $A$,

and contains $A \cap B$ and $A \cup B$ whenever it contains both $A$ and $B$. In the case where $\Theta$ is finite, there is a one-to-one correspondence between partitions and fields. Given a partition, we obtain the field by taking all unions of elements of the partition. Given a field, we obtain the partition by taking the *atoms* of the field, the elements of the field that contain no smaller element of the field other than the empty set. In the case where $\Theta$ is infinite, however, this correspondence breaks down. Continuous probability measures on infinite sets usually involve fields that do not contain all unions of their atoms (Halmos [13]).

In this article we are concerned only with the case where $\Theta$ is finite. We therefore emphasize partitions and treat fields, which are more complex, as secondary and derivative.

## The Lattice of Partitions

We will now establish a notation for talking about partitions and review some well-known facts about partitions and variables. Throughout our discussion, $\Theta$ will be a fixed finite nonempty set.

Let us begin by recalling that a set $\mathfrak{P}$ of subsets of $\Theta$ is a *partition* of $\Theta$ if the sets in $\mathfrak{P}$ are all nonempty and disjoint and their union is $\Theta$.

A *variable* on $\Theta$ is a mapping from $\Theta$ to some other set. A variable on $\Theta$ induces a partition of $\Theta$, the partition obtained by grouping together elements that are mapped to the same value by the variable. We will let $\mathfrak{P}_X$ denote the partition induced by the variable $X$:

$$\mathfrak{P}_X = \{X^{-1}(x) | x \in X(\Theta)\}$$

Given a partition $\mathfrak{P}$, it is easy to construct a variable $X$ that induces it; we choose a set $\Omega$ that has the same number of elements as $\mathfrak{P}$, we set up a one-to-one correspondence between the elements of $\mathfrak{P}$ and the elements of $\Omega$, and we have $X$ map each element $\theta$ of $\Theta$ to the element of $\Omega$ corresponding to the element of $\mathfrak{P}$ that contains $\theta$. Different choices of $\Omega$ produce different variables, but all these variables have $\mathfrak{P}$ as their induced partition.

Let us write $\mathfrak{P}_1 \leq \mathfrak{P}_2$ whenever $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are partitions of $\Theta$ and for every $P_1 \in \mathfrak{P}_1$ there exists $P_2 \in \mathfrak{P}_2$ such that $P_1 \subseteq P_2$. This means that each element of $\mathfrak{P}_2$ is a union of elements of $\mathfrak{P}_1$. When $\mathfrak{P}_1 \leq \mathfrak{P}_2$, we say that $\mathfrak{P}_1$ is a *refinement* of $\mathfrak{P}_2$ and $\mathfrak{P}_2$ is a *coarsening* of $\mathfrak{P}_1$. We also say that $\mathfrak{P}_1$ is *finer* than $\mathfrak{P}_2$ and $\mathfrak{P}_2$ is *coarser* than $\mathfrak{P}_1$.

One partition being coarser than another is the same as one variable being a function of another. Here is a more precise way of saying this: Suppose $X$ and $Y$ are variables. The variable $X$ maps $\Theta$ to $\Omega_1$, and the variable $Y$ maps $\Theta$ to $\Omega_2$. Then $\mathfrak{P}_Y$ is coarser than $\mathfrak{P}_X$ if and only if there exists a mapping $f$ from $\Omega_1$ to $\Omega_2$ such that $Y(\theta) = f(X(\theta))$ for every element of $\theta$ of $\Theta$.

Given a partition $\mathfrak{P}$ of $\Theta$, let $\mathfrak{P}^*$ denote the set consisting of all unions of elements of $\mathfrak{P}$. As we have already pointed out, $\mathfrak{P}^*$ is a *field* of subsets of $\Theta$;

it contains both $\Theta$ and the empty set, it contains $A$'s complement whenever it contains $A$, and it contains $A \cap B$ and $A \cup B$ whenever it contains both $A$ and $B$. Notice that $\mathfrak{P}_1 \leq \mathfrak{P}_2$ if and only if $\mathfrak{P}_2^* \subseteq \mathfrak{P}_1^*$. (This reversal may seem in appropriate to some readers; perhaps we should write $\mathfrak{P}_2 \leq \mathfrak{P}_1$ instead of $\mathfrak{P}_1 \leq \mathfrak{P}_2$ when $\mathfrak{P}_1$ is finer than $\mathfrak{P}_2$. However, it is standard in the mathematical literature to write $\mathfrak{P}_1 \leq \mathfrak{P}_2$ when $\mathfrak{P}_1$ is finer than $\mathfrak{P}_2$.)

The relations $\leq$ partially orders the set of all partitions of $\Theta$. With this partial order, these partitions form a lattice (Birkhoff [14]). This means that any two partitions $\mathfrak{P}_1$ and $\mathfrak{P}_2$ have a greatest lower bound and a least upper bound. The greatest lower bound $\mathfrak{P}_1 \wedge \mathfrak{P}_2$ is also called the *coarsest common refinement* of $\mathfrak{P}_1$ and $\mathfrak{P}_2$. Its elements are intersections of the elements of $\mathfrak{P}_1$ and $\mathfrak{P}_2$:

$$\mathfrak{P}_1 \wedge \mathfrak{P}_2 = \{P_1 \cap P_2 | P_1 \in \mathfrak{P}_1,\ P_2 \in \mathfrak{P}_2,\ \text{and } P_1 \cap P_2 \neq \emptyset\}$$

The least upper bound $\mathfrak{P}_1 \vee \mathfrak{P}_2$ is also called the *finest common coarsening* of $\mathfrak{P}_1$ and $\mathfrak{P}_2$. It is most easily described in terms of its corresponding field: $(\mathfrak{P}_1 \vee \mathfrak{P}_2)^* = \mathfrak{P}_1^* \cap \mathfrak{P}_2^*$.

Notice that $\mathfrak{P}_Z \geq \mathfrak{P}_X \wedge \mathfrak{P}_Y$ if and only if $Z$ is a function of $X$ and $Y$.

The lattice of partitions of $\Theta$ contains a coarsest partition and a finest partition. The coarsest partition is $\{\Theta\}$, the set whose only element is $\Theta$ itself. The finest partition is $\{\{\theta\} | \theta \in \Theta\}$, the set of singleton subsets of $\Theta$. (This partition contains the same numbr of elements as $\Theta$ does, but it is not exactly the same set as $\Theta$.)

If $\mathfrak{P}$ is a partition of $\Theta$, and $A$ is a subset of $\Theta$, then we can obtain a partition of $A$ by intersecting the elements of $\mathfrak{P}$ with $A$ and discarding those intersections that are empty. Let this partition of $A$ be denoted by $\mathfrak{P}(A)$;

$$\mathfrak{P}(A) = (A \cap P | P \in \mathfrak{P},\ A \cap P \neq \emptyset\}$$

If we learn that the true statement in $\Theta$ is actually in the subset $A$, then we have the option of thinking of $A$ as our frame in place of $\Theta$, and then the question previously formalized by the partition $\mathfrak{P}$ of $\Theta$ will be formalized by the partition $\mathfrak{P}(A)$ of $A$.

## Qualitative Independence

In this section, we introduce the concepts of qualitative independence and qualitative conditional independence. We use the adjective "qualitative" in order to distinguish these concepts from the analogous probabilistic concepts. Qualitative independence is a property of just the partitions or variables involved. Probabilistic independence, in contrast, depends on the choice of a particular probability distribution; whether two random variables are independent depends on their joint probability distribution.

Two partitions $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are *qualitatively independent* if $P_1 \cap P_2 \neq \emptyset$ whenever $P_1 \in \mathfrak{P}_1$ and $P_2 \in \mathfrak{P}_2$. This means that knowing which element of $\mathfrak{P}_1$ contains the truth tells us nothing about which element of $\mathfrak{P}_2$ contains the

truth; being in one particular element of $\mathfrak{P}_1$ does not rule out being in any particular element of $\mathfrak{P}_2$. Two partitions $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are *qualitatively conditionally independent* given a third partition $\mathfrak{P}$ (or simply qualitatively independent given $\mathfrak{P}$) if $\mathfrak{P}_1(P)$ and $\mathfrak{P}_2(P)$ are qualitatively independent for every element $P$ of $\mathfrak{P}$. This means that once we are told which element of $\mathfrak{P}$ contains the truth, knowing which element of $\mathfrak{P}_1$ contains the truth tells us nothing further about which element of $\mathfrak{P}_2$ contains the truth. It follows directly from these definitions that $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent given $\mathfrak{P}$ if and only if $P \cap P_1 \cap P_2 \neq \emptyset$ whenever $P \in \mathfrak{P}$, $P_1 \in \mathfrak{P}_1$, $P_2 \in \mathfrak{P}_2$, $P \cap P_1 \neq \emptyset$, and $P \cap P_2 \neq \emptyset$.

We write $[\mathfrak{P}_1, \mathfrak{P}_2]\dashv$ to indicate that $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent. We write $[\mathfrak{P}_1, \mathfrak{P}_2]\dashv\mathfrak{P}$ to indicate that $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent given $\mathfrak{P}$. Notice that $[\mathfrak{P}_1\ \mathfrak{P}_2]\dashv$ is equivalent to $[\mathfrak{P}_1\ \mathfrak{P}_2]\dashv\{\Theta\}$.

Much of the theory we are about to study could just as well be stated in terms of variables. Instead of saying that $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent given $\mathfrak{P}$ and writing $[\mathfrak{P}_1, \mathfrak{P}_2]\dashv\mathfrak{P}$, we could say that $X_1$ and $X_2$ are qualitatively independent given $X$ and write $[X_1, X_2]\dashv X$, where $X_1$, $X_2$, and $X$ are variables such that

$$\mathfrak{P}_{X_1} = \mathfrak{P}_1, \qquad \mathfrak{P}_{X_2} = \mathfrak{P}_2 \qquad \text{and } \mathfrak{P}_X = \mathfrak{P}$$

Notice that there are many variables corresponding to a given partition, and the same qualitative independence relations hold for all of them.

Previous work in the theory of belief functions has used a different terminology for qualitative independence. Instead of saying that $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent given $\mathfrak{P}$, this work (Shafer [1,3] and Shafer and Logan [2]) has said that $\mathfrak{P}$ discerns the interaction relevant to itself between $\mathfrak{P}_1$ and $\mathfrak{P}_2$. This terminology obscures the analogy to probabilistic independence, but it makes explicit one aspect of the role the concept of qualitative independence plays in the management of evidence. As we shall see later (Theorem 4), we can use $\mathfrak{P}$ as our frame for combining evidence bearing directly on $\mathfrak{P}_1$ and $\mathfrak{P}_2$ if $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent given$\mathfrak{P}$.

Qualitative conditional independence is also equivalent to what is called "embedded multivalued dependency" is the theory of relational databases (Fagan [15] and Maier [16]). In that theory the relation $[X_1\ X_2]\dashv X$ is written $X \twoheadrightarrow X_1 | X_2$. Some of the properties that we will derive here have also been derived in that theory , but the different notation and point of view lead to different emphases. In particular, the work in relational databases has not brought out the analogy with probabilistic conditional independence that we emphasize.

Qualitative independence relations arise nationally in the multivariate formalism. Here are three examples.

EXAMPLE 1 *Suppose $\Theta_1$ and $\Theta_2$ are nonempty sets, and set $\Theta$ equal to the Cartesian product $\Theta_1 \times \Theta_2$. Set $\mathfrak{P}_1 = \{\{\theta_1\} \times \Theta_2 | \theta_1 \in \Theta_1\}$, and set $\mathfrak{P}_2 = \{\Theta_1 \times \{\theta_2\} | \theta_2 \in \Theta_2\}$. The partitions $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent, since $(\{\theta_1\} \times \Theta_2) \cap (\Theta_1 \times \{\theta_2\} = \{(\theta_1, \theta_2) \neq \emptyset$.*

EXAMPLE 2 *Suppose* $\Theta_1$, $\Theta_2$, *and* $\Theta_3$ *are nonempty sets, and set* $\Theta$ *equal to the Cartesian product* $\Theta_1 \times \Theta_2 \times \Theta_3$. Set $\mathfrak{P}_1 = \{\{\theta_1\} \times \{\theta_2\} \times \Theta_3 | \theta_1 \in \Theta_1, \theta_2 \in \Theta_2\}$, $\mathfrak{P}_2 = \{\Theta_1 \times \{\theta_2\} \times \{\theta_3\} | \theta_2 \in \Theta_2, \theta_3 \in \Theta_3\}$, and $\mathfrak{P} = \{\Theta_1 \times \{\theta_2\} \times \Theta_3 | \theta_2 \in \Theta_2\}$. *The partitions* $\mathfrak{P}_1$ *and* $\mathfrak{P}_2$ *are not qualitatively independent. However* $\mathfrak{P}_1$ *and* $\mathfrak{P}_2$ *are qualitatively independent given* $\mathfrak{P}$.

EXAMPLE 3 *Suppose we formalize a problem along the lines of Example 2, and afterwards we obtain evidence that rules out certain values for the pair* $(\theta_1, \theta_2)$ *and other evidence that rules out certain values for the pair* $(\theta_2, \theta_3)$. *Let* A *denote the subset of* $\Theta_1 \times \Theta_2$ *consisting of elements not ruled out by the first item of evidence, and let* B *denote the subset* $\Theta_2 \times \Theta_3$ *consisting of elements not ruled out by the second item of evidence. Then we may decide to simplify the formalization of our problem by adopting* $\Theta' = (A \times \Theta_3) \cap (\Theta_1 \times B)$, *a subset of* $\Theta$, *as our frame. If we do this, then the three questions that were formalized by* $\mathfrak{P}_1$, $\mathfrak{P}_2$, *and* $\mathfrak{P}$ *will now be formalized by* $\mathfrak{P}_1(\Theta')$, $\mathfrak{P}_2(\Theta')$, *and* $\mathfrak{P}(\Theta')$. *The qualitative independence relation will be preserved, however;* $\mathfrak{P}_1(\Theta')$ *and* $\mathfrak{P}_2(\Theta')$ *will be qualitatively independent given* $\mathfrak{P}(\Theta')$.

*There is another interesting conditional independence relation in this example. Define partitions* $\mathfrak{Q}_1$, $\mathfrak{Q}_2$, *and* $\mathfrak{Q}_3$ *of* $\Theta_1 \times \Theta_2 \times \Theta_3$ *by setting* $\mathfrak{Q}_1 = \{\{\theta_1\} \times \Theta_2 \times \Theta_3 | \theta_1 \in \Theta_1\}$, *etc. Then* $\mathfrak{Q}_1(\Theta')$ *and* $\mathfrak{Q}_3(\Theta')$ *will not, in general, be qualitatively independent. They will, however, be qualitatively independent given* $\mathfrak{Q}_2(\Theta')$.

The definition of qualitative independence involves categorical relations, not probabilities. But, as the following lemma shows, probabilistic independence for random variables does imply qualitative independence for the induced partitions.

LEMMA 1 *Suppose* X *and* Y *are variables defined on* $\Theta$. *Then* $[\mathfrak{P}_X, \mathfrak{P}_Y]\dashv$ *if and only if there exists a probability distribution Pr on* $\Theta$ *such that (i)* X *and* Y *are independent with respect to Pr and (ii)* $Pr(\{\theta\}) > 0$ *for all* $\theta \in \Theta$. *(Recall our assumption that* $\Theta$ *is finite.)*

Proof First assume that there exists such a probability distribution Pr. Let $P \in \mathfrak{P}_X$ and $Q \in \mathfrak{P}_Y$. Since $X$ and $Y$ are independent with respect to Pr, $\Pr(P \cap Q) = \Pr(P)\Pr(Q)$. Since $\Pr(P) > 0$ and $\Pr(Q) > 0$, it follows that $\Pr(P \cap Q) > 0$. Hence $P \cap Q \neq \varnothing$.

Now assume $[\mathfrak{P}_X, \mathfrak{P}_Y]\dashv$. We must construct a probability distribution Pr satisfying the stated conditions. For every $\theta \in \Theta$, there exist unique elements $P$ of $\mathfrak{P}_X$ and $Q$ of $\mathfrak{P}_Y$ such that $\theta \in P \cap Q$. Set

$$\Pr(\{\theta\}) = |P \cap Q|^{-1} \, |\mathfrak{P}_X|^{-1} \, \{\mathfrak{P}_Y|^{-1}$$

(Here $|A|$ denotes the number of elements in $A$.) The numbers $\Pr(\{\theta\})$ are positive. It is easy to see that they add to 1, thus defining a probability distribution, and that $X$ and $Y$ are independent with respect to this distribution. ■

COROLLARY *Supposee* X, Y, *and* Z *are variables defined on* $\Theta$. *Then* $[\mathfrak{P}_X, \mathfrak{P}_Y]\perp\mathfrak{P}_Z$ *if and only if there exists a probability distribution Pr on* $\Theta$ *such that (i)* X *and* Y *are independent given* Z *with respect to Pr and (ii)* $Pr(\{\theta\}) > 0$ *for all* $\theta \in \Theta$.

The concepts of qualitative independence generalize readily to collections of more than two partitions. We say that the partitions $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are qualitatively independent if $P_1 \cap \cdots \cap P_n \neq \emptyset$ whenever $P_i \in \mathfrak{P}_i$, for $i = 1, \ldots, n$. We say that the partitions $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are qualitatively independent given $\mathfrak{P}$ if the partitions $\mathfrak{P}_1(P), \ldots, \mathfrak{P}_n(P)$ are qualitatively independent for every $P$ in $\mathfrak{P}$. Notice that when $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are qualitatively independent (or qualitatively independent given $\mathfrak{P}$), any smaller group selected from them are also.

We write $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\perp$ to indicate that $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are qualitatively independent, and we write $[\mathfrak{P}_i]_{i\in N}\perp$ to indicate that the partitions in an indexed collection $\{\mathfrak{P}_i\}_{i\in N}$ are qualitatively independent. (This notation allows some of the $\mathfrak{P}_i$ to be identical, but in fact two identical partitions of $\Theta$ cannot be independent unless they are both equal to $\{\Theta\}$.) We write $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\perp\mathfrak{P}$ and $[\mathfrak{P}_i]_{i\in N}\perp\mathfrak{P}$ to indicate relations of qualitative conditional independence.

## Projections

Set

$$A^{\mathfrak{P}} = \cup\{P | P\in\mathfrak{P},\ P\cap A\neq\emptyset\}$$

and

$$A_{\mathfrak{P}} = \cup\{P | P\in\mathfrak{P},\ P\subseteq A\}$$

Notice that $A_{\mathfrak{P}} \subseteq A \subseteq A^{\mathfrak{P}}$. In fact, $A^{\mathfrak{P}}$ is the smallest element of $\mathfrak{P}^*$ containing $A$, and $A_{\mathfrak{P}}$ is the largest element of $\mathfrak{P}^*$ contained in $A$. Notice also that $A^{\mathfrak{P}} \subseteq B$ if and only if $A \subseteq B_{\mathfrak{P}}$.

Shafer [1, pp. 117–119] called $A_{\mathfrak{P}}$ and $A^{\mathfrak{P}}$ "outer and inner reductions," respectively. Pawlak [17, 18] called the pair $(A_{\mathfrak{P}}, A^{\mathfrak{P}})$ a "rough set." Here we will be interested primarily in $A^{\mathfrak{P}}$, which we will call the *projection* of $A$ onto the field $\mathfrak{P}^*$.

EXAMPLE 4 *Set* $\Theta = \Theta_1 \times \Theta_2$, *and* $\mathfrak{P} = \{\{\theta_1\} \times \Theta_2 | \theta_1 \in \Theta_1\}$. *Then the projection* $A^{\mathfrak{P}}$ *is equal to* $B \times \Theta_2$, *where* $B = \{\theta_1 \in \Theta_1 | (\theta_1, \theta_2) \in A$ *for some* $\theta_2 \in \Theta_2\}$.

LEMMA 2 *The following statements are equivalent.*

1. $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\perp\mathfrak{P}$.
2. If $P \in \mathfrak{P}$, $S_i \in \mathfrak{P}_i^*$ and $P \cap S_i \neq \emptyset$ *for* $i = 1, \ldots, n$, *then* $P \cap (S_1 \cap \cdots \cap S_n) \neq \emptyset$.
3. $S_1^{\mathfrak{P}} \cap \cdots \cap S_n^{\mathfrak{P}} = (S_1 \cap \cdots \cap S_n)^{\mathfrak{P}}$ *whenever* $S_i$ *is an element of* $\mathfrak{P}_i^*$, $i = 1, \ldots, n$.

**4.** $\mathrm{P}_1^{\mathfrak{P}} \cap \cdots \cap \mathrm{P}_n^{\mathfrak{P}} = (\mathrm{P}_1 \cap \cdots \cap \mathrm{P}_n)^{\mathfrak{P}}$ *whenever* $\mathrm{P}_i$ *is an element of* $\mathfrak{P}_i$, $\mathrm{i} = 1, \ldots, \mathrm{n}$.

Proof We will show that statement 1 implies 2, 2 implies 3, 3 implies 4, and 4 implies 1.

Suppose statement 1 holds. Suppose $P \in \mathfrak{P}$, $S_i \in \mathfrak{P}_i^*$ and $P \cap S_i \neq \varnothing$ for $i = 1, \ldots, n$. For each $i$, we can choose $P_i \in \mathfrak{P}_i$ such that $P_i \subseteq S_i$ and $P \cap P_i \neq \varnothing$. By statement 1, $P \cap (P_1 \cap \cdots \cap P_n) \neq \varnothing$. Since $P \cap (P_1 \cap \cdots \cap P_n) \subseteq P \cap (S_1 \cap \cdots \cap S_n)$, it follows that $P \cap (S_1 \cap \cdots \cap S_n) \neq \varnothing$.

Now suppose statement 2 holds, and choose $S_i$ in $\mathfrak{P}_i^*$ for $i = 1, \ldots, n$. We have

$$(S_1 \cap \cdots \cap S_n)^{\mathfrak{P}} = \cup \{P | P \in \mathfrak{P},\ P \cap (S_1 \cap \cdots \cap S_n) \neq \varnothing\}$$

and

$$S_1^{\mathfrak{P}} \cap \cdots \cap S_n^{\mathfrak{P}} = \cap \{\cup \{P | P \in \mathfrak{P},\ P \cap S_i \neq \varnothing\} | i = 1, \ldots, n\}$$
$$\cup \{P | P \in \mathfrak{P},\ P \cap S_i \neq \varnothing \text{ for } i = 1, \ldots, n\}$$

By statement 2, these are equal.

It is obvious that statement 3 implies statement 4, because any element of $\mathfrak{P}_i$ is also an element of $\mathfrak{P}_i^*$.

Finally, suppose statement 4 holds. Suppose $P \in \mathfrak{P}$, $P_i \in \mathfrak{P}_i$, and $P \cap P_i \neq \varnothing$ for $i = 1, \ldots, n$. Then $P \subseteq P_1^{\mathfrak{P}} \cap \cdots \cap P_n^{\mathfrak{P}}$. Hence $P \subseteq (P_1 \cap \cdots \cap P_n)^{\mathfrak{P}}$, and hence $P \cap P_1 \cap \cdots \cap P_n \neq \varnothing$. ■

Suppose $[\mathfrak{P}_1, \mathfrak{P}_2] \dashv \mathfrak{P}$, and suppose we are interested in projecting an element $S_1$ of $\mathfrak{P}_1^*$ to field $\mathfrak{P}_2^*$. The next lemma tells us that we will get the right answer if we first project $S_1$ to $\mathfrak{P}^*$ and then project the resulting projection to $\mathfrak{P}_2^*$. In order to understand the practical implications of this lemma, think of $S_1$ as information about the question represented by $\mathfrak{P}_1$. The lemma tells us that in order to deduce the implications of this information for the question represented by $\mathfrak{P}_2$, we do not need to work in the frame $\Theta$; instead we can work in the smaller frame $\mathfrak{P}$. This lemma will be used later to prove Theorem 5, which extends the idea from the categorical information represented by an element of $\mathfrak{P}_1^*$ to the probabilistic information represented by a belief function carried by $\mathfrak{P}_1$.

LEMMA 3 *Suppose* $[\mathfrak{P}_1, \mathfrak{P}_2] \dashv \mathfrak{P}$, *and* $\mathrm{S}_1 \in \mathfrak{P}_1^*$. *Then* $\mathrm{S}_1^{\mathfrak{P}_2} = (\mathrm{S}_1^{\mathfrak{P}})^{\mathfrak{P}_2}$.

Proof Using the definitions together with statement 3 of Lemma 2, we obtain

$$\begin{aligned}(S_1^{\mathfrak{P}})^{\mathfrak{P}_2} &= \cup \{P_2 | P_2 \in \mathfrak{P}_2,\ P_2 \cap S_1^{\mathfrak{P}} \neq \varnothing\} \\ &= \cup \{P_2 | P_2 \in \mathfrak{P}_2,\ (P_2 \cap S_1)^{\mathfrak{P}} \neq \varnothing\} \\ &= \cup \{P_2 | P_2 \in \mathfrak{P}_2,\ P_2 \cap S_1 \neq \varnothing\} \\ &= S_1^{\mathfrak{P}_2}\end{aligned}$$

■

## Properties of Qualitative Independence

We now study some properties of qualitative independence. These properties are analogous to properties of probabilistic independence, and many of them could be proved using Lemma 1, but we will give more elementary proofs.

We know that functions of independent random variables are independent. If $X_1$ and $X_2$ are independent random variables, then $f(X_1)$ and $g(X_2)$ are also independent random variables. More generally, if $X_1, \ldots, X_n$ are independent, $\{N_1, \ldots, N_k\}$ is a partition of the set $\{X_1, \ldots, X_n\}$, and $Y_j$ is a function of the $X_i$ in $N_j$, then $Y_1, \ldots, Y_k$ are independent. Yet, more generally, if the $X_i$ are independent given $Z$ and each $Y_j$ is a function of $Z$ together with the $X_i$ in $N_j$, then the $Y_i$ are independent given $Z$. The following lemma and corollary make analogous statements for qualitative independence.

LEMMA 4 *Suppose* $\mathfrak{P}_1, \ldots, \mathfrak{P}_n, \mathfrak{Q}_1, \ldots, \mathfrak{Q}_k$ *are partitions of* $\Theta$. *Suppose* $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv$, *and suppose* $\mathrm{N}_1, \ldots, \mathrm{N}_k$ *are disjoint subsets of* $\{1, \ldots, \mathrm{n}\}$. *If* $\mathfrak{Q}_j \geq \bigwedge\{\mathfrak{P}_i | \mathrm{i} \in \mathrm{N}_j\}$ for $\mathrm{j} = 1, \ldots, \mathrm{k}$, *then* $[\mathfrak{Q}_1, \ldots, \mathfrak{Q}_k]\dashv$.

Proof Suppose $Q_j \in \mathfrak{Q}_j$ for $j = 1, \ldots, k$. Since $\mathfrak{Q}_j \geq \bigwedge\{\mathfrak{P}_i | i \in N_j\}$, $Q_j$ is a union of sets of the form $\bigcap\{P_i | i \in N_j\}$. Choose one such set for ech $Q_j$. Since the $N_j$ are disjoint, this gives us a unique choice of $P_i$ for all $i$ in $\bigcup N_j$. Since $\{\mathfrak{P}_i\}_{i \in \bigcup N_j}\dashv$, $\bigcap\{P_i | i \in \bigcup N_j\} \neq \varnothing$. But $\bigcap\{P_i | i \in \bigcup N_j\} \subseteq Q_1 \cap \cdots \cap Q_k$. Hence $Q_1 \cap \cdots \cap Q_k \neq \varnothing$. ■

COROLLARY *If* $\mathrm{N}_1, \ldots, \mathrm{N}_k$ *are disjoint subsets of* $\{1, \ldots, \mathrm{n}\}$, $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv\mathfrak{P}$, *and* $\mathfrak{Q}_j \geq \mathfrak{P} \wedge (\bigwedge\{\mathfrak{P}_i | \mathrm{i} \in \mathrm{N}_j\})$, *then* $[\mathfrak{Q}_1, \ldots, \mathfrak{Q}_k]\dashv\mathfrak{P}$.

Proof If $P \in \mathfrak{P}$,

$$(\mathfrak{P} \wedge (\bigwedge\{\mathfrak{P}_i | i \in N_j\}))(P) = \bigwedge\{\mathfrak{P}_i(P) | i \in N_j\}$$

The hypotheses of the corollary therefore tell us that $[\mathfrak{P}_1(P), \ldots, \mathfrak{P}_n(P)]\dashv$ and that $\mathfrak{Q}_j(P) \geq \bigwedge\{\mathfrak{P}_i(P) | i \in N_j\}$. It follows from the lemma that $[\mathfrak{Q}_1(P), \ldots, \mathfrak{Q}_k(P)]\dashv$. ■

If $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv$, then $[\mathfrak{P}_i, \mathfrak{P}_j]\dashv$ for all distinct $i$ and $j$. The converse is not true, of course. If, for example, $\Theta = \{a,b,c,d\}$, $\mathfrak{P}_1 = \{\{a,b\},\{c,d\}\}$, $\mathfrak{P}_2 = \{\{a,c\},\{b,d\}\}$, and $\mathfrak{P}_3 = \{\{a,d\}\},\{b,c\}\}$, then the three relations $[\mathfrak{P}_1, \mathfrak{P}_2]\dashv$ $[\mathfrak{P}_1, \mathfrak{P}_3]\dashv$, and $[\mathfrak{P}_2, \mathfrak{P}_3]\dashv$ hold, but $[\mathfrak{P}_1, \mathfrak{P}_2, \mathfrak{P}_3]\dashv$ does not hold.

The following lemma shows that the qualitative independence of $n$ partitions is nevertheless equivalent to a set of pairwise qualitative independence relations.

LEMMA 5 *The following statements are equivalent.*

1. $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv$.
2. $[\mathfrak{P}_1, \ldots, \mathfrak{P}_{n-1}]\dashv$ *and* $[\mathfrak{P}_n, \wedge\{\mathfrak{P}_i | \mathrm{i} = 1, \ldots, \mathrm{n} - 1\}]\dashv$.

**3.** $[\mathfrak{P}_j, \bigwedge \{\mathfrak{P}_i | \mathrm{i} \in \mathrm{N} - \{\mathrm{j}\}\}]\dashv$ *for all* $\mathrm{j} \in \mathrm{N}$, *where* $\mathrm{N} = \{1, \ldots, \mathrm{n}\}$.
**4.** $[\mathfrak{P}_j, \bigwedge \{\mathfrak{P}_i | \mathrm{i} = 1, \ldots, \mathrm{j} - 1\}]\dashv$ *for* $\mathrm{j} = 2, \ldots, \mathrm{n}$.

Proof It follows directly from Lemma 4 that statement 1 implies statement 2, 2 implies 3, and 3 implies 4. So we need only show that statement 4 implies statement 1.

Assume statement 4 holds. Choose elements $P_i$ of $\mathfrak{P}_i$ for $i = 1, \ldots, n$. We need to show that $P_1 \cap \cdots \cap P_n \neq \varnothing$. Since $P_1 \neq \varnothing$, it suffices to show that if $P_1 \cap \cdots \cap P_{j-1} \neq \varnothing$, then $P_1 \cap \cdots \cap P_j \neq \varnothing$. But $P_1 \cap \cdots \cap P_{j-1}$, if it is nonempty, is an element of the partition $\bigwedge \{\mathfrak{P}_i | i = 1, \ldots, j - 1\}$, and hence the relation $[\mathfrak{P}_j, \bigwedge\{\mathfrak{P}_i | i = 1, \cdots, j - 1\}]\dashv$ implies that $P_1 \cap \cdots \cap P_j \neq \varnothing$. ■

COROLLARY *The four statements in the lemma remain equivalent if each independence relation is changed to conditional independence given* $\mathfrak{P}$.

In general, unconditional independence of random variables $X_1, \ldots, X_n$ does not imply their independence given another random variable $Z$. They will be independent given $Z$, however, if $Z$ is a function of one of the $X_i$. The situation is similar for qualitative independence:

LEMMA 6 *Suppose* $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv$, *and suppose* $\mathfrak{Q} \geq \mathfrak{P}_1$. *Then* $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv\mathfrak{Q}$.

Proof Choose elements $Q \in \mathfrak{Q}$ and $P_i \in \mathfrak{P}_i$ such that $Q \cap P_i \neq \varnothing$ for $i = 1, \ldots, n$. Since $\mathfrak{Q} \geq \mathfrak{P}_1$, $P_1 \subseteq Q$. Hence $Q \cap \mathrm{P}_1 \cap \cdots \cap P_n = P_1 \cap \cdots \cap P_n \neq \varnothing$. ■

COROLLARY *Suppose* $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv\mathfrak{P}$, *and suppose* $\mathfrak{P} \geq \mathfrak{Q} \geq \mathfrak{P} \wedge \mathfrak{P}_1$. *Then* $[\mathfrak{P}_1, \ldots, \mathfrak{P}_n]\dashv\mathfrak{Q}$.

Proof The hypotheses of this corollary can be restated by saying that for every element $P$ of $\mathfrak{P}$, $[\mathfrak{P}_1(P), \ldots, \mathfrak{P}_n(P)] \dashv$ and $\mathfrak{Q}(P) \geq \mathfrak{P}_1(P)$. By the lemma, this implies that $[\mathfrak{P}_1(P), \ldots, \mathfrak{P}_n(P)]\dashv\mathfrak{Q}(P)$. This means that $[\mathfrak{P}_1(P \cap Q), \ldots, \mathfrak{P}_n(P \cap Q)]\dashv$ whenever $P$ is in $\mathfrak{P}$ and $Q$ is in $\mathfrak{Q}$. Since $\mathfrak{P} \geq \mathfrak{Q}$, this is the same as saying that $[\mathfrak{P}_1(Q), \ldots, \mathfrak{P}_n(Q)]\dashv$ whenever $Q$ is in $\mathfrak{Q}$. And this is the conclusion of the corollary. ■

If $X$, $Y$, and $Z$ are random variables, $X$ and $Y$ are independent, and $X$ and $Z$ are independent given $Y$, then $X$ is independent of $Y$ and $Z$. Lemma 7 says the same thing for partitions.

LEMMA 7 *If* $[\mathfrak{P}_1, \mathfrak{P}_2]\dashv$ *and* $[\mathfrak{P}_1, \mathfrak{P}_3]\dashv\mathfrak{P}_2$, *then* $[\mathfrak{P}_1, \mathfrak{P}_2 \wedge \mathfrak{P}_3]\dashv$.

Proof Suppose $P_1 \in \mathfrak{P}_1$, $P_2 \in \mathfrak{P}_2$, and $P_3 \in \mathfrak{P}_3$ such that $P_2 \cap P_3 \neq \varnothing$. We need to show that $P_1 \cap P_2 \cap P_3 \neq \varnothing$. Since $[\mathfrak{P}_1, \mathfrak{P}_2]\dashv$, $P_1 \cap P_2 \neq \varnothing$. Since $P_2 \cap P_3 \neq \varnothing$ and we assumed that $[\mathfrak{P}_1, \mathfrak{P}_3]\dashv\mathfrak{P}_2$, we have $P_1 \cap P_2 \cap P_3 \neq \varnothing$. ■

The next lemma is bothersomely complex, too complex to correspond to familiar properties of probabilistic independence. We give it here because we will need it in the proof of Theorem 1.

LEMMA 8 *Suppose* $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ *are partitions of* $\Theta$. *Suppose* $i_1$ *and* $i_2$ *are elements of* $\{1, \ldots, n\}$, *and* $\mathfrak{N}_1$ *and* $\mathfrak{N}_2$ *are partitions of* $\{1, \ldots, n\}$. *Suppose* $[\wedge\{\mathfrak{P}_i | i \in \alpha\}]_{\alpha \in \mathfrak{N}_1} \dashv \mathfrak{P}_{i_1}$ *and* $[\wedge\{\mathfrak{P}_i | i \in \beta\}]_{\beta \in \mathfrak{N}_2} \dashv \mathfrak{P}_{i_2}$. *Then* $[\wedge\{\mathfrak{P}_i | i \in \delta\}]_{\delta \in \mathfrak{N}_1 \wedge \mathfrak{N}_2} \dashv \mathfrak{P}_{i_1} \wedge \mathfrak{P}_{i_2}$.

Proof Choose an element $Q$ of $\mathfrak{P}_{i_1} \wedge \mathfrak{P}_{i_2}$ and elements $Q_\delta$ of $\wedge\{\mathfrak{P}_i | i \in \delta\}$ such that $Q \cap Q_\delta \neq \varnothing$ for all $\delta \in \mathfrak{N}_1 \wedge \mathfrak{N}_2$. We need to show that $Q \cap (\cap\{Q_\delta | \delta \in \mathfrak{N}_1 \wedge \mathfrak{N}_2\}) \neq \varnothing$.

There exist unique elements $R_{i_1}$ of $\mathfrak{P}_{i_1}$ and $R_{i_2}$ of $\mathfrak{P}_{i_2}$ such that $Q = R_{i_1} \cap R_{i_2}$. And there exist unique elements $P_i$ of $\mathfrak{P}_i$ such that $Q_\delta = \cap\{P_i | i \in \delta\}$ for all $\delta \in \mathfrak{N}_1 \wedge \mathfrak{N}_2$. Since $Q \cap Q_\delta \neq \varnothing$, we have $R_{i_1} = P_{i_1}$ and $R_{i_2} = P_{i_2}$. So our task reduces to showing that if $P_i \in \mathfrak{P}_i$ and

$$P_{i_1} \cap P_{i_2} \cap (\cap\{P_i | i \in \delta\}) \neq \varnothing \qquad \text{for all } \delta \in \mathfrak{N}_1 \wedge \mathfrak{N}_2 \tag{1}$$

then $P_1 \cap \cdots \cap P_n \neq \varnothing$.

Choose $\alpha \in \mathfrak{N}_1$. Combining our hypothesis that $[\wedge\{\mathfrak{P}_i | i \in \beta\}]_{\beta \in \mathfrak{N}_2} \dashv \mathfrak{P}_{i_2}$ with Lemma 4 we have

$$[\wedge\{\mathfrak{P}_i | i \in \beta \cap (\alpha \cup \{i_1\})\}]_{\beta \in \mathfrak{N}_2} \dashv \mathfrak{P}_{i_2} \tag{2}$$

(We use the convention that $\wedge\{\mathfrak{P}_i | i \in \varnothing\} = \{\Theta\}$.) But Eq. (1) tells us that

$$\varnothing \neq P_{i_1} \cap P_{i_2} \cap (\cap\{P_i | i \in \alpha \cap \beta\}) = P_{i_2} \cap (\cap\{P_i | i \in \beta \cap (\alpha \cup \{i_1\})))$$

for all $\beta \in \mathfrak{N}_2$ such that $\alpha \cap \beta \neq \varnothing$. Hence (2) tells us that

$$\varnothing \neq P_{i_2} \cap (\cap\{P_i | i \in \alpha \cup \{i_1\}\}) = P_{i_1} \cap P_{i_2} \cap (\cap\{P_i | i \in \alpha\})$$

and therefore

$$P_{i_1} \cap (\cap\{P_i | i \in \alpha\}) \neq \varnothing \tag{3}$$

We have established (3) for all $\alpha \in \mathfrak{N}_1$. Since $[\wedge\{\mathfrak{P}_i | i \in \alpha\}]_{\alpha \in \mathfrak{N}_1} \dashv \mathfrak{P}_{i_1}$, it follows that $P_1 \cap \cdots \cap P_n \neq \varnothing$. ■

Notice that it is necessary, for the conclusion of the lemma to hold, that $\mathfrak{N}_1$ and $\mathfrak{N}_2$ each fully partition $\{1, \ldots, n\}$; it is not sufficient that they consist of disjoint subsets of $\{1, \ldots, n\}$.

The following corollary follows from Lemma 8 by induction on $k$.

COROLLARY *Suppose* $i_1, \ldots, i_k$ *are elements of* $\{1, \ldots, n\}$ *and* $\mathfrak{N}_1, \ldots, \mathfrak{N}_k$ *are partitions of* $\{1, \ldots, n\}$. *Suppose* $[\wedge\{\mathfrak{P}_i | i \in \alpha\}]_{\alpha \in \mathfrak{N}_j} \dashv \mathfrak{P}_{i_j}$ *for* $j = 1, \ldots, k$. *Then* $[\wedge\{\mathfrak{P}_i | i \in \delta\}]_{\delta \in \mathfrak{N}_1 \wedge \cdots \wedge \mathfrak{N}_k} \dashv \mathfrak{P}_{i_1} \wedge \cdots \wedge \mathfrak{P}_{i_k}$.

How many further properties of qualitative conditional independence are there?

Infinitely many. Perhaps there are really only a finite number, and the others can be deduced from these. Sagiv and Walecka [19] have shown, however, that the concept does not have a finite axiomatization using Horn clauses.

## QUALITATIVE MARKOV TREES

We call a tree of partitions a qualitative Markov tree if the structure of the tree indicates conditional independence relations among the partitions. Partitions and groups of partitions that are connected in the tree only through a given partition must be conditionally independent given that partition.

As we shall see, qualitative Markov trees arise naturally from diagnostic and causal trees. In other cases, we can construct them by collapsing networks.

Our interest in qualitative Markov trees will be justified in a later section, where we will show how belief functions can be propagated in such trees.

We begin this section by looking at alternative ways of stating precisely the condition that a tree of partitions be qualitative Markov. Then we study some ways of changing such trees without disturbing the Markov property, we show in detail how qualitative Markov trees arise from diagnostic trees, and we sketch some ways qualitative Markov trees arise in the multivariate formalism.

### Properties of Qualitative Markov Trees

We will use standard graph-theoretic terminology and notation for trees. Recall that a *graph* or *network* is a pair $(N, E)$, where $N$ is a finite set and $E$ is a set of unordered pairs of distinct elements of $N$. (In other words, each element of $E$ is a two-element subset of $N$.) The elements of $N$ are *nodes*, and the elements of $E$ are *edges*. A graph is a *tree* if it is connected and has no cycles. When $\{i, j\} \in E$, we say that $i$ and $j$ are *adjacent* and that each is a *neighbor* of the other. A node is a *leaf* if it has exactly one neighbor. Let $V_i$ denote the set of all neighbors of $i$.

If $M$ is a subset of $N$, $i$ is not in $M$, but $i$ is a neighbor of an element of $M$, then we say that $i$ is a neighbor of $M$. The set of nodes consisting of $M$ together with all its neighbors is called the *closure* of $M$. Let $V_M$ denote the set of all neighbors of $M$, and let $M^{\text{cl}}$ denote the closure of $M$; $M^{\text{cl}} = M \cup V_M$.

If $N_0, N_1, \ldots, N_r$ are subsets of $N$, and every path from a node in $N_i$ to a node in $N_j$ goes via some node in $N_0$ whenever $i \neq j$, we say that $N_0$ *separates* $N_1, \ldots, N_r$. If $N_0$ separates $N_1, \ldots, N_r$, then $N_1 - N_0, \ldots, N_r - N_0$ are disjoint.

If a node $n$ in a tree $(N, E)$ is not a leaf, then when we remove it (and the edges incident to it) from the tree, the graph that remains will consist of two or more isolated subtrees, one for each neighbor of $n$. Let $(N_{k,n}, E_{k,n})$ denote the subtree containing the neighbor $k$.

Suppose $\{\mathfrak{P}_i\}_{i\in N}$ is a finite collection of partitions, and suppose $(N, E)$ is a tree. (This notation permits some of the $\mathfrak{P}_i$ to be identical.) We say that $(N, E)$ is a *qualitative Markov tree* for $\{\mathfrak{P}_i\}_{i\in N}$ if for every $n \in N$,

$$[\wedge\{\mathfrak{P}_i | i\in N_{k,n}\}]_{k\in V_n} \dashv \mathfrak{P}_n \tag{4}$$

THEOREM 1 *Given a finite collection of partitions* $\{\mathfrak{P}_i\}_{i\in N}$ *and a tree* (N, E)*, the following conditions are equivalent.*

1. *Whenever* $N_1, \ldots, N_r$ *are separated by* $N_0$,

$$[\wedge\{\mathfrak{P}_i | i\in N_1\}, \ldots, \wedge\{\mathfrak{P}_i | i\in N_r\}] \dashv \wedge\{\mathfrak{P}_i | i\in N_0\}$$

2. *Whenever* $N_1$ *and* $N_2$ *are represented by* $N_0$,

$$[\wedge\{\mathfrak{P}_i | i\in N_1\}, \wedge\{\mathfrak{P}_i | i\in N_2\}] \dashv \wedge\{\mathfrak{P}_i | i\in N_0\} \tag{5}$$

3. *Whenever* $N_1$ *and* $N_2$ *are separated by* $N_0$ *and the three sets are disjoint, (5) holds.*
4. *For every subset* M *of* N,

$$[\wedge\{\mathfrak{P}_i | i\in M\}, \wedge\{\mathfrak{P}_i | i\in N - M^{cl}\}] \dashv \wedge\{\mathfrak{P}_i | i\in V_M\}$$

5. *(*N, E*) is qualitative Markov for* $\{\mathfrak{P}_i\}_{i\in N}$.

Proof Statement 2 is a special case of statement 1, 3 is a special case of 2, and 4 is a special case of 1. Hence it suffices to show that statement 4 implies 5 and that 5 implies 1.

To show that 4 implies 5, we use Lemma 5. Suppose $n \in N$, and suppose $k$ is a neighbor of $n$; $k \in V_n$. Then $n$ is the only neighbor of $N_{k,n}$. So substituting $N_{k,n}$ for $M$ in statement 4 yields

$$[\wedge\{\mathfrak{P}_i | i\in N_{k,n}\}, \wedge\{\mathfrak{P}_i | i\in N - (N_{k,n}\cup\{n\})\}] \dashv \mathfrak{P}_n$$

or

$$\{\wedge\{\mathfrak{P}_i | i\in N_{k,n}\} \wedge \{\wedge\{\mathfrak{P}_i | i\in N_{k',n}\} | k'\in V_n - \{k\}\}] \dashv \mathfrak{P}_n$$

Since this holds for every element $k$ of $V_n$, (4) follows from the corollary to Lemma 5.

To show that statement 5 implies statement 1, we use Lemmas 4 and 7. Suppose $N_1, \ldots, N_r$ are separated by $N_0$. For each node $n$ in $N_0$, set $\mathfrak{N}_n = \{n\} \cup \{N_{k,n} | k \in V_n\}$; this is the partition of $N$ consisting of $n$ together with the disjoint subtrees that remain when $n$ is removed from the tree. Let $\mathfrak{N}$ denote the partition $\wedge\{\mathfrak{N}_n | n \in N_0\}$; this partition consists of the singleton subsets of $N_0$ together with the disjoint subtrees that remain when $N_0$ is removed from the tree. Now use $N_1, \ldots, N_r$ to define a partition $\{R_0, R_1, \ldots, R_{r+1}\}$ of $\mathfrak{N}$. Let $R_0$ be the set of singleton subsets of $N_0$. For $j = 1, \ldots, r$, let $R_j$ be the set of elements of $\mathfrak{N}$ that are contained in $N_j$ but are disjoint from $N_0$. Let $R_{r+1} =$

$\mathfrak{R} - (R_0 \cup \cdots \cup R_r)$. Then $N_j \subseteq \cup\{\delta | \delta \in R_0 \cup R_j\}$, and hence

$$\begin{aligned} \wedge\{\mathfrak{P}_i | i \in N_j\} &\geq \wedge\{\wedge\{\mathfrak{P}_i | i \in \delta\} | \delta \in R_0 \cup R_j\} \\ &= (\wedge\{\mathfrak{P}_i | i \in N_0\}) \wedge (\wedge\{\{\mathfrak{P}_i | i \in \delta\} | \delta \in R_j\}) \end{aligned} \tag{6}$$

So if we assume statement 5 is true, then we have

$$[\wedge\{\mathfrak{P}_i | i \in \alpha\}]_{\alpha \in \mathfrak{R}_n} \dashv \mathfrak{P}_n$$

for all $n \in N_0$. So by the corollary to Lemma 8,

$$[\wedge\{\mathfrak{P}_i | i \in \delta\}]_{\delta \in \mathfrak{R}} \dashv \wedge\{\mathfrak{P}_i | i \in N_0\}$$

Statement 1 follows from (6) and the corollary to Lemma 4. ■

It turns out that conditions 1 thorugh 4 of Theorem 1 are also equivalent when the graph $(N, E)$ is not necessarily a tree. When these conditions hold, we call the graph a *qualitative Markov network*. We will discuss such networks briefly in the section on multivariate Markov trees. For more information, see Mellouli [10].

In a private communication, Pearl has pointed out two more characteristics of qualitative Markov trees:

**6.** *For any pair of nonadjacent nodes* n *and* m *and any set* $N_0$ *consisting of one or more nodes lying along the path between* n *and* m,

$$[\mathfrak{P}_n, \mathfrak{P}_m] \dashv \wedge\{\mathfrak{P}_i | i \in N_0\}$$

**7.** *There exists a node* r *in the tree such that if we think of movement away from* r *as descent, then for every node* n *in the tree,*

$$[\mathfrak{P}_n, \wedge\{\mathfrak{P}_i | i \in A(n)\}] \dashv \mathfrak{P}_{m(n)}$$

*where* m(n) *is the mother of* n, *and* A(n) *consists of the nondescendants of* n.

It is obvious from our thereom that these conditions are necessary for a tree to be qualitative Markov; in fact, the second condition is true for any node $r$. That either condition is sufficient for a tree to be qualitative Markov can be deduced from the lemmas in the preceding section (Partitions) or from theorems given by Pearl and Verma [20].

The remainder of this section can be omitted on a first reading by those who are willing to grant the importance of qualitative Markov trees and want to move on as quickly as possible to the computational scheme presented under the heading Propagation in Trees.

## Transformations of Qualitative Markov Trees

The questions for which we have evidence sometimes change. We may obtain evidence about new questions, or we may discard given evidence because it has been discredited. If we are using a qualitative Markov tree because its partitions correspond to questions for which we have evidence, then these changes may lead us to make changes in the tree. We will not want these changes to destroy the Markov property.

We now consider some ways we can change a qualitative Markov tree while preserving the Markov property.

SUBTREES The simplest thing we can do to a tree is remove a leaf. This does not affect the separations in the tree; if $N_0$, $N_1$, and $N_2$ are sets of nodes in the tree that remains, then $N_0$ separates $N_1$ and $N_2$ in the tree that remains if and only if it did so in the original tree. It follows from statement 2 of Theorem 1 that if the original tree was qualitative Markov, then the one that remains is also.

By successively removing leaves from a tree, we can obtain any subtree. Hence any subtree of a qualitative Markov tree is qualitative Markov.

CONTRACTION There is another way to make a qualitative Markov tree smaller. We replace a subtree of the tree with a single new node, say $m$, and associate with $m$ the coarsest common refinement of the partitions that had been associated with the nodes of the subtree; if $N_0$ is the set of nodes in the subtree, then $\wedge\{\mathfrak{P}_i | i \in N_0\}$ is the partition associated with $m$. We introduce edges connecting $m$ to those nodes outside the subtree that were neighbors of nodes in the subtree. Let us call the new tree a *contraction* of the old one. We can see that a contraction of a qualitative Markov tree is also a qualitative Markov tree by noting that (4) holds for every node $n$ in the new tree. Indeed, if $n$ is the new node $m$, then (4) is the same as statement 1 of Theorem 1, applied to the old tree, with $N_1 \ldots, N_r$ the nodes in the $r$ disjoint subtrees that remain when $N_0$ is removed. And if $n$ is not $m$—if $n$ is one of the nodes remaining from the old tree—then the partitions $\{\wedge\{\mathfrak{P}_i | i \in N_{k,n}\}\}_{k \in V_n}$ in the new tree are the same as in the old, hence (4) holds in the new tree because it holds in the old.

An important special case of contraction occurs when $N_0$ consists of two neighboring nodes $i$ and $j$ and $\mathfrak{P}_i$ is coarser than $\mathfrak{P}_j$. In this case the partition associated with the new node is simply $\mathfrak{P}_i$, and hence the contraction amounts to the removal of $j$ from the tree.

DELETION Suppose we replace a subtree by a single node $m$, as in the preceding paragraph. But we associate a coarser partition with $m$. Instead of the partition $\wedge\{\mathfrak{P}_i | i \in N_0\}$, we use the partition $\wedge\{\mathfrak{P}_i | i \in M\}$, where $M$ is the subset of $N_0$ consisting of those elements that have neighbors outside $N_0$. Then the result is

still a qualitative Markov tree. To see this, we review the reasoning of the preceding paragraph. We see that (4) holds when $n$ is equal to $m$, because $M$ is sufficient to separate the isolated subtrees that remain when $N_0$ is removed from the old tree. It holds when $n$ is not equal to $m$, because the partitions $\{\wedge\{\mathfrak{P}_i | i \in N_{k,n}\}\}_{k \in V_n}$ in the new tree are still the same as in the old tree, except for the one for the neighbor $k$ for which $N_{k,n}$ contains the new node, and the difference there is only that $N_0 - M$ is subtracted from $N_{k,n}$ and hence $\wedge\{\mathfrak{P}_i | i \in N_{k,n}\}$ is coarsened. Independent partitions remain independent when they are coarsened.

It is natural to call this kind of reduction of a qualitative Markov tree a *deletion*. We are deleting the nodes in $N_0 - M$, in the sense that we are no longer taking account of the partitions associated with these nodes. Removal of a leaf is a special case of deletion; here $N_0$ consists of the leaf together with its only neighbor and $M$ consists of just the neighbor.

We can do things intermediate between contraction and deletion. We can associate with the new node $m$ any partition $\mathfrak{P}_m$ that satisfies $\wedge\{\mathfrak{P}_i | i \in N_{k,n}\} \geq \mathfrak{P}_m \geq \wedge\{\mathfrak{P}_i | i \in M\}$.

INTERPOLATION OF REFINEMENTS Suppose $i$ and $j$ are neighbors in a qualitative Markov tree. Suppose we interpolate a new node, say $m$, in the edge between $i$ and $j$. (This means that we add $m$ to $N$, remove $\{i,j\}$ from $E$, and add $\{i,m\}$ and $\{j,m\}$ to $E$.) Suppose we associate with $m$ the partition $\mathfrak{P}_i \wedge \mathfrak{P}_j$. Then we again have a qualitative Markov tree. To see that this is true, we again think about (4) for each node $n$ of the new tree. If $n$ is not equal to $i$, $j$, or $m$, then these three nodes all lie on the same branch from $n$, and since $\mathfrak{P}_i$ and $\mathfrak{P}_j$ were already on this branch, the addition of $\mathfrak{P}_i \wedge \mathfrak{P}_j$ will not change the refinement of all the partitions on it. If $n$ is equal to $m$, $i$, or $j$, then we must appeal to Lemma 4. Consider two cases:

1. Supose $n$ is equal to $m$. Compare the branches separated by $m$ in the new tree with the branches separated by $\{i,j\}$ in the old tree. The only difference is the addition of $i$ to one branch and $j$ to another in the new tree. The refinements of the partitions along these branches in the old tree are conditionally independent given $\mathfrak{P}_i \wedge \mathfrak{P}_j$ by statement 1 of Theorem 1. By the corollary to Lemma 4, addition of $\mathfrak{P}_i$ or $\mathfrak{P}_j$ to one or more of the branches does not affect this conditional independence.
2. Now suppose $n$ is equal to $i$ or $j$, say $i$. Compare the branches separated by $i$ in the new tree with those separated by $i$ in the old tree. The only difference is the addition of $m$ to the branch containing $j$. This has the effect of adding $\mathfrak{P}_i$ to the partitions along this branch, and we can again appeal to the corollary to Lemma 4 to see that this does not affect the conditional independence given $\mathfrak{P}_i$.

Notice that if we interpolate refinements between every pair of neighbors,

except those pairs in which one of the partitions is already a refinement of the other, then we obtain a qualitative Markov tree in which moving from a node to a neighbor of that node is always a matter of either refinement or coarsening.

ATTACHMENT Finally, notice that the Markov property is preserved when we attach a new node $m$ to what was a leaf node $i$ and associate with $m$ a coarsening of the partition associated with $i$. If $\mathfrak{P}_m$ is coarser than $\mathfrak{P}_i$, then its addition to a refinement already involving $\mathfrak{P}_i$ will not change that refinement.

## Diagnostic Markov Trees

In this section we study qualitative Markov trees that arise from hierarchical structures for diagnostic problems.

Hierarchical structures are rooted trees. Recall that a *rooted tree* is a tree that is drawn downward from a topmost node, as in Figure 1. The topmost node is called the *root,* and node $j$ is called a *daughter* of node $i$ if $j$ is directly below $i$. A *hierarchical structure* for a frame of discernment $\Theta$ is a rooted tree in which the root is $\Theta$, the other nodes are subsets of $\Theta$, and the daughters of any nonleaf node form a partition of that node.

To construct a hierarchical structure for a diagnostic problem, we first construct a *tree of diagnoses*, a rooted tree whose nodes are diagnoses. We begin with a list of possible diagnoses. In the case of a car that will not start, for example, we might begin with the list

{faulty battery system, faulty fuel system,
faulty starting system, something else}

Let us assume that the diagnoses on the list are mutually exclusive and collectively

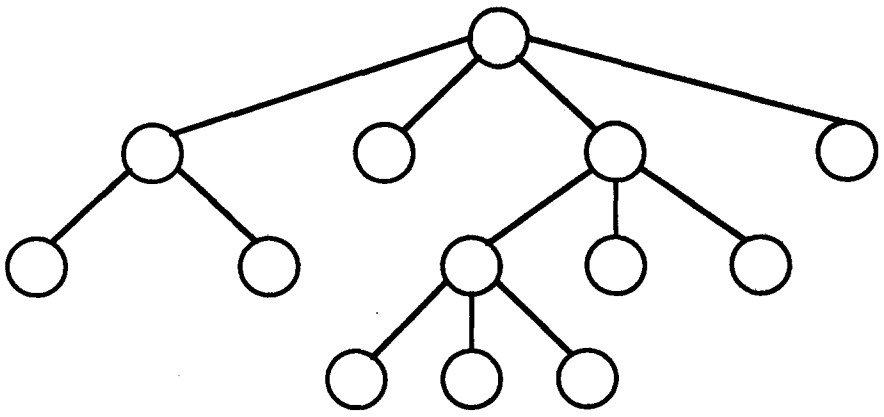

**Figure 1.** A Rooted Tree

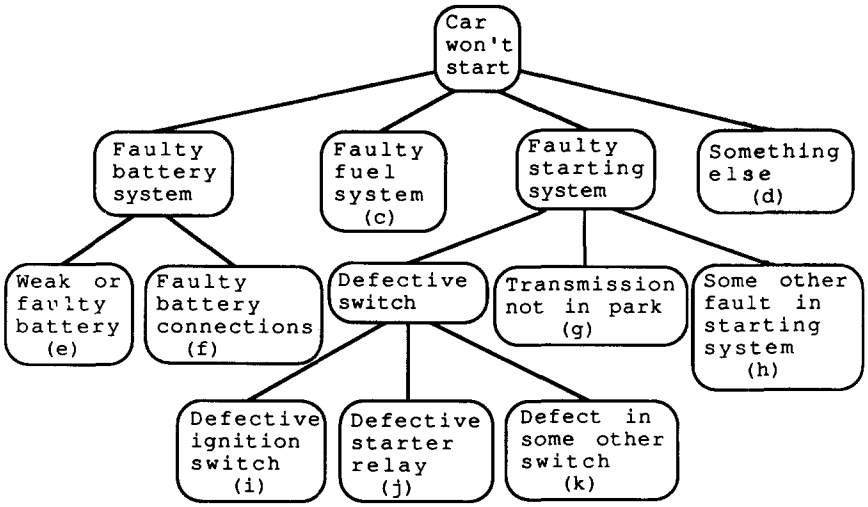


**Figure 2.** A Tree of Diagnoses

exhaustive; exactly one is correct. We use the list to construct an initial tree of diagnoses, consisting of just a root and its daughters. The root represents the problem, and the daughters are the diagnoses in the list. We then single out one of the diagnoses, say **d**, and we make a list of possibilities (again mutually exclusive and collectively exhaustive) for what more specifically might be true if **d** were correct. In the case of the car that will not start, for example, we might single out the diagnosis "faulty battery" and make the list

{weak or faulty battery, faulty battery connections}

We then enlarge our tree by attaching to the node corresponding to **d** a set of daughter nodes, one for each of the possibilities on this new list. We do this repeatedly, each time splitting some diagnosis **d** into more specific diagnoses, exactly one of which is correct if **d** is correct, and none of which are correct if **d** is not correct. Figure 2 shows a tree of diagnoses this might lead to in the case of the car.

Such a tree of diagnoses will have the property that its leaves themselves constitute a set of mutually exclusive and collectively exhaustive diagnoses. We can use this set as a frame of discernment. In Figure 2, for example, we can set $\Theta = \{c,d,e,f,g,h,i,j,k\}$. Moreover, we can easily transform the tree of diagnoses into a hierarchical structure for this frame; we simply replace each node with the set of the leaves that lie below it. In Figure 2, we replace "defective switch" with the set $N = \{i,j,k\}$, we replace "faulty starting system" with the set $M = \{g,h,i,j,k\}$, we replace "weak or faulty battery" with the singleton set $E = \{e\}$, and so on. This gives the hierarchical structure shown in Figure 3.

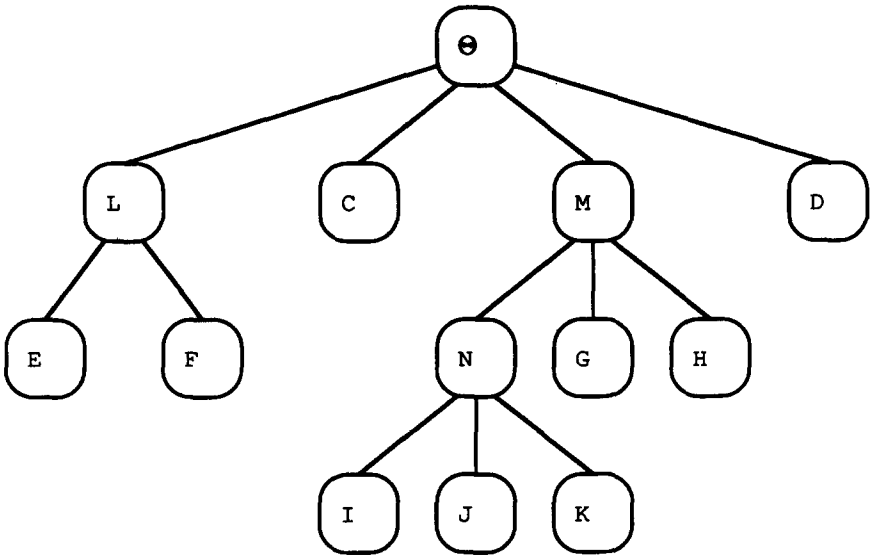


**Figure 3.** A Hierarchical Structure

We will be interested in a given hierarchical structure for a diagnostic problem if it matches the structure of the evidence we have for that problem. This means that each item of evidence we have either (1) directly supports or refutes some node in the tree or (2) bears on the question of which daughter of some node might be correct. An example of type 1 for Figure 3 would be evidence that the battery system is faulty. An example of type 2 would be experience about which switches in the starting system fail most often. Items of evidence of these two types bear most directly on relatively small partitions of the frame $\Theta$. Evidence of type 1 bears on partitions of the form $\{A,A^c\}$, where $A$ is a node of the tree. Evidence of type 2 bears on partitions of the form $\mathfrak{D}_A \cup \{A^c\}$, where $A$ is a nonleaf node of the tree and $\mathfrak{D}_A$ is the set of $A$'s daughters. Can these partitions be arranged in a qualitative Markov tree? Yes. In fact, given any hierarchical structure, we can construct two interesting qualitative Markov trees from these partitions. One of these we call the *tree of families*; the other we call the *tree of families and dichotomies*.

To construct the tree of families, we first remove the leaves from the hierarchical structure. We then associate a partition with each remaining node. With the root node $\Theta$, we associate the partition $\mathfrak{D}_\Theta$, and with each nonroot node $A$, we associate the partition $\mathfrak{D}_A \cup \{A^c\}$. Figure 4 shows the tree of families constructed in this way from the hierarchical structure in Figure 3.

To construct the tree of families and dichotomies, we enlarge the tree of families as follows. First we put a new node in the middle of each edge, and we associate with it the partition $\{A,A^c\}$, where $A$ is the node on the lower end

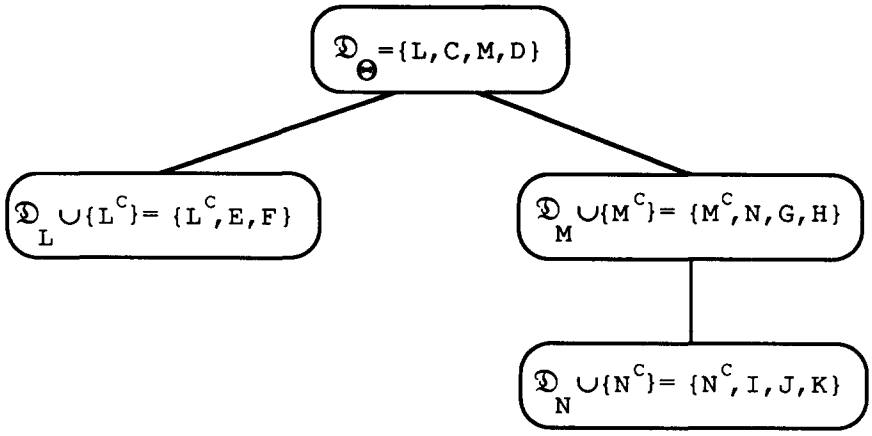


**Figure 4.** The Tree of Families for Figure 3

of the edge (the one now associated with the partition $\mathfrak{D}_A \cup \{A^c\}$). Then we add back to the tree each leaf $A$ that we removed from the hierarchical structure when we were constructing the tree of families, and we associate with it the partition $\{A,A^c\}$. Figure 5 shows the tree of families and dichotomies constructed in this way from Figure 4. Notice that this tree has a dichotomy $\{A,A^c\}$ corresponding to every nonroot node $A$ in the original hierarchical structure.

THEOREM 2 *The tree of families and the tree of families and dichotomies constructed from a hierarchical structure are qualitative Markov trees.*

To prove Theorem 2, it suffices to prove that the tree of families and dichotomies is qualitative Markov, since the tree of families is a contraction of it. (Each dichotomy $\{A,A^c\}$ has a refinement as a neighbor.) In order to prove that the tree of families and dichotomies is qualitative Markov, we need another lemma about conditional independence.

LEMMA 9 *If $\mathfrak{P}$ is a partition of $\Theta$, then $[\mathfrak{P}_A^\downarrow]_{A\in\mathfrak{P}} \dashv \mathfrak{P}$*, where

$$\mathfrak{P}_A^\downarrow = \{A^c\} \cup \{\{\theta\} | \theta \in A\}$$

Proof Suppose $P$ is an element of $\mathfrak{P}$, and for each element $A$ of $\mathfrak{P}$ choose an element $P_A$ of $\mathfrak{P}_A^\downarrow$ such that $P \cap P_A \neq \varnothing$. Then $P_P \subseteq P$, and $P_A = A^c$ for $A \neq P$. Hence $P \cap (\cap\{P_A | A \in \mathfrak{P}\}) = P_P \neq \varnothing$. ■

Proof of Theorem 2 We need to show that

$$[\wedge\{\mathfrak{P}_i | i \in N_{k,n}\}]_{k\in V_n} \dashv \mathfrak{P}_n \tag{7}$$

for every node $n$ in the tree of families and dichotomies. There are three cases: (1) $n$ is the root node, (2) $n$ is a family node, with associated partition $\mathfrak{D}_A \cup \{A^c\}$, or (3) $n$ is an interpolated or leaf node, with associated partition $\{A,A^c\}$. In case

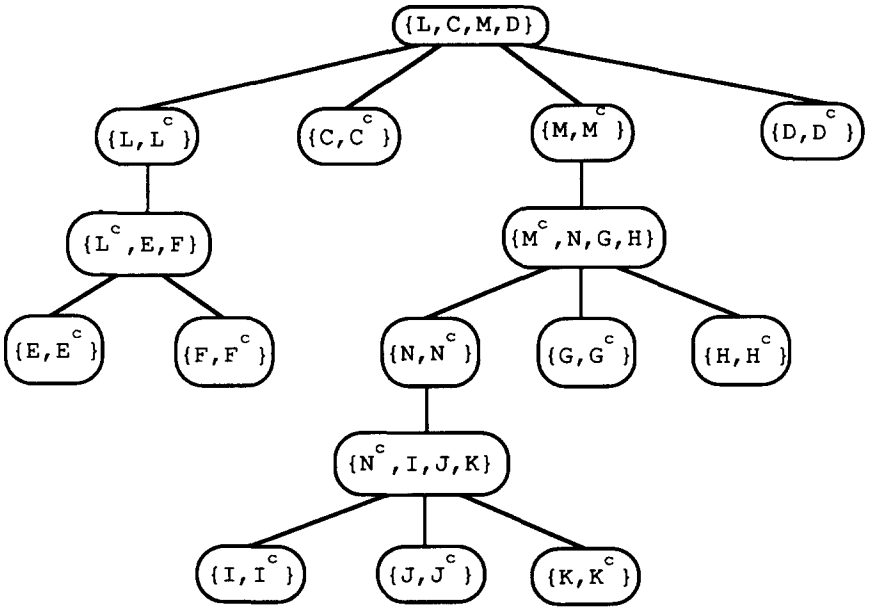


**Figure 5.** The Tree of Families and Dichotomies

1, (7) reduces to

$$[\mathfrak{P}_A^{\downarrow}]_{A\in\mathfrak{D}_\Theta} \dashv \mathfrak{D}_\Theta$$

In case 2, 7 reduces to

$$[\mathfrak{P}_B^{\downarrow}]_{B\in\mathfrak{D}_A\cup\{A^c\}} \dashv \mathfrak{D}_A\cup\{A^c\}$$

In case 3, (7) reduces to

$$[\mathfrak{P}_A^{\downarrow}, \mathfrak{P}_{A^c}^{\downarrow}] \dashv \{A,A^c\}.$$

These are all special cases of Lemma 9. ■

## Multivariate Markov Trees

We saw earlier that conditional independence relations do arise naturally in the multivariate formalism. In this section we will discuss some ways to construct qualitative Markov networks and trees in this formalism.

Suppose we begin with variables $X_1, \ldots, X_r$ with frames $\Theta_1, \ldots, \Theta_r$, respectively. If we initially specify no relations among these variables, then our overall frame $\Theta$ will be the Cartesian product $\Theta_1 \times \cdots \times \Theta_r$; otherwise it will be a subset of $\Theta_1 \times \cdots \times \Theta_r$.

Let $R$ denote the set of coordinates $R = \{1, \ldots, r\}$. Given any subset $V$ of $R$, we can construct a partition $\mathfrak{P}_V$ of $\Theta$ by grouping together those elements of

$\Theta$ that agree on all the coordinates in $V$; two elements $\theta$ and $\theta'$ of $\Theta$ are in the same element $P$ of $\mathfrak{P}_V$ if and only if $\theta_j = \theta'_j$ for all $j$ in $V$. We write $\mathfrak{P}_j$ for $\mathfrak{P}_{\{j\}}$. Elements of $\mathfrak{P}_V$ are often called *cylinder sets*. Notice that as we enlarge the subset $V$ of $R$, we refine the partition $\mathfrak{P}_V$. If $V_1 \subseteq V_2$, then $\mathfrak{P}_{V_1} \geq \mathfrak{P}_{V_2}$. And $\mathfrak{P}_{V_1} \wedge \mathfrak{P}_{V_2} = \mathfrak{P}_{V_1 \cup V_2}$.

THE UNRESTRICTED CASE Let us assume for the moment that we do not specify any relations among our variables, so that $\Theta = \Theta_1 \times \cdots \times \Theta_r$. In this case, generalizing Example 2, we see that $[\mathfrak{P}_{V_1}, \mathfrak{P}_{V_2}] \dashv \mathfrak{P}_V$ whenever $V_1 \cap V_2 \subseteq V$.

Generally we will be concerned only with partitions corresponding to certain subsets $V$ of $R$. If we let $W$ denote the set of these subsets, then we are dealing with a pair $(R, W)$, where $W$ is a set of subsets of $R$. In graph theory, such a pair is called a *hypergraph*. So instead of a graph $(N, E)$ with partitions associated with the nodes in $N$, we have a hypergraph $(R, W)$ with partitions associated with the *hyperedges* in $W$.

We can, however, use the elements of $W$ as nodes in a network. From the fact that $[\mathfrak{P}_{V_1}, \mathfrak{P}_{V_2}] \dashv \mathfrak{P}_V$ whenever $V_1 \cap V_2 \subseteq V$, we see that such a network will be qualitative Markov for the partitions $\{\mathfrak{P}_V\}_{V \in W}$ if $(\cup A) \cap (\cup B) \subseteq \cup C$ whenever $A$, $B$, and $C$ are sets of nodes in the network and $C$ separates $A$ from $B$.

One way to construct a network that has this property (and is therefore qualitative Markov) is to connect any two elements of $W$ that have a nonempty intersection; this gives the network $(N_1, E_1)$, where

$$N_1 = W \qquad \text{and} \qquad E_1 = \{\{w_1, w_2\} | w_1 \in W,\ w_2 \in W, \text{ and } w_1 \cap w_2 \neq \varnothing\}$$

Let us call this the *network of families* for the hypergraph.

Another way to construct a qualitative Markov network from the hypergraph $(R, W)$ is to take both the single coordinates in $R$ and the sets of coordinates in $W$ as nodes and join each coordinate to each set containing it. This gives the network $(N_2, E_2)$, where

$$N_2 = R \cup W \qquad \text{and} \qquad E_2 = \{\{j, w\} | w \in W \text{ and } j \in w\}$$

Let us call this the *network of families and variables* for the hypergraph.

Figure 6 shows a small hypergraph together with its network of families and its network of families and variables. The hypergraph is $(R, W)$, where $R = \{1,2,3,4,5\}$ and $W = \{\{1,2,3\},\{2,4\},\{3,5\},\{4,5\}\}$.

THE RESTRICTED CASE Now let us assume that we do specify relations among our variables, so that $\Theta$ is a subset of $\Theta_1 \times \cdots \times \Theta_r$. In Example 3, we saw that conditional independence relations can still hold in this case, provided relations among the variables follow a certain pattern. What pattern is needed in order for the networks we have just constructed to be qualitative Markov? We shall show that they will be qualitative Markov if $W$ is a *Kong pattern*.

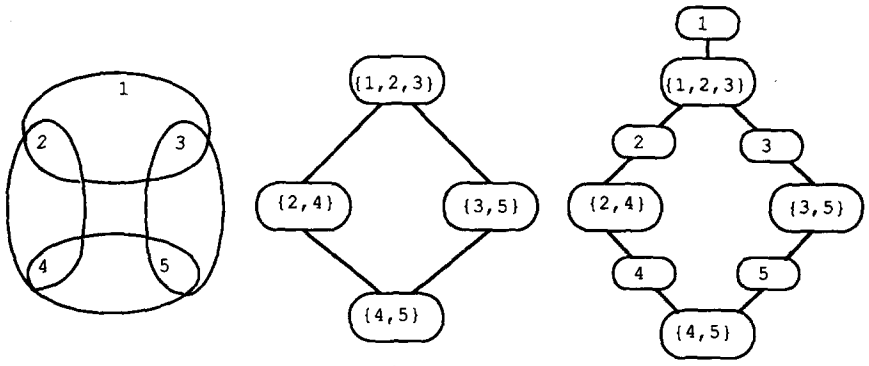


**Figure 6.** (a) A Hypergraph. (b) The Families. (c) The Families and Variables

Let us define this term. Specifying relations among variables means imposing restrictions on the values they can jointly take. We specify a relation among $X_{j_1}, \ldots, X_{j_k}$, for example, by specifying a subset $\Theta_{\{j_1, \ldots, j_k\}}$ of the Cartesian product $\Theta_{j_1} \times \cdots \times \Theta_{j_k}$ and imposing the restriction that $X_{j_1}, \ldots, X_{j_k}$ can jointly take the values $\theta_{j_1}, \ldots, \theta_{j_k}$, respectively, only if $(\theta_{j_1}, \ldots, \theta_{j_k})$ is in $\Theta_{\{j_1, \ldots, j_k\}}$. Let $W_0$ denote the set consisting of all the subsets of $R$ for which we impose restrictions; for each element $\{j_1, \ldots, j_k\}$ of $W_0$, we specify a subset $\Theta_{\{j_1 \ldots j_k\}}$ of the Cartesian product $\Theta_{j_1} \times \cdots \times \Theta_{j_k}$. We call $W_0$, and any set containing $W_0$, a *Kong pattern* for $\Theta$. To avoid trivialities, let us assume that every element of $W_0$ has at least two elements; $k \geq 2$ whenever $\{j_1, \ldots, j_k\} \in W$. Our overall frame of discernment is

$$\Theta = \{(\theta_1, \ldots, \theta_r) | (\theta_{j_1}, \ldots, \theta_{j_k}) \in \Theta_{\{j_1, \ldots, j_k\}} \text{ for each element } \{j_1, \ldots, j_k\} \in W_0\}$$

this is a subset of $\Theta_1 \times \cdots \times \Theta_r$.

THEOREM 3 *If* W *is a Kong pattern, then the network of families and the network of families and variables are qualitative Markov.*

To prove Theorem 3, it suffices to prove that the network of families and variables is qualitative Markov, since the network of families is a contraction of it. (The earlier reasoning about contractions of qualitative Markov trees applies also to qualitative Markov networks.) This is obvious from the following lemma.

LEMMA 10 *Suppose* W *is a Kong pattern. Suppose removal of a set of nodes from* $(N_2, E_2)$ *results in* k *connected subnetworks, and let* V *and* $V_1, \ldots, V_k$ *denote the sets of coordinates in the set of nodes removed and the remaining subnetworks, respectively. Then* $[\mathfrak{P}_{V_1}, \ldots, \mathfrak{P}_{V_s}] \dashv \mathfrak{P}_V$.

Proof Choose an element $P$ of $\mathfrak{P}_V$ and an element $P_k$ of $\mathfrak{P}_{V_k}$ such that $P \cap P_k \neq \varnothing$ for $k = 1, \ldots, s$. We want to show that $P \cap P_1 \cap \cdots \cap P_s \neq \varnothing$.

The choice of $P$ amounts to the choice of an element $\theta_j$ of $\Theta_j$ for each coordinate $j$ in $V$, and the choice of $P_k$ amounts to the choice of an element $\theta_j$ for each coordinate $j$ in $V_k$. Since $V, V_1, \ldots, V_s$ form a partition of $R$, we have chosen $\theta_j$ for every coordinate $j$; we have chosen an element $(\theta_1, \ldots, \theta_r)$ of $\Theta_1 \times \cdots \times \Theta_r$. We will have $P \cap P_1 \cap \cdots \cap P_s \neq \varnothing$ if and only if $(\theta_1, \ldots, \theta_r)$ is in $\Theta$.

Now $(\theta_1, \ldots, \theta_r)$ will be in $\Theta$ unless there is an element $\{j_1, \ldots, j_k\}$ of $W_0$ such that $(\theta_{j_1}, \ldots, \theta_{j_k})$ is not in $\Theta_{\{j_1, \ldots, j_k\}}$. But every element of $W_0$ is contained either in $V$ or else in exactly one union $V \cup V_k$. If $\{j_1, \ldots, j_k\}$ is contained in $V$, then $(\theta_{j_1}, \ldots, \theta_{j_k})$ is in $\Theta_{\{j_1, \ldots, j_k\}}$ by virtue of the fact that $P$ is nonempty. If $\{j_1, \ldots, j_k\}$ is contained in $V \cup V_k$, then $(\theta_{j_1}, \ldots, \theta_{j_k})$ is in $\Theta_{\{j_1, \ldots, j_k\}}$ by virtue of the fact that $P \cap P_k \neq \varnothing$. ■

If $(N_1, E_1)$ or $(N_2, E_2)$ happens to be a tree, then it is a qualitative Markov tree. In general, we cannot expect these networks to be trees. They will be trees, however, if the variables $X_1, \ldots, X_r$ with which we begin are already arranged in a rooted tree $(R, F)$ and each family in $W$ consists of a mother and her daughters:

$$W = \{\mathfrak{D}_j \cup \{j\} \mid j \in R,\ j \text{ is not a leaf in } (R, F)\}$$

In this case,the construction of $(N_1, E_1)$ and $(N_2, E_2)$ from $(R, F)$ is analogous to our construction of the tree of families and the tree of families and dichotomies in the section on diagnostic Markov trees.

In fact,the diagnostic problem can be put into the multivariate formalism so that Theorem 2 is a special case of Theorem 3. We take $R$ to be the set of nodes in the tree of diagnoses (e.g., Figure 2), and we take $X_j$ to be the variable that takes the value "yes" if the diagnosis $j$ is correct and "no" if it is not. The frame $\Theta_j$ for $X_j$ is the set {yes, no}, and the frame $\Theta_{\mathfrak{D}_j \cup \{j\}}$ for the element $\mathfrak{D}_j \cup \{j\}$ of $W$ is the subset of the Cartesian product $\times \{\Theta_{j'} \mid j' \in \mathfrak{D}_j \cup \{j\}\}$ consisting of (no, no, . . . , no) together with all elements that have a yes for the mother and a yes for exactly one daughter.

As Kong [9] has pointed out, the general multivariate framework does not require that we impose the restriction that the system we are diagnosing have only a single fault. We could impose weaker restrictions or no restrictions at all.

Diagnostic trees are not the only trees of variables that can be used to construct qualitative Markov trees in the way just described. Another class of examples is provided by causal trees, trees of variables where the mother-daughter relationship indicates direct causation from the mother to the daughter (Pearl [6]).

## BELIEF FUNCTIONS

A basic reference for the elementary aspects of the theory of belief functions (sometimes called the Dempster-Shafer theory in artificial intelligence) is Shafer

[1]. For more recent expositions and extensive bibliographies, see Shafer [21,22] and Shafer and Srivastava [23]. Expositions that discuss the theory's relevance to artificial intelligence include Garvey et al. [24] Gordon and Shortliffe [25], and Shafer [26]. See also Shafer [27–29] and Zhang [30].

Readers will need to turn to the references just cited for thorough expositions and for information about the theory's intuitive interpretation. Here we will quickly review the basics and then move on to some special topics. We continue to assume that the frame of discernment $\Theta$ is finite. For treatments of the infinite case, see Dempster [31–33], Shafer [34], and Strat [35].

Readers already familiar with the theory of belief functions will not need to read the next two subsections, which review the basics. They should, however, read Belief Functions and Partitions and Belief Functions and Qualitative Independence, which explain the notation that we will use in the final section and review the relevance of conditional independence to belief functions.

## Basic Definitions

A function Bel that assigns a degree of belief $\text{Bel}(A)$ to every subset $A$ of a frame of discernment $\Theta$ is called a *belief function* over $\Theta$ if there is a random nonempty subset $S$ of $\Theta$ such that $\text{Bel}(A) = \Pr[S \subseteq A]$ for every subset $A$ of $\Theta$. Intuitively, $\text{Bel}(A)$ is the total degree of belief committed to $A$ in light of given evidence.

A subset $A$ of $\Theta$ is called a *focal element* of Bel if $\Pr[S = A]$ is positive. The simplest belief function over $\Theta$ is the one whose only focal element is $\Theta$ itself; in this case $\Pr[S = \Theta] = 1$. This belief function is called the *vacuous belief function*. A belief function over $\Theta$ that has at most one focal element not equal to $\Theta$ itself is called a *simple support function*. If a simple support function does have a focal element not equal to $\Theta$ (i.e., if the simple support function is not vacuous), then this focal element is called the *focus* of the simple support function.

The information contained in a belief function can be expressed in several different ways. One way is in terms of the *basic probability assignment* $m$, defined by

$$m(A) = \Pr[S = A]$$

for every subset $A$ of $\Theta$. Since $S$ is nonempty, $m(\varnothing) = 0$, and since $\Theta$ is finite,

$$\sum\{m(A) | A \subseteq \Theta\} = 1$$

Intuitively, $m(A)$ measures the belief that is committed exactly to $A$ (and to nothing smaller). We can express Bel in terms of $m$ as follows:

$$\begin{aligned}\text{Bel}(A) &= \Pr[S \subseteq A] \\ &= \sum\{\Pr[S = B] | B \subseteq A\} = \sum\{m(B) | B \subseteq A\}\end{aligned}$$

It is shown in Shafer [1, Ch. 2] that we can also obtain $m$ from Bel:

$$m(A) = \sum\{(-1)^{|A-B|}\text{Bel}(B) | B \subseteq A\}$$

where $|A - B|$ denotes the number of elements in the set $A - B$.

Another way of expressing the information contained in a belief function Bel is in terms of the *plausibility function* Pl, which is given by

$$\text{Pl}(A) = 1 - \text{Bel}(A^c) = \Pr[S \cap A \neq \varnothing]$$

for every subset $A$ of $\Theta$. Intuitively $\text{Pl}(A)$ measures the extent to which given evidence fails to refute $A$. To recover Bel from Pl, we use the relation

$$\text{Bel}(A) = 1 - \text{Pl}(A^c) \tag{8}$$

Notice that $\text{Bel}(A) \leq \text{Pl}(A)$ for every subset $A$ of $\Theta$. Both Bel and Pl are monotone: $\text{Bel}(A) \leq \text{Bel}(B)$ and $\text{Pl}(A) \leq \text{Pl}(B)$ whenever $A \subseteq B$.

Finally, the information in Bel or $m$ or Pl is also contained in the *commonality function* $Q$, defined by

$$Q(A) = \Pr[S \supseteq A] = \sum\{m(B) | B \supseteq A\} \tag{9}$$

for every subset $A$ of $\Theta$. It is shown in Shafer [1, Ch. 2] that

$$Q(A) = \sum\{(-1)^{|B|+1}\text{Pl}(B) | \varnothing \neq B \subseteq A\} \tag{10}$$

and

$$\text{Pl}(A) = \sum\{(-1)^{|B|+1}Q(B) | \varnothing \neq B \subseteq A\} \tag{11}$$

for every nonempty subset $A$ of $\Theta$. We do not need formulas for the empty set, since $Q(\varnothing) = 1$ and $\text{Pl}(\varnothing) = 0$ for any belief function. Notice also that if the set $A$ contains only a single element, then (10) reduces to $Q(A) = \text{Pl}(A)$.

## Dempster's Rule of Combination

Dempster's rule of combination is a rule for forming a new belief function from two or more belief functions. Consider two random nonempty subsets $S_1$ and $S_2$. Suppose $S_1$ and $S_2$ are probabilistically independent, that is,

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

for all subsets $A_1$ and $A_2$ of $\Theta$. Suppose also that $\Pr[S_1 \cap S_2 \neq \varnothing] > 0$. Let $S$ be a random nonempty subset that has the probability distribution of $S_1 \cap S_2$ conditional on $S_1 \cap S_2 \neq \varnothing$, that is

$$\Pr[S = A] = \frac{\Pr[S_1 \cap S_2 = A]}{Pr[S_1 \cap S_2 \neq \varnothing]}$$

for every nonempty subset $A$ of $\Theta$. If $\text{Bel}_1$ and $\text{Bel}_2$ are the belief functions

corresponding to $S_1$ and $S_2$, then we call the belief function corresponding to $S$ the *orthogonal sum of* $\text{Bel}_1$ and $\text{Bel}_2$. The orthogonal sum of $\text{Bel}_1$ and $\text{Bel}_2$ is denoted by $\text{Bel}_1 \oplus \text{Bel}_2$. The rule form forming $\text{Bel}_1 \oplus \text{Bel}_2$ is called *Dempster's rule of combination*. If the bodies of evidence on which $\text{Bel}_1$ and $\text{Bel}_2$ are based are independent, then $\text{Bel}_1 \oplus \text{Bel}_2$ is supposed to represent the result of pooling these two bodies of evidence. Dempster's rule generalizes to the case where we wish to combine more than two belief functions; we merely use $S_1 \cap \cdots \cap S_n$ in place of $S_1 \cap S_2$.

It is obvious from the definitions that the operation $\oplus$ has the following properties:

- $\text{Bel}_1 \oplus \text{Bel}_2$ exists unless there is a subset $A$ of $\Theta$ such that $\text{Bel}_1(A) = 1$ and $\text{Bel}_2(A^c) = 1$.
- Commutativity: $\text{Bel}_1 \oplus \text{Bel}_2 = \text{Bel}_2 \oplus \text{Bel}_1$.
- Associativity: $(\text{Bel}_1 \oplus \text{Bel}_2) \oplus \text{Bel}_3 = \text{Bel}_1 \oplus (\text{Bel}_2 \oplus \text{Bel}_3)$.
- In general, $\text{Bel} \oplus \text{Bel} \neq \text{Bel}$. The belief function $\text{Bel} \oplus \text{Bel}$ will favor the same subsets as Bel, but it will do so with twice the weight of evidence, as it were.
- If $\text{Bel}_1$ is vacuous, then $\text{Bel}_1 \oplus \text{Bel}_2 = \text{Bel}_2$.

Dempster's rule can be expressed in terms of the probability mass assignment function as follows. Let the basic probability assignments for $\text{Bel}_1$, $\text{Bel}_2$, and $\text{Bel}_1 \oplus \text{Bel}_2$ be denoted by $m_1$, $m_2$, and $m$, respectively. Then for any nonempty subset $A$ of $\Theta$, we have

$$\begin{aligned} m(A) = \Pr[S = A] &= \frac{\Pr[S_1 \cap S_2 = A]}{\Pr[S_1 \cap S_2 \neq \varnothing]} \\ &= \frac{\sum\{m_1(B)m_2(C) \mid B \cap C = A\}}{\sum\{m_1(B)m_2(C) \mid B \cap C \neq \varnothing\}} \\ &= \frac{\sum\{m_1(B)m_2(C) \mid B \cap C = A\}}{1 - \sum\{m_1(B)m_2(C) \mid B \cap C = \varnothing\}} \end{aligned} \tag{12}$$

The formation of orthogonal sums by Dempster's rule corresponds to the multiplication of commonality functions. Indeed, if the commonality functions for $\text{Bel}_1$, $\text{Bel}_2$, and $\text{Bel}_1 \oplus \text{Bel}_2$ are denoted by $Q_1$, $Q_2$, and $Q$, respectively, then

$$\begin{aligned} Q(A) = \Pr[S \supseteq A] &= K \Pr[S_1 \cap S_1 \supseteq A] = K \Pr[S_1 \supseteq A \text{ and } S_2 \supseteq A] \\ &= K \Pr[S_1 \supseteq A] \Pr[S_2 \supseteq A] = KQ_1(A)Q_2(A) \end{aligned}$$

where $K$ does not depend on $A$. This result generalizes to the case where we combine more than two belief functions; if we are combining $\text{Bel}_1, \ldots, \text{Bel}_n$ with commonality functions $Q_1, \ldots, Q_n$, we obtain

$$Q(A) = KQ_1(A) \cdots Q_n(A) \tag{13}$$

where $K$ again does not depend on $A$.

Combining (13) and (11), we obtain an expression for Pl($A$), where Pl is the plausibility function corresponding to the orthogonal sum $\text{Bel}_1 \oplus \cdots \oplus \text{Bel}_n$:

$$\text{Pl}(A) = K \sum \{(-1)^{|B|+1} Q_1(B) \cdots Q_n(B) \mid \varnothing \neq B \subseteq A\} \tag{14}$$

Since $\text{Pl}(\Theta) = 1$, substituting $\Theta$ for $A$ in (14) results in an expression for $K$:

$$K^{-1} = \sum \{(-1)^{|B|+1} Q_1(B) \cdots Q_n(B) \mid \varnothing \neq B \subseteq \Theta\} \tag{15}$$

Formulas (14) and (15) can be used as the basis for a procedure for actually computing values of the orthogonal sum $\text{Bel}_1 \oplus \cdots \oplus \text{Bel}_n$. This procedure would begin by finding $Q_i(B)$ for each $i$ between 1 and $n$ and for each subset $B$ of $\Theta$. [If the $\text{Bel}_i$ are stored as lists of focal elements and associated $m$ values, then (9) might be used to do this.] Then it would use (14) and (15) to find Pl($A$) for the particular subsets $A$ that interest us. We could, if we wanted, then shift to actual values of $\text{Bel}_1 \oplus \cdots \oplus \text{Bel}_n$ using (8).

Unfortunately, this procedure is computationally expensive when $\Theta$ is large. The number of terms in (14) depends on the size of the subset $A$, but (15) involves a term for every nonempty subset $B$ of $\Theta$, and the number of these subsets increases exponentially with the size of $\Theta$. This means that we face a computation of exponential complexity even if we are trying to find the value of the orthogonal sum $\text{Bel}_1 \oplus \ldots \oplus \text{Bel}_n$ only for a single subset $A$ of $\Theta$.

This computational complexity seems to be intrinsic to Dempster's rule. It is possible in some cases to exploit special structure in the belief functions being combined in order to reduce the complexity, but there does not seem to be any general way of implementing the rule that will always involve fewer computations than are involved in (14) and (15) (Barnett [36]).

## Belief Functions and Partitions

As we explained in the introduction, a partition $\mathfrak{P}$ of a frame of discernment $\Theta$ can itself be regarded as a frame of discernment, a frame that is concerned with a narrower question than $\Theta$ is concerned with. A belief function Bel over $\Theta$ can also be narrowed so that it gives degrees of belief only for this narrower question. When we do this, we obtain a belief function that explicitly or implicitly uses $\mathfrak{P}$ instead of $\Theta$ as its frame. This shift from $\Theta$ to the simpler (and smaller) frame $\mathfrak{P}$ may have both conceptual and computational advantages. A simpler frame may be easier to think about, and a smaller frame may be necessary to make Dempster's rule computationally feasible. We must always ask, however, whether the simplification invalidates our reasoning. Under what conditions will we continue to get valid results from Dempster's rule when we substitute $\mathfrak{P}$ for $\Theta$?

In this section we will develop a notation that will help us discuss this question. This notation will make the shift from $\Theta$ to $\mathfrak{P}$ implicit rather than explicit. Given a belief function Bel over $\Theta$, we will represent the shift to $\mathfrak{P}$ by replacing Bel

with another belief function, denoted by $\text{Bel}_{\mathfrak{P}}$ and called the *coarsening* of Bel to $\mathfrak{P}$. Formally, the coarsening $\text{Bel}_{\mathfrak{P}}$ will be another belief function over $\Theta$, but it will have the simpler structure of a belief function over $\mathfrak{P}$, and it will be interpreted as a belief function over $\mathfrak{P}$ at the level of implementation. The advantage of this implicit approach is that it relieves us of the need for a notation that tracks what frames different belief functions are defined over. Formally, all our belief functions will be defined over a single frame.

How can a belief function over $\Theta$ have the simpler structure of a belief function over the partition $\mathfrak{P}$ of $\Theta$? To answer this, recall that a subset $A$ of $\mathfrak{P}$ correponds in meaning to the subset $\cup A$ of $\Theta$. (The symbol $\cup A$ denotes the union of the elements of $A$. Thus if $A = \{P_1, \ldots, P_n\}$, then $\cup A = P_1 \cup \cdots \cup P_n$. The two sets correspond in meaning because the answer to the question considered by $\mathfrak{P}$ is in $A$ if and only the answer to the question considered by $\Theta$ is in $\cup A$.) Because of this correspondence, we may say that a belief function over $\Theta$ has the simpler structure of a belief function over $\mathfrak{P}$ if all its focal elements are of the form $\cup A$ for some $A$ in $\mathfrak{P}$, that is, if all its focal elements are in the field $\mathfrak{P}^*$ of subsets of $\Theta$.

As the following lemma shows, there are a number of ways of expressing the condition that a belief function over $\Theta$ must have the simpler structure of a belief function over a partition $\mathfrak{P}$.

LEMMA 11 *Suppose Bel is a belief function over $\Theta$, with plausibility function Pl and random subset S, and suppose $\mathfrak{P}$* is a partition of $\Theta$. *Then the following statements are all equivalent.*

1. $Pr[S \in \mathfrak{P}^*] = 1$.
2. *Bel's focal elements are all in $\mathfrak{P}^*$.*
3. $Bel(A) = Bel(A_{\mathfrak{P}})$ *for every subset* A.
4. $Pl(A) = Pl(A^{\mathfrak{P}})$ *for every subset* A.
5. $Bel(A) = max\{Bel(B) | B \subseteq A$ *and* $B \in \mathfrak{P}^*\}$ *for every subset* A.

Proof The equivalence of statements 1 and 2 is obvious from the definitions. The equivalence of statements 3 and 4 follows from the fact that $(A^c)^{\mathfrak{P}} = (A_{\mathfrak{P}})^c$. And the equivalence of statements 3 and 5 follows from the montonicity of Bel and the fact that $A_{\mathfrak{P}}$ is the largest element of $\mathfrak{P}^*$ contained in $A$. So we need only show that statements 2 and 3 are equivalent.

Since $A_{\mathfrak{P}}$ is the largest element of $\mathfrak{P}^*$ contained in $A$, the elements of $\mathfrak{P}^*$ contained in $A$ are the same as those contained in $A_{\mathfrak{P}}$.

$$\sum\{m(B) | B \in \mathfrak{P}^* \text{ and } B \subseteq A\} = \sum\{m(B) | B \in \mathfrak{P}^* \text{ and } B \subseteq A_{\mathfrak{P}}\}$$

If all the focal elements are in $\mathfrak{P}^*$, then this means that $\text{Bel}(A) = \text{Bel}(A^{\mathfrak{P}})$. So statement 2 implies statement 3.

Suppose, on the other hand, that statement 3 is true; $\text{Bel}(A_{\mathfrak{P}})$ is just as large as $\text{Bel}(A)$ for all $A$. This means that all the focal elements contained in $A$ are

also contained in $A_{\mathfrak{P}}$. So if $A$ itself is a focal element, $A$ must be contained in and hence equal to $A_{\mathfrak{P}}$. Hence $A$ is in $\mathfrak{P}^*$. So statement 3 implies statement 2. ■

Let us say that a belief function Bel over $\Theta$ is *carried* by the partition $\mathfrak{P}$ if the conditions of Lemma 11 are satisfied. In general, a belief function Bel over $\Theta$ will be carried by many different partitions of $\Theta$, but there will be a coarsest partition that carries it, namely, the finest common coarsening of all the partitions that carry it. (*Proof* Recall that Bel is carried by a partition if and only if all Bel's focal elements are in the corresponding field. The intersection of all the fields of subsets that contain all Bel's focal elements certainly itself contains them and is itself a field. Hence it is the smallest field containing them, and hence it is the smallest field that carries Bel. But as we noted in the introduction, the partition corresponding to the intersection of a collection of fields is the finest common coarsening of the corresponding partitions.)

If we are working with belief functions carried by $\mathfrak{P}$, then we can think of them as belief functions that use $\mathfrak{P}$ as their frame. This is because operations on belief functions can be all defined in terms of their focal elements, and these are in $\mathfrak{P}^*$, which is isomorphic to the set of all subsets of $\mathfrak{P}$.

Notice, in particular, that the combination by Dempster's rule of two or more belief functions carried by $\mathfrak{P}$ is also carried by $\mathfrak{P}$.

We are now in a position to explain the idea of coarsening a belief function to a partition. If $S$ is a random subset of $\Theta$ and $\mathfrak{P}$ is a partion of $\Theta$, then $S^{\mathfrak{P}}$ is also a random subset of $\Theta$. If Bel is the belief function corresponding to $S$, then let $\text{Bel}_{\mathfrak{P}}$ denote the belief function corresponding to $S^{\mathfrak{P}}$. We call $\text{Bel}_{\mathfrak{P}}$ the *coarsening* of Bel to $\mathfrak{P}$. Since $S^{\mathfrak{P}}$ is in $\mathfrak{P}^*$, $\text{Bel}_{\mathfrak{P}}$ is carried by $\mathfrak{P}$.

LEMMA 12 *Suppose Bel and Bel' are belief functions over $\Theta$, with plausibility functions Pl and Pl', respectively. Then the following statements are all equivalent.*

1. *Bel' is the coarsening of Bel to $\mathfrak{P}$.*
2. *$Bel'(\text{A}) = Bel(\text{A}_{\mathfrak{P}})$ for every subset* A.
3. *Bel' agrees with Bel on $\mathfrak{P}^*$* and is carried by $\mathfrak{P}$.
4. *$Pl'(\text{A}) = Pl(\text{A}^{\mathfrak{P}})$* for every subset A.
5. *$Bel'(\text{A}) = max\{Bel(\text{B}) | \text{B} \subseteq \text{A}$ and $\text{B} \in \mathfrak{P}^*\}$ for every subset* A.

Proof We will show that the first three statements are equivalent by showing that statement 1 implies statement 2, 2 implies 3, and 3 implies 1.

First, statement 1 implies statement 2. Recall that $B^{\mathfrak{P}} \subseteq A$ if and only if $B \subseteq A_{\mathfrak{P}}$. If we substitute the random subset $S$ for $B$ in this relation, we obtain $\Pr[S^{\mathfrak{P}} \subseteq A] = \Pr[S \subseteq A_{\mathfrak{P}}]$. If statement 1 holds, then this is just another way of writing statement 2.

Second, statement 2 implies statement 3. Recall that $A = A_{\mathfrak{P}}$ for $A \in \mathfrak{P}^*$. If statement 2 holds, then this tells us that Bel' agrees with Bel on $\mathfrak{P}^*$. It then follows from statement 2 and Lemma 11 that Bel is carried by $\mathfrak{P}$.

Third, statement 3 implies statement 1. We have just established that the coarsening of Bel to $\mathfrak{P}$ is carried by $\mathfrak{P}$ and agrees with Bel on $\mathfrak{P}^*$. It is clear from Lemma 11 that only one belief function can have these properties. So any belief function Bel′ having them must be the coarsening of Bel to $\mathfrak{P}$.

To complete the proof, we note that statements 2 and 4 are equivalent because $(A^c)^{\mathfrak{P}} = (A_{\mathfrak{P}})^c$, while statements 2 and 5 are equivalent because Bel is monotonic and $A^{\mathfrak{P}}$ is the largest element of $\mathfrak{P}^*$ contained in $A$. ■

Condition 3 is crucial; it tell us that as far as $\mathfrak{P}$ is concerned, the coarsening $\text{Bel}_{\mathfrak{P}}$ says the same thing as Bel. Lemma 12 tells us, inter alia, that there is only one belief function that satisfies condition 3; saying that a belief function is carried by $\mathfrak{P}$ and agrees with Bel on $\mathfrak{P}^*$ identifies this belief function as the coarsening of Bel to $\mathfrak{P}$.

As we have emphasized, we usually think of $\text{Bel}_{\mathfrak{P}}$ as a belief function over the frame $\mathfrak{P}$. This means that the shift from Bel to $\text{Bel}_{\mathfrak{P}}$ involves a shift from $\Theta$ to $\mathfrak{P}$. Of course, Bel might already be carried by some partition $\mathfrak{P}_1$ of $\Theta$, and then we may really be making a shift from the frame $\mathfrak{P}_1$ to the frame $\mathfrak{P}$. In this case we may call $\text{Bel}_{\mathfrak{P}}$ the *projection* of Bel from $\mathfrak{P}_1$ to $\mathfrak{P}$.

## Belief Functions and Qualitative Independence

At the beginning of the previous section we asked under what conditions we will continue to get valid results from Dempster's rule if we use $\mathfrak{P}$ instead of $\Theta$ as our frame. We can now make the question more precise. Suppose we want to combine two belief functions $\text{Bel}_1$ and $\text{Bel}_2$, and suppose we are really interested in values of the orthogonal sum $\text{Bel}_1 \oplus \text{Bel}_2$ only for certain elements of the field $\mathfrak{P}^*$. Will we get the right answers for these values if we coarsen both belief functions to $\mathfrak{P}$ before combining them? The following theorem tells us that we will if there are partitions $\mathfrak{P}_1$ and $\mathfrak{P}_2$ such that $\text{Bel}_1$ is carried by $\mathfrak{P}_1$, $\text{Bel}_2$ is carried by $\mathfrak{P}_2$, and $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent given $\mathfrak{P}$.

THEOREM 4 *If $\text{Bel}_1$ is carried by $\mathfrak{P}_1$, $\text{Bel}_2$ is carried by $\mathfrak{P}_2$, and $[\mathfrak{P}_1, \mathfrak{P}_2] \dashv \mathfrak{P}$, then*

$$(\text{Bel}_1 \oplus \text{Bel}_2)_{\mathfrak{P}} = \text{Bel}_{1\mathfrak{P}} \oplus \text{Bel}_{2\mathfrak{P}}$$

Proof If $\text{Bel}_1$ is carried by $\mathfrak{P}_1$ and $\text{Bel}_2$ is carried by $\mathfrak{P}_2$, then the random nonempty subsets $S_1$ and $S_2$ corresponding to these belief functions are always in $\mathfrak{P}_1^*$ and $\mathfrak{P}_2^*$, respectively. So from Lemma 2 and the assumption that $[\mathfrak{P}_1, \mathfrak{P}_2] \dashv \mathfrak{P}$, we obtain $(S_1 \cap S_2)^{\mathfrak{P}} = S_1^{\mathfrak{P}} \cap S_2^{\mathfrak{P}}$. The theorem follows, because $(\text{Bel}_1 \oplus \text{Bel}_2)_{\mathfrak{P}}$ corresponds to a random nonempty subset that has the probability distribution of $(S_1 \cap S_2)^{\mathfrak{P}}$ conditional on $(S_1 \cap S_2)^{\mathfrak{P}} \neq \varnothing$, and $\text{Bel}_{1\mathfrak{P}} \oplus \text{Bel}_{2\mathfrak{P}}$ corresponds to a random nonempty subset that has the distribution of $S_1^{\mathfrak{P}} \cap S_2^{\mathfrak{P}}$ conditional on $S_1^{\mathfrak{P}} \cap S_2^{\mathfrak{P}} \neq \varnothing$.

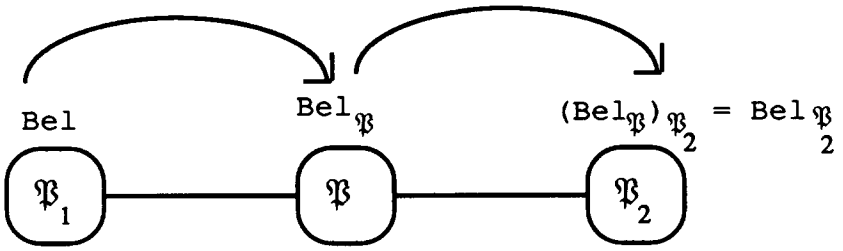


**Figure 7.** Projecting a Belief Function From $\mathfrak{P}_1$ to $\mathfrak{P}_2$

The next theorem tells us that if $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are qualitatively independent given $\mathfrak{P}$ and we want to project a belief function from $\mathfrak{P}_1$ to $\mathfrak{P}_2$, then it does no harm to project it first to $\mathfrak{P}$ (see Figure 7).

THEOREM 5 *If Bel is carried by* $\mathfrak{P}_1$ *and* $[\mathfrak{P}_1, \mathfrak{P}_2] \dashv \mathfrak{P}$, *then*

$$Bel_{\mathfrak{P}_2} = (Bel_{\mathfrak{P}})_{\mathfrak{P}_2}$$

Proof If Bel is carried by $\mathfrak{P}_1$, then the random subset $S$ corresponding to Bel is always in $\mathfrak{P}_1^*$. Hence, by Lemma 3, $S^{\mathfrak{P}_2} = (S^{\mathfrak{P}})^{\mathfrak{P}_2}$.

## An Explicit Description of Coarsening

For the benefit of those readers who may be uncomfortable with our implicit notation for coarsening, we will now describe the process with a notation that makes the shift from one frame to another explicit. Readers who are satisfied with the implicit approach need not read this section in order to understand the remainder of the article. Readers who want to see yet another way of handling the transition from one frame to another can consult Shafer [1, Ch. 6]. The correspondence between a subset $A$ of $\mathfrak{P}$ and the subset $\cup A$ of $\Theta$ is the key to seeing both how to move from a belief function over $\Theta$ to one over $\mathfrak{P}$ and how to move from one over $\mathfrak{P}$ to one over $\Theta$.

1. Suppose we begin with a belief function Bel over $\Theta$, with corresponding random nonempty subset $S$ and basic probability assignment $m$. In order to derive a belief function over $\mathfrak{P}$, we need only note that since the subset $A$ of $\mathfrak{P}$ corresponds to the subset $\cup A$ of $\Theta$, it should be assigned a degree of belief equal to Bel($\cup A$). Let Bel$_{\downarrow}$ denote the set function defined by

$$\text{Bel}_{\downarrow}(A) = \text{Bel}(\cup A) \tag{16}$$

for every subset $A$ of $\mathfrak{P}$. We call Bel$_{\downarrow}$ the *restriction* of Bel to $\mathfrak{P}$. It is indeed a belief function; it corresponds to the random subset $S_{\downarrow}$ of $\mathfrak{P}$ that is defined by

$$S_{\downarrow} = \{P \in \mathfrak{P} | P \cap S \neq \emptyset\}$$

(To prove this, check that $S_\downarrow \subseteq A$ is equivalent to $S \subseteq \cup A$.) The corresponding basic probability assignment, $m_\downarrow$, is given by

$$m_\downarrow(A) = \Pr[S_\downarrow = A] = \sum \{\Pr[S = B] | A = \{P \in \mathfrak{P} | P \cap B \neq \varnothing\}\}$$
$$= \sum \{m(B) | A = \{P \in \mathfrak{P} | P \cap B \neq \varnothing\}\} \quad (17)$$

In general, the restriction $\text{Bel}_\downarrow$ contains only part of the information contained by Bel; it gives degrees of belief only for propositions that correspond to subsets of $\mathfrak{P}$.

2. Suppose we begin with a belief function Bel over $\mathfrak{P}$, with corresonding random subset $S$ and basic probability assignment $m$. From the random subset $S$ of $\mathfrak{P}$, we may construct the random subset $\cup S$ of $\Theta$. Let $\text{Bel}^\uparrow$ denote the belief function over $\Theta$ corresponding to $\cup S$. We call $\text{Bel}^\uparrow$ the *vacuous extension* of Bel to $\Theta$. The focal elements of $\text{Bel}^\uparrow$ are all in $\mathfrak{P}^*$ Indeed, if $m^\uparrow$ denotes the basic probability assignment for $\text{Bel}^\uparrow$, then

$$m^\uparrow(A) = \Pr[\cup S = A] = \Pr[S \text{ equals some } B \in \mathfrak{P} \text{ such that } \cup B = A]$$

or

$$m^\uparrow(A) = \begin{cases} m(B) & \text{if } A = \cup B, \text{ where } B \in \mathfrak{P} \\ 0 & \text{if } A \text{ is not in } \mathfrak{P}^* \end{cases} \quad (18)$$

The values of $\text{Bel}^\uparrow$ itself are given by

$$\text{Bel}^\uparrow(A) = \sum \{m(B) | B \subseteq \mathfrak{P},\ \cup B \subseteq A\}$$
$$= \sum \{m(B) | B \subseteq \{P \in \mathfrak{P} \subseteq A\}\}$$
$$= \text{Bel}(\{P \in \mathfrak{P} | P \subseteq A\}) \quad (19)$$

Formula (19) tells us in particular that for every $B \subseteq \mathfrak{P}$,

$$\text{Bel}^\uparrow(\cup B) = \text{Bel}(B) \quad (20)$$

Thus $\text{Bel}^\uparrow$ contains all the information contained by Bel. $\text{Bel}^\uparrow$ also gives degrees of belief for subsets of $\Theta$ that are not of the form $\cup B$, that is, subsets of $\Theta$ that are not in $\mathfrak{P}^*$. But $\text{Bel}^\uparrow$ derives its degrees of belief for these subsets from its degrees of belief for elements of $\mathfrak{P}^*$; for any $A \subseteq \Theta$,

$$\text{Bel}^\uparrow(A) = \text{Bel}(\{P \in \mathfrak{P} | P \subseteq A\}) = \text{Bel}^\uparrow(\cup \{P \in \mathfrak{P} | P \subseteq A\}) = \text{Bel}^\uparrow(A_\mathfrak{P}) \quad (21)$$

So $\text{Bel}^\uparrow$ does not really contain any more information than Bel does.

Suppose Bel is a belief function over $\Theta$. Then $(\text{Bel}_\downarrow)^\uparrow$ is the belief function $\text{Bel}_\mathfrak{P}$ defined in the preceding section, the coarsening of Bel to $\mathfrak{P}$. To see this, we need only note that $(\text{Bel}_\downarrow)^\uparrow$ has all its focal elements in $\mathfrak{P}^*$ (because it is a vacuous extension from $\mathfrak{P}$) and agrees with Bel on $\mathfrak{P}^*$ [if $A$ is in $\mathfrak{P}^*$, then $(\text{Bel}_\downarrow)^\uparrow(\cup A) = \text{Bel}_\downarrow(A) = \text{Bel}(\cup A)$ by (20) and (16)]. To put the matter more verbally, in order to coarsen a belief function from $\Theta$ to $\mathfrak{P}$, we first restrict it

to $\mathfrak{P}$ and then vacuously extended it back ot $\Theta$. The second step, the vacuous extension back to $\Theta$, is only for notational and terminological convenience and can be ignored in an implementation.

For the sake of completeness, we might ask what happens when we combine vacuous extension and restriction in the opposite order, first vacuously extending and then restricting. The answer is that we get back what we started with; if Bel is a belief function over $\mathfrak{P}$, then $(\text{Bel}^{\uparrow})_{\downarrow} = \text{Bel}$. [*Proof* Suppose $A \subseteq \mathfrak{P}$. By (16), $(\text{Bel}^{\uparrow})_{\downarrow}(A) = \text{Bel}^{\uparrow}(\cup A)$. But by (20), $\text{Bel}^{\uparrow}(\cup A) = \text{Bel}(A)$.]

What are the computational costs of restriction and vacuous extension?

From a purely mathematical point of view, restriction is a simplification, but this mathematical simplification may involve a computational cost. If our belief function Bel over $\Theta$ were stored as an explicit specification of the numbers $\text{Bel}(B)$ for all subsets $B$ of $\Theta$, and if there were no computational cost in matching $A$ to $\cup A$, then formula (16) would make the shift from Bel to $\text{Bel}_{\downarrow}$ computationally trivial. But if Bel has been obtained by an application of Dempster's rule, then it is more likely to be stored as a commonality function or as a specification of focal elements and their $m$-values, and in this case the shift will involve a cost such as that suggested by (17).

Conversely, vacuous extension may involve relatively little computational cost. Since $\Theta$ is larger than $\mathfrak{P}$, we might think of any belief function over $\Theta$ as inherently more complicated than one over $\mathfrak{P}$. But (18) makes it clear that if we store a belief function over $\mathfrak{P}$ in terms of its focal elements and their $m$-values, then vacuous extension to $\Theta$ involves little more than changing the names of the focal elements.

## PROPAGATION IN TREES

In this section, we show how to take advantage of the structure of a qualitative Markov tree when using Dempster's rule for combining belief functions carried by partitions in the tree.

Here is the setting. We are concerned with $r$ related questions, $Q_1, \ldots, Q_r$. We represent these questions by partitions of an overall frame $\Theta$. As usual, we let $\mathfrak{P}_i$ denote the partition representing $Q_i$. We arrange these partitions in a qualitative Markov tree, say $T = (N, E)$, where $N = \{1, \ldots, r\}$. We have $r$ independent items of evidence, one bearing directly on each of the questions. We represent the evidence bearing directly on $Q_i$ by a belief function $\text{Bel}_i$; at the level of implementation, $\text{Bel}_i$ will be a belief function over $\mathfrak{P}_i$, but we may think of it as a belief function over $\Theta$ carried by $\mathfrak{P}_i$.

We want to use Dempster's rule to combine these belief functions. How can we do so efficiently?

Formally, we are interested in $\oplus\{\text{Bel}_i | i \in N\}$, the belief function over $\Theta$ that represents all our evidence. If $\Theta$ is large, it may be not feasible to compute all

the values of this belief function. In fact, it may not be feasible to compute any of them. We will show, however, that if the individual partitions are not too large, then it is feasible to compute values of $\oplus\{\text{Bel}_i | i \in N\}$ for subsets of $\Theta$ that relate to the original questions, that is, it is possible to compute the coarsening of $\oplus\{\text{Bel}_i | i \in N\}$ to each partition $\mathfrak{P}_i$.

How much does the scheme we present here reduce the computational complexity of Dempster's rule? The scheme reduces the problem of implementing the rule on the overall frame $\Theta$ to the problem of implementing it on the partitions $\mathfrak{P}_i$. Hence it reduces the problem from one exponential in the size of $\Theta$ (see the section on Dempster's rule of combinations) to one exponential in the size of the $\mathfrak{P}_i$. Typically, the feasibility of the scheme will depend on the size of the largest of the partitions.

We begin this discussion with a scheme for computing the coarsening to a single partition. Then, we present a general scheme, inspired by the work of Judea Pearl, for the parallel computation of coarsenings to all the partitions. Following that we argue that though this general scheme appears to be a control strategy, it should be understood primarily as a design. Finally, we discuss the generality and flexibility of the propagation scheme, and we sketch how schemes described by Shafer and Logan [2], Shafer [3], and Pearl [6] can be seen as special cases.

## Computation for a Single Partition

For convenience, let $\text{Bel}^T$ denote the result of combining all our belief functions by Dempster's rule:

$$\text{Bel}^T = \oplus\{\text{Bel}_i | i \in N\}$$

The symbol $T$ refers to our tree; $T = (N, E)$. Our task is to compute the coarsening $\text{Bel}^T_{\mathfrak{P}_n}$ for a particular node $n$.

If the tree were small enough, we would have no problem. If, for example, it consisted of a single node, there would be nothing to do. This trivial point gives us a hint. We can compute $\text{Bel}^T_{\mathfrak{P}_n}$ recursively if we can reduce the computation to similar computations for strictly smaller trees.

We need some more notation. For any subtree $U = \{N_U, E_U\}$ of $T$, let $\text{Bel}^U$ denote the orthogonal sum $\oplus\{\text{Bel}_i | i \in N_U\}$. Recall that $V_n$ denotes the set of $n$'s neighbors and that $T_{k,n} = (N_{k,n}, E_{k,n})$ denotes the subtree containing $k$ that remains when $n$ is removed from $T$.

The following theorem spells out how to reduce the computation of $\text{Bel}^T_{\mathfrak{P}_n}$ to similar computations for strictly smaller trees.

THEOREM 6 *Let* $\text{T} = (\text{N}, \text{E})$ *be a qualitative Markov tree for* $\{\mathfrak{P}_i | \text{i} \in \text{N}\}$ *and let* $\text{Bel}_i$ *be carried by* $\mathfrak{P}_i$ *for each* i *in* N. *Then*

$$\text{Bel}^T_{\mathfrak{P}_n} = \text{Bel}_n \oplus ( \oplus \{(\text{Bel}^{T_{k,n}}_{\mathfrak{P}_k})_{\mathfrak{P}_n} | k \in V_n\}) \tag{22}$$

Proof Since $\mathrm{Bel}^T = \mathrm{Bel}_n \oplus (\oplus\{\mathrm{Bel}^{T_{k,n}} | k \in V_n\})$, it follows from Theorem 4 that

$$\mathrm{Bel}^T_{\mathfrak{P}_n} = \mathrm{Bel}_n \oplus (\{\mathrm{Bel}^{T_{k,n}}_{\mathfrak{P}_n} | k \in V_n\})$$

Since

$$[\mathfrak{P}_n, \wedge\{\mathfrak{P}_j | j \in N_{k,n}\}] \dashv \mathfrak{P}_k$$

for every $k \in V_n$, it follows from Theorem 5 that

$$\mathrm{Bel}^{T_{k,n}}_{\mathfrak{P}_n} = (\mathrm{Bel}^{T_{k,n}}_{\mathfrak{P}_k})_{\mathfrak{P}_n}$$ ■

The trees $T_{k,n} = (N_{k,n}, E_{k,n})$ are subtrees of $T$, and hence they are strictly smaller than $T$. If we can solve our problem for these smaller trees—if for each neighbor $k$ of $n$ we can compute the coarsening $\mathrm{Bel}^{T_{k,n}}_{\mathfrak{P}_k}$ on the frame $\mathfrak{P}_k$—then Eq. (22) tells us that we need only perform two more tasks. We need to project each of these belief functions from its frame $\mathfrak{P}_k$ to the frame $\mathfrak{P}_n$, and then we need to combine them, together with $\mathrm{Bel}_n$, on the frame $\mathfrak{P}_n$. [Both $\oplus$'s in (22) can be interpreted as directions to combine belief functions on the frame $\mathfrak{P}_n$, because all the belief functions being combined, $\mathrm{Bel}_n$ and $(\mathrm{Bel}^{T_{k,n}}_{\mathfrak{P}_k})_{\mathfrak{P}_n}$ for $k \in V_n$, are carried by $\mathfrak{P}_n$]. These two tasks should be feasible. Since $k$ and $n$ are neighbors, we presumably understand the relations between the questions $Q_k$ and $Q_n$, and hence we should be able to manage the projection. If the frame $\mathfrak{P}_n$ is not too large, we should be able to handle the combination.

So our problem is solved. We begin at the leaves of our tree and move step by step toward $n$. We synchronize our paths inward from the different leaves by delaying the step to a given node $j$ until we have passed through all the neighbors of $j$ except the one that lies in the direction of $n$. As we pass through $j$, we compute $\mathrm{Bel}^{U_j}_{\mathfrak{P}_j}$, where $U_j$ is the subtree consiting of $j$ and all the branches of the tree from $j$ except the branch toward $n$.

Figure 8 shows a simple example in which the nodes of $T$ are numbered 1 to 15, and we want to compute $\mathrm{Bel}^T_{\mathfrak{P}_{15}}$. We move from the leaves to node 15 in five steps. At each step, we compute $\mathrm{Bel}^{U_j}_{\mathfrak{P}_j}$ for those $j$ such that we already computed $\mathrm{Bel}^{U_k}\mathfrak{P}_k$ for all $k$ in $U_j = \{j\}$.

**Step 1.** We begin with the leaves; $j = 1, \ldots, 9$. For a leaf $j$, $U_j$ consists just of the leaf; $U_j = (\{j\}, \varnothing)$. So $\mathrm{Bel}^{U_j}_{\mathfrak{P}_j} \doteq \mathrm{Bel}_j$.

**Step 2.** Next, we use versions of (22) to compute $\mathrm{Bel}^{U_j}_{\mathfrak{P}_j}$ for $j = 10$, 11, and 12. We compute $\mathrm{Bel}^{U_{11}}_{\mathfrak{P}_{11}}$, for example, using

$$\mathrm{Bel}^{U_{11}}_{\mathfrak{P}_{11}} = \mathrm{Bel}_{11} \oplus ((\mathrm{Bel}^{U_3}_{\mathfrak{P}_3})_{\mathfrak{P}_{11}} \oplus (\mathrm{Bel}^{U_4}_{\mathfrak{P}_4})_{\mathfrak{P}_{11}})$$

**Step 3.** Now we deal with node 13;

$$\mathrm{Bel}^{U_{13}}_{\mathfrak{P}_{13}} = \mathrm{Bel}_{13} \oplus ((\mathrm{Bel}^{U_8}_{\mathfrak{P}_8})_{\mathfrak{P}_{13}} \oplus (\mathrm{Bel}^{U_{12}}_{\mathfrak{P}_{12}})_{\mathfrak{P}_{13}})$$

**Step 4.** Now we compute $\mathrm{Bel}^T_{\mathfrak{P}_{14}}$ from $\mathrm{Bel}_{14}$, $\mathrm{Bel}^{U_{11}}_{\mathfrak{P}_{11}}$, and $\mathrm{Bel}^{U_{13}}_{\mathfrak{P}_{13}}$.

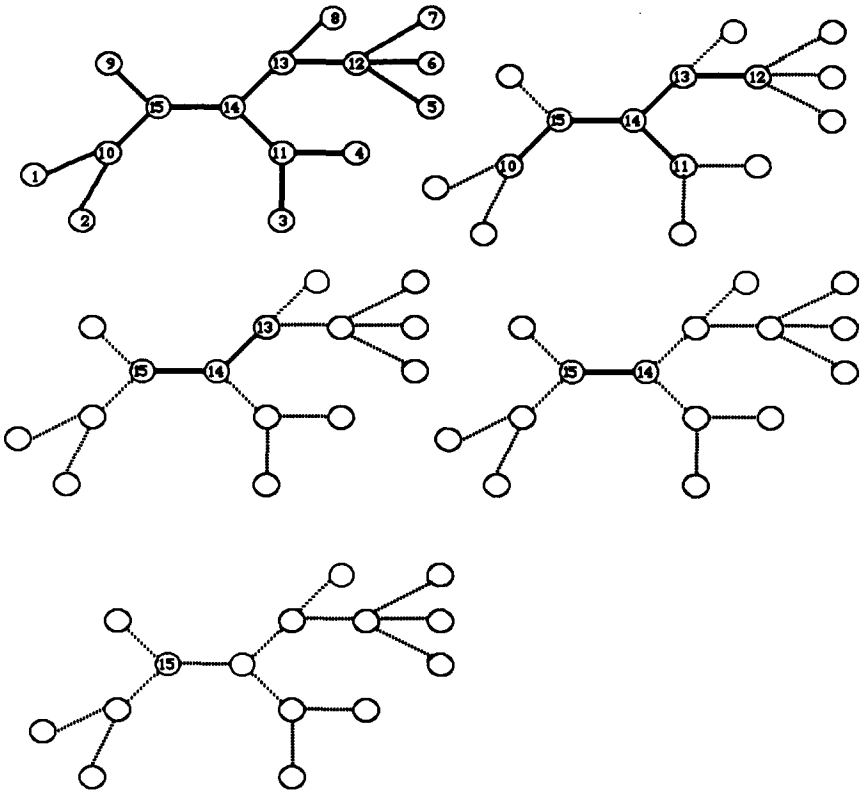


**Figure 8.** Computation for a Single Partition

**Step 5.** Finally, we compute $\text{Bel}^{T}_{\mathfrak{P}_{15}}$ from $\text{Bel}_{15}$, $\text{Bel}^{U_9}_{\mathfrak{P}_9}$, $\text{Bel}^{U_{10}}_{\mathfrak{P}_{10}}$, and $\text{Bel}^{U_{14}}_{\mathfrak{P}_{14}}$. Notice that the nodes $j$ for which we compute $\text{Bel}^{U_j}_{\mathfrak{P}_j}$ at a given step are those nodes that would become leaves were we to delete all the nodes we have dealt with on earlier steps, except that we wait until the last step to deal with node $n$ (node 15 in this case).

The simplicity of this scheme is somewhat obscured by the notation. Here is an alternative notation that may make its simplicity and recursive nature clearer. For any two neighboring nodes $i$ and $j$ in the tree $T$, set

$$\text{Bel}_{j\to i} = (\text{Bel}^{T_{j,i}}_{\mathfrak{P}_j})_{\mathfrak{P}_i} \tag{23}$$

$\text{Bel}_{j\to i}$ is the information that we need from the neighbor $j$ when we are computing $\text{Bel}^{U_i}_{\mathfrak{P}_i}$. If we use this new notation on the right-hand side of (22), we obtain

$$\text{Bel}^{T}_{\mathfrak{P}_n} = \text{Bel}_n \oplus (\oplus \{\text{Bel}_{k\to n} | k \in V_n\}) \tag{24}$$

Putting the tree $T_{j,i}$ in the role of $T$ and the node $j$ in the role of $n$ in (24), we

obtain

$$\mathrm{Bel}_{\mathfrak{B}_j}^{T_{j,i}} = \mathrm{Bel}_j \oplus (\oplus\{\mathrm{Bel}_{k\to j} | k \in (V_j - \{i\})\})$$

Substituting this in (23), we obtain

$$\mathrm{Bel}_{j\to i} = (\mathrm{Bel}_j \oplus (\oplus\{\mathrm{Bel}_{k\to j} | k \in (V_j - \{i\})\}))_{\mathfrak{B}_i} \tag{25}$$

for any neighboring nodes $i$ and $j$.

Notice that if $j$ is a leaf, then $i$ is its only neighbor, and the set $V_j - \{i\}$ is empty; in this case Eq. (25) reduces to $\mathrm{Bel}_{j\to i} = (\mathrm{Bel}_j)_{\mathfrak{B}_i}$.

Formulas (24) and (25) constitute a recursive program for computing $\mathrm{Bel}_{\mathfrak{B}_n}^T$.

This program is easily implemented in a forward-chaining production system. Begin with a working memory that contains $\mathrm{Bel}_i$ for each node $i$ in $N$, and put two rules in the rule base:

RULE 1 If j $\in$ (N $-$ {n}), i $\in$ $V_j$, *$Bel_{k\to j}$ is present in working memory for every* k *in* $V_j$ $-$ {i}, *and $Bel_j$ is present in working memory,* then *use (25) to compute $Bel_{j\to i}$, and place it in working memory.*

RULE 2: If *$Bel_{k\to n}$ is present in working memory for every* k *in* $V_n$, *and $Bel_n$ is present in working memory,* then *use Eq. (24) to compute $Bel_{\mathfrak{B}_n}^T$, and print it.*

Notice that $i$ and $j$ are variables in Rule 1, whereas $n$ is a constant in both rules; $n$ is the node for which we want to compute $\mathrm{Bel}_{\mathfrak{B}_n}^T$. Initially, Rule 1 will fire only for leaves, since initially no $\mathrm{Bel}_{k\to j}$ are in working memory. But eventually Rule 1 will fire once for every edge, resulting in $\mathrm{Bel}_{j\to i}$, where $j$ is the node on the edge farther from $n$, and $i$ is the one closer to (or equal to) $n$. After Rule 1 has fired for every edge, Rule 2 will fire once, completing the computation. Since a tree with $|N|$ nodes has $|N| - 1$ edges, the total number of firings is equal to $|N|$. (We assume a refractory principle that prevents a rule from firing again for the same instantiation of the antecedent.)

## Parallel Computation of Coarsenings

We have been discussing how to compute $\mathrm{Bel}_{\mathfrak{B}_n}^T$ for a single node $n$. If we want to compute $\mathrm{Bel}_{\mathfrak{B}_i}^T$, for all $i$, then we can achieve economies by working through the computations simultaneously, so that a particular $\mathrm{Bel}_{j\to i}$ is computed only once. Implementing this simultaneous computation in a production system requires only a slight modification of our two rules. Modify Rule 1 so that it can fire for all edges in both directions, and modify Rule 2 so that it fires for all nodes:

RULE 1′: If j $\in$ N, i $\in$ $V_j$, *$Bel_{k\to j}$ is present in working memory for every*

k *in* $V_j - \{i\}$, *and* $Bel_j$ *is present in working memory,* then *use (25) to compute* $Bel_{j \to i}$, *and place it in working memory.*

RULE 2′: If i ∈N, $Bel_{k \to i}$ *is present in working memory for every* k *in* $V_i$, *and* $Bel_i$ *is present in working memory,* then *use (24) to compute* $Bel^T_{\mathfrak{B}_i}$, *and print it.*

We again assume a refractory principle that prevents repetitions. Rule 1′ will eventually fire in both directions for every edge $\{i, j\}$, producing both $\text{Bel}_{j \to i}$ and $\text{Bel}_{i \to j}$. Rule 2′ will eventually fire for every $i$. Thus the total number of firings will be $2(|N| - 1) + |N| = 3|N| - 2$.

The speed of both our computational schemes (the one for a single $\text{Bel}^T_{\mathfrak{B}_n}$ and the one for all $\text{Bel}^T_{\mathfrak{B}_i}$) can be enhanced by parallel implementation. In the spirit of Pearl [6], we can make this potential for parallelism graphic by imagining that a separate processor is assigned to each node. The processor assigned to node $j$ computes $\text{Bel}^T_{\mathfrak{B}_j}$ and $\text{Bel}_{j \to k}$ using (24) and (25), respectively. To do this, it must be able to combine belief functions using $\mathfrak{B}_j$ as a frame, and it must be able to project belief functions from $\mathfrak{B}_j$ to any neighbor $\mathfrak{B}_k$.

Since the processor assigned to $\mathfrak{B}_j$ communicates directly with the processor devoted to $\mathfrak{B}_k$ only when $k$ is a neighboring of $i$, the Markov tree is a picture of the architecture of the parallel machine; the nodes are processors, and the edges are communication lines. In this machine, the working memory of the production system is replaced by local memory registers on the edges. We may assume that there are two registers on each edge—one for communication in each direction. On the edge between $j$ and $k$, say, there will be a register where $j$ writes $\text{Bel}_{j \to k}$ for $k$ to read and another where $k$ writes $\text{Bel}_{k \to j}$ for $j$ to read. Each processor $j$ also has an input register, where $\text{Bel}_j$ is written from outside the machine, and an output register, where it writes $\text{Bel}^T_{\mathfrak{B}_j}$. Figure 9 shows a typical processor, with three neighbors.

As Pearl [6] has emphasized, a parallel machine of this type could operate successfully under many different control regimes. In general, no global control is necessary. We can imagine that the machine initially has vacuous belief functions in all its input registers, resulting in vacuous belief functions in the other registers as well. The vacuous belief functions in the input registers can be replaced by nonvacuous belief functions as evidence is obtained and assessed. As long as each processor occasionally checks its inputs and recomputes its outputs, the belief functions in the output registers will eventually be the desired coarsenings.

If we want to complete the computation as quickly as possible, with no wasted effort, then we will require that the processor at $i$ begin work on the computations it is authorized to perform as soon as it receives the necessary inputs. It must begin work on $\text{Bel}_{i \to j}$ as soon as it receives $\text{Bel}_i$ and $\text{Bel}_{k \to i}$ for all $k \in (V_j - \{j\})$, and it must begin work on $\text{Bel}^T_{\mathfrak{B}_i}$ as soon as it receives $\text{Bel}_i$ and $\text{Bel}_{k \to i}$ for all $k \in V_i$. We further require that it not repeat computations. If we impose this

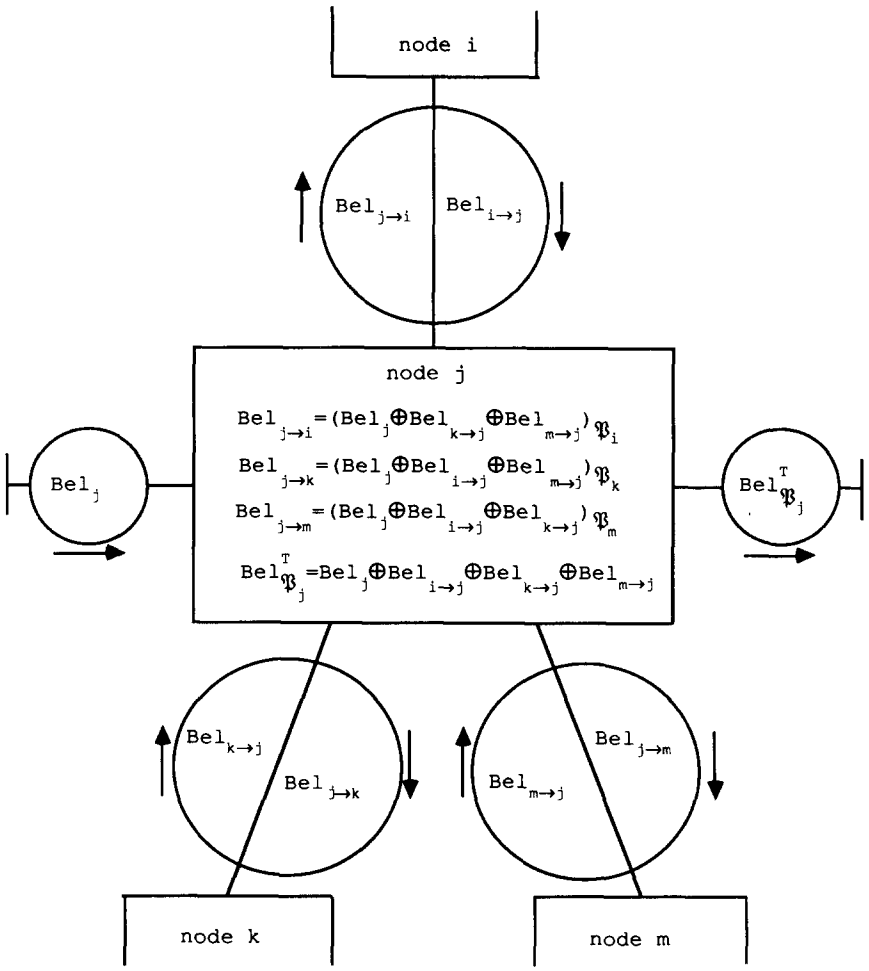


**Figure 9.** A Typical Processor (with Three Neighbors)

regime, and we input all the $Bel_i$ before turning the processors on, then the machine will operate in fundamentally the same way as the production system we described above. But its parallelism will make it faster.

To fix ideas, assume that a processor requires just one unit of time to perform and report all the computations it is authorized to perform with any given set of inputs. Then the time required to compute all the coarsenings will be equal to the diameter of the tree, the length of the longest path through the tree. In fact, the time required to compute the coarsening for a given node $n$ will be equal to the distance to the node farthest from $n$.

To see that this is true, notice first that the total time required to compute the

coarsening for $n$ must be at least the time to the node farthest from $n$, since this much time is required to get information from that node to $n$. On the other hand, if we move along the path from this farthest node as the computations are done, we will encounter no delays; since we are on the longest path, all the other inputs needed for our next step will always be on hand when we arrive.

### Control and Design

The most striking aspect of the general computational schemes we have just discussed is the absence of global control. First we imagine a foward-chaining production system in which there is no explicit control over the order in which rules are fired, and then we imagine autonomous processors acting on their own with no global control.

This lack of global control should not always be taken seriously, however. As Clancey [37] has pointed out, the absence of explicit control that is so attractive in pure production systems must usually be abandoned in actual expert systems. In general, part of an expert's knowledge is control knowledge—principles that help him or her decide what to do next. As Cohen et al. [38] have pointed out, such global control knowledge is necessary for two reasons. First, our problem may be too large and complex for it to be feasible to propagate all our evidence. We may need to decide which part to work on, and this may require global control. Second, the evidence may not be free. We have assumed that the evidence bearing directly on each partition is fixed; we have even assumed that it has already been represented by a belief function. In practice, however, we may need to decide which evidence to pay for, and this also may require global control.

If we do not take the absence of global control seriously, then what value remains in our picture of belief functions propagated through a Markov tree? The answer is that this picture is a design, in the sense of Shafer and Tversky [39]. It tells us the structure of our argument, to the extent that we do obtain the evidence and implement that argument. It gives the structure of legitimate inferences.

We have presented the qualitative Markov tree as a computational device. Conceptually, we combine our belief functions on the overall frame $\Theta$, but it is computationally infeasible to do this, so we propagate them in the tree instead. In fact, however, the structure represented by the tree can be much more than a computational device. It can be a framework for organizing our knowledge and thinking about our evidence. It can be a knowledge-engineering tool.

### The Generality of the Scheme

Here we will sketch the special cases of our general scheme that have been given by Shafer and Logan [2], Shafer [3], and Pearl [6]. We will also discuss whether the scheme needs to be generalized further.

DIAGNOSTIC TREES The work of Shafer and Logan [2] and Shafer [3] is concerned with propagation of belief functions in diagnostic trees. As we mentioned earlier in our discussion of qualitative independence, these articles use the language of partitions, but instead of using the terminology of qualitative independence, they talk about "discernment of interaction." Moreover, they do not use qualitative Markov trees to describe the diagnostic problem. Instead, they work with the hierarchical structure. The algorithms they give conform, however, to the general scheme we have presented here. These algorithms work up the hierarchical structure to its root node and then back down again; this corresponds to working in toward the middle of the tree of families and dichotomies and then back out again.

When we use the tree of families and dichotomies for belief-function propagation, we can include in the propagation any belief function over any of the family frames, $\mathfrak{D}_\Theta$ and the $\mathfrak{D}_A \cup \{A^c\}$. These family frames can be relatively large, and the belief functions on them can be relativel complicated. Shafer [3] allows this full generality. Shafer and Logan [2] are concerned, however, with the special case where each belief function is carried by a dichotomy $\{A, A^c\}$. Propagation in this special case is still best understood as propagation through the tree of families and dichotomies, but the belief functions being combined on the family frames are all very simple; they are "simply support functions" focused on singletons and their complements. Barnett [36] has shown that when we are combining such functions we can implement Dempster's rule very efficiently, with the number of operations being linear rather than exponential in the size of the frame. By taking advantage of Barnett's algorithm, Shafer and Logan obtain an algorithm for propagation with complexity proportional to the total number of terminal nodes in the diagnostic tree.

PROBABILISTIC MARKOV TREES Consider a frame $\Theta$ with a probability distribution Pr such that $\Pr(\{\theta\}) > 0$ for all $\theta \in \Theta$. Suppose $T$ is a tree of partitions of $\Theta$, and suppose separation in the tree implies independence with respect to Pr. In other words, the conditions of Theorem 1 are satisfied when $[\mathfrak{P}', \mathfrak{P}''] \dashv \mathfrak{P}$ is interpreted to mean that $\mathfrak{P}'$ and $\mathfrak{P}''$ are independent given $\mathfrak{P}$ with respect to Pr. Then we call $T$ a *probabilistic Markov tree.* (See Lauritzen and Spiegelhalter [11] and the references therein—e.g., Moussouris [40]). It is evident from the corollary to Lemma 1 that a probabilistic Markov tree is also qualitative Markov.

Conversely, given a qualitative Markov tree $T = (N, E)$, we can easily construct a probability distribution Pr that makes $T$ probabilistic Markov. We arbitrarily choose a node $n$ and think of $T$ as a rooted tree with $n$ as its root. This makes every element of $E$ a mother-daughter pair. We then supply "prior probabilities" for the root and "transition probabilities" for each mother-daughter pair. This means we supply

$$\Pr(A) \text{ for every element } A \text{ of } \mathfrak{P}_n \tag{26}$$

and for every mother-daughter pair $(i, j)$ we supply

$$\Pr(C|B) \text{ for every } B \text{ in } \mathfrak{P}_i \text{ and } C \text{ in } \mathfrak{P}_j \tag{27}$$

Then for every selection of elements $P_i$ from the partitions $\mathfrak{P}_i$, we set

$$\Pr(\cap\{P_i|i\in N\}) = \Pr(P_n)\prod_{(i,j)}\Pr(P_j|P_i) \tag{28}$$

where the product is over every mother-daughter pair $(i, j)$. If we assume, for the sake of simplicity, that the refinement of all the $\mathfrak{P}_i$ is the set of singleton subsets of $\Theta$, then (28) completely determines a probability distribution over $\Theta$.

More interesting in the present context than (28) is the obvious fact that (26) and (27) allow us to compute the unconditional distributions for the $\mathfrak{P}_i$ step by step as we move down the tree from its root. We initially have the ingredients to compute the distribution for any daughter $i$ of the root $n$;

$$\Pr(B) = \sum\{\Pr(A)\Pr(B|A)|A\in\mathfrak{P}_n\}. \tag{29}$$

for each element $B$ of $\mathfrak{P}_i$. Then we can compute the distribution for any daughter $j$ of $i$;

$$\Pr(C) = \sum\{\Pr(B)\Pr(C|B)|B\in\mathfrak{P}_i\}. \tag{30}$$

for each element $C$ of $\mathfrak{P}_j$; and so on.

This step-by-step computation of probability distributions can be interpreted as a special case of our general scheme for propagating belief functions. To see this, we need to make the following observations:

- Any probability distribution qualifies as a belief function, and hence a probability distribution for $\mathfrak{P}_i$ can be thought of as a belief function carried by $\mathfrak{P}_i$ (Shafer [1]).
- Any set of transition probabilities for $\mathfrak{P}_i$ to $\mathfrak{P}_j$ can be represented by a belief function. More precisely, given transition probabilities $\Pr(C|B)$, $B \in \mathfrak{P}_i$, and $C \in \mathfrak{P}_j$, there will be a belief function $\text{Bel}_{ij}$ (actually, there will be several) that has a vacuous projection on $\mathfrak{P}_i$ and yet satisfies

  $$(\text{Bel}_i \oplus \text{Bel}_{ij})(C) = \sum\{\text{Bel}_i(B)\Pr(C|B)|B\in\mathfrak{P}_i\}$$

  in agreement with (30), whenever $\text{Bel}_i$ is a belief function carried by $\mathfrak{P}_i$ and happening to be a probability distribution there.
- If between every mother-daughter pair $(i, j)$ in $T$ we interpolate a node corresponding to $\mathfrak{P}_i \wedge \mathfrak{P}_j$, then we still have a Markov tree, qualitatively and probabilistically (see the section Transformations of Qualitative Markov Trees and Figure 10).

Putting these observations together, we see that the step-by-step calculation of the distributions for the $\mathfrak{P}_i$ by (29) and (30) correspond to belief-function propagation after entering nonvacuous belief functions at the root node and the interpolated nodes and vacuous belief functions elsewhere. Nonvacuous belief

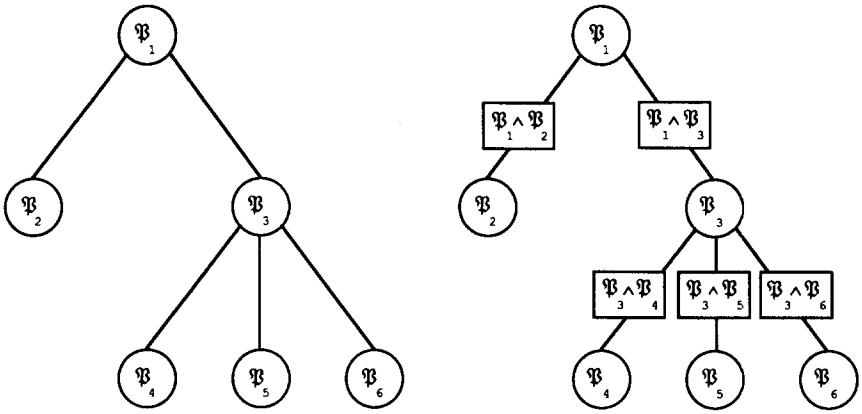

**Figure 10.** Interpolation in a Rooted Markov Tree

functions can also be input at these other nodes. In particular, we can input "categorical" belief functions—belief functions representing observations of variables corresponding to some of the partitions (Shafer [1]; Kong [9]).

Figure 11 shows a more general form of Bayesian propagation in a Markov tree. Here we allow dependent children and multiple parents. Each box contains conditional probabilities for the partition or partitions below given the partition or partitions above. Prior probabilities must be input to all root nodes (both 1 and 2 in the figure). Multiple parents are assumed to be unconditionally independent. The belief-function interpretation we just gave carries over to this more general form of propagation.

Bayesian propagation of the type shown in Figures 10 and 11 has been studied in detail by Pearl [6]. Pearl's pictures are somewhat different, however. His trees have nodes or processors only for individual variables (individual $\mathfrak{P}_i$)—none for joint variables (refinements)—and hence the action within these processors is more complex.

Though Bayesian propagation is a special case of belief-function propagation, it is not a case likely to arise when a problem is approached in a belief-function spirit. Belief functions representing transition probabilities tend to be complex and unnatural, and we will tend to prefer simpler belief functions on the refinements. We may sometimes use a belief function with just two focal elements, one equal to $\Theta$ and the other equal to a particular intersection $P_i \cap P_j$, indicating suspicision that $P_i$ and $P_j$ go together.

In general, Bayesian propagation in trees like those of Figures 10 and 11 is computationally easier than general belief-function propagation. The Bayesian computation is linear in the size of the refinements in the boxes, while the belief-function computation may be exponential in these sizes. This computational advantage must be balanced, however, against the greater demands of the Baye-

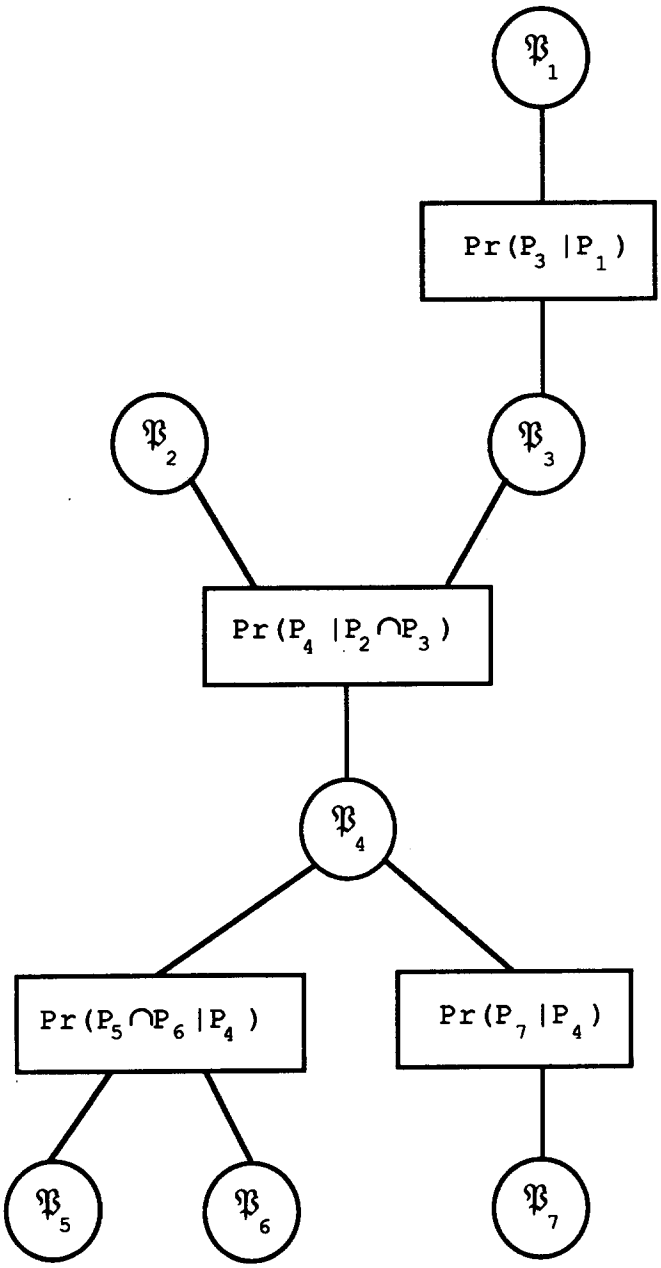


**Figure 11.** General Bayesian Propagation

sian method at the knowledge-engineering stage. The many transition probablities demanded by the Bayesian method may be unavailable.

THE GENERAL PROBLEM Suppose we are concerned with a number of equations, say $Q_1, \ldots, Q_r$. The theory of belief functions suggests that our study of these questions should begin with an analysis of our evidence. How can we sort our evidence into independent items, and on which of these questions do these items of evidence bear?

As Kong [9] has emphasized, such an analysis will usually lead not to a tree but to a hypergraph (see the section Multivariate Markov Trees). Usually each item of evidence will bear on the relationship among some subset of our questions. These subsets, together with the Kong pattern representing the categorical relations among the questions, will constitute a hypergraph on the nodes $\{1, \ldots, r\}$.

As we saw earlier, there are several natural ways to construct qualitative Markov networks from hypergraphs. But unless we begin with a diagnostic or causal tree, these networks will usually not be trees.

It would be nice if we could find simple and efficient computational schemes for networks that extend our scheme for trees. This does not appear possible, however. Escape from the computational complexity of Dempster's rule seems to require a tree.

If it is true, then our task will often be to find the best ways to embed a hypergraph in a tree. In the notation of multivariate Markov trees, the problem is this. Given a hypergraph $(R, W)$, where $R$ indexes a set of partitions of a frame $\Theta$ and $W$ is a Kong pattern for these partitions, find a qualitative Markov tree with subsets of $R$ as nodes and refinements of the partitions in the subsets as the associated partitions, such that each element of $W$ is contained in one of the nodes. We want each element of $W$ to be contained in a node so that we can input the belief function based on the corresponding item of evidence into that node. Notice that embedding $(R, W)$ in a tree in this way corresponds to collapsing, in a certain sense, the network of families to a tree.

Embeddings always exist. We can embed any hypergraph $(R, W)$ in the qualitative Markov tree consisting of the single node $R$. But we want the nodes in the tree to be as small as possible, in the hope that the associated partitions will be small enough for propagation to be feasible. Finding embeddings with the smallest possible nodes is the main question addressed by Kong [9], Mellouli [10], and Lauritzen and Spiegelhalter [11].

INDEPENDENCE Throughout this article we have assumed that the items of evidence on which our different belief functions are based are independent, so that it is legitimate to combine these belief functions by Dempster's rule. Many authors have questioned this assumption and have called for generalizing Dempster's rule to the case of dependent evidence.

Some progress has been made in generalizing Dempster's rule; see Shafer [22, 26]. It should be noted, however, that independence is not so much an assumption about our knowledge as a knowledge-engineering strategy. We do not begin, in general, with well-defined evidence divided into distinct items and set out for inspection. Usually we must search for evidence. We must decide what qualifies as evidence. We must turn hunches, dry facts, and confused arguments into evidence. At this knowledge-engineering stage, it is useful to look for independent uncertainties and to build up evidence around them. It is this belief-function strategy that leads to independent items of evidence in qualitative Markov trees.

## References

1. Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.

2. Shafer, G., and Logan, R., Implementing Dempster's rule for hierarchical evidence, *AI*, 1987, to appear.

3. Shafer, G., Hierarchical evidence, in *Proceedings of the 2nd Conference on AI Applications*, Miami Beach, FL, IEEE Computer Society Press, 16–25, 1985.

4. Pearl, J., Reverend Bayes on inference engines: a distributed hierarchical approach, in *Proceedings of the 2nd National Conference on AI, AAAI-82*, Pittsburg, PA, 133–136, 1982.

5. Pearl, J., On evidential reasoning in a hierarchy of hypotheses, *AI* 28, 9–15, 1986.

6. Pearl, J., Fusion, propagation, and structuring in Bayesian networks, *AI* 29, 241–288, 1986.

7. Pearl, J., Distributed revision of belief commitment in multi-hypotheses interpretation, in *Proceedings of the 2nd AAAI Workshop on Uncertainty in AI*, Philadelphia, PA, 201–209, 1986.

8. Shenoy, P. P., and Shafer, G., Propagating belief functions with local computations, *IEEE Expert* 1(3), 43–52, 1986.

9. Kong, A., Multivariate belief functions and graphical models, doctoral dissertation, Department of Statistics, Harvard University, 1986.

10. Mellouli, K., On the propagation of beliefs in networks using the Dempster-Shafer theory of evidence, doctoral dissertation, School of Business, University of Kansas, 1987.

11. Lauritzen, S. L., and Spiegelhalter, D. J., Fast manipulation of probabilities with local representations—with applications to expert systems, Report R-87-7, Institute of Electronic Systems, Aalborg Univeristy, Aalborg, Denmark, 1987.

12. Breiman, L., *Probability*, Addison-Wesley, Reading, MA, 1968.

13. Halmos, P. R., *Measure Theory*, Van Nostrand, New York, 1950.

14. Birkhoff, G., *Lattice Theory*, 3rd ed., American Mathematical Society, Providence, RI, 1967.

15. Fagin, R., Multivalued dependencies and a new normal form for relational databases, *Assoc. Computer Mach. Trans. Database Syst.* 2, 272–278, 1977.

16. Maier, D., *The Theory of Relational Databases* Computer Science Press, Rockville, MD, 1983.

17. Pawlak, Z., Rough sets, Int J. *Inf. Computer Sci.* 11, 341–356, 1982.

18. Pawlak, Z., Rough classification, *Int. J. Man-Machine Studies* 20, 469–483, 1984.

19. Sagiv, Y., and Walecka, S. F., Subset dependencies and a completeness result for a subclass of embedded multivalued dependencies. *J. Assoc. Comput. Mach.* 29, 103–117, 1982.

20. Pearl, J., and Verma, T., The logic of representing dependencies by directed graphs, *Proceedings of the 6th National Conference on AI,* AAAI-87, Seattle, WA, 374–379, 1987.

21. Shafer, G., The combination of evidence, *Int. J. Intelligent Syst.* 1, 155–179, 1986.

22. Shafer, G., Belief functions and possibility measures, in *The Analysis of Fuzzy Information*, Vol. 2, J. Bezdek (Ed.), CRC Press, Boca Raton, FL, 1987.

23. Shafer, G., and Srivastava, R., The Bayesian and belief-function formalisms: a general perspective for auditing, Working Paper No. 189, School of Business, University of Kansas, 1987.

24. Garvey, T. D., Lowrance, J. D., and Fischler, M. A., An inference technique for integrating knowledge from disparate sources, in *Proceedings of the 7th International Joint Conference on AI*, Vancounver, BC, 1981, pp. 487–495.

25. Gordon, J., and Shortliffe, E. H., The Dempster-Shafer theory of evidence, in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, B. G. Buchanan and E. H. Shortliffe, Eds., Addison-Wesley, Reading, MA, 272–292, 1984.

26. Shafer, G., Probability judgment in artificial intelligence and expert systems (with discussion), *Stat. Sci.,* 2, 3–44, 1987.

27. Shafer, G., Constructive probability, *Synthese* 48, 1–60, 1981.

28. Shafer, G., Lindley's paradox (with discussion), *J. Amer. Stat. Assoc.* 77, 325–351, 1982.

29. Shafer, G., Belief functions and parametric models (with discussion), *J. Roy. Stat. Soc. Ser. B* 44, 322–352, 1982.

30. Zhang, L., Weights of conflict and internal conflict for support functions, *Inf. Sci.* 38, 205–212, 1986.

31. Dempster, A. P., Upper and lower probabilities generated by a random closed interval, *Ann. Math. Stat.* 39, 957–966, 1968.

32. Dempster, A. P., A generalization of Bayesian inference (with discussion), *J. Roy. Stat. Soc. Ser. B* 30, 205–247, 1968.

33. Dempster, A. P., Upper and lower probability inferences for families of hypotheses with monotone density ratios, *Ann. Math. Stat.* 40, 953–969, 1969.

34. Shafer, G., Allocations of probability, *Ann. Probab.* 7, 827–839, 1979.

35. Strat, T. M., Continuous belief functions for evidential reasoning, in *Proceedings of the 4th National Conference on, AAAI-84*, Austin, TX, 308–313, 1984.

36. Barnett, J. A., Computational methods for a mathematical theory of evidence, in *Proceedings of the 7th International Joint Conference on AI*, Vancouver, BC, 868–875, 1981.

37. Clancey, W.J., The advantages of abstract control knowledge in expert system design, in *Proceedings of the 3rd National Conference on AI, AAAI-83*, Washington, DC, 1983, pp. 64–78.

38. Cohen, P., Day, D., DeLisio, J., Greenberg, M., Kjeldsen, R., Suthers, D., and Berman, P., Management of uncertainty in medicine, *Int. J. Approximate Reasoning*, 1, 103–116, 1987.

39. Shafer, G., and Tversky, A., Language and designs for probability judgment, *Cognitive Sci.* 9, 309–339, 1985.

40. Moussouris, J., Gibbs and Markov random systems with constraints, *J. Stat. Phys.* 10, 11–33, 1974.