Belief Functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem

Philippe Smets
IRIDIA—Université Libre de Bruxelles, Brussels, Belgium

ABSTRACT

We generalize the Bayes' theorem within the transferable belief model framework. The generalized Bayesian theorem (GBT) allows us to compute the belief over a space Θ given an observation $x \subseteq X$ when one knows only the beliefs over X for every $\theta_i \in \Theta$. We also discuss the disjunctive rule of combination (DRC) for distinct pieces of evidence. This rule allows us to compute the belief over X from the beliefs induced by two distinct pieces of evidence when one knows only that one of the pieces of evidence holds. The properties of the DRC and GBT and their uses for belief propagation in directed belief networks are analyzed. The use of the discounting factors is justified. The application of these rules is illustrated by an example of medical diagnosis.

KEYWORDS: ***belief functions, Bayes' theorem, disjunctive rule of combination.***

1. INTRODUCTION

This paper presents the disjunctive rule of combination (DRC) and the generalized Bayesian theorem (GBT) within the framework of the transferable belief model, a model for quantifying beliefs using belief functions. Their use is illustrated by a typical application in the domain of the medical diagnostic process.

Address correspondence to Philippe Smets, IRIDIA, Université Libre de Bruxelles, 50 av. Roosevelt, CP194 / 6, Brussels, Belgium.
This work has been partially funded by CEC-ESPRIT II Basic Research Project 3085 (DRUMS) and the Belgian National Incentive-Program for Fundamental Research in Artificial Intelligence.
Received January 1, 1992; accepted November 25, 1992.

International Journal of Approximate Reasoning 1993; 9:1–35

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

Suppose $\text{bel}_1: 2^\Omega \to [0,1]$ is a belief function induced on the frame of discernment Ω by a piece of evidence E_1. Suppose $\text{bel}_2: 2^\Omega \to [0,1]$ is a belief function induced on the same frame of discernment Ω by another piece of evidence E_2. Suppose E_1 and E_2 are distinct pieces of evidence (Shafer [1], Smets [2], Smets [3]). Shafer introduced Dempster's rule of combination to compute:

$$\text{bel}_{12} = \text{bel}_1 \oplus \text{bel}_2 \quad .$$

where bel_{12} is the belief function induced on Ω by the conjunction "E_1 and E_2."

We present a combination rule, the DRC, that permits the derivation of the belief function induced on Ω by the disjunction of two pieces of evidence. It corresponds to a situation where you could assess your belief on Ω if E_1 were true, your belief on Ω if E_2 were true, but you only know that the disjunction "E_1 or E_2" is true.

As an example of an application of the DRC, consider the medical diagnosis process. Let X be the domain of symptoms, each x in X being a particular symptom. Let Θ be the domain of diseases, each θ_i in Θ being a particular disease. The diseases θ_i are so defined that they are mutually exclusive and exhaustive. Suppose we have assessed our belief over the symptoms for every disease θ_i and we want to assess our belief over the symptoms knowing only that the patient has either disease θ_1 or disease θ_2. This is the case when it is known that all the diseases excepting θ_1 and θ_2 can be excluded. The DRC provides the solution when the *a priori* belief over θ_1 and θ_2 is vacuous. Its extension to the case where there is a non-vacuous *a priori* over θ_1 and θ_2 can also be obtained.

Simultaneously with the DRC, we derive the GBT. Bayes' theorem is central for probabilistic inference. In the medical diagnostic process considered, let $P(x|\theta_i)$ be the probability of the symptoms given each diagnostic $\theta_i \in \Theta$, and let our *a priori* belief over Θ be quantified by the probability distribution function P_0. After observing the symptom $x \subseteq X$, the probability distribution on Θ is updated into $P(\theta_i|x)$, the *a posteriori* probability distribution on Θ, by the application of Bayes' theorem:

$$P(\theta_i|x) = \frac{P(x|\theta_i)P_0(\theta_i)}{\sum_j P(x|\theta_j)P_0(\theta_j)} \qquad \forall \theta_i \in \Theta$$

In other words, from the probability over X given each $\theta_i \in \Theta$ (and the *a priori* probability on Θ), Bayes' theorem allows us to derive the probability over Θ given any $x \subseteq X$.

The GBT is a generalization of Bayes' theorem where all conditional probabilities are replaced by belief functions and the *a priori* belief function on Θ is vacuous. A further generalization for non-vacuous *a priori* belief on Θ is also presented.

The use of the GBT for medical diagnosis resolves the problem of how to select uncommitted *a priori* probabilities on Θ that can represent the absence of any *a priori* commitment towards any disease. The *vacuous belief* that characterizes a state of total ignorance is used on the disease space Θ. Such a state of ignorance cannot be represented within probability theory; indeed total ignorance means that any strict subset of the disease set Θ should receive the *same* degree of belief. No probability function can describe such a belief state once $|\Theta| > 2$, as the *same* probability should be given to every θi, but also to every $\theta_i \cup \theta_j \ldots$ (any strict subset of Θ).

1.1. Belief Propagation in Directed Networks

Belief networks described by Shafer *et al.* [4] are undirected hyper-graphs. Hyper-nodes represent sets of variables (e.g., the symptoms and the diseases) and hyper-edges are weighted with belief functions defined on the product space of the variables represented by the nodes attached to the hyper-edges. In Pearl's approach (Pearl [5])—concerning only probability functions—the edges are directed and weighted by the conditional probabilities over (the variables represented by) the child node given (the variables represented by) the parent nodes.

In this paper, we provide the tools necessary to use belief functions (instead of probability functions) in directed graphs similar to those considered by Pearl. An edge between a parent node Θ and a child node X will be weighted by conditional belief functions over X for each value θ_i of Θ. Our approach is less general than Shafer's, but we feel that in practice the loss of generality is not important. Indeed we agree with Pearl [5] who argues that it is more "natural" and "easier" to assess conditional probabilities (and conditional beliefs) over X given θ_i than the joint probabilities (and beliefs) over the space $Xx\Theta$, and that in most real life cases only conditional beliefs will be collected.

The DRC can be used for forward propagation in directed networks. Consider two parent nodes, Θ and Ψ, of node X and the conditional belief functions $\mathrm{bel}_X(.|\theta_i)$ and $\mathrm{bel}_X(.|\psi_i)$ on X given each $\theta_i \in \Theta$ and given each $\psi_j \in \Psi$. The conjunctive rule of combination provides the belief function on X given "θ_i and ψ_j." The disjunctive rule of combination provides the belief function on X given "θ_i or ψ_j."

The GBT can be used for backward propagation of beliefs in directed networks between a child node X and its parent node Θ. Given the conditional belief over X given each $\theta_i \in \Theta$, the GBT computes the belief induced on Θ for any $x \subseteq X$.

1.2. Content

In section 2, we define the principle of minimal commitment, the generalized likelihood principle, and the concept of conditional cognitive independence. The first formalizes the idea that one should never give more belief to something than is justified. The second formalizes the idea that the belief induced by a disjunction of several pieces of evidence is a function of the beliefs induced by each piece of evidence. The third extends the idea of stochastic independence to belief functions.

In section 3, we derive the DRC and the GBT. In section 4, we show that they can also be derived through constructive approaches based on the principle of minimal commitment. In section 5, we present some properties of the GBT and some of its limitations. We show in particular that the GBT becomes the classical Bayes' theorem when all the belief functions happen to be probability functions. In section 6, we present the use of the DRC and the GBT for the propagation of beliefs in directed belief networks. In section 7, we present an example of the use of the DRC and GBT for a medical diagnosis problem. In section 8, we summarize the major results and conclude.

1.3. Historical Notes

Smets [6] derived initially both the DRC and the GBT by the technique presented in section 4. Most theorems described here are proved in Smets [6]. The GBT was also presented in Smets [2, 7, 8], discussed at full length in Shafer [9]. The DRC was presented in Moral [10], Dubois and Prade [11], [12], Smets [2], and Cohen et al. [13]. The present paper not only details both rules and many of their properties, but it also provides normative requirements that justify them.

2. BELIEF FUNCTIONS

We present some necessary material concerning belief functions and proceed to expound the following three principles: the principle of minimal commitment, the generalized likelihood principle, and the conditional cognitive independence. Belief functions are used to quantify someone's beliefs. They cover the same domain as subjective probabilities, but do not use the additivity axiom required for probability measures. The existence of "basic belief masses" (bbm) allocated to subsets of a frame of discernment Ω is postulated. For $A \subseteq \Omega$, the bbm $m(A)$ quantifies the portion of belief that supports A without supporting any strict subset of A, and that could be transferred to subsets of A if further information justifies it. This model is at the core of the **transferable belief model**, our interpretation of Dempster–Shafer theory (Smets [2, 14], Smets and Kennes [15], Smets [16]). Our results can be easily transferred to other interpretations of Dempster–Shafer theory, like the hints theory (Kohlas and Monney [17]) or the context model (Gebhardt and Kruse [18]).

2.1. Background

Let Ω be a finite non empty set called the **frame of discernment**. The mapping $\mathrm{bel}: 2^{\Omega} \to [0,1]$ is an (unnormalized) **belief function** iff there exists a set of **basic belief assignment** (bba) $\mathrm{m}: 2^{\Omega} \to [0,1]$ such that:

$$\sum_{A \subseteq \Omega} \mathrm{m}(A) = 1$$

and

$$\mathrm{bel}(A) = \sum_{B \subseteq A,\, B \neq \varnothing} \mathrm{m}(B).$$

(Note that $\mathrm{bel}(\varnothing) = 0$.)

The values of $m(A)$ for A in Ω are called the **basic belief masses** (bbm). $m(\varnothing)$ may be positive; when $m(\varnothing) = 0$ (hence $\mathrm{bel}(\Omega) = 1$), bel is called a *normalized belief function*. In Shafer's presentation, he asserts that $m(\varnothing) = 0$, or equivalently that $\mathrm{bel}(\Omega) = 1$, and consequently, belief combination and conditioning are normalized by dividing the results by appropriate scaling factors. The difference between Shafer's definition and ours was introduced when we considered the difference between the *open-world* and *closed-world* assumptions (Smets [2]). The nature of $m(\varnothing) > 0$ is fully discussed in Smets [19].

Our presentation is developed under the open-world assumption, as described in the transferable belief model. However the whole presentation is still valid under the more restrictive assumption of a closed-world.

Belief functions are in one-to-one correspondence with **plausibility functions** $\mathrm{pl}: 2^{\Omega} \to [0,1]$ and **commonality functions** $\mathrm{q}: 2^{\Omega} \to [0,1]$ where for all $A \subseteq \Omega, A \neq \varnothing$,

$$pl(A) = \mathrm{bel}(\Omega) - \mathrm{bel}(\bar{A}) \quad \text{and} \quad pl(\varnothing) = 0$$

$$q(A) = \sum_{A \subseteq B} m(B) \quad \text{and} \quad q(\varnothing) = 1$$

where $\bar{A}$ is the complement of A relative to Ω.

A **vacuous belief function** is a normalized belief function such that $\text{bel}(A) = 0\ \forall A \neq \Omega$. It quantifies our belief in a state of total ignorance as no strict subset of Ω receives any support.

Suppose bel quantifies our belief about the frame of discernment Ω and we learn that $\bar{A} \subseteq \Omega$ is false. The resulting conditional belief function $\text{bel}(.\,¦A)$ is obtained through **the unnormalized rule of conditioning** (see remark 1 for the use of ¦ for the unnormalized conditioning). $\text{bel}(B¦A)$ can be read as the (degree of) belief of B given A or the belief of B in a context where A holds):

$$
\begin{aligned}
m(B¦A) &= \sum_{X \subseteq \bar{A}} m(B \cup X) && \text{if } B \subseteq A \subseteq \Omega \\
&= 0 && \text{otherwise} \\
\text{bel}(B¦A) &= \text{bel}(B \cup \bar{A}) - \text{bel}(\bar{A}) && \forall B \subseteq \Omega \\
pl(B¦A) &= pl(A \cap B) && \forall B \subseteq \Omega \qquad (2.1)
\end{aligned}
$$

The origin of this relation is to be found in the nature of the transferable belief model itself. A mass $m(B)$ given to B is transferred by conditioning on A to $A \cap B$. Other justifications can also be advanced. $\text{bel}(.\,¦A)$ is the minimal commitment specialization of bel, such that $\text{pl}(\bar{A}¦A) = 0$ (Klawonn and Smets [20]). It can also be derived as the minimal commitment solution where bel("$B|A$") is considered to be the belief in the conditional object "$B|A$" (Nguyen and Smets [21]). Note that these derivations are obtained without ever considering the concept of "combination of distinct pieces of evidence," hence without requiring any definition of the notions of distinctness, combination, and probability).

Consider two belief functions bel_1 and bel_2 induced by two distinct pieces of evidence on Ω. The belief function bel_{12} that quantifies the combined impact of the two pieces of evidence is obtained through the **conjunctive rule of combination:** $\text{bel}_{12} = \text{bel}_1 \textcircled{\scriptsize$\wedge$} \text{bel}_2$ where $\textcircled{\scriptsize$\wedge$}$ represents the conjunctive combination operator. Its computation is based on the basic belief assignment m_{12}:

$$
\forall A \subseteq \Omega \qquad m_{12}(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \qquad (2.2)
$$

Expressed with the commonality functions, it becomes:

$$
q_{12}(A) = q_1(A) q_2(A)
$$

It can also be represented as: (Dubois and Prade [22] proved the relation for m_{12}.)

$$m_{12}(A) = \sum_{B\subseteq\Omega} m_1(A\,\vdots\,B)m_2(B)$$

$$\mathrm{bel}_{12}(A) = \sum_{B\subseteq\Omega} \mathrm{bel}_1(A\,\vdots\,B)m_2(B)$$

$$\mathrm{pl}_{12}(A) = \sum_{B\subseteq\Omega} \mathrm{pl}_1(A\,\vdots\,B)m_2(B)$$

$$q_{12}(A) = \sum_{B\subseteq\Omega} q_1(A\,\vdots\,B)m_2(B) \qquad (2.3)$$

Note that no **normalization factor** appears in these rules.

REMARK 1: DEFINITIONS AND SYMBOLS Almost all authors working with belief functions consider only normalized belief functions, whereas we consider mainly unnormalized belief functions. In order to avoid confusion, we propose to keep the names of Dempster's rule of conditioning and Dempster's rule of combination for the normalized forms of conditioning and conjunctive combination, as was introduced by Shafer [1]. For the unnormalized rules, we propose to use the names of unnormalized rule of conditioning for 2.1, conjunctive rule of combination for 2.2, and disjunctive rule of combination for the rule introduced in section 3.

We also propose to use the following symbols to represent these operations.

Dempster's rule of conditioning:	$\mid$	$\mathrm{bel}(A\mid B)$
unnormalized rule of conditioning:	$\vdots$	$\mathrm{bel}(A\,\vdots\,B)$
Dempster's rule of combination:	$\oplus$	$\mathrm{bel}_{12} = \mathrm{bel}_1 \oplus \mathrm{bel}_2$
conjunctive rule of combination:	$\textcircled{\scriptsize$\wedge$}$	$\mathrm{bel}_{12} = \mathrm{bel}_1 \textcircled{\scriptsize$\wedge$} \mathrm{bel}_2$
disjunctive rule of combination:	$\textcircled{\scriptsize$\vee$}$	$\mathrm{bel}_{12} = \mathrm{bel}_1 \textcircled{\scriptsize$\vee$} \mathrm{bel}_2$

The difference between the elements of the two pairs $(\mid, \vdots)$ and $(\oplus, \textcircled{\scriptsize$\wedge$})$ results only from the normalization factors applied in $\mid$ and $\oplus$. $\textcircled{\scriptsize$\vee$}$ does not have a specific counterpart in Shafer's presentation (indeed once bel_1 and bel_2 are normalized, $\mathrm{bel}_1 \textcircled{\scriptsize$\vee$} \mathrm{bel}_2$ is also normalized). Note that $\mathrm{bel}(.\,\vdots\,B)$ could be a normalized belief function. In fact $\vdots$ is a generalization of $\mid$.

REMARK 2: NOTATION Given two spaces Θ and X, we write $\text{bel}_X(.\mid\theta)$ and $\text{pl}_X(.\mid\theta)$ to represent the belief and plausibility functions induced on space X in a context where $\theta \subseteq \Theta$ is the case, and $\text{bel}_{Xx\Theta}, \text{pl}_{Xx\Theta}$ to represent belief and plausibility functions on the space $Xx\Theta$. We write $x \cap \theta$ as a shorthand for the intersection of the cylindrical extensions of $x \subseteq X$ and $\theta \subseteq \Theta$ over the product space $Xx\Theta$ (i.e., $x \cap \theta$ means $\text{cyl}(x) \cap \text{cyl}(\theta)$). Similarly $x \cup \theta$ means $\text{cyl}(x) \cup \text{cyl}(\theta)\ldots$

Subscripts of bel and pl represent their domain and are omitted when there is no ambiguity as in $\text{bel}(x\mid\theta), \text{bel}(\theta), \ldots$.

REMARK 3: Our notation will not distinguish between elements like θ_i where $\theta_i \in \Theta$ and their corresponding singleton $\{\theta_i\} \subseteq \Theta$. The context should always make clear which is intended, and the notation is seriously lightened.

The following lemmas will be useful:

LEMMA 1: *If* $pl: 2^\Omega \to [0,1]$ *is a plausibility function, then the corresponding commonality function* q *is* $q(A) = \Sigma_{B \subseteq A}(-1)^{|B|+1} pl(B)$.

Proof immediate by replacing $\text{bel}(\bar{B})$ by $\text{pl}(\Omega) - \text{pl}(B)$ in the relation between q and bel given in Shafer [1, p. 41]. QED

LEMMA 2: $\forall x \subseteq X, \forall \theta \subseteq \Theta, \forall \theta_i \in \theta$: $pl(x\mid\theta) \geq pl(x\mid\theta_i)$.

Proof Let $\text{cyl}(x)$ and $\text{cyl}(\theta)$ be the cylindrical extensions of x and θ on the space $Xx\Theta$. Then $\text{pl}_X(x\mid\theta) = \text{pl}_{Xx\Theta}(\text{cyl}(x)\mid\text{cyl}(\theta)) = \text{pl}_{Xx\Theta}(\text{cyl}(x) \cap \text{cyl}(\theta)) \geq \text{pl}_{Xx\Theta}(\text{cyl}(x) \cap \text{cyl}(\theta_k)) = \text{pl}_X(x\mid\theta_k)$ where $\theta_k \in \theta$. QED

2.2. The Principle of Minimal Commitment

We introduce the *principle of minimal commitment*. Given a belief function derived on Ω, this principle induces the construction of new belief functions: 1) on **refined spaces Ω'** where every element of Ω is split into several elements of Ω' and 2) on **extended spaces Ω''**, where Ω'' contains all the elements of Ω and some new elements. These two processes are called the **vacuous extension** and the **ballooning extension**, respectively. In this paper, the vacuous extension transforms a belief function over Θ into a belief function over $Xx\Theta$ and the ballooning extension transforms a conditional belief function $\text{bel}_X(.\mid\theta_i)$ defined on X for $\theta_i \in \Theta$ into a new belief function over $Xx\Theta$.

In order to understand the principle of minimal commitment, we must consider the meaning of $\text{bel}(A)$ and $\text{pl}(A)$. Within the transferable belief model, the degree of belief $\text{bel}(A)$ given to a subset A quantifies the amount of *justified specific support* to be given to A, and the degree of plausibility pl(A) given to a subset A quantifies the maximum amount of *potential specific support* that could be given to A.

$$\text{bel}(A) = \sum_{\emptyset \neq X \subseteq \Omega} m(X) \qquad \text{pl}(A) = \sum_{A \cap X \neq 0} m(X) = \text{bel}(\Omega) - \text{bel}(\bar{A}).$$

We say *specific* because $m(\emptyset)$ is neither included in bel(A) nor in pl(A). The bbm's $m(X)$ included in bel(A) are only those given to the subsets of A that are not subsets of $\bar{A}$. $m(\emptyset)$ is not included because $\emptyset$ is a subset of both A and $\bar{A}$.

We say *justified* because we include in bel(A) *only* the bbm's given to subsets of A. For instance, consider two distinct elements x and y of Ω. The bbm $m(\{x, y\})$ given to $\{x, y\}$ could support x if further information indicates this. However given the available information the bbm can only be given to $\{x, y\}$.

We say *potential* because the bbm included in pl(A) could be transferred to non-empty subsets of A if some new information could justify such a transfer. It would be the case if we learn that $\bar{A}$ is impossible. After conditioning on A, note that bel($A|A$) = pl(A). Large plausibilities given to all subsets reflect the lack of commitment of our belief; we are ready to give a large belief to *any* subset.

Consider now the case where there is ambiguity about the amount of plausibility that should be given to the subsets of Ω. The ambiguity could be resolved by giving the largest possible plausibility to every subsets.

The principle of minimal commitment formalizes this idea: one should never give more support than justified to any subset of Ω. It satisfies a form of skepticism, noncommitment, or conservatism in the allocation of belief. In spirit, it is not far from what probabilists attempt to achieve with the maximum entropy principle. The concept of commitment was already introduced to create an ordering on the set of belief functions defined on a frame of discernment Ω (see Moral [10], Yager [23], Dubois and Prade [11, 24], Delgado and Moral [25], Kruse and Schwecke [26], Hsia [27].

To define the principle, let pl_1 and pl_2 be two plausibility functions on Ω such that:

$$\text{pl}_1(A) \leq \text{pl}_2(A) \qquad \forall A \subseteq \Omega. \tag{2.4}$$

We say that pl_2 is **no more committed** than pl_1 (and less committed if there is at least one strict inequality). The same qualification is extended to the related bba and belief functions. The least committed belief function is the vacuous belief function ($m(\Omega) = 1$). The most committed belief function is the contradictory belief function ($m(\emptyset) = 1$).

The **principle of minimal commitment** indicates that, given two equally supported beliefs, only one of which can apply, the most appropriate is the least committed.

For unnormalized belief functions, the principle is based on the plausibility function. The inequalities 2.4 expressed in terms of belief functions become:

$$\text{bel}_1(A) + m_1(\emptyset) \geq \text{bel}_2(A) + m_2(\emptyset) \qquad \forall A \subseteq \Omega. \tag{2.5}$$

To define the principle by requiring that:

$$\text{bel}_1(A) \geq \text{bel}_2(A) \qquad \forall A \subseteq \Omega \tag{2.6}$$

is inappropriate as seen in the following example. Let:

$$\text{bel}_1(A) = 0 \; \forall A \neq \Omega, \quad \text{and} \quad \text{bel}_1(\Omega) = .7$$

If bel_2 is a vacuous belief function, it is less committed than bel_1. It is not the case that $\text{bel}_2(A) \leq \text{bel}_1(A)\ \forall \text{A} \subseteq \Omega$. However, one has $\text{pl}_1(A) = .7 \leq \text{pl}_2(A) = 1\ \forall A \subseteq \Omega$ as required.

Under the closed-world assumption, the principle can be similarly defined with plausibility inequalities 2.4 or belief function inequalities 2.6. The last definition is historically the oldest. This explains why we maintain the "minimal commitment" name even though it could be argued that the principle would be better named the principle of "maximal plausibility" or "maximal skepticism."

The principle of minimal commitment is not used to derive the DRC and the GBT in section 3. However during the constructive derivations of the GBT in section 4, we will encounter plausibility functions pl whose values are known only for a set $\mathcal{F}$ of subsets of Ω. In most cases, one can build a plausibility function pl* such that $\text{pl}^*(A) = \text{pl}(A)\ \forall A \in \mathcal{F}$ and pl* is nevertheless known everywhere on Ω. This is achieved by committing the largest possible plausibility to every subset of Ω that is not an element of $\mathcal{F}$. This application of the principle of minimal commitment is translated into the following property.

THE PRINCIPLE OF MINIMAL COMMITMENT FOR PARTIALLY DEFINED PLAUSIBILITY FUNCTIONS

Let $\mathcal{F}$ be a set of subsets of a frame of discernment Ω, and let pl be a plausibility function whose value is known only for those subsets of Ω in $\mathcal{F}$. Let $\mathcal{P}$ be the set of all the plausibility functions pl′ on Ω such that $\text{pl}'(A) = \text{pl}(A)$ for all A in $\mathcal{F}$. The maximal element pl* of $\mathcal{P}$, when it exists, is the plausibility function pl* such that $\forall \text{pl}'$ in $\mathcal{P}$: $\text{pl}^*(\text{B}) \geq \text{pl}'(\text{B})\ \forall \text{B} \subseteq \Omega$.

Two special cases of the principle will be used here: the vacuous extension and the"ballooning" extension.

1. Let Ω be a frame of discernment and let pl be defined for every subset of Ω. Let Ω' be a refinement[2] R of Ω. The plausibility function pl′ on Ω' induced by pl that satisfies the principle of minimal commitment is the **vacuous extension** of pl on Ω via R. Its bbms are defined as follows (Shafer [1, p. 146] et seq.). Let m and m' be the bba underlying pl and pl′. Then $m'(R(A)) = m(A)$, $\forall A \subseteq \Omega$, and $m'(B) = 0$ otherwise.
2. Let Θ and X be two finite spaces, $\text{bel}_X(.\,|\,\theta)$ be a conditional belief function on X given some $\theta \in \Theta$ and $\mathscr{B}\text{el}^*$ be the set of belief functions $\text{bel}_{Xx\Theta}$ over space $Xx\Theta$ such that their conditioning given θ is equal to $\text{bel}_X(.\,|\,\theta)$. The element of $\mathscr{B}\text{el}^*$ that satisfies the principle of minimal commitment is the belief function $\text{bel}^*_{Xx\Theta}$ such that:

$$\text{bel}^*_{Xx\Theta}((\text{cyl}(x) \cap \text{cyl}(\theta)) \cup \text{cyl}(\bar{\theta})) - \text{bel}^*_{Xx\Theta}(\text{cyl}(\bar{\theta})) = \text{bel}_X(x\,|\,\theta)$$

where $\text{cyl}(x)$ and $\text{cyl}(\theta)$ are the cylindrical extensions of x and θ on the space $Xx\Theta$, and $\text{bel}^*_{Xx\Theta}(\text{cyl}(\bar{\theta})) = M_X(\emptyset\theta)$

It can be informally rewritten as:

$$\text{bel}^*_{Xx\Theta}(x \cup \bar{\theta}) = \text{bel}_X(x\,|\,\theta) + M_X(\emptyset\,|\,\theta)$$

We call this transformation between bel and bel* the **deconditionalization process** (Smets [6]). bel* is called the **"ballooning extension"** of $\text{bel}(x\,|\,\theta)$ on $Xx\Theta$ as each mass $m(x\,|\,\theta)$ is given after deconditionalization to the

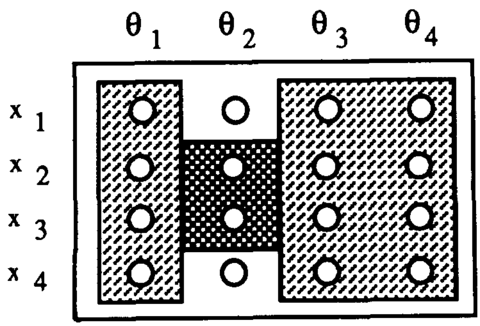

Figure 1. Ballooning of the bbm $m(x_2 \cup x_3\,|\,\theta_2)$ (dark area) onto $Xx\Theta$ (shaded area). The white dots correspond to the 16 elements of $Xx\Theta$.

[2] The mapping R from Ω to Ω' is a refinement if every element of Ω is mapped by R into one or more elements of Ω' and the images $R(\omega)$ of the elements ω of Ω under the refinement R partition Ω'.

largest subset of $Xx\Theta$ so that its intersection with cyl(θ) is the set cyl(x) $\cap$ cyl(θ) (Shafer [9]) called bel* the "conditional embedding" of bel($x\,|\,\theta$)). (Note the similarity between this ballooning extension and the passage from a conjunction cyl(x) $\cap$ cyl(θ) to a material implication cyl(x) $\rightarrow$ cyl(θ).)

2.3. Conditional Cognitive Independence

In our derivation of the GBT and the DRC, we need to determine the belief induced by two "independent" observations given the belief induced by each observation. The concept of "independence" is defined as follows. Let X and Y be two spaces from which we collect observations (pieces of evidence). The two variables X and Y are said to be "independent" if *the knowledge of the particular value taken by one of them does not change our belief about the value that the second could take*, i.e., $\mathrm{bel}_X(A\,|\,y) = \mathrm{bel}_X(A\,|\,y')$ $\forall A \subseteq X$, $\forall y$, $y' \in Y$, $y \neq y'$ and $\mathrm{bel}_Y(B\,|\,x) = \mathrm{bel}_Y(B\,|\,x')$ $\forall B \subseteq Y$, $\forall x$, $x' \in X$, $x \neq x'$.

We use this concept of independent observations in order to derive the DRC and the GBT as we claim that two independent observations induce two belief functions that can be combined by the conjunctive rule of combination. More specifically, suppose a set $\Theta = \{\theta_i\colon\ i = 1\ldots n\}$ of contexts θ_i. Suppose we collect two observations that are independent whatever the context θ_i. Such two observations are said to be conditionally independent. Each observation induces a belief on Θ and constitutes thus a piece of evidence relative to Θ. We claim that two observations that are conditionally independent constitute two pieces of evidence relative to Θ that are distinct. The satisfaction of that claim was often asked for—it motivated the development of the GBT in Smets [6]—and authors complain of its non-satisfaction by other attempts to define an equivalent of the GBT (e.g., see Halpern and Fagin [28]).

Once that claim is admitted, the properties underlying the concept of cognitive independence, detailed here below, are deduced as a spin-off of the DRC. But in fact the concept of independent observations is already sufficient to deduce the properties underlying the concept of cognitive independence within the TBM, therefore without regard to the DRC and the GBT.

In the transferable belief model framework, the concept of two independent variables X and Y translates as follows: the ratio of the plausibilities on X should not depend on $y \subseteq Y$:

$$\frac{\mathrm{pl}_X(x_1\,|\,y)}{\mathrm{pl}_X(x_2\,|\,y)} = \frac{\mathrm{pl}_X(x_1)}{\mathrm{pl}_X(x_2)} \qquad \forall x_1, x_2 \subseteq X, \forall y \subseteq Y. \tag{2.7}$$

As $\text{pl}_X(x\,|\,y) \equiv \text{pl}_{XxY}(x \cap y)$, the independence requirement becomes:

$$\frac{\text{pl}_{XxY}(x_1 \cap y)}{\text{pl}_{XxY}(x_2 \cap y)} = \frac{\text{pl}_X(x_1)}{\text{pl}_X(x_2)} \qquad \forall x_1, x_2 \subseteq X, \forall y \subseteq Y.$$

These ratio constraints imply that (the proof is given under lemma 3 in the appendix):

$$\text{pl}_{XxY}(x \cap y) = \text{pl}_X(x)\text{pl}_Y(y) \qquad \forall x \subseteq X, \forall y \subseteq Y. \tag{2.8}$$

Two variables (X and Y) that satisfy this requirement are said to satisfy the **cognitive independence** property. This definition was introduced in Shafer [1, p. 150]. It extends the classical stochastic independence.

The cognitive independence concept can be extended in a straightforward manner when the plausibility functions are conditional plausibility functions. If the two variables X and Y are independent in each context θ_i, for all $\theta_i \in \Theta$, then they satisfy the **conditional cognitive independence** (CCI) property if:

$$\text{pl}_{XxY}(x \cap y\,|\,\theta_i) = \text{pl}_X(x\,|\,\theta_i)\text{pl}_Y(y\,|\,\theta_i) \qquad \forall x \subseteq X, \forall y \subseteq Y, \forall \theta_i \in \Theta \tag{2.9}$$

The previous independence definitions are based on plausibility functions. They could have been based as well on belief functions. Two variables X and Y are CCI iff the ratio of their belief functions satisfy the dual of (2.7)

$$\frac{\text{bel}_X(x_1\,|\,y)}{\text{bel}_X(x_2\,|\,y)} = \frac{\text{bel}_X(\pi_1)}{\text{bel}_X(\pi_2)} \qquad \forall x_1, x_2 \subseteq X, \forall y \subseteq Y. \tag{2.10}$$

In fact, both definitions are equivalent as (2.7) is equivalent to (2.10). A proof is given in the appendix (see lemma 4).

2.4. The Generalized Likelihood Principle

In order to derive the DRC and the GBT, we need to generalize the likelihood principle within the transferable belief model. It simply postulates that the belief function induced by the disjunction of two pieces of evidence is only a function of the belief functions induced by each piece of evidence. We will build $\text{pl}_X(.\,|\,\theta)$ on X for any subset θ of Θ, even though we only know the conditional plausibility functions $\text{pl}_X(.\,|\,\theta_i)$ over $X, \forall \theta_i \in \Theta$.

To help in understanding the principle, we present the likelihood principle as described in probability theory. The likelihood $1(\theta_i|x)$ (sometimes called the relative plausibility) of the "simple" hypothesis $\theta_i, \forall \theta_i \in \Theta$, given the data $x \subseteq X$ is defined as being equal to the conditional probability $p(x|\theta_i)$ of the data x given the simple hypothesis θ_i (Edwards [29])

$$1(\theta_i|x) = p(x|\theta_i)$$

The likelihood of the disjunction $\theta \subseteq \Theta$ of several simple hypotheses $\theta_i, i = 1, 2 \dots k$ where $\theta = \{\theta_1 \cup \theta_2 \cup \dots \cup \theta_k\}$ is defined as a function of the likelihoods of the simple hypothesis $\theta_i \in \theta$:

$$1(\theta|x) = f(\{1(\theta_i|x) \colon \theta_i \in \theta\})$$

where f is the maximum operator ($f(a, b, \dots) = \max(a, b, \dots)$). The link between the likelihood functions extended to disjunction of hypothesis and possibility functions (Zadeh [30], Dubois and Prade [31]) was shown in Smets [32].

A form of this principle was already proposed in Shafer [1, p. 239] when he studied statistical inference in the context of belief functions. He proposed to define $pl(\theta|x) = \max_{\theta_i \in \theta} pl(\theta_i|x)$. This solution is not satisfactory for statistical inference, as it does not satisfy Requirement R1 in section 3, a requirement for which satisfaction is often asked (Smets [6], Halpern and Fagin [28]).

The likelihood principle is defined for probability functions. We broaden it into the **generalized likelihood principle** applicable to plausibility function within the transferable belief model:

$$\forall \theta \subseteq \Theta, \forall x \subseteq X, pl(x \,\vdots\, \theta) \text{ depends only on } \{pl(x \,\vdots\, \theta_i), pl(\bar{x} \,\vdots\, \theta_i) \colon \theta_i \in \theta\}.$$

The maximum operator is not assumed. The need of both $pl(x \,\vdots\, \theta_i)$ and $pl(\bar{x} \,\vdots\, \theta_i)$ reflects the non-additivity of the plausibility functions.

The origin of the principle can be justified by requiring that:

1. $pl(x \,\vdots\, \theta)$ is the same after the frame X has been transformed by coarsening into the frame with only two elements: x and $\bar{x}$. This explains why only those values of $pl(. \,\vdots\, \theta_i)$ for x and $\bar{x}$ are used.
2. the values of $pl(x \,\vdots\, \theta_j)$ for $\theta_j \in \theta$ are irrelevant to the values of $pl(x \,\vdots\, \theta)$. Hence only the $\theta_i \in \theta$ are used.

3. THE DISJUNCTIVE RULE OF COMBINATION AND THE GENERALIZED BAYESIAN THEOREM

We proceed with the derivation of the DRC and the GBT. Let X and Θ be two finite non-empty sets. Suppose all we know about X is represented initially by the set $\{bel_X(. \,\vdots\, \theta_i) \colon \theta_i \in \Theta\}$ of belief functions $bel_X(. \,\vdots\, \theta_i)$ on X. We only know the beliefs on X when we know which element of Θ holds. We do not know these beliefs on X when we only know that the prevailing element of Θ belongs to a given subset θ of Θ. The DRC permits to build the belief function $bel_X(. \,\vdots\, \theta)$ on X for any $\theta \subseteq \Theta$.

Simultaneously we derive the GBT that permits to build $\text{bel}_\Theta(.\|x)$ for any $x \subseteq X$ from the conditional belief functions $\text{bel}_X(.\|\theta_i)$, as the DRC and the GBT are linked through the relation:

$$\text{pl}_X(x\|\theta) = \text{pl}_\Theta(\theta\|x), \qquad \forall\theta \subseteq \Theta, \forall x \subseteq X.$$

The derivation of the DRC and the GBT is based on the following ideas. Let X and Y be two frames of discernment. For each $\theta_i \in \Theta$, let $\text{bel}_X(.\|\theta_i)$ quantify our belief on X given θ_i, and $\text{bel}_Y(.\|\theta_i)$ quantify our belief on Y given θ_i. θ_i can be interpreted as a context. We assume there is no other knowledge about X and Y except these conditional belief functions on X and Y known for each $\theta_i \in \Theta$. It implies among others that we do not have any *a priori* belief on Θ, i.e., we have the vacuous *a priori* belief function bel_0 on Θ (this condition will be relaxed in section 5).

Suppose we learn then that $x_0 \subseteq X$ holds. What is the belief function $\text{bel}_\Theta(.\|x_0)$ on Θ induced by the knowledge of the conditional belief functions $\text{bel}_X(.\|\theta_i)$ $\forall\theta_i \in \Theta$ and of the fact that x_0 holds? As we assume that every state of knowledge induces a unique belief on any variable, the belief function $\text{bel}_\Theta(.\|x_0)$ on Θ exists and is unique. Hence $\text{bel}_\Theta(.\|x_0)$ is a function F of x_0 and the $\text{bel}_X(.\|\theta_i)$ for $\theta_i \in \Theta$:

$$\text{bel}_\Theta(.\|x_0) = F(x_0, \{\text{bel}_X(.\|\theta_i): \theta_i \in \Theta\})$$

Similarly if we learn that $y_0 \subseteq Y$ holds, the belief function $\text{bel}_\Theta(.\|y_0)$ on Θ is a function F of Y_0 and the $\text{bel}_Y(.\|\theta_i)$ for $\theta_i \in \Theta$:

$$\text{bel}_\Theta(.\|y_0) = F(y_0, \{\text{bel}_Y(.\|\theta_i): \theta_i \in \Theta\})$$

Finally, if we learn that the joint observation $(x_0, y_0) \subseteq XxY$, $x_0 \subseteq X$, $y_0 \subseteq Y$, is the case, we could build the belief function $\text{bel}_\Theta(.\|x_0, y_0)$ on Θ based on (x_0, y_0) if we knew the conditional belief functions $\text{bel}_{XxY}(.\|\theta_i)$ for $\theta_i \in \Theta$:

$$\text{bel}_\Theta(.\|x_0, y_0) = F((x_0, y_0), \{\text{bel}_{XxY}(.\|\theta_i): \theta_i \in \Theta\})$$

Suppose the observations $x_0 \subseteq X$ and $y_0 \subseteq Y$ are conditionally independent whatever context $\theta_i \in \Theta$ holds. The conditional independence of X and Y implies that the observations x_0 and y_0 are two distinct pieces of evidence relative to Θ. Each piece of evidence induces a belief on Θ: $\text{bel}_\Theta(.\|x_0)$ and $\text{bel}_\Theta(.\|y_0)$. The belief $\text{bel}_\Theta(.\|x_0, y_0)$ that x_0 and y_0 jointly induce on Θ can be obtained by the conjunctive rule of combination: $\text{bel}_\Theta(.\|x_0, y_0) = \text{bel}_\Theta(.\|x_0) \textcircled{\wedge} \text{bel}_\Theta(.\|y_0)$.

In Requirement R, we ask that the belief function $\text{bel}_\Theta(.\|x_0, y_0)$ induced on Θ by two pieces of evidence x_0 and y_0 that correspond to two independent observations $x_0 \subseteq X$ and $y_0 \subseteq Y$ is the same as the belief function $\text{bel}_\Theta(.\,|\,x_0) \bigcirc\!\!\!\!\wedge\ \text{bel}_\Theta(.\,|\,y_0)$ on Θ computed by the conjunctive combination of the individual belief functions $\text{bel}_\Theta(.\,|\,x_0)$ and $\text{bel}_\Theta(.\,|\,y_0)$. We also ask that $\text{pl}_X(.\,|\,\theta)$, $\text{pl}_Y(.\,|\,\theta)$ and $\text{pl}_{XxY}(.\,|\,\theta)$, $\theta \subseteq \Theta$, satisfy the generalized likelihood principle.

REQUIREMENT R

Given:

three frames of discernment X, Y and Θ.

our knowledge on X, Y and Θ is represented by $\text{bel}_X(.\,|\,\theta_i)$ and $\text{bel}_Y(.\,|\,\theta_i)$ $\forall \theta_i \in \Theta$.

X and Y are conditionally independent given $\theta_i, \forall \theta_i \in \Theta$

$\forall x \subseteq X$ and $\forall y \subseteq Y$, there is a function F such that

$$\text{bel}_\Theta(.\,|\,x) = F(x, \{\text{bel}_X(.\,|\,\theta_i) \colon \theta_i \in \Theta\})$$

$$\text{bel}_\Theta(.\,|\,y) = F(y, \{\text{bel}_Y(.\,|\,\theta_i) \colon \theta_i \in \Theta\})$$

$$\text{bel}_\Theta(.\,|\,x, y) = F((x, y), \{\text{bel}_{XxY}(.\,|\,\theta_i) \colon \theta_i \in \Theta\})$$

Then:

Requirement R1:

$$\text{bel}_\Theta(.\,|\,x, y) = \text{bel}_\Theta(.\,|\,x) \bigcirc\!\!\!\!\wedge\ \text{bel}_\Theta(.\,|\,y)$$

Requirement R2:

$$\text{pl}_X(x\,|\,\theta) = g(\{\text{pl}_X(x\,|\,\theta_i), \text{pl}_X(\bar{x}\,|\,\theta_i) \colon \theta_i \in \theta\}) \qquad \forall x \subseteq X, \forall \theta \subseteq \Theta$$

$$\text{pl}_Y(y\,|\,\theta) = g(\{\text{pl}_Y(y\,|\,\theta_i), \text{pl}_Y(\bar{y}\,|\,\theta_i) \colon \theta_i \in \theta\}) \qquad \forall y \subseteq Y, \forall \theta \subseteq \Theta$$

$$\text{pl}_{XxY}(w\,|\,\theta) = g(\{\text{pl}_{XxY}(w\,|\,\theta_i), \text{pl}_{XxY}(\bar{w}\,|\,\theta_i) \colon \theta_i \in \theta\}) \qquad \forall w \subseteq XxY, \forall \theta \subseteq \Theta.$$

The functions F and g will be deduced from Requirement R in Theorems 1 to 4. This allows us to build:

1. $\text{bel}_X(.\,|\,\theta)$ and $\text{bel}_Y(.\,|\,\theta)$, $\theta \subseteq \Theta$, (the DRC)
2. $\text{bel}_\Theta(.\,|\,x)$ and $\text{bel}_\Theta(.\,|\,y)$, (the GBT) and
3. $\text{bel}_{XxY}(.\,|\,\theta)$, $\theta \subseteq \Theta$, (the CCI),

from the set of conditional belief functions $\text{bel}_X(.\,|\,\theta_i)$, and $\text{bel}_Y(.\,|\,\theta_i)$, $\theta_i \in \Theta$.

The derivation of the DRC and the GBT are presented successively, first when the belief functions $\text{bel}_X(.|\theta_i)$, and $\text{bel}_Y(.|\theta_i)$, $\theta_i \in \Theta$ are normalized (i.e., $\text{bel}_X(X|\theta_i) = 1$ and $\text{bel}_Y(Y|\theta_i) = 1$), then when they are not. The CCI is a by-product of the DRC derivation. All proofs are given in the appendix. We present only the formulas for $\text{bel}_X(.\,|\,\theta)$, $\theta \subseteq \Theta$ and $\text{bel}_\Theta(.\,|\,x)$, $x \subseteq X$, (and their related pl, m and q functions). The same formulas can be written for $\text{bel}_Y(.\,|\,\theta)$, $\theta \subseteq \Theta$ and $\text{bel}_\Theta(.\,|\,y)$, $y \subseteq Y$.

THEOREM 1. THE DISJUNCTIVE RULE OF COMBINATION, *normalized beliefs. Given the Requirement R and its antecedents. Given* $bel_X(X|\theta_i) = 1$ *and* $bel_Y(Y|\theta_i) = 1, \forall\theta_i \in \Theta$. *Then* $\forall\theta \subseteq \Theta, \forall x \subseteq X$;

$$\mathrm{bel}_X(x|\theta) = \mathrm{bel}_X(x\dot{:}\theta) = \prod_{\theta_i \in \theta} \mathrm{bel}_X(x|\theta_i) \tag{3.1}$$

$$\mathrm{pl}_X(x|\theta) = \mathrm{pl}_X(x\dot{:}\theta) = 1 - \prod_{\theta_i \in \theta} (1 - \mathrm{pl}_X(x|\theta_i)) \tag{3.2}$$

$$m_X(x|\theta) = m_X(x\dot{:}\theta) = \sum_{\cup_{i:\,\theta_i \in \theta} x_i = x} \ \prod_{i:\,\theta_i \in \theta} m_X(x_i|\theta_i) \tag{3.3}$$

The relation 3.3 shows the dual nature of the conjunctive and disjunctive rules of combination (Dubois and Prade [11]). Suppose two belief functions with their basic belief assignments m_1 and m_2 on Ω. When combined, the product $m_1(A)m_2(B), A \subseteq \Omega, B \subseteq \Omega$, is allocated to $A \cap B$ in the conjunctive rule of combination, and to $A \cup B$ in the disjunctive rule of combination. One has $\forall C \subseteq \Omega$:

1. conjunctive rule of combination (CRD)

$$m_1 \textcircled{\wedge} m_2(C) = \sum_{A \cap B = C} m_1(A)m_2(B)$$

$$q_1 \textcircled{\wedge} q_2(C) = q_1(C)q_2(C)$$

2. disjunctive rule of combination (DRC)

$$m_1 \textcircled{\vee} m_2(C) = \sum_{A \cup B = C} m_1(A)m_2(B)$$

$$\mathrm{bel}_1 \textcircled{\vee} \mathrm{bel}_2(C) = \mathrm{bel}_1(C)\mathrm{bel}_2(C)$$

The $\cap$ and $\cup$ operators encountered in the relations for the basic belief assignments explain the origin of the symbols $\textcircled{\wedge}$ and $\textcircled{\vee}$. These relations shows also the dual role of bel and q. Indeed bel(C) is the sum of the basic belief masses given to the subsets of C and q(C) as the sum of the basic belief masses given to the supersets of C (beware of the comments after theorem 3).

Once the DRC is known, the GBT is derived thanks to the rclation:

$$\mathrm{pl}_\theta(\theta\dot{:}x) = \mathrm{pl}_X(x\dot{:}\theta) \qquad \forall\theta \subseteq \Theta, \forall x \subseteq X.$$

as confirmed by the equality between 3.2 and 3.5.

THEOREM 2. THE GENERALIZED BAYESIAN THEOREM, *normalized beliefs. Given the Requirement R and its antecedents. Given* $bel_X(X|\theta_i) = 1$ *and* $bel_Y(Y|\theta_i) = 1$, $\forall\theta_i \in \Theta$. *Then* $\forall\theta \subseteq \Theta$, $\forall x \subseteq X$;

$$\mathrm{bel}_\Theta(\theta \mathrel{\vdots} x) = \prod_{\theta_i \in \bar{\theta}} \mathrm{bel}_X(\bar{x}|\theta_i) - \prod_{\theta_i \in \Theta} \mathrm{bel}_X(\bar{x}|\theta_i) \tag{3.4}$$

$$\mathrm{bel}_\Theta(\theta|x) = K.\mathrm{bel}_\Theta(\theta \mathrel{\vdots} x)$$

$$\mathrm{pl}_\Theta(\theta \mathrel{\vdots} x) = 1 - \prod_{\theta_i \in \theta} (1 - \mathrm{pl}_X(x|\theta_i)) \tag{3.5}$$

$$\mathrm{pl}_\Theta(\theta|x) = K.\mathrm{pl}_\Theta(\theta \mathrel{\vdots} x)$$

$$q_\Theta(\theta \mathrel{\vdots} x) = \prod_{\theta_i \in \Theta} pl_X(x|\theta_i) \tag{3.6}$$

$$q_\Theta(\theta|x) = K.q_\Theta(\theta \mathrel{\vdots} x)$$

where $$K^{-1} = 1 - \prod_{\theta_i \in \Theta} \mathrm{bel}_X(\bar{x}|\theta_i) = 1 - \prod_{\theta_i \in \Theta} (1 - \mathrm{pl}_X(x|\theta_i))$$

As announced the CCI is derived as a by-product of the DRC. Note that 3.2 and 3.5 are identical, reflecting the equality between $\mathrm{pl}_X(x \mathrel{\vdots} \theta)$ and $\mathrm{pl}_\Theta(\theta \mathrel{\vdots} x)$.

LEMMA 5. THE CONDITIONAL COGNITIVE INDEPENDENCE *Under theorem 1 conditions,:*

$$\mathrm{pl}_{XxY}(x \cap y \mathrel{\vdots} \theta_i)) = \mathrm{pl}_X(x \mathrel{\vdots} \theta_i)\mathrm{pl}_Y(y \mathrel{\vdots} \theta_i) \qquad \forall x \subseteq X, \forall y \subseteq Y, \theta_i \in \Theta.$$

We proceed with the derivation of the DRC and the GBT when the initial conditional belief functions are not normalized. Given a belief function bel: $2^\Omega \to [0,1]$, we define a function $b: 2^\Omega \to [0,1]$ such that $b(A) = \mathrm{bel}(A) + m(\emptyset)$. This b function is the real dual of the commonality function q. The real difference between theorems 1–2 and 3–4 concerns the computation of $\mathrm{bel}_X(x \mathrel{\vdots} \theta)$ and $\mathrm{bel}_\Theta(\theta \mathrel{\vdots} x)$.

THEOREM 3. THE DISJUNCTIVE RULE OF COMBINATION, *general case. Given the Requirement R and its antecedents. Then* $\forall\theta \subseteq \Theta$, $\forall x \subseteq X$;

$$b_X(x \mathrel{\vdots} \theta) = \prod_{\theta_i \in \theta} b_X(x \mathrel{\vdots} \theta_i) \tag{3.7}$$

$$\mathrm{bel}_X(x \mathrel{\vdots} \theta) = b_X(x \mathrel{\vdots} \theta) - b_X(\emptyset \mathrel{\vdots} \theta) \tag{3.8}$$

$$\mathrm{pl}_X(x \mathrel{\vdots} \theta) = 1 - \prod_{\theta_i \in \theta} (1 - \mathrm{pl}_X(x \mathrel{\vdots} \theta_i)) \tag{3.9}$$

$$m_X(x \mathrel{\vdots} \theta) = \sum_{\cup_{i:\, \theta_i \in \theta} x_i = x} \;\prod_{i:\, \theta \in \Theta} m_X(x_i|\theta_i) \tag{3.10}$$

The real dual of q is b, not bel: indeed in the disjunctive rule of combination one multiplies the b functions, not the bel functions. b(C) is the sum of the basic belief masses given to the subsets of C, including $\varnothing$. Another way to see the dual nature of the DRC and CRC consists in building the "complementary" basic belief assignment $\overline{\mathrm{m}}: 2^{\Omega} \to [0,1]$ of a basic belief assignment $\mathrm{m}: 2^{\Omega} \to [0,1]$ with $\overline{\mathrm{m}}(A) = \mathrm{m}(\overline{A})$ for every $A \subseteq \Omega$. Then $\overline{\mathrm{b}}(A) = \mathrm{q}(\overline{A})$ (Dubois and Prade [11]).

THEOREM 4. THE GENERALIZED BAYESIAN THEOREM, *general case. Given the Requirement R and its antecedents. Then* $\forall \theta \subseteq \Theta$, $\forall x \subseteq X$;

$$b_{\Theta}(\theta \,|\, x) = \prod_{\theta_i \in \overline{\theta}} b_X(\bar{x} \,|\, \theta_i)$$

$$\mathrm{bel}_{\Theta}(\theta \,|\, x) = b_{\Theta}(\theta \,|\, x) - b_{\Theta}(\varnothing \,|\, x) \tag{3.11}$$

$$\mathrm{pl}_{\Theta}(\theta \,|\, x) = 1 - \prod_{\theta_i \in \theta} (1 - \mathrm{pl}_X(x \,|\, \theta_i)) \tag{3.12}$$

$$q_{\Theta}(\theta \,|\, x) = \prod_{\theta_i \in \theta} \mathrm{pl}_X(x \,|\, \theta_i) \tag{3.13}$$

4. CONSTRUCTIVE DERIVATIONS OF THEOREMS 3 AND 4 (Smets [6])

In theorems 3 and 4 we derive the DRC and the GBT from general principles. These relations can also be obtained in a constructive way by the application of the principle of minimal commitment. We present three different ways to derive both the DRC and the GBT. These constructions help in understanding the nature of the solutions.

4.1.

For each $\theta_i \in \Theta$, build the ballooning extension $\mathrm{bel}^{(i)}_{X x \Theta}$ of $\mathrm{bel}_X(.\,|\,\theta_i)$ on $X x \Theta$. Combine these belief functions $\mathrm{bel}^{(i)}_{X x \Theta}$ by the conjunctive rule of combination. Let $\mathrm{bel}_{X x \Theta} = \mathrm{bel}^{(1)}_{X x \Theta} \textcircled{\scriptsize$\wedge$} \mathrm{bel}^{(2)}_{X x \Theta} \textcircled{\scriptsize$\wedge$} \ldots \textcircled{\scriptsize$\wedge$} \mathrm{bel}^{(n)}_{X x \Theta}$ be the resulting belief function on $X x \Theta$. Let $\omega \subseteq X x \Theta$ and let x_i be the projection of $\omega \cap \mathrm{cyl}(\theta_i)$ on X. Then

$$\mathrm{bel}_{X x \theta}(\omega) = \prod_{\theta_i \in \Theta} b_X(x_i \,|\, \theta_i) - \prod_{\theta_i \in \Theta} b_X(\varnothing \,|\, \theta_i)$$

$$m_{X x \Theta}(\omega) = \prod_{\theta_i \in \Theta} m_X(x_i \,|\, \theta_i)$$

$$q_{X x \Theta}(\omega) = \prod_{\theta_i \in \Theta} q_X(x_i \,|\, \theta_i)$$

(all proofs are given in Smets [6, p. 163] et seq.)

The relations of Theorems 3 and 4 are obtained by conditioning $\mathrm{bel}_{Xx\Theta}$ on cyl(x) or cyl(θ) and marginalizing the results on X or Θ.

Suppose the conditional belief functions $\mathrm{bel}_X(.|\theta_i)$ are normalized for all $\theta_i \in \Theta$, then any subset of $Xx\Theta$ whose projection of Θ is not Θ itself receives a zero belief, i.e., the only knowledge of the normalized conditional belief functions $\mathrm{bel}_X(.|\theta_i)$ induces a vacuous belief on Θ.

4.2.

Results of theorems 3 and 4 can also be derived by individually considering the ballooning extension bel_i of each conditional belief function $\mathrm{bel}_X(.\vert\theta_i)$, $i = 1, 2 \dots n$ $(n = |\Theta|)$ on space $Xx\Theta$. Then the bel_i are conditionated on $x \subseteq X$. The marginalization on Θ of the resulting conditional belief function is the (normalized) simple support function with basic belief masses

$$m\left(\bar{\theta}_i \vert x\right) = \mathrm{bel}_X(\bar{x} \vert \theta i) + m_X(\emptyset \vert \theta i)$$

$$m(\Theta \vert x) = \mathrm{bel}_X(X \vert \theta_i) - \mathrm{bel}_X(\bar{x} \vert \theta_i)$$

The conjunctive combination of these simple support functions on Θ obtained for each $\theta_i \in \Theta$ are the relations 3.11 to 3.13.

4.3.

Finally one can also consider that each θ_i $(i = 1, 2 \dots n)$ is the value of a variable Θ_i that can take only two values: θ_i and $\bar{\theta}_i$. Given $\mathrm{bel}_X(.\vert\theta_i)$, apply the principle of minimal commitment to build the belief function on the space $Xx\Theta_i$ (i.e., build the ballooning extension). Then vacuously extend these belief functions obtained on each $Xx\Theta_i$ onto the space $Xx\Theta_1 x\Theta_2 x \dots x\Theta_n$ by again applying the principle of minimal commitment (i.e., build their vacuous extensions on $Xx\Theta_1 x\Theta_2 x \dots x\Theta_n$). Combine all these belief functions on $Xx\Theta_1 x\Theta_2 x \dots x\Theta_n$ by the conjunctive rule of combination and call the resulting belief function bel_{Xn}. Let Θ be the space whose elements τ_i are the intersections (of the cylindrical extensions) of the complements of all θ_ν: $\nu \neq i$ and θ_i: so $\tau_i = \bar{\theta}_1 \cap \bar{\theta}_2 \dots \cap \theta_i \dots \cap \bar{\theta}_n$. Condition bel_{Xn} on the space $Xx\Theta$. The belief function induced on that space $Xx\Theta$ is the same as the one deduced in section 4.1.

Note that the belief function bel_X on X induced by the conditioning of bel_{Xn} on $\theta_1 \cap \theta_2 \dots \cap \theta_n$ is the belief function one would have derived by applying the conjunctive rule of combination to the individual conditional belief functions: $\mathrm{bel}_X = \mathrm{bel}_X(.\vert\theta_1) \textcircled{\wedge} \mathrm{bel}_X(.\vert\theta_2) \textcircled{\wedge} \dots \textcircled{\wedge} \mathrm{bel}_X(.\vert\theta_n)$.

5. PROPERTIES OF GBT

5.1.

Assume there exists some *a priori* belief bel_0 over Θ distinct from the belief induced by the set of conditional belief functions $\text{bel}_X(.\,|\,\theta_i)$, $\theta_i \in \Theta$. Combining bel_0 with the belief function induced on the space $Xx\Theta$ leads to a generalization of the DRC. By (2.3)

$$\text{bel}_X(x) = \sum_{\theta \subseteq \Theta} m_0(\theta)\text{bel}_X(x\,|\,\theta) \tag{5.1}$$

$$= \sum_{\theta \subseteq \Theta} m_0(\theta)\left(\prod_{\theta_i \in \theta} b_X(x\,|\,\theta_i) - \prod_{\theta_i \in \theta} b_X(\varnothing\,|\,\theta_i)\right) \tag{5.2}$$

$$\text{pl}_X(x) = \sum_{\theta \subseteq \Theta} m_0(\theta)\text{pl}_X(x\,|\,\theta) \tag{5.3}$$

$$= \sum_{\theta \subseteq \Theta} m_0(\theta)\left(1 - \prod_{\theta_i \in \theta} (1 - \text{pl}_X(x\,|\,\theta_i))\right) \tag{5.4}$$

Proof The solution is obtained by $\textcircled{\scriptsize\wedge}$-combining the vacuous extension of bel_0 on $Xx\Theta$ with $\text{bel}_{Xx\Theta}$ and marginalizing them on X, using then $\text{bel}_X(x\,|\,\theta)$ as given by equation 3.8. The full proof is given in Smets [6, p. 178]. QED

Equations 5.1 and 5.3 are particular cases of equation 2.3. They can be used to speed up computation of beliefs in beliefs networks.

To obtain the belief function induced on Θ given some $x \subseteq X$, we $\textcircled{\scriptsize\wedge}$-combine bel_0 with the belief function deduced on Θ by the GBT. The results are the same as those obtained if we combine the vacuous extension of bel_0 with the belief function $\text{bel}_{Xx\Theta}$ induced on $Xx\Theta$ by the set of conditional belief functions $\text{bel}_X(.\,|\,\theta_i)$, $\theta_i \in \Theta$ (see section 4.1) and then condition the result on x. (Proofs in Smets [6, p. 177]).

5.2.

Assume we have some belief bel_{X0} on X. The GBT becomes

$$\text{bel}_\Theta(\theta) = \sum_{x \subseteq X} m_{X0}(x)\text{bel}_\Theta(\theta\,|\,x) \tag{5.5}$$

where $\text{bel}_\Theta(\theta\,|\,x)$ is given by equation 3.11.

Proof Build the vacuous extension of bel_{X0} on $X x \Theta$, Ⓐ-combine it with $\text{bel}_{X x \Theta}$ as derived in section 4.1., and marginalize the result on Θ. QED

Note that equation 5.5 enables the backward propagation of belief based on doubtful observations.

5.3.

If each $\text{bel}_X(.\,\vdots\,\theta_i)$ happens to be a probability function $P(.|\theta_i)$ on X then the GBT for $|\theta| = 1$ becomes:

$$\text{pl}_\Theta(\theta\,\vdots\,x) = P(x|\theta) \qquad \forall x \subseteq X$$

That is, on the singletons θ of Θ, $\text{pl}_\Theta(.\,\vdots\,x)$ reduces to the likelihood of θ given x. The analogy stops there as the solution for the likelihood of subsets of Θ are different (see section 2.4).

If, furthermore, the *a priori* belief on θ is also a probability function $P_0(\theta)$, then the normalized GBT becomes:

$$\text{bel}_\Theta(\theta|x) = \frac{\sum_{\theta_i \in \theta} P(x|\theta_i)P_0(\theta_i)}{\sum_{\theta_i \in \Theta} P(x|\theta_i)P_0(\theta_i)} = P(\theta|x)$$

i.e., the (normalized) GBT reduces itself into the classic Bayesian theorem, which explains the origin of its name.

5.4.

Assume $\text{bel}_X(.\,\vdots\,\theta)$ is known not on each singleton of Θ, but on the elements of a partition of Θ. Then redefine Θ by creating the coarsening Θ' of Θ such that the elements of Θ' are the elements of the partition of Θ and proceed as before on the space Θ'.

5.5.

Assume $\text{bel}_X(.\,\vdots\,\theta)$ is known on subsets of Θ which are not mutually exclusive. For instance assume one knows $\text{bel}_X(.\,\vdots\,\theta_1)$, $\text{bel}_X(.\,\vdots\,\theta_2)$ and $\text{bel}_X(.\,\vdots\,\theta_1 \cup \theta_2)$. We must determine whether $\text{bel}_X(.\,\vdots\,\theta_1 \cup \theta_2)$ is compatible with the generalized likelihood principle (accepting some *a priori* belief on Θ) i.e., does there exist some *a priori* belief function bel_0 on Θ such that for all $x \subseteq X$:

$$\begin{aligned}\text{bel}_X(x\,\vdots\,\theta_1 \cup \theta_2) = m_0(\theta_1)\text{bel}_X(x\,\vdots\,\theta_1) + m_0(\theta_2)\text{bel}_X(x\,\vdots\,\theta_2) \\ + m_0(\theta_1 \cup \theta_2)(b_X(x\,\vdots\,\theta_1)b(x\,\vdots\,\theta_2) \\ -b(\emptyset\,\vdots\,\theta_1)b(\emptyset\,\vdots\,\theta_2))\end{aligned}$$

(see section 5.1.). A m_0 must be found that satisfies these constraints. This search will not always be successful in which case the DRC and the GBT do not apply. Failure reflects the fact that $\text{bel}_X(.\!\!\mid\!\theta_1 \cup \theta_2)$ is based on more information than the one represented by $\text{bel}_X(.\!\!\mid\!\theta_1)$, $\text{bel}_X(.\!\!\mid\!\theta_2)$ and some bel_0. Difficulties can also appear when there are several solutions m_0 that satisfy the constraints. We will not discuss them further here as, fortunately, in typical cases, $\text{bel}_X(.\!\!\mid\!\theta)$ is known for the singletons θ of Θ (or for subsets θ of Θ that constitutes a partition of Θ). Then both the DRC and the GBT apply.

5.6.

When one has an *a priori* belief function bel_{X0} on X, one could compute

$$\text{bel}^*_{Xi} = \text{bel}_X(.\!\!\mid\!\theta_i) \textcircled{\scriptsize$\wedge$} \text{bel}_{X0}$$

for each θ_i, i.e., our belief over X that combines both pieces of evidence, the one related to the θ_i and the one related to the prior on X. But it is erroneous to use the bel^*_{Xi} in the GBT directly. Indeed, bel^*_{Xi} and bel^*_{Xj}, $i \neq j$, do not result from distinct pieces of evidence as they share the same *a priori* bel_{X0}. The correct computation consists in isolating each $\text{bel}_X(.\!\!\mid\!\theta_i)$, ballooning them on $X x \Theta$, $\textcircled{\scriptsize$\wedge$}$-combining them and marginalizing them on X and then $\textcircled{\scriptsize$\wedge$}$-combining the result with bel_{X0}. Through this technique, each piece of evidence is taken into consideration once and only once.

5.7. Discounting a Belief Function

Consider an evidence that induces a normalized belief function bel_Ω on Ω. When the evidence as a whole is itself affected by some uncertainty (unreliability), Shafer [1, p. 251 et seq.] suggested "discounting" bel_Ω in order to take this new uncertainty into account. Let $1 - \alpha$ be the degree of trust (reliability) in the evidence as a whole, where $0 \leq \alpha \leq 1$. The discounted belief function bel^α_Ω on Ω is defined by Shafer [1, p. 251] such that:

$$\forall A \subseteq \Omega,\ A \neq \Omega, \qquad \text{bel}^\alpha_\Omega(A) = (1 - \alpha)\text{bel}_\Omega(A)$$

and

$$\text{bel}^\alpha_\Omega(\Omega) = \text{bel}_\Omega(\Omega) = 1$$

Shafer considers this concept of discounting as simple and useful but did not explain the origin of bel^α_Ω within his theory. It can be explained using the same ideas as those that lead to the GBT.

Let $\mathscr{E}$ be a frame with two elements E and $\bar{E}$, where E means "I know the evidence," and $\bar{E}$ means "I do not know the evidence." Assume that these are the only pieces of evidence available. By definition, the belief function $\text{bel}_\Omega(.|E)$ induced on Ω by E is bel_Ω. The belief function $\text{bel}_\Omega(.|\bar{E})$ induced by $\bar{E}$ on Ω is vacuous—not knowing an evidence leaves us in a state of total ignorance. Thus for each element in $\mathscr{E}$, one has a belief over Ω: $\text{bel}_\Omega(.|E) = \text{bel}_\Omega(.)$ and $\text{bel}_\Omega(.|\bar{E})$ is the vacuous belief function. Lemma 2 shows that $\text{bel}_\Omega(.|E \text{ or } \bar{E})$ is vacuous as $\text{bel}_\Omega(.|\bar{E})$ is vacuous (and this irrespective of the DRC).

Let $1 - \alpha$ be my degree of belief over $\mathscr{E}$ that E holds (i.e., my degree of belief that the source of the evidence E is reliable). So one has the bba over $\mathscr{E}$ with $m_{\mathscr{E}}(E) = 1 - \alpha$ and $m_{\mathscr{E}}(\mathscr{E}) = \alpha$.

Let bel^*_Ω be the belief induced on Ω by the conditional belief functions $\text{bel}_\Omega(.|E)$, $\text{bel}_\Omega(.|\bar{E})$ and $\text{bel}_\Omega(.|E \text{ or } \bar{E})$, and the prior bba $m_{\mathscr{E}}$ on $\mathscr{E}$. The application of (5.1) leads to:

$$\begin{aligned}
\text{bel}^*_\Omega(A) &= m_{\mathscr{E}}(E)\text{bel}_\Omega(.|E) + m_{\mathscr{E}}(\bar{E})\text{bel}_\Omega(.|\bar{E}) + m_{\mathscr{E}}(\mathscr{E})\text{bel}_\Omega(.|E \text{ or } \bar{E}) \\
&= (1 - \alpha)\text{bel}_\Omega(A) \qquad \forall A \subseteq \Omega, A \neq \Omega \\
&= 1 \qquad A = \Omega
\end{aligned}$$

Hence $\text{bel}^*_\Omega = \text{bel}^\alpha_\Omega$. The relation is always true as it is derived from (5.1) which always holds and not from (5.2) which is derived from the GBT. The discounted belief function bel^α_Ω can thus be justified within the TBM.

Informally, the discounted belief function bel^α_Ω results from the idea that I have a degree of belief $(1 - \alpha)$ that E is a legitimate (reliable) piece of evidence, in which case my belief on Ω is quantified by bel_Ω. The remaining bbm α is given to the fact that E might be but is not necessarily a legitimate piece of evidence, in which case my belief on Ω can be quantified by any belief function, including bel_Ω. In such a state of ignorance, the principle of minimal commitment justifies the use of the vacuous belief function to quantify my belief on Ω. bel^α_Ω results from the combination of the initial belief function bel_Ω on Ω and the belief built on $\mathscr{E}$.

Discounting can also be seen as the result of the impact of a meta-belief over the set $\mathscr{B}$ of belief functions on Ω. It fits with a very special but important case of a general theory of meta-beliefs. $1 - \alpha$ is the meta-bbm (the basic belief mass related to the meta-belief function) given to the particular element bel_Ω of the set $\mathscr{B}$ of belief functions on Ω. α is the meta-bbm given to $\mathscr{B}$ itself. The discounting operation corresponds to the collapse of the meta-beliefs over the set of belief functions on Ω into a belief function on Ω.

6. BELIEF NETWORKS

We now introduce some possible applications of the GBT and the DRC. All belief functions considered here are induced by distinct pieces of evidence.

Consider the simplest directed belief network with two nodes Θ and X representing binary variables. The weights on the edge are the conditional plausibility functions on X given θ and $\bar{\theta}$.

$$\Theta \longrightarrow X$$

$$\left| \begin{matrix} \mathrm{pl}(x \mid \theta) & \mathrm{pl}(\bar{x} \mid \theta) & \mathrm{pl}(x \cup \bar{x} \mid \theta) \\ \mathrm{pl}(x \mid \bar{\theta}) & \mathrm{pl}(\bar{x} \mid \bar{\theta}) & \mathrm{pl}(x \cup \bar{x} \mid \bar{\theta}) \end{matrix} \right|$$

Forward propagation:

Assume there is some basic belief masses on Θ: $\mathrm{m}(\theta)$, $\mathrm{m}(\bar{\theta})$ and $\mathrm{m}(\theta \cup \bar{\theta})$. Then we can compute the plausibility induced on X by equation 5.4:

$$\mathrm{pl}(x) = m(\theta)\mathrm{pl}(x \mid \theta) + m(\bar{\theta})\mathrm{pl}(x \mid \bar{\theta}) + m(\theta \cup \bar{\theta})(1 - (1 - \mathrm{pl}(x \mid \theta))(1 - \mathrm{pl}(x \mid \bar{\theta})))$$

$$\mathrm{pl}(\bar{x}) = m(\theta)\mathrm{pl}(\bar{x} \mid \theta) + m(\bar{\theta})\mathrm{pl}(\bar{x} \mid \bar{\theta}) + m(\theta \cup \bar{\theta})(1 - (1 - \mathrm{pl}(\bar{x} \mid \theta))(1 - \mathrm{pl}(\bar{x} \mid \bar{\theta})))$$

$$\mathrm{pl}(x \cup \bar{x}) = m(\theta)\mathrm{pl}(x \cup \bar{x} \mid \theta) + m(\bar{\theta})\mathrm{pl}(x \cup \bar{x} \mid \bar{\theta}) + m(\theta \cup \bar{\theta})(1 - (1 - \mathrm{pl}(x \cup \bar{x} \mid \theta))(1 - \mathrm{pl}(x \cup \bar{x} \mid \bar{\theta})))$$

Backward propagation:

Should we receive a plausibility on X instead, we could compute the belief on Θ by equation (3.3)

$$\mathrm{pl}(\theta) = m(x)\mathrm{pl}(x \mid \theta) + m(\bar{x})\mathrm{pl}(\bar{x} \mid \theta) + m(x \cup \bar{x})\mathrm{pl}(x \cup \bar{x} \mid \theta)$$

$$\mathrm{pl}(\bar{\theta}) = m(x)\mathrm{pl}(x \mid \bar{\theta}) + m(\bar{x})\mathrm{pl}(\bar{x} \mid \theta) + m(x \cup \bar{x})\mathrm{pl}(x \cup \bar{x} \mid \bar{\theta})$$

$$\mathrm{pl}(\theta \cup \bar{\theta}) = m(x)(1 - (1 - \mathrm{pl}(x \mid \theta))(1 - \mathrm{pl}(x \mid \bar{\theta})) + m(\bar{x})(1 - (1 - \mathrm{pl}(\bar{x} \mid \theta))(1 - \mathrm{pl}(\bar{x} \mid \bar{\theta}))) + m(x \cup \bar{x})(1 - (1 - \mathrm{pl}(x \cup \bar{x} \mid \theta))(1 - \mathrm{pl}(x \cup \bar{x} \mid \bar{\theta})))$$

Propagation in both directions:

Should one receive both a belief bel_Θ on Θ and a belief bel_X on X, then

for the X node: apply forward propagation using bel_Θ and the conditional plausibilities and $\textcircled{\scriptsize\wedge}$-combine the result with bel_X.

for the Θ *node*: apply backward propagation using bel_X and the conditional plausibilities and $\textcircled{\scriptsize\wedge}$-combine the result with bel_Θ.

Notice the strong symmetry between the above two sets of formula; it reflects the fact that unnormalized conditional plausibilities are symmetrical in their two arguments. Computing the corresponding belief function is immediate. Computing the corresponding basic belief masses or the commonality function should be done with the Fast Moebius Transform (Kennes and Smets [33]) to optimize computation time.

For more complicated acyclic belief networks, the computation is similar. Each node stores the beliefs induced by its immediate neighbors. Once a node X indicates that its belief has changed, it propagates its new belief to all its neighbors. Each neighbor updates the belief induced by X by $\textcircled{\scriptsize\wedge}$-combine with its stored beliefs, using commonality functions for efficiency reasons. They then propagate the updated belief to Y's neighbors that have not yet been updated. This propagation is in fact identical to the one encountered in Shafer, Shenoy, and Mellouli's algorithm (Shafer et al. [4]). The advantage of our method is that storage on the edge is smaller (at most $|\Theta|2^{|X|}$ values) and propagation between nodes is accelerated. The only weakness of our method is that it does not cover *all* possible belief functions between two variables; it is restricted to those belief functions that can be represented through the set of conditional belief functions, thus a subset of the set of all belief functions. We believe that this loss of generality is not serious, as far as most natural cases correspond to those where only the conditional belief functions are received. Finally, our computation is faster and requires less memory than the Shafer-Shenoy-Mellouli algorithm.

7. EXAMPLE

In order to illustrate the use of the GBT and the DRC, we consider an example of a medical diagnosis process. Let $\Theta = \{\theta_1, \theta_2, \theta_\omega\}$ be a set of diseases with three mutually exclusive and exhaustive diseases. θ_1 and θ_2 are two "well-known" diseases, i.e., we have some beliefs on what symptoms could hold when θ_1 holds or when θ_2 holds. θ_ω corresponds to the complement of $\{\theta_1, \theta_2\}$ relative to all possible diseases. θ_ω represents not only all the "other" diseases but also those not yet known. In such a context, our belief on the symptoms can only be vacuous. What do we know about the symptoms caused by a still unknown disease? Nothing of course, hence the vacuous belief function.

We consider two sets X and Y of symptoms with $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2\}$. Tables 1 and 2 present the beliefs over X and Y when each of the individual diseases holds. They also show the beliefs over the symptoms when we only know that either θ_1 or θ_2 holds. They are derived from theorem 3. The beliefs translate essentially the facts that θ_1 "causes" (supports) x_3 and y_2, and θ_2 "causes" x_1 or x_2 (without preference) and y_1. When we only know that θ_1 or θ_2 holds, then we have a balanced support over X, and some support in favor of y_1.

Table 3 presents the beliefs induced on Θ by the individual observation of symptom x_3 or of symptom y_2, respectively. We assume that the symptoms are independent within each disease, hence the GBT can be applied. The independence assumption means that if we knew which disease holds the observation of one of the symptoms, it would not change our belief about the status of the other symptom. The right half of Table 3 presents the beliefs induced on Θ by the joint observation of symptom x_3 and of symptom y_2. The beliefs are computed by the application of theorem 4. The symptoms individually and jointly support essentially $\{\theta_1, \theta_\omega\}$. The meaning of $\text{bel}(\theta_\omega \| x_3, y_2) = 0.27$ merits some consideration. It quantifies our belief that the joint symptoms x_3 and y_2 are not "caused" by θ_1 nor by θ_2. It supports the fact that the joint observation is "caused" by another disease or by some still unknown disease. A large value for

Table 1. Conditional beliefs (bel) and bbm (m) on the symptoms $x \subseteq X$ within each of the mutually exclusive and exhaustive diagnosis θ_1, θ_2, and $\theta_\omega \in \Theta$. The right part of the table present the beliefs (and bbm) on X given the disease is either θ_1 or θ_2.

	$\{\theta_1\}$		$\{\theta_2\}$		$\{\theta_\omega\}$		$\{\theta_1, \theta_2\}$	
X	m	bel	m	bel	m	bel	m	bel
$\{x_1\}$	.0	.0	.0	.0	.0	.0	.00	.00
$\{x_2\}$	.0	.0	.0	.0	.0	.0	.00	.00
$\{x_3\}$	.5	.5	.2	.2	.0	.0	.10	.10
$\{x_1, x_2\}$	.2	.2	.6	.6	.0	.0	.12	.12
$\{x_1, x_3\}$	.0	.5	.1	.3	.0	.0	.05	.15
$\{x_2, x_3\}$	.0	.5	.1	.3	.0	.0	.05	.15
$\{x_1, x_2, x_3\}$	.3	1.0	.0	1.0	1.0	1.0	.68	1.00

Table 2. Conditional beliefs (bel) and bbm (m) on the symptoms $y \subseteq Y$ within each of the mutually exclusive and exhaustive diagnosis θ_1, θ_2, and $\theta_\omega \in \Theta$. The right part of the table present the beliefs (and bbm) on Y given the disease is either θ_1 or θ_2.

	$\{\theta_1\}$		$\{\theta_2\}$		$\{\theta_\omega\}$		$\{\theta_1, \theta_2\}$	
Y	m	bel	m	bel	m	bel	m	bel
$\{y_1\}$	.1	.1	.6	.6	.0	.0	.12	.12
$\{y_2\}$	.7	.7	.0	.0	.0	.0	.00	.00
$\{y_1, y_2\}$	.1	.9	.4	1.0	1.0	1.0	.88	1.00

$\text{bel}(\theta_\omega \mid x_3, y_2)$ somehow supports the fact that we might be facing a new disease. In any case it should induce us in looking for other potential causes to explain the observations.

Table 4 presents the beliefs induced on $\{\theta_1, \theta_2\}$ when we condition our beliefs on Θ on $\{\theta_1, \theta_2\}$, or when we have some *a priori* belief on Θ. The results are obtained by the application of the conjunctive rule of combination applied to the *a priori* belief on Θ and the belief induced by the joint observations. The belief functions presented are normalized.

8. CONCLUSIONS

We have presented the GBT and the DRC built on the knowledge of a set of conditional belief functions $\text{bel}_X(.\mid\theta)$ on X for each θ in Θ where the θ's constitute a partition of Θ. Distinct beliefs on X and/or Θ can be

Table 3. Left part: the basic belief masses (m) and the related commonality functions (q) induced on Θ by the observation of symptom x_3 or of symptom y_2. Right part: the basic belief masses (m) and the related belief function (bel), plausibility function (pl) and commonality function (q) induced on Θ by the joint observation of x_3 and y_2.

	$\mid x_3$		$\mid y_2$		$\mid x_3, y_2$			
Θ	m	q	m	q	m	q	bel	pl
$\{\theta_1\}$	.00	.80	.00	.80	.00	.64	.00	.64
$\{\theta_2\}$	.00	.40	.00	.60	.00	.24	.00	.24
$\{\theta_\omega\}$	.12	1.00	.08	1.00	.27	1.00	.27	1.00
$\{\theta_1, \theta_2\}$	.00	.32	.00	.48	.00	.15	.00	.73
$\{\theta_1, \theta_\omega\}$	.48	.80	.32	.80	.49	.64	.76	1.00
$\{\theta_2, \theta_\omega\}$	.08	.40	.12	.60	.09	.24	.36	1.00
$\{\theta_1, \theta_2, \theta_\omega\}$	.32	.32	.48	.48	.15	.15	1.00	1.00

Table 4. The basic belief masses (m) and the related (normalized) belief function (bel_n) induced on Θ by the joint observation of x_3 and y_2, and based on three different *a priori* beliefs on Θ: an *a priori* that rejects θ_ω, a probabilistic *a priori* on $\{\theta_1, \theta_2\}$, and a simple support function on $\{\theta_1, \theta_2\}$.

$\mid x_3, y_2$	$m(\theta_1, \theta_2) = 1$		$m(\theta_1) = .3$ $m(\theta_2) = .7$		$m(\theta_1) = .3$ $m(\theta_1, \theta_2) = .7$	
Θ	m	bel_n	m	bel_n	m	bel_n
{ }	.30	.00	.70	.00	.32	
$\{\theta_1\}$	.54	.77	.19	.63	.57	.84
$\{\theta_2\}$	.06	.09	.11	.37	.04	.06
$\{\theta_1, \theta_2\}$	.10	1.00	.00	1.00	.07	1.00

included. Beside the direct relevance of these theorems for inference and the combination of distinct disjunctive pieces of evidence, they are also useful when building belief networks: the assessment of conditional beliefs on X given each θ is more natural and easier than the direct assessment of the joint belief on the space $Xx\Theta$. The loss of generality does not appear to be of any practical importance. In any case, even for the general one, one can always speed up computation and reduce memory requirements thanks to equations 5.1 and 5.3 that are always valid. Instead of storing the general belief function $bel_{Xx\Theta}$, store the set of conditional belief functions $bel_X(.\mid\theta)\ \forall\theta \subseteq \Theta$. The total amount of stored data is at most $2^{|X|+|\Theta|}$ instead of $2^{|X|.|\Theta|}$, a serious gain once $|X|$ and $|\Theta|$ become large.

The appropriate use of the GBT and the DRC resolves many of the problems that were raised in Pearl [34] as supposedly counterexamples against the Dempster–Shafer theory (see Smets [35] for an in-depth re-analysis of these examples).

One should take care not to apply the GBT and the DRC blindly. The generalized likelihood principle is not always satisfied; its applicability must be verified. As a counterexample, consider a set of urns with ten balls among which some (n) are white, the others black. Suppose there is an urn with six white balls ($n = 6$). Let bel(W$\mid n = 6$) be your belief that the next ball extracted from that urn is white knowing there are six white balls. You are free to give any value to bel(W$\mid n = 6$). Hacking's frequency principle (Hacking [36]) supports that bel(W$\mid n = 6$) should be $6/10$. It provides a reference scale to quantify beliefs, but any monotonous transformation could be as good. Nevertheless bel(W$\mid n = 6$) and bel(W$\mid n = 7$) are related: once bel(W$\mid n = 6$) is given, bel(W$\mid n = 7$) may not be smaller, (if you have the least amount of coherence). Only in the world of "Absurdia" could one accept that the knowledge of bel(W¦n = 6) does not induce any constraint on the value of bel(W¦n = 7). We accept—hopefully—that we are not living in Absurdia. Hence bel(W¦n = 6) and bel(W¦n = 7) are related by extra constraints, and these constraints must be incorporated into the model. Blindly applying the GBT in such a context without due regard to the constraints that exist between the conditional belief functions would lead to erroneous answers.

APPENDIX

LEMMA 3: *If* $pl_X(x|z)/pl_Y(y|z) = pl_X(x)/pl_Y(y)\ \forall x,\ y \subseteq X,\ \forall z \subseteq Z$, *then* $pl_X(x|z) = pl_X(x)pl_Z(z)$.

Proof By hypothesis, $\mathrm{pl}_X(x|z)/\mathrm{pl}_X(x) = \mathrm{pl}_Y(y|z)/\mathrm{pl}_Y(y)$. So these ratios do not depend on x (nor on y). Let the ratio be equal to f(z). Hence $\mathrm{pl}_X(x|z) = \mathrm{pl}_X(x)\,\mathrm{f}(z)$. As $\mathrm{pl}_X(x|z) = \mathrm{pl}_{XxZ}(x \cap z) = \mathrm{pl}_Z(z|x)$, then f($z$) $= \mathrm{pl}_Z(z)$. QED

LEMMA 4: *Let X and Y be two frames of discernment. Let* pl_X *and* pl_Y *be plausibility functions over the frames of discernment X and Y, respectively. Let* pl_{XxY} *be the plausibility function on XxY such that:* $pl_{XxY}(x \cap y) = pl_X(x)pl_Y(y)$. *Then*

$$\mathrm{bel}_X(x|y) = \mathrm{bel}_X(x)\mathrm{pl}_Y(y) \qquad \forall x \subseteq X, \forall y \subseteq Y.$$

and

$$\frac{\mathrm{bel}_X(x_1|y)}{\mathrm{bel}_X(x_2|y)} = \frac{\mathrm{bel}_X(x_1)}{\mathrm{bel}_X(x_2)} \qquad \forall x_1, x_2 \subseteq X, \forall y \subseteq Y.$$

Proof

One has:

$$\mathrm{pl}_Y(y) = \frac{\mathrm{pl}_Y(y)(\mathrm{pl}_X(X) - \mathrm{pl}_X(\bar{x}))}{\mathrm{pl}_X(X) - \mathrm{pl}_X(\bar{x})} = \frac{\mathrm{pl}_{XxY}(X \cap y) - \mathrm{pl}_{XxY}(\bar{x} \cap y)}{\mathrm{bel}_X(x)}$$

$$= \frac{\mathrm{pl}_X(X|y) - \mathrm{pl}_X(\bar{x}|y)}{\mathrm{bel}_X(x)} = \frac{\mathrm{bel}_X(x|y)}{\mathrm{bel}_X(x)}$$

what proves the first equality. The second is then immediate. QED

Proof of Theorem 1:

Let X and Y be two finite spaces. Let $\{\mathrm{bel}_X(.|\theta_i), \theta_i \in \Theta\}$ and $\{\mathrm{bel}_Y(.|\theta_i), \theta_i \in \Theta\}$, be two sets of normalized belief functions on X and Y, respectively. Let $\mathrm{pl}(\theta | x)$, $q(\theta | x)$ and $\mathrm{pl}(\theta | y)$, $q(\theta | y)$ be the plausibility and commonality functions induced on Θ by the two distinct pieces of evidence $x \subseteq X$ and $y \subseteq Y$. Requirement R1 is:

$$q(\theta | x, y) = q(\theta | x).q(\theta | y). \qquad \forall \theta \subseteq \Theta \tag{A.1}$$

It becomes by lemma 1:

$$\sum_{\theta' \subseteq \theta} (-1)^{|\theta'|+1} \mathrm{pl}(\theta' | x, y) = \left[\sum_{\theta' \subseteq \theta} (-1)^{|\theta'|+1} \mathrm{pl}(\theta' | x) \right] \cdot \left[\sum_{\theta' \subseteq \theta} (-1)^{|\theta'|+1} \mathrm{pl}(\theta' | y) \right] \tag{A.2}$$

We analyze successively the cases $|\theta| = 1$, 2 and n.

1. When $|\theta| = 1$, equation A.2 becomes: $\mathrm{pl}(\theta | x, y) = \mathrm{pl}(\theta | x)\mathrm{pl}(\theta | y)$ or equivalently

$$\mathrm{pl}_{XxY}(x \cap y | \theta) = \mathrm{pl}_X(x | \theta)\mathrm{pl}_Y(y | \theta) \tag{A.3}$$

So x and y are CCI (see Section 2.3).

2. Assume $\theta = \theta_1 \cup \theta_2$ with θ_1, $\theta_2 \in \Theta$, $\theta_1 \neq \theta_2$. For $i = 1,2$, let $\alpha_i = \mathrm{pl}_X(x | \theta_i)$, $\gamma_i = \mathrm{pl}_Y(y | \theta_i)$, $\bar{\alpha}_i = \mathrm{pl}_X(\bar{x} | \theta_i)$, $\bar{\gamma}_i = \mathrm{pl}_Y(\bar{y} | \theta_i)$, $\mathrm{f}_i = \mathrm{pl}_{XxY}(\bar{x} \cup \bar{y} | \theta_i)$. By A.3, $\mathrm{pl}_{XxY}(x \cap y | \theta_1) = \alpha_1 \gamma_1$ and $\mathrm{pl}_{XxY}(x \cap y | \theta_2) = \alpha_2 \gamma_2$. By the generalized likelihood principle, there exists a g function such that

$$\mathrm{pl}_X(x | \theta) = g(\alpha_1, \bar{\alpha}_1, \alpha_2, \bar{\alpha}_2) \quad \text{and} \quad \mathrm{pl}_Y(y | \theta) = g(\gamma_1, \bar{\gamma}_1, \gamma_2, \bar{\gamma}_2)$$

Equation A.2 becomes:

$$\alpha_1\gamma_1 + \alpha_2\gamma_2 - g(\alpha_1\gamma_1, f_1, \alpha_2\gamma_2, f_2) = (\alpha_1 + \alpha_2 - g(\alpha_1, \bar{\alpha}_1, \alpha_2, \bar{\alpha}_2)) \cdot (\gamma_1 + \gamma_2 - g(\gamma_1, \bar{\gamma}_1, \gamma_2, \bar{\gamma}_2)) \tag{A.4}$$

Let $\mathrm{pl}_X(. | \theta_1)$ be vacuous. Hence $\alpha_1 = \bar{\alpha}_1 = 1$ and $f_1 = 1$ as $f_1 = \mathrm{pl}_{XxY}(\bar{x} \cup \bar{y} | \theta_1) \geq \mathrm{pl}_X(\bar{x} | \theta_1) = 1$. One has also $g(1, 1, \alpha_2, \bar{\alpha}_2) = 1$ as $\mathrm{pl}_X(x | \theta) \geq \mathrm{pl}_X(x | \theta_1) = 1$.

Equation A.4 becomes:

$$\gamma_1 = \alpha_2\gamma_2 - g(\gamma_1, 1, \alpha_2\gamma_2, f_2) = \alpha_2(\gamma_1 + \gamma_2 - g(\gamma_1, \bar{\gamma}_1, \gamma_2, \bar{\gamma}_2)$$

So g does not depend on its second parameter. Identically g does not depend on its fourth parameter.

Let: $k(\alpha, \gamma) = g(\alpha,., \gamma,.)$.

One has $\mathrm{pl}(\theta_1 \cup \theta_2 | x) = k(\mathrm{pl}(\theta_1 | x), \mathrm{pl}(\theta_2 | x))$, or identically, $\mathrm{pl}_X(x | \theta_1 \cup \theta_2) = k(\mathrm{pl}_X(x | \theta_1), \mathrm{pl}_X(x | \theta_2))$. Let $\mathrm{pl}_X(x | \theta_1) = 1$. As $\mathrm{pl}_X(x | \theta_1 \cup \theta_2) \geq$

$\mathrm{pl}_X(x|\theta_1)$ by lemma 2, then $k(1, \gamma) = 1 = k(\gamma, 1)$ as k is symmetrical in its arguments.

Let $\alpha_1 = \gamma_2 = 1$. Then equation A.4 becomes:

$$\gamma_1 + \alpha_2 - k(\gamma_1, \alpha_2) = (1 - \alpha_2 - 1).(\gamma_1 + 1 - 1))$$

hence,

$$k(\gamma_1, \alpha_2) = \gamma_1 + \alpha_2 - \alpha_2\gamma_1 = 1 - (1 - \gamma_1)(1 - \alpha_2)$$

and

$$\begin{aligned}\mathrm{pl}_X(x|\theta) &= k(\alpha_1, \alpha_2) = 1 - (1 - \alpha_1)(1 - \alpha_2) \\ &= 1 - (1 - \mathrm{pl}_X(x|\theta_1)).(1 - \mathrm{pl}_X(x|\theta_2))\end{aligned}$$

3. By iteration one gets $\mathrm{pl}_X(x|\theta)$. Assume $\theta = \bigcup_{i=1}^{n}\theta_i$ where $\theta_i \cap \theta_j = \emptyset\ \forall i \neq j$. Assume

$$\mathrm{pl}_X(x|\theta) = 1 - \prod_{\theta_i \in \theta}(1 - \mathrm{pl}_X(x|\theta_i)) = 1 - \prod_{i=1}^{n}(1 - \alpha_i).$$

Consider part 2 of the proof, but replace θ_1 by θ and θ_2 by θ_{n+1}. The proof can proceed as in 2:

$$\begin{aligned}\mathrm{pl}_X(x|\theta \cup \theta_{n+1}) &= \mathrm{pl}_X(x|\theta) + \mathrm{pl}_X(x|\theta_{n+1}) \\ &\quad - \mathrm{pl}_X(x|\theta)\mathrm{pl}_X(x|\theta_{n+1}) \\ &= 1 - \prod_{\theta_i \in \theta \cup \theta_{n+1}}(1 - \mathrm{pl}_X(x|\theta_i)) \qquad \text{(A.5)}\end{aligned}$$

The relation for $\mathrm{bel}_X(x|\theta)$ and $m_X(x|\theta)$ are deduced from A.5. The results are normalized. QED

Proof of Theorem 2:

Derive directly from $\mathrm{pl}(\theta|x) = \mathrm{pl}_X(x|\theta)$ and $\mathrm{bel}(\theta|x) = \mathrm{pl}(\Theta|x) - \mathrm{pl}(\bar{\theta}|x)$ and normalize by dividing by $\mathrm{bel}(\Theta|x)$. QED

Proof of Theorem 3:

$m_X(\emptyset|\theta_i)$ (and/or $\mathrm{m}_Y(\emptyset|\theta_i)$) might be non-null. To see the impact of such non-null basic belief masses, enlarge the X space into X' where $X' = X \cup \omega$ and $X \cap \omega = \emptyset$. Apply the same proof as for theorem 1 with normalized belief functions on X' and condition all results on X. As such conditioning is idempotent, one can apply it at the level of $\mathrm{pl}_X(.|\theta)$ or at the level of each $\mathrm{pl}_X(.|\theta_i)$. For all $x \subseteq X$, the plausibilities before and after conditioning are the same. So the generalized likelihood principle still applies for all $x \subseteq X$. But after the conditioning has been applied, the functions $\mathrm{pl}_X(.|\theta_i)$ are un-normalized plausibility functions. QED

Proof of Theorem 4:

Derive directly from $\mathrm{pl}(\theta|x) = \mathrm{pl}_X(x|\theta)$ and $\mathrm{bel}(\theta|x) = \mathrm{pl}(\Theta|x) - \mathrm{pl}(\bar{\theta}|x)$. QED

ACKNOWLEDGMENTS

The author is indebted to Didier Dubois, Yen-Teh Hsia, Frank Klawonn, Rudolf Kruse, Victor Poznanski, and an anonymous referee for many suggestions for improving the presentation, and Glen Shafer who, in 1978, indicated to the author that the ballooning extension satisfies the principle of minimal commitment.

References

1. Shafer, G., *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, NJ, 1976.
2. Smets, Ph., Belief functions, in (Ph. Smets, A. Mamdani, D. Dubois, and H. Prade, Eds.), *Nonstandard Logics for Automated Reasoning*. Academic Press, London, 253–286, 1988.
3. Smets, Ph. The concept of distinct evidence. *IPMU 92 Proceedings*, 789–794, 1992.
4. Shafer, G., Shenoy, P. P., and Mellouli, K., Propagating belief functions in qualitative Markov trees, *Int. J. Approx. Reas.* 1:349–400, 1987.
5. Pearl, J., Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Pub., San Mateo, CA, 1988.
6. Smets, Ph., Un modèle mathématico-statistique simulant le processus du diagnostic médical. Doctoral dissertation, Université Libre de Bruxelles, Bruxelles, 1978 (available through University Microfilm International, 30–32 Mortimer Street, London W1N7RA, thesis 80-70,003).
7. Smets, Ph., Medical diagnosis: Fuzzy sets and degrees of belief, *Fuzzy Sets Syst.* 5:259–266, 1981.
8. Smets, Ph., Bayes' theorem generalized for belief functions, *Proc. ECAI-86*, vol. II., 169–171, 1986.

9. Shafer, G., Belief functions and parametric models, *J. Roy. Statist. Soc.* B44 322–352, 1982.

10. Moral, S., Informacion difusa. Relationes eentre probabilidad y possibilidad, Tesis Doctoral, Universidad de Granada, 1985.

11. Dubois, D., and Prade, H., A set theoretical view of belief functions, *Int. J. Gen. Syst.* 12:193–226, 1986.

12. Dubois, D., and Prade, H., *Computational Intell.* 4:244–263, 1988.

13. Cohen, M. S., Laskey, K. B., and Ulvila, J. W., The management of uncertainty in intelligence data: a self-reconciling evidential database. Falls Church, VA: Decision Science Consortium, Inc., 1987.

14. Smets, Ph., The combination of evidence in the transferable belief model, *IEEE–Pattern Analysis and Machine Intelligence* 12:447–458, 1990.

15. Smets, Ph., and Kennes, R., The transferable belief model. Technical Report: *TR-IRIDIA-90-14.* To be published in *Artificial Intelligence.*

16. Smets, Ph. The transferable belief model and other interpretations of Dempster–Shafer's Model, in (P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, Eds.), *Uncertainty in Artificial Intelligence 6.* North Holland, Amsterdam, 375–384, 1991.

17. Kohlas, J., and Monney, P. A., Modeling and reasoning with hints, *Technical Report.* Inst. Automation and OR. Univ. Fribourg, 1990.

18. Gebhardt, F., and Kruse, R., The context model: an integrating view of vagueness and uncertainty. To appear in *Int. J. Approx. Reas.*.

19. Smets, Ph., The nature of the unnormalized beliefs encountered in the transferable belief model, in (D. Dubois, M. P. Wellman, B. d'Ambrosio, and Ph. Smets, Eds.), *Uncertainty in AI 92.* Morgan Kaufmann, San Mateo, CA, 292–297, 1992.

20. Klawonn, F. and Smets, Ph., The dynamic of belief in the transferable belief model and specialization-generalization matrices, in (D. Dubois, M. P. Wellman, B. d'Ambrosio, and Ph. Smets, Eds.), *Uncertainty in AI 92.* Morgan Kaufmann, San Mateo, CA, 130–137, 1992.

21. Nguyen, T. H., and Smets, Ph., On dynamics of cautious belief and conditional objects. TR/IRIDIA/91-13. *Int. J. Approx. Reas.* 8:89–104 (1993).

22. Dubois, D., and Prade, H., On the unicity of Dempster rule of combination, *Int. J. Intell. Syst.* 1:133–142, 1986.

23. Yager, R., The entailment principle for Dempster–Shafer granules, *Int. J. Intell. Syst.* 1:247–262, 1986.

24. Dubois, D., and Prade, H., The principle of minimum specificity as a basis for evidential reasoning, in *Uncertainty in Knowledge-based Systems,* (B. Bouchon and R. Yager, Eds.), Springer Verlag, Berlin, 75–84, 1987.

25. Delgado, M., and Moral, S., On the concept of possibility-probability consistency, *Fuzzy Sets Syst.* 21:311–3018, 1987.

26. Kruse, R., and Schwecke, E., Specialization: a new concept for uncertainty handling with belief functions, *Int. J. Gen. Syst.* 18:49–60, 1990.

27. Hsia, Y.-T., Characterizing belief with minimum commitment, *IJCAI* 91:1184–1189, 1991.

28. Halpern, J. Y., and Fagin, R., Two views of belief: Belief as generalized probability and belief as evidence. *Proc. Eighth Natl. Conf. AI*, 112–119, 1990.

29. Edwards, A. W. F., *Likelihood.* Cambridge University Press, Cambridge, UK, 1972.

30. Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets Syst.* 1:3–28, 1978.

31. Dubois, D., and Prade, H., *Theorie des Possibilités*. Masson, Paris, 1985.

32. Smets, Ph., Possibilistic inference from statistical data, in *Second World Conference on Mathematics at the Service of Man*, (A. Ballester, D. Cardus, and E. Trillas, Eds.), Universidad Politecnica de Las Palmas, 611–613, 1982.

33. Kennes, R., and Smets, Ph., Computational aspects of the Möbius transform, *Proc. 6th Conf. Uncertainty AI*, Cambridge, USA, 1990.

34. Pearl, J., Reasoning with belief functions: an analysis of compatibility. *Int. J. Approx. Reas.* 4:363–390, 1990.

35. Smets, Ph., Resolving misunderstandings about belief functions: A response to the many criticisms raised by J. Pearl, *Int. J. Approx. Reas.* 6:321–344, 1992.

36. Hacking, I., *Logic of Statistical Inference*. Cambridge University Press, Cambridge, U.K., 1965.