A new method for representing and solving Bayesian decision problems

11

P.P. Shenoy

INTRODUCTION

The main goal of this chapter is to describe a new method for representing and solving Bayesian decision problems. A new representation of a decision problem called a **valuation-based system** is described. A graphical depiction of a valuation-based system is called a **valuation network**. Valuation networks are similar in some respects to influence diagrams. Like influence diagrams, valuation networks are a compact representation emphasizing qualitative features of symmetric decision problems. Also, like influence diagrams, valuation networks allow representation of symmetric decision problems without any preprocessing. But there are some differences. Whereas influence diagrams emphasize conditional independence among random variables, valuation networks emphasize factorizations of joint probability distributions. Also, the representation method of influence diagrams allows only conditional probabilities. While conditional probabilities are readily available in pure causal models, they are not always readily available in other graphical models (see, for example, Darroch *et al.*, 1980; Wermuth and Lauritzen, 1983; Edwards and Kreiner, 1983; Kiiveri *et al.*, 1984; Whittaker, 1990). The representation method of valuation-based systems is more general and allows direct representation of all probability models.

We also describe a new computational method for solving decision problems called a **fusion algorithm**. The fusion algorithm is a hybrid of local computational methods for computation of marginals of joint probability distributions and local computational methods for discrete optimization. Local computational methods for computation of marginals of joint probability distributions have been proposed by, for example, Pearl (1988), Lauritzen and Spiegelhalter (1988), Shafer and Shenoy (1988; 1990), and Jensen *et al.* (1990). Local computational methods for discrete optimization are also called **non-serial dynamic programming** (Bertele and Brioschi, 1972). Viewed abstractly using the framework of valuation-based systems, these two local computational methods are actually similar. Shenoy and Shafer (1990) and Shenoy (1991b) show that the same three axioms justify the use of local computation in both these cases.

Artificial Intelligence Frontiers in Statistics: AI and statistics III. Edited by D.J. Hand. Published in 1993 by Chapman & Hall, London. ISBN 0 412 40710 8

Valuation-based systems are described in Shenoy (1989; 1991c). In valuation-based systems, we represent knowledge by functions called **valuations.** We draw inferences from such systems using two operations called **combination** and **marginalization**. Drawing inferences can be described simply as marginalizing all variables out of the joint valuation. The joint valuation is the result of combining all valuations. The framework of valuation-based systems is powerful enough to include also the Dempster–Shafer theory of belief functions (Shenoy and Shafer, 1986; 1990), Spohn's theory of epistemic beliefs (Shenoy, 1991a; 1991c), possibility theory (Zadeh, 1979; Dubois and Prade, 1990), propositional logic (Shenoy, 1990), and constraint satisfaction problems (Shenoy and Shafer, 1988).

Our method for representing and solving decision problems has many similarities to influence diagram methodology (Howard and Matheson, 1984; Olmsted, 1983; Shachter, 1986; Ezawa, 1986; Tatman, 1986). But there are also many differences both in representation and solution. A detailed comparison of our method with decision tree and influence diagram methods is given in Shenoy (1991d).

The fusion algorithm described in this chapter applies to representations that contain only one utility valuation. The assumption of one utility valuation is similar to the assumption of one value node in influence diagrams (Shachter, 1986). If the utility valuation decomposes multiplicatively into several smaller utility valuations, then the fusion algorithm described in this chapter applies also to this case. But if the utility valuation decomposes additively into several smaller utility valuations, then the fusion algorithm does not apply directly to this case. Before we can apply the fusion algorithm described in this chapter, we have to combine the valuations to obtain one joint utility valuation. Thus the fusion algorithm described in this chapter is unable to take computational advantage of an additive decomposition of the utility valuation. Shenoy (1992) describes a modified fusion algorithm that is capable of taking advantage of an additive decomposition of the utility valuation. The modification involves some divisions.

We describe our new method using a diabetes diagnosis problem. The next section gives a statement of this problem and shows a decision tree representation and solution. The third section describes a valuation-based representation of a decision problem. The fourth describes the process of solving valuation-based systems. The fifth describes a fusion algorithm for solving valuation-based systems using local computation. Finally, the sixth section contains proofs of all results.

A DIABETES DIAGNOSIS PROBLEM

A medical intern is trying to decide a policy for treating patients suspected of suffering from diabetes. The intern first observes whether a patient exhibits two symptoms of diabetes—blue toe and glucose in the urine. After she observes the presence or absence of these symptoms, she then either prescribes a treatment for diabetes or does not.

Table 11.1 shows the intern's utility function. For the population of patients served by the intern, the prior probability of diabetes is 10%. Furthermore, for patients known to suffer from diabetes, 1.4% exhibit blue toe, and 90% exhibit glucose in the urine.

Table 11.1 The intern's utility function for all state–act pairs

		Act	
	Intern's utilities (v)	treat for diabetes (t)	not treat ($\sim t$)
State	has diabetes (d)	10	0
	no diabetes ($\sim d$)	5	10

On the other hand, for patients known not to suffer from diabetes, 0.6% exhibit blue toe, and 1% exhibit glucose in the urine.

Consider three random variables *D*, *B*, and *G* representing diabetes, blue toe, and glucose in the urine, respectively. Each variable has two possible values. $D = d$ will represent the proposition *patient has diabetes*, and $D = \sim d$ will represent the proposition *patient does not have diabetes*; similarly for *B* and *G*. Assume that variables *B* and *G* are conditionally independent given *D*.

Figure 11.1 shows the preprocessing of probabilities in the decision tree. Figure 11.2 shows a decision tree representation and solution of this problem. Thus an optimal strategy for the medical intern is to treat the patient for diabetes if and only if there is glucose in the patient's urine. The expected utility of this strategy is 9.86.

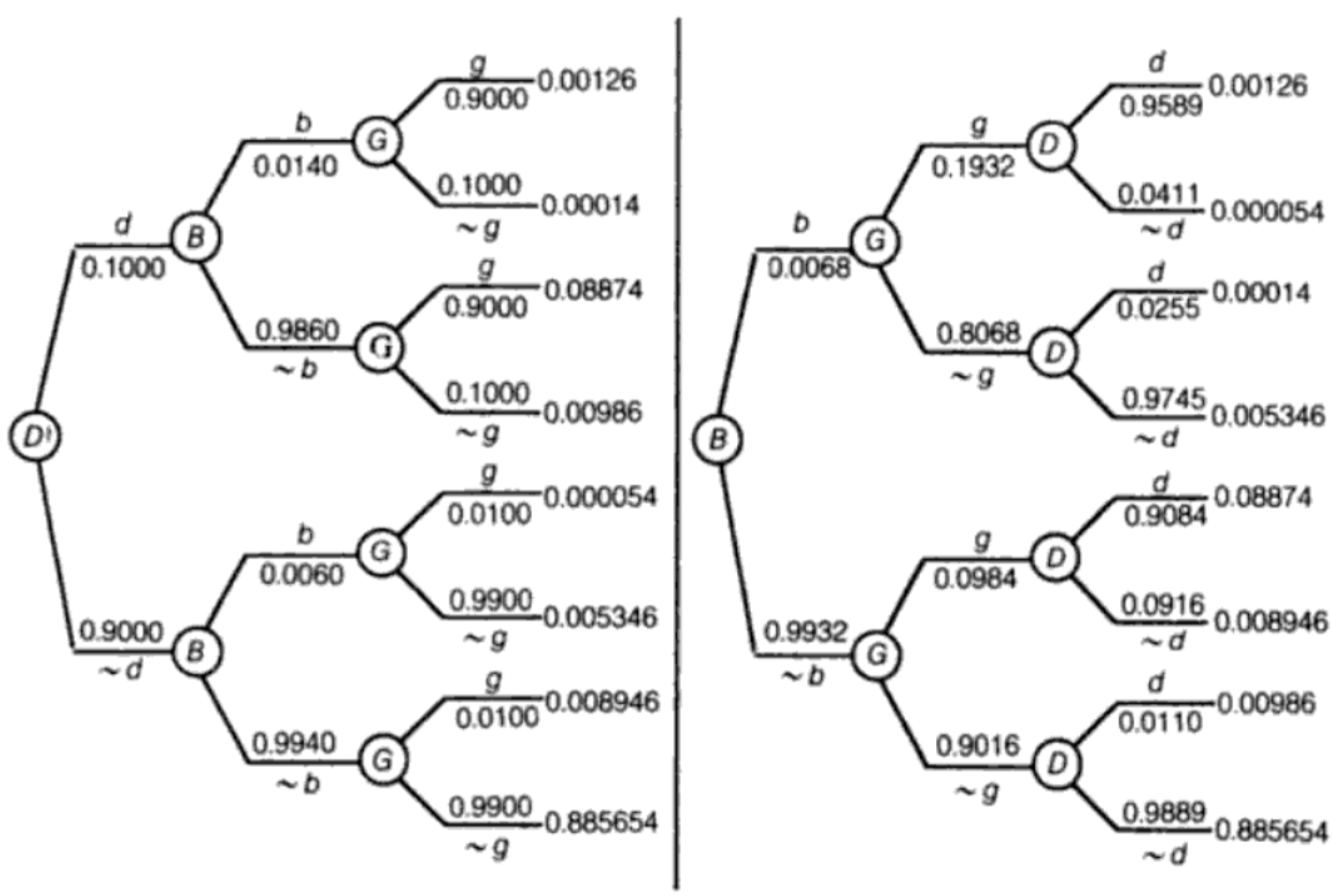

Figure 11.1 The preprocessing of probabilities in the diabetes diagnosis problem.

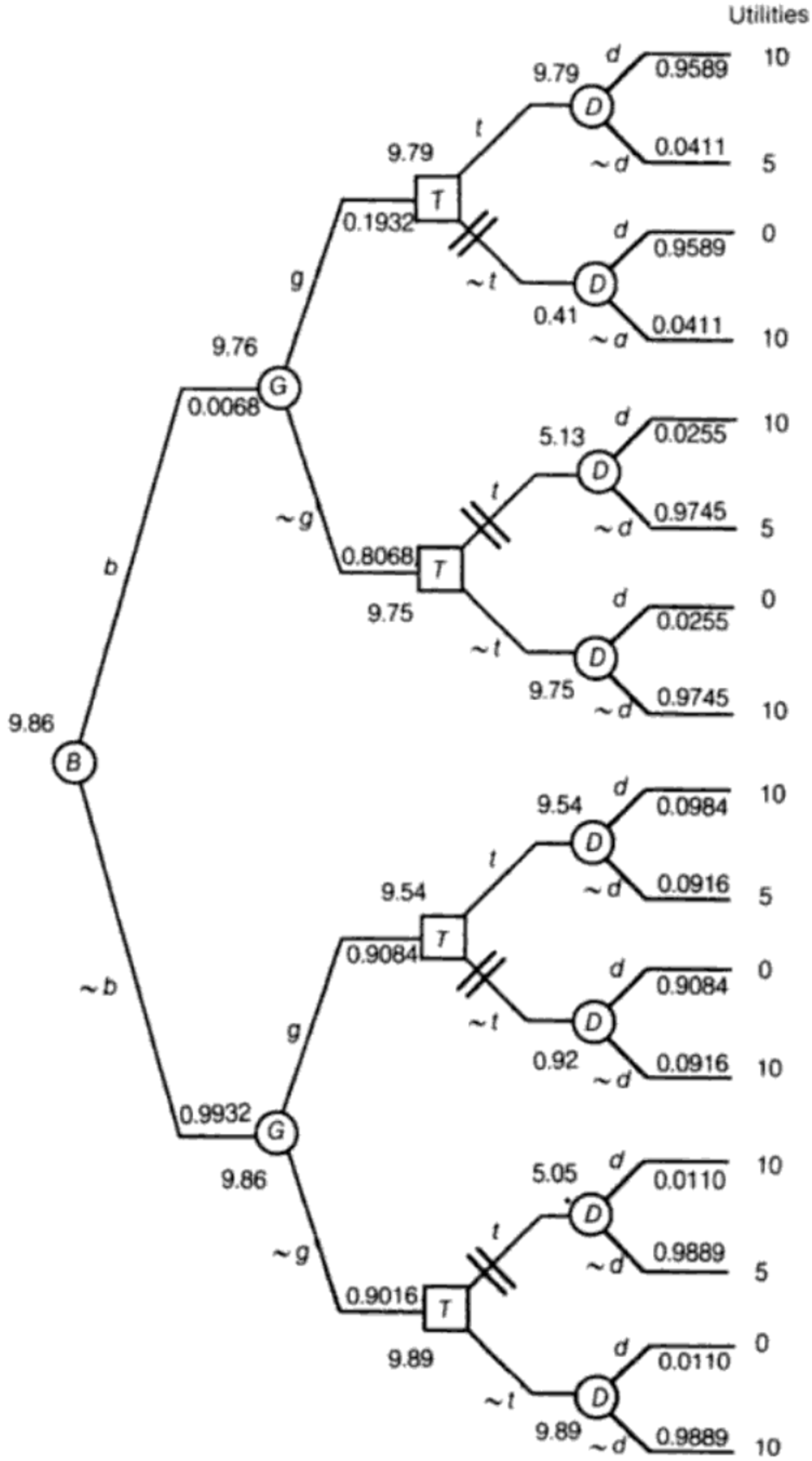

Figure 11.2 A decision tree representation and solution of the diabetes diagnosis problem.

A count of the computations involved in the solution of this problem (including the preprocessing of probabilities) reveals that we do 17 additions, 38 multiplications, 12 divisions, and 4 comparisons, for a total of 71 operations. If we solve this problem using Shachter's (1986) arc-reversal method, we do exactly the same operations (and hence the same number of operations) as the decision tree method. As we will see in a later section, we will solve this problem using our method with only 43 operations.

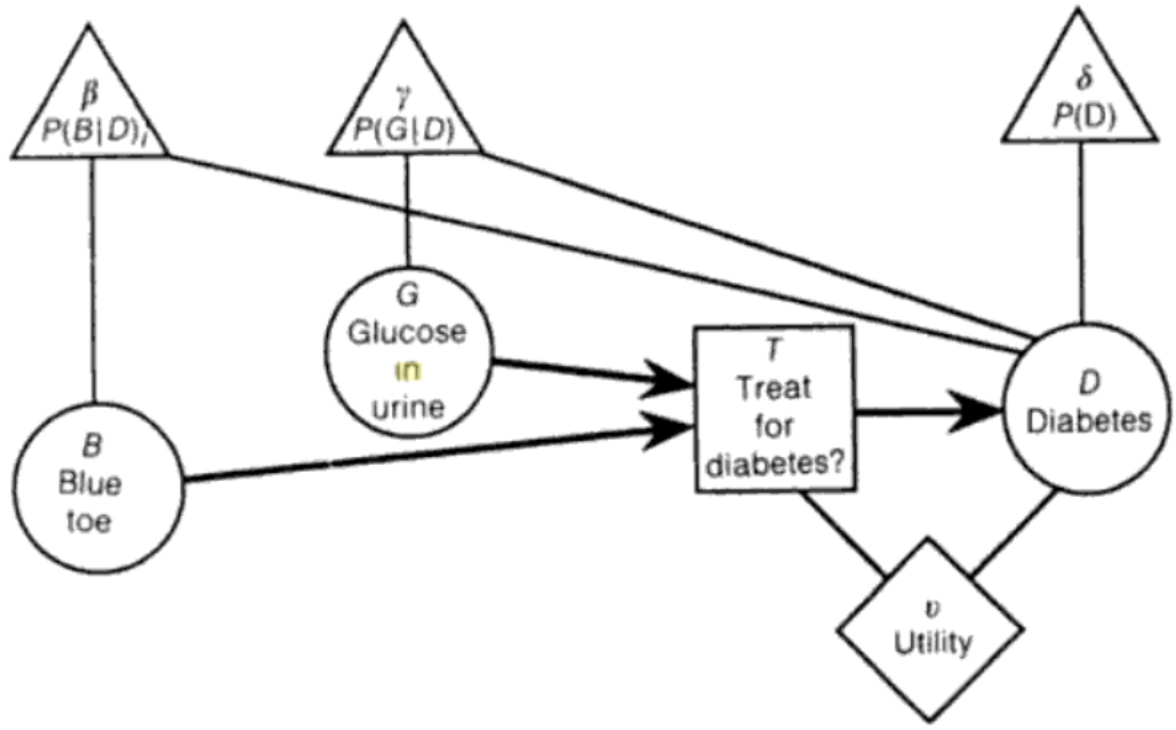

Figure 11.3 A valuation network for diabetes diagnosis problem.

VALUATION-BASED SYSTEM REPRESENTATION

In this section, we describe a valuation-based system (VBS) representation of a decision problem. A VBS representation consists of decision variables, random variables, frames, utility valuations, potentials, and precedence constraints. A graphical depiction of a VBS is called a **valuation network.** Figure 11.3 shows a valuation network for the diabetes diagnosis problem.

VARIABLES, FRAMES AND CONFIGURATIONS

A decision node is represented as a variable. The possible values of a decision variable represent the acts available at that point. We use the symbol $\mathscr{W}_D$ for the set of possible values of decision variable D. We assume that the decision-maker has to pick one and only one of the elements of $\mathscr{W}_D$ as a decision. We call $\mathscr{W}_D$ the **frame** for D. Decision variables are represented in valuation networks by rectangular nodes. In the *diabetes diagnosis* problem, there is one decision node T. The frame for T has two elements: Treat the patient for diabetes (t), and not treat ($\sim t$).

If R is a random variable, we will use the symbol $\mathscr{W}_R$ to denote its possible values. We assume that one and only one of the elements of $\mathscr{W}_R$ can be the true value of R. We can $\mathscr{W}_R$ the frame for R. Random variables are represented in valuation networks by circular nodes. In the diabetes diagnosis problem, there are three random variables: blue toe (B), glucose in the urine (G) and diabetes (D). Each variable has a frame consisting of two elements.

Let $\mathscr{X}_D$ denote the set of all decision variables, let $\mathscr{X}_R$ denote the set of all random variables, and let $\mathscr{X} = \mathscr{X}_D \cup \mathscr{X}_R$ denote the set of all variables. We will often deal with non-empty subsets of variables in $\mathscr{X}$. Given a non-empty subset h of $\mathscr{X}$, let $\mathscr{W}_h$ denote the Cartesian product of $\mathscr{W}_X$ for X in h, i.e. $\mathscr{W}_h = \times \{\mathscr{W}_X | X \in h\}$. We can think of the set $\mathscr{W}_h$ as the set of possible values of the joint variable h. Accordingly, we call $\mathscr{W}_h$

the frame for h. Also, we will refer to elements of $\mathscr{W}_h$ as **configurations** of h. We will use lower-case, bold-faced letters such as **x, y**, etc. to denote configurations. Also, if **x** is a configuration of g, **y** is a configuration of h, and $g \cap h = \varnothing$, then (**x, y**) will denote a configuration of $g \cup h$.

It will be convenient to extend this terminology to the case where the set of variables h is empty. We will adopt the convention that the frame for the empty set $\varnothing$ consists of a single configuration, and we will use the symbol ◆ to name that configuration; $\mathscr{W}_{\varnothing} = \{◆\}$. To be consistent with our notation above, we will adopt the convention that if **x** is a configuration for g, then (**x**, ◆) = **x**.

VALUATIONS

Suppose $h \subseteq \mathscr{X}$. A **utility valuation** υ for h is a function from $\mathscr{W}_h$ to $\mathbb{R}$, where $\mathbb{R}$ denotes the set of real numbers. The values of utility valuations are utilities. If $h = d \cup r$ where $d \subseteq \mathscr{X}_D$ and $r \subseteq \mathscr{X}_R$, $\mathbf{x} \in \mathscr{W}_d$, and $\mathbf{y} \in \mathscr{W}_r$, then $\upsilon(\mathbf{x}, \mathbf{y})$ denotes the utility to the decision-maker if the decision-maker chooses configuration **x** and the true configuration of r is **y**. If υ is a utility valuation for h and $X \in h$, then we will say that υ **bears on** X.

In a valuation network, a utility valuation is represented by a diamond-shaped node. To permit the identification of all valuations that bear on a variable, we will draw undirected edges between the utility valuation node and all the variable nodes it bears on. In the diabetes diagnosis problem, there is one utility valuation υ as shown in Fig. 11.3. Table 11.1 shows the values of this utility valuation.

Suppose $h \subseteq \mathscr{X}$. A **probability valuation** (or, simply, a **potential**) ρ for h is a function from $\mathscr{W}_h$ to the unit interval [0, 1]. The values of potentials are probabilities. In a valuation network, a potential is represented by a triangular node. Again, to identify the variables related by a potential, we draw undirected edges between the potential node and all the variable nodes it bears on. In the diabetes diagnosis problem, there are three potentials β, γ, and δ, as shown in Figure 11.3. Table 11.2 shows the details of these potentials. Note that β is a potential for $\{B, D\}$, γ is a potential for $\{G, D\}$, and δ is a potential for $\{D\}$.

PRECEDENCE CONSTRAINTS

Besides acts, states, probabilities, and utilities, an important ingredient of problems in decision analysis is the chronology, or structure, of information constraints. Some

Table 11.2 The potentials δ, β and γ

D	δ	β		B: b	B: $\sim b$	γ		G: g	G: $\sim g$
d	0.10	D	d	0.014	0.986	D	d	0.90	0.10
$\sim d$	0.90	D	$\sim d$	0.006	0.994	D	$\sim d$	0.01	0.99

decisions have to be made before the observation of some uncertain states, and some decisions can be postponed until after some states are observed. In the diabetes diagnosis problem, for example, the medical intern does not know whether the patient has diabetes or not. And the decision whether to treat the patient for diabetes or not may be postponed until after the observation of the blue toe and glucose in urine.

If a decision-maker expects to be informed of the true value of random variable R before making a decision D, then we represent this situation by the binary relation $R \to D$ (read as R **precedes** D). On the other hand, if a random variable R is only revealed after a decision D is made or perhaps never revealed, then we represent this situation by the binary relation $D \to R$. In the diabetes diagnosis problem, we have the precedence constraints $B \to T, G \to T, T \to D$. The decision whether to treat the patient for diabetes or not (T) is only made after observing blue toe (B) and glucose in the urine (G). And, diabetes (D) is not known at the time the decision whether to treat the patient for diabetes (T) has to be made.

Suppose $>$ is a binary relation on $\mathscr{X}$ such that it is the transitive closure of $\to$, i.e. $X > Y$ if either $X \to Y$, or there exists a $Z \in \mathscr{X}$ such that $X > Z$ and $Z > Y$. First, we will assume that $>$ is a partial order of $\mathscr{X}$ (otherwise the decision problem is ill defined and not solvable). Second, we will require that this partial order $>$ is such that for any $D \in \mathscr{X}_D$ and any $R \in \mathscr{X}_R$, either $D > R$ or $R > D$. We will refer to this second condition as the **perfect recall condition**. The reason for the perfect recall condition is as follows. Given the meaning of the precedence relation $\to$, for any decision variable D and any random variable R, either R is known when decision D has to be made, or not. This translates to either $R > D$ or $D > R$.

Next, we will define two operations called combination and marginalization. We use these operations to solve the valuation-based system representation. First we start with some notation.

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) onto $\{W, X\}$ 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 we will let $\mathbf{x}^{\downarrow h}$ denote the projection of $\mathbf{x}$ onto h. The projection $\mathbf{x}^{\downarrow h}$ is always a configuration of h. If $h = g$ and $\mathbf{x}$ is a configuration of g, then $\mathbf{x}^{\downarrow h} = \mathbf{x}$. If $h = \emptyset$, then $\mathbf{x}^{\downarrow h} = \blacklozenge$.

COMBINATION

The definition of **combination** will depend on the type of valuations being combined. Suppose h and g are subsets of $\mathscr{X}$, suppose υ_i is a utility valuation for h, and suppose ρ_j is a potential for g. Then the combination of υ_i and ρ_j, denoted by $\upsilon_i \otimes \rho_j$, is a utility valuation for $h \cup g$ obtained by pointwise multiplication of υ_i and ρ_j, i.e. $(\upsilon_i \otimes \rho_j)(\mathbf{x}) = \upsilon_i(\mathbf{x}^{\downarrow h})\rho_j(\mathbf{x}^{\downarrow g})$ for all $\mathbf{x} \in \mathscr{W}_{h \cup g}$. See Table 11.3 for an example.

Table 11.3 The computation of the combinations $\beta \oplus \gamma \oplus \delta$ and $\upsilon \oplus \beta \oplus \gamma \oplus \delta$

$\mathcal{W}_{\{B,G,T,D\}}$				υ	β	γ	δ	$\beta \oplus \gamma \oplus \delta$	$\upsilon \oplus \beta \oplus \gamma \oplus \delta = \tau$
b	g	t	d	10	0.014	0.90	0.10	0.00126	0.0126
b	g	t	~d	5	0.006	0.01	0.90	0.000054	0.00027
b	g	~t	d	0	0.014	0.90	0.10	0.00126	0
b	g	~t	~d	10	0.006	0.01	0.90	0.000054	0.00054
b	~g	t	d	10	0.014	0.10	0.10	0.00014	0.0014
b	~g	t	~d	5	0.006	0.99	0.90	0.005346	0.02673
b	~g	~t	d	0	0.014	0.10	0.10	0.00014	0
b	~g	~t	~d	10	0.006	0.99	0.90	0.005346	0.05346
~b	g	t	d	10	0.986	0.90	0.10	0.08874	0.8874
~b	g	t	~d	5	0.994	0.01	0.90	0.008946	0.04473
~b	g	~t	d	0	0.986	0.90	0.10	0.08874	0
~b	g	~t	~d	10	0.994	0.01	0.90	0.008946	0.08646
~b	~g	t	d	10	0.986	0.10	0.10	0.00986	0.0986
~b	~g	t	~d	5	0.994	0.99	0.90	0.885654	4.42827
~b	~g	~t	d	0	0.986	0.10	0.10	0.00986	0
~b	~g	~t	~d	10	0.994	0.99	0.90	0.885654	8.86554

Suppose h and g are subsets of $\mathcal{X}$, suppose ρ_i is a potential for h, and suppose ρ_j is a potential for g. Then the combination of ρ_i and ρ_j, denoted by $\rho_i \otimes \rho_j$, is a potential for $h \cup g$ obtained by pointwise multiplication of ρ_i and ρ_j, i.e. $(\rho_i \otimes \rho_j)(\mathbf{x}) = \rho_i(\mathbf{x}^{\downarrow h})\rho_j(\mathbf{x}^{\downarrow g})$ for all $\mathbf{x} \in \mathcal{W}_{h \cup g}$. See Table 11.3 for an example.

Note that combination is commutative and associative. Thus, if $\{\alpha_1, \ldots, \alpha_k\}$ is a set of valuations, we will write $\otimes\{\alpha_1, \ldots, \alpha_k\}$ or $\alpha_1 \otimes \cdots \otimes \alpha_k$ to mean the combination of valuations in $\{\alpha_1, \ldots, \alpha_k\}$ in some sequence.

MARGINALIZATION

Suppose h is a subset of variables, and suppose α is a valuation for h. Marginalization is an operation where we reduce valuation α to a valuation $\alpha^{\downarrow(h-\{X\})}$ for $h - \{X\}$. $\alpha^{\downarrow(h-\{X\})}$ is called the **marginal** of α for $h - \{X\}$. Unlike combination, the definition of marginalization does not depend on the nature of α. But the definition of marginalization does depend on whether X is a decision or a random variable.

If R is a random variable, $\alpha^{\downarrow(h-\{R\})}$ is obtained by summing α over the frame for R, i.e. $\alpha^{\downarrow(h-\{R\})}(\mathbf{c}) = \Sigma\{\alpha(\mathbf{c}, \mathbf{r}) | \mathbf{r} \in \mathcal{W}_R\}$ for all $\mathbf{c} \in \mathcal{W}_{h-\{R\}}$. Here, α could be either a utility valuation or a potential. See Table 11.4 for an example.

Table 11.4 The computation of $\tau^{\downarrow\{B,G,T\}}$, $\tau^{\downarrow\{B,G\}}$, Ψ_T, $\tau^{\downarrow\{B\}}$, and $\tau^{\downarrow\emptyset}(\blacklozenge)$. τ denotes the joint valuation $\upsilon\otimes\beta\otimes\gamma\otimes\delta$

$\mathcal{W}_{\{B,G,T,D\}}$				τ	$\tau^{\downarrow\{B,G,T\}}$	$\tau^{\downarrow\{B,G\}}$	Ψ_T	$\tau^{\downarrow\{B\}}$	$\tau^{\downarrow\emptyset}(\blacklozenge)$
b	g	t	d	0.0126	0.01287	0.01287	t	0.06633	9.864
b	g	t	~d	0.00027					
b	g	~t	d	0	0.00054				
b	g	~t	~d	0.00054					
b	~g	t	d	0.0014	0.02813	0.05346	~t		
b	~g	t	~d	0.02673					
b	~g	~t	d	0	0.05346				
b	~g	~t	~d	0.05346					
~b	g	t	d	0.8874	0.93213	0.93213	t	9.79767	
~b	g	t	~d	0.04473					
~b	g	~t	d	0	0.08646				
~b	g	~t	~d	0.08646					
~b	~g	t	d	0.0986	4.52687	8.86554	~t		
~b	~g	t	~d	4.42827					
~b	~g	~t	d	0	8.86554				
~b	~g	~t	~d	8.86554					

If D is a decision variable, $\alpha^{\downarrow(h-\{D\})}$ is obtained by maximizing α over the frame for D, i.e. $\alpha^{\downarrow(h-\{D\})}(\mathbf{c}) = \max\{\alpha(\mathbf{c},\mathbf{d}) \mid \mathbf{d}\in\mathcal{W}_D\}$ for all $\mathbf{c}\in\mathcal{W}_{h-\{D\}}$. Here, α must be a utility valuation. See Table 11.4 for an example.

We now state three lemmas regarding the marginalization operation. Lemma 1 states that in marginalizing two decision variables out of a valuation, the order in which the variables are eliminated does not affect the result. Lemma 2 states a similar result for marginalizing two random variables out of a valuation. Lemma 3 states that in marginalizing a decision variable and a random variable out of a valuation, the order in which the two variables are eliminated may make a difference.

Lemma 11.1
Suppose h is a subset of $\mathcal{X}$ containing decision variables D_1 and D_2, and suppose α is a utility valuation for h. Then

$$(\alpha^{\downarrow(h-\{D_1\})})^{\downarrow(h-\{D_1,D_2\})}(\mathbf{c}) = (\alpha^{\downarrow(h-\{D_2\})})^{\downarrow(h-\{D_1,D_2\})}(\mathbf{c})$$

for all $\mathbf{c}\in\mathcal{W}_{h-\{D_1,D_2\}}$.

Lemma 11.2
Suppose h is a subset of $\mathcal{X}$ containing random variables R_1 and R_2, and suppose

α is a valuation for h. Then

$$(\alpha^{\downarrow(h-\{R_1\})})^{\downarrow(h-\{R_1,R_2\})}(\mathbf{c}) = (\alpha^{\downarrow(h-\{R_2\})})^{\downarrow(h-\{R_1,R_2\})}(\mathbf{c})$$

for all $\mathbf{c} \in \mathcal{W}_{h-\{R_1,R_2\}}$.

Lemma 11.3
Suppose h is a subset of $\mathcal{X}$ containing decision variable D and random variable R, and suppose α is a utility valuation for h. Then

$$(\alpha^{\downarrow(h-\{D\})})^{\downarrow(h-\{R,D\})}(\mathbf{c}) \geqslant (\alpha^{\downarrow(h-\{R\})})^{\downarrow(h-\{R,D\})}(\mathbf{c})$$

for all $\mathbf{c} \in \mathcal{W}_{h-\{R,D\}}$.

It is clear from Lemma 11.3, that in marginalizing more than one variable, the order of elimination of the variables may make a difference. As we shall see shortly, we need to marginalize all variables out of the joint valuation. What sequence should we use? This is where the precedence constraints come into play. We will define marginalization such that variable Y is marginalized before X whenever $X > Y$.

Suppose h and g are non-empty subsets of $\mathcal{X}$ such that g is a proper subset of h, suppose α is a valuation for h, and suppose $>$ is a partial order on $\mathcal{X}$ satisfying the perfect recall condition. The **marginal** of α for g **with respect to the partial order** $>$, denoted by $\alpha^{\downarrow g}$, is a valuation for g defined as follows:

$$\alpha^{\downarrow g} = (((\alpha^{\downarrow(h-\{X_1\})})^{\downarrow(h-\{X_1,X_2\})})\ldots)^{\downarrow(h-\{X_1,X_2,\ldots,X_k\})} \tag{11.1}$$

where $h-g=\{X_1,\ldots,X_k\}$ and $X_1X_2\ldots X_k$ is a sequence of variables in $h-g$ such that with respect to the partial order $>$, X_1 is a minimal element of $h-g$, X_2 is a minimal element of $h-g-\{X_1\}$, etc.

The marginalization sequence $X_1X_2\ldots X_k$ may not be unique since $>$ is only a partial order. But, since $>$ satisfies the perfect recall condition, it is clear from Lemmas 1 and 2 that the definition of $\alpha^{\downarrow g}$ in Equation (11.1) is well defined.

STRATEGY

The main objective in solving a decision problem is to compute an optimal strategy. What constitutes a strategy? Intuitively, a strategy is a choice of an act for each decision variable D as a function of configurations of random variables R such that $R > D$. Let $\Pr(D) = \{R \in \mathcal{X}_R \mid R > D\}$. We shall refer to $\Pr(D)$ as the **predecessors** of D. Thus a **strategy** σ is a collection of functions $\{\xi_D\}_{D\in\mathcal{X}_D}$ where $\xi_D\colon \mathcal{W}_{\Pr(D)} \to \mathcal{W}_D$.

SOLUTION FOR A VARIABLE

Computing an optimal strategy is a matter of bookkeeping. Each time we marginalize a decision variable out of a utility valuation using maximization, we store a table of optimal values of the decision variable where the maxima are achieved. We can think of this table as a function. We will call this function a **solution** for the decision variable. Suppose h is a subset of variables such that decision variable $D \in h$, and suppose υ is a utility valuation for h. A function Ψ_D: $\mathscr{W}_{h-\{D\}} \rightarrow \mathscr{W}_D$ is called a solution for D (with respect to υ) if $\upsilon^{\downarrow(h-\{D\})}(\mathbf{c}) = \upsilon(\mathbf{c}, \Psi_D(\mathbf{c}))$ for all $\mathbf{c} \in \mathscr{W}_{h-\{D\}}$. See Table 11.4 for an example.

SOLVING A VALUATION-BASED SYSTEM

Suppose $\Delta = \{\mathscr{X}_D, \mathscr{X}_R, \{\mathscr{W}_X\}_{X \in \mathscr{X}}, \{\upsilon_1\}, \{\rho_1, \ldots, \rho_n\}, \rightarrow\}$ is a VBS representation of a decision problem consisting of one utility valuation and n potentials. What do the potentials represent? And how do we solve Δ? We will answer these two related questions in terms of a canonical decision problem.

CANONICAL DECISION PROBLEM

A **canonical decision problem** Δ_C consists of a single decision variable D with a finite frame $\mathscr{W}_D$, a single random variable R with a finite frame $\mathscr{W}_R$, a single utility valuation υ for $\{D, R\}$, a single potential ρ for $\{R, D\}$ such that

$$\Sigma\{\rho(\mathbf{d}, \mathbf{r}) | \mathbf{r} \in \mathscr{W}_R\} = 1 \qquad \text{for all } \mathbf{d} \in \mathscr{W}_D \tag{11.2}$$

and a precedence relation $\rightarrow$ defined by $D \rightarrow R$. Figure 11.4 shows a valuation network and a decision tree representation of the canonical decision problem.

The meaning of the canonical decision problem is as follows. The elements of $\mathscr{W}_D$ are acts, and the elements of $\mathscr{W}_R$ are states of nature. The potential ρ is a family of probability distributions for R, one for each act $\mathbf{d} \in \mathscr{W}_D$, i.e. $\Sigma\{\rho(\mathbf{d}, \mathbf{r}) | \mathbf{r} \in \mathscr{W}_R\} = 1$ for all $\mathbf{d} \in \mathscr{W}_D$. In other words, the probability distribution of random variable R is conditioned on the act $\mathbf{d}$ chosen by the decision-maker. The probability $\rho(\mathbf{d}, \mathbf{r})$ can be interpreted as the conditional probability of $R = \mathbf{r}$ given that $D = \mathbf{d}$.

The utility valuation υ is a utility function—if the decision-maker chooses act $\mathbf{d}$ and the state of nature $\mathbf{r}$ prevails, then the utility to the decision-maker is $\upsilon(\mathbf{d}, \mathbf{r})$.

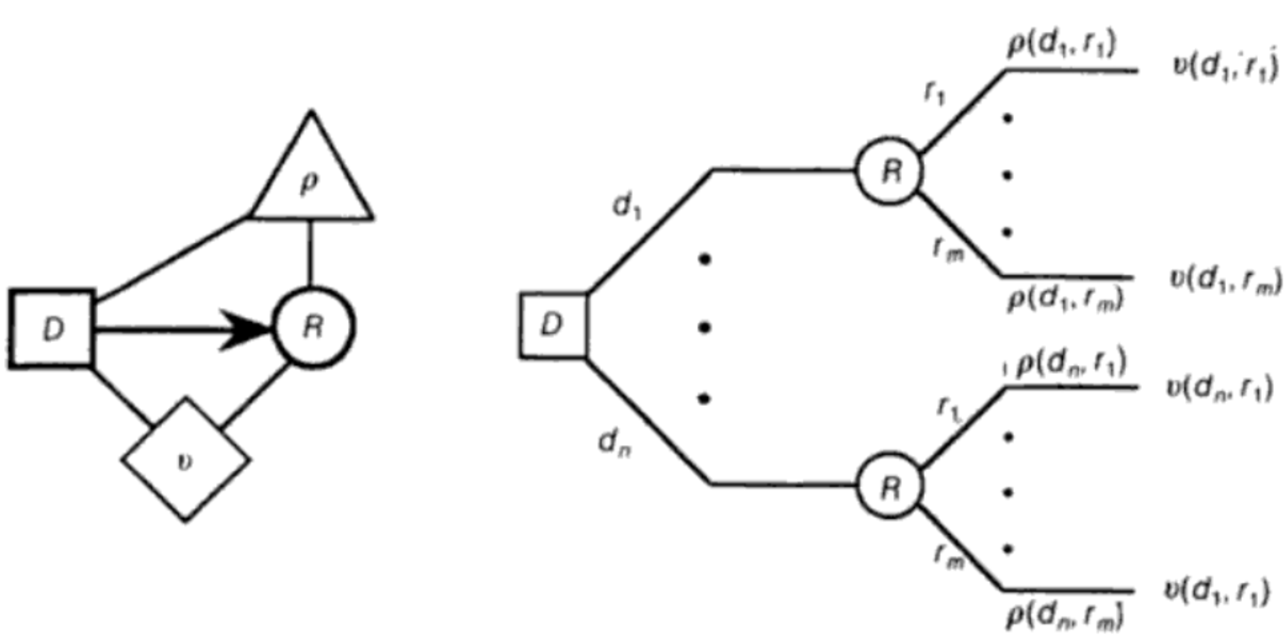

Figure 11.4 A valuation network and a decision tree representation of the canonical decision problem.

The precedence relation $\rightarrow$ states that the true state of nature is revealed to the decision-maker only after the decision-maker has chosen an act.

Solving a canonical decision problem using the criterion of maximizing expected utility is easy. The expected utility associated with act **d** is $\Sigma\{(\upsilon\oplus\rho)(\mathbf{d},\mathbf{r})|\mathbf{r}\in\mathcal{W}_R\} = (\upsilon\oplus\rho)^{\downarrow\{D\}}(\mathbf{d})$. The maximum expected utility (associated with an optimal act, say $\mathbf{d}^*$) is max $\{(\upsilon\oplus\rho)^{\downarrow\{D\}}(\mathbf{d})|\mathbf{d}\in\mathcal{W}_D\} = ((\upsilon\oplus\rho)^{\downarrow\{D\}})^{\downarrow\varnothing}(\blacklozenge) = (\upsilon\oplus\rho)^{\downarrow\varnothing}(\blacklozenge)$. Finally, act $\mathbf{d}^*$ is optimal if and only if $(\upsilon\oplus\rho)^{\downarrow\{D\}}(\mathbf{d}^*) = (\upsilon\oplus\rho)^{\downarrow\varnothing}(\blacklozenge)$.

Consider the decision problem $\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{\mathcal{W}_X\}_{X\in\mathcal{X}}, \{\upsilon_1\}, \{\rho_1,\ldots,\rho_n\}, \rightarrow\}$. We will explain the meaning of Δ by reducing it to an equivalent canonical decision problem $\Delta_C = \{\{D\},\{R\},\{\mathcal{W}_D, \mathcal{W}_R\},\{\upsilon\},\{\rho\},\rightarrow\}$. To define Δ_C, we need to define $\mathcal{W}_D$, $\mathcal{W}_R$, υ and ρ. Define $\mathcal{W}_D$ such that, for each distinct strategy σ of Δ, there is a corresponding act $\mathbf{d}_\sigma$ in $\mathcal{W}_D$. Define $\mathcal{W}_R$ such that for each distinct configuration **y** of $\mathcal{X}_R$ in Δ, there is a corresponding configuration $\mathbf{r}_\mathbf{y}$ in $\mathcal{W}_R$.

Before we define utility valuation υ for $\{D, R\}$, we need some notation. Suppose $\sigma = \{\xi_D\}_{D\in\mathcal{X}_D}$ is a strategy, and suppose **y** is a configuration of $\mathcal{X}_R$. Then together σ and **y** determine a unique configuration of $\mathcal{X}_D$. We will let $\mathbf{a}_{\sigma,\mathbf{y}}$ denote this unique configuration of $\mathcal{X}_D$. By definition, $\mathbf{a}_{\sigma,\mathbf{y}}{}^{\downarrow\{D\}} = \xi_D(\mathbf{y}^{\downarrow\mathrm{Pr}(D)})$ for all $D\in\mathcal{X}_D$.

Consider the utility valuation υ_1 in Δ. Assume that the domain of this valuation includes all of $\mathcal{X}_D$. Typically the domain of this valuation will include also some (or all) random variables. Let p denote the subset of random variables included in the domain of the joint utility valuation, i.e. $p\subseteq\mathcal{X}_R$ such that υ_1 is a utility valuation for $\mathcal{X}_D\cup p$. Define a utility valuation υ for $\{D, R\}$ such that $\upsilon(\mathbf{d}_\sigma, \mathbf{r}_\mathbf{y}) = \upsilon_1(\mathbf{a}_{\sigma,\mathbf{y}}, \mathbf{y}^{\downarrow p})$ for all strategies σ of Δ, and all configurations $\mathbf{y}\in\mathcal{W}_{\mathcal{X}_R}$. Remember that $\mathbf{a}_{\sigma,\mathbf{y}}$ is the unique configuration of $\mathcal{X}_D$ determined by σ and **y**.

Consider the joint potential $\rho_1\otimes\ldots\otimes\rho_n$. Assume that this potential includes all random variables in its domain. Let q denote the subset of decision variables included in the domain of the joint potential, i.e. $q\subseteq\mathcal{X}_D$ such that $\rho_1\otimes\ldots\otimes\rho_n$ is a potential for $q\cup\mathcal{X}_R$. Note that q could be empty. Define potential ρ for $\{D, R\}$, such that $\rho(\mathbf{d}_\sigma, \mathbf{r}_\mathbf{y}) = (\rho_1\otimes\ldots\otimes\rho_n)(\mathbf{a}_{\sigma,\mathbf{y}}{}^{\downarrow q}, \mathbf{y})$ for all strategies σ and all configurations $\mathbf{y}\in\mathcal{W}_{\mathcal{X}_R}$. Δ_C, as defined above, will be a canonical decision problem only if ρ satisfies Condition (11.2). This motivates the following definition. Δ is a **well-defined VBS representation of a decision problem** if and only if $\Sigma\{(\rho_1\otimes\ldots\otimes\rho_n)(\mathbf{x},\mathbf{y})|\mathbf{y}\in\mathcal{W}_{\mathcal{X}_R}\} = 1$ for every $\mathbf{x} = \mathcal{W}_q$.

In summary, the potentials $\{\rho_1,\ldots,\rho_n\}$ represents the factors of a family of probability distributions. It is easy to verify that the VBS representation of the diabetes diagnosis problem is well defined since $\delta\otimes\beta\otimes\gamma$ is a joint probability distribution for $\{D, B, G\}$ (see Table 11.3).

THE DECISION PROBLEM

Suppose $\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{\mathcal{W}_X\}_{X\in\mathcal{X}}, \{\upsilon_1\}, \{\rho_1,\ldots,\rho_n\}, \rightarrow\}$ is a well-defined decision problem. Let $\Delta_C = \{\{D\},\{R\},\{\mathcal{W}_D, \mathcal{W}_R\},\{\upsilon\},\{\rho\},\rightarrow\}$ represents an equivalent canonical decision problem. In the canonical decision problem Δ_C, the two computations that are of interest are, first, the computation of the maximum expected value

$(\upsilon \otimes \rho)^{\downarrow \varnothing}(\blacklozenge)$, and second, the computation of an optimal act $\mathbf{d}_{\sigma^*}$ such that $(\upsilon \otimes \rho)^{\downarrow \{D\}}(\mathbf{d}_{\sigma^*}) = (\upsilon \otimes \rho)^{\downarrow \varnothing}(\blacklozenge)$. Since we know the mapping between Δ and Δ_C, we can now formally define the questions posed in a decision problem Δ. There are two computations of interest.

First, we would like to compute the maximum expected utility. The maximum expected utility is given by $(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow \varnothing}(\blacklozenge)$. Second, we would like to compute an optimal strategy σ^* that gives us the maximum expected value $(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow \varnothing}(\blacklozenge)$. A strategy σ^* of Δ is **optimal** if $(\upsilon \otimes \rho)^{\downarrow \{D\}}(\mathbf{d}_{\sigma^*}) = (\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow \varnothing}(\blacklozenge)$, where υ, ρ, and D refer to the equivalent canonical decision problem Δ_C.

In the diabetes diagnosis problem, we have four valuations υ, β, γ, and δ. Also, from the precedence constraints, we have $B > T, G > T, T > D$. Thus we need to compute either

$$((((\upsilon \otimes \beta \otimes \gamma \otimes \delta)^{\downarrow \{B,G,T\}})^{\downarrow \{B,G\}})^{\downarrow \{B\}})^{\downarrow \varnothing} \quad \text{or} \quad ((((\upsilon \otimes \beta \otimes \gamma \otimes \delta)^{\downarrow \{B,G,T\}})^{\downarrow \{B,G\}})^{\downarrow \{G\}})^{\downarrow \varnothing}$$

In either case, we get the same answers. Tables 11.3 and 11.4 display the former computations. As seen from Table 11.4, the optimal expected utility is 9.864. Also, from Ψ_T, the solution for T (shown in Table 11.4), the optimal act is to treat the patient for diabetes if and only if the patient exhibits glucose in the urine.

Note that no divisions were done in the solution process, only additions and multiplications. But both decision tree and influence diagram methodologies involve unnecessary divisions and unnecessary multiplications to compensate for the unnecessary divisions. It is this feature of valuation-based systems that makes it more efficient than decision trees and influence diagrams. In solving the diabetes diagnosis problem using our method, we do only 11 additions, 28 multiplications and 4 comparisons, for a total of 43 operations. This is a savings of 40% over the decision tree and influence diagram methodologies, which required a total of 71 operations.

A FUSION ALGORITHM FOR SOLVING VALUATION-BASED SYSTEMS USING LOCAL COMPUTATION

In this section, we will describe a fusion algorithm for solving a VBS using local computation. The solution for the diabetes diagnosis problem shown in Tables 11.3 and 11.4 involves combination on the space $\mathcal{W}_{\mathcal{X}}$. While this is possible for small problems, it is computationally not tractable for problems with many variables. Given the structure of the diabetes diagnosis problem, it is not possible to avoid the combination operation on the space of all four variables, B, G, T and D. But in some problems it may be possible to avoid such global computations.

The basic idea of the method is successively to delete all variables from the VBS. The sequence in which variables are deleted must respect the precedence constraints in the sense that if $X > Y$, then Y must be deleted before X. Since $>$ is only a partial order, a problem may allow several deletion sequences. Any allowable deletion sequence may be used. All allowable deletion sequences will 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 $\alpha_1,\ldots,\alpha_k$. Suppose α_i is a valuation for h_i. Let $\mathrm{Fus}_X\{\alpha_1,\ldots,\alpha_k\}$ denote the collection of valuations after fusing the valuations in the set $\{\alpha_1,\ldots,\alpha_k\}$ with respect to variable X. Then

$$\mathrm{Fus}_X\{\alpha_1,\ldots,\alpha_k\} = \{\alpha^{\downarrow(h-\{X\})}\} \cup \{\alpha_i | X \notin h_i\}$$

where $\alpha = \otimes\{\alpha_i | X \in h_i\}$, and $h = \cup(h_i | X \in h_i\}$. 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 main theorem.

Theorem 11.1
Suppose $\Delta = \{\mathscr{X}_D, \mathscr{X}_R, \{\mathscr{W}_X\}_{X\in\mathscr{X}}, \{\upsilon_1\}, \{\rho_1,\ldots,\rho_n\}, \rightarrow\}$ is a well-defined decision problem. Suppose $X_1, X_2 \ldots X_k$ is a sequence of variables in $\mathscr{X} = \mathscr{X}_D \cup \mathscr{X}_R$ such that, with respect to the partial order $>$, X_1 is a minimal element of $\mathscr{X}$, X_2 is a minimal element of $\mathscr{X} - \{X_1\}$, etc. Then $\{(\otimes\{\upsilon_1, \rho_1,\ldots,\rho_n\})^{\downarrow\varnothing}\} = \mathrm{Fus}_{X_k}\{\ldots\mathrm{Fus}_{X_2}\{\mathrm{Fus}_{X_1}\{\upsilon_1, \rho_1,\ldots, \rho_n\}\}\}$.

To illustrate Theorem 11.1, consider a VBS as shown in Fig. 11.5 for a generic medical diagnosis problem. In this VBS, there are three random variables, D, P, and S, and one decision variable, T. D represents a disease, P represents a pathological state caused by the disease, and S represents a symptom caused by the pathological state. We assume that S and D are conditionally independent given P. The potential δ is the prior probability of D, the potential π is the conditional probability of P given D, and the potential σ is the conditional probability of S given P. A medical intern first observes the symptom S and then either treats the patient for the disease

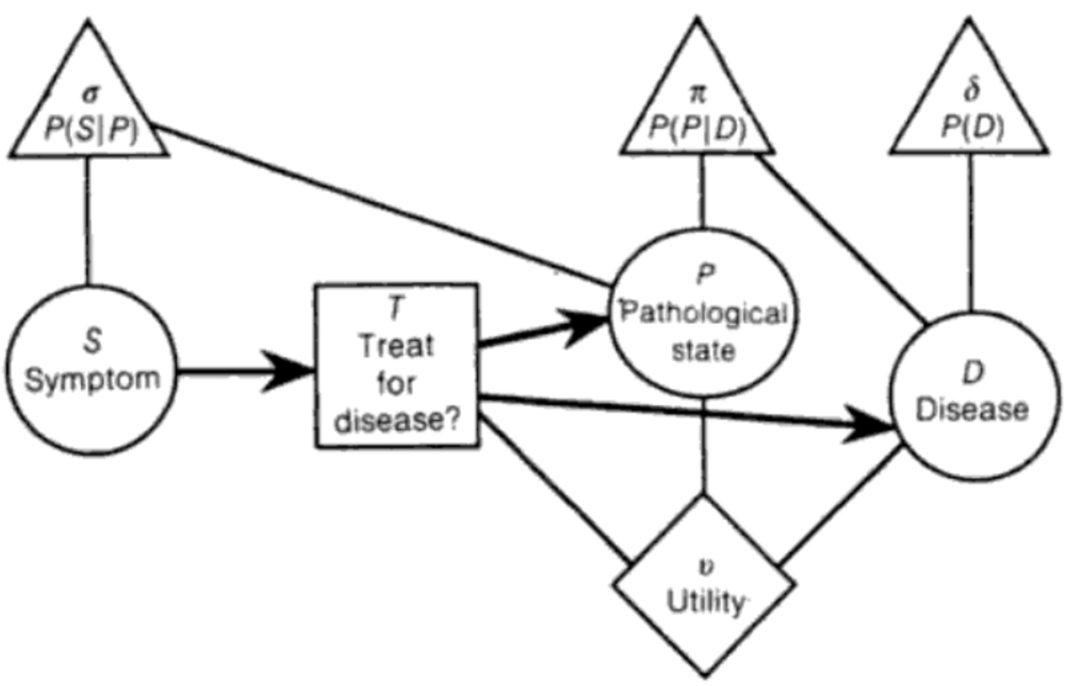

Figure 11.5 A valuation network for the medical diagnosis problem.

and pathological state or not. The utility valuation υ bears on the intern's action T, the pathological state P, and the disease variable D (see Shenoy, 1991d, for a more detailed description of this problem).

Figure 11.6 shows the results of the fusion algorithm for this problem. The deletion sequence used is *DPTS*. The valuation network labelled 0 in Fig. 11.6 is the same as the one in Fig. 11.5. The valuation network labelled 1 is the result after deletion of D and the resulting fusion. The combination involved in the fusion operation only involves variables D, P and T. The valuation network labelled 2 is the result after deletion of P. The combination operation involved in the corresponding fusion operation involves only three variables, P, T and S. The valuation network labelled 3 is the result after deletion of T. There is no combination involved here, only marginalization on the frame of $\{S, T\}$. The valuation network labelled 4 is the result after deletion of S. Again, there is no combination involved here, only marginalization on the frame of $\{S\}$. The maximum expected utility value is given by

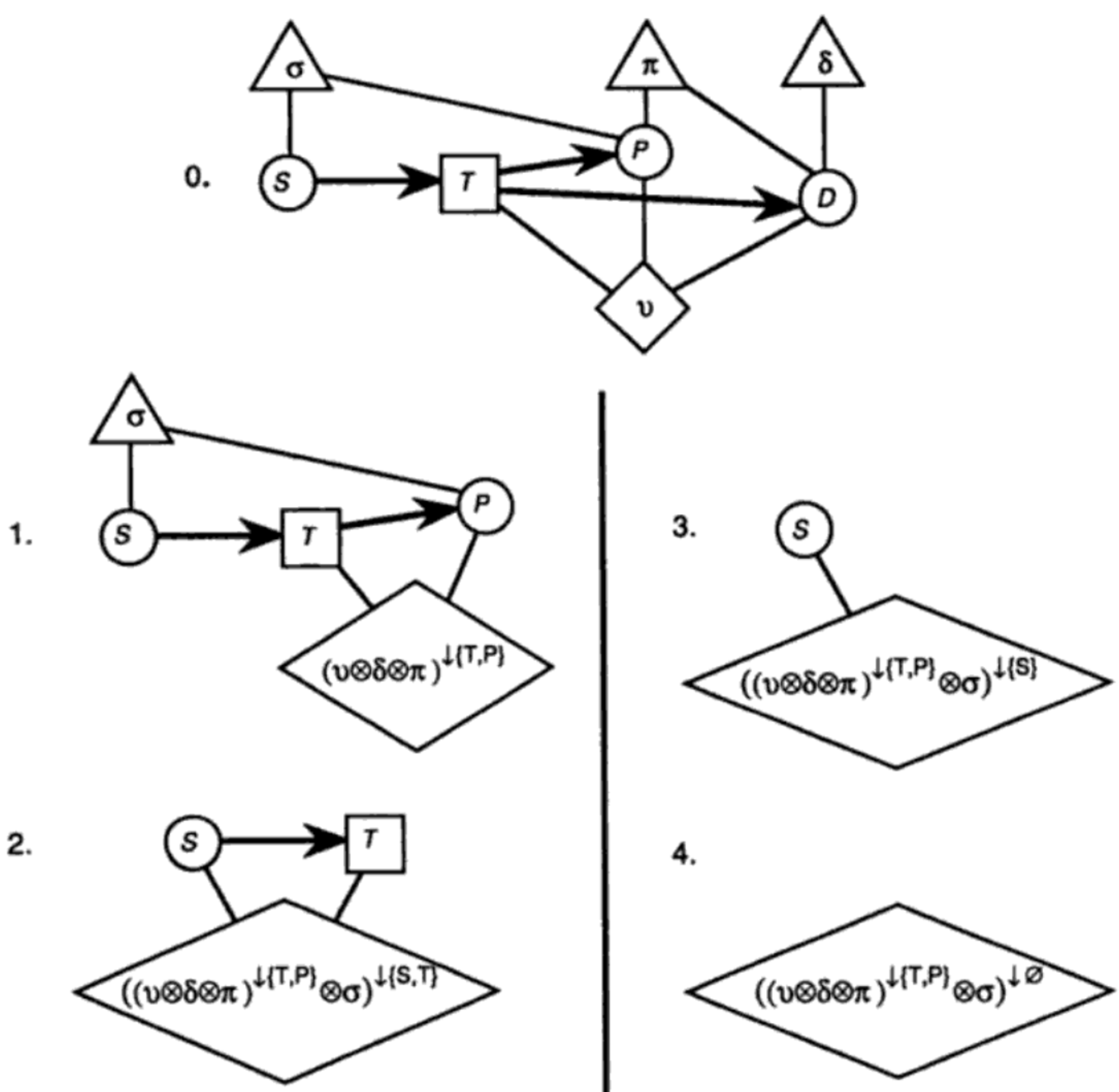

Figure 11.6 The fusion algorithm applied to the medical diagnosis problem.

$((\upsilon\otimes\delta\otimes\pi)^{\downarrow\{T,P\}}\otimes\sigma)^{\downarrow\varnothing}$ (◆). An optimal strategy is given by the solution for T with respect to $((\upsilon\otimes\delta\otimes\pi)^{\downarrow\{T,P\}}\otimes\sigma)^{\downarrow\{S,T\}}$ computed during fusion with respect to T. Note that in this problem, the fusion algorithm avoids computation on the frame of all four variables.

In solving the medical diagnosis problem using our method, we do a total of 31 operations (Shenoy, 1991d). On the other hand, for this problem, the decision tree solution method requires 59 operations (Shenoy, 1991d). Thus, for this problem, our method results in a savings of 47% over the decision tree methodology. If we use the influence diagram methodology for this problem, we do 49 operations (Shenoy, 1991d). Thus, for this problem, our method results in a savings of 20% over the influence diagram methodology.

The fusion method described in this section applies when there is one utility valuation in the VBS. This method will also apply unchanged in problems where the joint utility valuation factors multiplicatively into several utility valuations. In this case, we can define combination of utility valuations as pointwise multiplication, i.e. if υ_i is a utility valuation for h_i and υ_j is a utility valuation for h_j, then $\upsilon_i\otimes\upsilon_j$ is a utility valuation for $h_i\cup h_j$ defined by $(\upsilon_i\otimes\upsilon_j)(\mathbf{x})=\upsilon_i(\mathbf{x}^{\downarrow h_i})\,\upsilon_j(\mathbf{x}^{\downarrow h_j})$ for all $\mathbf{x}\in\mathcal{W}_{h_i\cup h_j}$. This method will not apply directly in problems where the joint utility valuation decomposes additively. In such problems, we will first have to combine all utility valuations before we apply the method described in this section. Thus the fusion method described in this section is unable to take computational advantage of an additive decomposition of the utility valuation. Shenoy (1992) describes a modification of the fusion algorithm that is able to take advantage of an additive decomposition of the utility function. The modification involves some divisions.

DELETION SEQUENCES

Since $>$ is only a partial order, in general, we may have many deletion sequences (sequences that satisfy the condition stated in Theorem 11.1). If so, which deletion sequence should one use? First, we note that all deletion sequences will lead to the same final result. This is implied in the statement of the theorem. Second, different deletion sequences may involve different computational efforts. For example, consider the VBS shown in Fig. 11.5. In this example, the deletion sequence *DPTS* involves less computational effort than *PDTS* as the former involves combinations on the frame of three variables only whereas the latter involves combination on the frame of all four variables. Finding an optimal deletion sequence is a secondary optimization problem that has been 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).

One such heuristic is called **one-step look ahead** (Kong, 1986). This heuristic tells us which variable to delete next from among those that qualify. 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 of Fig. 11.5, two variables qualify for first deletion, *P* and *D*. This heuristic would pick *D* over *P* since deletion of *P* involves combination over the frame of $\{S, D, P, T\}$ whereas deletion of D only involves combination over the frame of $\{T, P, D\}$. Thus, this heuristic would choose deletion sequence *DPTS*.

PROOFS

In this section we give proofs for all results in the chapter.

Proof of Lemma 11.1
The result follows directly from the definition of marginalization. ■

Proof of Lemma 11.2
The result follows directly from the definition of marginalization. ■

Proof of Lemma 11.3
The result follows directly from the definition of marginalization. ■

To prove Theorem 11.1, we need a lemma.

Lemma 11.4
Suppose $\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{\mathcal{W}_X\}_{X \in \mathcal{X}}, \{\upsilon_1\}, \{\rho_1, \ldots, \rho_n\}, \rightarrow\}$ is a well-defined decision problem. Suppose X is a minimal variable in $\mathcal{X} = \mathcal{X}_D \cup \mathcal{X}_R$ with respect to the partial order $>$, where $>$ is the transitive closure of $\rightarrow$. Then

$$(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow(\mathcal{X} - \{X\})} = \otimes \text{Fus}_X\{\upsilon_1, \rho_1, \ldots, \rho_n\}$$

Proof of Lemma 11.4
We will prove this result in two mutually exclusive and exhaustive cases. Suppose υ_1 is a payoff valuation for h_1, and suppose ρ_i is a potential for g_i, $i = 1, \ldots, n$.

Case 1
Suppose X is a decision variable. Without loss of generality, assume that υ_1, $\rho_1, \ldots, \rho_k$ are the only valuations that bear on X. Let $\upsilon = \upsilon_1 \otimes \rho_1 \otimes \ldots \otimes \rho_k$, let $h = h_1 \cup g_1 \cup \ldots. \cup g_k$, and let $\mathbf{c} \in \mathcal{W}_{\mathcal{X} - \{X\}}$. Then

$$(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow(\mathcal{X} - \{X\})}(\mathbf{c})$$

$$= \max\{[\upsilon_1(\mathbf{c}^{\downarrow h_1}, \mathbf{x})\rho_1(\mathbf{c}^{\downarrow g_1}, \mathbf{x}) \ldots \rho_k(\mathbf{c}^{\downarrow g_k}, \mathbf{x})\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}}) \ldots \rho_n(\mathbf{c}^{\downarrow g_n})] \mid \mathbf{x} \in \mathcal{W}_X\}$$

$$= \max\{[\upsilon_1(\mathbf{c}^{\downarrow h_1}, \mathbf{x})\rho_1(\mathbf{c}^{\downarrow g_1}, \mathbf{x}) \ldots \rho_k(\mathbf{c}^{\downarrow g_k}, \mathbf{x})] \mid \mathbf{x} \in \mathcal{W}_X\}[\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}}) \ldots \rho_n(\mathbf{c}^{\downarrow g_n})]$$

$$= \max\{\upsilon(\mathbf{c}^{\downarrow h - \{X\}}, \mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_X\}[\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}}) \ldots \rho_n(\mathbf{c}^{\downarrow g_n})]$$

$$= \upsilon^{\downarrow h - \{X\}}(\mathbf{c}^{\downarrow h - \{X\}})[\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}}) \ldots \rho_n(\mathbf{c}^{\downarrow g_n})]$$

$$= \otimes \text{Fus}_X\{\upsilon_1, \rho_1, \ldots \rho_n\}(\mathbf{c})$$

Case 2

Suppose X is a random variable. Without loss of generality, assume that $\upsilon_1, \rho_1, \ldots, \rho_k$ are the only valuations that bear on X. Let $\upsilon = \upsilon_1 \otimes \rho_1 \otimes \ldots \otimes \rho_k$, let $h = h_1 \cup g_1 \cup \ldots \cup g_k$, and let $\mathbf{c} \in \mathscr{W}_{\mathscr{X} - \{X\}}$. Then

$$
\begin{aligned}
&(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow(\mathscr{X} - \{X\})}(\mathbf{c}) \\
&= \Sigma\{[\upsilon_1(\mathbf{c}^{\downarrow h_1}, \mathbf{x})\rho_1(\mathbf{c}^{\downarrow g_1}, \mathbf{x})\ldots\rho_k(\mathbf{c}^{\downarrow g_k}, \mathbf{x})\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}})\ldots\rho_n(\mathbf{c}^{\downarrow g_n})] \mid \mathbf{x} \in \mathscr{W}_X\} \\
&= \Sigma\{[\upsilon_1(\mathbf{c}^{\downarrow h_1}, \mathbf{x})\rho_1(\mathbf{c}^{\downarrow g_1}, \mathbf{x})\ldots\rho_k(\mathbf{c}^{\downarrow g_k}, \mathbf{x})] \mid \mathbf{x} \in \mathscr{W}_X\}[\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}})\ldots\rho_n(\mathbf{c}^{\downarrow g_n})] \\
&= \Sigma\{\upsilon(\mathbf{c}^{\downarrow h - \{X\}}, \mathbf{x}) \mid \mathbf{x} \in \mathscr{W}_X\}[\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}})\ldots\rho_n(\mathbf{c}^{\downarrow g_n})] \\
&= \upsilon^{\downarrow h - \{X\}}(\mathbf{c}^{\downarrow h - \{X\}})[\rho_{k+1}(\mathbf{c}^{\downarrow g_{k+1}})\ldots\rho_n(\mathbf{c}^{\downarrow g_n})] \\
&= \otimes \mathrm{Fus}_X\{\upsilon_1, \rho_1, \ldots \rho_n\}(\mathbf{c})
\end{aligned}
$$

Proof of Theorem 11.1

By definition, $(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow\varnothing}$ is obtained by sequentially marginalizing a minimal variable. A proof of this theorem is obtained by repeatedly applying the result of Lemma 11.4. At each step, we delete a minimal variable and fuse the set of all valuations with respect to the minimal variable. Using Lemma 11.4, after fusion with respect to X_1, the combination of all valuations in the resulting VBS is equal to $(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow(\mathscr{X} - \{X_1\})}$. Again, using Lemma 11.4, after fusion with respect to X_2, the combination of all valuations in the resulting VBS is equal to $(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow(\mathscr{X} - \{X_1, X_2\})}$. And so on. When all the variables have been deleted, there will be a single valuation left. Using Lemma 11.4, this valuation will be $(\otimes\{\upsilon_1, \rho_1, \ldots, \rho_n\})^{\downarrow\varnothing}$.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under grant IRI-8902444. I am grateful for discussions with, and comments from Dan Geiger, Steffen Lauritzen, Pierre Ndilikilikesha, Anthony Neugebauer, Geoff Schemmel, Philippe Smets, Glenn Shafer and Po-Lung Yu.

REFERENCES

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

Bertele, U. and Brioschi, F. (1972) *Nonserial Dynamic Programming*, New York: Academic Press.

Darroch, J.N., Lauritzen, S.L. and Speed, T.P. (1980) Markov fields and log-linear interaction models for contingency tables. *Annals of Statistics*, **8**(3), 522–539.

Dubois, D. and Prade, H. (1990) Inference in possibilistic hypergraphs, in *Proceedings of the Third International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU-90)*, Paris, France, 228–230.

Edwards, D. and Kreiner, S. (1983) The analysis of contingency tables by graphical models. *Biometrika*, **70**, 553–565.

Ezawa, K.J. (1986) Efficient evaluation of influence diagrams. Ph.D. thesis, Department of Engineering-Economic Systems, Stanford University.

Howard, R.A. and Matheson, J.E. (1984) Influence diagrams, in *The Principles and Applications of Decision Analysis*, **2**, R.A. Howard and J.E. Matheson (eds), Menlo Park, CA: Strategic Decisions Group, pp. 719–762.

Jensen, F.V., Olesen, K.G. and Andersen, S.K. (1990) An algebra of Bayesian belief universes for knowledge-based systems. *Networks*, **20**, 637–659.

Kiiveri, H., Speed, T.P. and Carlin, J.B. (1984) Recursive causal models. *Journal of the Australian Mathematics Society, A*, **36**, 30–52.

Kong, A. (1986) Multivariate belief functions and graphical models. Ph.D. thesis, Department of Statistics, Harvard University.

Lauritzen, S.L. and Spiegelhalter, D.J. (1988) Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, B*, **50**(2), 157–224.

Mellouli, K. (1987) On the propagation of beliefs in networks using the Dempster–Shafer theory of evidence. Ph.D. thesis, School of Business, University of Kansas.

Olmsted, S.M. (1983) On representing and solving decision problems. Ph.D. thesis, Department of Engineering-Economic Systems, Stanford University.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan Kaufmann.

Shachter, R.D. (1986) Evaluating influence diagrams. *Operations Research*, **34**(6), 871–882.

Shafer, G. and Shenoy, P.P. (1988) Local computation in hypertrees. Working Paper No. 201, School of Business, University of Kansas.

Shafer, G. and Shenoy, P.P. (1990) Probability propagation. *Annals of Mathematics and Artificial Intelligence*, **2**, 327–352.

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

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

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 *Uncertainty in Artificial Intelligence*, **6**, P.P. Bonissone, M. Henrion, L.N. Kanal and J. Lemmer (eds), Amsterdam: North-Holland, pp. 385–400.

Shenoy, P.P. (1991c) Valuation-based systems: A framework for managing uncertainty in expert systems. In L.A. Zadeh and J. Kacprzyk (eds), *Fuzzy Logic for the Management of Uncertainty* New York: John Wiley & Sons, pp. 83–104.

Shenoy, P.P. (1991d) Valuation networks, decision trees, and influence diagrams: A comparison. Working Paper No. 227, School of Business, University of Kansas.

Shenoy, P.P. (1992) Valuation-based systems for Bayesian decision analysis. *Operations Research*, **40**(3), 463–484.

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

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

Shenoy, P.P. and Shafer, G. (1990) Axioms for probability and belief-function propagation, in *Uncertainty in Artificial Intelligence*, **4**, R.D. Shachter, T.S. Levitt, J.F. Lemmer and L.N. Kanal (eds), Amsterdam: North-Holland, pp. 169–198.

Tatman, J.A. (1986) Decision processes in influence diagrams: Formulation and analysis. Ph.D. thesis, Department of Engineering-Economic Systems, Stanford University.

Wermuth, N. and Lauritzen S.L. (1983) Graphical and recursive models for contingency tables. *Biometrika*, **70**, 537–552.

Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*, Chichester: John Wiley & Sons.

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

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