



ELSEVIER

European Journal of Operational Research 121 (2000) 579–608

EUROPEAN
JOURNAL
OF OPERATIONAL
RESEARCH

www.elsevier.com/locate/orms

Theory and Methodology

Valuation network representation and solution of asymmetric decision problems

Prakash P. Shenoy *

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

Received 24 May 1995; accepted 20 October 1998

Abstract

This paper deals with asymmetric decision problems. We describe a generalization of the valuation network representation and solution technique to enable efficient representation and solution of asymmetric decision problems. The generalization includes the concepts of indicator valuations and effective frames. We illustrate our technique by solving Howard's used car buyer's problem in complete detail. We highlight the contribution of this paper over the symmetric valuation network technique. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Asymmetric decision problems; Valuation networks; Influence diagrams; Indicator valuations; Effective frames

1. Introduction

This paper deals with asymmetric decision problems. An asymmetric decision problem can be defined using its decision tree representation. In a decision tree, a path from the root node to a leaf node is called a *scenario*. We say a decision problem is *asymmetric* if the number of scenarios in a decision tree representation is less than the cardinality of the Cartesian product of the state spaces of all chance and decision variables. In asymmetric decision problems, some scenarios may exclude

either some chance variables, or some decision variables, or both.

A decision problem can be represented formally in several different ways. A traditional representation of a decision problem is a decision tree model. Decision trees have their genesis in the extensive-form game representation defined by von Neumann and Morgenstern (1944). Decision trees graphically depict all possible scenarios. The sequence of variables in each scenario represents information constraints. The decision tree solution technique computes an optimal strategy using local computation, namely the backward recursion method of dynamic programming (Raiffa and Schlaifer, 1961).

A major disadvantage of the decision tree representation technique is the combinatorial

* Tel.: +1-785-864-7551.

E-mail address: pshenoy@ukans.edu (P.P. Shenoy).

explosion of the number of scenarios. Recently, Kirkwood (1993) has described an algebraic representation technique that addresses this problem. Another major disadvantage of the decision tree representation technique is the computation of the conditional probabilities it demands. A problem statement may describe a probability model for chance variables using one set of conditionals, but the decision tree representation may demand a different set of conditionals. The decision tree representation technique does not include an efficient method for computing the necessary conditionals. A brute-force computation of the conditionals by computing the joint may be intractable in problems that have many chance variables. Recently, Shenoy (1998) has proposed using information sets in a decision tree representation to allow the resulting representation – called a game tree representation – to always use the conditionals specified in the problem statement, i.e., no preprocessing is required to represent any decision problem as a game tree.

Another traditional representation is a strategy matrix model. Strategy matrices have their genesis in the normal form game representation defined by von Neumann and Morgenstern (1947). In a strategy matrix model, along one-dimension we list strategies, along the other dimension we list configurations of all chance variables with their respective probabilities, and for each strategy–configuration pair, we list a utility value. Solving a strategy matrix model is straightforward. We simply compute the expected utility for each strategy, and then identify an optimal strategy. Notice that the strategy matrix representation and solution technique is computationally less efficient than decision tree technique. The strategy matrix technique uses global computation both for the computation of the joint probability distribution and for the identification of an optimal strategy, whereas the decision tree technique uses local computation for the latter.

A more recent representation is an influence diagram model. Influence diagrams were initially proposed as a front-end for decision trees (Miller III et al., 1976; Howard and Matheson, 1981). A motivation behind the formulation of influence diagram representation was to find a method for

representing decision problems without any preprocessing. Subsequently, Olmsted (1983) and Shachter (1986) devised a method for solving influence diagrams directly, i.e., without first having to convert influence diagrams to decision trees. The influence diagram solution technique – called arc-reversal – uses local computation both for the computation of the conditionals and for the computation of an optimal strategy.

A most recent representation is a valuation network model (Shenoy, 1992a, 1993). Valuation networks are similar to influence diagrams in many ways. However, there are two main differences. First, unlike influence diagrams, valuation networks do not demand probabilities in the form of conditionals. Any probability model can be represented in the form of probability valuations. Second, the valuation network solution technique – called the fusion algorithm – is slightly more efficient than the arc-reversal technique of influence diagrams since it avoids unnecessary divisions (Shenoy, 1994b). Recently, Ndilikilikisha (1992, 1994) has translated the fusion algorithm into the influence diagram framework.

Both influence diagrams and valuation networks as originally conceived were designed for symmetric decision problems. For asymmetric decision problems, these techniques make an asymmetric problem symmetric by adding variables and dummy configurations to scenarios. In doing so, we increase the computational burden of solving the problem. For this reason, representing and solving asymmetric problems has been the subject of several studies in recent years.

In the influence diagram literature, four techniques have been proposed by Call and Miller (1990), Smith et al. (1993), Fung and Shachter (1990), and Covaliu and Oliver (1995), to deal with asymmetric decision problems. Each of these four techniques is a hybrid of influence diagram and decision tree techniques. In essence, influence diagram representation is used to capture the uncertainty information, and decision tree representation is used to capture the structural asymmetry information.

In this paper, we investigate the use of valuation networks to represent and solve asymmetric

decision problems. The structural asymmetry information is represented by indicator valuations. An indicator valuation is a special type of a probability valuation whose values are restricted to either 0 or 1. Indicator valuations contribute to the computational efficiency in three important ways. First, indicator valuations enable us to reduce the domain of probability valuations. Thus the probability information is broken down into smaller chunks. This means that the fusion algorithm (which uses local computation) operates more locally, and this contributes greatly to improving the computational efficiency in solving problems. Second, we use indicator valuations to define effective frames as subsets of frames of variables. All numeric information is specified only for effective frames. The solution technique is mostly the same as in the symmetric case. The main difference is that all computations are done on the smaller space of effective frames instead of on the larger space of frames. This contributes to the increased efficiency of the solution technique. Third, when restricted to effective frames, the values of indicator valuations are identically one, and therefore indicator valuations can be handled implicitly and this contributes further to the increased efficiency of the solution technique.

We compare the technique proposed here with the symmetric valuation network technique proposed earlier (Shenoy, 1992a). This helps to highlight the added contribution of this paper. Bielza and Shenoy (1996) compare the asymmetric valuation network technique with the influence diagram-based technique of Smith et al. (1993) and the sequential decision diagram technique of Covaliu and Oliver (1995).

An outline of the remainder of the paper is as follows. In Section 2, we give a verbal statement of the used car buyer's problem (Howard, 1962). This is a highly asymmetric decision problem. In Section 3, we describe the valuation network representation method for asymmetric decision problems and illustrate it using the used car buyer's problem. In Section 4, we describe what it means for a valuation network representation to be well-defined, and what it means to solve a well-defined valuation network representation. In Section 5, we describe a fusion algorithm for solving

well-defined valuation network representations. The fusion algorithm described here is an adaptation of the fusion algorithm for symmetric valuation network representations (Shenoy, 1992a). In Section 6, we compare our method to the symmetric valuation network technique (Shenoy, 1992a). Finally, in Section 7, we summarize and conclude with a brief statement of further research on this topic.

2. The used car buyer's problem

In this section, we give a complete statement of the used car buyer's (UCB) problem (Howard, 1962). This problem is highly asymmetric. Howard (1962) describes a decision tree representation and solution of this problem. Smith et al. (1993) describe a representation and solution of this problem based on a generalization of the symmetric influence diagram technique (Howard and Matheson, 1981; Olmsted, 1983; Shachter, 1986; Ezawa, 1986; Tatman, 1986).

A statement of the UCB problem is as follows. Joe is considering buying a used car from a dealer for \$1000. The market price of similar cars with no defects is \$1100. Joe is uncertain whether the particular car he is considering is a "peach" or a "lemon". Of the 10 major subsystems in the car, a peach has a serious defect in only one subsystem, whereas a lemon has a serious defect in six subsystems. The probability that the used car under consideration is a lemon is 0.2. The cost of repairing one defect is \$40, and the cost of repairing six defects is \$200.

For an additional \$60, Joe can buy the car from the dealer with an "anti-lemon guarantee". The anti-lemon guarantee will normally pay for 50% of the repair cost, but if the car is a lemon, then the guarantee will pay 100% of the repair cost.

Before buying the car, Joe has the option of having the car examined by a mechanic for an hour. In this time period, the mechanic offers three alternatives, t_1 , t_2 , and t_3 as follows:

- t_1 : Test the steering subsystem alone at a cost of \$9;
- t_2 : Test the fuel and electrical subsystems for a total cost of \$13;

t_3 : Do a two-test sequence in which Joe can authorize a second test after the result of the first test is known. In this alternative, the mechanic will first test the transmission subsystem at a cost of \$10 and report the results to Joe. If Joe approves, the mechanic will then proceed to test the differential subsystem at an additional cost of \$4.

All tests are guaranteed to find a defect in the subsystems if a defect exists. We assume that Joe’s utility for profit is linear in dollars.

A decision tree representation and solution of this problem is shown in Figs. 1 and 2. Fig. 1 shows the preprocessing of probabilities, and Fig. 2 shows a decision tree representation and solution. The optimal strategy is to do test t_2 ; if both systems are non-defective then buy with no guarantee, else buy with guarantee. The maximum expected utility is \$32.87.

3. Valuation network representation

In this section, we describe the valuation network representation technique and illustrate it

using the UCB problem. The representation technique described here is a generalization of the valuation network representation technique described in Shenoy (1992a, 1993) for symmetric decision problems. To deal with asymmetries in decision problems, we introduce the concepts of indicator valuations and effective frames. An indicator valuation is a special type of a probability valuation, and an effective frame is a subset of a frame.

A valuation network representation is specified at three levels – graphical, dependence, and numeric. This is somewhat analogous to Howard and Matheson’s (1981) relational, functional, and numerical levels of specification in influence diagrams. The graphical and dependence levels have qualitative (or symbolic) knowledge, whereas the numeric level has quantitative knowledge.

3.1. Graphical level

At the graphical level, a valuation network representation consists of a graph called a *valua-*

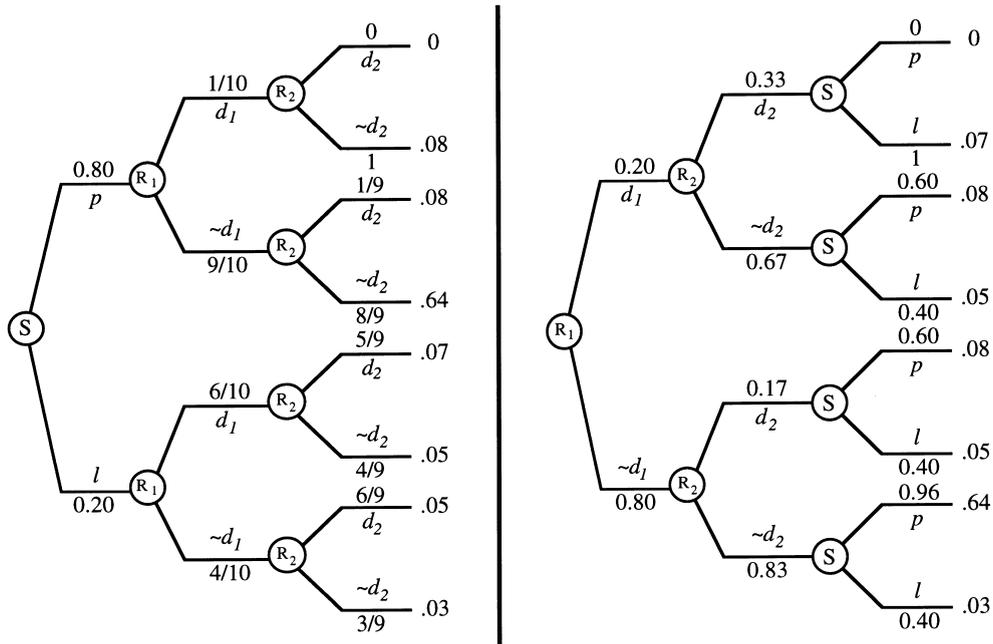


Fig. 1. The preprocessing of probabilities in the UCB problem.

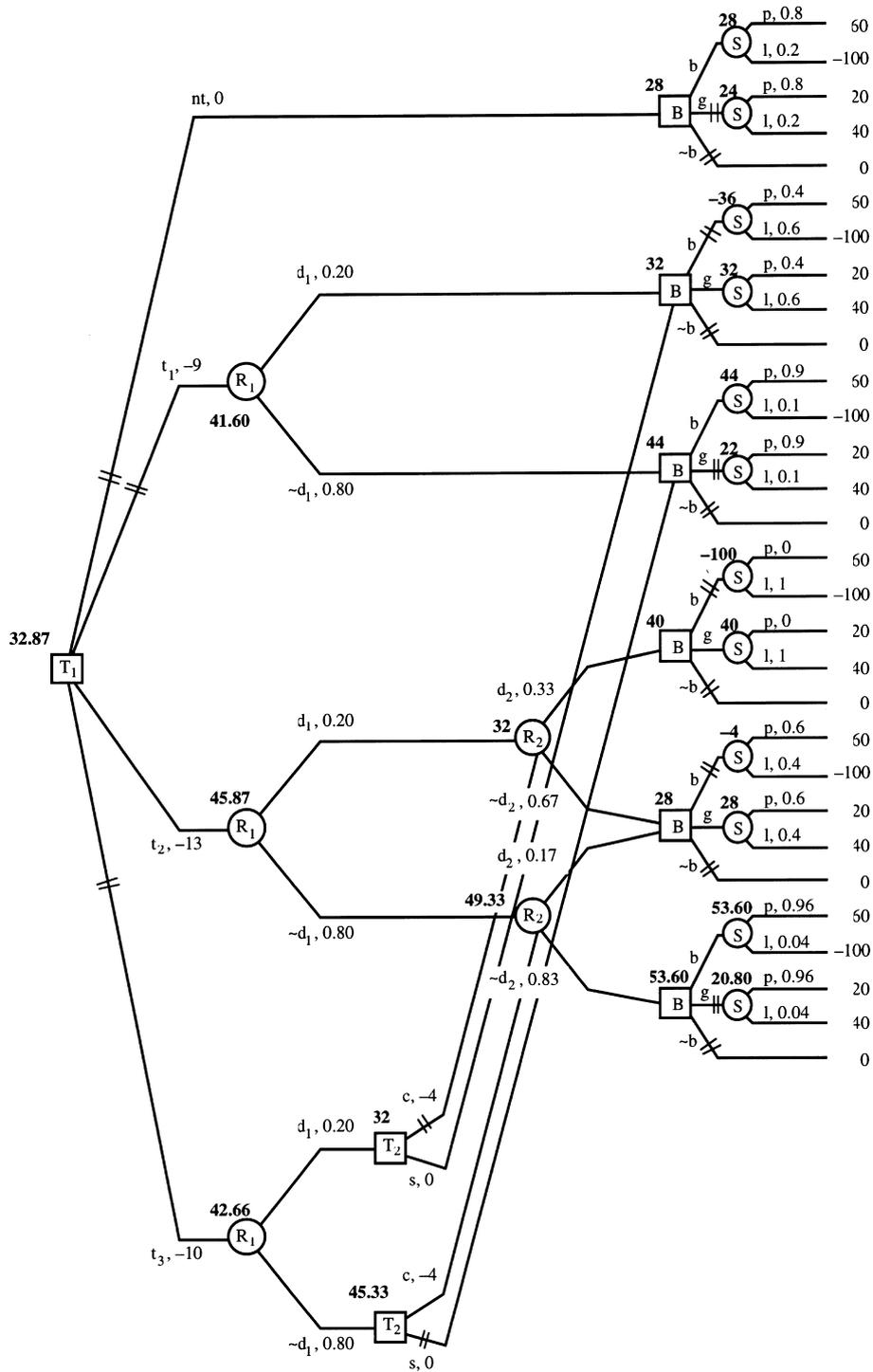


Fig. 2. A decision tree representation and solution of the UCB problem.

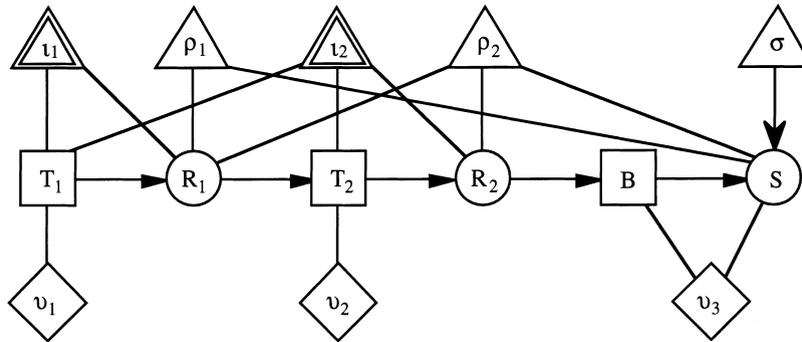


Fig. 3. A valuation network for the UCB problem.

tion network. Fig. 3 shows a valuation network for the UCB problem. A valuation network consists of two types of nodes – variable and valuation. Variables are further classified as either decision or chance, and valuations are further classified as either indicator, or probability, or utility. Thus in a valuation network, there are five different types of nodes—decision, chance, indicator, probability, and utility. Using knowledge-based systems terminology, the collection of decision and chance variables constitute a description of the problem at the propositional level, and the collection of indicator, probability, and utility valuations constitute a description of the problem at the knowledge level.

Decision nodes: Decision nodes correspond to decision variables and are depicted by rectangles. In the UCB problem, there are three decision nodes labeled T_1 , T_2 , and B . T_1 represents the first test decision, T_2 represents the second test decision, and B represents the buy car decision. The values of the decision variables (i.e., the alternatives) are not shown in the valuation network. The set of values of a variable is called the *frame* for that variable. The frames of decision variables are specified at the dependence level description of the valuation network representation (in Section 3.2).

Chance nodes: Chance nodes correspond to chance variables and are depicted by circles. In the UCB problem, there are three chance nodes labeled R_1 , R_2 , and S . R_1 represents the first test results, R_2 represents the second test results, and S

represents the state of the car. As in the case of decision variables, the values of chance variables are not shown in the valuation network. The frames of chance variables are specified in the dependence level description of the valuation network representation (in Section 3.2).

Let \mathcal{X}_D denote the set of all decision variables, let \mathcal{X}_R denote the set of all chance variables, and let $\mathcal{X} = \mathcal{X}_D \cup \mathcal{X}_R$ denote the set of all variables. In this paper we are concerned only with the case where \mathcal{X} is finite. We use upper-case italic alphabets to denote variables.

Indicator valuations: Indicator valuations represent qualitative constraints on the joint frames of decision and chance variables and are depicted by double-triangular nodes. The set of variables directly connected to an indicator valuation by undirected edges constitutes the domain of the indicator valuation. In the UCB problem, there are two indicator valuations labeled l_1 and l_2 . l_1 's domain is $\{T_1, R_1\}$, and l_2 's domain is $\{T_1, T_2, R_2\}$. l_1 represents the constraint that first test result is not available if Joe decides not to do any of three tests proposed by the mechanic at T_1 . l_2 represents the constraints that at T_2 , the option to stop or continue is available only if Joe decides on test t_3 proposed by the mechanic at T_1 , and that the second test result is available only if Joe decides on test t_2 at T_1 , or if he decides on test t_3 at T_1 and decides to continue at T_2 . The details of the indicator valuations are specified at the dependence level description of the valuation network representation (in Section 3.2).

Utility valuations: Utility valuations represent factors of the joint utility function and are depicted by diamond-shaped nodes in valuation networks. The set of variables directly connected to a utility valuation constitutes the domain of the utility valuation. Depending on whether the utility function decomposes additively or multiplicatively, the factors are additive or multiplicative (or perhaps some combination of the two). In the UCB problem, there are three additive utility valuations labeled v_1 , v_2 , and v_3 . v_1 's domain is $\{T_1\}$, v_2 's domain is $\{T_2\}$, and v_3 's domain is $\{B, S\}$. v_1 represents the cost of the first test, v_2 represents the cost of the second test, and v_3 represents the value of the car less the cost of buying the car and repairing the defects. The details of the utility valuations are specified at the numeric level description of the valuation network representation (in Section 3.3).

Probability valuations: Probability valuations represent multiplicative factors of the family of joint probability distributions of the chance variables in the problem, and are depicted by triangular nodes in valuation networks. The set of all variables directly connected to a probability valuation constitutes the domain of the probability valuation. If a probability valuation is a conditional, then this is indicated by making the edges between the probability valuation node and the variables in the head of the conditional directed toward the variables. In the UCB problem, there are three probability valuations labeled σ , ρ_1 , and ρ_2 . σ 's domain is $\{S\}$, ρ_1 's domain is $\{R_1, S\}$, and ρ_2 's domain is $\{R_1, R_2, S\}$. The arrow from σ to S indicates that σ is a conditional for $\{S\}$ given \emptyset . The details of the probability valuations are specified at the numeric level description of the valuation network representation (in Section 3.3). The precise meaning of the probability valuations is given in Section 3.3 and Section 4.2.

Information constraints: The specification of the valuation network at the graphical level includes directed arcs between pairs of distinct variables. These directed arcs represent information constraints. Suppose R is a chance variable and suppose D is a decision variable. An arc $R \rightarrow D$ means that the true value of R is known to the

decision maker (DM) when the DM chooses an alternative from D 's frame, and an arc from $D \rightarrow R$ means that the true value of R is not known to the DM at the time the DM has to choose an alternative from D 's frame.

Well-defined information constraints: We say the information constraints are *well-defined* if they satisfy the following condition: For any chance variable R and for any decision variable D , either there is a directed path from R to D , or there is a directed path from D to R , but not both. (We say there is a *directed path from X to Y* if either $X \rightarrow Y$ or there exists Z_1, \dots, Z_k for some $k \geq 1$ such that $X \rightarrow Z_1, Z_1 \rightarrow Z_2, \dots, Z_{k-1} \rightarrow Z_k, Z_k \rightarrow Y$.)

The rationale behind the definition of well-defined information constraints is as follows. The information constraints are used during the solution phase to solve the valuation network representation. The information constraints dictate what deletion sequences are valid for deleting variables during the solution phase. If there is a directed path from D to R , then R must be deleted before D , and if there is a directed path from R to D , then D must be deleted before R . If there is neither a directed path from D to R nor a directed path from R to D , then depending on whether R is deleted before D or not, we may get different solutions to the decision problem. On the other hand, if there is a directed path from D to R and a directed path from R to D , then there does not exist a valid deletion sequence, and we are unable to solve the problem.

We assume that the DM has perfect recall. This means that if there is a directed path from R to D , then the DM knows the true value of R at the time the DM has to choose an alternative from D 's frame, and if there is a directed path from D to R , then this means that the DM does not know the true value of R at the time the DM has to choose an alternative from D 's frame. In the UCB problem, for example, at the time Joe makes a buy decision, he knows the results of the first test and the second test, but not the state of the car. If the information constraints are well-defined, then given any chance variable R and any decision variable D , either the DM knows the true value of R (at the time the DM has to choose an alternative

from D 's frame) or not. If the information constraints are not well-defined, then this means that either the information constraints are incompletely specified or that the information constraints are contradictory.

3.2. Dependence level

Next, we specify valuation network representation at the dependence level. Like the graphical level, specification of a valuation network representation at the dependence level involves only qualitative (or symbolic) knowledge. In this sense, the dependence level of a valuation network representation is different from the function level of influence diagram representation since the latter may involve arithmetic operations.

Frames: Associated with each variable X is a frame \mathcal{W}_X . We assume that all variables have finite frames. The frame of a decision variable consists of alternatives available to the DM. The frame of a chance variable consists of all mutually exclusive and exhaustive values that the chance variable can assume. We use the terminology “frame” (as opposed to “sample space”) to emphasize the fact that the possibilities that comprise a frame are neither determined nor meaningful independent of our knowledge (Shafer, 1976, p. 36).

In the UCB problem, $\mathcal{W}_{T_1} = \{nt, t_1, t_2, t_3\}$, where nt denotes no test, and t_i denotes the i th test option offered by the mechanic, $i=1, 2$, and 3 ; $\mathcal{W}_{R_1} = \{n_1, d_1, \sim d_1\}$, where n_1 denotes no result, d_1 denotes the first subsystem tested is defective, and $\sim d_1$ denotes the first subsystem tested is non-defective; $\mathcal{W}_{T_2} = \{nc, s, c\}$, where nc denotes no choice, s denotes stop, and c denotes continue; $\mathcal{W}_{R_2} = \{n_2, d_2, \sim d_2\}$, where n_2 denotes no result, d_2 denotes the second subsystem tested is defective, and $\sim d_2$ denotes the second subsystem tested is non-defective; $\mathcal{W}_B = \{b, g, \sim b\}$, where b denotes buy with no guarantee, g denotes buy with guarantee, and $\sim b$ denotes do not buy; and $\mathcal{W}_S = \{p, l\}$ where p denotes the car is a peach, and l denotes the car is a lemon.

Configurations: We often deal with non-empty subsets of variables in \mathcal{X} . Given a non-empty subset h of \mathcal{X} , let \mathcal{W}_h denote the Cartesian product

of \mathcal{W}_X for X in h , i.e., $\mathcal{W}_h = \prod \{\mathcal{W}_X | X \in h\}$. We can think of \mathcal{W}_h as the set of possible values of the joint variable h . Accordingly, we call \mathcal{W}_h the *frame for h* . Also, we refer to elements of \mathcal{W}_h as *configurations of h* . We use this terminology even when h consists of a single variable, say X . Thus we refer to elements of \mathcal{W}_X as *configurations of X* . We use lower-case, bold-faced letters such as \mathbf{x} , \mathbf{y} , etc., to denote configurations. Also, if \mathbf{x} is a configuration of g , \mathbf{y} is a configuration of h , and $g \cap h = \emptyset$, then (\mathbf{x}, \mathbf{y}) denotes a configuration of $g \cup h$.

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

Indicator valuations: Earlier we described indicator valuations as qualitative constraints on the joint frames of variables. Here, we define them formally. Suppose s is a subset of variables. An *indicator valuation for s* is a function $v: \mathcal{W}_s \rightarrow \{0, 1\}$. The only values assumed by an indicator valuation are 0 and 1, hence the term indicator valuation. The values of indicator valuations can be treated as (degenerate) probabilities. An efficient way of representing an indicator valuation is simply to describe the elements of the frame that have value 1, i.e., we represent v by Ω_v where $\Omega_v = \{\mathbf{x} \in \mathcal{W}_s | v(\mathbf{x}) = 1\}$. Obviously, $\Omega_v \subseteq \mathcal{W}_s$. To minimize jargon, we also call Ω_v an indicator valuation for s .

In the UCB problem, we have two indicator valuations – t_1 (or Ω_{t_1}) with domain $\{T_1, R_1\}$, and t_2 (or Ω_{t_2}) with domain $\{T_1, T_2, R_2\}$. These indicator valuations are specified as follows:

$$\Omega_{t_1} = \{(nt, n_1), (t_1, d_1), (t_1, \sim d_1), (t_2, d_1), \\ (t_2, \sim d_1), (t_3, d_1), (t_3, \sim d_1)\};$$

and

$$\Omega_{t_2} = \{(nt, nc, n_2), (t_1, nc, n_2), (t_2, nc, d_2), \\ (t_2, nc, \sim d_2), (t_3, s, n_2), (t_3, c, d_2), \\ (t_3, c, \sim d_2)\}.$$

t_1 represents the constraint that the first test result is not available if Joe decides not to do any of the three tests proposed by the mechanic at T_1 . t_2 represents the constraints that at T_2 , the option to stop or continue is available only if Joe decides on test t_3 proposed by the mechanic at T_1 , and that the second test result is available only if Joe decides on test t_2 at T_1 , or if he decides on test t_3 at T_1 and decides to continue at T_2 . Fig. 4 depicts the two indicator valuations graphically.

Some indicator valuations constrain domains of chance variables. For example, t_1 constrains the domain of R_1 . Some indicator valuations constrain domains of decision variable and some indicator valuations constrain the domains of both decision and chance variables. For example t_2 constrains the domains of T_2 and R_2 .

The concept of indicator valuations is crucial to the computational efficiency of the solution technique. Using the indicator valuations in a problem, we define the “effective frame” for a subset of variables. The effective frame for s is a subset of the frame for s . The increased computational efficiency of the solution technique is partly the result of working on effective frames instead of working

on frames. In the remainder of this subsection, we introduce some notation and definitions to enable us to define the effective frame for a subset of variables.

Projection of configurations: Projection of configurations simply means dropping extra coordinates; if $(t_3, d_1, c, \sim d_2)$ is a configuration of $\{T_1, R_1, T_2, R_2\}$, for example, then the projection of $(t_3, d_1, c, \sim d_2)$ to $\{T_1, R_1\}$ is simply (t_3, d_1) , which is a configuration of $\{T_1, R_1\}$.

If g and h are sets of variables, $h \subseteq g$, and \mathbf{x} is a configuration of g , then let $\mathbf{x}^{|h}$ denote the projection of \mathbf{x} to h . The projection $\mathbf{x}^{|h}$ is always a configuration of h . If $h = g$ and \mathbf{x} is a configuration of h , then $\mathbf{x}^{|h} = \mathbf{x}$. If $h = \emptyset$, then $\mathbf{x}^{|\emptyset} = \diamond$.

Marginalization of indicator valuations: Suppose Ω_a is an indicator valuation for a , and suppose $b \subseteq a$. The marginalization of Ω_a to b , denoted by $\Omega_a^{|b}$, is an indicator valuation for b given by

$$\Omega_a^{|b} = \{\mathbf{x} \in \mathcal{W}_b \mid (\mathbf{x}, \mathbf{y}) \in \Omega_a \text{ for some } \mathbf{y} \in \mathcal{W}_{a-b}\}.$$

To illustrate this definition, consider the indicator valuation Ω_{t_2} for $\{T_1, T_2, R_2\}$ in the UCB

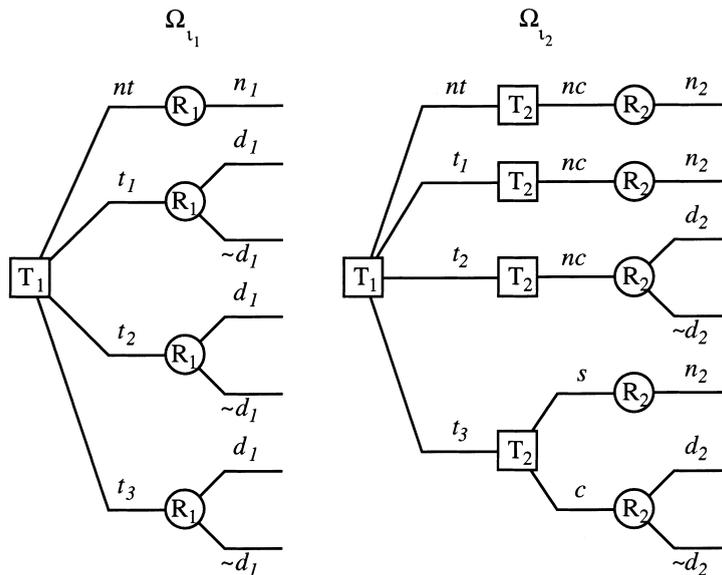


Fig. 4. A graphical depiction of Ω_{t_1} and Ω_{t_2} .

problem. The marginal of Ω_{i_2} for $\{T_1, T_2\}$ is given by the indicator valuation

$$\Omega_{i_2}^{\downarrow\{T_1, T_2\}} = \{(nt, nc), (t_1, nc), (t_2, nc), (t_3, s), (t_3, c)\}.$$

Combination of indicator valuations: Suppose Ω_{i_a} is an indicator valuation for a , and suppose Ω_{i_b} is an indicator valuation for b . The combination of Ω_{i_a} and Ω_{i_b} , denoted by $\Omega_{i_a} \otimes \Omega_{i_b}$, is an indicator valuation for $a \cup b$ given by

$$\Omega_{i_a} \otimes \Omega_{i_b} = \{\mathbf{x} \in \mathcal{W}_{a \cup b} \mid \mathbf{x}^{i_a} \in \Omega_{i_a} \text{ and } \mathbf{x}^{i_b} \in \Omega_{i_b}\}.$$

To illustrate this definition, consider the two indicator valuations Ω_{i_1} and Ω_{i_2} in the UCB problem. The combination $\Omega_{i_1} \otimes \Omega_{i_2}$ is an indicator valuation for $\{T_1, R_1, T_2, R_2\}$ given as follows:

$$\begin{aligned} \Omega_{i_1} \otimes \Omega_{i_2} &= \{(nt, n_1, nc, n_2), (t_1, d_1, nc, n_2), \\ &(t_1, \sim d_1, nc, n_2), (t_2, d_1, nc, d_2), (t_2, d_1, nc, \sim d_2), \\ &(t_2, \sim d_1, nc, d_2), (t_2, \sim d_1, nc, \sim d_2), \\ &(t_3, d_1, s, n_2), (t_3, d_1, c, d_2), (t_3, d_1, c, \sim d_2), \\ &(t_3, \sim d_1, s, n_2), (t_3, \sim d_1, c, d_2), \\ &(t_3, \sim d_1, c, \sim d_2)\}. \end{aligned}$$

Effective frames: Suppose $\{\Omega_{i_1}, \dots, \Omega_{i_p}\}$ is the set of indicator valuations in a given problem such that Ω_{i_j} is an indicator valuation for s_j , $j = 1, \dots, p$. Without loss of generality, assume that $s_1 \cup \dots \cup s_p = \mathcal{X}$. (If a variable, say X , is not included in the domain of some indicator valuation, include the vacuous indicator valuation Ω_i for $\{X\}$, i.e., $\Omega_i = \mathcal{W}_X$.) Suppose s is a subset of variables. The effective frame for s , denoted by Ω_s , is given by

$$\Omega_s = (\otimes \{\Omega_{i_k} \mid s_k \cap s \neq \emptyset\})^{\downarrow s}.$$

In words, the effective frame for s is defined in two steps as follows. First we combine indicator valuations whose domains include a variable in s . Second, we marginalize the resulting combination to eliminate variables not in s .

To illustrate this definition, consider the indicator valuations Ω_{i_1} for $\{T_1, R_1\}$, and Ω_{i_2} for $\{T_1, T_2, R_2\}$. Since B and S are not included in the

domains of the two indicator valuations, we need to introduce vacuous indicator valuations Ω_{i_3} for B and Ω_{i_4} for $\{S\}$. Then, for example, the effective frame for R_1, R_2, S is given by

$$\begin{aligned} \Omega_{\{R_1, R_2, S\}} &= (\Omega_{i_1} \otimes \Omega_{i_2} \otimes \Omega_{i_4})^{\downarrow\{R_1, R_2, S\}} \\ &= \{(n_1, n_2, p), (n_1, n_2, l), (d_1, n_2, p), \\ &(d_1, n_2, l), (d_1, d_2, p), (d_1, d_2, l), \\ &(d_1, \sim d_2, p), (d_1, \sim d_2, l), (\sim d_1, n_2, p), \\ &(\sim d_1, n_2, l), (\sim d_1, d_2, p), (\sim d_1, d_2, l), \\ &(\sim d_1, \sim d_2, p), (\sim d_1, \sim d_2, l)\}. \end{aligned}$$

As we will see shortly, all the numeric information in probability and utility valuations are specified on effective frames only. Also, in the solution phase, all numerical computations are done on effective frames. The definitions of marginalization and combination of indicator valuations allow us to define effective frames in terms of indicator valuations. In Section 5, we describe an efficient method to compute effective frames using local computation.

3.3. Numeric level

Finally, we specify a valuation network at the numeric level. At this level, we specify the details of the utility and probability valuations.

Utility valuations: Suppose $u \subseteq \mathcal{X}$. A utility valuation v for u is a function $v: \Omega_u \rightarrow \mathbb{R}$, where \mathbb{R} is the set of real numbers. The values of v are utilities. If v is a utility valuation for u , we say u is the domain of v .

Table 1
Utility valuations in the UCB problem

Ω_{T_1}	v_1	Ω_{T_2}	v_2	$\Omega_{\{B, S\}}$	v_3	
nt	0	nc	0	b	p	60
t_1	-9	c	-4	b	l	-100
t_2	-13	s	0	g	p	20
t_3	-10			g	l	40
				$\sim b$	p	0
				$\sim b$	l	0

In the UCB problem, there are three utility valuations v_1 for $\{T_1\}$, v_2 for $\{T_2\}$, and v_3 for $\{B, S\}$. Table 1 shows the details of these utility valuations.

What are the semantics of utility valuations? Each utility valuation is a “factor” of the joint utility valuation. Naturally, we need to specify how the utility valuations combine to define the joint utility valuation. In the UCB problem, the utility valuations combine by pointwise addition. A formal definition is as follows.

Combination of utility valuations: Suppose h and g are subsets of \mathcal{X} , suppose v_i is a utility valuation for h , and suppose v_j is a utility valuation for g . Then the combination of v_i and v_j , denoted by $v_i \otimes v_j$, is a utility valuation for $h \cup g$ defined as follows:

$$(v_i \otimes v_j)(\mathbf{x}) = v_i(\mathbf{x}^{\downarrow h}) + v_j(\mathbf{x}^{\downarrow g})$$

for all $\mathbf{x} \in \Omega_{h \cup g}$. Thus combination of utility valuations is pointwise addition. This definition assumes of course that v_i and v_j are additive factors of $v_i \otimes v_j$. (If v_i and v_j were multiplicative factors, we would have defined combination as pointwise multiplication.)

In the UCB problem, the joint utility function is given by the utility valuation $v_1 \otimes v_2 \otimes v_3$ for $\{T_1, T_2, B, S\}$. As we will see, it is not necessary to compute the joint utility valuation to solve a decision problem. The only reason for defining the

joint utility valuation is to explain what each utility valuation represents, namely a factor of the joint.

Probability valuations: Suppose $r \subseteq \mathcal{X}$. A *probability valuation* ρ for r is a function $\rho: \Omega_r \rightarrow [0, 1]$. The values of ρ are probabilities. If ρ is a valuation for r , then we say r is the *domain* of ρ .

In the UCB problem, there are three probability valuations: σ for $\{S\}$, ρ_1 for $\{S, R_1\}$, and ρ_2 for $\{S, R_1, R_2\}$. Table 2 shows the details of these probability valuations.

What are the semantics of probability valuations? In general, each probability valuation is a factor of the joint probability valuation. Naturally, we need to specify how the probability valuations combine to define the joint probability valuation. Also, to define conditionals – a special type of a probability valuation – we need to define marginalization of probability valuations and division for probability valuations. After we define these terms, we will explain precisely what the three probability valuations in the UCB problem mean.

Combination of probability valuations: Probability theory defines combination as an operation that combines probability valuations by pointwise multiplication. A formal definition is as follows. Suppose h and g are subsets of \mathcal{X} , suppose ρ_i is a probability valuation for h , and suppose ρ_j is a probability valuation for g . Then the combination

Table 2
Probability valuations in the UCB problem

Ω_S	σ	$\Omega_{\{S, R_1\}}$	ρ_1	$\Omega_{\{S, R_1, R_2\}}$	ρ_2
p	0.80	$p \quad n_1$	1	$p \quad n_1 \quad n_2$	1
l	0.20	$p \quad d_1$	1/10	$p \quad d_1 \quad n_2$	1
		$p \quad \sim d_1$	9/10	$p \quad \sim d_1 \quad n_2$	1
		$l \quad n_1$	1	$p \quad d_1 \quad d_2$	0/9
		$l \quad d_1$	6/10	$p \quad d_1 \quad \sim d_2$	9/9
		$l \quad \sim d_1$	4/10	$p \quad \sim d_1 \quad d_2$	1/9
				$p \quad \sim d_1 \quad \sim d_2$	8/9
				$l \quad n_1 \quad n_2$	1
				$l \quad d_1 \quad n_2$	1
				$l \quad \sim d_1 \quad n_2$	1
				$l \quad d_1 \quad d_2$	5/9
				$l \quad d_1 \quad \sim d_2$	4/9
				$l \quad \sim d_1 \quad d_2$	6/9
				$l \quad \sim d_1 \quad \sim d_2$	3/9

of ρ_i and ρ_j , denoted by $\rho_i \otimes \rho_j$ or $\rho_j \otimes \rho_i$, is a probability valuation for $h \cup g$ such that

$$(\rho_i \otimes \rho_j)(\mathbf{x}) = (\rho_j \otimes \rho_i)(\mathbf{x}) = \rho_i(\mathbf{x}^{lh})\rho_j(\mathbf{x}^{lg})$$

for all $\mathbf{x} \in \Omega_{h \cup g}$.

Marginalization of probability valuations: Probability theory defines marginalization as an operation that reduces the domain of a probability function. Suppose h is a subset of \mathcal{X} containing (chance or decision) variable X , and suppose ρ is a probability valuation for h . The *marginal of ρ for $h - \{X\}$* , denoted by $\rho^{l(h - \{X\})}$, is a probability valuation for $h - \{X\}$ such that

$$\rho^{l(h - \{X\})}(\mathbf{c}) = \sum\{\rho(\mathbf{c}, \mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_h\}$$

for all $\mathbf{c} \in \Omega_{h - \{X\}}$.

Suppose ρ is a probability valuation for r . We say ρ is a *probability distribution* for r if and only if $\rho^{l\emptyset}(\diamond) = 1$, i.e., if the values of ρ sum to 1.

Division: Finally, probability theory defines a division operation (for defining conditionals). Suppose α is a probability valuation for g , and suppose $h \subseteq g$. Then we define α/α^{lh} , called α *divided by α^{lh}* , to be a probability valuation for g defined as follows:

$$(\alpha/\alpha^{lh})(\mathbf{r}) = \alpha(\mathbf{r})/\alpha^{lh}(\mathbf{r}^{lh})$$

for all $\mathbf{r} \in \Omega_g$. If $\alpha(\mathbf{r}) = \alpha^{lh}(\mathbf{r}^{lh}) = 0$, then we consider $(\alpha/\alpha^{lh})(\mathbf{r}) = 0$. In all other respects, the division in the right-hand side of the definition above should be interpreted as the usual division of two real numbers. Since $\alpha^{lh}(\mathbf{r}^{lh}) \geq \alpha(\mathbf{r}) \geq 0$ for all $\mathbf{r} \in \Omega_g$, the values of α/α^{lh} lie in the interval $[0, 1]$.

Conditionals: Suppose α is a probability distribution for $a \cup b$, where a and b are disjoint subsets of variables. We call $\chi = \alpha/\alpha^{lb}$ a *conditional for a given b* . We call a the *head* of (the domain of) χ , and we call b the *tail* of χ . Suppose $\chi = \alpha/\alpha^{lb}$ is a conditional for a given b . Then it is easy to see that χ^{lb} is an identity for α , i.e., $\alpha \otimes \chi^{lb} = \alpha$. Also, it is easy to see that $\chi/\chi^{lb} = \chi$. We will exploit these properties of conditionals to avoid unnecessary division operations in the fusion algorithm. If a probability valuation is a conditional by itself, then this is specified in the

valuation network by directing the edges between a conditional and its head toward the head. In the UCB problem, for example, σ is a conditional for $\{S\}$ given \emptyset .

Combination of an indicator and a probability valuation: Suppose h and g are subsets of \mathcal{X} , suppose ι is an indicator valuation for h , and suppose ρ is a probability valuation for g . Then the *combination of ι and ρ* , denoted by $\iota \otimes \rho$ or $\rho \otimes \iota$, is a probability valuation for $h \cup g$ defined as follows:

$$(\iota \otimes \rho)(\mathbf{x}) = (\rho \otimes \iota)(\mathbf{x}) = \rho(\mathbf{x}^{lg})$$

for all $\mathbf{x} \in \Omega_{h \cup g}$. In combining an indicator valuation and a probability valuation, there is no computation involved since the values of an indicator valuation are identically one over the effective frame. However, indicator valuations do contribute domain information to the combination, i.e., the domain of $\iota \otimes \rho$ is $h \cup g$ whereas the domain of ρ is g .

Semantics of probability valuations in the UCB problem: In the UCB problem, the probability valuation σ for $\{S\}$ is the marginal for S . The probability valuation $\rho_1 \otimes \iota_1$ for $\{T_1, R_1, S\}$ is a conditional for $\{R_1\}$ given $\{S, T_1\}$ (see Table 3). The values of this conditional are specified implicitly in the statement of the problem. The conditional for $\{R_1\}$ given $\{S, T_1\}$ factors into ρ_1 and ι_1 . ρ_1 contains the numeric information of this conditional, and ι_1 contains the structural information of this conditional. ρ_1 can be given precise semantics as follows. For $\mathbf{x} \in \Omega_{\{R_1, S\}}$, $\rho_1(\mathbf{x})$ represents the conditional probability of \mathbf{x}^{lR_1} given \mathbf{x}^{lS} and the fact that \mathbf{x} is not ruled out by structural constraints. For example, $\rho_1(p, n_1) = 1$ because since no result (n_1) is not ruled out by structural constraints, and it is the only value of R_1 possible in this circumstance, its conditional probability must be 1.

The probability valuation $\rho_2 \otimes \iota_2$ for $\{T_1, R_1, T_2, R_2, S\}$ is a conditional for $\{R_2\}$ given $\{S, T_1, R_1, T_2\}$ (see Table 3). The values of this conditional are specified implicitly in the problem statement. As in the previous case, this conditional factors into ρ_2 and ι_2 . Also, using the same argument as in the case of σ , $\rho_2(\mathbf{x})$ represents the

Table 3
The conditionals in the UCB problem

$\Omega_{\{S,T_1,R_1\}}$			$\rho_1 \otimes t_1$	$\Omega_{\{S,T_1,R_1,T_2,R_2\}}$					$\rho_2 \otimes t_2$
p	nt	n_1	1	p	nt	n_1	nc	n_2	1
p	t_1	d_1	1/10	p	t_1	d_1	nc	n_2	1
p	t_1	$\sim d_1$	9/10	p	t_1	$\sim d_1$	nc	n_2	1
p	t_2	d_1	1/10	p	t_2	d_1	nc	n_2	1
p	t_2	$\sim d_1$	9/10	p	t_2	$\sim d_1$	nc	n_2	1
p	t_3	d_1	1/10	p	t_3	d_1	c	d_2	0/9
p	t_3	$\sim d_1$	9/10	p	t_3	d_1	c	$\sim d_2$	9/9
l	nt	n_1	1	p	t_3	d_1	s	n_2	1
l	t_1	d_1	6/10	p	t_3	$\sim d_1$	c	d_2	1/9
l	t_1	$\sim d_1$	4/10	p	t_3	$\sim d_1$	c	$\sim d_2$	8/9
l	t_2	d_1	6/10	p	t_3	$\sim d_1$	s	n_2	1
l	t_2	$\sim d_1$	4/10	l	nt	n_1	nc	n_2	1
l	t_3	d_1	6/10	l	t_1	d_1	nc	n_2	1
l	t_3	$\sim d_1$	4/10	l	t_1	$\sim d_1$	nc	n_2	1
				l	t_2	d_1	nc	n_2	1
				l	t_2	$\sim d_1$	nc	n_2	1
				l	t_3	d_1	c	d_2	5/9
				l	t_3	d_1	c	$\sim d_2$	4/9
				l	t_3	d_1	s	n_2	1
				l	t_3	$\sim d_1$	c	d_2	6/9
				l	t_3	$\sim d_1$	c	$\sim d_2$	3/9
				l	t_3	$\sim d_1$	s	n_2	1

conditional probability of $\mathbf{x}^{\setminus R_2}$ given $\mathbf{x}^{\setminus \{S,R_1\}}$ and the fact that $\mathbf{x}^{\setminus R_2}$ is not ruled out by structural constraints.

Summary: We have now completely defined a valuation network representation of a decision problem. In summary, a valuation network representation of a decision problem Δ consists of decision variables, chance variables, indicator valuations, probability valuations, utility valuations, and information constraints,

$$\Delta = \{ \mathcal{X}_D, \mathcal{X}_R, \{l_1, \dots, l_p\}, \{v_1, \dots, v_m\}, \{\rho_1, \dots, \rho_n\}, \rightarrow \}.$$

In the next section, we describe the meaning and solution of a valuation network representation.

4. Semantics and solution of a valuation network representation

The main goal of this section is to explain the meaning of utility valuations, indicator valuations, and probability valuations. This will enable us to

define when a valuation network representation is well-defined, and what it means to solve a well-defined valuation network representation. Although most of this material is stated in Shenoy (1992a), there are two points of departure. First we need to account for indicator valuations. Second, the definition of a well-defined valuation network representation stated here is weaker than the corresponding definition in Shenoy (1992a). This has an important implication in the solution technique.

This section is organized into two subsections. In Section 4.1, we define the concept of a canonical valuation network. In Section 4.2, we use the concept of a canonical valuation network to define what it means for a valuation network representation to be well-defined and to define what it means to solve a valuation network representation.

4.1. A canonical valuation network representation

A canonical valuation network $\Delta_C = \{ \{D\}, \{R\}, \{v\}, \{\rho\}, \rightarrow \}$ consists of a decision variable D with

a finite frame \mathcal{W}_D , a chance variable R with a finite frame \mathcal{W}_R , a utility valuation v for $\{D, R\}$, a conditional ρ for $\{R\}$ given $\{D\}$, and a precedence relation \rightarrow defined by $D \rightarrow R$. Fig. 5 shows a canonical valuation network and an equivalent decision tree representation.

The meaning of the canonical valuation network is as follows. The elements of \mathcal{W}_D are alternatives, and the elements of \mathcal{W}_R are states of nature. The conditional ρ for $\{R\}$ given $\{D\}$ is a family of probability distributions for R , one for each alternative $\mathbf{d} \in \mathcal{W}_D$. In other words, the probability distribution of chance variable R is conditioned on the alternative \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 $D = \mathbf{d}$. Since ρ is a conditional for $\{R\}$ given $\{D\}$, it follows that

$$\rho^{\{D\}}(\mathbf{d}) = 1 \quad \text{for all } \mathbf{d} \in \mathcal{W}_D. \tag{4.1}$$

The utility valuation v is a conditional utility function – if the decision maker chooses alternative \mathbf{d} and the state of nature \mathbf{r} prevails, then the utility to the decision maker is $v(\mathbf{d}, \mathbf{r})$. 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 alternative.

Before we can describe how to solve a canonical decision problem, we need some defini-

tions. We give formal definitions for combining a probability and a utility valuation, marginalizing a chance variable out of the domain of a utility valuation, and defining a decision function associated with marginalizing a decision variable out of the domain of a utility valuation.

Combination of a utility and a probability valuation: In the previous section, we defined combination of two utility valuations, and combination of two probability valuations. Here we define combination of a utility and a probability valuation, and combination of a utility and an indicator valuation. Suppose h and g are subsets of \mathcal{X} , suppose v is a utility valuation for h , and suppose ρ is a probability valuation for g . Then the *combination of v and ρ* , denoted by $v \otimes \rho$ (or $\rho \otimes v$), is a utility valuation for $h \cup g$ defined as follows:

$$(v \otimes \rho)(\mathbf{x}) = (\rho \otimes v)(\mathbf{x}) = v(\mathbf{x}^{lh})\rho(\mathbf{x}^{lg}) \tag{4.2}$$

for all $\mathbf{x} \in \Omega_h \cup g$.

Combination of a utility and an indicator valuation: Suppose h and g are subsets of \mathcal{X} , suppose v is a utility valuation for h , and suppose ι is an indicator valuation for g . Then the *combination of v and ι* , denoted by $v \otimes \iota$ (or $\iota \otimes v$), is a utility valuation for $h \cup g$ defined as follows:

$$(v \otimes \iota)(\mathbf{x}) = (\iota \otimes v)(\mathbf{x}) = v(\mathbf{x}^{lh}) \tag{4.3}$$

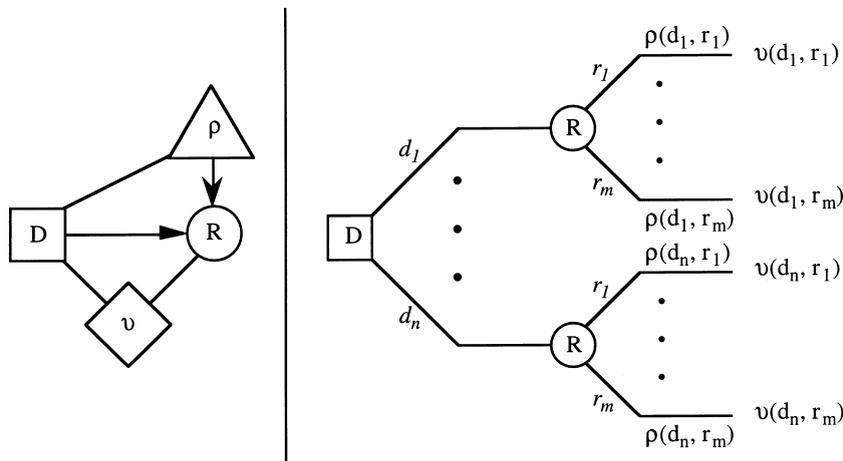


Fig. 5. A canonical valuation network and an equivalent decision tree representation.

for all $\mathbf{x} \in \Omega_{h \cup g}$. In combining an indicator valuation and a utility valuation, there is no computation involved (since the values of an indicator valuation are identically one over the effective frame). However, indicator valuations do contribute domain information to the combination, i.e., the domain of $v \otimes I$ is $h \cup g$ whereas the domain of v is h .

Marginalization of utility valuations: In the previous section, we defined marginalization for probability valuation. Here we define marginalization for utility valuations. The definition of marginalization of utility valuations depends on the nature of the variable being deleted.

Suppose h is a subset of \mathcal{X} containing chance variable R , and suppose α is a utility valuation for h . The *marginal of α for $h - \{R\}$* , denoted by $\alpha^{\downarrow(h-\{R\})}$, is a utility valuation for $h - \{R\}$ defined as follows:

$$\alpha^{\downarrow(h-\{R\})}(\mathbf{c}) = \sum \{ \alpha(\mathbf{c}, \mathbf{r}) \mid \mathbf{r} \in \Omega_R \ni (\mathbf{c}, \mathbf{r}) \in \Omega_h \} \tag{4.4}$$

for all $\mathbf{c} \in \Omega_{h-\{R\}}$.

Suppose h is a subset of \mathcal{X} containing decision variable D , and suppose α is a utility valuation for h . The *marginal of α for $h - \{D\}$* , denoted by $\alpha^{\downarrow(h-\{D\})}$, is a utility valuation for $h - \{D\}$ defined as follows:

$$\alpha^{\downarrow(h-\{D\})}(\mathbf{c}) = \text{MAX} \{ \alpha(\mathbf{c}, \mathbf{d}) \mid \mathbf{d} \in \Omega_D \ni (\mathbf{c}, \mathbf{d}) \in \Omega_h \} \tag{4.5}$$

for all $\mathbf{c} \in \Omega_{h-\{D\}}$.

Decision functions: Each time we marginalize a decision variable D from a utility valuation α for h , we implicitly determine a decision function. Suppose α is a utility valuation for h , and suppose D is a decision variable in h . A *decision function for D with respect to α* is a function $\xi_D: \Omega_{h-\{D\}} \rightarrow \Omega_D$ such that $\xi_D(\mathbf{c}) = \mathbf{d}$ whenever $\alpha^{\downarrow(h-\{D\})}(\mathbf{c}) = \alpha(\mathbf{c}, \mathbf{d})$.

Intuitively, a decision function is rule for selecting an alternative \mathbf{d} from the frame of D on the basis of the information \mathbf{c} about chance and decisions variables in $h - \{D\}$.

A decision function $\xi_D: \Omega_{h-\{D\}} \rightarrow \Omega_D$ can be encoded as an indicator valuation ζ_D for h as follows:

$$\zeta_D(\mathbf{c}, \mathbf{d}) = \begin{cases} 1 & \text{if } \xi_D(\mathbf{c}) = \mathbf{d} \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

for all $(\mathbf{c}, \mathbf{d}) \in \Omega_h$. We also call indicator valuation ζ_D a decision function for D with respect to α .

Strategy: A *strategy* σ is a collection of decision functions, one for each decision variable in \mathcal{X}_D , i.e., $\sigma = \{ \xi_D \}_{D \in \mathcal{X}_D}$. In the canonical valuation network, since there is only one decision variable, a strategy is simply one decision function. Also, since the true value of R is not known when decision D is made, a decision function for D , $\xi_D: \Omega_\emptyset \rightarrow \Omega_D$, is simply a configuration $\mathbf{d} = \xi_D(\blacklozenge)$ in Ω_D .

Suppose $\sigma = \{ \xi_D \}_{D \in \mathcal{X}_D}$ is a strategy, and suppose \mathbf{y} is a configuration of \mathcal{X}_R . Then σ and \mathbf{y} together determine a unique configuration of \mathcal{X}_D . Let $\mathbf{a}_{\sigma, \mathbf{y}}$ denote this unique configuration of \mathcal{X}_D . By definition, $\mathbf{a}_{\sigma, \mathbf{y}}^{\downarrow(D)} = \xi_D(\mathbf{y}^{\downarrow(h-\{D\})})$, where h is the domain of the valuation α with respect to which ξ_D is defined.

Valid deletion sequences: Suppose \rightarrow is a binary relation on \mathcal{X} representing well-defined information constraints. Suppose $g = \{X_1, \dots, X_k\}$ is a subset of \mathcal{X} . We call X_1 a *minimal* variable of g if there does not exist $X_j \in g$ such that $X_1 \rightarrow X_j$. We call $X_1 X_2 \dots X_k$ a *valid deletion sequence* of variables in g if and only if X_1 is a minimal variable in g , X_2 is a minimal variable in $g - \{X_1\}$, and X_{k-1} is a minimal variable in $g - \{X_1, \dots, X_{k-2}\}$.

Marginalizing a subset of variables from utility valuations: Suppose h and g are non-empty subsets of \mathcal{X} such that g is a proper subset of h , suppose α is a utility valuation for h , and suppose \rightarrow is a binary relation on \mathcal{X} representing well-defined information constraints. The *marginal of α for g with respect to the binary relation \rightarrow* , denoted by $\alpha^{\downarrow g}$, is a valuation for g defined as follows:

$$\alpha^{\downarrow g} = \left(\left(\left(\alpha^{\downarrow(h-\{X_1\})} \right)^{\downarrow(h-\{X_1, X_2\})} \right) \dots \right)^{\downarrow(h-\{X_1, X_2, \dots, X_k\})}, \tag{4.7}$$

where $h - g = \{X_1, \dots, X_k\}$, and $X_1 X_2 \dots X_k$ is a valid deletion sequence of variables in $h - g$.

Solving a canonical valuation network: Solving a canonical valuation network using the criterion of maximizing expected utility is easy. Informally, first, we combine ρ and v by pointwise

multiplication. The result is a utility valuation $\rho \otimes v$ for $\{D, R\}$. Next we marginalize R out of $\rho \otimes v$ using addition. The result is a utility valuation $(\rho \otimes v)^{\downarrow\{D\}}$ for D . Each value $(\rho \otimes v)^{\downarrow\{D\}}(\mathbf{d})$ represents expected utility if alternative \mathbf{d} is chosen by the DM. Next, we marginalize D out of $(\rho \otimes v)^{\downarrow\{D\}}$ using maximization. The result is a utility valuation $((\rho \otimes v)^{\downarrow\{D\}})^{\uparrow\emptyset}$ for \emptyset . The value $((\rho \otimes v)^{\downarrow\{D\}})^{\uparrow\emptyset}(\blacklozenge)$ is the maximum expected utility, and an optimal alternative is $\mathbf{d}^* \in \mathcal{W}_D$ such that $(\rho \otimes v)^{\downarrow\{D\}}(\mathbf{d}^*) = ((\rho \otimes v)^{\downarrow\{D\}})^{\uparrow\emptyset}(\blacklozenge)$.

Notice that we marginalize R before D . This is a consequence of the precedence relation $D \rightarrow R$. If the DM had perfect information, then this would be represented in the canonical valuation network by the precedence relation $R \rightarrow D$, and in this case, we would marginalize D before R , and the maximum expected utility is $((\rho \otimes v)^{\downarrow\{R\}})^{\uparrow\emptyset}(\blacklozenge)$. The difference $((\rho \otimes v)^{\downarrow\{R\}})^{\uparrow\emptyset}(\blacklozenge) - ((\rho \otimes v)^{\downarrow\{D\}})^{\uparrow\emptyset}(\blacklozenge)$ is called the *expected value of perfect information*.

In summary, the maximum expected utility associated with an optimal alternative is $(\rho \otimes v)^{\uparrow\emptyset}(\blacklozenge)$, and alternative \mathbf{d}^* is optimal if and only if $(\rho \otimes v)^{\downarrow\{D\}}(\mathbf{d}^*) = (\rho \otimes v)^{\uparrow\emptyset}(\blacklozenge)$.

4.2. Well-defined valuation network representations

Consider a valuation network representation of a decision problem, $\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{l_1, \dots, l_p\}, \{v_1, \dots, v_m\}, \{\rho_1, \dots, \rho_n\}, \rightarrow\}$. We explain the meaning of Δ by reducing it to an equivalent canonical valuation network

$$\Delta_C = \{\{D\}, \{R\}, \{v\}, \{\rho\}, \rightarrow\}.$$

To define Δ_C , we need to define $\mathcal{W}_D, \mathcal{W}_R, v$, and ρ . Define \mathcal{W}_D such that for each distinct strategy σ of Δ , there is a corresponding alternative \mathbf{d}_σ in \mathcal{W}_D . Define \mathcal{W}_R such that for each distinct configuration $\mathbf{y} \in \Omega_{\mathcal{X}_R}$ in Δ , there is a corresponding configuration $\mathbf{r}_\mathbf{y}$ in \mathcal{W}_R .

Before we define utility valuation v for $\{D, R\}$, we need some notation. Consider the joint utility valuation $v_1 \otimes \dots \otimes v_m$ in Δ . For convenience of exposition, suppose that the domain of this utility valuation includes all of \mathcal{X}_D (if not, we can always vacuously extend it so it does). Typically the

domain of this valuation also includes some (or all) chance variables. Let v denote the subset of chance variables included in the domain of the joint utility valuation, i.e., $v \subseteq \mathcal{X}_R$ such that $v_1 \otimes \dots \otimes v_m$ is a utility valuation for $\mathcal{X}_D \cup v$. Define utility valuation v for $\{D, R\}$ such that

$$v(\mathbf{d}_\sigma, \mathbf{r}_\mathbf{y}) = (v_1 \otimes \dots \otimes v_m)(\mathbf{a}_{\sigma, \mathbf{y}}, \mathbf{y}^{\downarrow v}) \tag{4.8}$$

for all strategy σ of Δ , and for all configuration $\mathbf{y} \in \Omega_{\mathcal{X}_R}$. Remember that $\mathbf{a}_{\sigma, \mathbf{y}}$ is the unique configuration of \mathcal{X}_D determined by σ and \mathbf{y} (see Section 4.1).

Well-defined probability valuations: Consider the joint probability valuation $l_1 \otimes \dots \otimes l_p \otimes \rho_1 \otimes \dots \otimes \rho_n$. For convenience of exposition, suppose that the domain of this probability valuation includes all of \mathcal{X}_R (if not, we can always vacuously extend it so it does). Let q denote the subset of decision variables included in the domain of the joint probability valuation, i.e., $q \subseteq \mathcal{X}_D$ such that $l_1 \otimes \dots \otimes l_p \otimes \rho_1 \otimes \dots \otimes \rho_n$ is a probability valuation for $q \cup \mathcal{X}_R$. Note that q could be empty. Define probability valuation ρ for $\{D, R\}$ such that

$$\rho(\mathbf{d}_\sigma, \mathbf{r}_\mathbf{y}) = (l_1 \otimes \dots \otimes l_p \otimes \rho_1 \otimes \dots \otimes \rho_n)(\mathbf{a}_{\sigma, \mathbf{y}}^{\downarrow q}, \mathbf{y}) \tag{4.9}$$

for all strategies σ , and for all configurations $\mathbf{y} \in \Omega_{\mathcal{X}_R}$. ρ is a conditional if it satisfies condition (4.1). This motivates the following definition. We say $\{l_1, \dots, l_p, \rho_1, \dots, \rho_n\}$ is *well-defined* in Δ if and only if

$$\sum_{\mathbf{y} \in \Omega_{\mathcal{X}_R}} \{(l_1 \otimes \dots \otimes l_p \otimes \rho_1 \otimes \dots \otimes \rho_n)(\mathbf{a}_{\sigma, \mathbf{y}}^{\downarrow q}, \mathbf{y})\} = 1 \tag{4.10}$$

for each strategy σ .

In general, checking whether or not a set of indicator and probability valuations are well defined is computationally intensive since we need to verify Eq. (4.10) for each strategy σ , and there are an exponential number of strategies. This is the case in the UCB problem, in which we have 63 distinct strategies. However, if a decision problem is in Howard canonical form (Howard, 1990), i.e., none of the indicator valuations or probability valuations include a decision variable in their domains (i.e., $q = \emptyset$), then (4.10) reduces to simply checking whether $l_1 \otimes \dots \otimes l_p \otimes \rho_1 \otimes \dots \otimes \rho_n$ is a probability distribution, and there are efficient

methods that employ local computation to do this (e.g. Shenoy, 1992b).

Well-defined valuation network representation: Suppose

$$\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{l_1, \dots, l_p\}, \{v_1, \dots, v_m\}, \{\rho_1, \dots, \rho_n\}, \rightarrow\}$$

is a decision problem. We say Δ is *well-defined* if and only if $\{l_1, \dots, l_p, \rho_1, \dots, \rho_n\}$ is well-defined, and \rightarrow is well-defined (see Section 3.1 for what it means for \rightarrow to be well-defined).

In summary, in a valuation network representation Δ of a decision problem, the utility valuations $\{v_1, \dots, v_m\}$ represent the factors of a joint utility function v , and the indicator and probability valuations $\{l_1, \dots, l_p, \rho_1, \dots, \rho_n\}$ represent the factors of a family of joint probability distributions ρ .

Solving a decision problem: Suppose $\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{l_1, \dots, l_p\}, \{v_1, \dots, v_m\}, \{\rho_1, \dots, \rho_n\}, \rightarrow\}$ is a well-defined valuation network representation of a decision problem. Let $\Delta_C = \{\{D\}, \{R\}, \{v\}, \{\rho\}, \rightarrow\}$ denote the equivalent canonical valuation network representation. In the canonical valuation network Δ_C , the two computations that are of interest are (1) the computation of the maximum expected value $(\rho \otimes v)^{\downarrow 0}(\blacklozenge)$; and (2) the computation of an optimal alternative \mathbf{d}_{σ^*} such that $(\rho \otimes v)^{\downarrow \{D\}}(\mathbf{d}_{\sigma^*}) = (\rho \otimes v)^{\downarrow 0}(\blacklozenge)$. Since we know the mapping between Δ and Δ_C , we can formally define what it means to solve decision problem Δ . There are two computations of interest.

First, we would like to compute the maximum expected utility. The maximum expected utility, denoted by u^* , is given by $u^* = ((\otimes\{v_1, \dots, v_m\}) \otimes (\otimes\{l_1, \dots, l_p, \rho_1, \dots, \rho_n\}))^{\downarrow 0}(\blacklozenge)$.

Second, we would like to compute an optimal strategy σ^* that gives us the maximum expected utility u^* . A strategy σ^* of Δ is *optimal* if and only if $(v \otimes \rho)^{\downarrow \{D\}}(\mathbf{d}_{\sigma^*}) = u^*$, where v , ρ , and D refer to the equivalent canonical valuation network representation Δ_C .

5. A fusion algorithm

In this section, we describe a fusion algorithm for solving valuation network representations of

decision problems. The fusion algorithm described here is a slight generalization of the fusion algorithm described in (Shenoy, 1992a). The main difference is in the treatment of indicator valuations. Indicator valuations are accounted for explicitly and through the use of effective frames. All computations are done on effective frames only.

Computing effective frames: If we treat an indicator valuation as a function whose values are either 0 or 1, then marginalization of indicator valuation (defined in Section 3.2) is equivalent to Boolean addition over the frame of deleted variables, and combination of indicator valuations (also defined in Section 3.2) is equivalent to pointwise Boolean multiplication. We have shown elsewhere (Shenoy, 1994a) that Boolean addition and Boolean multiplication satisfy the three axioms that allow the use of local computation in computing marginals (Shenoy and Shafer, 1990). Thus we can use local computation to compute effective frames (see Shenoy (1994a) for a fusion algorithm that can be used for computing effective frames using local computation). Consider, for example, the computation of the effective frame for $\{R_1, R_2, S\}$. In Section 3.2, we saw that the effective frame for $\{R_1, R_2, S\}$ is defined as $(\Omega_{i_1} \otimes \Omega_{i_2} \otimes \Omega_{i_4})^{\downarrow \{R_1, R_2, S\}}$. Notice that $\Omega_{i_1} \otimes \Omega_{i_2} \otimes \Omega_{i_4}$ is an indicator valuation for $\{T_1, R_1, T_2, R_2, S\}$, and in marginalizing the combination to $\{R_1, R_2, S\}$, we delete $\{T_1, T_2\}$. Since T_2 is only in the domain of i_2 , it is easy to show that $(\Omega_{i_1} \otimes \Omega_{i_2} \otimes \Omega_{i_4})^{\downarrow \{R_1, R_2, S\}} = (\Omega_{i_1} \otimes \Omega_{i_2}^{\downarrow \{T_1, R_2\}})^{\downarrow \{R_1, R_2\}} \otimes \Omega_{i_4}$. (A brief explanation is as follows: $\text{Fus}_{T_2}\{\Omega_{i_1}, \Omega_{i_2}, \Omega_{i_4}\} = \{\Omega_{i_1}, \Omega_{i_2}^{\downarrow \{T_1, R_2\}}, \Omega_{i_4}\}$, and $\text{Fus}_{T_1}\{\Omega_{i_1}, \Omega_{i_2}^{\downarrow \{T_1, R_2\}}, \Omega_{i_4}\} = \{(\Omega_{i_1} \otimes \Omega_{i_2}^{\downarrow \{T_1, R_2\}})^{\downarrow \{R_1, R_2\}}, \Omega_{i_4}\}$). The most expensive combination in $(\Omega_{i_1} \otimes \Omega_{i_2} \otimes \Omega_{i_4})^{\downarrow \{R_1, R_2, S\}}$ is on the frame for $\{T_1, R_1, T_2, R_2, S\}$ whereas the most expensive combination in $(\Omega_{i_1} \otimes \Omega_{i_2}^{\downarrow \{T_1, R_2\}})^{\downarrow \{R_1, R_2\}} \otimes \Omega_{i_4}$ is on the frame for $\{T_1, R_1, R_2\}$. Thus it is much more efficient to compute $(\Omega_{i_1} \otimes \Omega_{i_2}^{\downarrow \{T_1, R_2\}})^{\downarrow \{R_1, R_2\}} \otimes \Omega_{i_4}$ than it is to compute $(\Omega_{i_1} \otimes \Omega_{i_2} \otimes \Omega_{i_4})^{\downarrow \{R_1, R_2, S\}}$.

Computation of effective frames is done during the process of solving valuation networks. It is also done prior to specification of the numerical details of the probability and utility valuations. As we saw earlier, the numerical details of the probability and

utility valuations are only specified for effective frames.

Solving valuation networks: We define a fusion operation that is used to solve a valuation network representation of a decision problem. The fusion operation depends on the type of variable being eliminated. Consider a set of valuations consisting of j utility valuations, v_1, \dots, v_j , s indicator valuations, l_1, \dots, l_s , and k probability valuations, ρ_1, \dots, ρ_k . Suppose v_i is a utility valuation for g_i , l_i is an indicator valuation for h_i , and suppose ρ_i is a probability valuation for r_i . Let $\text{Fus}_X\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$ denote the collection of valuations after fusing the valuations in the set $\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$ with respect to variable X . We define $\text{Fus}_X\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$ in two distinct cases as described below.

Case 1: Suppose D is a decision variable in $g_1 \cup \dots \cup g_j \cup h_1 \cup \dots \cup h_s \cup r_1 \cup \dots \cup r_k$. Then $\text{Fus}_D\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$ is defined as follows:

$$\begin{aligned} &\text{Fus}_D\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\} \\ &= \{v_i | D \notin g_i\} \cup \{v_i^{l(g-D)}\} \cup \{l_i | D \notin h_i\} \\ &\quad \cup \{\rho_i | D \notin r_i\} \cup \{(\rho \otimes \zeta_D)^{l((r \cup g) - \{D\})}\}, \end{aligned} \quad (5.1)$$

where

$$\begin{aligned} v &= \otimes\{v_i | D \in g_i\}, \quad g = \cup\{g_i | D \in g_i\}, \\ \rho &= (\otimes\{l_i | D \in h_i\}) \otimes (\otimes\{\rho_i | D \in r_i\}), \\ r &= (\cup\{h_i | D \in h_i\}) \cup (\cup\{r_i | D \in r_i\}), \end{aligned}$$

and ζ_D is the indicator valuation representation of the decision function for D with respect to v .

In this case, after fusion, the set of valuations is changed as follows. The indicator, utility, and probability valuations that do not include D in their domains remain unchanged. All utility valuations that include D in their domain are combined together, and the resulting utility valuation v is marginalized (using maximization) such that D is eliminated from its domain. A new indicator valuation ζ_D corresponding to the decision function for D is created. All probability and indicator valuations that include D in their domain are

combined together and the resulting probability valuation is combined with ζ_D and the result is marginalized so that D is eliminated from its domain.

In Eq. (5.1), since ζ_D is an indicator valuation representation of the decision function for D , if $g \subseteq r$, then it is not necessary to either combine ζ_D or marginalize D out to compute $(\rho \otimes \zeta_D)^{l((r \cup g) - \{D\})}$ in Eq. (5.1). From the definition of ζ_D in Eq. (4.6), it follows that

$$\begin{aligned} &(\rho \otimes \zeta_D)^{l((r \cup g) - \{D\})}(\mathbf{c}) = \rho(\mathbf{c}, \zeta_D(\mathbf{c}^{l(g - \{D\})})) \\ &\text{for all } \mathbf{c} \in \Omega_{r - \{D\}}, \end{aligned} \quad (5.2)$$

where ζ_D is the decision function for D with respect to v . In words, $(\rho \otimes \zeta_D)^{l((r \cup g) - \{D\})}$ is the “projection” of ρ from r to $r - \{D\}$.

In Eq. (5.1), if none of the probability and indicator valuations include D in their domains, then Eq. (5.1) simplifies as follows:

$$\begin{aligned} &\text{Fus}_D\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\} \\ &= \{v_i | D \notin g_i\} \cup \{v_i^{l(g - \{D\})}\} \cup \{l_1, \dots, l_s\} \\ &\quad \cup \{\rho_1, \dots, \rho_k\}, \end{aligned} \quad (5.3)$$

where v and g are as defined in Eq. (5.1). In words, after fusion, the set of valuations is changed as follows. All utility valuations that include D in their domain are combined together, and the resulting utility valuation v is marginalized such that D is eliminated from its domain. The indicator, utility and probability valuations that do not include D in their domains remain unchanged.

Case 2: Suppose R is a chance variable in $g_1 \dots \cup g_j \cup h_1 \cup \dots \cup h_s \cup r_1 \cup \dots \cup r_k$. In this case, $\text{Fus}_R\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$ is defined as follows:

$$\begin{aligned} &\text{Fus}_R\{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\} \\ &= \{v_i | R \notin g_i\} \cup \left\{ [v \otimes (\rho / \rho^{l(r - \{R\})})]^{l((g \cup r) - \{R\})} \right\} \\ &\quad \cup \{l_i | R \notin h_i\} \cup \{\rho_i | R \notin r_i\} \cup \{\rho^{l(r - \{R\})}\}, \end{aligned} \quad (5.4)$$

where

$$v = \otimes \{v_i | R \in g_i\}, \quad g = \cup \{g_i | R \in g_i\},$$

$$\rho = (\otimes \{l_i | R \in h_i\}) \otimes (\otimes \{\rho_i | R \in r_i\}), \text{ and}$$

$$r = (\cup \{h_i | R \in h_i\}) \cup (\cup \{r_i | R \in r_i\}).$$

In this case, after fusion, the set of valuations is changed as follows. The indicator, utility, and probability valuations whose domains do not include R remain unchanged. A new probability valuation, $\rho^{l(r-\{R\})}$, is created. Finally, we combine all indicator and probability valuations whose domains include R , divide the resulting probability valuation by the new probability valuation that was created, combine the resulting probability valuation with the utility valuations whose domains include R , and finally marginalize the resulting utility valuation such that R is eliminated from its domain.

There are several cases in which Eq. (5.4) can be made computationally more efficient.

Case 2.1: In some instances, $\rho^{l(r-\{R\})}$ is a probability valuation whose values are identically one. This happens, for example, if R is the only chance variable in $g_1 \cup \dots \cup g_j \cup h_1 \cup \dots \cup h_s \cup r_1 \cup \dots \cup r_k$, or if ρ is a conditional for R given $r-\{R\}$. In such cases, Eq. (5.4) simplifies to

$$\text{Fus}_R \{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$$

$$= \{v_i | R \notin g_i\} \cup \{(v \otimes \rho)^{l((g \cup r) - \{R\})}\} \cup \{l_i | R \notin h_i\}$$

$$\cup \{\rho_i | R \notin r_i\}, \tag{5.5}$$

where v, ρ, g , and r are as defined in Eq. (5.4).

Case 2.2: Suppose none of the utility valuations include R in their domains. In this case Eq. (5.4) simplifies as follows (Shenoy, 1992a):

$$\text{Fus}_R \{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$$

$$= \{v_1, \dots, v_j\} \cup \{l_i | R \notin h_i\} \cup \{\rho_i | R \notin r_i\}$$

$$\cup \{\rho^{l(r-\{R\})}\}, \tag{5.6}$$

where ρ and r are as defined in Eq. (5.4). In words, after fusion, the set of valuations is changed as follows. The utility valuations remain unchanged. The indicator and probability valuations whose domains do not include R remain unchanged. The indicator and probability valuations whose do-

main include R are combined together and the resulting probability valuation is marginalized such that R is eliminated from its domain.

Case 2.3: Suppose R is in the domain of all j utility valuations. In this case, Eq. (5.4) simplifies as follows (Shenoy, 1992a):

$$\text{Fus}_R \{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$$

$$= \{(v \otimes \rho)^{l((g \cup r) - \{R\})}\} \cup \{l_i | R \notin h_i\} \cup \{\rho_i | R \notin r_i\}, \tag{5.7}$$

where v, g, ρ , and r are as defined in Eq. (5.4). In words, after fusion, the set of valuations is changed as follows. All utility valuations and those indicator and probability valuations whose domains include R are combined together, and the resulting utility valuation is marginalized such that R is eliminated from its domain. The indicator and probability valuations whose domains do not include R remain unchanged.

We are now ready to state the main theorem.

Theorem 5.1 (Fusion algorithm). *Suppose $\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{l_1, \dots, l_p\}, \{v_1, \dots, v_m\}, \{\rho_1, \dots, \rho_n\}, \rightarrow\}$ is a well-defined valuation network representation of a decision problem. Suppose $X_1 X_2 \dots X_k$ is a valid deletion sequence of variables in $\mathcal{X} = \mathcal{X}_D \cup \mathcal{X}_R$. Then*

$$(\otimes (\otimes \{v_1, \dots, v_m\}) \otimes (\otimes \{l_1, \dots, l_p,$$

$$\rho_1, \dots, \rho_n\}))^{l\emptyset} (\blacklozenge)$$

$$= (\otimes \text{Fus}_{X_k} \{ \dots \text{Fus}_{X_2} \{ \text{Fus}_{X_1} \{v_1, \dots, v_m,$$

$$l_1, \dots, l_p, \rho_1, \dots, \rho_n\} \} \} \} (\blacklozenge).$$

A proof of this theorem is given in Section 8. We will illustrate the fusion algorithm using the UCB problem. Notice that the only valid deletion sequence is $\text{SBR}_2 T_2 R_1 T_1$.

Fusion with respect to S : First we fuse valuations in $\{v_1, v_2, v_3, l_1, l_2, \sigma, \rho_1, \rho_2\}$ with respect to S (see Fig. 6). Since S is a chance variable, we use the definition of fusion in Eq. (5.4), i.e.,

$$\text{Fus}_S \{v_1, v_2, v_3, l_1, l_2, \sigma, \rho_1, \rho_2\}$$

$$= \left\{ v_1, v_2, (v_3 \otimes [(\sigma \otimes \rho_1 \otimes \rho_2) / \right.$$

$$(\sigma \otimes \rho_1 \otimes \rho_2)^{l\{R_1, R_2\}}])^{l\{R_1, R_2, B\}}, l_1, l_2,$$

$$\left. (\sigma \otimes \rho_1 \otimes \rho_2)^{l\{R_1, R_2\}} \right\}.$$

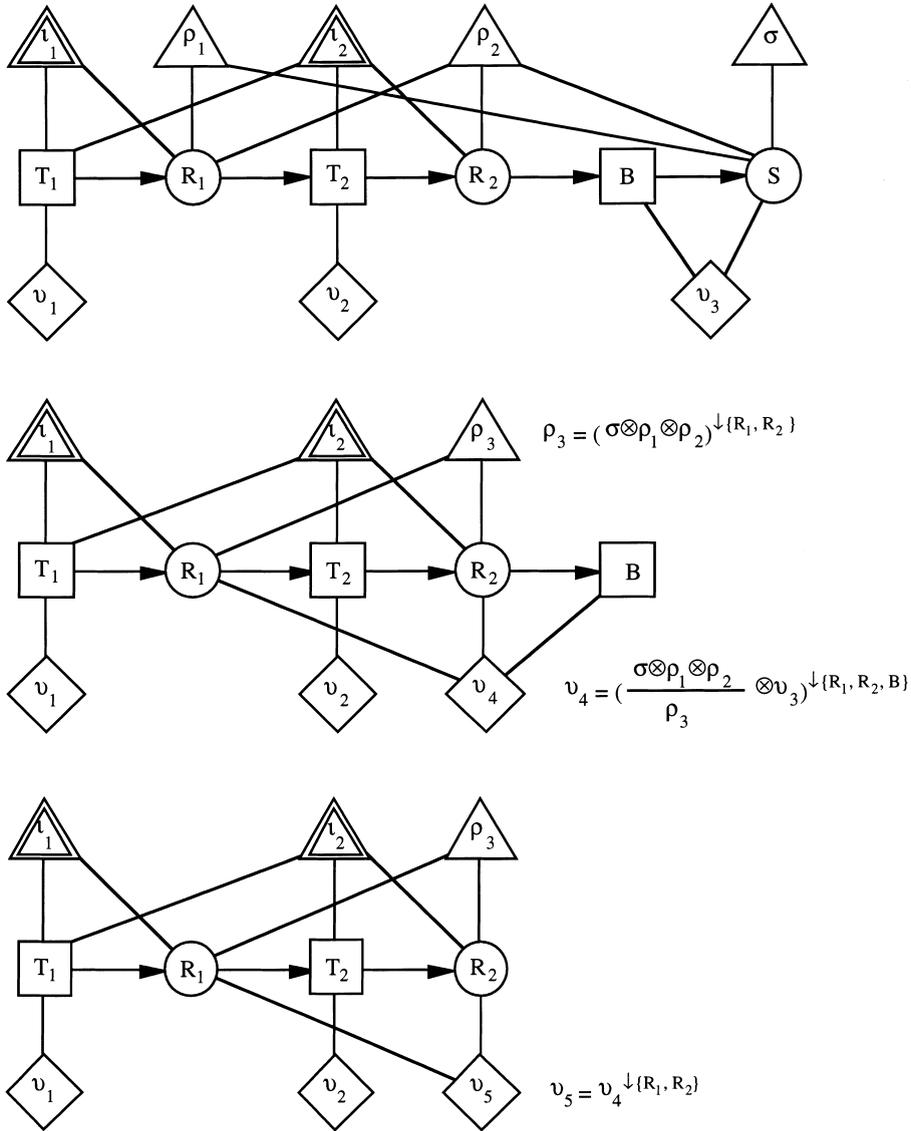


Fig. 6. Top: original VN; middle: after fusion wrt S; bottom: after fusion wrt B.

Let ρ_3 denote $(\sigma \otimes \rho_1 \otimes \rho_2)^{\downarrow\{R_1, R_2\}}$ and let v_4 denote $(v_3 \otimes [(\sigma \otimes \rho_1 \otimes \rho_2) / \rho_3])^{\downarrow\{R_1, R_2, B\}}$. The result of fusion with respect to S is shown graphically in Fig. 6. The details of the numerical computation involved in the fusion operation are shown in Tables 4 and 5. Notice that the combination and marginalization operations are done on effective frames that are computed using the local compu-

tation procedure outlined in Eq. (5.1). Also, the numerical results shown in Tables 4–8 were done in a spreadsheet with 32-bit precision, but the results are only shown after rounding off to two decimal places.

Fusion with respect to B: Next, we fuse the valuations in $\{v_1, v_2, v_4, l_1, l_2, \rho_3\}$ with respect to B (see Fig. 6). Since B is a decision variable and there

Table 4
Details of fusion with respect to S (continued in Table 5)

$\Omega_{\{R_1, R_2, S\}}$			σ	ρ_1	ρ_2	$\sigma \otimes \rho_1 \otimes \rho_2 = \rho$	$\rho^{\downarrow\{R_1, R_2\}} = \rho_3$	$\rho/\rho_3 = \rho'$
n_1	n_2	p	0.80	1	1	0.80	1	0.80
n_1	n_2	l	0.20	1	1	0.20		0.20
d_1	n_2	p	0.80	0.10	1	0.08	0.20	0.40
d_1	n_2	l	0.20	0.60	1	0.12		0.60
d_1	d_2	p	0.80	0.10	0	0	0.07	0
d_1	d_2	l	0.20	0.60	0.56	0.07		1
d_1	$\sim d_2$	p	0.80	0.10	1	0.08	0.13	0.60
d_1	$\sim d_2$	l	0.20	0.60	0.44	0.05		0.40
$\sim d_1$	n_2	p	0.80	0.90	1	0.72	0.80	0.90
$\sim d_1$	n_2	l	0.20	0.40	1	0.08		0.10
$\sim d_1$	d_2	p	0.80	0.90	0.11	0.08	0.13	0.60
$\sim d_1$	d_2	l	0.20	0.40	0.67	0.05		0.40
$\sim d_1$	$\sim d_2$	p	0.80	0.90	0.89	0.64	0.67	0.96
$\sim d_1$	$\sim d_2$	l	0.20	0.40	0.33	0.03		0.04

are no probability valuations that include B in their domains, we use the definition of fusion in Eq. (5.3), i.e.,

$$\begin{aligned} \text{Fus}_B\{v_1, v_2, v_4, l_1, l_2, \rho_3\} \\ = \{v_1, v_2, v_4^{\downarrow\{R_1, R_2\}}, l_1, l_2, \rho_3\}. \end{aligned}$$

Let v_5 denote $v_4^{\downarrow\{R_1, R_2\}}$. The result of fusion with respect to B is shown graphically in Fig. 6. The computational details of the fusion operation and the decision function for B are shown in Table 6.

Fusion with respect to R_2 : Next we fuse $\{v_1, v_2, v_5, l_1, l_2, \rho_3\}$ with respect to R_2 (see Fig. 7). Since R_2 is a chance variable, we use the definition of fusion in Eq. (5.4), i.e.,

$$\begin{aligned} \text{Fus}_{R_2}\{v_1, v_2, v_5, l_1, l_2, \rho_3\} \\ = \{v_1, v_2, [v_5 \otimes (l_2 \otimes \rho_3)] \\ (l_2 \otimes \rho_3)^{\downarrow\{T_1, R_1, T_2\}} \downarrow\{T_1, R_1, T_2\}}, l_1, (l_2 \otimes \rho_3)^{\downarrow\{T_1, R_1, T_2\}}\} \end{aligned}$$

Let ρ_4 denote $(l_2 \otimes \rho_3)^{\downarrow\{T_1, R_1, T_2\}}$ and let v_6 denote $[v_5 \otimes (l_2 \otimes \rho_3)]/\rho_4^{\downarrow\{T_1, R_1, T_2\}}$. The result of fusion with respect to R_2 is shown graphically in Fig. 7. The details of the numerical computation involved in the fusion operation are shown in Table 7.

Fusion with respect to T_2 : Next, we fuse $\{v_1, v_2, v_6, l_1, \rho_4\}$ with respect to T_2 (see Fig. 7).

Since T_2 is a decision variable, we use the definition of fusion in Eq. (5.1), i.e.,

$$\begin{aligned} \text{Fus}_{T_2}\{v_1, v_2, v_6, l_1, \rho_4\} \\ = \{v_1, (v_2 \otimes v_6)^{\downarrow\{T_1, R_1\}}, l_1, (\rho_4 \otimes \zeta_{T_2})^{\downarrow\{T_1, R_1\}}\}, \end{aligned}$$

where ζ_{T_2} is the decision function for T_2 . Let v_7 denote $(v_2 \otimes v_6)^{\downarrow\{T_1, R_1\}}$, and let ρ_5 denote $(\rho_4 \otimes \zeta_{T_2})^{\downarrow\{T_1, R_1\}}$. The result of fusion with respect to T_2 is shown graphically in Fig. 7. The details of the numerical computation involved in the fusion operation are shown in Table 7. Notice that we can use Eq. (5.2) to identify the values of $\rho_5 = (\rho_4 \otimes \zeta_{T_2})^{\downarrow\{T_1, R_1\}}$ and no computations are necessary for this operation (see Table 7).

Fusion with respect to R_1 : Next, we fuse $\{v_1, v_7, l_1, \rho_5\}$ with respect to R_1 (see Fig. 8). Since R_1 is a chance variable, and it is the only chance variable, we use the definition of fusion in Eq. (5.5), i.e.,

$$\text{Fus}_{R_1}\{v_1, v_7, l_1, \rho_5\} = \{v_1, (v_7 \otimes l_1 \otimes \rho_5)^{\downarrow T_1}\}.$$

Let v_8 denote $(v_7 \otimes l_1 \otimes \rho_5)^{\downarrow T_1}$. The result of fusion with respect to R_1 is shown in Fig. 8. The details of the computation involved in the fusion operation are shown in Table 8.

Fusion with respect to T_1 : Next we fuse $\{v_1, v_8\}$ with respect to T_1 (see Fig. 8). Since T_1 is a

Table 5
 Details of fusion with respect to S (continued from Table 4)

$\Omega_{\{R_1, R_2, B, S\}}$				ρ'	v_3	$\rho' \otimes v_3$	$(\rho' \otimes v_3)^{\downarrow\{R_1, R_2, B\}} = v_4$
n_1	n_2	b	p	0.80	60	48	28
n_1	n_2	b	l	0.20	-100	-20	
n_1	n_2	g	p	0.80	20	16	24
n_1	n_2	g	l	0.20	40	8	
n_1	n_2	$\sim b$	p	0.80	0	0	0
n_1	n_2	$\sim b$	l	0.20	0	0	
d_1	n_2	b	p	0.40	60	24	-36
d_1	n_2	b	l	0.60	-100	-60	
d_1	n_2	g	p	0.40	20	8	32
d_1	n_2	g	l	0.60	40	24	
d_1	n_2	$\sim b$	p	0.40	0	0	0
d_1	n_2	$\sim b$	l	0.60	0	0	
d_1	d_2	b	p	0	60	0	-100
d_1	d_2	b	l	1	-100	-100	
d_1	d_2	g	p	0	20	0	40
d_1	d_2	g	l	1	40	40	
d_1	d_2	$\sim b$	p	0	0	0	0
d_1	d_2	$\sim b$	l	1	0	0	
d_1	$\sim d_2$	b	p	0.60	60	36	-4
d_1	$\sim d_2$	b	l	0.40	-100	-40	
d_1	$\sim d_2$	g	p	0.60	20	12	28
d_1	$\sim d_2$	g	l	0.40	40	16	
d_1	$\sim d_2$	$\sim b$	p	0.60	0	0	0
d_1	$\sim d_2$	$\sim b$	l	0.40	0	0	
$\sim d_1$	n_2	b	p	0.90	60	54	44
$\sim d_1$	n_2	b	l	0.10	-100	-10	
$\sim d_1$	n_2	g	p	0.90	20	18	22
$\sim d_1$	n_2	g	l	0.10	40	4	
$\sim d_1$	n_2	$\sim b$	p	0.90	0	0	0
$\sim d_1$	n_2	$\sim b$	l	0.10	0	0	
$\sim d_1$	d_2	b	p	0.60	60	36	-4
$\sim d_1$	d_2	b	l	0.40	-100	-40	
$\sim d_1$	d_2	g	p	0.60	20	12	28
$\sim d_1$	d_2	g	l	0.40	40	16	
$\sim d_1$	d_2	$\sim b$	p	0.60	0	0	0
$\sim d_1$	d_2	$\sim b$	l	0.40	0	0	
$\sim d_1$	$\sim d_2$	b	p	0.96	60	57.60	53.60
$\sim d_1$	$\sim d_2$	b	l	0.04	-100	-4	
$\sim d_1$	$\sim d_2$	g	p	0.96	20	19.20	20.80
$\sim d_1$	$\sim d_2$	g	l	0.04	40	1.60	
$\sim d_1$	$\sim d_2$	$\sim b$	p	0.96	0	0	0
$\sim d_1$	$\sim d_2$	$\sim b$	l	0.04	0	0	

decision variable, and there are no probability valuations that bear on T_1 , we use the definition of fusion in Eq. (5.3), i.e.,

$$\text{Fus}_{T_1}\{v_1, v_8\} = \{(v_1 \otimes v_8)^{\downarrow\emptyset}\}.$$

Let v_9 denote $(v_1 \otimes v_8)^{\downarrow\emptyset}$. The result after fusion with respect to T_1 is shown in Fig. 8. The details of the computation involved in the fusion operation are shown in Table 8.

This completes the fusion algorithm. An optimal strategy can be pieced together from the

Table 6
Details of fusion with respect to B

$\Omega_{\{R_1, R_2, B\}}$			v_4	$v_4^{\{R_1, R_2\}} = v_5$	ξ_B
n_1	n_2	b	28	28	b
n_1	n_2	g	24		
n_1	n_2	$\sim b$	0		
d_1	n_2	b	-36		
d_1	n_2	g	32	32	g
d_1	n_2	$\sim b$	0		
d_1	d_2	b	-100		
d_1	d_2	g	40	40	g
d_1	d_2	$\sim b$	0		
d_1	$\sim d_2$	b	-4		
d_1	$\sim d_2$	g	28	28	g
d_1	$\sim d_2$	$\sim b$	0		
$\sim d_1$	n_2	b	44	44	b
$\sim d_1$	n_2	g	22		
$\sim d_1$	n_2	$\sim b$	0		
$\sim d_1$	d_2	b	-4		
$\sim d_1$	d_2	g	28	28	g
$\sim d_1$	d_2	$\sim b$	0		
$\sim d_1$	$\sim d_2$	b	53.60	53.60	b
$\sim d_1$	$\sim d_2$	g	20.80		
$\sim d_1$	$\sim d_2$	$\sim b$	0		

Table 7
Details of fusion with respect to R_2 and T_2 . ρ_4 denotes $(t_2 \otimes \rho_3)^{\{T_1, R_1, T_2\}}$, v_6 denotes $[(t_2 \otimes \rho_3) / \rho_4] \otimes v_5^{\{T_1, R_1, T_2\}}$, v_7 denotes $(v_6 \otimes v_2)^{\{T_1, R_1\}}$, and ρ_5 denotes $(\rho_4 \otimes \xi_{T_2})^{\{T_1, R_1\}}$

$\Omega_{\{T_1, R_1, T_2, R_2\}}$				$t_2 \otimes \rho_3$	ρ_4	$\frac{t_2 \otimes \rho_3}{\rho_4}$	v_5	$\frac{t_2 \otimes \rho_3}{\rho_4} \otimes v_5$	v_6	v_2	$v_6 \otimes v_2$	v_7	ξ_{T_2}	ρ_5
nt	n_1	nc	n_2	1	1	1	28	28	28	0	28	28	nc	1
t_1	d_1	nc	n_2	0.20	0.20	1	32	32	32	0	32	32	nc	0.20
t_1	$\sim d_1$	nc	n_2	0.80	0.80	1	44	44	44	0	44	44	nc	0.80
t_2	d_1	nc	d_2	0.07	0.20	0.33	40	13.33	32	0	32	32	nc	0.20
t_2	d_1	nc	$\sim d_2$	0.13		0.67	28	18.67						
t_2	$\sim d_1$	nc	d_2	0.13	0.80	0.17	28	4.67	49.33	0	49.33	49.33	nc	0.80
t_2	$\sim d_1$	nc	$\sim d_2$	0.67		0.83	53.60	44.67						
t_3	d_1	s	n_2	0.20	0.20	1	32	32	32	0	32	32	s	0.20
t_3	d_1	c	d_2	0.07	0.20	0.33	40	13.33	32	-4	28			
t_3	d_1	c	$\sim d_2$	0.13		0.67	28	18.67						
t_3	$\sim d_1$	s	n_2	0.80	0.80	1	44	44	44	0	44			
t_3	$\sim d_1$	c	d_2	0.13	0.80	0.17	28	4.67	49.33	-4	45.33	45.33	c	0.80
t_3	$\sim d_1$	c	$\sim d_2$	0.67		0.83	53.60	44.67						

decision functions $\xi_{T_1}, \xi_{T_2}, \xi_B$. An optimal strategy is do test t_2 ; if the test results from the two subsystems are either (d_1, d_2) or $(d_1, \sim d_2)$, or $(\sim d_1, d_2)$, then

buy with guarantee, and if the test results are $(\sim d_1, \sim d_2)$, then buy with no guarantee. The expected utility of the optimal strategy is \$32.87.

Table 8
Details of fusion with respect to R_1 and T_1

$\Omega_{\{T_1, R_1\}}$		$t_1 \otimes \rho_5 = \rho_6$	v_7	$\rho_6 \otimes v_7$	$(t_1 \otimes \rho_5 \otimes v_7)^{\downarrow T_1}$ $= v_8$	v_1	$v_1 \otimes v_8$	$v_1 \otimes v_8^{\downarrow \emptyset(\blacklozenge)}$ $= v_9(\blacklozenge)$	$\xi_{T_1}(\blacklozenge)$
nt	n_1	1	28	28	28	0	28		
t_1	d_1	0.20	32	6.40	41.60	-9	32.60		
t_1	$\sim d_1$	0.80	44	35.20					
t_2	d_1	0.20	32	6.40	45.87	-13	32.87	32.87	t_2
t_2	$\sim d_1$	0.80	49.33	39.47					
t_3	d_1	0.20	32	6.40	42.67	-10	32.67		
t_3	$\sim d_1$	0.80	45.33	36.27					

The fusion algorithm described here is similar to the message-passing algorithm described by Lauritzen and Spiegelhalter (1988) in the sense that our algorithm takes advantage of the conditional independencies in the underlying probability distribution. The fusion algorithm can be described as a message-passing algorithm where the messages are passed between neighboring nodes in a join tree (see, e.g., Jensen et al., 1994; Cowell, 1994). However, since in a decision problem we are only interested in one marginal (for the empty set), the junction tree data structure does not contribute much to computational efficiency. It does, however, add some overhead to the computational cost.

6. Comparison with symmetric valuation network technique

In this section, we compare the technique proposed here with the symmetric valuation network technique described in Shenoy (1992a, 1993). This will explain the contribution and significance of this work.

The valuation network representation described in Shenoy (1992a, 1993) is designed for symmetric decision problems. It includes neither the notion of indicator valuations nor the related notion of effective frames. Like the technique described here, the symmetric valuation network representation technique makes a problem symmetric by adding dummy configurations to variables. This strategy enlarges the size of the problem. In the asymmetric valuation network representation, the enlarged size is reduced using the notions of indicator val-

uations and effective frames. In comparison, the symmetric valuation network technique is unable to reduce the size of the problem, and this results in a huge computational penalty in solving the problem.

There is a second more fundamental advantage of the technique described here for asymmetric decision problems that is best illustrated by an example. Fig. 9 shows the symmetric valuation network for the UCB problem. If we compare this symmetric valuation network to the asymmetric valuation network shown in Fig. 3, the differences are in the probability valuations. In the symmetric valuation network, the joint conditional probability distribution factors into three probability valuations – ρ'_1 , ρ'_2 , and σ – whereas in the asymmetric valuation network, the joint probability distribution factors into five valuations – three probability valuations, ρ_1 , ρ_2 , and σ , and two indicator valuations, t_1 and t_2 . The relation between the two probability models is as follows: $\rho'_1 = t_1 \otimes \rho_1$, and $\rho'_2 = t_2 \otimes \rho_2$. Thus we can think of an indicator valuation as a qualitative factor of a probability valuation. The factoring of an indicator valuation out of a probability valuation results in reducing the domain of the resulting probability valuation. For example, while the domain of ρ'_2 is $\{T_1, R_1, T_2, R_2, S\}$, the domain of ρ_2 is only $\{R_1, R_2, S\}$. Thus, in general, the valuations in an asymmetric valuation network are on smaller domains than the valuations in symmetric valuation networks. This fact has a severe implication in the computational cost of solving a problem. For example, in the UCB problem, if we apply the fusion algorithm to the symmetric valuation network, since S has to be deleted first and S is in-

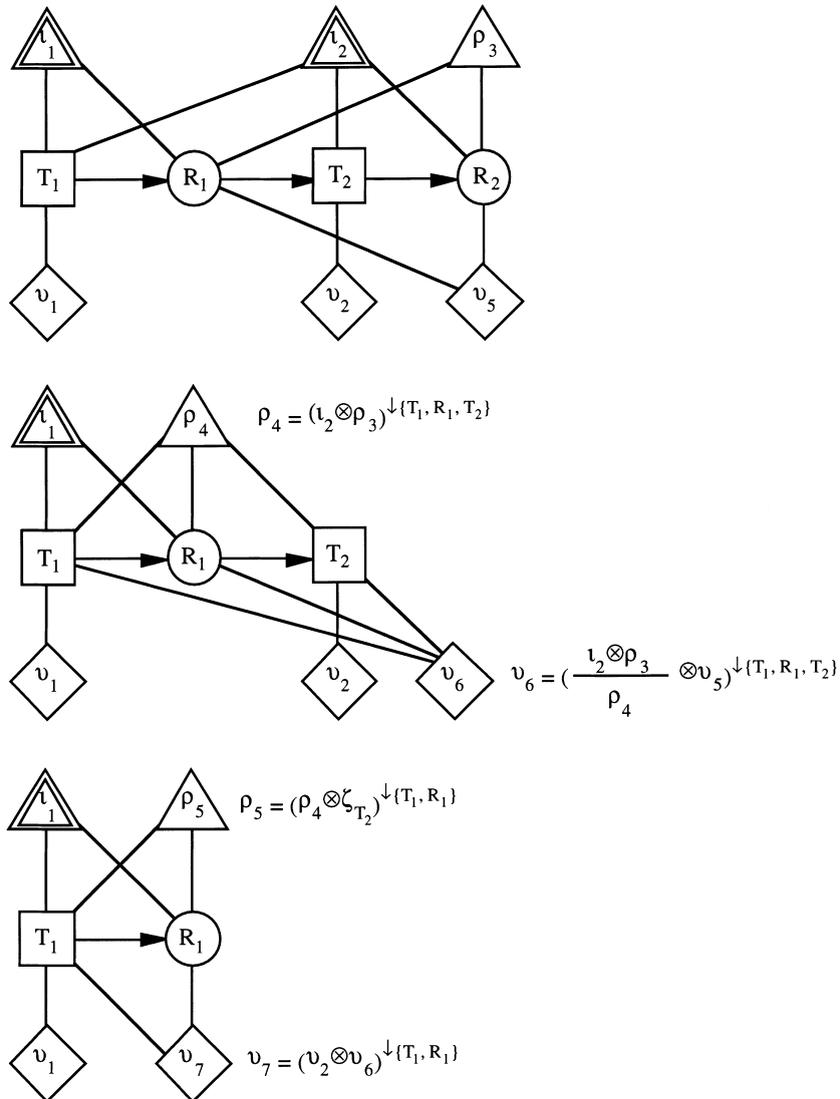


Fig. 7. Top: the UCB valuation network after fusion with respect to S and B ; middle: after fusion with respect to R_2 ; bottom: after fusion with respect to T_2 .

cluded in the domain of ρ'_1, ρ'_2 , and σ' , the fusion with respect to S is done on the global frame of all six variables. In comparison, in the asymmetric valuation network, the fusion with respect to S is done on the effective frame of only four variables, namely $\{R_1, R_2, B, S\}$.

In summary, the asymmetric valuation network representation is more expressive than the symmetric valuation network representation. The

concept of indicator valuations allows us to represent structural asymmetry in a decision problem. Also, the asymmetric valuation network representation leads to a more efficient solution for two reasons. First, we do all computations on effective frames (instead of frames). Second, in general, the computations are more local because the probability valuations are on smaller domains.

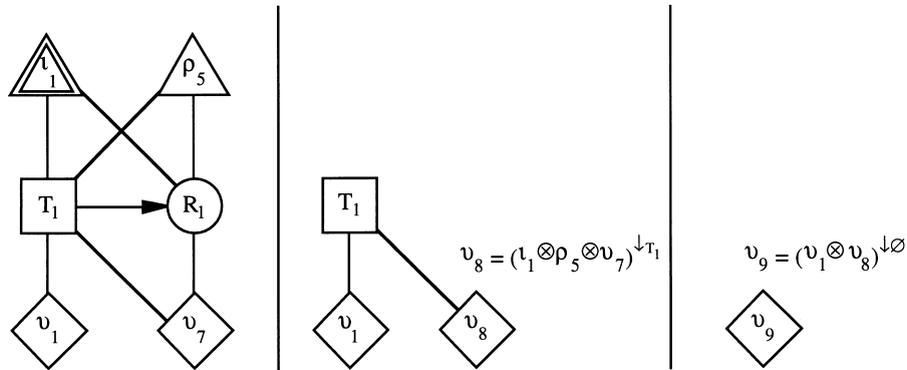


Fig. 8. Left: the VN for the UCB after fusion with respect to S, B, R_2 and T_2 ; middle: after fusion with respect to R_1 ; right: after fusion with respect to T_1 .

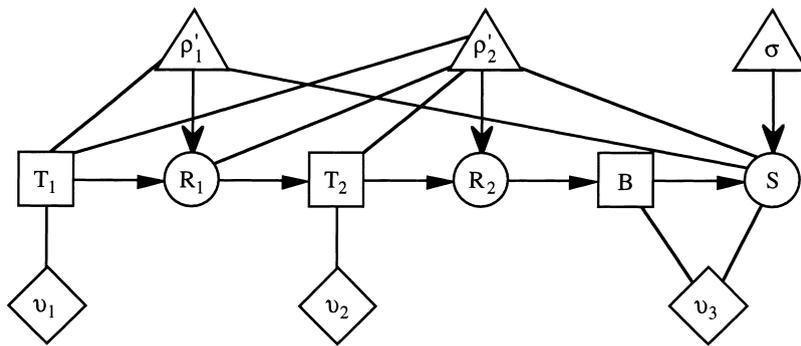


Fig. 9. A symmetric valuation network for the UCB problem.

7. Summary and conclusion

The main contribution of this paper is an augmentation of the valuation network technique for representing and solving asymmetric decision problems. The structural asymmetry in a decision problem is represented by indicator valuations. An indicator valuation is a special type of a probability valuation. Indicator valuations allow us to reduce the domain of probability valuations. This contributes to the efficiency of the solution technique. Also, indicator valuations are used to define effective frames. An effective frame is a subset of a frame. All computations are done on effective frames,

and this contributes also to the efficiency of the solution technique.

The fusion algorithm is essentially the same as in the symmetric case. The main difference is in how indicator valuations are handled. Indicator valuations are treated as probability valuations. However since indicator valuations are identically one on effective frames, there are no computations involved in combining indicator valuations. This contributes to the efficiency of the solution technique. Indicator valuations do contribute domain information and cannot be totally ignored.

A disadvantage of our technique is the adding of dummy configurations to variables that

increases the computational burden of solving the problem. If we can find a way of making a problem symmetric without enlarging the frames of variables, then our technique can be improved. One possibility is to adapt Smith, Holtzman and Matheson’s (Smith et al., 1993) sharing method to valuation networks. This can be done by assigning probabilities to subsets of frames as in the theory of belief functions (Bordley, 1993; Liu and Shenoy, 1993).

8. Proofs

In this section we give a proof for Theorem 5.1. First we state and prove a lemma needed to prove Theorem 5.1.

Lemma 8.1. *Suppose $\Delta = \{\mathcal{X}_D, \mathcal{X}_R, \{l_1, \dots, l_s\}, \{v_1, \dots, v_j\}, \{\rho_1, \dots, \rho_k\}, \rightarrow\}$ is a well-defined valuation network representation of a 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*

$$((\otimes\{v_1, \dots, v_j\}) \otimes (\otimes\{l_1, \dots, l_s, \rho_1, \dots, \rho_k\}))^{\downarrow(\mathcal{X}-\{X\})} = \otimes \text{Fus}_X \{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}.$$

Proof of Lemma 8.1. We prove this result in two mutually exclusive and exhaustive cases, each corresponding to the definition of the fusion operation in Eqs. (5.1) and (5.4), respectively. Suppose l_i is an indicator valuation for h_i ; v_i is a utility valuation for g_i , and ρ_i is a probability valuation for r_i .

Case 1: Suppose X is a decision variable. Without loss of generality, assume that v_1, \dots, v_j are the only utility valuations that have X in their domains, and $l_1, \dots, l_s, \rho_1, \dots, \rho_k$ are the only indicator and probability valuations that have X in their domains. Let v denote $\otimes\{v_1, \dots, v_j\}$; g the $\cup\{g_1, \dots, g_j\}$; ρ the $\otimes\{l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$, and r denote $\cup\{h_1, \dots, h_s, r_1, \dots, r_k\}$. Suppose $\mathbf{c} \in \Omega_{\mathcal{X}-\{X\}}$. Then

$$\begin{aligned} & ((\otimes\{v_1, \dots, v_j\}) \\ & \otimes (\otimes\{l_1, \dots, l_s, \rho_1, \dots, \rho_k\}))^{\downarrow(\mathcal{X}-\{X\})}(\mathbf{c}) \\ & = \text{MAX}\{[v_1(\mathbf{c}^{\downarrow g_1-\{X\}}, \mathbf{x}) + \dots + v_j(\mathbf{c}^{\downarrow g_j-\{X\}}, \mathbf{x}) \\ & \quad + v_{j+1}(\mathbf{c}^{\downarrow g_{j+1}}) + \dots + v_j(\mathbf{c}^{\downarrow g_j})] \\ & \quad \times [l_1(\mathbf{c}^{\downarrow h_1-\{X\}}, \mathbf{x}) \dots l_s(\mathbf{c}^{\downarrow h_s-\{X\}}, \mathbf{x}) l_{s+1} \\ & \quad (\mathbf{c}^{\downarrow h_{s+1}}) \dots l_s(\mathbf{c}^{\downarrow h_s}) \rho_1(\mathbf{c}^{\downarrow r_1-\{X\}}, \mathbf{x}) \dots \\ & \quad \rho_k(\mathbf{c}^{\downarrow r_k-\{X\}}, \mathbf{x}) \rho_{k+1}(\mathbf{c}^{\downarrow r_{k+1}}) \dots \rho_k(\mathbf{c}^{\downarrow r_k})]\} \\ & \quad \mathbf{x} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_{\mathcal{X}}. \end{aligned}$$

Since Δ is a well-defined, it can be shown that

$$\begin{aligned} & l_1(\mathbf{c}^{\downarrow h_1-\{X\}}, \mathbf{x}) \dots l_s(\mathbf{c}^{\downarrow h_s-\{X\}}, \mathbf{x}) \rho_1(\mathbf{c}^{\downarrow r_1-\{X\}}, \mathbf{x}) \dots \\ & \quad \rho_k(\mathbf{c}^{\downarrow r_k-\{X\}}, \mathbf{x}) \\ & = l_1(\mathbf{c}^{\downarrow h_1-\{X\}}, \mathbf{y}) \dots l_s(\mathbf{c}^{\downarrow h_s-\{X\}}, \mathbf{y}) \\ & \quad \rho_1(\mathbf{c}^{\downarrow r_1-\{X\}}, \mathbf{y}) \dots \rho_k(\mathbf{c}^{\downarrow r_k-\{X\}}, \mathbf{y}) \end{aligned}$$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}), (\mathbf{c}, \mathbf{y}) \in \Omega_{\mathcal{X}}$

(otherwise one can prove that $l_1, \dots, l_s, \rho_1, \dots, \rho_k$ do not satisfy the definition of well-defined potentials in Eq. (4.10)). Therefore,

$$\begin{aligned} & ((\otimes\{v_1, \dots, v_j\}) \\ & \otimes (\otimes\{l_1, \dots, l_s, \rho_1, \dots, \rho_k\}))^{\downarrow(\mathcal{X}-\{X\})}(\mathbf{c}) \\ & = [\text{MAX}\{v_1(\mathbf{c}^{\downarrow g_1-\{X\}}, \mathbf{x}) + \dots + v_j(\mathbf{c}^{\downarrow g_j-\{X\}}, \mathbf{x}) | \mathbf{x} \\ & \quad \in \mathcal{W}_X \ni (\mathbf{c}^{\downarrow g-\{X\}}, \mathbf{x}) \in \Omega_{\mathcal{X}}\} + v_{j+1}(\mathbf{c}^{\downarrow g_{j+1}}) \\ & \quad + \dots + v_j(\mathbf{c}^{\downarrow g_j})][l_1(\mathbf{c}^{\downarrow h_1-\{X\}}, \mathbf{x}) \dots l_s(\mathbf{c}^{\downarrow h_s-\{X\}}, \mathbf{x}) \\ & \quad \rho_1(\mathbf{c}^{\downarrow r_1-\{X\}}, \mathbf{x}) \dots \rho_k(\mathbf{c}^{\downarrow r_k-\{X\}}, \mathbf{x}) l_{s+1}(\mathbf{c}^{\downarrow h_{s+1}}) \\ & \quad \dots l_s(\mathbf{c}^{\downarrow h_s}) \rho_{k+1}(\mathbf{c}^{\downarrow r_{k+1}}) \dots \rho_k(\mathbf{c}^{\downarrow r_k})] \\ & = [v^{\downarrow g-\{X\}}(\mathbf{c}^{\downarrow g-\{X\}}) + v_{j+1}(\mathbf{c}^{\downarrow g_{j+1}}) + \dots + v_j(\mathbf{c}^{\downarrow g_j})] \\ & \quad [l_1(\mathbf{c}^{\downarrow h_1-\{X\}}, \mathbf{y}) \dots l_s(\mathbf{c}^{\downarrow h_s-\{X\}}, \mathbf{y}) \rho_1(\mathbf{c}^{\downarrow r_1-\{X\}}, \mathbf{y}) \\ & \quad \dots \rho_k(\mathbf{c}^{\downarrow r_k-\{X\}}, \mathbf{y}) l_{s+1}(\mathbf{c}^{\downarrow h_{s+1}}) \\ & \quad \dots l_s(\mathbf{c}^{\downarrow h_s}) \rho_{k+1}(\mathbf{c}^{\downarrow r_{k+1}}) \dots \rho_k(\mathbf{c}^{\downarrow r_k})] \\ & \quad (\text{where } \mathbf{y} = \zeta_X(\mathbf{c}^{\downarrow g-\{X\}})) \\ & = [v^{\downarrow g-X}(\mathbf{c}^{\downarrow g-\{X\}}) + v_{j+1}(\mathbf{c}^{\downarrow g_{j+1}}) + \dots + v_j(\mathbf{c}^{\downarrow g_j})] \\ & \quad [(\rho \otimes \zeta_X)^{\downarrow r-\{X\}}(\mathbf{c}^{\downarrow r-\{X\}}) l_{s+1}(\mathbf{c}^{\downarrow h_{s+1}}) \dots l_s(\mathbf{c}^{\downarrow h_s}) \\ & \quad \rho_{k+1}(\mathbf{c}^{\downarrow r_{k+1}}) \dots \rho_k(\mathbf{c}^{\downarrow r_k})] \\ & = \otimes(\{v^{\downarrow g-\{X\}}, v_{j+1}, \dots, v_j\} \\ & \quad \cup\{(\rho \otimes \zeta_X)^{\downarrow r-\{X\}}, l_{s+1}, \dots, l_s, \rho_{k+1}, \dots, \rho_k\})(\mathbf{c}) \end{aligned}$$

$$= \otimes \text{Fus}_X \{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}(\mathbf{c})$$

(from (5.1)).

Case 2: Suppose X is a random variable. Without loss of generality, assume that v_1, \dots, v_j are the only utility valuations that have X in their domains, and $l_1, \dots, l_s, \rho_1, \dots, \rho_k$ are the only indicator and probability valuations that have X in their domains. Let v denote $\otimes \{v_1, \dots, v_j\}$; $g \cup \{g_1, \dots, g_j\}$; $\rho \otimes \{l_1, \dots, l_s, \rho_1, \dots, \rho_k\}$, and r denote $\cup \{h_1, \dots, h_s, r_1, \dots, r_k\}$. Suppose $\mathbf{c} \in \Omega_{\mathcal{X}-\{X\}}$. Then

$$\begin{aligned} & ((\otimes \{v_1, \dots, v_j\}) \otimes (\otimes \{l_1, \dots, l_s, \rho_1, \dots, \rho_k\}))^{l(\mathcal{X}-\{X\})}(\mathbf{c}) \\ &= \sum [v_1(\mathbf{c}^{lg_1-\{X\}}, \mathbf{x}) + \dots + v_j(\mathbf{c}^{lg_j-\{X\}}, \mathbf{x}) + v_{j+1}(\mathbf{c}^{lg_{j+1}}, \mathbf{x}) + \dots + v_j(\mathbf{c}^{lg_j}) \\ & \quad [l_1(\mathbf{c}^{lh_1-\{X\}}, \mathbf{x}) \dots l_s(\mathbf{c}^{lh_s-\{X\}}, \mathbf{x}) l_{s+1}(\mathbf{c}^{lh_{s+1}}) \dots l_s(\mathbf{c}^{lh_s}) \rho_1(\mathbf{c}^{lr_1-\{X\}}, \mathbf{x}) \dots \rho_k(\mathbf{c}^{lr_k-\{X\}}, \mathbf{x}) \\ & \quad \rho_{k+1}(\mathbf{c}^{lr_{k+1}}) \dots \rho_k(\mathbf{c}^{lr_k})] | \mathbf{x} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_{\mathcal{X}} \\ &= \left(\sum \{ [v(\mathbf{c}^{lg-\{X\}}, \mathbf{x}) + v_{j+1}(\mathbf{c}^{lg_{j+1}}) + \dots + v_j(\mathbf{c}^{lg_j})] [\rho(\mathbf{c}^{lr-\{X\}}, \mathbf{x})] | \mathbf{x} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_{\mathcal{X}} \right) l_{s+1}(\mathbf{c}^{lh_{s+1}}) \dots l_s(\mathbf{c}^{lh_s}) \rho_{k+1}(\mathbf{c}^{lr_{k+1}}) \dots \rho_k(\mathbf{c}^{lr_k}) \\ &= \left(\sum \{ [v(\mathbf{c}^{lg-\{X\}}, \mathbf{x}) + v_{j+1}(\mathbf{c}^{lg_{j+1}}) + \dots + v_j(\mathbf{c}^{lg_j})] [\rho(\mathbf{c}^{lr-\{X\}}, \mathbf{x}) / \rho^{lr-\{X\}}(\mathbf{c}^{lr-\{X\}})] | \mathbf{x} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_{\mathcal{X}} \right) l_{s+1}(\mathbf{c}^{lh_{s+1}}) \dots l_s(\mathbf{c}^{lh_s}) \rho_{k+1}(\mathbf{c}^{lr_{k+1}}) \dots \rho_k(\mathbf{c}^{lr_k}) \\ &= \left(\sum \{ [v(\mathbf{c}^{lg-\{X\}}, \mathbf{x}) (\rho / \rho^{lr-\{X\}})(\mathbf{c}^{lr-\{X\}}, \mathbf{x}) + [(v_{j+1}(\mathbf{c}^{lg_{j+1}}) + \dots + v_j(\mathbf{c}^{lg_j})) (\rho / \rho^{lr-\{X\}})(\mathbf{c}^{lr-\{X\}}, \mathbf{x})] | \mathbf{x} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_{\mathcal{X}} \right) l_{s+1}(\mathbf{c}^{lh_{s+1}}) \dots l_s(\mathbf{c}^{lh_s}) \rho_{k+1}(\mathbf{c}^{lr_{k+1}}) \dots \rho_k(\mathbf{c}^{lr_k}) \rho^{lr-\{X\}}(\mathbf{c}^{lr-\{X\}}) \end{aligned}$$

$$\begin{aligned} &= \left(\sum \{ [v(\mathbf{c}^{lg-\{X\}}, \mathbf{x}) (\rho / \rho^{lr-\{X\}})(\mathbf{c}^{lr-\{X\}}, \mathbf{x})] | \mathbf{x} \in W_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_X \} \right) \\ & \quad + [(v_{j+1}(\mathbf{c}^{lg_{j+1}}) + \dots + v_j(\mathbf{c}^{lg_j})) (\rho / \rho^{lr-\{X\}})(\mathbf{c}^{lr-\{X\}}, \mathbf{x}) | \mathbf{x} \in \mathcal{W}_X \ni (\mathbf{c}, \mathbf{x}) \in \Omega_X] l_{s+1}(\mathbf{c}^{lh_{s+1}}) \dots l_s(\mathbf{c}^{lh_s}) \rho_{k+1}(\mathbf{c}^{lr_{k+1}}) \dots \rho_k(\mathbf{c}^{lr_k}) \rho^{lr-\{X\}}(\mathbf{c}^{lr-\{X\}}) \\ &= [(v \otimes \rho / (\rho^{lr-\{X\}}))^{l(g \cup r) - \{X\}}(\mathbf{c}^{l(g \cup r) - \{X\}}) + v_{j+1}(\mathbf{c}^{lg_{j+1}}) + \dots + v_j(\mathbf{c}^{lg_j})] l_{s+1}(\mathbf{c}^{lh_{s+1}}) \dots l_s(\mathbf{c}^{lh_s}) \rho_{k+1}(\mathbf{c}^{lr_{k+1}}) \dots \rho_k(\mathbf{c}^{lr_k}) \rho^{lr-\{X\}}(\mathbf{c}^{lr-\{X\}}) \\ &= \otimes \left((v \otimes \rho / (\rho^{lr-\{X\}}))^{l(g \cup r) - \{X\}}, v_{j+1}, \dots, v_j \right) \cup \{l_{s+1}, \dots, l_s, \rho_{k+1}, \dots, \rho_k, \rho^{lr-\{X\}}\}(\mathbf{c}) \\ &= \otimes \text{Fus}_X \{v_1, \dots, v_j, l_1, \dots, l_s, \rho_1, \dots, \rho_k\}(\mathbf{c}) \end{aligned}$$

(from (5.4)).

We have now shown the result for both cases. Therefore the result follows.

Proof of Theorem 5.1. By definition, $(\otimes \{v_1, \dots, v_m\}) \otimes (\otimes \{l_1, \dots, l_p, \rho_1, \dots, \rho_n\})^{l_0}$ is obtained by sequentially marginalizing a minimal variable. A proof of this theorem is obtained by repeatedly applying the result of Lemma 8.1. At each step, we delete a minimal variable and fuse the set of all valuations with respect to the minimal variable. It is easy to show that after fusion, the resulting valuation network is well-defined. Using Lemma 8.1, after fusion with respect to X_1 , the combination of all valuations in the resulting valuation network is equal to $(\otimes \{v_1, \dots, v_m\}) \otimes (\otimes \{l_1, \dots, l_p, \rho_1, \dots, \rho_n\})^{l(\mathcal{X}-\{X_1\})}$. Again, using Lemma 8.1, after fusion with respect to X_2 , the combination of all valuations in the resulting VBS is equal to $(\otimes \{v_1, \dots, v_m\}) \otimes (\otimes \{l_1, \dots, l_p, \rho_1, \dots, \rho_n\})^{l(\mathcal{X}-\{X_1, X_2\})}$ and so on. When all the variables have been deleted, there will be a single valuation left. Using Lemma 8.1, this valuation will be $(\otimes \{v_1, \dots, v_m\}) \otimes (\otimes \{l_1, \dots, l_p, \rho_1, \dots, \rho_n\})^{l_0}$.

Acknowledgements

This work is based upon work supported in part by the National Science Foundation under Grant No. SES-9213558. I am grateful to Rui Guo, Diane Lander, Liping Liu, Concha Bielza, Hong Xu, and Riza Demirer for comments and discussions. Comments and questions from Concha Bielza have resulted in elimination of several errors in previous drafts of this paper. The paper has benefited immensely from the comments of two anonymous referees on an earlier draft of this paper.

References

- Bielza, C., Shenoy, P.P., 1996. A comparison of graphical techniques for asymmetric decision problems. Working paper no. 271, School of Business, University of Kansas, Lawrence, KS. Available by ftp from <ftp.bs.school.ukans.edu/home/pshenoy/wp271.pdf>.
- Bordley, R.F., 1993. Modelling unforeseen events with similarity templates: When Bayesian probabilities resemble Shafer's probabilities. Unpublished manuscript. Operating Sciences Department, General Motors Research Labs, Warren, MI.
- Call, H.J., Miller, W.A., 1990. A comparison of approaches and implementations for automating decision analysis. *Reliability Engineering and System Safety* 30, 115–162.
- Covaliu, Z., Oliver, R.M., 1995. Representation and solution of decision problems using sequential decision diagrams. *Management Science* 41 (12), 1860–1881.
- Cowell, R.G., 1994. Decision networks: A new formulation for multistage decision problems. Research report no. 132, University College London, UK.
- Ezawa, K.J., 1986. Efficient evaluation of influence diagrams. Ph.D. thesis, Department of Engineering–Economic Systems, Stanford University, Stanford, CA.
- Fung, R.M., Shachter, R.D., 1990. Contingent influence diagrams. Working paper, Department of Engineering–Economic Systems, Stanford University, Stanford, CA.
- Howard, R.A., 1962. The used car buyer. In: Howard, R.A., Matheson J.E. (Eds.), *The Principles and Applications of Decision Analysis*, vol. 2. Strategic Decisions Group, Menlo Park, CA, pp. 719–62 (1984).
- Howard, R.A., 1990. From influence to relevance to knowledge. In: Oliver, R.M., Smith, J.Q. (Eds.), *Influence Diagrams, Belief Nets and Decision Analysis*. Wiley, Chichester, pp. 3–23.
- Howard, R.A., Matheson, J.E., 1981. Influence diagrams. In: Howard R.A., Matheson, J.E. (Eds.), 1984. *The Principles and Applications of Decision Analysis*, vol. 2. Strategic Decisions Group, Menlo Park, CA, pp. 719–762.
- Jensen, F., Jensen, F.V., Dittmer, S.L., 1994. From influence diagrams to junction trees. In: Mantaras, R.L., Poole, D. (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the 10th Conference*, Morgan Kaufmann, San Mateo, CA, pp. 367–373.
- Kirkwood, C.W., 1993. An algebraic approach to formulating and solving large models for sequential decisions under uncertainty. *Management Science* 39 (7), 900–913.
- Lauritzen, S.L., 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, Series B* 50 (2), 157–224.
- Liu, L., Shenoy, P.P., 1993. Belief potential networks for asymmetric decision problems. Working paper no. 256, School of Business, University of Kansas, Lawrence, KS.
- Miller III, A.C., Merkhofer, M.W., Howard, R.A., Matheson, J.E., Rice, T.R., 1976. Development of decision aids for decision analysis. Final technical report DO # 27742 Stanford Research Institute, Menlo Park, CA.
- Ndikilikiesha, P.C., 1992. A study of influence diagrams and their generalizations. Ph.D. thesis, School of Business, University of Kansas, Lawrence, KS.
- Ndikilikiesha, P.C., 1994. Potential influence diagrams. *International Journal of Approximate Reasoning* 10 (3), 251–285.
- Olmsted, S.M., 1983. On representing and solving decision problems. Ph.D. thesis, Department of Engineering–Economic Systems, Stanford University, Stanford, CA.
- Raiffa, H., Schlaifer, R.O., 1961. *Applied Statistical Decision Theory*, Harvard Business School, Cambridge, MA.
- Shachter, R.D., 1986. Evaluating influence diagrams. *Operations Research* 34 (6), 871–882.
- Shafer, G., 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.
- Shenoy, P.P., 1992a. Valuation-based systems for Bayesian decision analysis. *Operations Research* 40 (3), 463–484.
- Shenoy, P.P., 1992b. Valuation-based systems: A framework for managing uncertainty in expert systems. In: Zadeh, L.A., Kacprzyk, J. (Eds.), *Fuzzy Logic for the Management of Uncertainty*. Wiley, New York, pp. 83–104.
- Shenoy, P.P., 1993. A new method for representing and solving Bayesian decision problems. In: Hand, D.J. (Ed.), *Artificial Intelligence Frontiers in Statistics: AI and Statistics III*. Chapman and Hall, London, pp. 119–138.
- Shenoy, P.P., 1994a. Consistency in valuation-based systems. *ORSA Journal on Computing* 6 (3), 281–291.
- Shenoy, P.P., 1994b. A comparison of graphical techniques for decision analysis. *European Journal of Operational Research* 78 (1), 1–21.
- Shenoy, P.P., 1998. Game trees for decision analysis. *Theory and Decision* 44, 149–171.
- Shenoy, P.P., Shafer, G., 1990. Axioms for probability and belief-function propagation. In: Shachter, R.D., Levitt, T.S., Lemmer, J.F., Kanal, L.N. (Eds.), *Uncertainty in Artificial Intelligence*, vol. 4. North-Holland, Amsterdam,

- pp. 169–98 Reprinted by Shafer, G., Pearl, J. (Eds.) 1990. Readings in Uncertain Reasoning, vol. 4. Morgan Kaufmann, San Mateo, CA, pp. 575–610.
- Smith, J.E., Holtzman, S., Matheson, J.E., 1993. Structuring conditional relationships in influence diagrams. *Operations Research* 41 (2), 280–297.
- Tatman, J.A., 1986. Decision processes in influence diagrams: Formulation and analysis. Ph.D. thesis, Department of Engineering–Economic Systems, Stanford University, Stanford, CA.
- von Neumann, J., Morgenstern, O., 1944. *Theory of Games and Economic Behavior*, 1st ed. (2nd ed. 1947, 3rd ed. 1953). Princeton University Press, Princeton, NJ.