Target Identification Based on the Transferable Belief Model Interpretation of Dempster-Shafer Model. Pars I: Methodology.

François Delmotte
LAMIH, Université de Valenciennes
Le Mont Houy, 59313 Valenciennes Cedex 9, FRANCE
francois.delmotte@univ-valenciennes.fr
and
Philippe Smets
IRIDIA - Université Libre de Bruxelles
50 av. Roosevelt, CP 194-6, 1050 Bruxelles, Belgium
psmets@ulb.ac.be
http://iridia.ulb.ac.be/~psmets

Abstract

This paper explains how multisensor data fusion and target identification can be performed within the transferable belief model, a model for the representation of quantified uncertainty based on belief functions. The paper is presented in two parts: methodology and application. In this part, we present the underlying theory, in particular the General Bayesian Theorem needed to transform likelihoods into beliefs and the pignistic transformation needed to build the probability measure required for decision making. We end with a simple example. More sophisticated examples and some comparative studies are presented in Part II. The results presented here can be extended directly to many problems of data fusion and diagnosis.

Keywords: Belief functions, transferable belief model, General Bayesian Theorem, pignistic probabilities, target identification, data fusion.

I. Introduction

Classically, uncertainty is represented by probability functions, but other models like those based on belief functions have been proposed since a few years. Belief functions are proposed to represent quantified uncertainty in a supposedly better way than probability functions. Indeed the latter are not well adapted to represent states of full or partial ignorance. The ability of belief functions to represent various types of uncertainty, ambiguity or imprecision in a flexible way is claimed by some people, correctly or not, to be counterbalanced by a larger mathematical and computational complexity.

Most discussions for comparing the probability and belief function approaches are focusing on theoretical considerations and toys examples. Instead in this paper, we study a practical example of target identification, considering it as a prototypical example of diagnostic process. We show how to apply the transferable belief model to such a problem, comparing it to the probability model, its obvious contender.

The transferable belief model (TBM for short) has been developed to provide a model for the representation of quantified beliefs. It is based on belief functions. It corresponds to an interpretation of the model developed by Shafer in his book [1]. The model has been presented in [2], (see also [3], [4]). The TBM intends to represent weighted opinions, called here beliefs, and is not to be confused with the models based on lower probabilities. Many tools are presented in Shafer's book like the combination rules for distinct pieces of evidence and the discounting factors. Since then, we have developed several concepts.

Those useful for this presentation are:

• the open world concept [5] that acknowledges that the frame of discernment might not be exhaustive, what implies that positive mass can be allocated to the empty set.

• the General Bayesian Theorem [6] that permits to transform likelihoods into belief functions, and its closely related ballooning extension that corresponds to a deconditioning of a belief function.

• the separation of a credal mental level where beliefs are held and a pignistic mental level where beliefs are used for decision making [2]. The link between the two levels is achieved by the pignistic transformation that permits the construction of the probability function required to make decisions [7], [8].

• the least commitment principle [9], [10], [3] that justifies the selection of the belief function that does not create unjustified beliefs.

• the disjunctive rules of combinations [6] needed to combine two pieces of evidence when all we know is that at least one must be considered.

The computational efficiency of the belief function approach and its comparison with the probability approach has been studied in [11], [12] on a problem of target identification. The comparison is based on the number of identical data that must be collected in order to reach a 'firm' decision. Unfortunately, these authors use belief functions in a way we feel to be ad hoc. These authors conclude that the probability approach is more efficient. We present in details how we consider belief functions should be applied to such problems. The essential tools are the General Bayesian Theorem and the pignistic transformation. We repeat the efficiency study of Buede and Girardi [12] and show, on the contrary, that the transferable belief model approach is computationally *more* efficient that the probability one, even if the gain is small.

Most of the theoretical material is scattered over many papers, often difficult to access. So we feel it useful for the reader to regroup them here in a synthetic way. All details and proofs can be found in the original papers.

This paper is organized as follows. In Section II, we present the general problem of target identification by multisensors and the way the TBM can be used for such a purpose. We present the theoretical concepts concerning the TBM that are necessary for the present problem of target identification. In Section III, we present the General Bayesian Theorem that is central for the method we propose for target identification. In Section IV, we apply the General Bayesian Theorem to a simple problem.

In the second part, we show through an example that the General Bayesian Theorem and the probability solutions can be diametrically opposed indicating thus that the choice of the model is not just an academic exercise, but an issue which consequences can be very serious. We then reproduce in part the study presented in [11], [12] and show that the TBM approach is computationally never less efficient that its probability analogous. In the last three sections, we show how to use the TBM for multisensor target identification problems, and conclude.

II. The Transferable Belief Model for Identification

A. Sensors and Identification

A sensor can be seen essentially as a piece of equipment that observes some data x and transmits some 'opinion' about the actual value of a parameter of interest h. The simplest form arises when x and h are in one-to-one correspondence. In that case the sensor observes x and communicates the corresponding h. In more complex cases, the relation between x and h is not that simple, and the relation between x and h is represented by a set of probability distributions on x given h, one distribution for each possible value of h. Let $P(x|h)$ denote the probability measure on X given $h \in H$, where X and H are the observation and parameter domains, respectively. After observing x, the sensor communicates its opinion on the value of h under the form of a 'likelihood' vector. Let $l(h|x)$ denote the likelihood of the hypothesis h given the observed data is x: by definition $l(h|x) = P(x|h)$[1].

This classical representation is based on the idea that for each h, the value that might be observed is uncertain and this uncertainty can be represented by a probability measure on X. If this probability measure results from the observation of the value of x for many h objects, and if the sampling procedure used when collecting these data is well established

[1]This notation is unfortunate as it leads to confusions. It would have been better to write: if $P(x|h) = a$, then $l(h|x) = a$. Likelihoods and probabilities share the same values, but a likelihood is not a probability as one might erroneously deduce from the usual notation.

so that probabilistic inference can be used in order to build $P(x|h)$, the representation is perfectly valid. In practice, a (hopefully random and representative) sample is used and $P(x|h)$ is equated to the proportion of x data observed in the sample. The validity of such a simple procedure can be questioned and there are of course more sophisticated methods to estimate $P(x|h)$.

Nevertheless it might be argue that the uncertainty about x given h is not probabilist. For instance, the sample might be neither random nor 'representative', or even worse, it may happen there is no sample available. Non-probabilist representations have been suggested, like those based on possibility or belief functions. In that case, the probability measure $P(x|h)$ is replaced by a conditional possibility function or a conditional belief function defined on X for each h. We feel that the possibility approach is appropriate when some fuzziness pervades the data or the hypothesis or our knowledge about $P(x|h)$. Fuzziness covers imprecision, whereas we are concerned here only with uncertainty [13]. In this study, there will be no fuzziness involved, all sets will be crisp, and therefore we will not further consider possibility functions. Instead, we focus on belief functions.

B. Belief Functions for Uncertainty Representation

The idea of using belief functions to represent quantified uncertainty was first introduced by Shafer [1] who was building his theory from Dempster's research [14], [15], [16]. Later Gordon and Shortliffe [17] coined the term 'Dempster-Shafer theory'. Unfortunately, this term turns out to be ambiguous as it covers essentially two models, one based on upper and lower probability functions, the other on what we call the TBM. Missing to distinguish between these models has created confusion in the literature [18].

In the first model, one assumes the existence of some probability measure, denoted P, that represents the uncertainty. Because of a lack of adequate information, all that is available or known to the user about P is that it belongs to some family of probability functions, denoted Π. The lower envelop of Π is defined as the function P_* over X with $P_*(x) = min_{P\in\Pi} P(x)$. This lower envelop is usually called the lower probability function. In the classical case, Π is convex in which case the lower envelop of Π fully describes the set Π. This theory of imprecise probabilities is fully detailed in Walley [19]. In some cases like those studied initially by Dempster and later by Kohlas and Monney in their hint model [20], this lower probability function is a belief function.

The second model, the TBM, does not consider any underlying probability function. It accepts that uncertainty does not have to be additive as in probability theory, and that it is represented by a belief function (that plays then the same role as the probability function in the probability models). This model is close to what we feel Shafer presented in his book [1], but to avoid confusion we have called it the TBM. The axiomatic justification of the model can be found in [21], [22], and its description in [2], [4], [3]. How decisions must be taken in the TBM is explained and justified in [7], [2] (see Section II-J).

The TBM not only represents quantified beliefs in a static way, but also its dynamic, i.e., how beliefs change:

- when data are collected by several sensors and must be combined (see Section II-G on combination rules)
- when data are collected from partially reliable sources and must be somehow discounted (see Section II-H on belief discounting)
- when beliefs must be inverted in order to transform a belief over X given $h \in H$ into a belief over H given $x \subseteq X$, generalizing thus Bayes theorem (see Section III on the Generalized Bayesian Theorem)

It might seem odd that sensors would produce their report in the form of a 'belief'. The term belief is strongly human oriented, and a sensor is usually anything but human. The term 'belief' is maybe unfortunate in this context, but we feel that for historical reasons we must keep it. Of course when we will state that a sensor believes something given it knows something else, we only mean that the sensor has collected the 'something else' and produced a belief function about the value of the 'something'. No psychological, philosophical or religious connotation must be given here to the term 'belief'.

C. The Transferable Belief Model

The TBM is an interpretation of the so-called Dempster-Shafer theory. It is a model for the representation of the quantified beliefs that results from the data collected by some agent, a sensor in the present context. The beliefs concern the actual value h_0 of some variable. Let H be the set of possible values of that variable. In an identification problem, H is the set of possible values that the object to identify can take. H is called the frame of discernment. Usually one assumes that $h_0 \in H$ (what we call the closed world assumption) but we can also consider the case where H is not defined in an exhaustive way (open world assumption).

The central element of the TBM is the basic belief assignment (bba), denoted m. For $A \subseteq H$, $m(A)$ is the part of belief that supports A (i.e. the fact that $h_0 \in A$), and that, due to a lack of information, does not support any strict subset of A. The initial total belief is scaled to 1, and thus $m(A) \in [0,1]$, with $\sum_{A \subseteq H} m(A) = 1$. We do not require $m(\emptyset) = 0$ as in Shafer's work.

The degree of belief $bel(A)$ is defined as: $bel : 2^H \rightarrow [0,1]$ with, for all $A \subseteq H$,

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \tag{1}$$

It quantifies the total amount of 'justified specific' support given to A. The term 'justified' means that B supports A, thus $B \subseteq A$, and the term 'specific' means that B does not support $\overline{A}$, thus $B \nsubseteq \overline{A}$ or equivalently $B \neq \emptyset$.

The degree of plausibility $pl(A)$ is defined as: $pl : 2^H \rightarrow [0,1]$ with, for all $A \subseteq H$,

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = bel(H) - bel(\overline{A}). \tag{2}$$

It quantifies the maximum amount of 'potential specific' support that could be given to A. The term 'potential' means that B might come to support A without supporting $\overline{A}$ if further piece of evidence is taken into consideration, thus $B \cap A \neq \emptyset$.

The functions m, bel and pl are in one to one correspondence. Two other useful functions also in one to one correspondence with m are the commonality function q and the implicability function b with:

$$q : 2^H \rightarrow [0,1], \qquad q(A) = \sum_{B : A \subseteq B} m(B),\ \forall A \subseteq H \tag{3}$$

$$b : 2^H \rightarrow [0,1], \qquad b(A) = \sum_{B : B \subseteq A} m(B),\ \forall A \subseteq H \tag{4}$$

$$= bel(A) + m(\emptyset) \tag{5}$$

Their meaning are not considered here, but they are computationally very useful to combine two belief functions conjunctively for q or disjunctively for b.

Useful inversion formulas are for all $A \subseteq H$:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} bel(B), \; A \neq \emptyset \tag{6}$$

$$m(\emptyset) = 1 - bel(H) \tag{7}$$

$$m(A) = \sum_{A \subseteq B} (-1)^{|B|-|A|} q(B) \tag{8}$$

D. Notation

In order to enhance the fact that we work with non-normalized belief functions ($m(\emptyset)$ can be positive), we use the notation *bel* and *pl*, whereas Shafer uses the notation *Bel* and *Pl*. The latter two are kept for normalized belief and plausibility functions, i.e., where $m(\emptyset) = 0$.

Besides we use the next conventions that we have found convenient, even though they might seem cumbersome in some cases. The full notation for *bel* and its related functions is:

$$bel^{H,\Re}_{S,t}[F_{S,t}](h_0 \in A) = x.$$

It denotes that x is the value of the degree of belief provided by the sensor S at time t that the actual value h_0 of H belongs to the set of values A, where A is a subset of the frame of discernment H and $A \in \Re$ where $\Re$ is a Boolean algebra of subsets of H. The belief is based on the facts $F_{S,t}$, where $F_{S,t}$ represents all facts taken into consideration by the sensor S at time t.

In practice many indices can be omitted for simplicity sake. In this paper we assume that: 1) $\Re$ is 2^H, the power set of H, 2) '$h_0 \in A$' is denoted as 'A', 3) S, t and/or H are omitted when the values of the missing elements are clearly defined from the context. So $bel^H[F](A)$ or even $bel(A)$ are often used.

In the above notation, *bel* can be replaced by any of m, b, pl, q, etc... The indices should made it clear what the links are. So $m^{H,\Re}_{S,t}[F_{S,t}]$ and $pl^{H,\Re}_{S,t}[F_{S,t}]$ are the bba and the plausibility function related to $bel^{H,\Re}_{S,t}[F_{S,t}]$.

Note that $bel^{H,\Re}_{S,t}[F_{S,t}]$ (as well as its simplified forms) denotes the belief function, and can be understood as a finite vector of length $|\Re|$, which components are the values of $bel^{H,\Re}_{S,t}[F_{S,t}](A)$ for every $A \in \Re$.

E. Axiomatic Justifications

A study of rationality properties that should be satisfied by a function which purpose is to quantify someone's beliefs leads to the use of belief functions [22], [21] and to the derivation of Dempster's rule of conditioning [23], [24], [25]. From this construction, we have derived (and often justified) many other concepts like:

- the conjunctive combination rule (that is Dempster's rule of combination except for its normalization) to compute $bel[E_1 \wedge E_2]$ from $bel[E_1]$ and $bel[E_2]$ (see Section II-G).
- the least commitment principle: 'never give more support than justified' what means that we should select the belief function which values of $pl(A)$ are as large as possible for every $A \subseteq H$, (see Section II-F).
- the General Bayesian Theorem to compute the belief function induced by the observation $x \subseteq X$ over the frame H from the set of conditional belief functions over the frame X given every $h_i \in H$. (see Section III).
- the pignistic transformation to build the probability function needed for taking 'optimal' decisions using the expected utility theory, (see Section II-J).
- very efficient algorithms, called the Fast Möbius Transforms, to transform any of m, b, bel, pl, q into each others [26], [27], [28]. Versions for MatLab 5.2 can be downloaded from the web site at: http://iridia.ulb.ac.be/˜psmets

The TBM is a largely extended model inspired by what is described in Shafer's book [1] (note that some of Shafer's later papers enhance other interpretations).

F. Vacuous and Ballooning Extensions

F.1 The Least Commitment Principle.

Let two bba m_1 and m_2 be defined on a frame of discernment H. We say that m_1 is less committed than m_2 iff:

$$pl_1(h) \geq pl_2(h) \qquad \forall h \subseteq H$$

It means that m_2 gives stronger support to each $h \subseteq H$ than m_1.

F.2 The Vacuous Belief Function.

The least committed of all belief functions defined on H is the so called vacuous belief function defined by $pl(h) = 1, \forall h \subseteq H$, or equivalently $m(h) = 1$ if $h = H$, and 0 otherwise.

F.3 The Vacuous Extension.

[1] Suppose a bba m^H defined on a frame of discernment H, and let H' be a refinement R of H, i.e., every element h_i of H is mapped by R into one to several elements of H', and their image on H' are the elements of a partition of H'. It just means that H' is more detailed than H. Suppose there is a bba m^H on H. This bba can be extended on H' in order to build a bba on H' that expresses the same information as contained in m^H. This transformation is called the vacuous extension of m^H on H', denoted by $m^{H\uparrow H'}$ and its values are given by:

$$
\begin{aligned}
m^{H\uparrow H'}(h') &= m^H(h) && \text{if } h' = R(h) \\
&= 0 && \text{otherwise.}
\end{aligned}
$$

where $R(h)$ is the image of h under R.

F.4 Coarsening.

Suppose a bba $m^{H'}$ defined on H'. Let H be a coarsening of H', i.e., H' is a refinement R of H. The bba induced on H by $m^{H'}$ is denoted by $m^{H'\downarrow H}$, and the values of its related $bel^{H'\downarrow H}$ are:

$$bel^{H'\downarrow H}(h) = bel^{H'}(R(h)) \qquad \forall h \subseteq H$$

Marginalization is a special case of coarsening when H can be represented as the product space of two variables X and Y, and the bba defined on $X \times Y$ is transformed into a bba on X; indeed X is a coarsening of $X \times Y$.

F.5 The Ballooning Extension.

[29], [6]

Suppose a frame of discernment H and let H' be a subset of H. Suppose a bba $m^{H'}$ defined on H' and we need a bba on H. The least committed bba on H such that its conditioning on H' is $m^{H'}$ is given by the so called 'ballooning' extension, denoted $m^{H' \Uparrow H}$, and its values are:

$$\begin{aligned} m^{H' \Uparrow H}(h) &= m^{H'}(h') \qquad \text{if } h' \subseteq H', h = h' \cup \overline{H'} \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

The ballooning extension is useful when the received beliefs were build on a limited frame and we discover that some alternatives had not been taken into consideration when the sensor produced the bba on the limited frame. We can thus build a bba on the larger frame from the one collected on the limited frame. It is repeatedly used to derive the General Bayesian Theorem presented in Section III.

G. Conjunctive Combinations

Let E_1 and E_2 be two 'distinct' pieces of evidence and let $m^H[E_1]$ and $m^H[E_2]$ be the bba's they induce on H. Remember the symbols between [and] denote the pieces of evidence taken in consideration when building the belief functions. We want to build the bba that would result from the combination of the two pieces of evidence. There are two families of combinations:

- the conjunctive combinations that build the bba given it is accepted that both sources are fully reliable.
- the disjunctive combinations that build the bba given at least one source is accepted as fully reliable but which one is so is unknown.

More complex combinations also exist but they are not considered here.

G.1 Conjunctive and Disjunctive Combinations

Let m_1 and m_2 be two bba's on H induced by two distinct pieces of evidence. Then their conjunctive combination $m_1 \textcircled{\cap} m_2$ (also denoted $m_{1 \textcircled{\cap} 2}$) is defined by:

$$m_{1 \textcircled{\cap} 2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \ \forall A \subseteq H \tag{9}$$

The same result can be conveniently expressed with the commonality function.

$$q_{1 \bigcirc\!\!\!\!\cap\, 2}(A) = q_1(A) q_2(A) \ \forall A \subseteq H \tag{10}$$

This rule is associative and commutative.

Dempster's rule of combination is obtained by normalizing the result of the conjunctive combination rule, i.e., by dividing all results by $(1 - m_{1 \bigcirc\!\!\!\!\cap\, 2}(\emptyset))$. The result is denoted then $m_{1\oplus 2}$.

G.2 Conditioning

A special, but essential, case of conjunctive combination rule is the conditioning rule. Let m_A be so that $m_A(X) = 1$ if $X = A$, and 0 otherwise. The result of the conjunctive combination of m with m_A produces a new bba $m[A]$ with:

$$m[A](B) = \sum_{C \subseteq \overline{A}} m(A \cup C) \text{ if } B \subseteq A \tag{11}$$

$$= 0 \text{ otherwise} \tag{12}$$

$$bel[A](B) = bel(B \cup \overline{A}) - bel(\overline{A}), \ \forall B \subseteq H \tag{13}$$

$$pl[A](B) = pl(A \cap B), \ \forall B \subseteq H. \tag{14}$$

This operation represents the impact of the information '$h_0 \notin \overline{A}$', which differs from $h_0 \in A$ in the open world context as h_0 might then not belong in H.

G.3 Useful Relations

We present first some relations relative to the conjunctive rule of combination that are very useful to simplify computation and to understand the nature of the conjunction operator (see Smets, [6]).

Let m_0 and m_1 be two bba's on H. Let $m_2 = m_1 \textcircled{\cap} m_0$. For all $h \subseteq H$,

$$m_2(h) = \sum_{A \cap B = h} m_1(A) m_0(B) \tag{15}$$

$$= \sum_{B \subseteq H} m_1[B](h) m_0(B) \tag{16}$$

$$bel_2(h) = \sum_{B \subseteq H} bel_1[B](h) m_0(B) \tag{17}$$

$$pl_2(h) = \sum_{B \subseteq H} pl_1[B](h) m_0(B) \tag{18}$$

where $m_1[B], bel_1[B], pl_1[B]$ are the result of the conditioning of m_1, bel_1, pl_1 on $B \subseteq H$, respectively, using relations (12), (13) and (14).

G.4 Distinctness.

The concept of 'distinct' pieces of evidence is often left undefined. We propose that it means that once the agent knows the bba m_1 produced by one source of information, what the agent knows about the bba that could be produced by the other source alone is unchanged in comparison to what it was before the agent learns the value of m_1 (see also [30], [3]).

H. Discounting Beliefs

Suppose a source tells that the belief function on H is m, but the user, denoted You, is not sure the source is really reliable (maybe it refers to another question than the one You are interested in). Suppose You believe at level α that the source is reliable, and $1 - \alpha$ it is not, then Your belief m^* on H becomes:

$$m^*(A) = \alpha \; m(A), \; \forall A \neq H \tag{19}$$

$$m^*(H) = 1 - \alpha + \alpha \; m(H). \tag{20}$$

This operation is called a discounting by Shafer [1]. Its interest comes from the possibility to 'discount' sources of information when You feel they are not fully reliable.

Beware that combination and discounting do not commute, so the order with which they are applied is important.

I. Credal and Pignisitc Levels

The TBM is based on the assumption that beliefs manifest themselves at two mental levels: the 'credal' level where beliefs are entertained and the 'pignistic' level where beliefs are used to make decisions (from 'credo' I believe and 'pignus' a bet, both in Latin).

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 held by an ideal agent involved in some decision contexts [31], [32], [33]. This result is accepted here, except that these probability functions quantify the uncertainty only when a decision is really involved.

At the credal level, we defend that beliefs are represented by belief functions. When a decision must be made, the beliefs held at the credal level induce a probability function at the pignistic level. This probability function is needed to compute the expected utilities, and we call it the pignistic probability function, denoted by $BetP$. The transformation between the belief function and the pignistic probability function is called the pignistic transformation (see Section II-J).

One might argue that all that counts in practice are the decisions, not the beliefs, and that the credal level is just an intellectual subtlety. This happens not to be the case as the introduction of the credal level modifies the dynamic of the decisions. Indeed when conditioning is applied, the pignistic transformation of the revised beliefs is not equal to the revised probabilities computed in probability theory.

J. The Pignistic Probabilities for Decision Making

Suppose a bba m^H that quantifies Your beliefs on H. When a decision must be made that depends on the actual value h_0 where $h_0 \in H$, You must construct a probability function in order to take the optimal decision, i.e., the one that maximizes the expected utility. This construction is achieved by the pignistic transformation. Its nature and its justification are defined in [7], [3], [2], [8].

Let $BetP^H$ denote the pignistic probability function You will use to bet on the alternatives in H. $BetP^H$ is a function of H and m^H,

$$BetP^H = \Gamma(m^H, H).$$

We show that the only transformation from m^H to $BetP^H$ that satisfies some rationality requirements is the so called pignistic transformation given by:

$$BetP^H(h) = \sum_{A: h \in A \subseteq H} \frac{m^H(A)}{|A|(1 - m^H(\emptyset))}, \qquad \forall h \in H \tag{21}$$

where $|A|$ is the number of elements of H in A.

It is easy to show that the function $BetP^H$ is indeed a probability function and the pignistic transformation of a probability function is the probability function itself. We call it 'pignistic' in order to avoid the confusion that would consist in interpreting $BetP^H$ as a measure representing Your beliefs on H.

This approach has been shown to resist to the Dutch book argument used by the Bayesians to justify the probabilist approach [2], [34].

III. The General Bayesian Theorem

In probability theory, Bayes theorem permits the computation of a probability function over space H given the value of some variable $x \in X$ from the knowledge of the probabilities over X given each $h_i \in H$, and some *a priori* probability function over H. The same idea has been extended in the TBM context where we will build a belief function over H given an observation $x \subseteq X$ from the knowledge of the belief function over X given each $h_i \in H$ and a vacuous *a priori* belief over H, i.e., an *a priori* describing a state of total ignorance (therefore solving the delicate problem of choosing an appropriate *a priori*). This generalization, called the General Bayesian Theorem (or GBT for short) has been described in [29], [35], [6] and is also discussed in [36]. Identical results have been derived independently by Appriou [37].

Incorporating an *a priori* information on H is achieved by combining the belief induced by x on H with the *a priori* belief using the conjunctive combination rule.

A. The Probability Solution

Suppose the finite spaces X and H. Suppose that for each $h_i \in H$, there is a probability function on X, usually denoted $P(.|h_i)$ but for coherence reasons we denote it as $P^X[h_i]$. Furthermore suppose there is an *a priori* probability function p_0 over H describing the initial beliefs. The problem will be to determine what is the probability function induced on H if $x \subseteq X$ is observed, given the set of conditional probability functions $P^X[h_i]$ and the *a priori* probability function p_0 on H.

The *a posteriori* probability function over H given x, denoted here $P^H[x]$, is given by:

$$P^H[x](h_i) = \frac{P^X[h_i](x)p_0(h_i)}{\sum_{h_j \in H} P^X[h_j](x)p_0(h_j)}, \ \forall h_i \in H, x \subseteq X.$$

It can as well be written as:

$$P^H[x](h_i) = \frac{l(h_i|x)p_0(h_i)}{\sum_{h_j \in H} l(h_j|x)p_0(h_j)}$$

A constant *a priori* probability function is often accepted, even though this attitude may be seriously questioned. In that case, one obtains essentially:

$$P^H[x](h_i) \propto l(h_i|x)$$

a relation that explains why likelihoods are often erroneously confused with probabilities.

When all the ingredients required by this theorem are available and justified, its applicability does not raise any problem. In practice, the likelihoods are usually justified, and result from previous studies. The same cannot be said about the *a priori* probability function p_0. Too often p_0 is ad hoc, if not arbitrary, in which case the applicability of the theorem becomes questionable. The GBT was developed in order to answer to this criticism.

B. The TBM Solution

The GBT performs the same task as the Bayesian theorem but within the TBM context. The major point is that the needed *a priori* belief over H can be a vacuous belief function, what is the perfect representation of total ignorance. We avoid thus one of the major criticisms against the Bayesian approach. Of course, should some *a priori* belief over H be available, it would be combined by the conjunctive combination rule with the result obtained by the GBT.

For the GBT, all that is needed from the sensor after it observes x is the vector of plausibilities $pl^X[h_i](x)$ for all $h_i \in H$. In many cases, the conditional belief over X given h_i is in fact represented by a probability function, in which case $pl^X[h_i](x) = P^X[h_i](x)$.

In order to keep with the tradition, we call $pl^X[h_i](x)$ the likelihood of h_i given x, what we denote by $l(h_i|x)$.

Given the the likelihoods $l(h_i|x)$ for every $h_i \in H$, then for $x \subseteq X$ and for every $A \subseteq H$, Smets [29] proves:

$$m^H[x](A) = \prod_{h_i \in A} l(h_i|x) \prod_{h_i \in \overline{A}} (1 - l(h_i|x)) \tag{22}$$

$$bel^H[x](A) = \prod_{h_i \in \overline{A}} (1 - l(h_i|x)) - \prod_{h_i \in H} (1 - l(h_i|x)) \tag{23}$$

$$pl^H[x](A) = 1 - \prod_{h_i \in A} (1 - l(h_i|x)) \tag{24}$$

$$q^H[x](A) = \prod_{h_i \in A} l(h_i|x) \tag{25}$$

Should You have some non vacuous beliefs on H, represented by $m^H[E_0]$, then this belief is simply combined with $m^H[x]$ by the application of the conjunctive rule of combination.

The GBT has been derived axiomatically by Smets [29], [35], [6] and by Appriou [37].

C. Computational Aspects

C.0.a bel and pl.. Realizing the classical (and sometimes unjustified) criticisms against the computation load required when using belief functions, we produce the MatLab code for computing the previous functions. By definition the vectors m, bel, pl and q are defined on H. Their lengths are $2^{|H|}$ and the position in the vector denotes the subset of H with the convention they are binary ordered. So m(1) is the mass given to $\emptyset$, m(2) is given to $\{h_1\}$, m(3) to $\{h_2\}$, m(4) to $\{h_1, h_2\}$, m(5) to $\{h_3\}$, etc ... This particular order allows us to write computationally efficient - but memory greedy - algorithms.

Let lik denote the likelihood vector with lik(i)= $l(h_i|x)$. The algorithm to compute pl, m and q according to relations (22), (24), and (25) is[2]:

[2]Suppose the vectors $a = \{a_1, a_2 \dots a_n\}$ and $b = \{b_1, b_2 \dots b_m\}$ then in MatLab [a b] denotes the vector $\{a_1, a_2 \dots a_n, b_1, b_2 \dots b_m\}$.

```
m =prod(1-lik); pl = 1; q = 1;
    for i =1:cardinalH
        m=[m m*lik(i)/(1-lik(i))];
        pl = [pl (1-lik(i))*pl];
        q = [q lik(i)*q];
    end
pl = 1-pl;
```

The bel vector is computed directly from the pl vector, using $bel(A) = pl(H) - pl(\overline{A})$.

In case some likelihoods are 1, the set H is transformed into a new set H^* so that $\forall h_i \in H^*,\ l(h_i|x) < 1$ and the computation is performed on H^*. The ballooning extensions of these resulting functions from H^* to H produces the correct results.

As far as in practice, users focus essentially on the belief and the plausibility given to the elements h_i of H, we produce some useful simplified formula. We have:

$$m^H[x](\emptyset) = \prod_{h_i \in H} (1 - l(h_i|x)), \tag{26}$$

then

$$m^H[x](h_i) = bel^H[x](h_i) = m^H[x](\emptyset) \frac{l(h_i|x)}{1 - l(h_i|x)} \tag{27}$$

$$pl^H[x](h_i) = q^H[x](h_i) = l(h_i|x)) \tag{28}$$

Algorithms can be simplified by eliminating those h_i which plausibility is below some threshold α. It consists in fact in a conditioning on $H^\alpha = \{h_i : l(h_i|x) \geq \alpha\}$ and applying the algorithm on H^α which cardinality is usually much smaller then the one of H.

C.0.b BetP.. The computation of $BetP$ can be done using the general formula, but given the special nature of the bba $m^H[x]$, a more efficient algorithm can be defined. Let

$$s(i) = \sum_{|A|=i} \prod_{h_i \in A} l(h_i|x).$$

Let $m0 = m^H[x](\emptyset)$ be computed by relation (26). The MatLab code to compute $BetP(h_i)$ is:

```
x = 1; p = 1; cs = -1;
    for k = 2:cardinalH
        x = s(k-1) - lik(i)*x;
        p = p + cs*x/k;
        cs = - cs;
    end
BetP(i) = lik(i)*p/(1-m0);
```

D. Some Properties

Some particular properties of the GBT are worth mentioning.

D.1 A Way to Derive the GBT

The GBT can be derived from the ballooning extensions (see Section II-F). We start from the bba $m^X[h_i]$ collected for $h_i \in H$. We build its ballooning extension on $X \times H$, and then conjunctively combine these bba's over the h_i's. The result is then marginalized on H and is exactly the bba derived by the GBT.

Formally we have:

$$\begin{aligned} m^H[x] &= (\textcircled{\cap}_{h_i \in H}\ m^X[h_i]^{\Uparrow X \times H})[x]^{\downarrow H} \\ &= \textcircled{\cap}_{h_i \in H}\ (m^X[h_i]^{\Uparrow X \times H}[x]^{\downarrow H}) \end{aligned}$$

The combination can be performed before or after conditioning on x and marginalizing, results are identical.

Furthermore, the bba $m^X[h_i]^{\Uparrow X \times H}[x]^{\downarrow H}$ happens to be a simple support functions[3] on H with a mass $1 - l(h_i|x)$ given to $\overline{h_i}$ and a mass $l(h_i|x)$ given to H [6].

D.2 Independent Observations

We consider the case of two 'independent' observations x defined on X and y defined on Y, and the inference on H obtained from their joint observation.

[3]A Simple Support Function is a belief function where all bbm's are null except for one set and the whole frame.

Suppose the two variables X and Y are conditionally 'independent'. If the beliefs over X and over Y given $h_i \in H$ are represented by probability functions, it means that X and Y are conditionally stochastically independent. In the more general case where the beliefs over X and over Y given $h_i \in H$ are represented by belief functions, the 'independence' requirement becomes what is called the Conditional Cognitive Independence [6], [38]. In both cases, the next property is satisfied;

$$l(h|x, y) = l(h|x)l(h|y) \; \forall h \in H, x \in X, y \in Y. \tag{29}$$

where the likelihood is either the conditional probability or the conditional plausibility of x given h.

The GBT could be applied in two different ways.

Let $pl^H[x]$ and $pl^H[y]$ be computed by the GBT (with a vacuous *a priori* belief on H) from the likelihoods obtained from x and y, separately. They are then combined by the conjunctive rule of combination in order to build $pl^H[x, y]$.

We could as well consider the likelihoods directly obtained from the joint observation x, y, using the product rule (29). We then compute $pl^H[x, y]$ from them using the GBT.

Both results are the same. This property is essential and in fact at the core of the axiomatic derivations of the rule.

D.3 New Hypothesis

The mass given to the empty set by the application of the GBT can receive a nice and useful interpretation. Suppose we accept that H is not exhaustive, thus there are unthought-of hypotheses. Let h^* denote all of them. Then $m^H[x](\emptyset)$ is equal to $bel^{H\cup h^*}[x](h^*)$, thus the degree of belief that the data x supports that none of the hypothesis in H holds, and that we are facing a case where a new previously unthought-of hypothesis must be considered. Identically we have for all $h \subseteq H$:

$$\begin{aligned} bel^{H\cup h^*}[x](h \cup h^*) &= bel^H[x](h) + m^H[x](\emptyset), \\ bel^{H\cup h^*}[x](h) &= 0. \end{aligned}$$

This result is what the GBT produces if we add an extra hypothesis h^*, and define $bel^X[h^*]$ as the vacuous belief function. This is the natural solution, as it is obvious that the user knows nothing about the conditional beliefs over the data when the actual hypothesis belongs to h^*, the set of unthought-of hypotheses.

D.4 The Bayesian Degradation

If for each $h_i \in H$, $pl^X[h_i]$ is a probability function $P^X[h_i]$ on X, then the GBT for $|h_i| = 1$ becomes:

$$pl^H[x](h_i) = P(x|h_i), \qquad \forall x \subseteq X.$$

That is, on the singletons h_i of H, $pl^H[x]$ reduces to the likelihood of h_i given x. The analogy stops there as the solution for the likelihood of subsets of H are different.

If, furthermore, the *a priori* belief on H is also a probability function $p_0(h)$, then the normalized GBT becomes:

$$bel^H[x](A) = \frac{\sum_{h_i \in A} P(x|h_i)p_0(h_i)}{\sum_{h_i \in H} P(x|h_i)p_0(h_i)} = P[x](A)$$

i.e. the (normalized) GBT reduces itself into the classical Bayesian theorem (as it should). This explains the origin of its name.

D.5 Most Plausible and Most Probable Hypothesis

Suppose we know $pl^X[h_i](x)$ for all $x \subseteq X$ and all $h_i \in H$. Data $x \subseteq X$ is collected and want to select which hypothesis $h_i \in H$ is 'best supported' given the observed data. Two strategies have been proposed, one based on selecting the hypothesis with the largest pignistic probability (see Section II-J), the other based on selecting the most plausible hypothesis [37], [39]. It happens that the selected hypothesis is the same for both approaches.

Theorem III.1: Given $x \subseteq X$ and $pl^X[h_i](x)$ for all $h_i \in H$, let $pl^H[x]$ be the plausibility function defined on H and computed by the GBT (relations (22) to (25)), and $BetP^H[x]$ be the pignistic probability function constructed on H from $pl^H[x]$ (relation (21)), then:

$$BetP^H[x](h_i) > BetP^H[x](h_j) \qquad \text{iff} \qquad pl^X[h_i](x) > pl^X[h_j](x).$$

Proof. Let $l(h_k|x) = pl^X[h_k](x)$. Suppose $l(h_k|x) < 1$ for all $h_k \in H$. Let $r_k = l(h_k|x)/(1 - l(h_k|x))$ and $\alpha = \prod_{h_k \in H}(1 - l(h_k|x))$. Then by the GBT (see relation (22)), we have for $h \subseteq H$:

$$m^H[x](h) = \prod_{h_i \in h} l(h_i|x) \prod_{h_i \in \overline{h}} (1 - l(h_i|x)) = \alpha \prod_{h_k \in h} r_k.$$

With $K = \frac{1}{1-m^H[x](\emptyset)}$, we have $BetP^H[x](h_i)$:

$$\begin{aligned} BetP^H[x](h_i) &= K \sum_{h \subseteq \overline{h_i}} \frac{1}{|h|+1} m^H[x](h_i \cup h) \\ &= \alpha K \; r_i \sum_{h \subseteq \overline{h_i}} \frac{1}{|h|+1} \prod_{h_k \in h} r_k \\ &= \alpha K \sum_{h \subseteq \overline{h_i \cup h_j}} \prod_{h_k \in h} r_k \; (\frac{r_i}{|h|+1} + \frac{r_i \; r_j}{|h|+2}) \text{ where } j \neq i \end{aligned}$$

In that case:

$$\begin{aligned} BetP^H[x](h_i) - BetP^H[x](h_j) &= \alpha \; K \sum_{h \subseteq \overline{h_i \cup h_j}} \prod_{h_k \in h} r_k \; (\frac{r_i}{|h|+1} - \frac{r_j}{|h|+1}) \\ &= \alpha \; K(r_i - r_j) \sum_{h \subseteq \overline{h_i \cup h_j}} \prod_{h_k \in h} r_k \; (\frac{1}{|h|+1}) \end{aligned}$$

As $r_k \geq 0$, the product terms are non negative, and so is their sum. The sum is positive as the term with $h = \emptyset$ has a product equal to 1. Hence the sign of the difference is the same as the sign of $r_i - r_j$:

$$BetP^H[x](h_i) > BetP^H[x](h_j) \text{ iff } r_i > r_j.$$

As $r_i > r_j$ iff $l(h_i|x) > l(h_j|x)$, i.e., iff $pl^X[h_i](x) > pl^X[h_j](x)$, the largest value of $BetP^H[x](h_\nu)$ is obtained for the hypothesis h_ν for which $pl^X[h_\nu](x)$ is maximal.

If for $k \in H_0$, $l(h_k|x) = 1$, then every positive bbm on H is given to a superset of h_0, and thus the pignistic probabilities given to the $h_k \in H_0$ are equal and maximal. Simultaneously, $pl^X[h_k](x)$ is always less or equal to 1, so the hypothesis $h_k \in H_0$ are those with a maximal plausibility, hence the theorem. □

This property is very useful when the only purpose is to take a decision and the *a priori* belief on H is vacuous. Indeed all computation can be avoided as all that is needed is $pl^X[h_k](x)$. Of course, the whole computation is still needed when expected utilities and other results are required. This result does not hold when a non vacuous *a priori* belief on H is introduced.

IV. Example 1: A Simple Example of Data Fusion

In order to illustrate the use of the GBT and the pignistic transformation, we present a simple problem of target identification by two sensors. Our examples are inspired by those in [12].

Let S_1 and S_2 be two sensors, an Electronic Support Measure (ESM) and a Radar sensor, respectively. Let X and Y be the domains of the data they can observed, respectively. Let $H = \{F, M, B\}$ be the set of possible targets where the letters denote a F-15, a Mig-27 and a Boeing 747 aircraft, respectively. Table I presents the values of the conditional plausibility functions for $x \subseteq X$ and $y \subseteq Y$, where x and y are the observations made by the two sensors, respectively. Table II presents the computation performed by the GBT. We list the plausibility function induced by data x on H (line $pl^H[x]$), and its related bba (line $m^H[x]$) and commonality function (line $q^H[x]$). For instance,

$$pl^H[x](F, M) = 1 - (1 - .7)(1 - .4) = .82$$

We do the same with data y. We then conjunctively combine the two belief functions by a pointwise multiplication of their commonality functions (line $q^H[x, y]$). We then present the bba (line $m^H[x, y]$) and the normalized belief (line $Bel^H[x, y]$) and plausibility functions (line $Pl^H[x, y]$) ‘ related to the commonality function and that result thus from the conjunctive combination of the belief functions induced on H by x and by y. Finally Table III presents the pignistic probability functions computed after collecting data x alone, data y alone and data x and y jointly. For example,

$$BetP^H[x](F) = (.378 + .252/2 + .041/ + .028/3) \ / \ (1 - .162) = .638$$

The ESM sensor supports the hypothesis that the object is a F-15, whereas the Radar sensor supports that it is a Boeing. But together, they support more strongly that the object is a F-15. This fits with what Table I tells. Hypothesis M and B are each rejected by one sensor, and F is the only hypothesis somehow supported by both sensors. For comparison purposes we also present in Table III what would be the posterior probabilities obtained with the same data using equi *a priori* probabilities on H. Results are very similar in this case. Such a similarity is not always encountered as shown in the next section where the TBM conclusions and the probability conclusions diverge completely.

Sensor	data	F	M	B
ESM	x	.7	.4	.1
Radar	y	.5	.2	.6

TABLE I

EXAMPLE 1. VALUES OF THE CONDITIONAL PLAUSIBILITY FUNCTIONS ON $x \subseteq X$ AND $y \subseteq Y$ GIVEN THE 3 HYPOTHESES F, M AND B IN H. x AND y ARE THE OBSERVATIONS MADE BY THE ESM AND THE RADAR SENSORS, RESPECTIVELY.

V. Conclusions

We have explained how to use the transferable belief model for problems of target identification, data fusion and diagnosis. The transferable belief model is a model developed to represent quantified uncertainty based on belief functions. The major components of the model needed for the present problems are the General Bayesian Theorem that permits to pass from the likelihoods to the posterior beliefs, and the pignistic transformation that permits the construction of the probabilities needed for decision making. These two essential components have received full justifications in [2], [6].

The method we present in fact mimics exactly what is done in probability theory except that every probability function is replaced by a belief function. The latter is more flexible than the former and is able to represent forms of uncertainty difficult to realize within the framework of probability theory. This flexibility allows in particular to solve the problem of the choice of an adequate prior encountered with the classical probabilist approach. Indeed belief functions can represented all forms of uncertainty, from total ignorance to full knowledge. Probability functions occupy an intermediate level in that domain.

The second part of this paper explains how the transferable belief model can be used in practice. It also discusses computational efficiency issues.

H	$\emptyset$	F	M	B	F,M	F,B	M,B	F,M,B
$pl^H[x]$	0	0.7	0.4	0.1	0.82	0.73	0.46	0.838
$m^H[x]$	0.162	0.378	0.108	0.018	0.252	0.041	0.012	0.028
$q^H[x]$	1	0.7	0.4	0.1	0.28	0.07	0.040	0.028
$pl^H[y]$	0	0.5	0.2	0.6	0.6	0.8	0.68	0.84
$m^H[y]$	0.16	0.16	0.039	0.24	0.04	0.24	0.060	0.06
$q^H[y]$	1	0.5	0.2	0.6	0.1	0.3	0.12	0.06
$q^H[x,y]$	1	0.35	0.08	0.06	0.028	0.021	0.004	0.001
$m^H[x,y]$	0.562	0.302	0.048	0.035	0.026	0.019	0.003	0.001
$Bel^H[x,y]$	0	0.691	0.111	0.081	0.862	0.817	0.200	1
$Pl^H[x,y]$	0	0.799	0.182	0.137	0.918	0.888	0.308	1

TABLE II

Example 1. Computation performed by the GBT in order to compute the belief and plausibility functions on H given the x and y data.

References

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

[2] Ph. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.

[3] Ph. Smets, "The transferable belief model for quantified belief representation," In Gabbay and Smets [40], pp. 267–301.

[4] P. Smets and R. Kruse, "The transferable belief model for quantified belief representation," in *Uncertainty in Information Systems: From Needs to Solutions*, A. Motro and P. Smets, Eds., pp. 343–368. Kluwer, Boston, MA, 1997.

[5] Ph. Smets, "Belief functions," in *Non Standard Logics for Automated Reasoning*, Ph. Smets, E. H. Mamdani, D. Dubois, and H. Prade, Eds. 1988, pp. 253–286, Academic Press, London.

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

[7] Ph. Smets, "Constructing the pignistic probability function in a context of uncertainty," in *Uncertainty in Artificial Intelligence 5*, M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, Eds. 1990, pp. 29–40, North Holland, Amsterdam.

[8] Ph. Smets, "Decision making in a context where uncertainty is represented by belief functions," in *Belief Functions in Business Decisions*, R. P. Srivastava, Ed. Physica-Verlag, Forthcoming, 2002.

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

Sensor		F	M	B
ESM	$BetP^H[x]$	0.638	0.298	0.065
Radar	$BetP^H[y]$	0.381	0.131	0.488
ESM+Radar	$BetP^H[x, y]$	0.745	0.147	0.109
ESM+Radar	$P^H[x, y]$	0.714	0.163	0.122

TABLE III

EXAMPLE 1. PIGNISTIC PROBABILITIES COMPUTED ON H GIVEN OBSERVED DATA ARE x, y AND x, y, RESPECTIVELY. THE LAST LINE PRESENTS THE PROBABILITIES ONE WOULD OBTAINED WHEN APPLYING A CLASSICAL PROBABILITY APPROACH, USING AN EQUI *a priori* PROBABILITY FUNCTION ON H.

[10] Y. T. Hsia, "Characterizing belief with minimum commitment," in *Int. Joint Conf. on Artificial Intelligence*, IJCAI-91, Ed. 1991, pp. 1184–1189, Morgan Kaufman, San Mateo, Ca.

[11] E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Boston, 1990.

[12] D. M. Buede and P. Girardi, "A target identification comparison of Bayesian and Dempster-Shafer multisensor fusion," *IEEE Trans. SMC: A*, vol. 27, pp. 569–577, 1997.

[13] Ph. Smets, "Probability, possibility, belief: Which and where?," In Gabbay and Smets [40], pp. 1–24.

[14] A. P. Dempster, "Upper and lower probabilities induced by a multiple valued mapping," *Ann. Math. Statistics*, vol. 38, pp. 325–339, 1967.

[15] A. P. Dempster, "A generalization of Bayesian inference," *J. Roy. Statist. Soc. B.*, vol. 30, pp. 205–247, 1968.

[16] A. P. Dempster, "A class of random convex polytopes.," *Annals of Mathematical Statistics*, vol. 43, pp. 250–272, 1972.

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

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

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

[20] 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, 1995.

[21] Ph. Smets, "An axiomatic justifiaction for the use of belief function to quantify beliefs," in *Int. Joint Conf. on Artificial Intelligence*, IJCAI-93, Ed. 1993, pp. 598–603, Morgan Kaufman, San Mateo, Ca.

[22] Ph. Smets, "The normative representation of quantified beliefs by belief functions," *Artificial Intelligence*, vol. 92, pp. 229–242, 1997.

[23] Ph. Smets, "The combination of evidence in the transferable belief model," *IEEE Pattern analysis and Machine Intelligence*, vol. 12, pp. 447–458, 1990.

[24] F. Klawonn and E. Schwecke, "On the axiomatic justification of Dempster's rule combination," *International Journal of Intelligent Systems*, vol. 7, pp. 469–478, 1992.

[25] F. Klawonn and Ph. Smets, "The dynamic of belief in the transferable belief model and specialization- generalization matrices," in *Uncertainty in Artificial Intelligence 92*, D. Dubois, M. P. Wellman, B. D'Ambrosio, and Ph. Smets, Eds. 1992, pp. 130–137, Morgan Kaufman, San Mateo, Ca.

[26] R. Kennes and Ph. Smets, "Computational aspects of the Möbius transform," in *Uncertainty in Artificial Intelligence 6*, P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, Eds. 1991, pp. 401–416, North Holland, Amsterdam.

[27] R. Kennes and Ph. Smets, "Fast algorithms for Dempster-Shafer theory," in *Uncertainty in knowledge bases*, B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, Eds. 1990, pp. 14–23, Springer Verlag, Berlin.

[28] R. Kennes, "Computational aspects of the Möbius transform of a graph," *IEEE-SMC*, vol. 22, pp. 201–223, 1992.

[29] Ph. Smets, *Un modèle mathématico-statistique simulant le processus du diagnostic médical*, Ph.D. thesis, Université Libre de Bruxelles, (available through University Microfilm International, 30-32 Mortimer street, London W1N 7RA, thesis 80-70,003), 1978.

[30] Ph. Smets, "The concept of distinct evidence," in *Information Processing and Management of Uncertainty*, IPMU-92, Ed., 1992, pp. 89–94.

[31] F. P. Ramsey, "Truth and probability," in *Studies in Subjective Probability*, H. E. Kyburg, Jr. and H. E. Smokler, Eds. 1964, pp. 61–92, Wiley, New York.

[32] L. J. Savage, *Foundations of Statistics*, Wiley, New York, 1954.

[33] M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.

[34] Ph. Smets, "No Dutch book can be built against the TBM even though update is not obtained by bayes rule of conditioning," in *Workshop on Probabilisitic Expert Systems*, R. Scozzafava, Ed. 1993, pp. 181–204, Societa Italiana di Statistica , Roma.

[35] Ph. Smets, "Bayes' theorem generalized for belief functions," in *Proc. ECAI-86*, 1986, pp. 169–171.

[36] G. Shafer, "Belief functions and parametric models," *J. Roy. Statist. Soc. B.*, vol. 44, pp. 322–352, 1982.

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

[38] B. Ben Yaghlane, Ph. Smets, and K. Mellouli, "On conditional belief function independence," in *Independence, Conditioning, Inference*, R. Scozzafava and B. Vantaggi, Eds., pp. 0–0. Workshop on Partial Knowledge and Uncertainty, Rome, Italy., 2000.

[39] A. Nifle and R. Raynaud, "Un argument pour le choix entre décision pignistique et maximum de plausibilité en théorie de l'évidence," *GRETSI'97, Grenoble*, 1997.

[40] D. M. Gabbay and Ph. Smets, Eds., *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 1*, vol. 1. Kluwer, Doordrecht, The Netherlands, 1998.