# BELIEF FUNCTIONS AND THE TRANSFERABLE BELIEF MODEL

## PHILIPPE SMETS

ABSTRACT. Belief functions have been proposed for modeling someone's degrees of belief. They provide alternatives to the models based on probability functions or on possibility functions. There are several interpretations of belief functions: the lower probabilities model, Dempster's model, the hint model, the probability of modal propositions model, the transferable belief model. All these models are unfortunately clustered under the generic name of Dempster–Shafer theory, which hides their differences and explains most of the confusion and errors that appear in the literature.

## 1. INTRODUCTION

1.1. **Dempster–Shafer theories.** Dempster–Shafer theory covers several models that use the mathematical object called 'belief function'. Usually their aim is in the *modeling of someone's degrees of belief*, where a degree of belief is understood as a strength of opinion. We do not study the problem of vagueness and ambiguity for which fuzzy set theory and possibility theory are appropriate.

Beliefs result from *uncertainty*. Uncertainty sometimes results from a random process (the objective probability case), it sometimes results only from the lack of information that induces some 'belief' ('belief' must be contrasted to 'knowledge' as what is believed can be false). The uncertainty studied here concerns the one usually quantified by probability functions, as done by the Bayesians.

We will consider the upper and lower probabilities (ULP) model, Dempster's model, the theory of hints, the transferable belief model (TBM), the probability of provability model. Each model corresponds to a different understanding of the concept of uncertainty.

None of these models receives the *qualification of 'Dempster–Shafer'* and I do not think any model deserves it. This qualification has an historical origin (Gordon and Shortliffe, 1984) [9] but is misleading. Some people qualify any model that uses belief functions as Dempster–Shafer. This might be acceptable provided they did not blindly accept the applicability of both Dempster's rules of conditioning and combination. Such uncritical — and in fact often inappropriate — use of these two rules explains most of the errors encountered in the so-called Dempster–Shafer literature (Pearl, 1990) [23].

Both Dempster and Shafer introduced models but I do not think they would recognise their model as a 'Dempster–Shafer' model. Dempster's seminal work was not orientated toward modeling someone's beliefs. The idea of Shafer was to use the mathematical model introduced by Dempster in order to model someone's belief

(Shafer, 1976a) [26]. In his 1976 book, Shafer models beliefs without considering probability theory, but in later papers, he often used the concepts of random sets, one-to-many mappings and upper and lower probabilities, creating confusion. We introduced the concept of the transferable belief model (TBM) in order to justify the use of belief functions (including Dempster's rule of conditioning and Dempster's rule of combination) to model someone's belief (Smets, 1988b) [32]. The TBM is a theory in itself, not an adaptation of probability theory. It seems to fit essentially with what Shafer developed in his book.

This presentation will successively introduce the transferable belief model, then the upper and lower probability model, Dempster's model, the model of hints, and finally probabilities defined on modal propositions. We start with a definition of what these models intend to represent.

1.2. **The representation of quantified beliefs.** We start from a finite set of worlds $\Omega$ called the frame of discernment. One of its worlds, denoted $\omega_0$, corresponds to the **actual world**. The term 'world' is used in a general sense. It covers concepts like 'state of affairs', 'state of nature', 'situation', 'context', 'value of a variable' ...

An agent, denoted You (but it might be a robot, a piece of software), does not know which world in $\Omega$ corresponds to the actual world $\omega_0$. Nevertheless, You have some idea, some opinion about which world might be the actual one. So for every subset $A$ of $\Omega$, You can express the strength of Your opinion that the actual world $\omega_0$ belongs to $A$. This strength is denoted belief($A$), and belief means (weighted) opinions. The larger belief($A$), the stronger You believe $\omega_0 \in A$.

Formally, we suppose a finite propositional language **L**, supplemented by the tautology and the contradiction. Extensions to infinite language could be considered but they are useless here as we focus on understanding, not on mathematical generality. Let $\Omega$ denote the set of worlds that correspond to the interpretations of **L**. It is built in such a way that no two worlds in $\Omega$ denote logically equivalent propositions, i.e., for every pair of worlds in $\Omega$, there exists a proposition in the language **L** that is true in one world and false in the other (this avoids useless repetition of worlds denoting logically equivalent propositions).

Among the worlds in $\Omega$, a particular one corresponds to the actual world $\omega_0$. Because the data available to You are imperfect, You do not know exactly which world in a set of possible worlds is the actual world $\omega_0$. All You can express is Your 'opinion', represented by belief($A$) for $A \subseteq \Omega$, about the fact that the actual world $\omega_0$ belongs to the various subsets $A$ of $\Omega$. The major problem is in the choice of the properties that the function 'belief' should satisfy.

We first assume that for every pair of subsets $A$ and $B$ of $\Omega$, belief($A$) and belief($B$) are comparable, i.e.,

$$\text{belief}(A) \leq \text{belief}(B) \text{ or belief}(A) \geq \text{belief}(B).$$

Such an assumption leads easily to the fact that belief($A$) is a real number. It is not so innocuous as it eliminates, among others, the theories based on sets of probability functions as proposed by Kyburg (1987) [18] or Voorbraak (1991) [42].

Monotony with respect to inclusion is the next and quite obvious requirement:

$$\text{if } A \subseteq B \subseteq \Omega \text{ then belief}(A) \leq \text{belief}(B).$$

That belief is non-negative and bounded are normal requirements, and using the closed interval $[0, 1]$ is just as good as any other choice: so, belief: $2^\Omega \rightarrow [0, 1]$ (where $2^\Omega$ denotes the power set of $\Omega$). Other properties are less obvious, and the differences between the models proposed for representing quantified belief result from them.

1.3. **The static and dynamic components.** Any model that wants to represent quantified beliefs has, at least, two components: one, static, that describes Your state of belief given the information available to You, and the other, dynamic, that explains how to update Your beliefs given new pieces of information become available to You. We insist on the fact that both components must be considered in order to compare these models. Unfortunately, too many publications are restricted to the static component and fail to detect the differences between the models based on belief functions. In fact, these differences are essentially observed at the dynamic level.

1.4. **Credal versus pignistic levels.** Intrinsically, belief is not a directly observable property. Once a decision must be made, its impact can be observed. Studying beliefs without considering the decision making process is useless. Any model could be elaborated, but its values could not be assessed if it weren't for the fact that decisions can be observed. So both processes of holding beliefs and making decisions must be described.

We have described a two-level mental model in order to distinguish between two aspects of beliefs, belief as weighted opinions, and belief for decision making. The two levels are: the credal level, where beliefs are entertained, and the pignistic level, where beliefs are used to make decisions (credal and pignistic derive from the latin words '*credo*', I believe and '*pignus*', a wage, a bet).

Usually these two levels are not distinguished and probability functions are used to quantify beliefs at both levels. The justification for the use of probability functions is usually linked to "rational" behavior to be exhibited by an ideal agent involved in some betting or decision contexts. If decisions must be "coherent", the uncertainty over the possible alternatives must be represented by a probability function. This result justifies that a probability function must be used at the pignistic level in order to compute the expected utilities that must be maximized by selecting the 'best' act. Nevertheless that probability function does not represent the agent's beliefs, it is only the function needed to derive the best decision. What is rejected when we distinguish between the two mental levels is the assumption that the probability function used during the decision making process represents the uncertainty at the credal level. We assume that the pignistic and the credal levels are different in which case the justification for using probability functions at the credal level does not hold anymore. At the credal level, we advocate in the transferable belief model that beliefs are represented by belief functions. Of course, we will have to explain and justify how decisions can be achieved, i.e., how to derive the probability function needed at the pignistic level from the belief function that represents Your beliefs at the credal level (see Section 6).

## 2. The transferable belief model

The transferable belief model (TBM) is a model developed to represent quantified beliefs (Smets and Kennes, 1994). It covers the same domain as the Bayesian–subjectivist probabilities except that it is based on belief functions (Shafer, 1976a) [26] (Smets, 1988b) [32] (Smets and Magrez, 1987) [40] (Smets, 1994b) [38], instead of probability functions as usually advocated by the Bayesians.

The TBM departs from the Bayesian approach in that we do not assume the additivity encountered in probability theory. It is replaced by inequalities like:

$$\mathrm{bel}(A \cup B) \geq \mathrm{bel}(A) + \mathrm{bel}(B) - \mathrm{bel}(A \cap B), \tag{1}$$

where $\mathrm{bel}(A)$ denotes belief$(A)$.

In the TBM, one assumes that bel is a 'capacity monotone of order infinity', i.e., bel satisfies the following inequalities:

$$\forall n > 1, \forall A_1, A_2, \ldots A_n \subseteq \Omega, \mathrm{bel}(A_1 \cup A2 \cup \ldots A_n) \geq$$

$$\sum_i \mathrm{bel}(A_i) - \sum_{i>j} \mathrm{bel}(A_i \cap A_j) \cdots - (-1)^n \mathrm{bel}(A_1 \cap A_2 \cap \ldots A_n). \quad (2)$$

As such, the meaning of these inequalities is not obvious, except when $n = 2$ (see (1)).

2.1. **Basic belief assignment.** The understanding of the inequalities (2) is clarified once the concept of a **basic belief assignment** (bba) is introduced. A basic belief assignment (bba) is a function $m\colon 2^\Omega \to [0, 1]$ that satisfies $\sum_{A\colon A \subseteq \Omega} m(A) = 1$. The term $m(A)$ is called the **basic belief mass** (bbm) given to $A$.

The bbm $m(A)$ represents that part of Your belief that supports $A$ — i.e., the fact that the actual world $\omega_0$ belongs to $A$ — without supporting any more specific subset, by lack of adequate information.

As an example, consider that You learn that $\omega_0$ belongs to $A$, and You know nothing else about the value of $\omega_0$. Then some part of Your beliefs will be given to $A$, but no subset of $A$ will get any positive support. In that case, You would have $m(A) > 0$ and $m(B) = 0$ for all $B \subseteq A$ and $B \neq A$, a property that could not be satisfied by a probability measure.

2.2. **Belief functions.** The bbm $m(A)$ does not in itself quantify Your belief, that we denote by $\mathrm{bel}(A)$, that the actual world $\omega_0$ belongs to $A$. Indeed, the bbm $m(B)$ given to some subset $B$ of $A$ also supports that $\omega_0 \in A$. Hence, the degree of belief $\mathrm{bel}(A)$ is obtained by summing all the bbm $m(B)$ for $B \subseteq A$. We have:

$$\mathrm{bel}(\emptyset) = 0 \text{ and } \mathrm{bel}(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \quad \forall A \subseteq \Omega, A \neq \emptyset. \quad (3)$$

The degree of belief $\mathrm{bel}(A)$ for $A \subseteq \Omega$ quantifies the total amount of *justified specific support* given to the fact that the actual world $\omega_0$ belongs to $A$. We say *justified* because we include in $\mathrm{bel}(A)$ *only* the basic belief masses given to subsets of $A$. For instance, consider two distinct elements $x$ and $y$ of $\Omega$. The basic belief mass $m(\{x, y\})$ given to $\{x, y\}$ could support $x$ if further information indicates this. However given the available information the basic belief mass can only be given to $\{x, y\}$. We say *specific* because the basic belief mass $m(\emptyset)$ is not included in $\mathrm{bel}(A)$ as it is given to the subset $\emptyset$ that supports not only $A$ but also its complement $\overline{A}$.

*Note:* Shafer assumes $m(\emptyset) = 0$. In the TBM, such a requirement is not assumed.

The function bel so defined satisfies the inequalities (2). Thanks to the natural interpretation that can be given to the basic belief masses, the meaning of the inequalities (2) becomes somehow clearer. The originality of the model comes from the non-null masses that may be given to non-singletons of $\Omega$. Indeed, when $m(A) = 0$ for all $A \subseteq \Omega$ with $|A| \neq 1$ ($|A|$ denotes the number of elements in $A$), then the inequalities of (2) become equalities, and the function bel is a probability function. In that last case, the TBM reduces to the Bayesian theory. Of course, in general, this requirement is not satisfied in the TBM.

The dual of bel is called a plausibility function $\mathrm{pl}\colon 2^\Omega \to [0, 1]$. It is defined as:

$$\mathrm{pl}(A) = \mathrm{bel}(\Omega) - \mathrm{bel}(\overline{A}), \quad \text{for all } A \subseteq \Omega,$$

or

$$\mathrm{pl}(A) = \sum_{X \subseteq \Omega\colon X \cap A \neq \emptyset} m(X), \quad \text{for all } A \subseteq \Omega.$$

The degree of plausibility pl(A) for $A \subseteq \Omega$ quantifies the maximum amount of potential specific support that could be given to the fact that the actual world $\omega_0$ belongs to A. We say potential because the basic belief masses included in pl(A) could be transferred to non-empty subsets of A if new information could justify such a transfer. It would be the case if we learn that the actual world cannot be in $\overline{A}$.

2.3. **Related functions.** Two other functions related to bel are also defined: the commonality function q: $2^\Omega \rightarrow [0,1]$ and the implicability function b: $2^\Omega \rightarrow [0,1]$ with:

$$q(A) = \sum_{X \subseteq \Omega:\, A \subseteq X} m(X), \quad \text{for all } A \subseteq \Omega \tag{4}$$

$$b(A) = \text{bel}(A) + m(\emptyset) = \sum_{X \subseteq \Omega:\, X \subseteq A} m(X), \quad \text{for all } A \subseteq \Omega. \tag{5}$$

Their major interest will appear when conditioning and combination will be introduced.

It must be emphasized that each of these five functions is in one-to-one correspondence, so none gives information not included in the others. Their interest derives from the fact they enhance different aspects of the same underlying beliefs, and that some are sometime computationally very convenient.

The advantage of the TBM over the classical Bayesian approach resides in its large flexibility, its ability to represent every state of partial beliefs, up to the **state of total ignorance**. In the TBM, total ignorance is represented by the so-called **vacuous belief function**, i.e., a belief function such that $m(\Omega) = 1$, $m(A) = 0$ for all A with $A \subseteq \Omega$, $A \neq \Omega$. Hence $\text{bel}(\Omega) = 1$ and $\text{bel}(A) = 0$ for all A with $A \subseteq \Omega$, $A \neq \Omega$. It expresses that all You know is that $\omega_0 \in \Omega$. The representation of total ignorance in probability theory is hard to achieve adequately, most proposed solutions being doomed to contradictions. With the TBM, we claim we+ can represent every state of belief, full ignorance, partial ignorance, probabilistic-additive beliefs, or even full belief ($m(A) = 1$ corresponds to A is certain).

**Example 1.** Let us consider a somewhat reliable witness in a murder case who testifies to You that the killer is male. Let $\alpha = .7$ be the reliability You give to the testimony. Suppose furthermore that *a priori* You have an equal belief that the killer is male or female.

A classical probability analysis would compute the probability $P(M)$ of $M =$ 'the killer is male' given the witness testimony as:

$$P(M) = P(M|\text{Reliable})P(\text{Reliable}) + P(M|\text{NotReliable})P(\text{NotReliable})$$
$$= 1 \times 0.7 + 0.5 \times 0.3 = 0.85,$$

where Reliable and Not Reliable refer to the witness' reliability. The value 0.85 is the sum of the probability of M given that the witness is reliable (1) weighted by the probability that the witness is reliable (0.7) plus the probability of M given that the witness is not reliable (0.5) weighted by the probability that the witness is not reliable (0.3).

The TBM analysis is different. It goes as follows. You have some reason to believe that the killer is a male, as the witness said so. But this belief is not total (maximal) as the witness might be wrong. The 0.7 is the belief You give to the fact that the witness tells the truth (is reliable), in which case the killer is male. The remaining 0.3 mass is given to the fact that the witness is not really telling the truth (he lies or he might have seen a male, but this was not the killer). In

that last case, the testimony does not tell You anything about the killer's sex. So the TBM analysis will give a belief 0.7 to $M$: $\mathrm{bel}(M) = 0.7$ (and $\mathrm{bel}(\mathrm{Not}\ M) = 0$). The information relative the population of killers is not relevant to Your problem. The fact that almost all crimes are committed by the members of some particular group of individuals may not be used in court. It does not allow You (nor any judge) to conclude to the culpability of a suspect.

In the probability analysis where $P(M) = 0.7 + 0.15$, the 0.7 value can be viewed as the justified component of the probability given to $M$ (called the belief or the support) whereas the 0.15 value can be viewed as the random component of that probability. It would become relevant to $\mathrm{bel}(M)$ only if the murderer had been really selected by a random process from a population where 50% are male. In our example, such a random selection does not apply, so the random component is not considered when building Your beliefs. The transferable belief model deals only with the justified components.

2.4. **Conditioning.** Suppose You have some belief on $\Omega$ represented by the basic belief assignment $m$. Then some further evidence becomes available to You and this piece of information implies that the actual world cannot be in one of the worlds in $\overline{A}$. Then the mass $m(B)$ that initially was supporting that the actual world is in $B$ now supports that the actual world is in $B \cap A$ as every world in $\overline{A}$ must be 'eliminated'. So $m(B)$ is transferred to $B \cap A$ after conditioning on $A$. (The model gets its name from this transfer operation.)

This operation leads to the conditional basic belief assignment $m_A$, belief function $\mathrm{bel}_A$ and plausibility function $\mathrm{pl}_A$ with:

$$m_A(B) = \sum_{C \subseteq \overline{A}} m(B \cap C)$$

$$\mathrm{bel}_A(B) = \mathrm{bel}(B \cup \overline{A}) - \mathrm{bel}(\overline{A}), \quad \mathrm{pl}_A(B) = \mathrm{pl}(B \cap A), \quad \mathrm{b}_A(B) = \mathrm{b}(B \cup \overline{A})$$

$$\mathrm{q}_A(B) = \begin{cases} \mathrm{q}(B) & \text{if } B \subseteq A \\ 0 & \text{otherwise} \end{cases}$$

The rule by which these equations are built is called Dempster's rule of conditioning.

*Note:* As already mentioned above, an important difference between the TBM and the model introduced by Shafer is that we do not require that $m(\emptyset) = 0$, or equivalently that $\mathrm{bel}(\Omega) = 1$. Its origin can be explained in two ways: the open world assumption (Smets, 1992c) [36] and the quantified conflict (Smets, 1988b) [32]. This difference explains why we do not normalize (we omit the denominator in) Dempster's rules of conditioning (and combination, see Section 4.1).

The open world assumption reflects the idea that $\Omega$ might not contain the actual world. We do not consider such a generalization in this contribution.

The second interpretation of $m(\emptyset) > 0$ is that there is some underlying conflict between the sources that are combined in order to produce the bba $m$. Consider for example a bba $m_0$ defined on $\Omega$ with $m_0(\emptyset) = 0$ and $\mathrm{bel}_0(\overline{A}) > 0$, where $\overline{A}$ is the complement of $A$ relative to $\Omega$. Suppose you collect another piece of evidence, a conditioning one, that just states that $A$ is true for sure. Its related bba is given by $m_A$ with $m_A(A) = 1$. You had given some belief to $\overline{A}$ and now you learn that no belief should have been given to $\overline{A}$. So a conflict appears between the first belief $\mathrm{bel}_0$ and the new one $m_A$. The larger $\mathrm{bel}_0(\overline{A})$, the larger the conflict. The worst conflict between two pieces of evidence would be encountered if $\mathrm{bel}_0(\overline{A}) = 1$, as it means that you were sure that $\overline{A}$ holds and now you learn for sure that it is false. This leads to a contradiction, the conflict encountered in classical logic. The best case would be $\mathrm{bel}_0(\overline{A}) = 0$, in which case there is no conflict between $\mathrm{bel}_0$ and the

new piece of information. After combining the two pieces of information, we get a new bba $m = m_0 \oplus m_A$ with $m(\emptyset) = \mathrm{bel}_0(\overline{A})$ (see also Section 4.1). So $m(\emptyset)$ can be understood as the amount of conflict between $m_0$ and the conditioning evidence represented by $m_A$. This can be generalized to any pair of belief functions, and we can understand $m(\emptyset)$ as the amount of conflict present in $m$, and that results from the pieces of evidence that were taken into account when building $m$.

**Example 1: Continuation.** Continuing with the murder case, suppose there are only two potential male suspects: Phil and Tom, so $m(\{\mathrm{Phil}, \mathrm{Tom}\}) = 0.7$. Then You learn that Phil is not the killer. The initial testimony now supports that the killer is Tom. The reliability $\alpha = 0.7$ You gave to the testimony initially supported 'the killer is Phil or Tom'. The new information about Phil implies that the value $0.7$ now supports 'the killer is Tom', hence $m_{\mathrm{not\_Phil}}(\{\mathrm{Tom}\}) = 0.7$.

2.5. **The vacuous belief function.** Total ignorance is represented by the vacuous belief function, i.e. a belief function such that $m(\Omega) = 1$, hence $\mathrm{bel}(A) = 0$ for all $A \subseteq \Omega$, $A \neq \Omega$, and $\mathrm{bel}(\Omega) = 1$. The origin of this particular quantification for representing a state of total ignorance can be justified. Suppose that there are three propositions written on three pieces of paper labeled $A$, $B$ and $C$. You know that one and only one of these propositions is true, but You even do not know what the propositions are. You only know their number and the labels written on the papers. By symmetry arguments, Your beliefs about their truth are equal: $\mathrm{bel}(A) = \mathrm{bel}(B) = \mathrm{bel}(C) = \alpha$ for some $\alpha \in [0, 1]$ where $\mathrm{bel}(A)$ is the belief You give to the fact that the proposition written on the paper labeled $A$ is true, ... . Furthermore, You have no reason to put more (or less) belief on $C$ than on $A \cup B$ (that one of the propositions written on $A$ or on $B$ is true): $\mathrm{bel}(A \cup B) = \mathrm{bel}(C) = \alpha$. This requirement leads to the equalities: $\mathrm{bel}(A \cup B) = \mathrm{bel}(A) = \mathrm{bel}(B) = \alpha$. The vacuous belief function ($\alpha = 0$) is the only belief function that satisfies these equalities (and no probability function can satisfy simultaneously these equalities and those obtained by symmetry).

2.6. **The Principle of Minimal Commitment.** Let Me (or I) be another agent, different from You. Suppose You only know that My belief function over $\Omega = \{a, b, c\}$ is such that $\mathrm{bel}_{\mathrm{Me}}(\{a\}) = 0.3$ and $\mathrm{bel}_{\mathrm{Me}}(\{b, c\}) = 0.5$, and You do not know the value given to $\mathrm{bel}_{\mathrm{Me}}$ for the other subsets of $\Omega$. Suppose You have no other information on $\Omega$ and You are ready to adopt My belief as Yours. How to build Your beliefs given these partial constraints? Many belief functions can satisfy them. If you adopt the principle that subsets of $\Omega$ should not get more support than justified, then Your belief on $\Omega$ will be such that $m_{\mathrm{You}}(\{a\}) = 0.3$, $m_{\mathrm{You}}(\{b, c\}) = 0.5$ and $m_{\mathrm{You}}(\{a, b, c\}) = 0.2$. Among all belief functions compatible with the constraints given by known values of $\mathrm{bel}_{\mathrm{Me}}$, $\mathrm{bel}_{\mathrm{You}}$ is the one that gives the smallest degree of belief to every subset of $\Omega$. The principle evoked here is called the **Principle of Minimal Commitment**. It is really at the core of the TBM, where degrees of belief are degrees of 'justified' supports.

With un-normalized belief functions (where $m(\emptyset)$ can be positive), the definition of the principle is based on the plausibility functions. Consider two plausibility functions $\mathrm{pl}_1$ and $\mathrm{pl}_2$ such

$$\mathrm{pl}_1(A) \leq \mathrm{pl}_2(A) \quad \forall A \subseteq \Omega.$$

We say that $\mathrm{pl}_2$ is not more committed than $\mathrm{pl}_1$ (and less committed if there is at least one strict inequality). The same qualification is extended to their related basic belief assignments and belief functions. Among all belief functions on $\Omega$, the least committed belief function is the vacuous belief function where $\mathrm{pl}(A) = 1$ for all $A \neq \emptyset$.

When expressed with belief functions, the principle becomes:

$$b_1(A) \geq b_2(A) \quad \forall A \subseteq \Omega$$

$$\text{i.e.,} \quad \text{bel}_1(A) + m_1(\emptyset) \geq \text{bel}_2(A) + m_2(\emptyset) \quad \forall A \subseteq \Omega$$

The concept of 'least commitment' permits the construction of a partial order $\sqsubseteq$ on the set of belief functions (Yager, 1986) [46] (Dubois and Prade, 1986) [5].

We write:

$$\text{pl}_1 \sqsubseteq \text{pl}_2$$

to denote that $\text{pl}_1$ is equal to or more committed than $\text{pl}_2$. By analogy the derived notations $m_1 \sqsubseteq m_2$ and $\text{bel}_1 \sqsubseteq \text{bel}_2$ have the same meaning.

The **Principle of Minimal Commitment** consists in selecting the least committed belief function in a set of equally justified belief functions. The principle formalizes the idea that one should never give more support than justified to any subset of $\Omega$. It satisfies a form of skepticism, of uncommitment, of conservatism in the allocation of our belief. In its spirit, it is not far from what the probabilists try to achieve with the maximum entropy principle, see (Dubois and Prade, 1987) [4] (Hsia, 1991) [11].

This principle does not work in every situation. The set of belief functions compatible with a given set of constraints does not always admit a unique least committed solution. In that case extra requirements must be introduced, like those based on the information content of a belief function (Klir, 1987) [15] (Klir and Wierman, 1998) [14] (Pal et al., 1992, 1993) [20, 21] (Smets, 1983) [30].

2.7. **The Generalized Bayesian Theorem.** Consider the finite spaces $X$ and $\Theta$. Suppose that for each $\theta \in \Theta$, there is a basic belief assignment on $X$, denoted $m^X[\theta]$. Given this set of basic belief assignments, what is the belief induced on $\Omega$ if You come to know that $x \subseteq X$ holds?

In statistics, this is the classical inference problem where $\Theta$ is a parameter space and $X$ an observation space. You know the probability functions on $X$ given each $\theta$ in $\Theta$, You observe $x \subseteq X$ and You compute, by applying Bayes' theorem, the probability function on $\Theta$ given $x$, using also an *a priori* probability function on $\Theta$.

The Generalized Bayesian Theorem (GBT) performs the same task with belief functions. The main point is that the needed prior can be a vacuous belief function, which is the perfect representation of total ignorance. No informative prior belief is needed, avoiding thus one of the major criticisms against the classical approach, in particular when used for diagnostic applications.

Given the set of basic belief assignments $m^X[\theta]$ known for every $\theta \in \Theta$ and their related b-functions $b^X[\theta]$ and plausibility functions $\text{pl}^X[\theta]$, then for $x \subseteq X$ and for every $A \subseteq \Theta$:

$$b^\Theta[x](A) = \prod_{\theta \in \Theta \setminus A} b^X[\theta](\overline{x})$$

$$\text{pl}^\Theta[x](A) = 1 - \prod_{\theta \in A} (1 - \text{pl}^X[\theta](x))$$

$$q^\Theta[x](A) = \prod_{\theta \in A} \text{pl}^X[\theta](x)$$

where $[x]$ is the piece of evidence stating that '$x$ holds'.

Should You have some non-vacuous beliefs on $\Theta$, represented by $m^\Theta[E_0]$, then this belief is simply combined with $m^\Theta[x](A)$ by the application of the conjunctive rule of combination (see Section 4.1).

This rule has been derived axiomatically by Smets (1978, 1993) [29, 37] and by Appriou (1991) [1].

*Some particular cases are worth mentioning.* 1. We consider the case of two 'independent' observations $x$ defined on $X$ and $y$ defined on $Y$, and the inference on $\Theta$ obtained from their joint observation.

Suppose the two variables X and Y satisfy the **Conditional Cognitive Independence** property defined as:

$$\mathrm{pl}^{X \times Y}[\theta](x, y) = \mathrm{pl}^{X}[\theta](x)\mathrm{pl}^{Y}[\theta](y) \quad \forall x \subseteq X, \forall y \subseteq Y, \forall \theta \in \Theta.$$

This property is a generalization of the classical independence described in probability theory. It reduces to the classical conditional independence property when all plausibility functions are probability functions.

Let $\mathrm{b}^{\Theta}[x]$ and $\mathrm{b}^{\Theta}[y]$ be computed by the GBT (with a vacuous *a priori* belief on $\Theta$) from the set of basic belief assignment $m^{X}[\theta]$ and $m^{Y}[\theta]$ known for every $\theta \in \Theta$. We then combine by the conjunctive rule of combination (see Section 4.1) these two functions in order to build the belief $\mathrm{b}^{\Theta}[x, y]$ on $\Theta$ induced by the pair of observations.

We could as well consider the basic belief assignment $m^{X \times Y}[\theta]$ built on the space $X \times Y$ thanks to the Conditional Cognitive Independence property, and compute $\mathrm{b}^{\Theta}[x, y]$ from it using the GBT. Both results are the same, a property that is essential and at the core of the axiomatic derivation of the rule.

2. If for each $\theta \in \Theta$, $\mathrm{b}^{X}[\theta]$ is a **probability function** $P(\cdot|\theta)$ **on** $X$, then the GBT for $|\theta| = 1$ becomes:

$$\mathrm{pl}^{\Theta}[x](\theta) = P(x|\theta) \quad \forall x \subseteq X$$

That is, on the singletons $\theta$ of $\Theta$, $\mathrm{pl}^{\Theta}[x]$ reduces to the likelihood of $\theta$ given $x$. The analogy stops there as the solution for the likelihood of subsets of $\Theta$ is different.

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

$$\mathrm{bel}^{\Theta}[x](A) = \frac{\sum_{\theta \in A} P(x|\theta)P_0(\theta)}{\sum_{\theta \in \Theta} P(x|\theta)P_0(\theta)} = P(A|x)$$

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

## 3. CONDITIONING

Besides the representation of beliefs by a belief function, the most characteristic component of the TBM is the use of Dempster's rule of conditioning to represent the impact of a conditioning piece of evidence. Conditioning Your belief on $A \subseteq \Omega$ means that You accept as true the conditioning piece of evidence $Ev_A$ that states that the actual world is not one of those in $\overline{A}$. This piece of evidence usually means that the actual world is in $A$, but this last expression states a little more than the previous one. It states that the actual world is really in $\Omega$, what is not necessarily required in the TBM; see (Smets, 1988b) [32] where the difference between the open and the closed world assumptions is described, and (Smets, 1992c) [36], where the meaning of $m(\emptyset) > 0$ is analysed.

Given the importance of the conditioning process, we present an example that illustrates its nature. Differentiating between factual and generic revisions is used, details can be found in Dubois and Prade (1994) [6] and Smets (1998) [28].

**Example 2: Failure diagnosis.** Suppose some electric equipment has failed and You know that one and only one circuit has failed. There are two types of circuits, the $A$- and the $B$-circuits made at the $F_A$ and $F_B$ factories, respectively. You know that circuits made at factory $F_A$ are of high quality whereas those made at factory $F_B$ are of a lower quality. Hence You might have good reasons to believe

that the broken circuit is a $B$-circuit, even though it might be an $A$-circuit. Let $\mathrm{bel}_0$ represents Your belief about which type of circuit is broken, with $\mathrm{bel}_0(A)$ and $\mathrm{bel}_0(B)$ being the degree of belief given by You to the fact that the broken circuit is an $A$- or a $B$-circuit, respectively. $\mathrm{bel}_0$ is built on the frame $\Omega = \{A, B\}$.

Then You learn that the $A$-circuits are painted in green ($G$) and the $B$-circuits are painted in white ($W$) and pink ($P$). Let $2^{\Omega'}$ be the power set of a new frame of discernment $\Omega' = \{G, W, P\}$. Let $R$ denote the operator that transforms the elements of $\Omega$ into those of $\Omega'$. $R$ is called a refinement.

**Refinement**: A refinement $R$ is a mapping from a frame $\Omega$ to a frame $\Omega'$ such that 1) each element of $\Omega$ is mapped into one or several elements of $\Omega'$, 2) each element of $\Omega'$ is derived from one and only one element of $\Omega$, and 3) $R(\emptyset) = \emptyset$.

By construction, we have in the example: $R(A) = G$ and $R(B) = \{P, W\}$.

For You, the color has nothing to do with failure (as far as You know), thus from Your point of view, $R$ is an uninformative refinement, i.e., $R$ reflects only a change of the frame on which beliefs are held. Let $\mathrm{bel}'$ quantify Your beliefs about the color of the broken circuit. The uninformative nature of the refinement $R$ is reflected by the fact that we require:

$$\mathrm{bel}'(G) = \mathrm{bel}_0(A), \ \mathrm{bel}'(P \cup W) = \mathrm{bel}_0(B) \text{ and } \mathrm{bel}'(G \cup P \cup W) = \mathrm{bel}_0(A \cup B),$$

i.e., subsets denoting equivalent propositions receive the same beliefs.

**Example 2: The conditioning process.** We consider now what happens to $\mathrm{bel}_0$ and $\mathrm{bel}'$ when new pieces of evidence become available to You. We restrict ourselves to the form of changes encountered in probability theory, i.e., conditioning. Two forms of revisions are presented. Each one must be considered independently of the other.

*Example 2: Generic Revision.* You learn that none of the circuits made at factory $F_B$ used in the failed equipment was painted pink. This piece of evidence is denoted $Ev_1$. So knowing that the circuit has been made at factory $F_B$ is now equivalent to knowing that the circuit is white. The impact of the conditioning information $Ev_1$ results in a transformation of $\mathrm{bel}'$ into a new belief represented by $\mathrm{bel}_1$. To keep adequate coherence between the various beliefs, $\mathrm{bel}_1$ must satisfy certain constraints. We have now:

$$\mathrm{bel}_1(G) = \mathrm{bel}_0(A), \ \mathrm{bel}_1(P) = 0, \ \mathrm{bel}_1(W) = \mathrm{bel}_0(B)$$
$$\mathrm{bel}_1(G \cup P) = \mathrm{bel}_0(A), \ \mathrm{bel}_1(G \cup W) = \mathrm{bel}_0(A \cup B), \ \mathrm{bel}_1(P \cup W) = \mathrm{bel}_0(B)$$

and

$$\mathrm{bel}_1(G \cup P \cup W) = \mathrm{bel}_0(A \cup B). \tag{6}$$

*Example 2: Factual Revision.* Suppose we are in the situation as described in the beginning of this section, so the revision information $Ev_1$ considered above is not taken into consideration. Instead, You possess a fully reliable sensor that is only able to detect whether the color of a circuit is pink or not, so it cannot distinguish between green and white circuits. You learn that Your sensor has been applied to the broken circuit and has reported that the broken circuit is not pink. This new piece of evidence is denoted $Ev_2$.

Under $Ev_2$, $B$ and $W$ denote equivalent propositions, as knowing that the broken circuit has been made at factory $F_B$ is now equivalent to knowing that the broken circuit is white. Let $\mathrm{bel}_2$ be the belief function obtained after conditioning $\mathrm{bel}'$ on $Ev_2$. To keep adequate coherence between the various beliefs, $\mathrm{bel}_2$ must satisfy

certain constraints. We have now:

$$\text{bel}_2(G) = \text{bel}_0(A), \ \text{bel}_2(P) = 0, \text{bel}_2(W) = bel_0(B)$$

$$\text{bel}_2(G \cup P) = \text{bel}_0(A), \ \text{bel}_2(G \cup W) = \text{bel}_0(A \cup B), \ \text{bel}_2(P \cup W) = \text{bel}_0(B)$$

and

$$\text{bel}_2(G \cup P \cup W) = \text{bel}_0(A \cup B). \tag{7}$$

The constraints on $\text{bel}_1$ and $\text{bel}_2$ happen to be identical. The difference between the pieces of evidence that lead to the two revisions resides in the fact that the second concerns only the broken circuit, whereas the first concerns all circuits made at factory $F_B$. *But as far as Your beliefs concern only the broken circuit, the two cases are equivalent for the problem You try to solve.*

The two cases would be different if You had to select one circuit at random and bet on its color. In the case of generic revision, You would start with some probability that the circuit that will be selected has been made in $F_A$ (in $F_B$). Then learning about the three colors, You would build a probability measure over the three colors. Finally, learning that all $B$-circuits are in fact white, You would reassess Your beliefs over the two remaining colors and obtain the same solution as we obtained (6). What You had built over the three colors was based on the assumption that there were three colors, an assumption that turns out to be erroneous, and thus probabilities must be reassessed from scratch, i.e., from the state of belief You had before learning about the three colors. The generic revision solution is uncontroversial and will not be further discussed.

In the case of factual revision, You would build the same probability measure over the three colors as in the previous case. Then You would learn that the selected circuit is not pink. You would condition Your beliefs over the two remaining colors through the Bayesian conditioning rule. This is where the Bayesian model seems to diverge from ours.

But these are not the stories we are considering. In the factual revision case, we do not have any underlying random selection: there is a broken circuit and we learn some information about it. For instance, does the information about the colors and the fact that the broken circuit happens not to be painted in pink give any reason to modify Your belief that the broken circuit is an $A$-circuit? We don't think so. You had some reasons to believe that the broken circuit was an $A$-circuit, and the factual information should not change Your beliefs about it, i.e.,

$$\text{bel}_2(G) = \text{bel}_0(A),$$

By a similar argument, we get the equalities (7).

The Bayesian solution is also derived from the TBM once the random element has been taken in due consideration. The analysis that leads to the equalities (7) is not based on the assumption of any underlying random selection. It translates the process by which we try to discover which circuit is broken and we analyze the impact of new information.

The mathematical consequences of the equalities (7) are enormous. They almost imply the mathematical structure of both the conditioning and the uninformative refinement processes. In the example we get:

- for the uninformative refinement process:

$$\text{bel}'(P) = \text{bel}'(W) = 0,$$

$$\text{bel}'(G \cup P) = \text{bel}'(G \cup W) = \text{bel}_0(A), \qquad \dots$$

- for the conditioning process:

$$\mathrm{bel}_2(G) = \mathrm{bel}'(G \cup P) - \mathrm{bel}'(P) = \mathrm{bel}_0(A)$$

$$\mathrm{bel}_2(W) = \mathrm{bel}'(W \cup P) - \mathrm{bel}'(P) = \mathrm{bel}_0(B), \qquad \ldots$$

The solution for the conditioning process is the so-called Dempster's rule of conditioning encountered in belief functions-based theories. It is the solution described in the TBM and Dempster-Shafer theory, and detailed in Section 2.4.

## 4. General combination rules

### 4.1. **Non-interactive combinations.**

*Notation:* let the symbols between [ and ] denote the pieces of evidence taken in consideration when building Your belief function on a frame $\Omega$. So $\mathrm{bel}[E_1 \wedge E_2]$ is the belief function built when both the pieces of evidence $E_1$ and $E_2$ are taken in consideration. $\mathrm{bel}[E_1](A)$ denotes then the value of the belief function $\mathrm{bel}[E_1]$ at $A \subseteq \Omega$.

All pieces of evidence are considered as 'distinct', an ill-defined concept but some justifications can be found in Shafer and Tversky (1985) [27] and Smets (1990a) [33]. Mathematically, it means that $\mathrm{bel}[E_1 \wedge E_2]$ is a function of $\mathrm{bel}[E_1]$ and $\mathrm{bel}[E_2]$.

Let $m[E_1]$ and $m[E_2]$ be two basic belief assignments on $2^\Omega$, induced by the distinct pieces of evidence $E_1$ and $E_2$, respectively. Then:

$$m[E_1 \wedge E_2](C) = \sum_{A \subseteq \Omega, B \subseteq \Omega, A \cap B = C} m[E_1](A)m[E_2](B) \qquad \text{for all } C \subseteq \Omega,$$

in which case:

$$\mathrm{q}[E_1 \wedge E_2](A) = \mathrm{q}[E_1](A)\mathrm{q}[E_2](A) \qquad \text{for all } A \subseteq \Omega.$$

This rule is called *Dempster's rule of combination.* It corresponds to a conjunctive combination rule, as it results from the conjunction of the two pieces of evidence. We also denote use the following standard notation: $m[E_1 \wedge E_2] = m[E_1] \oplus m[E_2]$.

Other rules can be defined.

A disjunction of two pieces of evidence could occur if You only know that the disjunction '$E_1$ or $E_2$' holds. Then:

$$m[E_1 \vee E_2](C) = \sum_{A \subseteq \Omega, B \subseteq \Omega, A \cup B = C} m[E_1](A)m[E_2](B) \qquad \text{for all } C \subseteq \Omega,$$

in which case:

$$\mathrm{b}[E_1 \vee E_2](A) = \mathrm{b}[E_1](A)\mathrm{b}[E_2](A) \qquad \text{for all } A \subseteq \Omega.$$

This rule is called the disjunctive rule of combination (Smets, 1993b) [37].

The concept of negation has also been described. Let $\breve{E}$ denote the fact that the source of the piece of evidence $E$ lies, what means that whenever it says that the actual world belongs to $A$, it means that it belongs to $\overline{A}$. One has:

$$m[\breve{E}](A) = m[E](\overline{A}) \qquad \text{for all } A \subseteq \Omega,$$

in which case: $\mathrm{b}[\breve{E}](\overline{A}) = \mathrm{q}[E](A)$.

Finally the exclusive disjunctive rule of combination can also be defined. It applies when You accept that one source of evidence tells the truth and the other lies, but You don't know which is which. We have:

$$m[E_1 \underline{\vee} E_2](C) = \sum_{A \subseteq \Omega, B \subseteq \Omega, A \underline{\cup} B = C} m[E_1](A)m[E_2](B) \qquad \text{for all } C \subseteq \Omega.$$

where $\underline{\vee}$ denotes the exclue disjunction and $\underline{\cup}$ denotes the disjoint union, or symmetric difference.

These three rules can be extended to any number of pieces of evidence and any combination formula that states which source You accept as telling the true. So let $E_1$, $E_2$, $\ldots E_n$ be a set of pieces of evidence produced by the sources $S_1$, $S_2$, $\ldots S_n$, with $\mathrm{bel}[E_i]$, $i = 1, 2, \ldots n$, being the belief functions induced by each piece of evidence individually. Suppose the pieces of evidence are **non-interactive**, maybe a better name than distinct, and by which we mean only that the belief function built from the combination of the pieces of evidence is a function of the individual belief functions $\mathrm{bel}[E_i]$). For instance, suppose all You accept is that $(E_1 \wedge E_2) \vee E_3)\underline{\vee}(E_4 \wedge E_1)$ holds. It means You accept that one and only one of the two following cases holds: $(E_1 \wedge E_2) \vee E_3$ or $E_4 \wedge E_1$. In the first case, You accept that at least one of the next two cases holds: $(E_1 \wedge E_2)$ or $E_3$. It means that You accept that either $S_1$ and $S_2$ tell the truth or $S_3$ tells the truth, in a non-exclusive way. In the second case, You accept that both $S_4$ and $S_1$ tell the truth. Given this complex piece of evidence, the basic belief assignment related to the belief function $\mathrm{bel}[((E_1 \wedge E_2) \vee E_3)\underline{\vee}(E_4 \wedge E_1)]$ is:

$$m[((E_1 \wedge E_2) \vee E_3)\underline{\vee}(E_4 \wedge E_1)](A)$$
$$= \sum_{X,Y,Z,T \subseteq \Omega, ((X \cap Y) \cup Z)\underline{\cup}(T \cap X) = A} m[E_1](X)m[E_2](Y)m[E_3](Z)m[E_4](T)$$

for all $A \subseteq \Omega$. This result was known for long (e.g., Dubois and Prade, 1986) [5].

Further generalization has been achieved in Smets (1997a) [39] where conjunctions and disjunctions are even weighted. The practical meaning of these so-called $\alpha$-junctions is still unclear.

4.2. **Cautious combinations.** Consider 6 agents (You, $\mathrm{You}_1$ and $\mathrm{You}_2$, $\mathrm{Witness}_1$, $\mathrm{Witness}_2$, $\mathrm{Witness}_3$). The three witnesses are three distinct sources of evidence on $\Omega$. Let $\mathrm{bel}[W_i]$, $i = 1, 2, 3$, be the belief functions built on $\Omega$ by each witness (denoted $W_i$). Suppose $\mathrm{You}_1$ collects the beliefs of $W_1$ and $W_2$, and these were his only sources of beliefs over $\Omega$. Let $\mathrm{bel}_1$ be the belief function built by $\mathrm{You}_1$ by conjunctively combining $\mathrm{bel}[W_1]$ and $\mathrm{bel}[W_2]$, so $\mathrm{bel}_1 = \mathrm{bel}[W_1 \wedge W_2]$. Similarly let $\mathrm{bel}_2 = \mathrm{bel}[W_2 \wedge W_3]$.

Then You collect the beliefs produced by $\mathrm{You}_1$ and $\mathrm{You}_2$. If You blindly apply Dempster's rule of combination on $\mathrm{bel}_1$ and $\mathrm{bel}_2$, what You get is $\mathrm{bel}[W_1 \wedge W_2 \wedge W_2 \wedge W_3]$, whereas You should have computed $\mathrm{bel}[W_1 \wedge W_2 \wedge W_3]$ (Note that Dempster's rule of combination is not idempotent: $\mathrm{bel}[E \wedge E] \neq \mathrm{bel}[E]$).

Should You know the $\mathrm{bel}[W_i]$'s, You would have just combined them in order to compute $\mathrm{bel}[W_1 \wedge W_2 \wedge W_3]$, but the real situation we want to model is the case where You know there is some evidence that has been used by both $\mathrm{You}_1$ and $\mathrm{You}_2$, but You don't know which one.

Being cautious, and applying somehow the least commitment principle, what You can do is to compute the set $B$ of belief functions that could be obtained from $\mathrm{bel}_1$ and $\mathrm{bel}_2$ through their combination with some belief functions:

$$B = \left\{ \mathrm{bel} \colon \exists \mathrm{bel}', \mathrm{bel}'' \text{ on } \Omega, \mathrm{bel} = \mathrm{bel}_1 \oplus \mathrm{bel}' = \mathrm{bel}_2 \oplus \mathrm{bel}'' \right\}.$$

Then applying the least commitment principle, You might find the 'minimal' element of $B$. Special solutions have already been derived; the general solution is based on the information content of a belief function (Smets, forthcoming).

## 5. SPECIALIZATION

The concept of specialization is at the core of the transferable belief model. Let $m_0$ be the basic belief assignment that represents Your belief on $\Omega$. The impact of a new piece of evidence $Ev$ induces a change in $m_0$ characterized by a redistribution

of $m_0$ such that $m_0(A)$ is distributed among the subsets of $A$. Let $s(B, A) \in [0, 1]$ be the proportion of $m(A)$ that flows into $B \subseteq A$ when You learn the new piece of evidence $Ev$. In order to conserve the whole mass $m(A)$ after this transfer, the coefficients $s(B, A)$ must satisfy:

$$\sum_{B \subseteq \Omega} s(B, A) = 1 \qquad \forall A \subseteq \Omega$$

As masses can flow only to subsets, $s(B, A) = 0$, $\forall B \not\subseteq A$. The matrix $\mathbf{S}$ of the coefficients $s(B, A)$ for $A, B \subseteq \Omega$ is called a specialization matrix on $\Omega$; see (Yager, 1986) [46] (Dubois and Prade, 1986) [5], (Kruse and Schwecke, 1990) [17] (Moral, 1985) [19] (Delgado and Moral, 1987) [2]. In order to simplify the notation, we switch to the classical matrix notation. By convention the lines and columns of the matrices and the elements of the vectors are ordered as follow: $\emptyset$, $\{a\}$, $\{b\}$, $\{a, b\}$, $\{c\}$, $\{a, c\}$, $\{b, c\}$, $\{a, b, c\}$, $\{d\}$, $\{a, d\}$, etc. The vectors whose components are the values of a basic belief assignment, belief function, plausibility function, commonality function, implicability function are vertical vectors denoted $m$, bel, pl, q, b, respectively.

After learning $Ev$, the basic belief assignment $m_0$ is transformed into the new basic belief assignment $m$ such that:

$$m(B) = \sum_{A \subseteq \Omega} s(B, A) m_0(A) \qquad \text{for all } B \subseteq \Omega.$$

or in matricial notation,

$$m = \mathbf{S} \cdot m_0.$$

The basic belief assignment $m$ is called a specialization of $m_0$.

Whenever a bba $m$ is a specialization of a bba $m_0$, then $m$ is at least as committed as $m_0$ (Yager, 1986) [46]. So $\mathbf{S} \cdot m_0 \sqsubseteq m_0$ for any bba $m_0$ and any specialization matrix $\mathbf{S}$ (all defined on $\Omega$).

It is easy to show that the effects of both Dempster's rules can be obtained by specialization matrices.

Let $m \colon 2^{\Omega} \to [0, 1]$ be a bba and $\Sigma$ be the set of specialization matrices on $\Omega$.

**1. Dempster's rule of conditioning:** let $C \subseteq \Omega$ and $\mathbf{S}_C$ be the specialization matrix such that

$$s_C(B, A) = \begin{cases} 1 & \text{if } B = A \cap C \\ 0 & \text{otherwise.} \end{cases}$$

Let $m_C$ be the bba obtained after conditioning the bba m on $C$ by Dempster's rule of conditioning. Then $m_C = \mathbf{S}_C \cdot m$. We call $\mathbf{S}_C$ the $C$-conditionating specialization matrix. We define $\Sigma_{cond} \subseteq \Sigma$ as the set of $C$-conditionating specializations where $C \subseteq \Omega$.

**2. Dempster's rule of combination:** let $m$ be a bba on $\Omega$ and let $\mathbf{S}_m$ be a specialization matrix with coefficients

$$s_m(B, A) = m_A(B) \qquad \forall A, B \subseteq \Omega$$

where $m_A$ is the bba obtained after conditioning the bba $m$ on $A$ by Dempster's rule of conditioning. The coefficients $s_m(B, A)$ satisfy:

$$s_m(B, \Omega) = m(B) \qquad \forall B \subseteq \Omega$$

and $\forall A \subset \Omega$,

$$\begin{cases} s_m(B, A) = \sum_{X \subseteq \overline{A}} s_m(B \cup X, \Omega) & \text{if } B \subseteq \Omega \\ 0 & \text{otherwise} \end{cases}$$

Consider two bba $m$ and $m'$. One can prove that $m' \oplus m = \mathbf{S}_m \cdot m'$.

$\mathbf{S}_m$ is called the Dempsterian specialization matrix associated with $m$ as it updates any $m'$ on $\Omega$ into $m \oplus m'$. We define $\Sigma_D \subseteq \Sigma$ as the set of the Dempsterian specialization matrices.

In particular, if $m$ is such that $m(C) = 1$ (the bba that corresponds to a conditioning on $C$), then $\mathbf{S}_C = \mathbf{S}_m$, so $\Sigma_{cond} \subseteq \Sigma_D$.

We consider that the expansion procedure (the revision of a belief by a new piece of evidence) achieved by a specialization is one of the fundamental ideas for the dynamic part of the transferable belief model. Accepting that every expansion is defined by a specialization matrix, we have shown that:

(1) When conditioning on $A \subseteq \Omega$, $\mathbf{S}_A \in \Sigma_{cond}$ is the specialization matrix that induces the least committed revised plausibility on $\Omega$ such that the updated plausibility given to $\overline{A}$ is 0. The requirement $\mathrm{pl}(\overline{A}) = 0$ after expansion translates the fact that all elements of $\Omega$ in $\overline{A}$ are impossible.

(2) $\Sigma_D$ is the largest family of specialization matrices that commute, which includes $\Sigma_{cond}$. The commutativity translates the idea that the combination of two pieces of evidence should lead to the same result whatever the order with which they are considered.

These requirements provide excellent justifications for Dempster's rules.

*Note:* let $m$ be a bba on $\Omega$, with q its related commonality function. Let $\mathbf{S}_m$ be the Dempsterian specialization matrix generated by $m$. Let $\mathbf{T}$ be the matrix that transforms a bba into a commonality function, where the elements $t_{A,B}$ of $\mathbf{T}$ are:

$$t_{A,B} = \begin{cases} 1 & \text{if } A \subseteq B \subseteq \Omega \\ 0 & \text{otherwise,} \end{cases}$$

in which case $q = \mathbf{T} \cdot m$.

It can be shown that:

(1) the commonalities $q(A)$, $A \subseteq \Omega$, are the diagonal elements and the eigenvalues of $\mathbf{S}_m$; and

(2) the columns of $\mathbf{T}^{-1}$ are the eigenvectors of $\mathbf{S}_m$. One has the representation:

$$\mathbf{S}_m = \mathbf{T}\Lambda\mathbf{T}^{-1}$$

where $\Lambda$ is a diagonal matrix with elements $\lambda_{A,A} = q(A)$, $A \subseteq \Omega$. Among others, this property may be useful when $m(\Omega) = 0$, in which case the theory of generalized inverses described in matrix calculus can be helpful.

**Example 3.** We present a matrix $\mathbf{S}_m$ and show the computation when $\Omega = \{a, b, c\}$. The table presents the subsets that correspond to each row and column (column 1 and top row), the $m_0$ vector (column 11), the specialization matrix $\mathbf{S}_m$ (columns 3 to 10), and $m_1 = \mathbf{S}_m \cdot m_0$ (column 2). The $m$ vector is given in the $\{a, b, c\}$ column of the $\mathbf{S}_m$ matrix.

| 1 | 2 | | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $2^\Omega$ | $m_1$ | | $\emptyset$ | {a} | {b} | {a,b} | {c} | {a,c} | {b,c} | {a,b,c} | | $m_0$ |
| $\emptyset$ | 0.4227 | | 1 | 0.62 | 0.42 | 0.3 | 0.55 | 0.25 | 0.18 | 0.11 | | 0.08 |
| {a} | 0.1186 | | 0 | 0.38 | 0 | 0.12 | 0 | 0.3 | 0 | 0.07 | | 0.18 |
| {b} | 0.1394 | = | 0 | 0 | 0.58 | 0.32 | 0 | 0 | 0.37 | 0.14 | - | 0.07 |
| {a,b} | 0.0628 | | 0 | 0 | 0 | 0.26 | 0 | 0 | 0 | 0.23 | | 0.10 |
| {c} | 0.1825 | | 0 | 0 | 0 | 0 | 0.45 | 0.37 | 0.24 | 0.19 | | 0.20 |
| {a,c} | 0.0152 | | 0 | 0 | 0 | 0 | 0 | 0.08 | 0 | 0.05 | | 0.09 |
| {b,c} | 0.0540 | | 0 | 0 | 0 | 0 | 0 | 0 | 0.21 | 0.18 | | 0.12 |
| {a,b,c} | 0.0048 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.03 | | 0.16 |

**Generalization of the specialization concept.** Suppose two spaces $X$ and $Y$ and a $2^{|X|}$ by $2^{|Y|}$ matrix $M$ whose columns are basic belief assignments on $2^X$. These vectors of basic belief masses are denoted by $m^X[y]$ for $y \subseteq Y$. The vector $m^X[y]$ would correspond to Your belief on $X$ given You only knew that the actual value $y_0$ belongs to subset $y$ of $Y$. Let $m^Y$ be Your belief about the actual value of $y_0$ (this would correspond to the *a priori* belief on $Y$ as described by the Bayesians). It can be shown that the belief $m^X[M, m^Y]$ induced on $X$ by all the basic belief assignments in $M$ and Your *a priori* belief $m^Y$ on $Y$ is:

$$m^X[M, m^Y] = M \cdot m^Y.$$

Several special cases of this equation are worth considering.

**Markov chains.** Suppose that $y_0$ is the actual state of a system at time $t$, and that $x_0$ is the state of the same system at $t+1$. Then $|Y| = |X|$. The bba $m^Y$ describes thus the belief about the state of the system at $t$, $M$ is the transition matrix, and $m^X[M, m^Y]$ describes the belief about the state of the system at $t+1$. The relation could be written as:

$$m_{t+1} = M \cdot m_t.$$

What we have here is the generalization of the Markov process within the realm of belief functions.

**Specializations.** Then consider that $Y = X$. In that case the vector $m^X[x]$ of $M$ described Your belief on $X$ given You only know that $x_0$ belongs to $x \subseteq X$. Then $m^X[x](A) = 0$ whenever $A \not\subseteq x$, what just means that $M$ is a specialization matrix.

**Dempsterian specializations.** If furthermore You had some belief $m_0^X$ on $X$, and the $m^X[x]$ results from the conditioning of $m_0^X$ on $x$, then the specialization matrix becomes Dempsterian.

Results similar to those obtained from the concepts of specialization are derivable when the conjunctive combination is replaced by other types of combinations like the disjunctive combination and the $\alpha$-junctions (Smets, 1997a) [39]. Concepts of generalization matrices appear in the disjunctive case (Klawonn and Smets, 1992) [13].

## 6. Decision Making

6.1. **The pignistic probability function for decision making.** Consider a frame of discernment $\Omega$ with bel quantifying Your beliefs on $\Omega$ at the credal level. When a decision *must* be made that depends on the actual state of affairs $\omega_0$ where $\omega_0 \in \Omega$, You must construct a probability function on $2^\Omega$ in order to make the optimal decision, i.e., the one that maximizes the expected utility. If You did not, Your behavior could be shown to be 'irrational'. We assume that the probability function defined on $2^\Omega$ is a function of the belief function bel also defined on $2^\Omega$. It translates the saying that beliefs guide our actions. Hence one must transform bel into a probability function, denoted BetP. This transformation is called the pignistic transformation and denoted $\Gamma$; it depends on both bel and the betting frame $F$ on which bets are built.

6.2. **The betting frame.** The pignistic transformation depends on the structure of the frame on which decisions must be made. The betting frame $F$ on $\Omega$ is the set of 'atoms' on which stakes will be allocated. Bets can then be built only on the elements of the power set of that frame. The granularity of the frame $F$ is defined so that a stake could be given to each atom of $F$ independently of the stakes given to the other atoms of $F$. Suppose one starts with a credibility function on a frame $F_0$. If the stakes given to atoms $A$ and $B$ of $F_0$ must necessarily be always equal,

both $A$ and $B$ belong to the same granule of the betting frame $F$. The betting frame is organized so that its granules are the atoms of $F$. $F$ results from the application of a sequence of coarsenings and/or refinements on $F_0$. The pignistic probability BetP is then built from the belief function bel so derived on $F$, and we write:

$$\text{BetP} = \Gamma(\text{bel}, F).$$

We call BetP a pignistic probability to insist on the fact that it is a probability measure used to make decisions (Bet is for betting). Of course BetP is a classical probability measure.

6.3. **The pignistic transformation.** The mathematical structure of the pignistic transformation is derived from the following scenario.

**Example 4: Buying Your friend a drink.** Suppose You have two friends, $G$ and $J$. You know they will toss a fair coin and the winner will visit You tonight. You want to buy the drink Your friend would like to have tonight: coke, wine or beer. You can only buy one drink. Let $D = \{\text{coke}, \text{wine}, \text{beer}\}$.

Let $\text{bel}_G(d)$, for all $d \subseteq D$, quantify Your belief about the drink $G$ will ask for. Given $\text{bel}_G$, You build the pignistic probability $\text{BetP}_G$ about the drink $G$ will ask by applying the (still to be defined) pignistic transformation. You build in the same way the pignistic probability $\text{BetP}_J$ based on $\text{bel}_J$, Your belief about the drink $J$ will ask for. The two pignistic probability distributions $\text{BetP}_G$ and $\text{BetP}_J$ are the conditional probability distributions about the drink that will be asked for, given $G$ or $J$ comes. The pignistic probability distribution $\text{BetP}_{GJ}$ about the drink that Your visitor will ask for is then:

$$\text{BetP}_{GJ}(d) = 0.5\text{BetP}_G(d) + 0.5\text{BetP}_J(d) \qquad \text{for all } d \subseteq D$$

You will use these pignistic probabilities $\text{BetP}_{GJ}(d)$ to decide which drink to buy.

But You might as well reconsider the whole problem and first compute Your belief $\text{bel}_V$ about the drink Your visitor ($V$) would like to have. It can be proved that $\text{bel}_V$ satisfies:

$$\text{bel}_V(d) = 0.5\text{bel}_G(d) + 0.5\text{bel}_J(d) \qquad \text{for all } d \subseteq D$$

Given $\text{bel}_V$, You could then build the pignistic probability $\text{BetP}_V$ You should use to decide which drink to buy. It seems reasonable to assume that $\text{BetP}_V$ and $\text{BetP}_{GJ}$ must be equal. In such a case, the pignistic transformation is uniquely defined.

Formally, we have assumed:

*Linearity Axiom:* Let a betting frame $F$ and let $\text{bel}_1$ and $\text{bel}_2$ be two belief functions on $2^F$. Let $\Gamma$ be the pignistic transformation that transforms a belief function over $2^F$ into a probability function BetP over $F$. Then $\Gamma$ satisfies, for any $\alpha \in [0, 1]$,

$$\Gamma(\alpha\text{bel}_1 + (1-\alpha)\text{bel}_2, F) = \alpha\Gamma(\text{bel}_1, F) + (1-\alpha)\Gamma(\text{bel}_2, F)$$

Two technical axioms must be added that are hardly arguable. Informally, they state:

*Anonymity Axiom:* The pignistic probability given to the image of $A \subseteq F$ after a permutation of the atoms of $F$ is the same as the pignistic probability given to $A$ before applying the permutation.

*Impossible Event Axiom:* The pignistic probability of an impossible event is zero.

Under these three axioms, it is possible to derive $\Gamma$. It is easily expressed when using the basic belief assignment $m$ related to bel.

**The pignistic transformation.** Let $m$ be the basic belief assignment on $2^F$ related to the belief function bel also defined on $2^F$. Let $\text{BetP} = \Gamma(\text{bel}, F)$. Then:

$$\text{BetP}(\omega) = \sum_{A:\,\omega \in A \subseteq \Omega} \frac{m(A)}{|A|(1 - m(\emptyset))} \tag{8}$$

where $|A|$ is the number of elements of $\Omega$ in $A$ and $n = |\Omega|$.

It is easy to show that the function BetP is indeed a probability function and the pignistic transformation of a probability function is the probability function itself.

6.4. **Betting under total ignorance.** To show the potency of our approach, let us consider one of those disturbing examples based on total ignorance.

**Example 5.** Consider a guard in a huge power plant. On the emergency panel, alarms $A_1$ and $A_2$ are both on. The guard never heard about these two alarms. They were hidden in a remote place. He takes the instruction book and discovers that alarm $A_1$ is on when circuit $C$ is either in state $C_1$ or in state $C_2$ and that alarm $A_2$ is on when circuit $D$ is one state of $D_1$, $D_2$ or $D_3$. He never heard about these $C$ and $D$ circuits. Therefore, his beliefs on the $C$ circuits will be characterized by a vacuous belief function on space $\Omega_C = \{C_1, C_2\}$. By the application of (8) his pignistic probability will be given by $\text{BetP}_C(C_1) = \text{BetP}_C(C_2) = 1/2$. Similarly for the $D$ circuit, the guard's belief on space $\Omega_D = \{D_1, D_2, D_3\}$ will be vacuous and the pignistic probabilities are $\text{BetP}_D(D_1) = \text{BetP}_D(D_2) = \text{BetP}_D(D_3) = 1/3$. Now, by reading the next page on the manual, the guard discovers that circuits $C$ and $D$ are so made that whenever circuit $C$ is in state $C_1$, circuit $D$ is in state $D_1$ and vice-versa. So he learns that $C_1$ and $D_1$ are equivalent and that $C_2$ and $D_2$ or $D_3$ are also equivalent. This information does modify neither his belief nor his pignistic probability.

If the guard had been a trained Bayesian, he would have assigned values for $P_C(C_1)$ and $P_D(D_1)$ (given the lack of any information, they would probably be $1/2$ and $1/3$, but any value could be used). Once he learns about the equivalence between $C_1$ and $D_1$, he must adapt his probabilities because they must give the same probabilities to $C_1$ and $D_1$. Which set of probabilities he is going to update: $P_C$ or $P_D$, and why?, especially since it must be remembered that he has no knowledge whatsoever about what the circuits are. In a probabilistic approach, the difficulty raised by this type of example results from the requirement that equivalent propositions should receive identical beliefs, and therefore identical probabilities. This reflects that the credal and the pignistic levels are not distinguished.

Within the transferable belief model, the only requirement is that equivalent propositions should receive equal beliefs (it is satisfied as $\text{bel}_C(C_1) = \text{bel}_D(D_1) = 0$). Pignistic probabilities depend not only on these beliefs but also on the structure of the betting frame. The difference between $\text{BetP}_C(C_1)$ and $\text{BetP}_D(D_1)$ reflects the difference between the two betting frames.

6.5. **The family of probability functions compatible with** bel**.** The literature dealing with belief functions is poisoned by a serious confusion that often leads to erroneous results. In the TBM, the values of bel do not result from some probability. The theory for quantifying the strength of Your belief that the actual world belongs to the subsets of $\Omega$ is developed and justified without considering the existence of some underlying, maybe hidden, probability. In Shafer's (1976a) book [26], the same approach prevails. But in the early 80's, authors understood the approaches based on belief functions as a theory of lower probabilities. Indeed it is mathematically true that given a normalized (i.e., where $m(\emptyset) = 0$) belief function bel on $\Omega$, it is

always possible to define a family $\Pi$ of probability functions $P$ defined on $\Omega$ that satisfy the following constraints:

$$\forall P \in \Pi, \forall A \subseteq \Omega, \quad \mathrm{bel}(A) \leq P(A) \leq \mathrm{pl}(A)$$

This property has often been used to claim that belief functions are just lower probability functions. The danger comes from the fact that some authors generalize this statement and claim that belief functions concern an ill-known probability function. In that case, the existence of a probability function $P$ that belongs to $\Pi$ is assumed and $P$ represents 'something', the 'something' being of course understood as Your degree of belief on $\Omega$. At the static level, the difference is the following. In the TBM, bel represents Your beliefs. In the lower probability approach, one assumes that Your belief is represented by a probability function, whose value is only known to belong to $\Pi$, and bel is just the lower limit of that family $\Pi$. (Note there exist families $\Pi$ such that their lower envelopes are not belief functions.)

The difference becomes more obvious once conditioning on an event $X$ is introduced. In the TBM, conditioning of bel on $X$ into $\mathrm{bel}_X$ is achieved by Dempster's rule of conditioning, hence by the transfer of the basic belief masses as explained above. In the lower probability approach, the conditioning is obtained by considering every probability function $P$ in $\Pi$, conditioning $P$ on $X$ and collecting them in a new family $\Pi_X$ of conditional probability functions $P_X$. The results are different: indeed $\mathrm{bel}_X$ is not the lower envelope of $\Pi_X$ (Kyburg, 1987) [18] (Voorbraak, 1993) [43].

The family $\Pi$ of probability functions compatible with a given belief function has nevertheless a meaning in the TBM, but quite different from the one considered in the lower probabilities approach. Given a belief function, the probability function used to compute the expected utilities at the pignistic level when a decision is involved is computed by the so-called pignistic transformation. The result depends of course on bel, but also on the betting frame, the set of elementary options considered in the decision process. Suppose we consider all the possible betting frames. For each possible betting frame we get a probability function. Collect all these probability functions into a family. This family is the same as the family $\Pi$ (Wilson, 1993) [45]. So we can derive $\Pi$ in the TBM. The difference with the lower probability approach is that we start with bel and derive $\Pi$ as a by-product, whereas the lower probability approach starts with $\Pi$ and derive bel as a by-product.

## 7. Non-standard probabilities and belief functions

7.1. **Upper and lower probabilities.** Smith (1961) [41], Good (1962) [8] and Walley (1991) [44] suggested that personal degrees of belief cannot be expressed by a single number but that one can only assess intervals that bound them. The interval is described by its boundaries called the upper and lower probabilities. Such an interval can easily be obtained in a two-agent situation when one agent, $Y_1$, communicates the probability of some events in $\Omega$ to a second agent, $Y_2$, by only saying that, for each $A \subseteq \Omega$, the probability $P(A)$ belongs to some interval. Suppose $Y_2$ has no other information about the probability on $\Omega$. In that case, $Y_2$ can only build a set $\Pi$ of probability measures on $\Omega$ compatible with the boundaries provided by $Y_1$. All that is known to $Y_2$ is that there exists a probability measure $P$ and that $P \in \Pi$. Should $Y_2$ learn then that an event $A \in \Omega$ has occurred, $\Pi$ should be updated to $\Pi_A$ where $\Pi_A$ is this set of conditional probability measures obtained by conditioning the probability measures $P \in \Pi$ on $A$ (Smets, 1987) [31] (Fagin and Halpern, 1991a) [7] (Jaffray, 1992) [12].

One obtains a similar result by assuming that one's belief is not described by a single probability measure as do the Bayesians but by a family of probability

measures (usually the family is assumed to be convex). Conditioning on some event $A \subseteq B \subseteq \Omega$ is obtained as in the previous case.

7.2. **Dempster's model.** A special case of upper and lower probabilities has been described by Dempster (1967) [3]. He assumes the existence of a probability measure on a space $X$ and a one-to-many mapping $M$ from $X$ to $Y$. Then the lower probability of $A$ in $Y$ is equal to the probability of the largest subset of $X$ such that its image under $M$ is included in $A$. The upper probability of $A$ in $Y$ is the probability of the largest subset of $X$ such that the images under $M$ of all its elements have a non-empty intersection with $A$. In the Artificial Intelligence community, this theory is what people often call the Dempster–Shafer theory.

7.3. **The theory of hints.** Kohlas and Monney (1995) [16] have proposed a theory of hints. They assume Dempster's original structure $(\Omega, P, \Gamma, \Theta)$ where $\Omega$ and $\Theta$ are two sets, $P$ is a probability measure on $\Omega$ and $\Gamma$ is a one-to-many mapping from $\Omega$ to $\Theta$. They assume a question, whose answer is unknown. The set $\Theta$ is the set of possible answers to the question. One and only one element of $\Theta$ is the correct answer to the question. The goal is to make assertions about the answer in the light of the available information. We assume that this information allows for several different interpretations, depending on some unknown circumstances. These interpretations are regrouped into the set $\Omega$ and there is exactly one correct interpretation. Not all interpretations are equally likely and the known probability measure $P$ on $\Omega$ reflects our information in that respect. Furthermore, if the interpretation $\omega \in \Omega$ is the correct one, then the answer is known to be in the subset $\Gamma(\omega) \subseteq \Theta$. Such a structure $\mathbf{H} = (\Omega, P, \Gamma, \Theta)$ is called a hint. An interpretation $\omega \in \Omega$ supports the hypothesis $H$ if $\Gamma(\omega) \subseteq H$ because in that case the answer is necessarily in $H$. The degree of support of $H$, denoted $sp(H)$, is defined as the probability of all supporting interpretation of $H'$ (Kohlas and Monney, 1995, page vi) [16].

The theory of hints corresponds to Dempster's original approach. Kohlas and Monney call their measure a degree of support, instead of belief, to avoid personal, subjective connotation, but degrees of support and degrees of belief are mathematically equivalent and conceptually very close. In the hints theory, the primitive concept is the hint from which degrees of supports are deduced, whereas the TBM and Shafer's initial approach (Shafer, 1976a) [26], consider the degrees of belief as a primitive concept. The theory of hints is quite similar to the probability of provability theory (see Section 7.5). All details on the theory of hints can be found in Kohlas and Monney (1995) [16].

7.4. **Inner and outer measures.** Halpern and Fagin (1992) [10], and Voorbraak (1993) [43] have studied the following problem. Suppose two algebras $A$ and $B$ defined on the set $\Omega$, where $A$ is a subalgebra of $B$. Suppose the values of the probability measure are known only on the elements of the algebra $A$. Fagin and Halpern try to determine the values of the probability measure on the subsets of the algebra $B$. Because of the missing information, only the inner and outer measures for every event $B$ in $B$ can be determined. By construction, the inner (outer) measure is a lower (upper) probability function, and even a special one as the inner (outer) measure is a belief (plausibility) function, an obvious result when $\Omega$ is finite, and that is easily derived once Dempster's one-to-many relation is considered.

7.5. **Probabilities defined on modal propositions.** Classically probability theory is defined on propositional logic. The whole presentation of probability theory could be realized by using propositions instead of events and subsets. So for a proposition $p$, $P(p)$ would be the probability that $p$ is true (hence that $p$ is true

in the actual world). Extending the domain of the probability functions to modal propositions is quite feasible. Ruspini (1986) [24] examines the 'probability of knowing'. Pearl (1988) [22] examines the 'probability of provability'. Both approaches fit essentially with the same ideas.

The probability $P(\Box p)$ is the probability that $\Box p$ is true in the actual world. The worlds of $\Omega$ can be partitioned in three categories: those where $\Box p$ holds, those where $\Box \neg p$ holds, and those where neither $\Box p$ nor $\Box \neg p$ hold. Hence,

$$P(\Box p) + P(\Box \neg p) + P(\neg \Box p \& \neg \Box \neg p) = 1. \tag{9}$$

Suppose You define bel($p$) as $P(\Box p)$, i.e., You define bel($p$) as the probability that $p$ is proved, is known, is necessary, depending on the meaning given to the $\Box$ operator. The equality (9) becomes then:

$$\mathrm{bel}(p) + \mathrm{bel}(\neg p) \leq 1,$$

Similarly the other inequalities described with belief functions (2) are also satisfied. This approach provides a nice interpretation of bel as the probability of provability, of knowing, etc. Nevertheless the theory so derived is not the TBM, as seen once conditioning is involved (Smets, 1991b) [34]. The probability $P(\Box p | \Box q)$ of knowing $p$ when knowing $q$ is:

$$P(\Box p | \Box q) = \frac{P(\Box p \& \Box q)}{P(\Box q)} = \frac{P(\Box (p \& q))}{P(\Box q)} = \frac{\mathrm{bel}(p \& q)}{\mathrm{bel}(q)}$$

This is not Dempster's rule of conditioning. It happens to be the so-called geometrical rule of conditioning (Shafer, 1976b) [25]. Dempster's rule of conditioning is obtained if the impact of the conditioning event results in an adaptation of the accessibility relation underlying the modal logic (Smets, 1991c) [35].

## WHERE TO FIND WHAT?

**Major conferences where papers on belief functions are presented:**
- IPMU: Information Processing and Management of Uncertainty, in Europe, every two years, since 1986
- UAI: Uncertainty in Artificial Intelligence, mainly USA, yearly, since 1985
- ECSQARU: European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty, in Europe, every two years, since 1991
- ISIPTA: International Symposium on Imprecise Probabilities and Their Applications, in Europe and the USA, every two years, since 1999

**Major journals publishing on belief functions:**
- International Journal of Approximate Reasoning
- International Journal of Intelligent Systems
- International Journal of General Systems
- IEEE Transactions on Pattern analysis and Machine Intelligence (PAMI)
- IEEE Transactions on Systems, Machines and Cybernetics (SMC)
- Artificial Intelligence
- International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.

For **books and papers** on belief functions, see the IPP Bibliography.

## REFERENCES

[1] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, 11:27–40, 1991.

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

[3] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[4] D. Dubois and H. Prade. The principle of minimum specificity as a basis for evidential reasoning. In B. Bouchon and R. R. Yager, editors, *Uncertainty in Knowledge-Based Systems*, pages 75–84. Springer-Verlag, Berlin, 1987.

[5] D. Dubois and H. Prade. On the unicity of Dempster's rule of combination. *International Journal of Intelligent Systems*, 1:133–142, 1986.

[6] D. Dubois and H. Prade. Focusing versus updating in belief function theory. In R. R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 71–95. Wiley, New York, 1994.

[7] R. Fagin and J. Halpern. A new approach to updating beliefs. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 347–374. North Holland, Amsterdam, 1991.

[8] I. J. Good. *Probability and the Weighing of Evidence*. Griffin, London, 1950.

[9] J. Gordon and E. H. Shortliffe. The Dempster-Shafer theory of evidence. In B. G. Buchanan and E. H. Shortliffe, editors, *Rule-Based Expert Systems: the MYCIN Experiments of the Stanford Heuristic Programming Project*, pages 272–292. Addison-Wesley, Reading, MA, 1984.

[10] J. Halpern and R. Fagin. Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–318, 1992.

[11] Y. T. Hsia. Characterizing belief with minimum commitment. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI-91*, pages 1184–1189. Morgan Kaufman, San Mateo, CA, 1991.

[12] J. Y. Jaffray. Bayesian updating and belief functions. *IEEE Transactions on Systems, Man and Cybernetics*, 22:1144–1152, 1992.

[13] F. Klawonn and Ph. Smets. 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, editors, *Uncertainty in Artificial Intelligence 92*, pages 130–137. Morgan Kaufman, San Mateo, CA, 1992.

[14] G. Klir and M. J. Wierman. *Uncertainty-based-information: elements of generalized information theory*. Physica-Verlag, Heidelberg, NY, 1998.

[15] G. J. Klir. Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like. *Fuzzy Sets and Systems*, 24:141–160, 1987.

[16] J. Kohlas and P. A. Monney. *A Mathematical Theory of Hints: An Approach to Dempster-Shafer Theory of Evidence*. Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag, Berlin, 1995.

[17] R. Kruse and E. Schwecke. Specialization: a new concept for uncertainty handling with belief functions. *International Journal of General Systems*, 18:49–60, 1990.

[18] H. E. Kyburg, Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–294, 1987.

[19] S. Moral. *Informacion difusa. Relaciones entre probabilidad y possibilidad*. PhD thesis, Universidad de Granada, Granada, Spain, 1985.

[20] N. Pal, J. Bezdek, and R. Hemasinha. Uncertainty measures for evidential reasoning I: a review. *International Journal of Approximate Reasoning*, 7:165–183, 1992.

[21] N. Pal, J. Bezdek, and R. Hemasinha. Uncertainty measures for evidential reasoning II: new measure of total uncertainty. *International Journal of Approximate Reasoning*, 8:1–16, 1993.

[22] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[23] J. Pearl. Reasoning with belief functions: an analysis of compatibility. *International Journal of Approximate Reasoning*, 4:363–390, 1990.

[24] E. H. Ruspini. The logical foundations of evidential reasoning. Technical note 408, SRI International, Menlo Park, CA, 1986.

[25] G. Shafer. A theory of statistical evidence. In W. L. Harper and C. A. Hooker, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, vol. 2, pages 365–436. Reidel, Dordrecht, 1976.

[26] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[27] G. Shafer and A. Tversky. Languages and designs for probability. *Cognitive Science*, 9:309–339, 1985.

[28] Ph. Smets. The transferable belief model for quantified belief representation. In D. M. Gabbay and Ph. Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol. 1, pages 267–301. Kluwer, Dordrecht, 1998.

[29] Ph. Smets. *Un modèle mathématico-statistique simulant le processus du diagnostic médical.* PhD thesis, Université Libre de Bruxelles, 1978.

[30] Ph. Smets. Information content of an evidence. *International Journal of Man–Machine Studies*, 19:33–43, 1983.

[31] Ph. Smets. Belief functions and generalized Bayes theorem. In *Proceedings of the Second IFSA Congress*, pages 404–407, Tokyo, Japan, July 20–25 1987.

[32] Ph. Smets. Belief functions. In Ph. Smets, A. Mamdani, D. Dubois, and H. Prade, editors, *Non standard logics for automated reasoning*, pages 253–286. Academic Press, London, 1988.

[33] Ph. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 12:447–458, 1990.

[34] Ph. Smets. Probability of provability and belief functions. *Logique et Analyse*, 133:177–195, 1991.

[35] Ph. Smets. 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, editors, *Uncertainty in Artificial Intelligence 6*, pages 375–384. North Holland, Amsterdam, 1991.

[36] Ph. Smets. The nature of the unnormalized beliefs encountered in the transferable belief model. In D. Dubois, M. P. Wellman, B. D'Ambrosio, and Ph. Smets, editors, *Uncertainty in Artificial Intelligence 92*, pages 292–297. Morgan Kaufman, San Mateo, CA, 1992.

[37] Ph. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.

[38] Ph. Smets. What is Dempster-Shafer's model? In R. R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 5–34. Wiley, New York, 1994.

[39] Ph. Smets. The alpha-junctions: combination operators applicable to belief functions. In D.M Gabbay, R. Kruse, A. Nonnengart, and H.J. Ohlbach, editors, *Qualitative and quantitative practical reasoning*, pages 131–153. Springer, 1997.

[40] Ph. Smets and P. Magrez. La représentation de la croyance. *Revue de l'Intelligence Artificielle*, 1:31–46, 1987.

[41] C. A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society B*, 23:1–37, 1961.

[42] F. Voorbraak. On the justification of Dempster's rule of combination. *Artificial Intelligence*, 48:171–197, 1991.

[43] F. Voorbraak. *As Far as I Know: Epistemic Logic and Uncertainty.* Dissertation, Utrecht University, 1993.

[44] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[45] N. Wilson. Decision making with belief functions and pignistic probabilities. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 364–371, Granada, November 1993. Springer Verlag.

[46] R. Yager. The entailment principle for Dempster-Shafer granules. *International Journal of Intelligent Systems*, 1:247–262, 1986.