Appeared in : Yager, R. R., M. Federizzi, and J. Kacprzyk, eds., Advances in the Dempster-Shafer Theory of Evidence, 1994, pp. 395–414, John Wiley & Sons, New York, NY.

USING DEMPSTER-SHAFER'S BELIEF-FUNCTION THEORY IN EXPERT SYSTEMS

by

Prakash P. Shenoy

School of Business, University of Kansas, Lawrence, KS 66045-7585, USA
Tel: (785)-864-7551, Fax: (785)-864-5328, Email: pshenoy@ku.edu

ABSTRACT

The main objective of this paper is to describe how Dempster-Shafer's (DS) theory of belief functions fits in the framework of valuation-based systems (VBS). Since VBS serve as a framework for managing uncertainty in expert systems, this facilitates the use of DS belief-function theory in expert systems.

Keywords: Dempster-Shafer's theory of belief functions, valuation-based systems, expert systems

1. INTRODUCTION

The main goal of this paper is to describe how Dempster-Shafer's (DS) theory of belief function can be used for managing uncertainty in expert systems. We describe how belief-function theory fits in the framework of valuation-based systems (VBS). Since VBS serve as a framework for managing uncertainty in expert systems, this facilitates the use of belief-function theory in expert systems.

The theory of belief functions was first described by Dempster (1967) and further developed by Shafer (1976). In belief-function theory, the basic representational unit is called a basic probability assignment (bpa) function. The two main operations for manipulating bpa functions are marginalization and Dempster's rule of combination.

The framework of VBS was first defined by Shenoy (1989) as a general language for incorporating uncertainty in expert systems. It was further elaborated in (Shenoy, 1991c) to include axioms that permit local computation in solving a VBS, and a fusion algorithm for solving a VBS using local computation. VBS encode knowledge using functions called valuations. VBS include two operators called combination and marginalization that operate on valuations. Combination corresponds to aggregation of knowledge. Marginalization corresponds to coarsening of knowledge. The process of reasoning in VBS can be described simply as finding the marginal of the joint valuation for each variable in the system. The joint valuation is the valuation obtained by combining all valuations. In systems with many variables, it is computationally intractable to explicitly compute the joint valuation. However, if combination and marginalization satisfy certain axioms, it is possible to compute the marginals of the joint valuation without explicitly computing the joint valuation.

The framework of VBS is general enough to represent many domains such as Bayesian probability theory (Shenoy, 1991c), Spohn's theory of epistemic beliefs (Spohn, 1988; Shenoy, 1991a), Zadeh's possibility theory (Zadeh, 1979; Shenoy, 1991d), discrete optimization (Shenoy 1991b), propositional logic (Shenoy, 1990a,b), constraint satisfaction (Shenoy and Shafer, 1988), and Bayesian decision theory (Shenoy, 1990c,d). Saffiotti and Umkehrer (1991) describe an efficient implementation of VBS called Pulcinella.

The correspondence between belief-function theory and VBS is as follows. Dempster's rule of combination in belief-function theory corresponds to the combination operation in VBS. And the marginalization operation in belief-function theory corresponds to the marginalization operation in VBS. The framework of VBS was inspired by the axiomatic study of the computational theory in Bayesian probability theory and belief-function theory (Shenoy and Shafer, 1990).

An outline of this paper is as follows. Section 2 describes the framework of VBS. Section 3 describes the main features of belief-function theory in terms of the framework of VBS. Section 4 describes a small example illustrating the use of belief-function theory for managing uncertainty in expert systems. Section 5 contains some concluding remarks. Finally, section 6 contains proofs of all results in the paper.

2. VALUATION-BASED SYSTEMS

In this section, we sketch the basic features of VBS. Also, we describe three axioms that permit the use of local computation, and describe a fusion algorithm for solving a VBS using local computation.

2.1. The framework

This subsection describes the framework of valuation-based systems. In a VBS, we represent knowledge by functions called valuations. We make inferences in a VBS using two operators called combination and marginalization. We use these operators on valuations.

Variables and Configurations. We use the symbol $\mathbf{W}_X$ for the set of possible values of a variable X, and we call $\mathbf{W}_X$ the *frame for X*. We assume that one and only one of the elements of $\mathbf{W}_X$ is the true value of X. We are concerned with a finite set $\mathbf{X}$ of variables, and we assume that all the variables in $\mathbf{X}$ have finite frames.

Given a nonempty set s of variables, let $\mathbf{W}_s$ denote the Cartesian product of $\mathbf{W}_X$ for X in s; $\mathbf{W}_s = \times\{\mathbf{W}_X \mid X \in s\}$. We call $\mathbf{W}_s$ the *frame for s*. We call the elements of $\mathbf{W}_s$ *configurations of s*.

Valuations. Given a subset s of variables, there is a set $\mathbf{V}_s$. We call the elements of $\mathbf{V}_s$ *valuations for s*. Let $\mathbf{V}$ denote the set of all valuations, i.e., $\mathbf{V} = \cup\{\mathbf{V}_s \mid s \subseteq \mathbf{X}\}$. If σ is a valuation for s, then we say s is the *domain* of σ.

Valuations are primitives in our abstract framework and as such require no definition. But as we shall see shortly, they are objects which can be combined and marginalized. Intuitively, a valuation for s represents some knowledge about the variables in s.

Nonzero valuations. For each $s \subseteq \mathbf{X}$, there is a subset $\mathbf{P}_s$ of $\mathbf{V}_s$ whose elements are called *nonzero valuations for s*. Let $\mathbf{P}$ denote $\cup\{\mathbf{P}_s \mid s \subseteq \mathbf{X}\}$, the set of all nonzero valuations.

Intuitively, a nonzero valuation represents knowledge that is internally consistent. The notion of nonzero valuations is important as it enables us to constrain the definitions of combination and marginalization to meaningful operators. An example of a nonzero valuation is a basic probability assignment (bpa) function.

Combination. We assume there is a mapping $\otimes: \mathbf{V} \times \mathbf{V} \to \mathbf{V}$, called *combination*, such that

(i) if ρ and σ are valuations for r and s, respectively, then $\rho \otimes \sigma$ is a valuation for $r \cup s$;

(ii) if either ρ or σ is not a nonzero valuation, then $\rho \otimes \sigma$ is not a nonzero valuation; and

(iii) if ρ and σ are both nonzero valuations, then $\rho \otimes \sigma$ may or may not be a nonzero valuation.

We call $\rho \otimes \sigma$ the *combination of ρ and σ.*

Intuitively, combination corresponds to aggregation of knowledge. If ρ and σ are valuations for r and s representing independent knowledge about variables in r and s, respectively, then $\rho \otimes \sigma$ represents the aggregated knowledge about variables in $r \cup s$. (The definition of independence is given in (Shenoy, 1991e).) For bpa functions, combination is Dempster's rule of combination (described in section 3).

Marginalization. We assume that for each $s \subseteq \mathbf{X}$, and for each $X \in s$, there is a mapping $\downarrow(s-\{X\}): \mathbf{V}_s \to \mathbf{V}_{s-\{X\}}$, called *marginalization to s–{X}* such that

(i) If σ is a valuation for s, then $\sigma^{\downarrow(s-\{X\})}$ is a valuation for s–{X}; and

(ii) $\sigma^{\downarrow(s-\{X\})}$ is a nonzero valuation if and only if σ is a nonzero valuation.

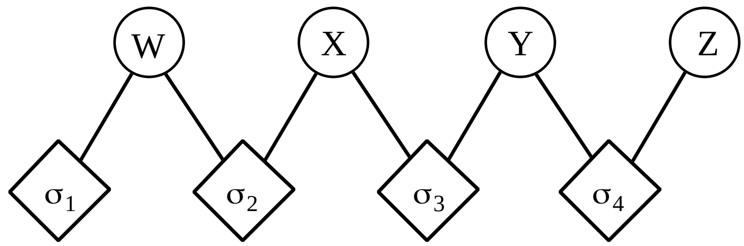
We call $\sigma^{\downarrow(s-\{X\})}$ the *marginal of σ for s–{X}*.

Intuitively, marginalization corresponds to coarsening of knowledge. If σ is a valuation for s representing some knowledge about variables in s, and $X \in s$, then $\sigma^{\downarrow(s-\{X\})}$ represents the knowledge about variables in s–{X} implied by σ if we disregard variable X. In the case of bpa functions, marginalization is addition.

In summary, a *valuation-based system* consists of a 3-tuple $\{\{\sigma_1, ..., \sigma_m\}, \otimes, \downarrow\}$ where $\{\sigma_1, ..., \sigma_m\}$ is a collection of valuations, $\otimes$ is the combination operator, and $\downarrow$ is the marginalization operator.

Valuation Networks. A graphical depiction of a valuation-based system is called a *valuation network*. In a valuation network, variables are represented by circular nodes, and valuations are represented by diamond-shaped nodes. Also, each valuation node is connected by an undirected edge to each variable node in its domain. Figure 1 shows a valuation network for a VBS that consists of valuations σ_1 for {W}, σ_2 for {W, X}, σ_3 for {X, Y}, and σ_4 for {Y, Z}.

Figure 1. A valuation network for a VBS consisting of valuations σ_1 for $\{W\}$, σ_2 for $\{W, X\}$, σ_3 for $\{X, Y\}$, and σ_4 for $\{Y, Z\}$.



Making Inference in VBS. In a VBS, the combination of all valuations is called the *joint valuation*. Given a VBS, we make inferences by computing the marginal of the joint valuation for each variable in the system.

If there are n variables in the system, and each variable has two configurations in its frame, then there are 2^n configurations of all variables. Hence, it is not computationally feasible to compute the joint valuation when there are a large number of variables. In section 2.3, we describe an algorithm for computing marginals of the joint valuation without explicitly computing the joint valuation, i.e., using only local computation. So that this algorithm gives us the correct answers, we require combination and marginalization to satisfy three axioms. The axioms and the algorithm are described in the next two subsections, respectively.

2.2. Axioms

In this section, we state three simple axioms that enable local computation of marginals of the joint valuation. These axioms were first formulated by Shenoy and Shafer (1990). The axioms are stated slightly differently here.

Axiom A1 (*Commutativity and associativity of combination*): Suppose ρ, σ, and τ are valuations for r, s, and t, respectively. Then

$$\rho\otimes\sigma = \sigma\otimes\rho, \text{ and } \rho\otimes(\sigma\otimes\tau) = (\rho\otimes\sigma)\otimes\tau.$$

Axiom A2 (*Order of deletion does not matter*): Suppose σ is a valuation for s, and suppose $X_1, X_2 \in s$. Then

$$(\sigma^{\downarrow(s-\{X_1\})})^{\downarrow(s-\{X_1,X_2\})} = (\sigma^{\downarrow(s-\{X_2\})})^{\downarrow(s-\{X_1,X_2\})}.$$

Axiom A3 (*Distributivity of marginalization over combination*): Suppose ρ and σ are valuations for r and s, respectively. Suppose $X \notin r$, and suppose $X \in s$. Then

$$(\rho\otimes\sigma)^{\downarrow((r\cup s)-\{X\})} = \rho\otimes(\sigma^{\downarrow(s-\{X\})}).$$

One implication of Axiom A1 is that when we have multiple combinations of valuations, we can write it without using parenthesis. For example, $(...((\sigma_1\otimes\sigma_2)\otimes\sigma_3)\otimes...\otimes\sigma_m)$ can be written simply as

$\otimes\{\sigma_i \mid i = 1, ..., m\}$ or as $\sigma_1 \otimes ... \otimes \sigma_m$, i.e., we need not indicate the order in which the combinations are carried out.

If we regard marginalization as a coarsening of a valuation by deleting variables, then axiom A2 says that the order in which the variables are deleted does not matter. One implication of this axiom is that $(\sigma^{\downarrow(s-\{X_1\})})^{\downarrow(s-\{X_1,X_2\})}$ can be written simply as $\sigma^{\downarrow(s-\{X_1,X_2\})}$, i.e., we need not indicate the order in which the variables are deleted.

Axiom A3 is the crucial axiom that makes local computation possible. Axiom A3 states that the computation of $(\rho\otimes\sigma)^{\downarrow((r\cup s)-\{X\})}$ can be accomplished without having to compute $\rho\otimes\sigma$. The combination operation in $\rho\otimes\sigma$ is on the frame for $r\cup s$ whereas the combination operation in $\rho\otimes(\sigma^{\downarrow(s-\{X\})})$ is on the frame for $(r\cup s)-\{X\}$. In the next subsection, we describe a fusion algorithm that applies this axiom repeatedly resulting in an efficient method for computing marginals.

2.3. A fusion algorithm

In this subsection, we describe a fusion algorithm for making inferences in a VBS using local computation. Suppose $\{\{\sigma_1, ..., \sigma_m\}, \otimes, \downarrow\}$ is a VBS with n variables and m valuations. Suppose that combination and marginalization satisfy the three axioms stated in section 2.2. Suppose we have to compute the marginal of the joint valuation for variable X, $(\sigma_1\otimes...\otimes\sigma_m)^{\downarrow\{X\}}$. The basic idea of the fusion algorithm is to successively delete all variables but X from the VBS. The variables may be deleted in any sequence. Axiom A2 tells us that all deletion sequences lead to the same answers. But, different deletion sequences may involve different computational costs. We will comment on good deletion sequences at the end of this section.

When we delete a variable, we have to do a "fusion" operation on the valuations. Consider a set of k valuations $\rho_1, ..., \rho_k$. Suppose ρ_i is a valuation for r_i. Let $Fus_X\{\rho_1, ..., \rho_k\}$ denote the collection of valuations after fusing the valuations in the set $\{\rho_1, ..., \rho_k\}$ with respect to variable X. Then

$$Fus_X\{\rho_1, ..., \rho_k\} = \{\rho^{\downarrow(r-\{X\})}\}\cup\{\rho_i \mid X\notin r_i\}$$

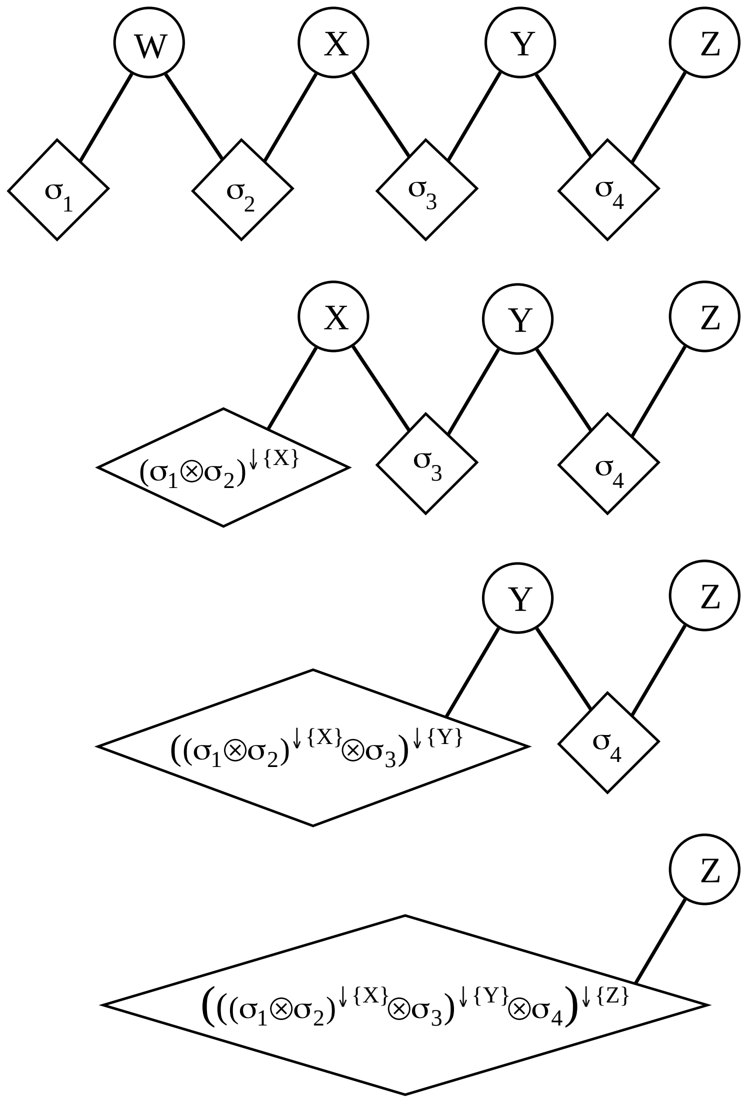
where $\rho = \otimes\{\rho_i \mid X\in r_i\}$, and $r = \cup\{r_i \mid X\in r_i\}$. After fusion, the set of valuations is changed as follows. All valuations that bear on X are combined, and the resulting valuation is marginalized such that X is eliminated from its domain. The valuations that do not bear on X remain unchanged.

We are ready to state the theorem which describes the fusion algorithm.

Theorem 1. Suppose $\{\{\sigma_1, ..., \sigma_m\}, \otimes, \downarrow\}$ is a VBS, where σ_i is a valuation for s_i, and suppose $\otimes$ and $\downarrow$ satisfy axioms A1–A3. Let $\mathbf{X}$ denote $s_1\cup...\cup s_m$. Suppose $X\in\mathbf{X}$, and suppose $X_1X_2...X_{n-1}$ is a sequence of variables in $\mathbf{X}-\{X\}$. Then

$$(\sigma_1\otimes...\otimes\sigma_m)^{\downarrow\{X\}} = \bigotimes\left\{Fus_{X_{n-1}}\{... Fus_{X_2}\{Fus_{X_1}\{\sigma_1, ..., \sigma_m\}\}\}\right\}.$$

Figure 2. The first valuation network shows the initial VBS. The second network is the result after fusion with respect to W. The third network is the result after fusion with respect to X. The fourth network is the result after fusion with respect to Y.



Example 1 (Fusion Algorithm). Consider a VBS with four variables: W, X, Y, and Z, and four valuations: σ_1 for {W}, σ_2 for {W, X}, σ_3 for {X,Y}, and σ_4 for {Y, Z}. Suppose we need to compute the marginal of the joint for Z, $(\sigma_1 \otimes ... \otimes \sigma_4)^{\downarrow\{Z\}}$. Consider the deletion sequence WXY. After fusion with respect to W, we have $(\sigma_1 \otimes \sigma_2)^{\downarrow\{X\}}$ for {X}, σ_3 for {X, Y}, and σ_4 for {Y, Z}. After fusion with respect to X, we have $((\sigma_1 \otimes \sigma_2)^{\downarrow\{X\}} \otimes \sigma_3)^{\downarrow\{Y\}}$ for {Y}, and σ_4 for {Y, Z}. Finally, after fusion with respect to Y, we have $\left(((\sigma_1 \otimes \sigma_2)^{\downarrow\{X\}} \otimes \sigma_3)^{\downarrow\{Y\}} \otimes \sigma_4\right)^{\downarrow\{Z\}}$ for {Z}. Theorem 1 tells us that $\left(((\sigma_1 \otimes \sigma_2)^{\downarrow\{X\}} \otimes \sigma_3)^{\downarrow\{Y\}} \otimes \sigma_4\right)^{\downarrow\{Z\}} = (\sigma_1 \otimes ... \otimes \sigma_4)^{\downarrow\{Z\}}$. Thus, instead of doing combinations on the frame for {W, X, Y, Z}, we do combinations on the frame for {W, X}, {X, Y}, and {Y, Z}. The fusion algorithm is shown graphically in Figure 2. ■

If we can compute the marginal of the joint valuation for one variable, then we can compute the marginals for all variables. We simply compute them one after the other. It is obvious, however, that this will involve much duplication of effort. Shenoy and Shafer (1990) describe an efficient algorithm for simultaneous computation of all marginals without duplication of effort. Regardless of the number of variables in a VBS, we can compute marginals of the joint valuation for all variables for roughly three times the computational effort required to compute one marginal.

Deletion Sequences. Different deletion sequences may involve different computational efforts. For example, consider the VBS in the above example. In this example, deletion sequence WXY involves less computational effort than, for example, XYW, as the former involves combinations on the frame for two variables only whereas the latter involves combination on the frame for three variables. Finding an optimal deletion sequence is a secondary optimization problem that has shown to be NP-complete (Arnborg et al., 1987). But, there are several heuristics for finding good deletion sequences (Kong, 1986; Mellouli, 1987; Zhang, 1988; Kjærulff, 1990).

One such heuristic is called one-step-look-ahead (Kong, 1986). This heuristic tells us which variable to delete next. As per this heuristic, the variable that should be deleted next is one that leads to combination over the smallest frame. For example, in the VBS described above, if we assume that each variable has a frame consisting of two configurations, then this heuristic would pick W over X and Y for first deletion since deletion of W involves combination on the frame for {W, X} whereas deletion of X involves combination on the frame for {W, X, Y}, and deletion of Y involves combination on the frame for {X, Y, Z}. After W is deleted, for second deletion, this heuristic would pick X over Y. Thus, this heuristic would choose deletion sequence WXY.

3 DS THEORY OF BELIEF FUNCTIONS

In this section, we describe the main features of DS theory of belief functions in terms of the framework of VBS described in the previous section. The basic unit of knowledge representation is called a basic probability assignment (bpa) function.

> ***Definition 1 (Bpa function)***. A *basic probability assignment (bpa) function* μ *for h* is a function $\mu: 2^{\mathbf{W}_h} \rightarrow [0, 1]$ such that $\Sigma\{\mu(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathbf{W}_h}\} = 1$. ($2^{\mathbf{W}_h}$ denotes the set of all nonempty subsets of $\mathbf{W}_h$).

Intuitively, $\mu(\mathbf{a})$ represents the degree of belief assigned exactly to $\mathbf{a}$ (the proposition that the true configuration of h is in the set $\mathbf{a}$) and to nothing smaller. A bpa function is the belief function equivalent of a probability mass function in probability theory. Whereas a probability mass function is restricted to assigning probability masses only to singleton configurations of variables, a bpa function is allowed to assign probability masses to sets of configurations without assigning any mass to the individual configurations contained in the sets.

Consider the following bpa function μ for h: $\mu(\mathbf{W}_h) = 1$, and $\mu(\mathbf{a}) = 0$ for all other $\mathbf{a} \in 2^{\mathbf{W}_h}$. We shall call such a bpa function *vacuous*. It represents a state of complete ignorance.

Belief Functions. The information contained in a bpa function can be expressed in several different ways. One way is in terms of the belief function defined as follows.

> ***Definition 2 (Belief function)***. A *belief function β for h* corresponding to bpa function μ for h is a function $\beta: 2^{\mathbf{W}_h} \to [0, 1]$ such that
> $$\beta(\mathbf{a}) = \Sigma\{\mu(\mathbf{b}) \mid \mathbf{b} \subseteq \mathbf{a}\} \tag{1}$$
> for each $\mathbf{a} \in 2^{\mathbf{W}_h}$.

Intuitively, whereas the quantity $\mu(\mathbf{a})$ measures the belief that one commits exactly to $\mathbf{a}$, the quantity $\beta(\mathbf{a})$ measures the total belief that one commits to $\mathbf{a}$.

It is shown in Shafer (1976, p. 39) that we can recover μ from β:

$$\mu(\mathbf{a}) = \sum\{(-1)^{|a-b|}\beta(\mathbf{b}) \mid \mathbf{b} \subseteq \mathbf{a}\},$$

where $|\mathbf{a}-\mathbf{b}|$ denotes the number of elements in the set $\mathbf{a}-\mathbf{b}$.

Plausibility Function. Another way of expressing the information contained in a bpa function μ is in terms of the *plausibility function* π defined as follows.

> ***Definition 3 (Plausibility function)***. A *plausibility function π for h* corresponding to bpa function μ for h is a function $\pi: 2^{\mathbf{W}_h} \to [0, 1]$ such that
> $$\pi(\mathbf{a}) = \Sigma\{\mu(\mathbf{b}) \mid \mathbf{b} \cap \mathbf{a} \neq \emptyset\} \tag{2}$$
> for each $\mathbf{a} \in 2^{\mathbf{W}_h}$.

Since the probability mass $\mu(\mathbf{b})$ can move into $\mathbf{a}$ if and only if $\mathbf{b} \cap \mathbf{a} \neq \emptyset$, $\pi(\mathbf{a})$ measures the total probability mass that can move into $\mathbf{a}$, i.e., $\pi(\mathbf{a})$ measures the extent to which one finds $\mathbf{a}$ plausible. Suppose $\sim\mathbf{a}$ denotes the complement of $\mathbf{a}$; $\sim\mathbf{a} = \mathbf{W}_h - \mathbf{a}$. Since $\{\mathbf{b} \in 2^{\mathbf{W}_h} \mid \mathbf{b} \subseteq \mathbf{a}\} \cup \{\mathbf{b} \in 2^{\mathbf{W}_h} \mid \mathbf{b} \cap \sim\mathbf{a} \neq \emptyset\} = 2^{\mathbf{W}_h}$, $\beta(\mathbf{a}) = 1 - \pi(\sim\mathbf{a})$, and $\pi(\mathbf{a}) = 1 - \beta(\sim\mathbf{a})$. Thus $\pi(\mathbf{a})$ also measures the extent to which the given evidence fails to refute $\mathbf{a}$.

Since $\{\mathbf{b} \in 2^{\mathbf{W}_h} \mid \mathbf{b} \subseteq \mathbf{a}\} \subseteq \{\mathbf{b} \in 2^{\mathbf{W}_h} \mid \mathbf{b} \cap \mathbf{a} \neq \emptyset\}$, $\beta(\mathbf{a}) \leq \pi(\mathbf{a})$ for every subset $\mathbf{a}$ of $\mathbf{W}_h$. Both β and π are monotone: $\beta(\mathbf{a}) \leq \beta(\mathbf{b})$ and $\pi(\mathbf{a}) \leq \pi(\mathbf{b})$ whenever $\mathbf{a} \subseteq \mathbf{b}$.

Although belief functions and plausibility functions are easier to interpret than bpa functions, bpa functions are easier to work with mathematically. Bpa functions correspond to nonzero valuations in VBS.

Before we can define combination and marginalization for bpa functions, we need the concepts of projection of configurations, and projection and extension of subsets of configurations.

Projection of configurations. Projection of configurations simply means dropping extra coordinates; if (w,x,y,z) is a configuration of {W,X,Y,Z}, for example, then the projection of (w,x,y,z) to {W,X} is simply (w,x), which is a configuration of {W,X}. If g and h are sets of variables, $h \subseteq g$, and $\mathbf{x}$ is a configuration of g, then $\mathbf{x}^{\downarrow h}$ denotes the projection of $\mathbf{x}$ to h.

Projection and Extension of Sets of Configurations. If g and h are sets of variables, $h \subseteq g$, and $\mathbf{g}$ is a nonempty subset of $\mathbf{W}_g$, then the *projection of* $\mathbf{g}$ *to h*, denoted by $\mathbf{g}^{\downarrow h}$, is the subset of $\mathbf{W}_h$ given by $\mathbf{g}^{\downarrow h} = \{\mathbf{x}^{\downarrow h} \mid \mathbf{x} \in \mathbf{g}\}$. For example, If $\mathbf{a}$ is a subset of $\mathbf{W}_{\{W,X,Y,Z\}}$, then the projection of $\mathbf{a}$ to {X,Y} consists of the elements of $\mathbf{W}_{\{X,Y\}}$ which can be obtained by projecting elements of $\mathbf{a}$ to $\mathbf{W}_{\{X,Y\}}$.

By extension of a subset of a frame to a subset of a larger frame, we mean a cylinder set extension. If g and h are sets of variables, h is a proper subset of g, and $\mathbf{h}$ is a nonempty subset of $\mathbf{W}_h$, then the *extension of* $\mathbf{h}$ *to g* is $\mathbf{h} \times \mathbf{W}_{g\text{-}h}$. Let $\mathbf{h}^{\uparrow g}$ denote the extension of $\mathbf{h}$ to g. For example, if $\mathbf{a}$ is a nonempty subset of $\mathbf{W}_{\{W,X\}}$, then the extension of $\mathbf{a}$ to {W,X,Y,Z} is $\mathbf{a} \times \mathbf{W}_{\{Y,Z\}}$.

Marginalization. Suppose μ is a bpa function for h. Suppose $X \in h$. We may be interested only in propositions about variables in h–{X}. In this case, we would like to marginalize μ to h–{X}.

> ***Definition 4 (Marginalization).*** Suppose μ is a bpa function for h, and suppose $X \in h$. The *marginal of* μ *for h–{X}*, denoted by $\mu^{\downarrow(h-\{X\})}$, is the bpa function for h–{X} defined as follows:
> $$\mu^{\downarrow(h-\{X\})}(\mathbf{a}) = \Sigma\{\mu(\mathbf{b}) \mid \mathbf{b} \subseteq \mathbf{W}_h \text{ such that } \mathbf{b}^{\downarrow(h-\{X\})} = \mathbf{a}\}$$
> for all nonempty subsets $\mathbf{a}$ of $\mathbf{W}_{h-\{X\}}$.

Theorem 2 states that the marginalization operation for belief function satisfies Axiom A2 stated in section 2.2.

> ***Theorem 2***. Suppose μ is a bpa function for h, and suppose $X_1, X_2 \in h$. Then
> $$(\mu^{\downarrow(h-\{X_1\})})^{\downarrow(h-\{X_1,X_2\})} = (\mu^{\downarrow(h-\{X_2\})})^{\downarrow(h-\{X_1,X_2\})}.$$

Example 2 (Bpa functions and marginalization). We would like to determine whether a stranger (about who we know nothing about) is a pacifist or not depending on whether he is a Republican or not and whether he is a Quaker or not. Consider three variables R, Q and P. R has two configurations: r (for Republican), ~r (not Republican); Q has two configurations: q (Quaker) and ~q (not Quaker); and P has two configurations: p (pacifist) and ~p (not pacifist). Our knowledge that most (at least 90 percent) Republicans are not pacifists and that most (at least 99 percent) Quakers are pacifists is represented by the bpa function μ for {R,Q,P} shown in Table 1 (the construction of this bpa function will be explained later in this section—see Example 3).

Note that the marginal of μ for R is the vacuous bpa function for R, i.e., $\mu^{\downarrow\{R\}}(\{r, \sim r\}) = 1$. Thus, we have no knowledge whether the stranger is a republican or not. Similarly, notice that the marginals of μ for {Q} and {P} are also vacuous. The marginal of μ for {R, P} is as follows:

$$\mu^{\downarrow\{R,P\}}(\{(r,\sim p), (\sim r,p), (\sim r,\sim p)\}) = 0.90,\ \mu^{\downarrow\{R,P\}}(\{(r,p), (r,\sim p), (\sim r,p), (\sim r,\sim p)\}) = 0.10$$

Thus the plausibility of a Republican pacifist is only 0.10. Similarly, notice that the marginal of μ for {Q, P} is as follows:

$$\mu^{\downarrow\{Q,P\}}(\{(q,p), (\sim q,p), (\sim q,\sim p)\}) = 0.99,$$
$$\mu^{\downarrow\{Q,P\}}(\{(q,p), (q,\sim p), (\sim q,p), (\sim q,\sim p)\}) = 0.01.$$

Table I.

A bpa function μ for {R,Q,P}.

$2^{\mathbf{W}_{\{R,Q,P\}}}$	μ
{(r,~q,~p), (~r,q,p), (~r,~q,p), (~r,~q,~p)}	0.891
{(r,q,~p), (r,~q,~p), (~r,q,p), (~r,~q,p), (~r,q,~p), (~r,~q,~p)}	0.009
{(r,q,p), (r,~q,p), (r,~q,~p), (~r,q,p), (~r,~q,p), (~r,~q,~p)}	0.099
{(r,q,p), (r,q,~p), (r,~q,p), (r,~q,~p), (~r,q,p), (~r,q,~p), (~r,~q,p), (~r,~q,~p)}	0.001

Thus the plausibility of a non-pacifist Quaker is only 0.01. Finally note that the marginal of μ for {R, Q} is as follows:

$$\mu^{\downarrow\{R,Q\}}(\{(r,\sim q), (\sim r,q), (\sim r,\sim q)\}) = 0.891,$$
$$\mu^{\downarrow\{R,Q\}}(\{(r,q), (r,\sim q), (\sim r,q), (\sim r,\sim q)\}) = 0.109$$

Thus the plausibility of a Republican Quaker is only 0.109. ■

Next, we state a rule for combining bpa functions. This rule is called Dempster's rule (Dempster, 1967, pp. 335–337).

Definition 5. Consider two bpa functions μ_1 and μ_2 for h_1 and h_2, respectively. Suppose $K = \Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow(h_1\cup h_2)})\cap(\mathbf{b}^{\uparrow(h_1\cup h_2)}) \neq \varnothing\}$. The *combination of μ_1 and μ_2*, denoted by $\mu_1\otimes\mu_2$, is the function for $h_1\cup h_2$ given by

$$(\mu_1\otimes\mu_2)(\mathbf{c}) = \begin{cases} K^{-1}\Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow(h_1\cup h_2)})\cap(\mathbf{b}^{\uparrow(h_1\cup h_2)}) = \mathbf{c}\} & \text{if } K \neq 0 \\ 0 & \text{if } K = 0 \end{cases}$$

for all nonempty $\mathbf{c}\subseteq\mathbf{W}_{h_1\cup h_2}$.

If K = 0, then $\mu_1\otimes\mu_2$ is not a bpa function. In this case, $\mu_1\otimes\mu_2$ is not a nonzero valuation. This means that the knowledge in μ_1 and μ_2 are inconsistent. If $K \neq 0$, then K is a normalization constant that ensures that $\mu_1\otimes\mu_2$ is a bpa function.

Example 3 (Dempster's rule of combination). Consider two pieces of independent knowledge as follows:

1. Most Republicans (at least 90 percent) are not pacifists.
2. Most Quakers (at least 99 percent) are pacifists.

If μ_1 is a bpa function representation of the first piece of knowledge, and μ_2 is a bpa function representation of the second piece of knowledge, then $\mu_1\otimes\mu_2$ represents the aggregation of these two pieces of knowledge. The first piece of knowledge can be represented by the bpa function μ_1 for {R,P} as follows:

$$\mu_1(\mathbf{W}_{\{R,P\}}-\{(r,p)\}) = 0.90,\ \mu_1(\mathbf{W}_{\{R,P\}}) = 0.10$$

(i.e., the plausibility of pacifist Republicans is only 0.10). Similarly, the second piece of knowledge can be represented by the bpa function μ_2 for {Q,P} as follows:

$$\mu_2(\mathbf{W}_{\{Q,P\}}-\{(q,\sim p)\}) = 0.99,\ \mu_2(\mathbf{W}_{\{Q,P\}}) = 0.01$$

(i.e., the plausibility of non-pacifist Quakers is only 0.01). Then the bpa function $\mu_1 \otimes \mu_2 = \mu$, say, shown in Table 1, represents the aggregate knowledge. ■

Theorem 3 (Shafer, 1976). Dempster's rule of combination has the following properties:
C1. (Commutativity) $\mu_1 \otimes \mu_2 = \mu_2 \otimes \mu_1$.
C2. (Associativity) $(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3)$.
C3. If μ_1 is vacuous, then $\mu_1 \otimes \mu_2 = \mu_2$.
C4. In general, $\mu_1 \otimes \mu_1 \neq \mu_1$. The bpa function $\mu_1 \otimes \mu_1$ believes in the same propositions as μ_1, but it will do so with twice the degree, as it were.

The bpa function $\mu_1 \otimes \mu_2$ represents aggregation of knowledge contained in bpa functions μ_1 and μ_2 only when the bpa functions μ_1 and μ_2 are independent, i.e., the bpa functions μ_1 and μ_2 do not contain some common knowledge. Property C4 tells us that double-counting of knowledge may lead to erroneous information. Thus it is important to ensure that the bpa functions being combined are independent.

We have already shown that axioms A1 and A2 are valid for bpa functions. Theorem 4 below states that axiom A3 is also satisfied.

Theorem 4. Suppose μ_1 and μ_2 are bpa functions for h_1 and h_2, respectively. Suppose $X \notin h_1$, and suppose $X \in h_2$. Then $(\mu_1 \otimes \mu_2)^{\downarrow((h_1 \cup h_2)-\{X\})} = \mu_1 \otimes (\mu_2{}^{\downarrow(h_2-\{X\})})$.

Since all three axioms required for local computation of marginals are satisfied, the fusion algorithm described in section 2.3 can be used for reasoning from knowledge expressed as bpa functions. The next section describes a small example to illustrate the use of the fusion algorithm to find marginals.

4. AN EXAMPLE

In this section, we describe an example in complete detail to illustrate the use of DS belief-function theory in managing uncertainty in expert systems.

Is Dick a Pacifist? Consider the following independent items of evidence. Most (at least 90 percent) Republicans are not pacifists. Most Quakers (at least 99 percent) are pacifists. Dick is a Republican (and we are more than 99.9 percent certain of this). Dick is a Quaker (and we are more than 99.9 percent certain of this). Is Dick a pacifist?

We will model the four items of evidence as bpa potentials as follows. Consider three variables R, Q, and P, each with two configurations in their respective frames. R=r represents the proposition that Dick is a Republican, and R=~r represents the proposition that Dick is not a Republican. Similarly for Q and P. The four items of evidence are represented by bpa potentials μ_1 for {R,P}, μ_2 for {Q, P}, μ_3 for {R}, and μ_4 for {Q}, respectively, as displayed in Table 2. Figure 3 shows the valuation network for this example.

If we apply the fusion algorithm using deletion sequence RQ, we get $(\mu_1 \otimes \mu_2 \otimes \mu_3 \otimes \mu_4)^{\downarrow\{P\}} = (\mu_1 \otimes \mu_3)^{\downarrow\{P\}} \otimes (\mu_2 \otimes \mu_4)^{\downarrow\{P\}}$. The details of the computations are shown in Table 3. The degree of belief that Dick is a pacifist is 0.9008. Notice that we can avoid the normalization operation in Dempster's rule in the intermediate stages and do it just once at the very end (the last combination operation). ■

The use of local computation in computing marginals of the joint bpa function has been widely studied. Some of the influential works in this area are those by Shenoy and Shafer (1986, 1990), Kong (1986), Shafer, Shenoy and Mellouli (1987), Mellouli (1987), and Dempster and Kong (1988). The fusion algorithm applied to the case of belief function theory is an abstraction of the methods described in these papers.

5. CONCLUSIONS

We have described the framework of valuation-based systems (VBS), we have described three axioms that permit the use of local computation in computing marginals of the joint valuation, and we have described a fusion algorithm for computing marginals using local computation. Next, we have described the essential features of DS theory of belief functions and described how this theory fits in the framework of VBS. Elsewhere, we have described how VBS serve as a language for constructing expert systems (Shenoy, 1989). Thus, the correspondence between the theory of belief functions and the framework of VBS should facilitate the use of DS belief-function theory in expert systems.

Table II.
The bpa potentials μ_1, μ_2, μ_3, and μ_4 in *Is Dick a Pacifist?* example.

$2^{\mathbf{W}_{\{R, P\}}}$	μ_1
{(r,~p), (~r,p), (~r,~p)}	.90
{(r,p), (r,~p), (~r,p), (~r,~p)}	.10

$2^{\mathbf{W}_R}$	μ_3
{r}	.999
{r, ~r}	.001

$2^{\mathbf{W}_{\{Q, P\}}}$	μ_2
{(q,p), (~q,p), (~q,~p)}	.99
{(q,p), (q,~p), (~q,p), (~q,~p)}	.01

$2^{\mathbf{W}_Q}$	μ_4
{q}	.999
{q, ~q}	.001

Figure 3. The valuation network for the *Is Dick a Pacifist*? example.

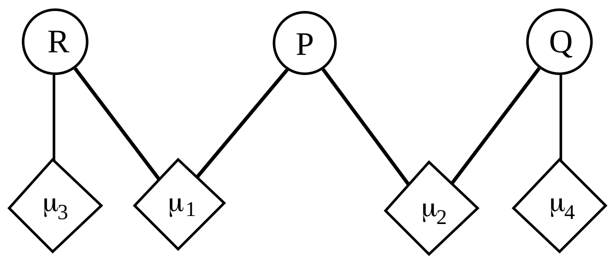


Table III.
The computation of $(\mu_1 \otimes \mu_3)^{\downarrow\{P\}} \otimes (\mu_2 \otimes \mu_4)^{\downarrow\{P\}}$.

$2^{\mathbf{W}_{\{R,P\}}}$	$\mu_1 \otimes \mu_3$
{(r,~p)}	.8991
{(r,p), (r,~p)}	.0999
{(r,~p), (~r,p), (~r,~p)}	.0009
{(r,p), (r,~p), (~r,p), (~r,~p)}	.0001

$2^{\mathbf{W}_{\{Q,P\}}}$	$\mu_2 \otimes \mu_4$
{(q,p)}	.98901
{(q,p), (q,~p)}	.00999
{(q,p), (~q,p), (~q,~p)}	.00099
{(q,p), (q,~p), (~q,p), (~q,~p)}	.00001

$2^{\mathbf{W}_P}$	$(\mu_1 \otimes \mu_3)^{\downarrow\{P\}}$	$(\mu_2 \otimes \mu_4)^{\downarrow\{P\}}$	$(\mu_1 \otimes \mu_3)^{\downarrow\{P\}} \otimes (\mu_2 \otimes \mu_4)^{\downarrow\{P\}}$
{p}	.0000	.98901	.9008
{~p}	.8991	.00000	.0892
{p, ~p}	.1009	.01099	.0100

6. PROOFS

In this section, we give proofs for all results in the paper. First we state and prove a lemma needed to prove Theorem 1.

> **Lemma 1**. Suppose $\{\{\sigma_1, ..., \sigma_m\}, \otimes, \downarrow\}$ is a VBS where σ_i is a valuation for s_i, and suppose $\otimes$ and $\downarrow$ satisfy axioms A1–A3. Let **X** denote $s_1\cup...\cup s_m$. Suppose $X\in\mathbf{X}$. Then
> $$(\sigma_1\otimes...\otimes\sigma_m\})^{\downarrow(x-\{X\})} = \otimes Fus_X\{\sigma_1, ..., \sigma_m\}.$$

Proof of Lemma 1. Suppose σ_i is a valuation for s_i, i = 1, ..., m. Let $s = \cup\{s_i \mid X\in s_i\}$, and let $r = \cup\{s_i \mid X\notin s_i\}$. Let $\rho = \otimes\{\sigma_i \mid X\notin s_i\}$, and $\sigma = \otimes\{\sigma_i \mid X\in s_i\}$. Note that $X\in s$, and $X\notin g$. Then

$$\begin{aligned}(\sigma_1\otimes...\otimes\sigma_m)^{\downarrow(x-\{X\})} &= (\rho\otimes\sigma)^{\downarrow((r\cup s)-\{X\})}\\ &= \rho\otimes(\sigma^{\downarrow(s-\{X\})}) \qquad \text{(using axiom A3)}\\ &= (\otimes\{\sigma_i \mid X\notin s_i\})\otimes(\sigma^{\downarrow(s-\{X\})})\\ &= \otimes Fus_X\{\sigma_1, ..., \sigma_m\}.\end{aligned}$$ ■

Proof of Theorem 1. By axiom A2, $(\sigma_1\otimes...\otimes\sigma_m)^{\downarrow\{X\}}$ is obtained by sequentially marginalizing all variables but X from the joint valuation. A proof of this theorem is obtained by repeatedly applying the result of Lemma 1. At each step, we delete a variable and fuse the set of all valuations with respect to this variable. Using Lemma 1, after fusion with respect to X_1, the combination of all valuations in the resulting VBS is equal to $(\sigma_1\otimes...\otimes\sigma_m)^{\downarrow(x-\{X_1\})}$. Again, using Lemma 1, after fusion with respect to X_2, the combination of all valuations in the resulting VBS is equal to $(\sigma_1\otimes...\otimes\sigma_m)^{\downarrow(x-\{X_1,X_2\})}$. And so on. When all variables but X have been deleted, we have the result. ■

Proof of Theorem 2. Suppose μ is a bpa function for h, and suppose $X_1, X_2 \in h$. Then

$$\begin{aligned}&(\mu^{\downarrow(h-\{X_1\})})^{\downarrow(h-\{X_1,X_2\})}(\mathbf{a})\\ &= \Sigma\{\mu^{\downarrow(h-\{X_1\})}(\mathbf{b}) \mid \mathbf{b} \subseteq \mathbf{W}_{h-\{X_1\}} \text{ such that } \mathbf{b}^{\downarrow(h-\{X_1,X_2\})} = \mathbf{a}\}\\ &= \Sigma\Big\{\Sigma\{\mu(\mathbf{c}) \mid \mathbf{c}\subseteq\mathbf{W}_h \text{ such that } \mathbf{c}^{\downarrow(h-\{X_1\})} = \mathbf{b}\} \Big| \mathbf{b}\subseteq\mathbf{W}_{h-\{X_1\}} \text{ such that } \mathbf{b}^{\downarrow(h-\{X_1,X_2\})} = \mathbf{a}\Big\}\\ &= \Sigma\{\mu(\mathbf{c}) \mid \mathbf{c}\subseteq\mathbf{W}_h \text{ such that } (\mathbf{c}^{\downarrow(h-\{X_1\})})^{\downarrow(h-\{X_1,X_2\})} = \mathbf{a}\}\\ &= \Sigma\{\mu(\mathbf{c}) \mid \mathbf{c}\subseteq\mathbf{W}_h \text{ such that } \mathbf{c}^{\downarrow(h-\{X_1,X_2\})} = \mathbf{a}\}.\end{aligned}$$

Similarly, we can show that

$$(\mu^{\downarrow(h-\{X_2\})})^{\downarrow(h-\{X_1,X_2\})} = \Sigma\{\mu(\mathbf{c}) \mid \mathbf{c}\subseteq\mathbf{W}_h \text{ such that } \mathbf{c}^{\downarrow(h-\{X_1,X_2\})} = \mathbf{a}\}.$$ ■

Proof of Theorem 3. All four properties follow trivially from the definition of Dempster's rule of combination. ■

Proof of Theorem 4. First, note that the normalization factor in the combination on the left-hand side, say K_1, is the same as the normalization factor in the combination on the right-hand side, say K_2, i.e., $K_1 = K_2$, as shown below.

$$
\begin{aligned}
K_2 &= \Sigma\{\mu_1(\mathbf{a})\mu_2^{\downarrow(h_2-\{X\})}(\mathbf{e}) \mid (\mathbf{a}^{\uparrow(h_1\cup(h_2-\{X\}))})\cap(\mathbf{e}^{\uparrow(h_1\cup(h_2-\{X\}))}) \neq \varnothing\} \\
&= \sum\Big\{\mu_1(\mathbf{a})\Sigma\{\mu_2(\mathbf{b}) \mid \mathbf{b}^{\downarrow(h_2-\{X\})} = \mathbf{e}\} \mid (\mathbf{a}^{\uparrow(h_1\cup(h_2-\{X\}))})\cap(\mathbf{e}^{\uparrow(h_1\cup(h_2-\{X\}))}) \neq \varnothing\Big\} \\
&= \Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow(h_1\cup(h_2-\{X\}))})\cap((\mathbf{b}^{\downarrow(h_2-\{X\})})^{\uparrow(h_1\cup(h_2-\{X\}))}) \neq \varnothing\} \\
&= \Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow(h_1\cup(h_2-\{X\}))})\cap(\mathbf{b}^{\uparrow(h_1\cup(h_2-\{X\}))}) \neq \varnothing\} \\
&= \Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow(h_1\cup h_2)})\cap(\mathbf{b}^{\uparrow(h_1\cup h_2)}) \neq \varnothing\} \\
&= K_1 = K, \text{ say.}
\end{aligned}
$$

Next,

$$
\begin{aligned}
&(\mu_1\otimes\mu_2)^{\downarrow((h_1\cup h_2)-\{X\})}(\mathbf{d}) = \\
&= \Sigma\{(\mu_1\otimes\mu_2)(\mathbf{c}) \mid \mathbf{c}^{\downarrow((h_1\cup h_2)-\{X\})} = \mathbf{d}\} \\
&= \sum\Big\{K^{-1}\Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow(h_1\cup h_2)})\cap(\mathbf{b}^{\uparrow(h_1\cup h_2)}) = \mathbf{c}\} \Big| \mathbf{c}^{\downarrow((h_1\cup h_2)-\{X\})} = \mathbf{d}\Big\} \\
&= K^{-1}\Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid ((\mathbf{a}^{\uparrow(h_1\cup h_2)})\cap(\mathbf{b}^{\uparrow(h_1\cup h_2)}))^{\downarrow((h_1\cup h_2)-\{X\})} = \mathbf{d}\} \\
&= K^{-1}\Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow((h_1\cup h_2)-\{X\})})\cap(\mathbf{b}^{\uparrow((h_1\cup h_2)-\{X\})}) = \mathbf{d}\} \\
&= K^{-1}\Sigma\{\mu_1(\mathbf{a})\mu_2(\mathbf{b}) \mid (\mathbf{a}^{\uparrow(h_1\cup(h_2-\{X\}))})\cap((\mathbf{b}^{\downarrow(h_2-\{X\})})^{\uparrow(h_1\cup(h_2-\{X\}))}) = \mathbf{d}\} \\
&= K^{-1}\sum\Big\{\mu_1(\mathbf{a})\Sigma\{\mu_2(\mathbf{b}) \mid \mathbf{b}^{\downarrow(h_2-\{X\})} = \mathbf{e}\}\Big|(\mathbf{a}^{\uparrow(h_1\cup(h_2-\{X\}))})\cap\mathbf{e}^{\uparrow(h_1\cup(h_2-\{X\}))}) = \mathbf{d}\Big\} \\
&= K^{-1}\sum\Big\{\mu_1(\mathbf{a})\mu_2^{\downarrow(h_2-\{X\})}(\mathbf{e})\Big|(\mathbf{a}^{\uparrow(h_1\cup(h_2-\{X\}))})\cap(\mathbf{e}^{\uparrow(h_1\cup(h_2-\{X\}))}) = \mathbf{d}\Big\} \\
&= (\mu_1\otimes(\mu_2^{\downarrow(h_2-\{X\})}))(\mathbf{d}). \qquad \blacksquare
\end{aligned}
$$

7. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grant IRI-8902444. I am grateful to Leen-Kiat Soh for comments.

8. REFERENCES

Arnborg, S., D. G. Corneil, and A. Proskurowski (1987), "Complexity of finding embeddings in a k-tree," *SIAM Journal of Algebraic and Discrete Methods*, **8**, 277–284.

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

Dempster, A. P. and A. Kong (1988), "Uncertain evidence and artificial analysis," *Journal of Statistical Planning and Inference*, **20**, 355–368.

Kjærulff, U. (1990), "Triangulation of graphs—Algorithms giving small total state space," Technical report R90-09, Institute for Electronic Systems, University of Aalborg, Denmark.

Kong, A. (1986), “Multivariate belief functions and graphical models,” Ph.D. dissertation, Department of Statistics, Harvard University, Cambridge, MA.

Mellouli, K. (1987), “On the propagation of beliefs in networks using the Dempster-Shafer's theory of evidence,” Ph.D. thesis, School of Business, University of Kansas, Lawrence, KS.

Saffiotti, A. and E. Umkehrer (1991), “Pulcinella: A general tool for propagating uncertainty in valuation networks,” in D’Ambrosio, B., P. Smets, and P. P. Bonissone (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Seventh Conference*, 323–331, Morgan Kaufmann, San Mateo, CA.

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

Shafer, G., P. P. Shenoy, and K. Mellouli (1987), “Propagating belief functions in qualitative Markov trees,” *International Journal of Approximate Reasoning*, **1**(4), 349–400.

Shenoy, P. P. (1989), “A valuation-based language for expert systems,” *International Journal of Approximate Reasoning*, **3**(5), 383–411, 1989.

Shenoy, P. P. (1990a), “Valuation-based systems for propositional logic,” in Ras, Z. W., M. Zemankova, and M. L. Emrich (eds.), *Methodologies for Intelligent Systems*, **5**, 305–312, North-Holland, Amsterdam.

Shenoy, P. P. (1990b), “Consistency in valuation-based systems,” Working Paper No. 216, School of Business, Lawrence, KS.

Shenoy, P. P. (1990c), “Valuation-based systems for Bayesian decision analysis,” Working Paper No. 220, School of Business, University of Kansas, Lawrence, KS. Appeared in 1992 in *Operations Research*, **40**(3), 463–484.

Shenoy, P. P. (1990d), “A new method for representing and solving Bayesian decision problems,” Working Paper No. 223, School of Business, University of Kansas, Lawrence, KS. To appear in 1992 in Hand, D. J. (ed.), *Artificial Intelligence and Statistics III*, Chapman & Hall, London.

Shenoy, P. P. (1991a), “On Spohn’s rule for revision of beliefs,” *International Journal of Approximate Reasoning*, **5**(2), 149–181.

Shenoy, P. P. (1991b), “Valuation-based systems for discrete optimization,” in Bonissone, P. P., M. Henrion, L. N. Kanal, and J. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, **6**, 385–400. North-Holland, Amsterdam.

Shenoy, P. P. (1991c), “Valuation-based systems: A framework for managing uncertainty in expert systems,” Working Paper No. 226, School of Business, University of Kansas, Lawrence, KS. Appeared in 1992 in Zadeh, L. A. and J. Kacprzyk (eds.), *Fuzzy Logic for the Management of Uncertainty*, 83–104, John Wiley & Sons, New York, NY.

Shenoy, P. P. (1991d), “Using possibility theory in expert systems,” Working Paper No. 233, School of Business, University of Kansas, Lawrence, KS. To appear in 1992 in *Fuzzy Sets and Systems*, **51**(3).

Shenoy, P. P. (1991e), “Independence in valuation-based systems,” Working Paper No. 236, School of Business, University of Kansas, Lawrence, KS.

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

Shenoy, P. P. and G. Shafer (1988), “Constraint propagation,” Working Paper No. 208, School of Business, University of Kansas, Lawrence, KS.

Shenoy, P. P. and G. Shafer (1990), “Axioms for probability and belief-function propagation,” in Shachter, R. D., T. S. Levitt, J. F. Lemmer and L. N. Kanal (eds.), *Uncertainty in Artificial Intelligence*, **4**, 169–198, North-Holland, Amsterdam. Reprinted in Shafer, G. and J. Pearl, eds. (1990), *Readings in Uncertain Reasoning*, 575–610, Morgan Kaufmann, San Mateo, CA.

Spohn, W. (1988), “Ordinal conditional functions: A dynamic theory of epistemic states,” in Harper, W. L. and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, **2**, 105–134, D. Reidel Publishing Company, Dordrecht.

Zadeh, L. A. (1979), "A theory of approximate reasoning," in Ayes, J. E., D. Mitchie, and L. I. Mikulich (eds.), *Machine Intelligence*, **9**, 149–194, Ellis Horwood, Chichester, U.K.

Zhang, L. (1988), "Studies on finding hypertree covers of hypergraphs," Working Paper No. 198, School of Business, University of Kansas, Lawrence, KS.

SELECTED WORKING PAPERS

Unpublished working papers are available from the respective authors at:

School of Business
University of Kansas
Summerfield Hall
Lawrence, KS 66045-2003 USA

No. 184. "Propagating Belief Functions with Local Computations," Prakash P. Shenoy and Glenn Shafer, February 1986. Appeared in *IEEE Expert*, **1**(3), 1986, 43–52.

No. 190. "Propagating Belief Functions in Qualitative Markov Trees," Glenn Shafer, Prakash P. Shenoy, and Khaled Mellouli, June 1987. Appeared in *International Journal of Approximate Reasoning*, **1**(4), 1987, 349–400.

No. 197. "AUDITOR'S ASSISTANT: A Knowledge Engineering Tool for Audit Decisions," Glenn Shafer, Prakash P. Shenoy, and Rajendra Srivastava, April 1988. Appeared in *Auditing Symposium IX: Proceedings of 1988 Touche Ross/University of Kansas Symposium on Auditing Problems*, 61–84, School of Business, University of Kansas, Lawrence, KS.

No. 199. "An Axiomatic Framework for Bayesian and Belief-Function Propagation," Prakash P. Shenoy and Glenn Shafer, July 1988. Appeared in *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN, 307–314.

No. 200. "Probability Propagation," Glenn Shafer and Prakash P. Shenoy, August 1988. Appeared in *Annals of Mathematics and Artificial Intelligence*, **2**(1–4), 1990, 327–352.

No. 203. "A Valuation-Based Language for Expert Systems," Prakash P. Shenoy, August 1988. Appeared in *International Journal of Approximate Reasoning*, **3**(5), 1989, 383-411.

No. 208. "Constraint Propagation," Prakash P. Shenoy and Glenn Shafer, November 1988.

No. 209. "Axioms for Probability and Belief-Function Propagation," Prakash P. Shenoy and Glenn Shafer, November 1988. Appeared in Shachter, R. D., M. Henrion, L. N. Kanal, and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, **4**, 1990, 169–198. Reprinted in Shafer, G. and J. Pearl (eds.), *Readings in Uncertain Reasoning*, 1990, 575–610, Morgan Kaufmann, San Mateo, CA.

No. 211. "MacEvidence: A Visual Evidential Language for Knowledge-Based Systems," Yen-Teh Hsia and Prakash P. Shenoy, March 1989. An 8-page summary of this paper appeared as "An evidential language for expert systems," in Ras, Z. W. (ed.), *Methodologies for Intelligent Systems*, **4**, 1989, 9–16, North-Holland, Amsterdam.

No. 213. "On Spohn's Rule for Revision of Beliefs," Prakash P. Shenoy, July 1989. Appeared in *International Journal of Approximate Reasoning*, **5**(2), 1991, 149–181.

No. 216. "Consistency in Valuation-Based Systems," Prakash P. Shenoy, February 1990, revised May 1991. To appear in *ORSA Journal on Computing*, Vol. 6, No. 3, Summer 1994.

No. 218. "Perspectives on the Theory and Applications of Belief Functions," Glenn Shafer, April 1990. Appeared in *International Journal of Approximate Reasoning*, **4**(5–6), 1990, 323–362.

No. 220. "Valuation-Based Systems for Bayesian Decision Analysis," Prakash P. Shenoy, April 1990, revised May 1991. Appeared in *Operations Research*, **40**(3), 1992, 463–484.

No. 221. "Valuation-Based Systems for Discrete Optimization," Prakash P. Shenoy, June 1990. Appeared in Bonissone, P. P., M. Henrion, L. N. Kanal, and J. F. Lemmer, eds., *Uncertainty in Artificial Intelligence*, **6**, 1991, 385–400, North-Holland, Amsterdam.

No. 223. "A New Method for Representing and Solving Bayesian Decision Problems," Prakash P. Shenoy, September 1990, revised February 1991. Appeared in: Hand, D. J. (ed.), *Artificial*

Intelligence Frontiers in Statistics: AI and Statistics III, 1993, 119–138, Chapman & Hall, London, England.

No. 226. "Valuation-Based Systems: A Framework for Managing Uncertainty in Expert Systems," Prakash P. Shenoy, March, 1991. Revised July 1991. Appeared in: Zadeh, L. A. and J. Kacprzyk (eds.), *Fuzzy Logic for the Management of Uncertainty*, 1992, 83–104, John Wiley and Sons, New York, NY.

No. 227. "Valuation Networks, Decision Trees, and Influence Diagrams: A Comparison," Prakash P. Shenoy, June 1991. Revised June 1992. A 12-page summary appeared in: Bouchon-Meunier, B., L. Valverde, and R. R. Yager (eds.), *Uncertainty in Intelligent Systems*, 1993, 3–14, North-Holland, Amsterdam. A complete version will appear as: "A Comparison of Graphical Techniques for Decision Analysis" in *European Journal of Operational Research*, Vol. 78, No. 1, October 1994.

No. 233. "Using Possibility Theory in Expert Systems," Prakash P. Shenoy, September 1991. Appeared in *Fuzzy Sets and Systems*, **52**(2), 1992, 129–142.

No. 234. "Using Dempster-Shafer's Belief-Function Theory in Expert Systems," Prakash P. Shenoy, September 1991. Appeared in: Yager, R. R., M. Federizzi, and J. Kacprzyk (eds.), *Advances in the Dempster-Shafer Theory of Evidence*, 1994, 395–414, John Wiley & Sons, New York, NY.

No. 236. "Conditional Independence in Valuation-Based Systems," Prakash P. Shenoy, September 1991. Revised November 1991. Revised September 1992. Appeared in *International Journal of Approximate Reasoning*, Vol. 10, No. 3, April 1994, 203–234.

No. 237. "A Learning Paradigm for Neural Network Architecture," Kenneth O. Cogger, June 1992.

No. 238. "Valuation Networks and Conditional Independence," Prakash P. Shenoy, September 1992. Revised January 1993. A 9-page summary appeared in Heckerman, D. and A. Mamdani (eds.), *Uncertainty in Artificial Intelligence*, **9**, 1993, 191–199, Morgan Kaufmann, San Mateo, CA. The complete paper will appear as "Representing Conditional Independence Relations by Valuation Networks" in *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 2, No. 2, June 1994.

No. 239. "Game Trees for Decision Analysis," Prakash P. Shenoy, February 1993. Revised February 1994. An 8-page summary titled "Information Sets in Decision Theory" appeared in Clarke, M., R. Kruse and S. Moral (eds.), *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, Lecture Notes in Computer Science No. 747, 1993, 318–325, Springer-Verlag, Berlin.

No. 241. "A Comparative Analysis of Artificial Neural Networks for Financial Distress Prediction," Kurt M. Fanning and Kenneth O. Cogger, February 1993.

No. 242. "A Theory of Coarse Utility," Liping Liu and Prakash P. Shenoy, February 1993. Revised September 1993.

No. 245. "Modeling Ignorance in Uncertainty Theories," Prakash P. Shenoy, April 1993.

No. 246. "Valuation Network Representation and Solution of Asymmetric Decision Problems," Prakash P. Shenoy, April 1993. Revised November 1993.

No. 247. "Inducing Attitude Formation Models Using TETRAD," Sanjay Mishra and Prakash P. Shenoy, May 1993. Revised October 1993. A 10-page summary titled "Attitude Formation Models: Insights from TETRAD" will appear in Cheeseman, P. and W. Oldford (eds.), *Artificial Intelligence and Statistics*, Lecture Notes in Computer Science, Springer-Verlag, Berlin.

No. 258. "A Note on Kirkwood's Algebraic Method for Decision Problems," Rui Guo and Prakash P. Shenoy, November 1993. Revised May 1994.

No. 261. "A New Pruning Method for Solving Decision Trees and Game Trees," Prakash P. Shenoy, March 1994.