# Removing Partial Inconsistency in Valuation-Based Systems*

Luis M. de Campos and Serafín Moral†‡
*Departamento de Ciencias de la Computación e I.A., Universidad de Granada, 18071 Granada, Spain*

This paper presents an abstract definition of partial inconsistency and one operator used to remove it: normalization. When there is partial inconsistency in the combination of two pieces of information, this partial inconsistency is propagated to all the information in the system thereby degrading it. To avoid this effect, it is necessary to apply normalization. Four different formalisms are studied as particular cases of the axiomatic framework presented in this paper: probability theory, infinitesimal probabilities, possibility theory, and symbolic evidence theory. It is shown how, in one of these theories (probability), normalization is not an important problem: a property which is verified in this case gives rise to the equivalence of all the different normalization strategies. Things are very different for the other three theories: there are a number of different normalization procedures. The main objective of this paper will be to determine conditions based on general principles indicating how and when the normalization operator should be applied. 

## 1. INTRODUCTION

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

Cano et al.[3] have proposed a modification of this axiomatic framework

*This work has been supported by the Commission of the European Communities under ESPRIT III BRA 6156: DRUMS 2.

†To whom correspondence should be addressed.
‡E-mail: smc@decsai.ugr.es.


 CCC 0884-8173/97/090629-25

adding three new axioms to generalize Pearl's propagation algorithm[4] in directed acyclic graphs (DAG's). The last axiom was the strangest at first sight. However, it turned out to be very important. It states that in the empty set all the valuations must be equal to the neutral element or to the contradiction. In general, propagation algorithms could be applied (with minor modifications) to cases in which this axiom is not verified. However, when this axiom is not verified there is a problem with the propagation of partial inconsistency in a process in which the rest of the information is degraded.

The objective of this paper is to make an abstract study of partial inconsistency and of the strategies to remove it. In Section 2, we shall start by considering general valuation systems in which the last axiom proposed by Cano et al.[3] is not necessarily verified. In Section 3, we shall show that this axiomatic system is verified by probability theory, the theory of infinitesimal probabilities, possibility theory, and the so called symbolic evidence theory.

Section 4 will be devoted to defining partial inconsistency and a set of Axioms N1–N4 to be verified by normalization operators. The normalization of a valuation is intended to be the same valuation in which partial inconsistency is removed. This axiomatic is applied to the particular cases introduced in Section 3.

Section 5 explains what the role of normalization is about and introduces the main problem associated with it: if we have several valuations, when should normalization be applied? The fact is that there is no single alternative, and depending on the one we choose, we shall obtain a different result.

Section 6 considers strategies to determine the application points of normalization operators. First we determine a simple case, in which probability is included. This case is precisely that in which the last axiom in Cano et al.[3] can be verified. The other theories studied in this paper, infinitesimal probabilities, possibility theory, and symbolic evidence theory, do not correspond to this simple case. The rest of this section is devoted to generating procedures for them. The idea will be to apply general principles for determining which group of valuations should be combined and normalized, with the objective of reducing the number of final alternatives. One of these principles will be that the selection of the valuations will be based on the frame in which they are defined.

Section 7 shows that the application of normalization is not an obstacle for the application of propagation algorithms. It is shown that the basic property introduced by Shafer and Shenoy[1,2] allowing the application of propagation algorithms is also verified in our formalism in which the partial inconsistency is removed.

## 2. VALUATION BASED SYSTEMS

Assume that we have an $n$-dimensional variable, $(X_1, \ldots, X_n)$, each dimension, $X_i$, taking values on a finite set $U_i$. First of all, we shall consider the notation we are going to follow:

- If $I \subseteq \{1, \ldots, n\}$, 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 $\prod_{i \in I} \mathrm{U}_I$, that is the set in which $X_I$ takes its values. Sometimes, for simplicity in the language, we shall identify set $I$ with variable $X_I$, and we shall talk about variable $I$.

- If $u \in U_i$, then we shall denote by $u_i$ the $i$th 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

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

In these conditions, a valuation is a primitive concept meaning the mathematical representation for a piece of information in a given uncertainty theory. We shall assume that for each $I \subseteq \{1, \ldots, n\}$ there is a set $\mathcal{V}_I$ of valuations defined on the cartesian product, $U_I$. If $V \in \mathcal{V}_I$ we shall say that $V$ is defined on $U_I$ or that $U_I$ is the frame of $V$. We shall also say that $V$ is defined on $I$.

$\mathcal{V}$ will be the set of all valuations $\mathcal{V} = \bigcup_{I \subseteq \{1, \ldots, n\}} \mathcal{V}_I$

Two basic operations are necessary (see Zadeh[5]; Shenoy and Shafer[1,2]):

- *Marginalization:* If $J \subseteq I$ and $V_1 \in \mathcal{V}_I$ then the marginalization of $V_1$ to $J$ is a valuation $V_1^{\downarrow J}$ in $\mathcal{V}_J$.
- *Combination:* If $V_1 \in \mathcal{V}_I$ and $V_2 \in \mathcal{V}_J$, then their combination is a valuation $V_1 \otimes V_2$ in $\mathcal{V}_{I \cup J}$

We shall assume that the valuations verify the following axioms (Axioms 1–3 are from Shenoy and Shafer[2] and Axioms 4–5 from Cano, Delgado, and Moral[3]).

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

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

**Axiom 3:** If $V_1 \in \mathcal{V}_I$, $V_2 \in \mathcal{V}_J$, then $(V_1 \otimes V_2)^{\downarrow I} = V_1 \otimes V_2^{\downarrow (J \cap I)}$.

**Axiom 4:** *Neutral Element.* For each $J \subseteq \{1, \ldots, n\}$, there exists one and only one valuation $V_0^J$ defined on $U_J$ such that $\forall\ V \in \mathcal{V}^I$, with $J \subseteq I$, $V_0^J \otimes V = V$.

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

The following axiom was considered in Ref. 3, but in our present setting it will not be assumed.

**Axiom 6:** $\forall\ V \in \mathcal{V}_\emptyset$, if $V \neq V_c^{\downarrow \emptyset}$, then $V = V_0^\emptyset$.

According to this axiom, in the frame corresponding to the empty set of variables, $U_\emptyset$, all valuations are the contradiction or the neutral element. In this sense, there are only two possible degrees of consistency: one corresponding to the contradiction (inconsistency) and the other corresponding to the neutral valuation (consistency), without intermediate degrees of consistency. That is not the case we are trying to represent. We want to study formalisms in which valuations have partial inconsistency, as is the case of the Theory of Possibility. To cope with partial inconsistency we are going to assume a new operation, normalization, which is applied to the valuations different from the contradiction, i.e., to the set of valuations,

$$\mathscr{V}' = \{V \in \mathscr{V} : V^{\downarrow\emptyset} \neq V_c^{\downarrow\emptyset}\} \tag{1}$$

For the cases in which it is possible, as in the Theory of Probability, Axiom 6 will be recovered later on.

## 3. PARTICULAR CASES

### 3.1. Probability Theory

In Probability Theory a valuation is the representation of a probabilistic piece of information about some of the variables, $X_I$, $I \subseteq \{1, \ldots, n\}$. 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 $X_2$,

$p(u_{31} \mid u_{21}) = 0.9 \quad p(u_{32} \mid u_{21}) = 0.1$

$p(u_{31} \mid u_{22}) = 0.6 \quad p(u_{32} \mid u_{22}) = 0.4$

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

$$p: U_I \to \mathbb{R}_0^+$$

where $\mathbb{R}_0^+$ denotes the non-negative real numbers.

If $p_1$ is a valuation on $U_I$ and $p_2$ is a valuation on $U_J$, then the combination of $p_1$ and $p_2$ denoted as $p_1 \otimes p_2$ is a valuation defined on $U_{I \cup J}$ by:

$$p_1 \otimes p_2(u) = p_1(u^{\downarrow I}) \cdot p_2(u^{\downarrow J}) \tag{2}$$

If $p$ is a valuation on $U_I$ and $J \subseteq I$, the marginalization $p^{\downarrow J}$ is defined by:

$$p^{\downarrow J}(v) = \sum \{p(u) : u^{\downarrow J} = v\} \tag{3}$$

The neutral valuation on $U_J$ is the mapping $p_0$ always taking the value 1.

The contradiction is the valuation taking the value 0 for all the elements of $U_1 \times \cdots \times U_n$.

### 3.2. Infinitesimal Probabilities

Here we describe a very simple version of infinitesimal probabilities. More sophisticated theories are to be found in the literature.[6,7]

It is considered that apart from the numerical probabilities, there is a value, $\varepsilon$, representing a very small probability. So a valuation is a mapping

$$p: U_I \to \mathbb{R}_0^+ \cup \{\varepsilon\}$$

The operations are defined in the same way as in the probabilistic case, considering that $\varepsilon$ is a number between 0 and all the positive real numbers, where:

$-\varepsilon + x = x, \qquad \forall x \in \mathbb{R}_0^+ \cup \{\varepsilon\}, x \neq 0$

$-\varepsilon + 0 = \varepsilon$

$-\varepsilon . x = \varepsilon, \qquad \forall x \in \mathbb{R}_0^+ \cup \{\varepsilon\}, x \neq 0$

$-\varepsilon . 0 = 0$

### 3.3. Possibility Theory

In possibility theory,[8,9] a valuation in $U_I$ is a possibility distribution on $U_I$, i.e., a mapping

$$\pi: U_I \to [0, 1] \tag{4}$$

If $\pi_1$ is a valuation on $U_I$ and $\pi_2$ is a valuation on $U_J$, then the combination of $\pi_1$ and $\pi_2$ denoted as $\pi_1 \otimes \pi_2$ is a valuation defined on $U_{I \cup J}$ by using the minimum rule:

$$\pi_1 \otimes \pi_2(u) = Min\,\{\pi_1(u^{\downarrow I}), \pi_2(u^{\downarrow J})\} \tag{5}$$

If $\pi$ is a valuation on $U_I$ and $J \subseteq I$, the marginalization $\pi^{\downarrow J}$ is defined by using the maximum operator in the following way,

$$\pi^{\downarrow J}(v) = Max\,\{\pi(u): u^{\downarrow J} = v\} \tag{6}$$

The neutral valuation is the possibility, $\pi_0$, defined by $\pi_0(u) = 1, \forall u \in U_J$. The contradiction is the possibility, $\pi_c$, defined by $\pi_c(u) = 0, \forall u \in U_{\{1, \ldots, n\}}$.

### 3.4. Symbolic Evidence Theory

The Symbolic Evidence Theory has been developed by Kohlas,[10] starting out from assumption-based propositional knowledge. If $P$ is a set of elementary propositional symbols, then $\mathcal{L}_P$ will denote the set of all well-formed propositional formulae using symbols from $P$.

Let $A$ be a finite set of symbols representing assumptions, a set of propositional symbols $P$ and a set of clauses $\Sigma = \{\xi_1, \ldots, \xi_m\}$ over $A \cup P$, representing the available knowledge and facts.

If $h$ is an element of $\mathcal{L}_P$, we are interested in the assumptions under which $h$ can be deduced from $\Sigma$. In particular, we say that a conjunction $a \in \mathcal{L}_A$ is a support set of $h$ if and only if

$$a, \Sigma \models h$$

A conjunction $a \in \mathcal{L}_A$ is said to be a contradiction set relating $\Sigma$ if and only if

$$a, \Sigma \models \bot$$

where $\bot$ stands for the contradiction.

Support sets $a$ for $h$ such that no subset is also a support set are called minimal support sets for $h$, similarly, contradiction sets $a$, such that no subset is still a contradiction set, are called minimal contradictions.

The support for a formula $h \in \mathcal{L}_P$ relating $\Sigma$ is given by the formula:

$$sp(h, \Sigma) = \bigvee \{a : a \text{ is a minimal support set of } h \text{ relating } \Sigma\}$$

The contradiction relating $\Sigma$ is

$$sp(\bot, \Sigma) = \bigvee \{a : a \text{ is a minimal contradiction relating } \Sigma\}$$

If $Q \subseteq P$, and $\mathcal{L}_Q$ is the language of propositional formulae on $Q$, then $\mathcal{L}_Q/\equiv$ will denote the set of equivalence classes of $\mathcal{L}_Q$ under the relation $\equiv$. If $h \in \mathcal{L}_Q$, then $[h]$ will stand for the set of all formulae equivalent to $h$.

Here, to talk in terms of valuations, we identify the set of variables with the set of propositional symbols, $P$, which will be assumed to be finite.

A valuation on $Q \subseteq P$ will be a mapping

$$m : \mathcal{L}_Q/\equiv \; \rightarrow \mathcal{L}_A$$

verifying

(1) If $[h_1]$, $[h_2] \in \mathcal{L}_Q/\equiv$ and $h_1 \not\equiv h_2$, then
$m(h_1) \wedge m(h_2) \equiv \bot$;

(2) $\Sigma \{m(h) : [h] \in \mathcal{L}_Q/\equiv\} \equiv \top$,

where $\top$ stands for the tautology and $\Sigma$ stands for the disjoint union.

For simplicity's sake, we write $m(h)$ instead of $m([h])$.

A valuation, $m$, can be defined from a set of clauses $\sigma$ on $A \cup P$ in the following way:

$$m(h) = sp(h, \Sigma) \wedge \left( \neg \bigvee \{sp(h', \Sigma) : h' \in \mathcal{L}_Q, h' \models h, h' \not\equiv h\} \right)$$

$m(h)$ is called the basic argument of $h$ relating $\Sigma$.[10]

If $m$ is defined from the knowledge base $\Sigma$ and $h \in \mathcal{L}_Q$, its support may be obtained in the following way:

$$sp(h, \Sigma) = \Sigma \{m(h') : [h'] \in \mathcal{L}_Q/\equiv, h' \models h, h' \not\equiv \bot\}$$

If $Q' \subseteq Q$ and $h \in \mathcal{L}_Q$ is expressed in disjunctive normal form, the projection of $h$ to $\mathcal{L}_{Q'}$ is the formula obtained by eliminating all literals in $h$ belonging to $Q - Q'$. This projection will be denoted as $h^{\downarrow Q'}$.

If $m$ is a valuation in $Q$ and $Q' \subseteq Q$, the marginalization of $m$ to $Q'$ will be the valuation on $Q'$ defined by:

$$m^{\downarrow Q'}(h') = \sum \{m(h) : [h] \in \mathcal{L}_Q/\equiv, h^{\downarrow Q'} \equiv h'\}$$

It may be verified that if $m$ is a valuation on $\mathcal{L}_Q$ obtained from a knowledge base $\Sigma$ on $A \cup P$, then $m^{\downarrow Q'}$ is the valuation induced by the same knowledge base on $\mathcal{L}_{Q'}$.[10]

Assume now that $Q_1, Q_2 \subseteq P$ and that $m_1, m_2$ are valuations on $Q_1$ and $Q_2$, respectively. The combination of $m_1$ and $m_2$ will be a valuation on $Q_1 \cup Q_2$ given by:

$$m_1 \otimes m_2(h) = \sum \{m_1(h_1) \wedge m_2(h_2) : [h_i] \in \mathcal{L}_{Q_i}/\equiv, h_1 \wedge h_2 \equiv h\}$$

Kohlas[10] justifies this rule verifying that if each $m_i$ comes from the knowledge base $\Sigma_i$ defined on $Q_i \cup A$, then $m_1 \otimes m_2$ is the valuation on $Q_1 \cup Q_2$ induced by the union of the knowledge bases $\Sigma_1 \cup \Sigma_2$.

The neutral valuation is the valuation $m_0$ on $P$ assigning the tautology on $\mathcal{L}_A$ to the tautology on $\mathcal{L}_P$ and the contradiction of $\mathcal{L}_A$ to the rest of the formulae on $\mathcal{L}_p$. That is, only the tautology on $\mathcal{L}_P$ is supported and it is fully supported: We only know that we do not know anything. The contradiction is the valuation $m_c$ on $P$ assigning the tautology on $\mathcal{L}_A$ to the contradiction on $\mathcal{L}_P$: only the contradiction has a basic argument.

## 4. AN AXIOMATIC VIEW OF PARTIAL INCONSISTENCY

In this framework, $\mathcal{V}_\emptyset$ is called the set of consistency values and denoted as $\mathcal{C}$. The application

$$CG: \mathcal{V} \to \mathcal{C} \tag{7}$$

given by $CG(V) = V^{\downarrow \emptyset}$ is the consistency mapping. $CG$ applies each valuation on its consistency degree.

Though we have to add more properties to the axiomatic system, below we give an example to illustrate these somewhat abstract ideas.

*Example* 1.

(1) *Probability Theory:* $U_\emptyset$ is the cartesian product of no set. This cartesian product has only one element: $U_\emptyset = \{e\}$ ($e$ would be the element ( )). In the case of probability theory a valuation, $p_\emptyset$, defined on $U_\emptyset$ is a mapping defined on this set $\{e\}$ which has only one element. To state it, we only have to specify a number $p_\emptyset(e)$. That is, we only need to provide a nonnegative real number. Thus, a valuation on $U_\emptyset$ can be identified with a number. If we have a probability, $p$, defined on $U_I$, the marginalization of this probability on the empty set of variables is equal to:

$$CG(p) = p^{\downarrow\varnothing}(e) = \sum\{p(u) : u^{\downarrow\varnothing} = e\} = \sum\{p(u) : u \in U_I\} \tag{8}$$

The last equality comes from the fact that the marginalization of a point $u \in U_I$ to the empty set, $u^{\downarrow\varnothing}$, is always equal to ( ) = $e$.

(2) *Infinitesimal Probability:* The definitions are the same as in probability theory with the only difference that now we may have a new value $\varepsilon$ and that we have to operate with this value according to the rules above.

(3) *Possibility Theory:* As in the case of probability, a possibility on $U_\varnothing$ is given by a single value. Therefore the set of consistency degrees is the interval [0, 1]. The degree of consistency of a possibility, $\pi$, defined on $U_I$ is given by

$$CG(\pi) = \pi^{\downarrow\varnothing}(e) = Max\{\pi(u) : u^{\downarrow\varnothing} = e\} = Max\{\pi(u) : u \in U_I\} \tag{9}$$

(4) *Symbolic Evidence:* For the empty set, $\mathscr{L}_\varnothing$ has only two formulae, the contradiction $\perp$ and the tautology $\top$. A valuation on this set, $m_\varnothing$, is a mapping from this set on $\mathscr{L}_A$, in such a way that $m_\varnothing(\perp) \equiv \neg m_\varnothing(\top)$. So a valuation is given by the degree of support of $\top$. The other value, $m_\varnothing(\perp)$ is its negation. As a proposition on $\mathscr{L}_A$ defines a degree of consistency and reciprocally, we may identify that the set of degrees of consistency is $\mathscr{L}_A$. When we have an arbitrary valuation, $m$, on $\mathscr{L}_Q$, its degree of consistency is calculated by

$$\begin{aligned} CG(m) = m^{\downarrow\varnothing}(\top) &= \sum\{m(h) : [h] \in \mathscr{L}_Q/\equiv, h^{\downarrow Q'} \equiv \top\} \\ &= \sum\{m(h) : h \not\equiv \perp\} = \neg m(\perp) \end{aligned} \tag{10}$$

The consistency of $m$ is the negation of the support of the contradiction.

The normalization function is a mapping $N : \mathscr{V}' \to \mathscr{V}'$, verifying the following properties,

**Axiom N1:** $\forall V \in \mathscr{V}'$, $V \in \mathscr{V}_I$, if and only if $N(V) \in \mathscr{V}_I$.

**Axiom N2:** $\forall V \in \mathscr{V}'$, $N(N(V)) = N(V)$.

**Axiom N3:** $\forall V \in \mathscr{V}'$, $V^{\downarrow\varnothing} = V_0$, if and only if $N(V) = V$.
Where $V_0$ is the neutral valuation on $V_\varnothing$.

**Axiom N4:** $\forall V \in \mathscr{V}'$, if $V \in \mathscr{V}_I$ and $J \subseteq I$, $N(V^{\downarrow J}) = N(V)^{\downarrow J}$.

We say that a valuation has partial inconsistency when $N(V) \neq V$ or equivalently, when $V^{\downarrow\varnothing} \neq V_0$. Essentially, a valuation and its normalization represent the same information. The normalized version has removed the partial inconsistency from the valuation.

According to Axiom N1 the normalization of a valuation is defined on the same set of variables as the original valuation. Axiom N2 specifies that the normalization of a normalized valuation is the same valuation. Axiom N3 states that the neutral valuation in the empty set and all the valuations with the same degree of consistency are normalized valuations: their normalization produces the same valuation. Axiom N4 states that the order in which a normalization and a marginalization are carried out is irrelevant.

We have not defined the normalization of the total contradiction. We might have considered that $N(V_c) = V_c$. But then Axiom N3 is not verified and we should modify the axiom making our system more complicated. In this paper the problem of total contradiction is not considered and when it is obtained, it should be solved by another procedure. Our assumption will be that it is never obtained. All the results in this paper are stated under the condition that all the combinations are different from the contradiction. If, in a particular case, the total contradiction is obtained then we should not apply these methods. We should consider other ways to remove inconsistency or simply report that we have inconsistency in our system.

Shafer[11] has given a different concept for normalization. It says that $N'(V)$ is a normalization of $V$ if

$$V^{\downarrow\emptyset} \otimes N'(V) = V \tag{11}$$

The idea behind this definition is quite different from the concept we have given in this paper. According to Shafer's definition $N'(V)$ may have inconsistency, because $(N'(V))^{\downarrow\emptyset}$ can be different from the neutral valuation. That is not possible in our case. For idempotent valuations (possibility theory and symbolic evidence) we could make $N'(V) = V$ verifying Shafer's definition and keeping the valuation as inconsistent as before applying normalization.

*Example* 2.

(1) *Probability Theory:* If $p$ is a valuation on $U_I$, other than the contradiction, its normalization is obtained by the following expression:

$$N(p)(u) = p(u)/CG(p) \tag{12}$$

(2) *Infinitesimal Probability:* If $p$ is a valuation on $U_I$, and $CG(p) \neq 0$, $\varepsilon$ its normalization is obtained by the same expression (12) as in the case of probabilities, taking into account that $\varepsilon/x = \varepsilon$, $\forall x > 0$:

When $CG(p) = 0$, then the valuation is the contradiction and the normalization is not defined. We should define the normalization when $CG(p) = \varepsilon$. A reasonable definition verifying the axioms is:

$$N(p)(u) = \begin{cases} 1/k & \text{if } p(u) = \varepsilon \\ 0 & \text{if } p(u) = 0 \end{cases} \tag{13}$$

where $k$ is the number of elements of $U_I$ for which $p$ takes the value $\varepsilon$.

(3) *Possibility Theory:* It is possible to find several normalization functions in Possibility Theory. It is possible to use a normalization analogous to the probabilistic one. Here we give the following one:

$$N(\pi)(u) = \begin{cases} \pi(u) & \text{if } \pi(U) < CG(\pi) \\ 1 & \text{otherwise} \end{cases} \tag{14}$$

(4) *Symbolic Evidence:* In this case the consistency degree of a valuation is $\neg m(\perp)$. A valuation is normalized when $m(\perp)$ is the contradiction: nothing supports the contradiction.

To normalize a valuation is to remove the support from $\perp$ and pass it onto another formula. The most neutral selection is to assign this support to the tautology. So we can define:

$$N(m)(h) = \begin{cases} m(h) & \text{if } h \not\equiv \perp, h \not\equiv \top \\ \perp & \text{if } h \equiv \perp \\ m(\perp) + m(\top) & \text{if } h \equiv \top \end{cases} \tag{15}$$

An immediate property of normalization operators is given by the following proposition.

PROPOSITION 1. *If $V_1 \in \mathcal{V}_I$, $V_2 \in \mathcal{V}_J$, where $I \cap J = \emptyset$, then*

$$N(N(V_1) \otimes N(V_2)) = N(V_1) \otimes N(V_2)$$

*Proof.* First let us calculate the value of $(N(V_1) \otimes N(V_2))^{\downarrow\emptyset}$. Following Axiom 2, we find,

$$(N(V_1) \otimes N(V_2))^{\downarrow\emptyset} = ((N(V_1) \otimes N(V_2))^{\downarrow I})^{\downarrow\emptyset} \tag{16}$$

By Axiom $N1$, $N(V_1)$ is defined on $\mathcal{V}_I$, then applying Axiom 3, we have,

$$(N(V_1) \otimes N(V_2))^{\downarrow\emptyset} = (N(V_1) \otimes N(V_2)^{\downarrow I \cap J})^{\downarrow\emptyset} = (N(V_1) \otimes N(V_2)^{\downarrow\emptyset})^{\downarrow\emptyset} \tag{17}$$

By applying Axiom 3 to the resulting expressions, we find

$$(N(V_1) \otimes N(V_2))^{\downarrow\emptyset} = N(V_1)^{\downarrow\emptyset} \otimes N(V_2)^{\downarrow\emptyset} \tag{18}$$

Now, following Axioms N2 and N3, we have that

$$N(V_i)^{\downarrow\emptyset} = V_0^{\downarrow\emptyset} \tag{19}$$

Introducing this equality in Eq. 18, we obtain

$$(N(V_1) \otimes N(V_2))^{\downarrow\emptyset} = V_0^{\downarrow\emptyset} \tag{20}$$

Now, by applying Axiom N3, we achieve the desired result,

$$N(N(V_1) \otimes N(V_2)) = N(V_1) \otimes N(V_2) \tag{21}$$

■

The main definitions relating to the calculus with valuations are given below. These are modifications of the ones given in Ref. 3 adapted to this axiomatic framework.

DEFINITION 1. *A valuation $V \in \mathcal{V}_I$ is said to be absorbent if and only if it is normalized* $(N(V) = V)$ *and* $(\forall V' \in \mathcal{V}_I)((N(V \otimes V') = V)$ *or* $(V \otimes V' = V_c))$.

If a valuation from $\mathcal{V}_I$ represents a piece of information about the values of variables $X_I$, then an absorbent valuation represents prefect knowledge about these values: it cannot be consistently refined by combination with another valuation: We can only obtain the same information with less consistency or the contradiction.

*Example* 3. In Probability Theory, absorbent valuations are probability distributions assigning a value of 1 to one element and 0 to the other elements. It represents an observation of this element.

The cases of Infinitesimal Probabilities and Possibility Theory are exactly the same.

In the Symbolic Theory of Evidence an absorbent valuation on $Q$ is a mapping $m$ such that there is a formula $h$ with the tautology as basic argument and verifying: $\forall p \in Q$, $h \models p$ or $h \models \neg p$. That is to say, it represents a perfect knowledge of the true values of all the formulae in this language. ■

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

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 to say, it does not say anything about $X_J$. A valuation on $U_I$ conditioned to $U_\emptyset$ is said to be an unconditional valuation about $U_I$. Unconditional valuations are always normalized. We only have to take into account its definition and Axiom N3.

It is also immediate that the combination of a valuation on $U_I$ conditioned to $U_J$ and an unconditional valuation on $U_K$, where $J \subseteq K$ gives rise to a normalized valuation.

## 5. THE ROLE OF NORMALIZATION

By combination, partial inconsistency of a valuation contaminates the rest of valuations degrading their information value. The following example shows this effect.

*Example* 4. Assume that we have two possibility distributions $\pi_1$ and $\pi_2$, defined for the variables $X_1$ and $X_2$, respectively. Assume also that $X_1$ takes values in $U = \{u_1, u_2\}$ and $X_2$ takes values on $V = \{v_1, v_2\}$, and that $\pi_1$, $\pi_2$ are given by,

$$\pi_1(u_1) = 0, \quad \pi_1(u_2) = 0.2 \tag{22}$$

$$\pi_2(v_1) = 1, \quad \pi_2(v_2) = 0.7 \tag{23}$$

In these conditions, $\pi_1 \otimes \pi_2$ is given by,

$$\begin{aligned} &\pi_1 \otimes \pi_2(u_1, v_1) = 0, \quad &\pi_1 \otimes \pi_2(u_1, v_2) = 0, \\ &\pi_1 \otimes \pi_2(u_2, v_1) = 0.2, \quad &\pi_1 \otimes \pi_2(u_2, v_2) = 0.2 \end{aligned} \tag{24}$$

And the marginalization of this product on $U_2$ is the valuation given by:

$$(\pi_1 \times \pi_2)^{\downarrow 2}(v_1) = 0.2, \quad (\pi_1 \times \pi_2)^{\downarrow 2}(v_2) = 0.2 \tag{25}$$

By combining $\pi_2$ with $\pi_1$ and then marginalizing to $U_2$, this valuation $\pi_2$ loses all its information value: it does not distinguish between $v_1$ and $v_2$. The partial inconsistency of $\pi_1$ has contaminated the product and $\pi_2$ has become noninformative.

Intuitively, $\pi_1$ and $\pi_2$ are defined for different sets of variables, so they should not influence each other. However, if we combine $\pi_1$ and $\pi_2$ and, for example, $\pi_1$ has partial inconsistency, then this partial inconsistency contaminates $\pi_2$.

The role of normalization is to remove the inconsistency before the combination, in such a way that such undesirable effects do not occur. The following example shows that the situation is different if we calculate with normalized valuations.

*Example* 5. If we normalize each one of the valuations and then we multiply the results, we find:

$$\begin{aligned} &N(\pi_1) \otimes N(\pi_2)(u_1, v_1) = 0, \quad &N(\pi_1) \otimes N(\pi_2)(u_1, v_2) = 0, \\ &N(\pi_1) \otimes N(\pi_2)(u_2, v_1) = 1, \quad &N(\pi_1) \otimes N(\pi_2)(u_2, v_2) = 0.7 \end{aligned} \tag{26}$$

The marginalization of this combination on $U_2$ gives rise to the same original valuation: $(N(\pi_1) \times N(\pi_2))^{\downarrow 2} = N(\pi_2)$.

The situation, in general, is not as easy as in this example. The problem is that when we have more than two valuations, the result may depend on the way we carry out the normalization. In effect, if we have three valuations $V_1$, $V_2$, and $V_3$, then the valuations $N(V_1) \otimes N(V_2) \otimes N(V_3)$, $N(V_1) \otimes N(N(V_2) \otimes N(V_3))$, $N(N(V_1) \otimes N(V_2)) \otimes N(V_3)$ are, in general different and nonnormalized. What option should we take? There is no clear answer at this stage. This paper will try to find reasonable ways of performing normalization, or at least to reduce the number of possible alternatives.

## 6. ISOLATING INCONSISTENCY

In general, in a classical reasoning system, as in probability theory, we start with a general knowledge about a population or problem. This knowledge is composed of several elementary pieces of information, each one of them relating some of the variables in the problem. For example, in Probability Theory, initially we have a family of a priori and conditional probability distributions. In general, there is no inconsistency in these rules or general knowledge. Then we 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 pieces of information and the observations for this particular case. This is the time when partial inconsistency arises and the correct application of normalization becomes relevant.

If the deduction system is defeasible, as most of the systems dealing with knowledge that is imprecise or uncertain are, then all the initial pieces of information have to be taken into account, because we cannot obtain a conclusion from part of the knowledge we have. In Probability Theory, this implies that deductions have to be obtained through a global probability distribution, built taking into account the initial elementary probability distributions.

The process of obtaining an unconditional probability for all the variables (a global probability) was generalized in Ref. 3 to the case of general valuations. This process relies on the consideration of independence relationships between variables as primitive concepts, that may be associated to different uncertainty representations, and that may be known prior to any numerical relationships between the variables in consideration.[4,12,13] Here we give some of the basic ideas for the construction of a global valuation from elementary valuations.

Let us assume that $(X_1, \ldots, X_n)$ is an $n$-dimensional variable and that $\sigma$ is a permutation on the set $\{1, \ldots, n\}$, then a global valuation can be built by means of the combination of the following valuations:

- An unconditional valuation about $X_{\sigma(1)}$.
- For each $i = 2, \ldots, n$, a valuation about $X_{\sigma(i)}$ conditioned on $X_{\{\sigma(1), \ldots, \sigma(i-1)\}}$.

Independence relationships may be used to reduce the size of the valuations to be combined, in the following way:

If $\{\sigma(1), \ldots, \sigma(i-1)\} = I \cup J$ and $X_{\sigma(i)}$ is independent of $X_J$ given the variables $X_I$ then we can consider a valuation about $X_{\sigma(i)}$ conditioned on $X_I$ instead of a valuation about $X_{\sigma(i)}$ conditioned on $X_{\{\sigma(1), \ldots, \sigma(i-1)\}}$.

This initial combination is always normalized.

PROPOSITION 2. *Assume that $V_i$ is a conditional valuation about $X_{\sigma(i)}$ given $X_I$ where $I \subseteq \{\sigma(1), \ldots, \sigma(i)\}$ then $V = \bigotimes_{i\in\{1,\ldots,n\}} V_i$ is an unconditional valuation for $(X_1, \ldots, X_n)$, or equivalently, it is a normalized valuation for these variables.*

*Proof.* The proof will be by induction on $n$.

For $n = 1$, the result is immediate, with $V_1$ being an unconditional valuation about $X_{\sigma(1)} = X_1$.

If the proposition is true for $n = k$, then for $n = k + 1$, we have

$$V = \bigotimes_{i\in\{1,\ldots,k,k+1\}} V_i = \left(\bigotimes_{i\in\{1,\ldots,k\}} V_i\right) \otimes V_{k+1} \tag{27}$$

Now, using the hypothesis of induction, $\bigotimes_{i\in\{1,\ldots,k\}} V_i$, is a normalized valuation on $\{\sigma(1), \ldots, \sigma(k)\}$. With $V_{k+1}$ being a valuation about $X_{\sigma(k+1)}$ conditioned to $X_I$ with $I \subseteq \{\sigma(1), \ldots, \sigma(k)\}$, then the product is a valuation on $\{\sigma(1), \ldots, \sigma(k), \sigma(k+1)\} = \{1, \ldots, n\}$. Furthermore, following Axioms 2 and 3, we obtain:

$$V^{\downarrow\emptyset} = \left(\bigotimes_{i\in\{1,\ldots,k,k+1\}} V_i\right)^{\downarrow\emptyset} = \left(\left(\bigotimes_{i\in\{1,\ldots,k\}} V_i\right) \otimes V_{k+1}\right)^{\downarrow\emptyset} \tag{28}$$

$$= \left(\left(\left(\left(\bigotimes_{i\in\{1,\ldots,k\}} V_i\right) \otimes V_{k+1}\right)^{\downarrow I}\right)\right)^{\downarrow\emptyset} = \left(\left(\bigotimes_{i\in\{1,\ldots,k\}} V_i\right) \otimes V_0^I\right)^{\downarrow\emptyset}$$

$$= \left(\bigotimes_{i\in\{1,\ldots,k\}} V_i\right)^{\downarrow\emptyset} = V_0^{\emptyset}. \tag{29}$$

■

Below we provide abstract definitions of observation and *a posteriori* valuation.

DEFINITION 3. *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 may be characterized (see Definition 1) as valuations for a variable that may not be refined by combination without obtaining the contradiction, i.e., they represent perfect knowledge about the value of these variables.

The combination of *a priori* information with observations gives rise to the so called *a posteriori* information which represents the particularization on the *a priori* valuation for the case given by the observations.[3]

DEFINITION 4. *If V is a global unconditional valuation about the variables* $(X_1, \ldots, X_n)$ *and* $\{O_i\}_{i\in I}$ *a family of observations about these variables, we call* $V_J^{PS} = ((\bigotimes_{i\in I} O_i) \otimes V)^{\downarrow J}$ *the a posteriori information about variables* $X_I$ *induced by V and* $\{O_i\}_{i\in I}$, *where* $J \subseteq \{1, \ldots, n\}$.

The problem with this definition is that the consideration of a global valuation and the combination without normalization allows the propagation of partial inconsistency. Our approach will be to use the components of the global valuation, the elementary valuations giving rise thereto, instead of the global valuation itself trying to normalize before combining. Under these conditions, assume that we have a family of valuations $\{V_i\}_{i\in\{1,\ldots,n\}}$ in such a way that the combination of the valuations is an unconditional valuation for all the variables (*a priori* information) and that $\{O_i\}_{i\in I}$ is a family of observations for a particular case, let $\mathcal{H} = \{V_i\}_{i\in\{1,\ldots,n\}} \cup \{O_i\}_{i\in I}$ be the family of all the valuations we want to combine, then we make the following definitions.

DEFINITION 5. *We say that* $V^P$ *is an improved particularization of* $\{V_i\}_{i\in\{1,\ldots,n\}}$ *to the observations* $\{O_i\}_{i\in I}$ *if and only if we can transform* $\mathcal{H}$ *by a successive application of the following rule:*

***Combine*** *Take* $\mathcal{H}' \subseteq \mathcal{H}$, *transform* $\mathcal{H}$ *on* $(\mathcal{H} - \mathcal{H}') \cup \{N(\otimes_{V \in \mathcal{H}'} V)\}$ *until* $\mathcal{H} = \{V^P\}$.

Nothing is said in this definition about how to select the subsets $\mathcal{H}'$ of $\mathcal{H}$ to which to apply the rules. Different criteria will give rise to different improved particularizations. Our main objective will be to determine which are the reasonable strategies according to general principles.

DEFINITION 6. *We say that* $V_J^{AP}$ *is an improved a posteriori information about* $X_J$ *of the a priori valuations* $\{V_i\}_{i \in \{1,\dots,n\}}$ *given the observations* $\{O_i\}_{i \in I}$ *if and only if* $V_J^{AP} = (V^P)^{\downarrow J}$, *where* $V^P$ *is an improved particularization induced by the same sets of valuations.*

### 6.1. The Strategy for the Simple Case

A special case in which we can determine a simple general strategy is when the following axiom is verified:

**Axiom N5.** $\forall V_1, V_2 \in \mathcal{V}'$, then $N(V_1 \otimes V_2) = N(N(V_1) \otimes N(V_2))$.

An equivalent version of this axiom is the following:

$$\forall V_1, V_2, V_3 \in \mathcal{V}' \text{ if } N(V_2) = N(V_3) \text{ then } N(V_1 \otimes V_2) = N(V_1 \otimes V_3) \tag{30}$$

When this is verified then we can prove the following theorem:

THEOREM 1. *If Axiom N5 is verified and* $V^P$ *is an improved particularization relating to* $\mathcal{H}$, *then*

$$V^P = N\left(\bigotimes_{V \in \mathcal{H}} V\right) \tag{31}$$

*Proof.* Equation 31 is trivial when $\mathcal{H}$ has only one element.

Let us prove the theorem showing that if $\mathcal{H}_{old}$ and $\mathcal{H}_{new}$ are the family $\mathcal{H}$ before and after applying the rule in Definition 5 for $\mathcal{H}' \subseteq \mathcal{H}_{old}$, then

$$N\left(\bigotimes_{V \in \mathcal{H}_{old}} V\right) = N\left(\bigotimes_{V \in \mathcal{H}_{new}} V\right) \tag{32}$$

This is a consequence of the following equalities, in which Axiom 5 is applied:

$$\begin{aligned} N\left(\bigotimes_{V \in \mathcal{H}_{new}} V\right) &= N\left(\left(\bigotimes_{V \in \mathcal{H}_{old} - \mathcal{H}'} V\right) \otimes N\left(\bigotimes_{V \in \mathcal{H}'} V\right)\right) \\ &= N\left(N\left(\bigotimes_{V \in \mathcal{H}_{old} - \mathcal{H}'} V\right) \otimes N\left(\bigotimes_{V \in \mathcal{H}'} V\right)\right) \\ &= N\left(\left(\bigotimes_{V \in \mathcal{H}_{old} - \mathcal{H}'} V\right) \otimes \left(\bigotimes_{V \in \mathcal{H}'} V\right)\right) = N\left(\bigotimes_{V \in \mathcal{H}_{old}} V\right) \end{aligned} \tag{33}$$

■

This theorem states that if Axiom N5 is verified then all the particularizations are the same and can be calculated by combining all the valuations and normalizing at the end. So, in this case, the partial inconsistency can be removed for all the valuations at the same time without problems of degrading the quality of final results. Unfortunately this axiom is not always verified.

*Example* 6.

(1) *Probability Theory*. In this case Axiom N5 is trivially verified. We just have to consider that $CG(N(p_1) \otimes N(p_2)) = CG(p_1 \otimes p_2)/(CG(p_1) \cdot CG(p_2))$, then

$$\begin{aligned} N(N(p_1) \otimes N(p_2))(u) &= \frac{(N(p_1) \otimes N(p_2))(u)}{CG(p_1 \otimes p_2)/(CG(p_1) \cdot CG(p_2))} \\ &= \frac{N(p_1)(u^{\downarrow I}) \cdot N(p_2)(u^{\downarrow J}) \cdot CG(p_1) \cdot CG(p_2)}{CG(p_1 \otimes p_2)} \\ &= \frac{p_1(u^{\downarrow I}) \cdot p_2(u^{\downarrow J})}{CG(p_1 \otimes p_2)} = \frac{(p_1 \otimes p_2)(u)}{CG(p_1 \otimes p_2)} = N(p_1 \otimes p_2)(u) \end{aligned} \tag{34}$$

(2) *Infinitesimal Probability*. The presence of infinitesimal values changes everything. In effect, consider $U = \{u_1, u_2\}$ and two infinitesimal probabilities, $p_1$ and $p_2$, defined on this set:

$$p_1(u_1) = \varepsilon, \qquad p_1(u_2) = \varepsilon$$
$$p_2(u_1) = 0.25, \quad p_2(u_2) = 0.75$$

It is very easy to verify that

$$N(p_1 \otimes p_2) = N(p_1) \neq p_2 = N(N(p_1) \otimes N(p_2)) \tag{35}$$

(3) *Possibility Theory*. For the same set as in infinitesimal probabilities, consider two possibilities, $\pi_1$ and $\pi_2$, given by:

$$\pi_1(u_1) = 0.1, \qquad \pi_1(u_2) = 0.1$$
$$\pi_2(u_1) = 0.25, \quad \pi_2(u_2) = 1$$

The same can be verified:

$$N(\pi_1 \otimes \pi_2) = N(\pi_1) \neq \pi_2 = N(N(\pi_1) \otimes N(\pi_2)) \tag{36}$$

(4) *Symbolic Evidence*. Consider $Q = \{p\}$, $A = \{A_1\}$ and the symbolic basic assignments, $m_1$ and $m_2$, given by

$$m_1(\emptyset) = A_1, m_1(p) = \bot, m_1(\neg p) = \bot, m_1(\top) = \neg A_1$$
$$m_2(\emptyset) = \bot, m_2(p) = A_1, m_2(\neg p) = \bot, m_2(\top) = \neg A_1$$

Axiom N5 is not verified either:

$$N(m_1 \otimes m_2) = N(m_1) \neq m_2 = N(N(m_1) \otimes N(m_2)) \tag{37}$$

When Axiom N5 is verified, we can define an equivalence relationship on the set of valuations, $\mathcal{V}$, given by

$$V_1 \sim V_2 \text{ if and only if } V_1, V_2 \neq V_c \text{ and } N(V_1) = N(V_2) \tag{38}$$

We can extend the operations of marginalization and combination to equivalence classes of valuations:

- *Marginalization:* If $J \subseteq I$ and $V_1 \in \mathcal{V}_I$ then, $[V_1]^{\downarrow J} = [V_1^{\downarrow J}]$.
- *Combination:* If $V_1 \in \mathcal{V}_I$ and $V_2 \in \mathcal{V}_J$, then $[V_1] \otimes [V_2] = [V_1 \otimes V_2]$

The result of Marginalization is independent of the valuation we select in the class because of Axiom N4. The combination is also independent of the chosen valuations when Axiom N5 is verified.

With this definition Axiom 6 is verified for classes of valuations: If $V \in \mathcal{V}_\emptyset$ and $V \neq V_c$, then by Axiom N3, $N(V)^{\downarrow\emptyset} = V_0$ and by Axiom N4, $N(V)^{\downarrow\emptyset} = N(V^{\downarrow\emptyset}) = N(V)$. Thus, $N(V) = V_0 = N(V_0)$ and $V \sim V_0$, or equivalently $[V] = [V_0]$.

This axiom states that all the valuations defined on the empty set of variables are the neutral element or the total contradiction: we do not have partial contradiction. In conclusion, under Axiom N5, we can define an equivalence relationship in such a way that partial contradiction disappears.

### 6.2. Particularization: The General Case

In general, for a family $\mathcal{H} = \{V_i\}_{\{1,\ldots,n\}} \cup \{O_i\}_{i \in I}$ there a number of different valuations which verify the definition of improved particularization with respect to $\mathcal{H}$. In this section we shall try to give some general rules or principles to reduce the number of improved particularizations.

The first idea for the selection of $\mathcal{H}'$ is that the frame in which a valuation is defined should be the basis for including a valuation on $\mathcal{H}'$. That is,

$$\text{If } V_1, V_2 \in \mathcal{V}_K, \text{ then } (V_1 \in \mathcal{H}' \Leftrightarrow V_2 \in \mathcal{H}') \tag{39}$$

The idea behind this property is that all the valuations giving information about the same set of variables should be combined and normalized at the same time. Variables should be grouped according to what they give information about, as a first step. For example, if we have two variables, $X_1$ and $X_2$, then all the information about $X_1$ should be combined and normalized, then all the information about $X_2$ and finally the result combined with the information relating the 2 variables. It does not make much sense to take a valuation about $X_1$, combine it with a valuation about the two variables, normalize and then combine the result with a valuation about $X_1$. We should take all the information about $X_1$ first, clarify (normalize) it and then proceed by combining the result with the rest of the information.

Following these ideas, we can extract another general rule: first we should select the valuations given in smaller frames. That is, if $J_1 \subset J_2$ we should take the valuations belonging to $\mathcal{V}_{J_1}$ before the valuations on $\mathcal{V}_{J_2}$ to combine and normalize. In other words, it does not make sense to consider a valuation relating two variables before the information about each one of them has been combined and normalized. We should carry on by clarifying information from small to large frames. The following example illustrates these ideas.

*Example* 7. Assume two variables: $X_1$ and $X_2$. $X_1$ is the input to a communication channel, and $X_2$ is the output. The values for input and output can be 0 or 1. We assume that the input is almost always 0 and that the channel almost always works properly, that is the output is 0 (1) if the input is 0 (1). This *a priori* information can be represented by means of two infinitesimal probabilities:

- One informing about the input, $X_1$:

$$p_1(0) = 1, \quad p_1(1) = \varepsilon$$

- One relating the input and the output: $p_2(u_i, v_j) = p(X_2 = v_j \mid X_1 = u_i)$:

$$p_2(0,0) = 1, \quad p_2(0,1) = \varepsilon, \quad p_2(1,0) = \varepsilon, \quad p_2(1,1) = 1$$

Furthermore, we have observed that the input is 1: $X_1 = 1$. This is represented by the valuation:

$$O_1(0) = 0, \quad O_1(1) = 1$$

The first thing that we have to do is to clarify the input, by combining $p_1$ and $O_1$ and normalizing the result: we obtain $O_1$ again. This information can be combined now with $p_2$ whereby we obtain the following particularization:

$$p_1^p(0,0) = 0, \quad p_1^p(0,1) = 0, \quad p_1^p(1,0) = \varepsilon, \quad p_1^p(1,1) = 1$$

which is a reasonable conclusion: the input is 1 and the output will surely be 1.

However, if we first combine $p_1$ and $p_2$, normalize and combine the result with $O_1$, the resulting particularization is:

$$p_2^p(0,0) = 0, \quad p_2^p(0,1) = 0, \quad p_2^p(1,0) = 1/2, \quad p_2^p(1,1) = 1/2$$

With this particularization we are still sure that the input is 1, but now we are not sure about the output: since we did not solve the partial inconsistency between $p_1$ and $O_1$ at the appropriate stage, this inconsistency has *contaminated* the information about the output.

Following these ideas, we make the definitions below:

DEFINITION 7. *We say that $V^p$ is a frame-based improved particularization of the family of valuations $\mathcal{H}$, if and only if it can be obtained according to the procedure in Definition 5, where every $\mathcal{H}'$ verifies condition* (39) *and the following condition:*

$$\text{If } V_1, V_2, V_3 \in \mathcal{H}', V_i \in \mathcal{V}_{K_i} (i = 1, 2, 3) \text{ and } K_1 \subseteq K_2 \subseteq K_3 \text{ then } K_2 = K_3 \quad (40)$$

DEFINITION 8. *We say that* $V_J^{AP}$ *is a frame-based improved a posteriori information about* $X_J$ *of the a priori valuations* $\{V\}_{i\in\{1,...,n\}}$ *given the observations* $\{O_i\}_{i\in I}$ *if and only if* $V_J^{AP} = (V^P)^{\downarrow J}$, *where* $V^P$ *is a frame-based improved particularization induced by the same sets of valuations.*

Condition 40 states that we should first solve the inconsistency in smaller frames: if $K_2 \neq K_3$, then the inconsistency between $V_1$ and $V_2$ should have been solved before the combination with $V_3$.

*Example* 8. If we have 2 variables, $X_1$, and $X_2$, and $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_{12}$ are the valuations from $\mathcal{H}$ informing about the first variable, the second variable and the two variables, respectively, then a frame based *a posteriori* information can be obtained, by following the steps below:

- Let $\mathcal{H}' = \mathcal{H}_1$ and $W_1 = N(\bigotimes_{V\in\mathcal{H}_1} V)$. After applying **Combine** we have $\mathcal{H} = \{W_1\} \cup \mathcal{H}_2 \cup \mathcal{H}_{12}$.
- Let $\mathcal{H}' = \mathcal{H}_2$ and $W_2 = N(\bigotimes_{V\in\mathcal{H}_2} V)$. After applying **Combine** we have $\mathcal{H} = \{W_1, W_2\} \cup \mathcal{H}_{12}$.
- Let $\mathcal{H}' = \mathcal{H}_{12} \cup \{W_1\}$ and $W_{12} = N((\bigotimes_{V\in\mathcal{H}_{12}} V) \otimes W_1)$. After applying **Combine** we have $\mathcal{H} = \{W_2, W_{12}\}$.
- Apply rule **Combine** to the set $\mathcal{H}' = \{W_2, W_{12}\}$. We finally obtain the frame based particularization $V^P = N(W_2 \otimes W_{12})$.

Note that in this example the information about $X_1$ has been combined with the information relating the two variables, before the information about the second variable. The reverse could also be possible (valuations about $X_2$ before valuations about $X_1$). This may seem a bit arbitrary: there is no reason to do one or the other.

The different possibilities of obtaining a frame-based particularized valuation can be significantly reduced if we impose the following condition on the selection of sets $\mathcal{H}'$:

$$\text{If } V_1 \in \mathcal{H}, V_2 \in \mathcal{H}', V_i \in \mathcal{V}_{K_i} (i = 1, 2) \text{ and } K_1 \subseteq K_2 \text{ then } V_1 \in \mathcal{H}' \tag{41}$$

That is, if we consider a valuation to combine, then we have to consider all the valuations defined for a smaller set of variables. Under these conditions, the case in Example 8 is not possible: When we combine the valuations on $\mathcal{H}_{12}$ we also have to consider $W_1$ and $W_2$, obtaining a particularization at this step.

Although condition 41 reduces the number of final possibilities, we have no guarantee that we are going to obtain a single final particularization, as the following example shows.

*Example* 9. Assume that we have 3 variables, $X_1$, $X_2$, $X_3$, and $\mathcal{H} = \{V_1, V_2, V_3, V_{12}, V_{23}\}$. Then we have two possibilities:

(a) • Select $\mathscr{H}' = \{V_1, V_2, V_{12}\}$. Result: $W_{12}$.
• Select $\mathscr{H}' = \{V_3, V_{23}\}$. Result: $W_{23}$.
• Select $\mathscr{H}' = \{W_{12}, W_{23}\}$. Result: $V^p$ (particularization).

(b) • Select $\mathscr{H}' = \{V_2, V_3, V_{23}\}$. Result: $W'_{23}$.
• Select $\mathscr{H}' = \{V_1, V_{12}\}$. Result: $W'_{12}$.
• Select $\mathscr{H}' = \{W'_{12}, W'_{23}\}$. Result: $V'^p$ (particularization).

$V^p$ and $V'^p$ are different frame-based improved particularizations verifying condition 41.

### 6.3. *A Posteriori* Information

In this case, we not only want to particularize a set of valuations to a set of observations, but also to marginalize the result to one or several variables, $X_J$. Does this make any difference? Our point is that the strategies are not the same if we want to combine all the pieces of information in a global valuation, or if our purpose is to marginalize the result in a set of variables. In the former, we want to integrate all the information. In the latter, we want to calculate the marginalization on a given objective: we have to combine and marginalize.

Consider the case of Example 8 with the difference that we now want to calculate the *a posteriori* information for $X_2$, instead of a particularization of all the variables. In this situation, condition 41 does not make much sense, and the case that was considered unreasonable: combination of the information about $X_1$ with the information relating $X_1$ and $X_2$, before considering the information about $X_2$, now makes more sense.

Now there is an asymmetry between the variables which was not present before: we want to reach $X_2$ and we start by considering all the information for $X_1$.

If we have more than two variables, independence relationships can be a very useful guide to find partial ordering among the variables, guiding the way in which we should choose the valuations to be combined and normalized. For example, if we have three variables, $X_1$, $X_2$, and $X_3$, and we want to calculate the *a posteriori* information on $X_3$ and we know that variables $X_1$ and $X_3$ are independent given variable $X_2$, then it seems reasonable to take all the valuations defined for $X_1$, then all the valuations for $X_2$, and finally the valuations defined for $X_3$.

The reason is that all the dependence between $X_1$ and $X_3$ goes through $X_2$. The flow of information is from $X_1$ to $X_2$ and then, from $X_2$ to $X_3$. As a consequence, it is more reasonable to combine all the valuations defined on {1}, remove inconsistency, combine the result with the valuations not combined before and defined on {1, 2}, remove inconsistency, and finally combine the result with the valuations not combined before and defined on subsets {1, 2. 3}, removing inconsistency. There is no valuation relating $X_1$ and $X_3$: the relation between these two variables is obtained through the relation of $X_1$ with $X_2$ and the relation of $X_2$ with $X_3$. The following practical example may help to understand the reasoning.

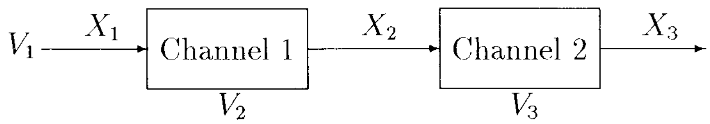


**Figure 1.** Message transmission in Example 6.3.

*Example* 10. Assume a message (see Fig. 1) that is transmitted through a channel and the message received is transmitted again using a different channel to a new place. We consider three variables:

$X_1$ The original message

$X_2$ The message received in the first place

$X_3$ The message received in the second place

This is a clear example in which $X_1$ and $X_3$ are conditionally independent given $X_2$. The general knowledge about the system can be given by the following valuations:

$V_1$ A valuation about the original message, $X_1$.

$V_2$ A valuation relating the input and the output to channel 1, $X_1$ and $X_2$.

$V_3$ A valuation relating the input and the output to channel 2, $X_2$ and $X_3$.

If we have an observation about $X_1$, $O_1$, that is we know the input to channel 1 and we want to calculate the *a posteriori* valuation about $X_3$, then the partial inconsistency between $O_1$ and $V_1$ should be removed before combining the result with $V_2$, to obtain the resulting information about $X_2$. Finally this information is combined with $V_3$ and normalized. There is an order among the variables to remove inconsistency: First information about $X_1$, then information about $X_2$, and finally information about $X_3$.

DEFINITION 9. *An ordered partition of* $\{1, \ldots, n\}$ *is a vector* $(J_1, \ldots, J_k)$ *verifying the following conditions:*

- $J_i \subseteq \{1, \ldots, n\}$, $\forall i = 1, \ldots, k$.
- $\bigcup_{i=1,\ldots,k} J_i = \{1, \ldots, n\}$.
- *If* $i \neq j$ *then* $J_i \cap J_j = \emptyset$.

An ordered partition will be the basis to remove inconsistency for calculating *a posteriori* valuations. Ordered partitions should be based on independence relationships.

DEFINITION 10. *We say that the ordered partition* $(J_1, \ldots, J_k)$ *is based on the independence relationships between variables* $\{X_1, \ldots, X_n\}$ *if and only if* $X_{J_i}$ *is independent of* $X_{L_{i-2}}$ *given* $X_{J_{i-1}}$, *for every* $i = 3, \ldots k$, *and where* $L_i = \bigcup_{j=1,\ldots,i} J_j$.

There is always a trivial ordered partition which is based on independence relationships: the one given by $(J_1)$ where $J_1 = \{1, \ldots, n\}$. This partition will not be very useful. We are interested in partitions that express a maximal set of independence relationships.

DEFINITION 11. *We say that $(J_1, \ldots, J_k)$ is a maximal ordered partition based on the independence relationships between variables $\{X_1, \ldots, X_n\}$ if and only if every ordered partition $(J_1, \ldots, J_{i-1}, J_i'^1, J_i'^2, J_{i+1}, \ldots, J_k)$, where $J_i'^1 \cap J_i'^2 = \emptyset$, $J_i'^1 \cup J_i'^2 = J_i$ is not based on independences between variables $X_{\{1,\ldots,n\}}$.*

DEFINITION 12. *We say that $V_J^{AP}$ is a frame-based improved a posteriori information about $X_J$ of the family of valuations $\mathcal{H}$ and based on independence relationships if and only if $V_J^{AP} = (V^P)^{\downarrow J}$, where $V^P$ is a frame-based improved particularization induced by the same set of valuations and there is a maximal ordered partition $(J_1, \ldots, J_k)$ based on independence relationships verifying:*

- $J = J_k$
- *The strategy to select $\mathcal{H}'$ verifies for every three different valuations $V_1, V_2, V_3 \in \mathcal{H}$ the following condition.*

$$\begin{aligned} &\textit{If } V_i \in \mathcal{V}_{K_i} (i = 1, 2, 3), \textit{ and } K_1 \cap L_j \neq \emptyset, K_2 \cap L_j \neq \emptyset, \\ &K_3 \subseteq \{1, \ldots, n\} - L_j \textit{ then } V_3 \notin \mathcal{H}' \end{aligned} \tag{42}$$

*where $L_j$ is as in Definition 10.*

This definition says that *a posteriori* valuations should be calculated by combining first all the valuations on $J_1$, then consider the valuations on $J_2$, and so on, until we reach the valuations defined for $J_k$, a set containing $J$.

## 7. PROPAGATION ALGORITHMS

The strategies presented in the previous section have the problem of being very complex from a computational point of view. Each time we combine valuations, the resulting valuation is defined on the union of the sets of the combined valuations. As a consequence, the set of definition of valuations is always increasing, and finally we obtain a valuation defined for all the variables, which is infeasible in most cases. Propagation algorithms for probabilities[4,14] and for general uncertainty formalisms[1,2] try to solve this problem, by marginalizing the valuations as soon as possible, without waiting till the end. The most important and basic rule for propagation algorithms is the following:

If we want to calculate the *a posteriori* information in $X_J$ and there is $L \subseteq \{1, \ldots, n\} - J$ such that there is a single valuation, $W$, on $\mathcal{H}$ such that its set of definition, $K$, has a non-empty intersection with $L$, then we can transform $W$ into $W^{\downarrow K-L}$, without affecting the final result.

This rule, which is valid for the case in which normalization is not used, is the basis for the reduction of calculations in propagation algorithms. We can select $l \in \{1, \ldots, n\} - J$, then multiply all the valuations which are defined for variable $X_l$ (and possibly more variables). If the result is defined for the variables $K$, the rule above can be applied and we can marginalize the result to $K - \{l\}$ (*delete l*). Marginalizing a valuation decreases the size of our data, i.e., the size of the valuations we have to combine and therefore the amount of calculations. Most propagation algorithms are based on an ordered application of this rule to delete variables.

Here we are going to prove that this rule is also valid to calculate improved *a posteriori* valuations. This will be a consequence of the following theorem.

THEOREM 2. *Assume that $\mathcal{H}_1$ is a family of valuations and that $J \subseteq \{1, \ldots, n\}$, $L \subseteq \{1, \ldots, n\} - J$ and that there is a single valuation, W, on $\mathcal{H}_1$ such that its set of definition, K, has a non-empty intersection with L, then if $\mathcal{H}_2$ is the set of valuations given by $\mathcal{H}_2 = (\mathcal{H}_1 - \{W\}) \cup \{W^{\downarrow K-L}\}$ then $V_J^{PS}$ is an a posteriori valuation on $X_J$ with respect to $\mathcal{H}_1$, if and only if it is an a posteriori valuation with respect to $\mathcal{H}_2$*

*Proof.* The proof will be based on the following points:

- $\mathcal{H}_1$ and $\mathcal{H}_2$ have the same number of valuations. The only difference is that $W$ in $\mathcal{H}_1$ is replaced with $W^{\downarrow K-L}$ in $\mathcal{H}_2$. So we can apply to the two families the same strategies for the selection of subsets of valuations $\mathcal{H}'$, making the appropriate transformation when necessary. Note that we are not talking about frame-based improved *a posteriori* valuations. In this case, as the frames of $W$ and $W^{\downarrow K-L}$ are different, it is possible that we can not choose equivalent subsets for the two families.
- We are going to prove that the property relating $\mathcal{H}_1$ and $\mathcal{H}_2$ is also verified after applying the **Combine** rule under equivalent sets $\mathcal{H}'$.
- If this condition is kept after each application of the **Combine** rule it will be true at the end, when we reach a single valuation on the sets: $\mathcal{H}_1 = \{V_1^P\}$, $\mathcal{H}_2 = \{V_2^P\}$, and $V_i^P$ is an improved particularization with respect to $\mathcal{H}_i$. The only possibility is that $W = V_1^P$. Therefore, we have $V_2^P = V_1^{P\downarrow(\{1,\ldots,n\}-L)}$. From Axiom 2 for valuations, the marginalization to $J$ of the 2 valuations is the same:

$$V_2^{P\downarrow J} = (V_1^{P\downarrow(\{1,\ldots,n\}-L)})^{\downarrow J} = V^{P\downarrow J} \tag{43}$$

We only have to prove the second point: that the property relating $\mathcal{H}_1$ and $\mathcal{H}_2$ is verified also after applying the **Combine** rule. Let $\mathcal{H}_1^*$ and $\mathcal{H}_2^*$ be the transformations of $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, after applying the **Combine** rule for equivalent $\mathcal{H}'$. Let $\mathcal{H}'_1$ be the subset for $\mathcal{H}_1$ and $\mathcal{H}'_2$ the subset for $\mathcal{H}_2$. We have two possibilities:

(1) $W \notin \mathcal{H}'_1$. In this case the sets $\mathcal{H}'_1 = \mathcal{H}'_2$ and it is immediate that after the transformation we have that $\mathcal{H}_2^* = (\mathcal{H}_1^* - \{W\}) \cup \{W^{\downarrow K-L}\}$.
(2) $W \in \mathcal{H}'_1$. Then $\mathcal{H}'_2 = (\mathcal{H}'_1 - \{W\}) \cup \{W^{\downarrow K-L}\}$ and we can prove the following:

$$\mathscr{H}_2^* = \left(\mathscr{H}_1^* - \left\{N\left(\bigotimes_{V\in\mathscr{H}_1'} V\right)\right\}\right) \cup \left\{N\left(\bigotimes_{V\in\mathscr{H}_2'} V\right)\right\} \tag{44}$$

The condition is now verified for the valuation $N(\bigotimes_{V\in\mathscr{H}_1'} V)$. It is the only valuation with a set of definition, $K$, with a non-empty intersection with $L$ and applying the Axioms for valuations

$$\left(N\left(\bigotimes_{V\in\mathscr{H}_1'} V\right)\right)^{\downarrow K-L} = N\left(\bigotimes_{V\in\mathscr{H}_1'} V\right)^{\downarrow K-L} = N\left(\left(\bigotimes_{V\in\mathscr{H}_1'-\{W\}} V\right)\otimes W\right)^{\downarrow K-L} \tag{45}$$

$$= N\left(\left(\bigotimes_{V\in\mathscr{H}_1'-\{W\}} V\right)\otimes W^{\downarrow K-L}\right) = N\left(\bigotimes_{V\in\mathscr{H}_2'} V\right) \tag{46}$$

That is, in both cases, the condition is verified after applying **Combine.**

In this theorem, we have proven that we can marginalize in an intermediate step and getting the same improved *a posteriori* valuation. There is no difficulty in generalizing this theorem showing that this marginalization can be applied several times in the process of obtaining an *a posteriori* valuation. So allowing the application of propagation algorithms.

## 8. CONCLUSIONS

In this paper, we have given an axiomatic framework for the normalization operator applied to general valuations. This operator can be used to remove partial inconsistency in calculations with valuations. The main problem with its usage is that the final result depends on the concrete places in which the normalization is applied. Furthermore, there is not an obvious way of determining where to apply the normalization. In this paper we support the idea that independence structures may be a good guide to determine where to normalize.

Shenoy[15] has also considered the problem of inconsistency in hypertrees, however, the point of view is different. He considers the problem of detecting maximally consistent sets of valuations in cases in which there is total inconsistency. In our case we consider the removing of partial inconsistency. The two works can be seen as complementary and a general model could be devised encompassing both methodologies.

Our objective has been to obtain the normalization points from general principles. However this objective has not been achieved in all its extension and, although the number of possible alternatives is not very great, some degree of freedom remains in normalization points. Our feeling is that it is very difficult to reduce it with additional general principles. At least, we have tried to do it without going further.

In this axiomatic we have tried to codify the main properties of a normalization operator. However, we may find particular cases in which this framework does not apply. In concrete, in Bolanos, De Campos and Moral[16] we have considered valuations representing probabilities with linguistic labels. There is a normalization mechanism for this type of valuations in which $N(V)$ is a set of valuations. Obviously, this case is not covered by our axiomatic system. We could have defined a more general model including this case, but this would have complicated even more the proposed methodology.

We have shown that the basic property to define propagation algorithms is verified. So we can achieve in this case the same efficiency as when normalization is not used. We have not given the details of how calculations can be organized in graphical structures, such as trees of cliques.[13] But the process would be analogous to the one used for general valuations.[1,2]

## References

1. G. Shafer and P.P. Shenoy, ''Local computation in hypertrees,'' working Paper N. 201. School of Business, University of Kansas, Lawrence, 1988.
2. P.P. Shenoy and G. Shafer, ''Axioms for probability and belief-functions propagation,'' in *Uncertainty in Artificial Intelligence, 4,* R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer, Eds., North-Holland, Amsterdam, 1990, 169–198.
3. J.E. Cano, M. Delgado, and S. Moral, ''An axiomatic framework for propagating uncertainty in directed acyclic graphs,'' *Int. J. Approx. Reasoning,* **8,** 253–280 (1993).
4. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* Morgan and Kaufmann, San Mateo, 1988.
5. L.A. Zadeh, ''A theory of approximate reasoning,'' in *Machine Intelligence,* **9,** J.E. Hayes and D. Mikulich, Eds., Elsevier, Amsterdam, 1979, 149–194.
6. J. Pearl, ''From conditional oughts to qualitative decision theory,'' in *Uncertainty in A.I. Proceedings of the 11th Conference,* D. Heckerman, A. Mamdani, Eds., Morgan and Kaufmann, San Mateo, 1993, 12–20.
7. N. Wilson, ''An order of magnitude calculus, in *Uncertainty in A.I. Proceedings of the 11th Conference,* Ph. Besnard and S. Hanks, Eds., Morgan and Kaufmann, San Mateo, 1995, 548–555.
8. D. Dubois and H. Prade, *Possibility Theory. An Approach to Computerized Processing of Information,* Plenum Press, New York, 1988.
9. L.A. Zadeh, ''Fuzzy sets as a basis for a Theory of Possibility,'' *Fuzzy Sets and Systems,* **1,** 3–28 (1978).
10. J. Kohlas, ''Symbolic evidence, arguments, supports and valuation networks,'' in *Symbolic and Quantitative Approaches to Reasoning and Uncertainty,* M. Clarke, R. Kruse, S. Moral, Eds., Springer Verlag, Berlin, 1993, 186–198.
11. G. Shafer, ''An axiomatic study of computation in hypertrees,'' working Paper N. 232. School of Business, University of Kansas, Lawrence, 1991.
12. D. Geiger and J. Pearl, ''Logical and algorithmic properties of conditional independence,'' Technical Report R-97, Cognitive Systems Laboratory, University of California, Los Angeles, 1988.
13. G. Verma and J. Pearl, ''Identifying Independence in Bayesian Networks,'' *Networks,* **20,** 507–534 (1990).
14. S.L. Lauritzen and D.J. Spiegelhalter, ''Local computation with probability on graphical structures and their applications to expert systems,'' *J.R. Stat. Soc., B,* **50,** 1988, 157–224.
15. P.P. Shenoy, ''Consistency in valuation-based systems,'' working Paper N. 216 School of Business, University of Kansas, Lawrence, 1990.
16. M.J. Bolanos, L.M. de Campos, and S. Moral, (1993) ''Propagation of linguistic labels in causal networks,'' *Second IEEE International Conference on Fuzzy Systems, Vol. II,* San Francisco, 1993, 863–870.