An Axiomatic Framework for Propagating Uncertainty in Directed Acyclic Networks

José Cano, Miguel Delgado, and Serafín Moral
Universidad de Granada, Granada, Spain

ABSTRACT

This paper presents an axiomatic system for propagating uncertainty in Pearl's causal networks, (Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, 1988 [7]). The main objective is to study all aspects of knowledge representation and reasoning in causal networks from an abstract point of view, independent of the particular theory being used to represent information (probabilities, belief functions or upper and lower probabilities). This is achieved by expressing concepts and algorithms in terms of valuations, an abstract mathematical concept representing a piece of information, introduced by Shenoy and Shafer [1, 2]. Three new axioms are added to Shenoy and Shafer's axiomatic framework [1, 2], for the propagation of general valuations in hypertrees. These axioms allow us to address from an abstract point of view concepts such as conditional information (a generalization of conditional probabilities) and give rules relating the decomposition of global information with the concept of independence (a generalization of probability rules allowing the decomposition of a bidimensional distribution with independent marginals in the product of its two marginals). Finally, Pearl's propagation algorithms are also developed and expressed in terms of operations with valuations.

KEYWORDS: ***causal network, uncertainty, hypertrees, PULCINELLA system, marginalization, combination, conditional information***

1. INTRODUCTION

Shenoy and Shafer [1, 2] have given an axiomatic framework for the propagation of uncertainty in hypergraphs. In this work the propagation

Address correspondence to Serafín Moral, Department of Artificial Intelligence, Campus de Fuente Nueva, University of Granada 18071, Granada, Spain.

Received January 1, 1992; accepted September 9, 1992.

International Journal of Approximate Reasoning 1993; 8:253–280

655 Avenue of the Americas, New York, NY 10010 0888-613X/93/$6.00

algorithms are abstracted from the particular theory being used to represent information. They introduce the primitive concept of valuation, which can be considered as the mathematical representation of a piece of information. A valuation may be particularized to a possibility distribution, a probability distribution, a belief function, etc. Then they develop and express propagation algorithms in terms of operations with valuations. These algorithms may be particularized to any concrete theory by translating valuations and operations to their special interpretation in this theory. These general algorithms have been implemented in the PULCINELLA system [3].

The notation used in this paper, and examples of what is a valuation in Probability Theory and Theory of Belief Functions, are as follows:

NOTATION Assume that we have an n-dimensional variable, $(X_1, \dots, X_n)$, each dimension, X_i, taking values on a finite set U_i. The following conventions will be followed:

- If $I \subseteq \{1, \dots, n\}$, we shall denote by X_I the $|I|$-dimensional variable ($|I|$ is the number of elements of set I), $(X_i)_{i \in I}$, and by U_I the cartesian product $\prod_{i \in I} U_i$, that is the set in which X_i takes its values.
- If $u \in U_I$, then we shall denote by u_i the ith coordinate of u, that is the element from U_i.
- If $u \in U_I$ and $J \subseteq I$, we shall denote by $u^{\downarrow J}$ the element from U_J obtained from u by dropping the extra coordinates; that is, the element given by $u_j^{\downarrow J} = u_j$, $\forall j \in J$.
- If $A \subseteq U_I$ and $J \subseteq I$, we shall denote by $A^{\downarrow J}$ the subset of U_J, given by

$$A^{\downarrow J} = \{v \in U_J | v = u^{\downarrow J}, u \in A\}.$$

EXAMPLE 1 In Probability Theory a valuation is the representation of a probabilistic piece of information about some of the variables, X_I, $I \subseteq \{1, \dots, n\}$. More concretely, if we have three variables (X_1, X_2, X_3) taking values on $U_1 \times U_2, \times U_3$, where $U_i = \{u_{i1}, u_{i2}\}$, $i = 1, 2, 3$, then a valuation may be a probability distribution about X_1,

$$p(u_{11}) = 0.8$$
$$p(u_{12}) = 0.2.$$

It may also be a conditional probability distribution about X_3 given X_2,

$$p(u_{31}|u_{21}) = 0.9 \qquad p(u_{32}|u_{21}) = 0.1$$
$$p(u_{31}|u_{22}) = 0.6 \qquad p(u_{32}|u_{22}) = 0.4.$$

From a mathematical point of view, a probabilistic valuation about variables X_I is a non-negative mapping,

$$p: U_I \to \Re_0^+,$$

where $\Re_0^+$ denotes the non-negative reals.

These mappings are not considered normalized, but are considered equivalent upon multiplication by a positive constant; that is, two valuations p_1, p_2 defined on the same frame U_I are considered equivalent if there exists a constant $\alpha > 0$, such that

$$\forall u \in U_I, \quad p_1(u) = \alpha \cdot p_2(u).$$

From strict mathematical point of view, a valuation should be considered an equivalence class on the set of non-negative mappings from U_I on $\Re_0^+$, under the above equivalence relation; however, to simplify the language and notation, we shall consider that a valuation is a mapping, but that two mappings are considered identical if they are equivalent.

EXAMPLE 2 For Belief Functions [4–6], a valuation about X_I is a non-necessarily normalized mass assignment on U_I, that is, a mapping

$$m: \mathscr{P}(U_I) \to \Re_0^+,$$

where $\mathscr{P}(U_I)$ is the set of all the subsets of U_I, and $m(\emptyset) = 0$.

Two basic operations are assumed to be defined among valuations: combination and marginalization. Combination is an operation to summarize in a single valuation the information of two valuations. If the two valuations to be combined are V_1 and V_2 defined on U_I and U_J, respectively, their combination will be denoted $V_1 \otimes V_2$ and will be defined on $U_{I \cup J}$.

Marginalization is an operation to calculate the information induced by a valuation defined on a frame U_I, on a less fine frame: U_J, where $J \subseteq I$. If V is the valuation defined on U_I, its marginalization to U_J is denoted by $V^{\downarrow J}$.

EXAMPLE 3 In the particular case of probabilistic valuations, *combination* is defined by point-wise multiplication. If p_1 and p_2 are non-negative functions defined on U_I and U_J, respectively, then $p_1 \otimes p_2$ is a mapping defined on $U_{I \cup J}$ to $\Re_0^+$ given by,

$$p_1 \otimes p_2(u) = p_1(u^{\downarrow I}) \cdot p_2(u^{\downarrow J}), \quad \forall u \in U_{I \cup J}.$$

This operation is used in probability to combine a marginal distribution with a conditional one to produce a bidimensional distribution, or used to calculate conditional information. Remember that as we are not concerned about normalization, conditioning to a set A may be considered as the multiplication with the likelihood associated to A (its characteristic function: $l_A(u) = 1$, if $u \in A$; $l_A(u) = 0$, otherwise).

Marginalization is defined in the usual way: If p is a valuation defined on U_I and $J \subseteq I$, then

$$p^{\downarrow J}(v) = \sum_{u^{\downarrow J} = v} p(u), \qquad \forall v \in U_J.$$

EXAMPLE 4 In Belief Functions [4, 5] combination is carried out by means of Dempster's rule:

$$m_1 \otimes m_2(A) = \sum_{B_1 \cap B_2 = A} m_1(B_1) \cdot m_2(B_2).$$

We do not normalize because two valuations are considered as equivalent if one is obtained from the other by multiplying by a positive constant.

If m is defined on U_I and $J \subseteq I$, then the marginalization of m to U_J is given by,

$$m^{\downarrow J}(A) = \sum_{B^{\downarrow J} = A} m(B)$$

Shenoy and Shafer [1, 2] show that if these two operations verify a system of three axioms, then the calculus with valuations may be done by means of propagation algorithms. More specifically, they show that if we have n valuations, $V_1, \ldots, V_n$, and we want to calculate $(V_1 \otimes V_2 \otimes \cdots \otimes V_n)^{\downarrow \{i\}}$ for each variable X_i, then we can do so without explicitly calculating the global valuation $V_1 \otimes V_2 \cdots \otimes V_n$, but by doing local computations among the initial valuations, arranged in an appropriate way. The advantage of avoiding the calculation of $V_1 \otimes V_2 \cdots \otimes V_n$ is that, in general, this calculation is very inefficient. For example, in the case of probabilities, if each variable X_i appears at least on a valuation V_j and we have m variables, then $V_1 \otimes V_2 \cdots \otimes V_n$ will be defined on $U_1 \times \cdots \times U_m$. If each U_i has k_i elements, we will need $\prod_{i=1}^{m} k_i$ values to specify this valuation. In the best case (all the k_i equal to 2) this number is 2^m.

Although Shenoy and Shafer, [1, 2], focus their work on the problem of calculus, there are other important aspects in the process of problem modeling and resolution that have not been considered. If we have two pieces of information represented by valuations V_1 and V_2, then their combination, $V_1 \otimes V_2$ does not always give rise to a meaningful or valid information for the problem. Consider, for example, two probabilistic valuations about variables X_1 and X_2: p, a bidimensional probability about (X_1, X_2), and p_1, a probability about X_1. The combination $p \otimes p_1$

makes no sense from a probabilistic viewpoint. It is not a valid probability for the two variables; however, if p is a conditional probability about X_2 given X_1, then $p_1 \otimes p$ is valid probability distribution about (X_1, X_2). It is a bidimensional probability. The concept of independence plays an important role in this kind of rules: If p_1 and p_2 are probability distributions for X_1 and X_2, respectively, and these variables are independent, then, $p_1 \otimes p_2$ is a valid probability for (X_1, X_2). If X_1 and X_2 are not independent, then this is not true.

Pearl [7] uses these rules in Probability Theory to show that in a causal network (directed acyclic graph), giving a conditional probability for each node given its parents determines one and only one global probability distribution for all the variables. In other words, the initial pieces of information are complete and coherent. These important issues are the main topics of the present paper. Shenoy, [8], also studies conditional independence for valuations in terms of factors of the joint valuation. In this paper, it is not assumed the existency of a joint valuation. We give rules to build more complex valuations from elemental ones using the given independence relationships. We also give conditions to determine one and only one joint valuation.

The graphical structures used to represent relationships among variables in our work are Pearl's causal networks, not Shenoy and Shafer's hypergraphs, because the former are more appropriate to represent independence relationships among variables in a direct way.

In the second section we introduce three new axioms for operations with valuations, and define, in an abstract way, the concepts of *conditional valuation*, *observation*, and '*a posteriori*' *information*. We then introduce rules to build new valuations from initial ones. In the third section we show that in a directed acyclic graph, having a valuation for each node, given its parents, determines one and only one global valuation valid for all the variables. In the fourth section, we obtain general propagation algorithms in directed acyclic graphs. These are a generalization of Pearl's algorithms, [7], but now are expressed in terms of operations with valuations, and, therefore, applicable to different uncertainty theories. Finally, in the last section we relate propagation algorithms with the independence relationships associated with the graph.

2. VALUATION-BASED SYSTEMS

In this section we describe how to represent information with valuations and how to do calculations with them, giving a method analogous to Bayes Theorem.

Let $X = (X_1, \ldots, X_n)$ be an n-dimensional variable such that each X_i

takes its values on a finite set U_i. A valuation is a primitive concept meaning the mathematical representation for a piece of information in a given uncertainty theory. We will assume that for each $I \subseteq \{1, \ldots, n\}$ there is a set V_I of valuations defined on the cartesian product, U_I. V will be the set of all valuations $V = \cup_{I \subseteq \{1,\ldots,n\}} V_I$.

Two basic operations are necessary (see Zadeh [9]; Shenoy, Shafer [1, 2]):

- *Marginalization.* If $J \subseteq I$ and $V_1 \in V_I$, then the marginalization of V_1 to J is a valuation $V_1^{\downarrow J}$ in V_J.
- *Combination.* If $V_1 \in V_I$ and $V_2 \in V_J$, then their combination is a valuation $V_1 \otimes V_2$ in $V_{I \cup J}$.

Shenoy and Shafer [1, 2], consider the following three axioms for these operations on valuations:

Axiom 1 $V_1 \otimes V_2 = V_2 \otimes V_1$, $(V_1 \otimes V_2) \otimes V_3 = V_1 \otimes (V_2 \otimes V_3)$.

Axiom 2 If $I \subseteq J \subseteq K$, and $V \in V_K$, then $(V^{\downarrow J})^{\downarrow I} = V^{\downarrow I}$.

Axiom 3 If $V_1 \in V_i$, $V_2 \in V_J$, then $(V_1 \otimes V_2)^{\downarrow I} = V_1 \otimes V_2^{\downarrow (J \cap I)}$.

We assume three more axioms,

Axiom 4 *Neutral Element.* There exists one and only one valuation V_0 defined on $U_1 \times \cdots \times U_n$ such that $\forall V \in V_I$, $\forall J \subseteq I$, $V_0^{\downarrow J} \otimes V = V$.

Axiom 5 *Contradiction.* There exists one and only one valuation, V_c, defined on $U_1 \times \cdots \times U_n$, such that $\forall V \in V$, $V_c \otimes V = V_c$.

Axiom 6 $\forall V \in V_\varnothing$, if $V \neq V_c^{\downarrow \varnothing}$, then $V = V_0^{\downarrow \varnothing}$.

The three first axioms provide the necessary conditions to deduce propagation algorithms. The third axiom is of particular importance to the development of propagation algorithms, as it allows us to calculate $(V_1 \otimes V_2)^{\downarrow I}$ without explicitly calculating $(V_1 \otimes V_2)$, a valuation defined on $U_{I \cup J}$. This can be done by calculating $V_2^{\downarrow (J \cap I)}$ and combining the result with V_1. In this last case we need only handle valuations on U_J, $U_{I \cap J}$, and U_I, which is much more efficient. Remember that, in general it is inefficient to handle valuations defined for a large number of variables.

The fourth axiom deals with the existence of the neutral element. This neutral element is considered by Shafer and Shenoy [1], but its existence is not postulated by an axiom. This axiom is essential in the present study to define the concept of conditional information in an abstract way. In the case of Probability Theory, the neutral element is a constant (non-zero) valuation:

$$p_0(u) = 1, \forall u \in U_1 \times \cdots \times U_n$$

The constant value is not important because probabilistic valuations are equivalent upon multiplication by a positive constant. In the case of *Belief Functions*, the neutral element is the so-called vacuous belief, [5, 6], given by

$$m_0(A) = \begin{cases} 1 & \text{if } A = U_1 \times \cdots \times U_n. \\ 0 & \text{otherwise} \end{cases}$$

In the Axiom 4, the existence of the neutral element is postulated on the frame corresponding to all the variables, $U_1 \times \cdots \times U_n$. In smaller frames, the neutral element is obtained by marginalization. Therefore, $V_0^{\downarrow K}$ will be called the *neutral element of* V_K, and when there is no chance of confusion, it will be denoted simply by V_0.

The contradiction is characterized in Axiom 5 as a valuation such that if it is combined with any other valuation, it produces the contradiction. In Probability Theory it is given by the zero-valued function

$$p_c(u) = 0, \qquad \forall u \in U_1 \times \cdots \times U_n$$

In Belief Functions Theory, the contradiction is the zero valued mass assignment,

$$m_c(A) = 0, \qquad \forall A \subseteq U_1 \times \cdots \times U_n$$

These valuations are obtained by combining two contradictory valuations. For example, in the case of Belief Functions, it can result from combining the masses, m_1 and m_2, defined on $\{u_1, u_2, u_3\}$ and given by

$m_1(\{u_1\}) = 1;$ $\qquad m_1(A) = 0$, otherwise

$m_2(\{u_2\}) = 0.5;$ $\qquad m_2(\{u_2, u_3\}) = 0.5;$ $\qquad m_2(A) = 0$, otherwise.

In Shenoy and Shafer [2], the contradiction is considered from a different standpoint: They define a subset of the set of valuations called the *family of proper valuations*. The elements of this set are the valuations different from the contradiction.

As in the case of the neutral element, $V_c^{\downarrow K}$ will be called the *contradiction of* V_K, and will sometimes be denoted by V_c.

According to Axiom 6, in the frame corresponding to the empty set of variables, $U_\varnothing$, all valuations are the contradiction or the neutral element. To better appreciate this, let us show its meaning in Probability Theory. The cartesian product $U_\varnothing$ has one element: $\{e\}$. A valuation, p, in this set is, therefore, a mapping from $\{e\}$ on $\Re_0^+$, that is, a number, $p(e)$. If this number is zero, then we have the contradiction; if it is different from zero, then we have the neutral element: combination with this valuation produces an equivalent valuation. Its meaning and consequences will be discussed below when we consider the relationships of the calculus with valuations and the concept of independence.

The following propositions illustrate some additional properties of valuations. We assume that Axioms 1–6 are verified.

PROPOSITION 1 $\forall V \in V_K,\ V \otimes V_c^{\downarrow K} = V_c^{\downarrow K}$

Proof On the basis of Axioms 5 and 3,

$$V_c^{\downarrow K} = (V_c \otimes V)^{\downarrow K} = V \otimes V_c^{\downarrow K}$$

Q.E.D. ■

PROPOSITION 2 $\forall V \in V_I$ being $I \subseteq K$, we have $V \otimes V_c^{\downarrow K} = V_c^{\downarrow K}$

Proof On the basis of Axioms 4, 1, and Proposition 1,

$$V \otimes V_c^{\downarrow K} = V \otimes (V_c^{\downarrow K} \otimes V_c^{\downarrow K}) = (V \otimes V_c^{\downarrow K}) \otimes V_c^{\downarrow K} = V_c^{\downarrow K}.$$

Proposition 1 has been applied in the last equality. ■

The following proposition is quite natural: if a valuation is marginalized on the variables in which it is defined, we obtain the same valuation; however, proof of this requires invocations of Axiom 4.

PROPOSITION 3 $\forall V \in V_I,\ V^{\downarrow I} = V$

Proof By applying Axioms 4, 3, 2, and 4,

$$V^{\downarrow I} = (V \otimes V_0^{\downarrow I})^{\downarrow I} = V \otimes (V_0^{\downarrow I})^{\downarrow I} = V \otimes V_0^{\downarrow I} = V.$$

■

The following is a technical proposition that will be used in later propagation algorithms.

PROPOSITION 4 If $V_1, V_2 \in V, I \subseteq \{1, 2, \ldots, n\}$, $V_1 \in V_J, V_2 \in V_K$ with $I \subseteq J \cup K, J \cap K \subseteq I$, then

$$(V_1 \otimes V_2)^{\downarrow I} = V_1^{\downarrow I \cap J} \otimes V_2^{\downarrow I \cap K}.$$

Proof On the basis of Axiom 2,

$$(V_1 \otimes V_2)^{\downarrow I} = \left((V_1 \otimes V_2)^{\downarrow J \cup I}\right)^{\downarrow I}.$$

As $(V_1 \otimes V_2)$ is a valuation on $J \cup K$ and we know that $I \subseteq J \cup K$, then by using Axiom 4, we obtain

$$(V_1 \otimes V_2)^{\downarrow I} = \left(((V_1 \otimes V_2) \otimes V_0^{\downarrow I})^{\downarrow J \cup I}\right)^{\downarrow I}.$$

On the basis of Axioms 1 and 3,

$$\begin{aligned}(V_1 \otimes V_2)^{\downarrow I} &= \left((V_1 \otimes V_2 \otimes V_0^{\downarrow I})^{\downarrow J \cup I}\right)^{\downarrow I} \\ &= \left(((V_1 \otimes V_0^{\downarrow I}) \otimes V_2)^{\downarrow J \cup I}\right)^{\downarrow I} \\ &= \left((V_1 \otimes V_0^{\downarrow I}) \otimes V_2^{\downarrow (J \cup I) \cap K}\right)^{\downarrow I} \\ &= (V_1 \otimes V_0^{\downarrow I} \otimes V_2^{\downarrow I \cap K})^{\downarrow I}\end{aligned}$$

The last equality is derived from the fact that if $(J \cap K) \subseteq I$, then $(J \cup I) \cap K = I \cap K$.

Now, using Axioms 1, 3, and 4 we get,

$$(V_1 \otimes V_2)^{\downarrow I} = \left((V_0^{\downarrow I} \otimes V_2^{\downarrow I \cap K}) \otimes V_1\right)^{\downarrow I} = (V_0^{\downarrow I} \otimes V_2^{\downarrow I \cap K} \otimes V_1^{\downarrow I \cap J}).$$

As the valuation $V_2^{\downarrow I \cap K} \otimes V_1^{\downarrow I \cap J}$ is defined on $(I \cap K) \cup (I \cap J) = I \cap (K \cup J)$, which—because $I \subseteq J \cup K$—is equal to I, this valuation is defined on I. Therefore, applying Axiom 4 we get

$$(V_1 \otimes V_2)^{\downarrow I} = (V_2^{\downarrow I \cap K} \otimes V_1^{\downarrow I \cap J}),$$

Q.E.D. ∎

Note that in this proof, multiplication by the neutral valuation on V_I, $V_0^{\downarrow I}$ is used, in general, to extend a valuation $V \in V_K$ to the set $V_{K \cup I}$. This method will be used several times throughout the article.

An immediate consequence of above proposition is the following one, which we give without proof.

PROPOSITION 5 If $V_1, V_2 \in V, I \subseteq \{1, 2, \ldots, n\}, V_1 \in V_J, V_2 \in V_K$ with $J \cap K = I$, then

$$(V_1 \otimes V_2)^{\downarrow I} = V_1^{\downarrow I} \otimes V_2^{\downarrow I}$$

The following proposition can be deduced using Axiom 6.

PROPOSITION 6 If $V_1, V_2 \in V, V_1 \in V_J, V_2 \in V_K$ with $J \cap K = \varnothing$ and $V_2^{\downarrow\varnothing} \neq V_c^{\downarrow\varnothing}$ then

$$(V_1 \otimes V_2)^{\downarrow J} = V_1.$$

According to this proposition, if we have two valuations, V_1 and V_2, given for disjoint sets of variables, J and K, and we combine them—obtaining a valuation for the set $J \cup K$—and if we then marginalize on the first set, then we obtain the same valuation V_1. This is not necessarily verified if Axiom 6 is not postulated as true. In that case, the combination of a valuation with a disjoint valuation, followed by marginalization, may affect this valuation. This may produce some incoherence between the concept of independence and the calculus with valuations to be given below. We shall consider this problem in more detail in section 5.

The main definitions relative to the calculus with valuations are given below.

DEFINITION 1 *A valuation $V \in V_I$ is said to be absorbent if and only if it is not the contradiction in V_I and* $(\forall V' \in V_I)((V \otimes V' = V)$ *or* $(V \otimes V' = V_c))$.

If a valuation from V_I represents a piece of information about the values of variables X_I, then an absorbent valuation represents perfect knowledge about these values: it cannot be consistently refined by combination with other valuation: We can obtain only the same information or the contradiction.

EXAMPLE 5 In Belief Functions, absorbent valuations are mass assignments with a positive value on an unitary set only:

$$m(\{u_0\}) = 1, \qquad m(A) = 0 \quad \text{otherwise}.$$

If one of these valuations is combined with another mass assignment, m', we then obtain the contradiction (if $m'(A) = 0, \forall A$, such that $u_0 \in A$), or an equivalent mass assignment.

DEFINITION 2 *If $V \in V_{I \cup J}$, then it is said that V is a valuation on U_I conditioned to U_J, if and only if $V^{\downarrow J} = V_0 \in V_J$, the neutral element on V_J. The subset of $V_{I \cup J}$ given by the valuations on U_I conditioned to U_J will be denoted by $V_{I|J}$.*

This is an abstract definition of conditional valuation. If V is a valuation on U_I conditioned to U_J, then it may give some information about variables X_I and their relationships with variables X_J, but not about variables X_J. It is thus defined as a valuation such that marginalizing it on U_J gives the neutral element; that is, it does not say anything about X_J.

EXAMPLE 6 Assume that we have two variables, X_1 and X_2, taking values on U_1 and U_2, where $U_i = \{u_{i1}, u_{i2}\}$, $i = 1,2$. A probabilistic conditional valuation on U_2 given U_1, is a mapping

$$p: U_1 \times U_2 \to [0,1]$$

such that, marginalized on U_1, we get the neutral valuation (a constant valuation). For example,

$$p(u_{11}, u_{21}) = 0.8 \qquad p(u_{11}, u_{22}) = 0.2$$
$$p(u_{12}, u_{21}) = 0.3 \qquad p(u_{12}, u_{22}) = 0.7$$

Marginalizing on U_1, we get the neutral element

$$p^{\downarrow\{1\}}(u_{11}) = 1; \qquad p^{\downarrow\{1\}}(u_{12}) = 1.$$

If we consider $p(u_{1i}, u_{2j})$ as $P(X_2 = u_{2j}|X_1 = u_{1i})$, then the condition of marginalization is equivalent to the condition that all the conditional probability distributions have the same normalization factor.

EXAMPLE 7 Consider the same variables and sets as in Example 6, but with mass assignments as valuations. A conditional valuation on U_2 given U_1 is a mass assignment in $U_1 \times U_2$ such that if it is marginalized on U_1, then we get the neutral element. As the neutral element is given by a valuation with a positive mass only on set U_1, and taking into account the definition of marginalization (see Example 4), a conditional valuation, m, is characterized by the following:

$$\textit{If } A \subseteq U_1 \times U_2 \quad \text{and} \quad m(A) > 0, \text{Then } A^{\downarrow\{1\}} = U_1,$$

that is, the projection of all the subsets with positive mass is equal to the whole set U_1. As an example, consider the mass assignment given by

$$m(\{(u_{11}, u_{22}), (u_{12}, u_{22})\}) = 0.3;$$

$$m(\{(u_{11}, u_{21}), (u_{12}, u_{22})\}) = 0.5;$$

$$m(\{(u_{11}, u_{21}), (u_{12}, u_{21}), (u_{12}, u_{22})\}) = 0.2$$

We base our definitions of dependence and conditional independence, on Pearl's assumption, [7, 10, 11], that they are primitive concepts. Given a set of variables, then they have an associated set of (in)dependence relationships among them. These relationships are of the type: 'X_1 and X_2 are independent, but they are dependent if we know X_3', '(X_1, X_2) is dependent of (X_4, X_5), but they are independent if we know X_6', etc.... This is knowledge of a qualitative nature. It may be deduced from numbers, if for example, we know a global probability distribution for all the variables of the problem, or we have a sample of the population in which statistical tests of independence can be carried out. However, this is not always the case. We may know the (in)dependencies of very well-structured problems, in which we know the relationships among the variables of the problem before any numerical value. Below we give Pearl's axiomatic definition of (in)dependence relationships among a set of variables.

DEFINITION 3 *Given a family of variables* $(X_1, \ldots, X_n)$, *a dependence structure on it is a mapping* D: $\mathcal{P}(\{1, \ldots, n\}) \times \mathcal{P}(\{1, \ldots, n\}) \times \mathcal{P}(\{1, \ldots, n\}) \to \{0, 1\}$, *where if* $D(I, J, K) = 0$, *then* X_I *is said to be independent of* X_K *given* X_J, *verifying the following axioms* [10, 11].

- *Symmetry. If* $D(I, J, K) = 0$, *then* $D(K, J, I) = 0$ *and viceversa.*
- *Decomposition. If* $D(I, J, K \cup L) = 0$, *then* $D(I, J, K) = 0$ *and* $D(I, J, L) = 0$.
- *Weak Union. If* $D(I, J, K \cup L) = 0$, *then* $D(I, J \cup L, K) = 0$.
- *Contraction. If* $D(I, J, K) = 0$ *and* $D(I, J \cup K, L) = 0$, *then* $D(I, J, K \cup L) = 0$.

An intuitive interpretation of these axioms can be found in Pearl, [7].

A valuation on U_I is the mathematical representation of information about how X_I takes its values. The following properties establish how to build more complex valuations from elemental ones. These properties are specified on the definition of system of information. A system of information for a problem aims to include all valuations representing the available initial information for this problem and, all valuations that may be deduced from them.

DEFINITION 4 *Let* $(X_1, \ldots, X_n)$ *be an n-dimensional variable taking values on* $U_1 \times \cdots \times U_n$, *and D be an associated dependence structure. A system of information about this variable with respect to D is a family* $H \subseteq V$ *and a mapping*

$$h: H \to \mathcal{P}(\{1, \ldots, n\}) \times \mathcal{P}(\{1, \ldots, n\})$$

with the following properties:

1. $\forall V \in H$, *if* $h(V) = (I, J)$, *then* $V \in V_{I|J}$ *and* $I \cap J = \emptyset$. *It is said that V is a valid valuation about variables* X_I *conditioned to (or given) variables* X_J. *If* $J = \emptyset$, *then it is said simply that V is a valid valuation about* X_I.
2. *If* $V_1, V_2 \in H$ *and* $h(V_1) = (J, \emptyset)$ *and* $h(V_2) = (I, J)$, *then* $V_1 \otimes V_2 \in H$ *with* $h(V_1 \otimes V_2) = (I \cup J, \emptyset)$.
3. *If* $V \in H$ *with* $h(V) = (I, J)$ *and* $D(I, J, K) = 0$, *then* $V \otimes V_0^{\downarrow K} \in H$ *with* $h(V \otimes V_0^{\downarrow K}) = (I, J \cup K)$.
4. *If* $V \in H$ *with* $h(V) = (I, J)$, *then if* $K \subseteq I$, $V^{\downarrow K \cup J} \in H$ *with* $h(V^{\downarrow K \cup J}) = (K, J)$.

Property 1 says that valid valuations are always associated with a set of variables with respect to another set of conditioning. Property 2 explains how to construct a valid valuation about $X_{I \cup J}$ from a valuation about X_J and another about X_I given X_J. Property 3 relates the system of information with the associated dependence structure. $V \otimes V_0^{\downarrow K}$ stands for the extension of $V \in V_{I|J}$ to the set $V_{I|(J \cup K)}$. It says that if X_I and X_K are independent given X_J, then a conditional valuation about X_I given X_J may be extended to a conditional valuation about X_I given $X_{J \cup K}$, by combination with the neutral element. Property 4 says that the marginalization of a valid valuation is a valid valuation.

There is no axiom saying that the contradiction V_c does not belong to H; however, it is clear that in a system that includes the contradiction, some of the initial pieces of information are not true. In this case it would be interesting to find methods to resolve the inconsistency, by changing the minimum number of valuations.

Another interesting problem is that of completeness considered in the following definition.

DEFINITION 5 *A system of information* (H, h), *about* $(X_1, \ldots, X_n)$ *with associated dependence structure D is said to be complete if and only if there is a valuation* $V \in H$ *such that* $h(V) = (\{1, \ldots, n\}, \emptyset)$.

That is, completeness is equivalent to the determination of a global valid unconditional valuation for all the variables.

EXAMPLE 8 If we have three variables, X_1,X_2, and X_3, and all of them are dependent, then if we have the system of information generated by the following valuations

- V_1 an unconditional valuation about X_1, $h(V_1) = (\{1\}, \emptyset)$,
- V_2 a valuation about X_2 given X_1, $h(V_2) = (\{2\}, \{1\})$,
- V_3 a valuation about X_3 given X_2, $h(V_3) = (\{3\},\{2\})$,

this system is not complete. $V_1 \otimes V_2$ is an unconditional valuation about (X_1, X_2), but there is no rule that could be applied to obtain a valid valuation for all three variables of the problem. If V_3 were a valuation about X_3 given (X_1, X_2), then the combination $V_1 \otimes V_2 \otimes V_3$ would be a global unconditional valuation for the three variables, $h(V_1 \otimes V_2 \otimes V_3) = (\{1, 2, 3\}, \emptyset)$ and the system would be complete.

DEFINITION 6 *A system of information* (H, h) *is said to be complete and deterministic if and only if there exists one and only one valuation* $V \in H$ *such that* $h(V) = (\{1, \ldots, n\}, \emptyset)$.

It is important that the systems of information be deterministic. In a non-deterministic system there is more than one valid valuation for all the variables. In some theories, for example, upper and lower probabilities, this may be possible. But in others, such as in Probability Theory, this is impossible: we can only have one global probability for all variables.

EXAMPLE 9 If we have two variables X_1 and X_2 and the system of information (H, h) generated by the following probabilistic valuations (the system of information containing these valuations and all the valuations obtained by applying any of the rules of Definition 4),

- p_1 and p_2 two unconditional probabilities about X_1 and X_2, respectively,
- p_3 and p_4 a conditional probability about X_2 given X_1 and about X_1 given X_2, respectively,

then $p_1 \otimes p_3$ and $p_2 \otimes p_4$ are two valid, unconditional valuations for all the variables. If they are different, then the system is incoherent.

Two more definitions are needed to complete this abstract description of valuations and conditional valuations. These definitions are used to introduce the role of observations. Thus far, a valuation has been considered a piece of information for all the elements of a population. In practice, we usually start with a complete and deterministic system of information, representing our background knowledge. We then have a particular case, about which we want to make some inferences: We observe some data, and we want to obtain the information about some variables that can be deduced from the initial general knowledge and the observations for this particular case.

DEFINITION 7 *A family of observations about an n-dimensional variable* $(X_1, \ldots, X_n)$ *is a set of valuations* $\{O_i\}_{i \in I}$, *where* $I \subseteq \{1, \ldots, n\}$, *and* O_i *is an absorbent valuation on* U_i.

Observations can be characterized (see Definition 1) as valuations for a variable that may not be refined by combination without obtaining the contradiction, that is, they represent perfect knowledge about the value of these variables.

DEFINITION 8 *If* (H, h) *is a complete and deterministic system of information about the variables* $(X_1, \ldots, X_n)$ *and* $\{O_i\}_{i \in I}$ *a family of observations about these variables, we call* $((\otimes_{i \in I} O_i) \otimes V)^{\downarrow J}$ *the 'a posteriori' information about variables* X_J *induced by* (H, h) *and* $\{O_i\}_{i \in I}$, *where* $J \subseteq \{1, \ldots, n\}$ *and V is the only valid global valuation for all the variables* $(X_1, \ldots, X_n)$.

3. VALUATIONS AND DIRECTED ACYCLIC GRAPHS

In the following, we relate valuations with directed acyclic graphs. We shall show that graphs of this type have an associated dependence structure (as given by Pearl [7]) and are very appropriate for specifying complete and deterministic systems of information.

DEFINITION 9 *A directed graph is a pair* (T, E) *where T is a finite set and E a subset of* $T \times T$. *The elements of T are called nodes or vertices, and if* $e = (t_1, t_2) \in E$, *then e is said to be an arc going from* t_1 *to* t_2, t_1 *is said to be a parent of* t_2, *and* t_2 *is said to be a son of* t_1.

DEFINITION 10 *A directed path on a graph* (T, E) *is a sequence of arcs,* $e_1, \ldots, e_k \in E$, *such that if* $e_i = (t_1^i, t_2^i)$, *then* $\forall i \in \{2, \ldots, k\}$, $t_1^i = t_2^{i-1}$.

DEFINITION 11 *An undirected path on a graph* (T, E) *is a sequence* $e_1, \ldots, e_k \in T \times T$, *such that if* $e_i = (t_1^i, t_2^i)$, *then* $\forall i \in \{1, \ldots, k\}$ $((t_1^i, t_2^i) \in E$ *or* $(t_2^i, t_1^i) \in E)$ *and* $\forall i \in \{2, \ldots, k\}(t_1^i = t_2^{i-1})$.

DEFINITION 12 *A directed acyclic graph (DAG) is a graph* (T, E) *such that for each directed path,* $e_1, \ldots, e_k$, *defined on it,* $t_1^1 \neq t_2^k$.

A directed acyclic graph is said to be associated with the n-dimensional variable $(X_1, \ldots, X_n)$ if the set of vertices is $\{X_1, \ldots, X_n\}$, that is, if there is a vertex for each variable. These graphs can be used to represent a dependence structure for these variables, as described by Pearl, [7]. This structure is given by the following definitions.

DEFINITION 13 *Given an n-dimensional variable* $(X_1, \ldots, X_n)$ *and* (T, E) *a graph with* $T = \{X_i\}_{i \in \{1,\ldots,n\}}$, *we call dependence structure associated with* (T, E) *to the structure given by* $D(I, J, K) = 0$ *if and only if for every undirected path between* $\{X_i\}_{i \in I}$ *and* $\{X_k\}_{k \in K}$, *one of the following conditions is true (d-separation criterion,* [7]):

- *There exists a node with converging arrows in the path that does not belong to* $\{X_j\}_{j \in J}$ *and its descendents do not belong to* $\{X_j\}_{j \in J}$.
- *There exists a node without converging arrows in the path that belong to* $\{X_j\}_{j \in J}$.

A DAG can form the basis to describe a system of information.

DEFINITION 14 *If* $(X_1, \ldots, X_n)$ *is an n-dimensional variable and* (T, E) *is a graph with* $T = \{X_i\}_{i \in \{1,\ldots,n\}}$, *then* (H, h) *is said to be a system of information defined on* (T, E) *if and only if* (H, h) *is the information system generated by a set of valuations* $\{V_i\}_{i \in \{1,\ldots,n\}}$ *where* $h(V_i) = (i, P(i))$ *and* $P(i) = \{j | (X_j, X_i) \in E\}$, *that is, the set of parents of* X_i, *and where the dependency structure is the one associated with* (T, E).

Systems of information defined on directed acyclic graphs are very appropriate because they are always complete and deterministic. By introducing for each variable a valuation conditioned to its parents, we are able to obtain a global valuation for all the variables, and this valuation is unique.

PROPOSITION 7 If (H, h) is a system of information defined on a graph (T, E) with $T = \{X_i\}_{i \in \{1,\ldots,n\}}$, then it is complete and deterministic about $(X_1, \ldots, X_n)$ where the structure of dependencies is that associated with (T, E).

Proof Assume that $\{V_i\}_{i \in I}$ are the valuations generating the system of information, (H, h), defined on (T, E). The proof is based on the system of information (H', h') given by the following valuations:

- $M_I = (V_1 \otimes V_2 \otimes \cdots \otimes V_n)^{\downarrow I}$, where $I \subseteq \{1, 2, \ldots, n\}$, with $h'(M_I) = (I, \emptyset)$.
- $R_{I,i} = (V_i \otimes V_0^{\downarrow I})$, where $I \subseteq \{1, 2, \ldots, n\} - D_i$, D_i is the set of descendents of X_i including i, and $h'(R_{I,i}) = (\{i\}, I \cup P(i))$, $P(i)$ being the set of parents of X_i.
- $S_{I,J} = (M_I \otimes V_0^{\downarrow J})$, where $I \subseteq \{1, 2, \ldots, n\}$, and $J \subseteq \{j | X_j$ is not connected with a variable from $\{X_i\}_{i \in I}\}$, $h'(S_{I,J}) = (I, J)$.

Then the following facts are verified.

a. (H', h') is actually a system of information.

b. $V_i \in H'$, $\forall i \in \{1, \ldots, n\}$, with $h'(V_i) = (i, P(i))$.

c. If for a system of information (H'', h'') with the same dependence structure, (**b**) is verified, then M_I, $R_{I,i}$, $S_{I,J}$ belong to H''.

d. The only global valuation, V, in H' with $h(V) = (\{1, 2, \ldots, n\}, \emptyset)$ is $V_1 \otimes V_2 \otimes \cdots \otimes V_n$.

If these points are proved, then from (**a**) and (**b**) we conclude that $H \subseteq H'$, H being the system of information generated by valuations $\{V_i\}_{i \in \{1, 2, \ldots, n\}}$. Taking into account (**c**), we have the equality $H = H'$, because H is a system of information that verifies (**b**). According to (**d**), there is one and only one global valuation defined for all the variables, $\{1, \ldots, n\}$: $V_1 \otimes V_2 \otimes \cdots \otimes V_n$. Now all we need is to prove these points.

To show (**a**), we need to prove the four properties of systems of information:

1. It is immediate by the way valuations M_I, R_I, $S_{I,J}$ are defined.
2. Property 2 may be applied only to valuations $R_{I,i}$, $S_{I,J}$ combined with valuations M_I. It is verified because the combination $R_{I,i} \otimes M_{I \cup P(i)} = M_{I \cup P(i) \cup \{i\}}$, and the combination $S_{I,J} \otimes M_J = M_{I \cup J}$, that is, elements from H'.
3. The combination of a valuation M_I with $V_0^{\downarrow J}$, with $D(I, \emptyset, J) = 0$, is precisely $S_{I,J}$. Applying this rule to valuations in the form of $R_{I,i}$ and $S_{I,J}$ yields elements of the same type.
4. Marginalization of elements M_I and $S_{I,J}$ produces elements of the same type. The only proper marginalization operations applicable to elements $R_{I,i}$ are to marginalize to $U_{I \cup P(i)}$ and to $U_{I \cup P(i) \cup \{i\}}$. In the first case, as $R_{I,i}$ is a valuation conditioned on variables $(X_j)_{j \in I \cup P(i)}$, yields the neutral valuation on $U_{I \cup P(i)}$ which is equal to $S_{\emptyset, P(i)}$. The second is not really a marginalization. We get the same valuation $R_{I,i}$.

The proof of (**b**), (**c**), and (**d**) is straightforward and will not be described in detail. ■

4. PROPAGATION OF VALUATIONS IN POLYTREES

Assume that $(X_1, \ldots, X_n)$ is an n-dimensional variable and (H, h) is a system of information defined on the graph (T, E) and given by valuations $\{V_i\}_{i \in \{1, \ldots, n\}}$. Under these conditions, the only possible global valuation for all the variables is $V_1 \otimes V_2 \otimes \cdots \otimes V_n$. Consider also a family of observations $\{O_i\}_{i \in I}$, and that our objective is to calculate the 'a posteriori' valuations for each single variable, that is, to calculate the following valuations:

$$PS_j = \left(\bigotimes_{i \in I} O_i \otimes V \right)^{\downarrow \{j\}}$$

for each $j \in \{1, \dots, n\}$, where $V = V_1 \otimes \cdots \otimes V_n$ ____ is the only existing global information.

The main problem, as indicated in the introduction, is to calculate, $V_1 \otimes \cdots \otimes V_n$. Propagation algorithms proposed by Pearl, [7], avoid the calculation of $V_1 \otimes \cdots \otimes V_n$ if the graph does not have undirected cycles (also called *loops*). These graphs are called polytrees, [7]. In this section we generalize these algorithms to the uncertainty expressed by means of valuations.

The notation will be based on the following valuations,

$$H_k = V_k, \text{ if } k \notin I$$
$$H_k = V_k \otimes O_k, \text{ if } k \in I$$
$$O'_j = V_0 \in V_{\{j\}}, \text{ if } j \notin I$$
$$O'_j = O_j, \text{ if } j \in I$$

and the following sets for each $j \in \{1, \dots, n\}$,

- I_j^-, the set of indexes i such that X_i is a descendent of X_j (there is a directed path from X_j to X_i).
- $I_j^+ = \{1, \dots, n\} - (I_j^- \cup \{j\})$.
- $P(j)$ is the set of parents of node X_j.
- For each $k \in P(j)$, I_{jk}^+ is the set of indexes of nodes X_i such that there is an undirected path from X_i to X_k or from X_k to X_i not containing X_j.
- $C(j)$, is the set of children of node X_j.
- For each $k \in C(j)$, I_{jk}^-, the set of indexes of nodes connected with X_k via an undirected path not containing X_j.
- I_{j0}^+ is the set of indexes of nodes not connected with node X_j.

The following propositions form the basis of propagation algorithms. In all of them, it is assumed that the graph has no loops.

PROPOSITION 8 $\forall j \in \{1, \dots, n\}$, $PS_j = \pi_j \otimes \lambda_j$, where

$$\pi_j = \left[\left(\bigotimes_{i \in I_j^+} H_i \right) \otimes V_j \right]^{\downarrow \{j\}}, \qquad \lambda_j = \left(\bigotimes_{i \in I_j^-} H_i \right)^{\downarrow \{j\}} \otimes O'_j$$

Proof Taking into account the expression of PS_j and the definition of H_j,

$$PS_j = (H_1 \otimes H_2 \otimes \cdots \otimes H_n)^{\downarrow \{j\}}$$

Taking into account Axioms 1 and 2,

$$PS_j = (H_1 \otimes H_2 \otimes \cdots \otimes H_n)^{\downarrow\{j\}}$$

$$= \left(\left(\bigotimes_{i \in I_j^+} H_i \right) \otimes H_j \otimes \left(\bigotimes_{i \in I_j^-} H_i \right) \right)^{\downarrow\{j\}}$$

$$= \left(\left(\bigotimes_{i \in I_j^+} H_i \otimes V_j \right) \otimes \left(O_j' \otimes \left(\bigotimes_{i \in I_j^-} H_i \right) \right) \right)^{\downarrow\{j\}}.$$

On the basis of Proposition 5, and taking into account that $(I_j^+ \cup \{j\}) \cap (I_j^- \cup \{j\}) = \{j\}$,

$$PS_j = \left(\left(\bigotimes_{i \in I_j^+} H_i \right) \otimes V_j \right)^{\downarrow\{j\}} \otimes \left(O_j' \otimes \left(\bigotimes_{i \in I_j^-} H_i \right) \right)^{\downarrow\{j\}}$$

$$= \left(\left(\bigotimes_{i \in I_j^+} H_i \right) \otimes V_j \right)^{\downarrow\{j\}} \otimes \left(O_j' \otimes \left(\bigotimes_{i \in I_j^-} H_i \right)^{\downarrow\{j\}} \right) = \pi_j \otimes \lambda_j.$$

Q.E.D. ■

PROPOSITION 9 If $(H_1 \otimes H_2 \otimes \cdots \otimes H_n)^{\downarrow\emptyset}$ is different from the contradiction, then

$$\forall j \in \{1, \ldots, n\}, \quad \pi_j = \left[V_j \otimes \left(\bigotimes_{k \in P(j)} \pi_j^k \right) \right]^{\downarrow\{j\}}; \qquad \lambda_j = \left(\bigotimes_{k \in C(j)} \lambda_j^k \right) \otimes O_j'$$

where

$$\pi_j^k = \left(\bigotimes_{i \in I_{jk}^+} H_i \right)^{\downarrow\{k\}}; \qquad \lambda_j^k = \left(\bigotimes_{i \in I_{jk}^-} H_i \right)^{\downarrow\{j\}}$$

Proof The expression for π_j is

$$\pi_j = \left[\left(\bigotimes_{i \in I_j^+} H_i \right) \otimes V_j \right]^{\downarrow\{j\}}.$$

Taking into account that the graph has no loops, we can show that

$I_j^+ = \bigcup_{k \in P(j)} I_{jk}^+$ with $I_{jk_1}^+ \cap I_{jk_2}^+ = \varnothing$ if $k_1 \neq k_2$. Therefore,

$$\pi_j = \left[\left[\bigotimes_{k \in P(j)} \left(\bigotimes_{i \in I_{jk}^+} H_i\right) \otimes V_j\right]^{\downarrow (\{j\} \cup P(j))}\right]^{\downarrow \{j\}}$$

and on the basis of Axiom 3,

$$\pi_j = \left[V_j \otimes \left[\bigotimes_{k \in P(j)} \left(\bigotimes_{i \in I_{jk}^+} H_i\right)\right]^{\downarrow P(j)}\right]^{\downarrow \{j\}}.$$

Now, by a repeated application of Proposition 4,

$$\pi_j = \left[V_j \otimes \left[\bigotimes_{k \in P(j)} \left(\bigotimes_{i \in I_{jk}^+} H_i\right)^{\downarrow \{k\}}\right]\right]^{\downarrow \{j\}}.$$

Taking into account the expression for π_j^k, we obtain the desired equality for π_j,

$$\pi_j = \left[V_j \otimes \left(\bigotimes_{k \in P(j)} \pi_j^k\right)\right]^{\downarrow \{j\}}.$$

The valuation λ_j was defined as

$$\lambda_j = \left(\bigotimes_{i \in I_j^-} H_i\right)^{\downarrow \{j\}} \otimes O_j'.$$

As the graph has no loops, $I_j^- = \bigcup_{k \in C(j) \cup \{0\}} I_{jk}^-$, with $I_{jk_1}^- \cap I_{jk_2}^- = \varnothing$, if $k_1 \neq k_2$. Therefore,

$$\lambda_j = \left[\bigotimes_{k \in C(j) \cup \{0\}} \left(\bigotimes_{i \in I_{jk}^-} H_i\right)\right]^{\downarrow \{j\}} \otimes O_j'.$$

By a repeated application of Proposition 5,

$$\lambda_j = \left[\bigotimes_{k \in C(j)} \left(\bigotimes_{i \in I_{jk}^-} H_i\right)^{\downarrow \{j\}}\right] \otimes \left(\bigotimes_{i \in I_{j0}^-} H_i\right)^{\downarrow \varnothing} \otimes O_j'.$$

Taking into account the expression of λ_j^k,

$$\lambda_j = \left(\bigotimes_{k \in C(j)} \lambda_j^k\right) \otimes \left(\bigotimes_{i \in I_{j0}^-} H_i\right)^{\downarrow \varnothing} \otimes O_j'.$$

As $(H_1 \otimes \cdots \otimes H_n)^{\downarrow \emptyset}$ is different from the contradiction then $(\otimes_{i \in I_{j0}^-} H_i)^{\downarrow \emptyset}$ is also different from the contradiction, and taking into account Axiom 6,

$$\left(\bigotimes_{i \in I_{j0}^-} H_i \right)^{\downarrow \emptyset} = V_0^{\downarrow \emptyset}.$$

Thus,

$$\lambda_j = \left(\bigotimes_{k \in C(j)} \lambda_j^k \right) \otimes V_0^{\downarrow \emptyset} \otimes O_j' = \left(\bigotimes_{k \in C(j)} \lambda_j^k \right) \otimes O_j'.$$

Q.E.D. ∎

We now have the valuations we want to calculate, PS_j, in terms of valuations π_j and λ_j, which are expressed in terms of valuations λ_j^k and π_j^k. In the next proposition, we show how we calculate these valuations in a node, assuming that the corresponding valuations in all the neighboring nodes are known.

PROPOSITION 10 If $j \in \{1, \ldots, n\}$, then $\forall k \in P(j)$,

$$\pi_j^k = \left[\pi_k \otimes O_k' \otimes \left(\bigotimes_{i \in C(k),\, i \neq j} \lambda_k^i \right) \right].$$

And $\forall k \in C(j)$,

$$\lambda_j^k = \left[\lambda_k \otimes O_k' \otimes V_k \otimes \left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^i \right) \right]^{\downarrow \{j\}}.$$

Proof If $k \in P(j)$, then

$$\pi_j^k = \left(\bigotimes_{i \in I_{jk}^+} H_i \right)^{\downarrow \{k\}}$$

$$= \left[\left(\bigotimes_{i \in I_k^+} H_i \right) \otimes H_k \otimes \left(\bigotimes_{i \in C(k),\, i \neq j} \left(\bigotimes_{l \in I_{ki}^-} H_l \right) \right) \right]^{\downarrow \{k\}}$$

$$= \left[\pi_k \otimes O_k' \otimes \left(\bigotimes_{i \in C(k),\, i \neq j} \lambda_k^i \right) \right].$$

Analogously, if $k \in C(j)$,

$$\lambda_j^k = \left(\bigotimes_{i \in I_{jk}^-} H_i \right)^{\downarrow \{j\}}$$

$$= \left[\left(\bigotimes_{i \in I_k} H_i \right) \otimes O_k' \otimes V_k \otimes \left(\bigotimes_{i \in P(k),\, i \neq j} \left(\bigotimes_{l \in I_{ki}^+} H_l \right) \right) \right]^{\downarrow \{j\}}$$

$$= \left[\lambda_k \otimes O_k' \otimes V_k \otimes \left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^i \right) \right]^{\downarrow \{j\}} .$$

Q.E.D. ■

The contradiction raises the following problem. If we have a graph with two disconnected parts, for example one with vertices $\{1, 2, \ldots, i\}$ and other with vertices $\{i + 1, \ldots, n\}$ (that is, there is no arc from vertices of one set to vertices of the other set), then if $H_1 \otimes \cdots \otimes H_i$ is the contradiction in $U_{\{1,\ldots,i\}}$, it can be shown that $H_1 \otimes \cdots \otimes H_n$ is the global contradiction and PS_j is the contradiction for each variable X_j, $j \in \{1, 2, \ldots, n\}$. However, if we use the above formulas, then the contradiction is not propagated between non-connected parts of the graph and what we obtain for PS_j, $j \in \{i + 1, \ldots, n\}$ is $(H_{i+1} \otimes \cdots \otimes H_n)^{\downarrow \{j\}}$, which is not PS_j if there is no contradiction in variables $X_{i+1}, \ldots, X_n$. What to do from a practical point of view? The answer is simple. If the graph is connected, then the propagation formulas are correct. We have used the lack of contradiction to show that

$$\left(\bigotimes_{i \in I_{j0}^-} H_i \right)^{\downarrow \emptyset} = V_0^{\downarrow \emptyset}$$

and in this case I_{j0}^- is empty.

If the graph is not connected, then for a global contradiction to exist, some of the connected parts have to be contradictory. In such a case, from a mathematical point of view all values of PS_j will be contradictory. But there is an alternative: not to consider the information from variables giving rise to the contradiction, removing such contradictory information from the system and keeping only non-contradictory valuations. Shenoy [12] describes an efficient way to isolate a maximal consistent set of valuations. With this restricted system, the formulas of propagation can be used.

Our objective is now to calculate the values π_j, λ_j, π_j^k $(k \in C(j))$, λ_j^k $(k \in P(j))$, for each $j \in \{1, \ldots, n\}$. If the set of observations, $\{O_i\}_{i \in I}$ is empty $(I = \emptyset)$, then this can be done very easily. First, we shall prove that all valuations λ_i are the neutral valuation.

PROPOSITION 11 If $I = \emptyset$ and $V_c \neq (V_1 \otimes \cdots \otimes V_n)^{\downarrow\emptyset}$ then $\lambda_j = V_0 \in V_{\{j\}}$, $\forall j \in \{1, \ldots, n\}$.

Proof For every leaf variable in (T, E), we have $\lambda_j = (\otimes_{k \in C(j)} \lambda_j^k) \otimes O_j' = V_0$. If X_j is a leaf (a node with no children) then $C(j) = \emptyset$, therefore, $O_j' = V_0$.

Now let us prove that if $\lambda_k = V_0^{\downarrow\{k\}} \in V_{\{k\}}$ for every $k \in C(j)$ then $\lambda_j = V_0^{\{j\}} \in V_{\{j\}}$.

$$\lambda_j = \left(\bigotimes_{k \in C(j)} \lambda_j^k \right) \otimes O_j' = \left(\bigotimes_{k \in C(j)} \lambda_j^k \right)$$

and

$$\lambda_j^k = \left[\lambda_k \otimes O_k' \otimes V_k \left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^i \right) \right]^{\downarrow\{j\}}$$

$$= \left[V_k \otimes \left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^i \right) \right]^{\downarrow\{j\}} = \left[\left(V_k \otimes \left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^i \right) \right)^{\downarrow P(k)} \right]^{\downarrow\{j\}}$$

$$= \left[\left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^i \right) \otimes V_k^{\downarrow P(k)} \right]^{\downarrow\{j\}}.$$

Now, taking into account that $V_k \in V_{\{k\}|P(k)}$,

$$\lambda_j^k = \left[V_0^{\downarrow P(k)} \otimes \left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^i \right) \right]^{\downarrow\{j\}}.$$

By a repeated application of Proposition 4 we get

$$\lambda_j^k = V_0^{\{j\}} \otimes \left(\bigotimes_{i \in P(k),\, i \neq j} \pi_k^{i\,\downarrow\emptyset} \right) = V_0^{\{j\}}$$

That is, $\lambda_j^k = V_0^{\downarrow\{j\}} \in V_{\{j\}}$, hence

$$\lambda_j = \left(\bigotimes_{k \in C(j)} \lambda_j^k \right) = V_0^{\downarrow\{j\}} \in V_{\{j\}}.$$

Q.E.D. ■

Under these conditions, if $I = \emptyset$, we can find π_j for every $j \in \{1, \ldots, n\}$ with the following algorithm, where it is assumed that if $j > i$, then X_j is not a parent of X_i (that is, the π_j value of a node will be calculated when corresponding π_j^k, $j \in P(k)$ are known). The steps are as follows:

- For every $j = 1, \ldots, n$ calculate

$\pi_j = [V_j \otimes (\otimes_{k \in P(j)} \pi_j^k)]^{\downarrow \{j\}}$
$\lambda_j = V_o^{\downarrow j}$
$PS_j = \pi_j \otimes \lambda_j$
$\forall k \in C(j),\ \pi_k^j = \pi_k$
$\forall i \in P(j),\ \lambda_i^j = V_o^{\downarrow \{i\}}$

Assume now that we want to calculate PS_j with a set $I \neq \varnothing$. We shall give, like Pearl, [7], an algorithm that calculates these values $(PS_j, \pi_j, \lambda_j, \lambda_j^k, \pi_j^k)$ for a set $I' = I \cup \{O_l\}$, assuming that we know these values for the set of observations I; that is, the algorithm updates the values of valulations in the light of a new observation.

For $k \in P(j)$, the π_j^k values are said to be messages from the parents of X_j to it, or incoming messages from its parents. Analogously, for $k \in C(j)$, the λ_j^k values are the messages that this node receives from its children.

On introducing a new observation, O_l, all incoming messages to a node do not change, unless there is an undirected path between this node and X_l. In this case, only the message arriving from this path changes; that is, λ_j^k or π_j^k change only when there is an undirected path between X_j and X_l via X_k.

As the graphs we are working with do not have loops, there is, at most, only one undirected path from each node to X_l. All incoming valuations for X_j with the set of observations I' are, therefore, the same as the valuations for set of observations I, except perhaps for the message coming from one of its children or parents. The outgoing message from X_j to this node does not change, but all the other outgoing messages are different (Figure 1).

For node X_l, the variable for which we have introduced the new observation, the situation is as follows: all incoming messages are the same and all outgoing messages are new (Figure 1). In the light of these considerations we can design an algorithm to perform the updating. The algorithm is based on the fact that when we arrive at a node, all the incoming messages are calculated, and we then calculate π_j, λ_j, PS_j, and the outgoing messages. Let J be the set of pairs (j_1, j_2) where j_1 is a node to be updated and j_2 the incoming node. Then the algorithm is as follows:

- $J = \{(l, -1)\}$
- While $J \neq \varnothing$

Choose $(j_1, j_2) \in J, J \leftarrow J - \{(j_1, j_2)\}$
calculate

* $\pi_{j1} \leftarrow [V_{j1} \otimes (\otimes_{i \in P(j_1)} \pi_{j1}^k)]^{\downarrow \{j_1\}}$
* $\lambda_{j1} \leftarrow (\otimes_{K \in C(j_1)} \lambda_{j1}^k) \otimes O'_{j1}$
* $PS_{j1} \leftarrow \pi_{j1} \otimes \lambda_{j1}$

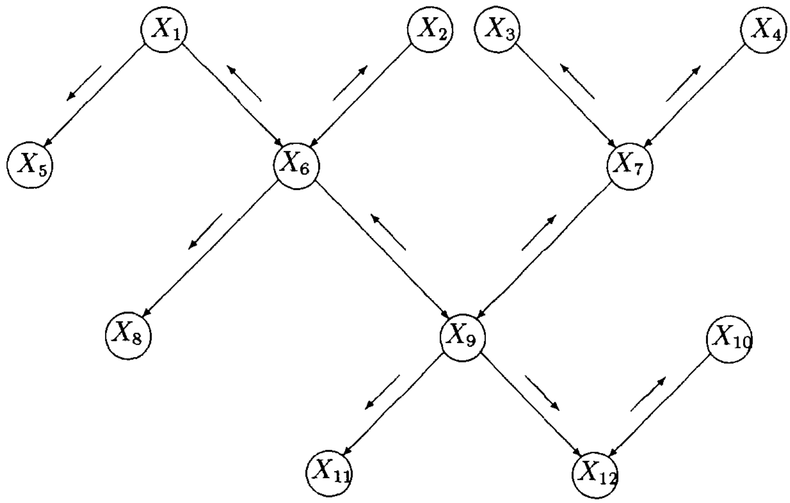

Figure 1. Changing messages with a new observation O_9.

For every $k \in C(j_1)$, $k \neq j_2$ calculate
- $\pi_k^{j1} \leftarrow [\pi_{j1} \otimes O'_{j1} \otimes (\otimes_{i \in C(j_1)_{i \neq k}} \lambda_{j1}^i)]$
- $J \leftarrow J \cup \{(k, j_1)\}$

For every $k \in P(j_1)$, $k \neq j_1$ calculate
- $\lambda_k^{j1} \leftarrow [\lambda_{j1} \otimes O'_{j1} \otimes V_{j1} \otimes (\otimes_{i \in P(j_1)_{i \neq k}} \pi_{j1}^i)]^{\downarrow \{k\}}$
- $J \leftarrow J \cup \{(k, j_1)\}$

It is immediately clear that this algorithm updates all the messages, and the values of PS_j corresponding to I'.

5. PROPAGATION ALGORITHMS AND CONDITIONAL INDEPENDENCE

In this section we prove that if all the axioms hold, then the propagation algorithms are coherent with the initial dependence structure associated with a graph. To prove this coherence, we need Axiom 6. In a theory of uncertainty representation in which this axiom does not hold, we could resort to use propagation formulas, but then we would have to admit that we are violating the coherence with the dependence structure associated with the graph.

PROPOSITION 12 Let (T, E) be a DAG without loops associated with n-dimensional variable $(X_1, \ldots, X_n)$, (H, h) a system of information defined on it, and $I, J, K \subseteq \{1, \ldots, n\}$ in such a way that $D(I, J, K) = 0$. Under these conditions, if we know the values of variables X_J, that is, if we have the observations $\{O_j\}_{j \in J}$ and the contradiction is never obtained, then the introduction of a new observation of a variable in $\{X_i\}_{i \in I}$ does not change the 'a posteriori' valuation in any of the variables in $\{X_k\}_{k \in K}$.

Proof Assume that we introduce a new observation, O_{i_0}, where $i_0 \in I$, and assume a node X_k, where $k \in K$. As the graph has no loops there is only one undirected path going from X_{i_0} to X_k. Along this undirected path travel the messages with the influence of O_{i_0} on X_k. If at a given moment, one of these messages does not change after introducing observation O_{i_0}, then this observation does not have any effect on X_k.

As we assume that $D(I, J, K) = 0$, then by the d-separation criterion (see Definition 13) we have that for this path one of the two following conditions holds

- There exists a node with converging arrows in the path that does not belong to $\{X_j\}_{j \in J}$ and its descendents do not belong to $\{X_j\}_{j \in J}$.
- There exists a node without converging arrows in the path that belongs to $\{X_j\}_{j \in J}$.

In the first case, let this node be X_l. None of its descendents is in $\{X_j\}_{j \in J}$. By applying a proof similar to that in Proposition 11, it can be shown that before and after introducing observation O_{i_0}, all messages that this node sends to its parents are the neutral valuation. As the chain of messages carrying the effect of O_{i_0} to X_k goes through X_l from one of its parents to a different parent, the outgoing message does not change. It is the neutral element, with no effect on X_k.

In the second case, we have a node, X_j, $j \in J$, without converging arrows in the path from X_{i_0} to X_k. As j is in the set of observations, we have an observauon O_j for this node, which is an absorbent valuation. The messages this node sends to its children, X_m, $m \in C(j)$ are

$$\pi_m^j = \left[\pi_j \otimes O_j \otimes \left(\bigotimes_{i \in C(j)_{i \neq m}} \lambda_j^i\right)\right]$$

As O_j is absorbent and the contradiction is never obtained,

$$\pi_m^j = O_j.$$

Thus, the messages from X_j to its children never change, and if the path from X_{i_0} to X_k passes through X_j from a parent to a child, or from a child to a different child, then it does not change in X_k and has no effect on X_k.

The messages that node X_j sends to its children, X_m, $m \in C(j)$ are

$$\lambda_m^j = \left[\lambda_j \otimes O_j \otimes V_j \otimes \left(\bigotimes_{i \in P(j)_{i \neq m}} \pi_j^i\right)\right]^{\downarrow \{m\}}.$$

As O_j is absorbent and the contradiction is never obtained, we have

$$\lambda_m^j = \left[O_j \otimes V_j \otimes \left(\bigotimes_{i \in P(j)_{i \neq m}} \pi_j^i\right)\right]^{\downarrow \{m\}}.$$

In this expression λ_m^j does not depend on the incoming messages from its children, it only depends on O_j, V_j and the incoming messages from its parent π_j^i. Therefore, if the path from X_{i_0} to X_k passes through X_j from a child to a parent it does not change and has no effect.

At this point we run out of alternatives as, in this second case, the path cannot go through X_j from a parent to another, because the path has no converging arrows on X_j. In all cases, the messages going from X_{i_0} to X_k are cut off on a given moment and the effect of O_{i_0} never reaches to X_k. ∎

6. CONCLUSIONS

In this paper we have shown that the propagation model developed by Pearl can be adapted to other formalisms for representing information. The initial idea of extending the procedures of probabilistic propagation to other systems used to represent uncertainty was introduced by Shafer and Shenoy, [1, 2], whose work centered mainly on calculus but did not consider some important aspects related with the representation of knowledge. Shenoy, [8], also studies the concept of independence in the valuation formalism from the point of view of factors of a joint or global valuation. Here we give a general framework to represent and use our knowledge about a problem integrating independence relationships and abstract valuations. The point of view is different: we study rules to build more complex valuations from elemental ones, using independence relationships. Conditional valuations have special relevance in our work. Two different types are considered: 'a priori' conditional valuations, given by the user as general knowledge about a population, and 'a posteriori' conditional valuations, calculated after some observations for a particular case.

The graphical structures have been directed acyclic graphs. These graphs are directly related with the concepts of conditional independence and dependence, which are fundamental for our problem.

The loops in the resulting algorithms pose a serious problem. Of the methods proposed by Pearl, [7], to surmount this difficulty in the probabilistic case, only the clustering method can be directly generalized. It is difficult to see how the conditioning method could be expressed in a general manner. Montecarlo algorithms can be applied only to particular theories of representation.

A particular important case is that posed when the combination of valuations is idempotent (see Dubois and Prade, [13]), as occurs with valuations based on the Theory of Possibility, with the minimum rule for combination. In this case loops are more amenable to resolution: we can combine the same information several times.

With regard to the verification of the axioms, all representations models we have studied verify Axioms 1–5; however, Axiom 6 is not verified in some cases, as Possibility Theory. This can be explained by recalling that Possibility Theory can represent different degrees of partial inconsistency, [14]. Therefore, apart from the contradictory valuation, there are other levels of contradiction. This has important consequences for the propagation algorithms, which deserve particular attention.

ACKNOWLEDGMENTS

The authors are very grateful to Luis M. de Campos and to the anonymous referees for their valuable and useful comments and suggestions.

This work was supported by the Commission of the European Communities under ESPRIT BRA 3085: DRUMS.

References

1. Shafer, G., and Shenoy, P. P., Local computation in hypertrees, Working Paper #201. School of Business, University of Kansas, Lawrence, 1988.

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

3. Saffiotti, A., and Umkehrer, E., PULCINELLA. A general tool for propagating uncertainty in valuation networks, in *Uncertainty in A. I. Proceedings of the 7th Conference*, (B. D. D'Ambrosio, Ph. Smets, and P. P. Bonissone, Eds.), Morgan and Kaufmann, San Mateo, 323–331, 1991.

4. Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Statist.* 38, 325–339, 1967.

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

6. Smets, Ph., Belief functions, in *Non-Standard Logics for Automated Reasoning*, (Ph. Smets, E. H. Mandani, D. Dubois, and H. Prade, Eds.), Academic Press, London, pp. 253–286, 1988.

7. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan and Kaufmann, San Mateo, 1988.

8. Shenoy, P. P., Conditional independence in uncertainty theories, in *Uncertainty in A.I. Proceedings of the 8th Conference*, (D. Dubois, M. P. Wellman, B. D. D'Ambrosio, and Ph. Smets, Eds., Morgan and Kaufmann, San Mateo, 284–291, 1992.

9. Shenoy, P. P., Consistency in valuation-based systems, Working Paper #216. School of Business, University of Kansas, Lawrence, 1990.

10. Geiger, D., and Pearl, J., Logical and algorithmic properties of conditional independence, Tech. Report R-97, Cognitive Systems Laboratory, University of California, Los Angeles, 1988.

11. Verma, G., and Pearl, J., Identifying independence in Bayesian Networks. *Networks* 20, 507–534, 1990.

12. Shenoy, P. P, Consistency in valuation-based systems, Working Paper #216. School of Business, University of Kansas, Lawrence, 1990.

13. Dubois, D., and Prade, H., Inference in possibilistic hypergraphs, in *Uncertainty in Knowledge Bases*, (B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, Eds.), Springer-Verlag, Berlin, 250–259, 1991.

14. Lang, J., Dubois, D., and Prade, H., A logic of graded possibility and certainty coping with partial inconsistency, in *Uncertainty in A.I. Proceedings of the 7th Conference*, (B. D. D'Ambrosio, Ph. Smets, and P. P. Bonissone, Eds.), Morgan and Kaufmann, San Mateo, 188–196, 1991.