The Transferable Belief Model for Belief Representation

Philippe SMETS[1]
IRIDIA,
Université Libre de Bruxelles

Rudolf KRUSE[2]
Department of Computer Science
University of Braunschweig

Abstract.
A survey of the use of belief functions to quantify the beliefs held by an agent, and in particular of their interpretation in the transferable belief model.

1. Introduction.

As shown in Smets (1995), there is a variety of imperfect data, be they uncertain or imprecise. Some models have been proposed for each form, but modeling combined forms of imperfect data is hardly achieved. It would nevertheless seems useful to have a common model that would integrate several forms of uncertainty. One way, could be by trying to simulate what might be the human approach of such a problem.

Imprecision about the actual value of an attribute induces some beliefs about it. The information that John is in the 40's induces a belief about John's age, e.g. an equiprobability over 40-49. Imprecision as well as uncertainty induce some beliefs, some subjective opinions held by an agent at a given time about what is the actual value of the variable under consideration.

Bayesians claim this belief is always quantified by a probability measure. Sometimes they go so far as to disregard any other models. The available information is summarized by the probability they generate. Even though the probability model is by far the oldest and the most popular for belief representation, we think it has some limitations that can be handled by some of the alternative or complementary models recently proposed.

The generalization of the Bayesian model has been achieved either by non-standard probability models or by non-probability models (Kohlas, 1994a).

The **non-standard probability models** are those models where beliefs are still somehow quantified by some underlying probability measures.

[1]This work has been partly supported by the CE-ESPRIT III Basic Research Project 6156 (DRUMS II), and the Communauté Française de Belgique, ARC 92/97-160 (BELON).
[2]The work has been partly supported by the CE-ESPRIT III Basic Research Project 6156 (DRUMS II) and by the Fraunhofer Gesellschaft Project No. T/R470/NO191/NO407.

1) Koopman (1940), Good (1962), Kyburg (1961) and Smith (1961) propose that beliefs are quantified by a family of probability measures.

2) Walley (1991) assumes that beliefs held by an agent are quantified by a unique probability measure but the agent cannot expressed the value of these probabilities. He can also assess intervals in which these values fall. Such an approach results in a theory of upper and lower probabilities.

3) Dempster (1967) assumes the existence of a domain X on which there is known probability measure P_X, of another domain Y and of a one-to-many mapping M from X to Y. For instance let $X = \{x_1, x_2\}$, $P_X(x_1) = .7$, $P_X(x_2) = .3$, $Y = \{y_1, y_2\}$, $M(x_1) = \{y_1\}$and $M(x_2) = \{y_1, y_2\}$. The probability measure P_Y on Y induced by P_X and M is not well-defined. Indeed $P_Y(y_1)$ is at least .7 and at most 1, and $P_Y(y_2)$ is at least .0 and at most .7. The imprecision about the value of the probability on the subsets of Y is due to the one-to-many nature of M. Indeed we have no information about how the .3 given to x_2 is distributed among y_1 and y_2. One can at most assess the upper and lower limits of the probabilities that could be allocated to the subsets of Y. The model obtained is Dempster's model. It assumes the existence of a probability measure on the space XxY that quantifies the agent's beliefs over the space XxY. That joint probability measure is partially known. Its marginalization on X is the given probability measure P_X. But given the one-to-many nature of M, it is impossible to decide on the particular distribution of each probability $P_X(\{x\})$ given to the singletons $\{x\}$ of X among the elements of Y. Hence one can only assess the upper and lower probabilities for the subsets of Y. The particularity of this model is resides in the fact that the lower probability function happens to be a belief function. Usually the Dempster-Shafer models described in the Artificial Intelligence literature corresponds to Dempster's model or to some equivalent model as it is the case with the models of related to the probability of knowing (Ruspini 1986), the probability of provability (Pearl 198), the probability of provability, of deductibility (Smets , 1991, 1993b).

4) Other modeling like the hint model of Kohlas (1993), Kohlas and Monney (1994b, 1995) and Shafer's model for evidential reasoning (1976, 1990) are based on similar ideas of an underlying probability space and a one-to-many mapping. Their description of the induced belief on Y is very similar to Dempster's solution except they don't acknowledge the existence of a probability measure over the XxY space that quantifies the agent's opinion about the subsets of XxY. That distinction gets its real importance when revision (conditioning) of beliefs are considered.

5) From the semantical point of view it is often reasonable to separate the representation of information sources from the representation of the data given by these information sources. A uniform approach to the handling of imprecise and uncertain data is discussed in Gebhardt and Kruse (1993).

The **non-probabilistic models** are represented, among others, by the possibilistic model (Bosc and Prade, 1995) and the transferable belief model that is described in this paper. No concept of probability measure is considered. When there is nevertheless some probability measure representing some forms of uncertainty, it is treated just as any other forms of imperfect data. It just induces some possibilities or some beliefs. These are of course related to this probability measure but they are not identical to it.

In this paper, we assume 1) that every form of uncertainty and imprecision **induces** a belief held by an agent at a given time, 2) this belief is quantified by a belief function and 3) the relation between the various forms of imperfect data can be achieved by considering the beliefs they induce[3]. We describe the model for belief representation based on belief functions, called the transferable belief model, and some of its application in Information Sciences. How to appropriately build the belief induced from imperfect data is hardly resolved today and will not be tackled here.

2. The Transferable Belief Model.

The transferable belief model is a model for representing the quantified beliefs held by an agent at a given time on a given frame of discernment Ω. One of the elements of Ω, denoted ϖ, corresponds to the actual state of nature, to the actual world. Unfortunately due to his limited intellectual abilities, the agent is not certain about which element of Ω is ϖ. The agent can only express his subjective opinion about the fact that such or such subset of Ω might contain ϖ. The belief bel(A) given by the agent Y at time t to a subset A of Ω express the strength of the agent's beliefs that ϖ is an element of A based on the information available to Y at time t. This degree of belief is usually quantified by a probability measure as it is the case in the Bayesian approach. The transferable belief model concerns the same problem as the one considered by the Bayesian model except it does not rely on probabilistic quantification but on belief functions.

The **transferable belief model** is based on a two-level model: the *credal level* and the *pignistic level (*from *credo* = I believe and *pignus* = a bet both in Latin). The credal level is the level of intellectual activity where beliefs are entertained. It is where I express the strength of my beliefs about the fact that ϖ belongs to the various subsets of Ω. It is where my knowledge is store, revised, updated, combined, etc......

Beside the credal level, there is also a *pignistic level.* It is also a level of intellectual activity, but completely oriented toward making decisions. When a decision must be made, the beliefs held at the credal level will transmit the needed information to the pignistic level such that optimal decisions are made. The pignistic level contains just the machinery that

[3]Note that we do not say that every form of uncertainty and imprecision is represented by a belief function. We only assume that they induce it.

transform your beliefs (as held at the credal level) into optimal decisions whenever decisions must be made. It has no activity when no decision must be made.

We will assume that the beliefs held at the credal level are quantified by belief functions that will be presented in this survey paper.

When decision must be made, we accept the Savage axioms (Savage, 1954) that claim that optimal decisions can only be achieved if the decider use additive weights to represent his uncertainty on the possible state of affair, and utilities that describes the consequence of each decision in each state of affair. The additive weights corresponds to a probability measure on the state of affairs, but Savage does not claim that such probability measure is THE way the decider beliefs are represented. It only says that when a decision must be made, the decider must generate a set of probabilities on the state of affairs, and make decision by maximizing the expected utility computed from the utilities and that probability measure.

The origin of the probability measure is not discussed in Savage approach. That it could be generate from the belief function that represents the decider's beliefs held at the credal level is perfectly acceptable, even under Savage's approach. Once two distinct levels are considered, it is necessary to describe the nature of the relation between the beliefs held at the credal level, and the probabilities needed at the pignistic level. This relation is described in section 3.9.

Bayesians argue usually that beliefs coexists necessarily with decisions, and do not exist by themselves. We do not share that opinion. That beliefs are necessary ingredients for our decisions does not mean that beliefs cannot be entertained without any revealing behavior manifestations (Smith and Jones, 1986, p.147). We claim that beliefs can indeed be entertained without any concept of decision, justifying thus a distinct credal state. For instances, I can entertain beliefs about meta-physical problems even though I am not going to make any decision about it. I can have some beliefs about the status of the traffic light down my street in Brussels even though I am not in Brussels for the moment and no decision will be made that depends on the color of the traffic light. Beside the consideration of a two-level model with beliefs quantified by belief functions at the credal level has an impact on decision to be made. There are examples where decisions are different depending on the fact one considers the credal level or not. So the distinction is not purely academic.

A full description of the model for beliefs representation based on belief functions can be found in Shafer's book (1976). A somehow revised version appears in Smets (1988). The transferable belief model is described in Smets and Kennes (1994). The axiomatic justification of the use of belief functions to quantify beliefs is given in Smets (1993c). Justifications of the conditioning rule can be found in Klawonn and Smets (1992), in Nguyen and Smets (1993) and in Kruse, Nauck and Klawonn (1991) where differences between various revision concepts are considered. Further results on Bayes theorem and the disjunctive rule of combination appear in Smets (1978, 1993a). Measures of uncertainty related to belief functions are surveyed in Pal et al. (1992, 1993). Many algebraic properties on belief functions can be found in Dubois and Prade (1986a, 1986b). Polemics on the use of belief functions in Artificial Intelligence, essentially in the context of logical deductions, is detailed in two special issues of the International Journal of Approximate Reasoning (in vol. 4, 1990 and vol. 6, 1992). Jeffreys rule of conditioning are presented in Smets (1993d) and the avoidance of Dutch Book is explained in Smets (1993e).

3. The Mathematics of the TBM.

As a didactic tool, we will analyze an example throughout the presentation of the transferable belief model. The sections dealing with the example end with ∇. This example deals with a murder story as it required no contextual background. Of course the example could easily be rephrased into a diagnostic problem, a prediction problem...Furthermore some sections are marked '**Note**' they corresponds to comments unnecessary for the overall understanding of the presentation, but useful for further studies.

Example: A man (Ron) has been murdered. You are the policeman in charge of finding who is (are) the killer(s) (note that the killing might have been committed by several persons jointly). There are three suspects: John, Paul and Sarah. ∇

3.1. The Frame of Discernment.

Let L be a finite **propositional language**.
Example: The three atomic propositions are: 'John is a murderer of Ron', 'Paul is a murderer of Ron', 'Sarah is a murderer of Ron'. They are denoted by J, P and S, respectively. ∇

Let $\Omega = \{\omega_1, \omega_2, ...\omega_n\}$ be the set of **worlds** that correspond to the interpretations of L.
Example: The eight possible worlds are:
$\omega_1 = \neg J \wedge \neg P \wedge \neg S$ (none of the three suspects is a murderer of Ron)
$\omega_2 = \neg J \wedge \neg P \wedge S$ (Sarah is the only murderer among the three suspects)
$\omega_3 = \neg J \wedge P \wedge \neg S$ (Paul is the only murderer among the three suspects)
$\omega_4 = \neg J \wedge P \wedge S$ (Paul and Sarah are murderers, John is not)
$\omega_5 = J \wedge \neg P \wedge \neg S$ (John is the only murderer among the three suspects)
$\omega_6 = J \wedge \neg P \wedge S$ (John and Sarah are murderers, Paul is not)
$\omega_7 = J \wedge P \wedge \neg S$ (John and Paul are murderers, Sarah is not)
$\omega_8 = J \wedge P \wedge S$ (John, Paul and Sarah are murderers) ∇

Propositions identify subsets of Ω.

Example: The proposition 'Paul is a murderer of Ron' identifies the subset $\{\omega_3, \omega_4, \omega_7, \omega_8\}$ of Ω. ∇

Beliefs and probabilities given to propositions can thus be identically considered as beliefs and probabilities given to subsets of Ω. The set notation will be used hereafter. By definition there is an **actual world** ϖ and it is an element of Ω. For $A \subseteq \Omega$, bel(A) and P(A) denote the degrees of belief and probability that the actual world ϖ belongs to A, respectively. For simplicity sake, we admit that bel and P will be defined for every subsets of Ω, so bel and P are function from 2^Ω to [0,1]. Ω is called the **frame of discernment**.

All beliefs entertained by You[4] at time t about which world is the actual world ϖ are defined relative to a given **evidential corpus** (EC_t^Y) i.e., the set of pieces of evidence in Your mind at time t. Our approach is normative: You is an ideal rational agent and EC_t^Y is deductively closed. Your **credal state** on a frame of discernment Ω describes Your subjective, personal judgment that $\varpi \in A$ for every subset A of Ω. By a classical abuse of language, the actual world ϖ is called the 'true' world, and we say that 'A is true' or 'the truth is in A' to mean that $\varpi \in A$. Your credal state results from EC_t^Y that induces in You some partial beliefs on the subsets of Ω. These partial beliefs quantify the strength of Your belief that $\varpi \in A$, $\forall A \subseteq \Omega$. It is an epistemic construct as it is relative to Your knowledge that is included in Your evidential corpus EC_t^Y.

Example: In order to simplify the example, suppose You know that one and only one of the three suspects is the murderer of Ron. So EC_t^Y contains the fact that Ron has been murdered by a single person, the murder is one of John, Paul and Sarah. The available information can be translated by $\varpi \in \{\omega_2, \omega_3, \omega_5\}$. ∇

3.2. The Basic Belief Masses.

Example: A partially reliable witness testifies to You that the murderer is a male. Let $\alpha = .7$ be the reliability You give to the testimony. Suppose that *a priori* You have an equal belief that the murderer is a male or a female. A classical probability analysis would compute the probability P(M) of M where M = 'the killer is a male': P(M) = .85 = .7 + .5 x.3 (the probability that the witness is reliable (.7) in which case M is true for sure, plus the probability of M given the witness is not reliable (.5) weighted by the probability that the witness is not reliable (.3)). The .7 can be viewed as the *justified* component of the probability given to M whereas the .15 can be viewed as the *aleatory* component of that probability.

This analysis is not the one proposed in the TBM. The TBM deals only with the justified components. It gives a belief (or support) .7 to M. The .7 and .3 are parts of an initial

[4]'You' is the agent that entertains the beliefs considered in this presentation.

unitary amount of belief that supports $J \vee P$ (the males) and $J \vee P \vee S$ (anybody), respectively. These parts are called the basic belief masses (bbm). They are denoted by $m(J \vee P) = .7$, $m(J \vee P \vee S) = .3$. The .7 that supports $J \vee P$ supports the fact that the murdered is either John or Paul. It is kept to the disjunction $J \vee P$ (i.e., $\{\omega_3,\omega_5\}$). It is not further distributed between $\{\omega_3\}$ and $\{\omega_5\}$. The .7 supports $\varpi \in \{\omega_3,\omega_5\}$ without supporting any proposition strictly more specific like $\{\omega_3\}$ and $\{\omega_5\}$. Identically the .3 is a 'specific' support given to $\varpi \in \{\omega_2,\omega_3,\omega_5\}$ that cannot be distributed more specifically among the subsets of $\{\omega_2,\omega_3,\omega_5\}$. This basic belief masses (the .7 and .3) allocation is at the core of the assumptions underlying TBM. ∇

Formally the TBM is defined as follows.

Basic Assumption.
*The **TBM** postulates that the impact of a piece of evidence on an agent is translated by an allocation of parts of an initial unitary amount of belief among the subsets of Ω. For $A \subseteq \Omega$, m(A) is a part of the agent's belief that supports A, i.e. that the 'actual world ϖ is in A, and that, due to lack of information, does not support any strict subset of A.*

The m(A) values, $A \subseteq \Omega$, are called the **basic belief masses** (bbm) and the m function is called the **basic belief assignment**[5].

Let $m: 2^{\Omega} \rightarrow [0,1]$ with

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

Every $A \subseteq \Omega$ such that $m(A)>0$ is called a focal proposition. The difference with probability models is that masses can be given to any subsets of Ω instead of only to the elements of Ω as it would be the case in probability theory.

3.3. Conditioning.

Example: Suppose You learn that Paul was not the murderer as he was dead the day before Ron was murdered (a perfect alibi). The world ω_3 is thus impossible and You know for sure that $\varpi \in \{\omega_2,\omega_5\}$. The bbm .7 that was initially allocated to $\{\omega_3,\omega_5\}$ is now supporting specifically that the murderer is John $\{\omega_5\}$, and the bbm .3 initially allocated to $\{\omega_2,\omega_3,\omega_5\}$ now supports that the murderer is John or Sarah, i.e., $\varpi \in \{\omega_2,\omega_5\}$. Indeed the reliability .7 You gave to the testimony initially supported 'the murderer is John or Paul'. The new information about Paul implies that the .7 now supports 'the murderer is John'. ∇

[5] Shafer speaks about basic probability masses and assignment. To avoid confusion, we have banned the word "probability" whenever possible.

The impact of the revision by the knowledge that $\varpi \in B \subseteq \Omega$ results in a transfer for each $A \subseteq \Omega$ of the bbm m(A) initially allocated to A to $A \cap B \subseteq \Omega$. Hence the origin of the name of the TBM (transferable belief model).

The transfer of belief described in the TBM corresponds to the **unnormalized rule of conditioning**[6]. Let m be a basic belief assignment on the frame of discernment Ω and suppose the conditioning evidence tells You that the truth is in $B \subseteq \Omega$, the basic belief assignment m is transformed into $m_B: 2^\Omega \rightarrow [0,1]$ with:

$$m_B(A) = \sum_{X \subseteq \overline{B}} m(A \cup X) \qquad \text{for } A \subseteq B \tag{3.1}$$

$$m_B(A) = 0 \qquad \text{for } A \not\subseteq B$$

Note: In this presentation we have accepted that a non null basic belief mass could be given to Ø. In most presentation of the models based on belief functions, it is assumed that m(Ø)=0. The meaning of the basic belief mass given to Ø is analyzed in Smets (1992b). It corresponds to the amount of contradiction present in the basic belief assignment m, as could be encountered when two sources of information give some support to contradictory hypothesis.

3.4. Belief and Plausibility Functions.

Given Ω, the **degree of belief** of $A \subseteq \Omega$, bel(A), quantifies the total amount of *justified specific support* given to A. It is obtained by summing all the basic belief masses given to propositions $X \subseteq A$ (and $X \neq Ø$). Let bel : $2^\Omega \rightarrow [0,1]$ with:

$$\text{bel}(A) = \sum_{Ø \neq X \subseteq A} m(X)$$

We say *justified* because we include in bel(A) *only* the basic belief masses given to subsets of A. For instance, consider two distinct elements ω_1 and ω_2 of Ω. The basic belief mass $m(\{\omega_1,\omega_2\})$ given to $\{\omega_1,\omega_2\}$ could support ω_1 if further information indicates this. However given the available information the basic belief mass can only be given to $\{\omega_1,\omega_2\}$. (Note: as m(Ø) might be positive, it should not be included in bel(A) (nor in pl(A), see below), as m(Ø) is given to the subset Ø that supports not only A but also $\overline{A}$. This is the origin of the *specific* support.)

The function bel is called a belief function. Belief functions satisfy the following inequalities (Shafer 1976):
$\forall n \geq 1,\ A_1, A_2, \ldots A_n \subseteq \Omega,$

---

[6] When m(Ø)=0 is assumed, the result is further normalized by dividing each term in m_B by $1-m_B(Ø)$. The resulting conditioning rule is then called the Dempster rule of conditioning..

$$\mathrm{bel}(A_1 \cup A_2 \cup \ldots A_n) \geq \sum_i \mathrm{bel}(A_i) - \sum_{i>j} \mathrm{bel}(A_i \cap A_j) \ldots . -(-1)^n \mathrm{bel}(A_1 \cap A_2 \cap \ldots A_n) \qquad (3.2)$$

The **degree of plausibility** of A, pl(A), quantifies the maximum amount of *potential specific support* that could be given to $A \subseteq \Omega$. It is obtained by adding all those basic belief masses given to propositions X compatible with A, i.e. such that $X \cap A \neq \emptyset$. Let pl : $2^{\Omega} \rightarrow [0,1]$ with:

$$\mathrm{pl}(A) = \sum_{X \cap A \neq \emptyset} m(X) = \mathrm{bel}(\Omega) - \mathrm{bel}(\overline{A})$$

We say *potential* because the basic belief masses included in pl(A) could be transferred to non-empty subsets of A if some new information could justify such a transfer. It would be the case if we learn that $\overline{A}$ is impossible.

The function pl is called a plausibility function. It is in one-to-one correspondence with the belief function induced by the same bba. It is just another way of presenting the same information and could be forgotten, except inasmuch as it provides a convenient alternate representation of our beliefs.

The unnormalized rule of conditioning expressed with bel and pl is:

$$\mathrm{bel}(A|B) = \mathrm{bel}(A \cup \overline{B}) - \mathrm{bel}(\overline{B}) \qquad \mathrm{pl}(A|B) = \mathrm{pl}(A \cap B)$$

Total ignorance is a state of belief that is hard to represent in probability theory. Bayesians will reject the existence of such a state, avoiding thus many problems related to its representation by a probability function. In the TBM, total ignorance is beautifully represented by a 'vacuous' belief function, i.e. a belief function such that $m(\Omega) = 1$, hence $\mathrm{bel}(A) = 0$ $\forall A \subseteq \Omega$, $A \neq \Omega$, and $\mathrm{bel}(\Omega) = 1$. That belief function does not support any particular subset of Ω. None of the subsets of Ω is supported (except Ω itself) and all subsets receive the same degree of belief, what should be the case under total ignorance. In particular when $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathrm{bel}(\{\omega_1\}) = \mathrm{bel}(\{\omega_2\}) = \mathrm{bel}(\{\omega_3\}) = \mathrm{bel}(\{\omega_1, \omega_2\})$, a representation that cannot be achieved in probability theory.

3.5. The Rules of Combination.

Example: A second witness, with reliability .6, tells You that the murderer is John or Sarah. How to combine the bba m_1 (that represent Your beliefs on Ω as induced by the testimony of the first witness) and m_2 (that represent Your beliefs on Ω as induced by the testimony of the second witness) into a new bba m_{12} that will represent Your beliefs on Ω as induced by the joint testimony of the two witnesses. When the two witnesses are fully reliable, their joint testimony support that the murderer is John, and the combined bbm $m_{12}(J) = .7 \times .6 = .42$ (the origin of the product is commented below). When the first witness is reliable and the second is not, their joint testimony supports that the murderer

John or Paul, so $m_{12}(J \wedge P) = .7 \times .4 = .28$. Similarly $m_{12}(J \wedge S) = .3 \times .6 = .18$ and $m_{12}(J \wedge P \wedge S) = .3 \times .4 = .12$. In fact the proposed combination rule consists in allocating the product of two bba $m_1(A) \times m_2(B)$ to the intersection $A \cap B$ of their focal elements A and B. ∇

Formally, suppose two belief functions bel_1 and bel_2 induced by two 'distinct' pieces of evidence. The question is to define a belief function $bel_{12} = bel_1 \oplus bel_2$ resulting from the **combination** of the two belief functions, where the $\oplus$ symbolizes the combination operator. Shafer proposed to use Dempster's rule of combination in order to derive bel_{12}. The underlying intuitive idea is that the product of two bbm $m_1(X)$ and $m_2(Y)$ induced by the two distinct pieces of evidence on Ω supports $X \cap Y$, hence:

$$m_{12}(A) = \sum_{X \cap Y = A} m_1(X) \cdot m_2(Y) \qquad (3.3)$$

Note: Dubois and Prade (1986a), Smets (1990a), Klawonn and Schwecke (1990), Klawonn and Smets (1992) and Hajek (1992) provide different justifications for the origin and the unicity of this rule. These justifications are obtained without introducing some underlying probability concepts. They are based essentially on the associativity and commutativity properties of the combination operator.

Note: Dempster's rule of combination is a rule to combine conjunctive pieces of information. Let bel_1 and bel_2 be the belief functions induced by the two distinct pieces of evidence E_1 and E_2, respectively. Then bel_{12} is the belief function induced on Ω by the conjunction 'E_1 and E_2'. In Smets (1993a) we present the disjunctive rule of combination that allows us to derive the belief function induced on Ω by the disjunction of E_1 and E_2. It corresponds to a situation where you could assess Your belief on Ω if E_1 were true, Your belief on Ω if E_2 were true, but You only know that the disjunction 'E_1 or E_2' is true.

Directly in relation to that disjunctive rule of combination, we derive (Smets 1993a) the **Generalized Bayesian Theorem**. Suppose You have a vacuous a priori (i.e., You are in a state of total ignorance) on a space Θ. Suppose that for each $\theta_i \in \Theta$, You know what would be Your beliefs on another space X if θ_i happened to be the case. Let $bel_X(.|\theta_i)$ and $pl_X(x|\theta_i)$ be these conditional belief and plausibility functions on X given each $\theta_i \in \Theta$. Suppose You learn that $x \subseteq X$ is the case. The Generalized Bayesian Theorem allows You to derive the conditional belief $bel_\Theta(.|x)$ and plausibility $pl_\Theta(\theta \mid x)$ on the frame of discernment Θ given an observation $x \subseteq X$. One has:

$$bel_\Theta(\theta \mid x) = \prod_{\theta_i \in \overline{\theta}} \left(bel_X(\overline{x} \mid \theta_i)) + m(\emptyset|\theta_i) \right) - \prod_{\theta_i \in \Theta} \left(bel_X(\overline{x} \mid \theta_i)) + m(\emptyset|\theta_i) \right)$$

$$pl_\Theta(\theta \mid x) = 1 - \prod_{\theta_i \in \theta} (\ 1 - pl_X(x|\theta_i)\ )$$

Note: The theorem has been further generalized when there is some non vacuous beliefs on Θ (Smets 1993a).

Note: All combinations were performed for beliefs induced by 'distinct' pieces of evidence. The concept of 'distinctness' is presented in Smets (1992c). It fits essentially and intuitively with the idea that the two pieces of evidence involved are 'unrelated', 'independent', that the knowledge of any of them does not interfere with the beliefs that would be specifically induced by the other.

Note: The problem on **how to combine non distinct pieces of evidence** has been first considered in Smets (1986). Ling and Rudd (1989) and Kennes (1991) introduce the concept of a **cautious rule of (conjunctive) combination**. It is based on the idea that each expert provides a belief function that results from his/her own expertise plus a common background. The rule permits to disentangle the underlying common background. It is idempotent. Ling and Rudd (1989) solved the case where the experts opinions are described by simple belief functions, i.e., belief functions with two focal elements, Ω and a non empty strict subset of Ω. Kennes (1991) presents the solution when the experts opinions are described by separable belief functions, i.e., belief functions obtained by applying Dempster's rule of combination to several simple support functions). The generalization to any pair of belief functions is under way. Its use for pooling expertise provided by experts who share a common background will be studied in a forthcoming paper. In the present paper we restrict ourselves to the idealized situation where the experts are 'independent' , i.e. the experts do not communicate together and do not use common evidence.

3.6. Justifications for the Use of Belief Functions.

The **use of belief function at the credal level** can be justified by at least three different approaches. Initially Shafer suggests to justify the use of belief functions by claiming that any measure of belief must satisfy the inequalities of relations 3.2. In the TBM, we prefer to start from the basic belief masses that represent a certain part of our belief allocated to a proposition and that cannot be allocated to more specific propositions. Finally in (Smets, 1993c), we proposes a set of axioms that should be satisfied by any measure of belief and this set of axioms justifies the use of belief functions to represent quantified beliefs.

3.7. Discounting.

Example: Let us forget about the previous testimonies. Suppose an agent H tells You that His beliefs over Ω is such that $m_H(\{\omega_1\}) = .2$, $m_H(\{\omega_1,\omega_2\}) = .3$, $m_H(\{\omega_1,\omega_2,\omega_3\}) = .5$.

Should You have no other evidence over Ω, You would adopt m_H as representing Your beliefs on Ω. But suppose now that You have some doubt about H reliability. Let .7 be Your degree of belief that H is reliable, or equivalently he strength with which You believe what H is saying. Let Your beliefs about H saying be represented by the bbm m_0(reliable) = .7 and m_0(reliable or not reliable) = .3. How to combine the beliefs induced in You by H's beliefs on Ω and Your beliefs about H's reliability. The idea (easily justified) is that You should discount H's beliefs by multiplying m_H by a factor .7 (m_0(reliable)) and transfer to Ω the part of m_H that has been lost. Let m_Y be the resulting bba. One obtains:

$m_Y(\{\omega_1\})$	= .2 x .7	= .14
$m_Y(\{\omega_1,\omega_2\})$	= .3 x .7	= .21
$m_Y(\{\omega_1,\omega_2,\omega_3\})$	= .5 x .7 + (1-.7)	= .65

∇

Formally, suppose You have no belief whatsoever on a frame of discernment Ω, but a somehow reliable agent communicates to You his beliefs on Ω, beliefs represented by the basic belief assignment m_Ω. Should the agent be fully reliable, You would accept his beliefs and would be tempted to adopt his beliefs as Yours. But the agent is not fully reliable. Let m_0 represents Your *a priori* beliefs about the reliability of this agent, with m_0(reliable) = 1-α and m_0(reliable or not reliable) = α. Combining Your *a priori* belief m_0 with m_Ω on Ω provided by the agent leads to the discounted belief bel_Ω^α that quantifies Your belief on Ω induced by both Your *a priori* and the agent's beliefs (Shafer, 1976 pg 251, introduces the concept of discounting factors, Smets (1993a) explains its origin). bel_Ω^α is such that :

$$\forall A\subseteq\Omega,\ A\neq\Omega, \qquad bel_\Omega^\alpha(A) = (1-\alpha)\ bel_\Omega(A)$$

$$\text{and} \qquad bel_\Omega^\alpha(\Omega) = bel_\Omega(\Omega).$$

3.8. Static and Dynamic Components.

It is important to note that the TBM includes **two components**: one **static**, the basic belief assignment, and one **dynamic**, the transfer process underlying the conditioning, the combination processes and the discounting processes. Many authors on Dempster-Shafer model consider only the basic belief assignment and discover that the basic belief masses are probabilities on the power set of Ω. But usually they do not study the dynamic component, and their comparisons are therefore incomplete, if not misleading. The transfer of belief masses is studied in several papers such as Kruse, Schwecke and Klawonn (1991) using the more general concept of specialization (Kruse and Schwecke, 1991).

3.9. The Pignistic Probability BetP.

Example: After the first testimony, if You had to bet on who is the murderer, or equivalently on which of the worlds in { } corresponds to the actual world ϖ, You would have to build a probability function BetP on Ω. Remember that Your bba was described by a bbm .7 on J$\vee$P and .3 on J$\vee$P$\vee$S. The .7 could as well be given to John or to Paul. So we could distribute it equally between John and Paul. Identically the .3 could be distributed equally among the three suspects. In that case the probability BetP would be:

BetP(John) = .7/2 + .3/3 = .45

BetP(Paul) = .7/2 + .3/3 = .45

BetP(Sarah) = .3/3 = .10

That such a solution is indeed the only adequate one is not discussed here. ∇

In Smets (1990b) and Smets and Kennes (1994), we show how to make **decisions** when the beliefs are quantified by belief functions. (see also Stratt (1989, 1990) and Jaffray (1988) for other solutions). The satisfaction of some natural rationality requirements leads to the derivation of a unique transformation between the belief functions and the probability functions that must be used once decisions must be made. We call this transformation the pignistic transformation. Let BetP(A) be the pignistic probability derived from the bba m : $2^\Omega \rightarrow [0,1]$. BetP is just a classical probability measure, but we denote it BetP to avoid any confusion. BetP deals with betting weights, not with beliefs quantification whereas bel deals with such beliefs quantification. One has:

$$\mathrm{BetP}(A) = \sum_{B \subseteq \Omega} m(B) \frac{|A \cap B|}{|B|} \qquad \forall A \subseteq \Omega \tag{3.4}$$

where $|A|$ is the number of elements of Ω of A. BetP is the appropriate probability function to be used to make decisions (using expected utilities theory).

BetP is the only probability measure that satisfies the following **rationality requirement** (Smets, 1990b). Suppose two source of evidence E_1 and E_2 that induce the belief functions bel_1 and bel_2 on the same frame Ω and a random device that select the source that will be available to You. If You knew the evidence that would be selected (E_1 or E_2), You would bet on Ω according to a probability function P_i induced from the corresponding bel_i (i=1,2) by the pignistic transformation. These probability functions P_i are the conditional probability functions on Ω given E_i. Prior to selecting the source, the probability measure P on Ω is then $P(A) = p\, P_1(A) + (1-p)\, P_2(A)$ for all $A \subseteq \Omega$. But prior to the selection of the source, Your belief on Ω is given by $bel(A) = p\, bel_1(A) + (1-p)\, bel_2(A)$ for all $A \subseteq \Omega$. The probability function induced from that combined belief function bel should be equal to P. The only transformation that satisfies this requirement is the pignistic transformation as described by relation 3.4. Note that Stratt and Jaffray's solutions do not satisfy the rationality requirement we have just described.

Note: In this transformation the bbm m(A) given to a focal element $A \subseteq \Omega$ is distributed equally among the elements of A. For $\omega \in \Omega$, BetP(ω) results from the addition of all these parts of masses allocated to ω. It is quite similar to the application of the Principle of Insufficient Reason at the level of each bbm, but its justification is NOT based on the assumption of some Insufficient Reason.

Note that BetP is not a representation of Your beliefs on Ω. It is the additive measure induced on Ω by Your beliefs held at the credal level (and quantified in the transferable belief model by a belief function) when decision must be made and that must be used to compute the expected utility to be maximized in order to select the optimal decision.

4. Applications in Data Bases.

We will present three applications related to data base, source reliability and medical diagnosis.

Uncertainty, and beliefs, can be encountered at two levels in a DB: it can be described at the attribute level or at the tuple level.

At the attribute level, it might occur that the actual value of an attribute in a given tuple is not exactly known. The crudest way to represent the available knowledge consists in presenting the value of the attribute as the disjunctive e set of possible values compatible with the available knowledge. The next level of sophistication consists in acknowledging that some values are 'better' than other, either more possible, more probable, more believable. One end up with a possibility, a probability or a belief distribution over the domain D of the value of the attribute. Conceptually the three representations are very similar. They diverge in their interpretations. The value of the attribute is made of the disjunction of several weighted subsets of the domain D of the value.

The TBM could be applied if the knowledge about the actual value of the attribute could only be described by our belief about it. In that case, the value of the attribute will be described by a basic belief assignment m (or any related functions) on the domain D.

Suppose the attribute is the age of the individual characterized by a tuple in a relation. Suppose the age of John is known only by a basic belief assignment on D = [0, 120]. For instance the available information about John's age is described by the following bbm : m([40-49]) = .6, m([30, 59]) = .3, m([0, 120])=.1. It could represent the information : John is in his forties (.6 given to [40, 49]), or almost (.3 given to [30, 59], but the source is not fully reliable (hence the .1 on D).

Suppose now you want to select those tuple where the age is between 35 and 52: should you select John? One way consist in creating a new relation R = 'Age 35-52', with all cases that satisfy the selection criteria. John' tuple would be included, with some weights that represent the degree of beliefs and plausibility that the John's tuple belongs to the relation. So the belief that John belongs to the relation is bel(John belongs to R) = .6 and pl(John belongs to R) = 1. So for each tuple one has a pair of weights that quantify the degree of belief and plausibility that the tuple belongs to the relation, what correspond to the second level where uncertainty can be encountered in a DB.

5. Application with Sources Reliability.

The source of the information introduced in a DB can re recorded and used for discounting data when relevant. Suppose the relation AGE given in table 5.1. It means that Paul tells that John is 45, that Peter tells that he believes that Henri is 45 at level .6 and has no more specific knowledge about Henri's age, that Peter tells that Jack is between 30 and 40... Information '45' is equivalent to m([45]) = 1.

Name	Age	Source
John	45	Paul
Henri	m([45])=.6	Peter
	m([0,120]) = .4	
Jack	m([30-40]) = 1	Peter
Jim	m([32]) = 1	Peter
Phil	m([54] = .5	Paul
	m([52, 55] = .3	
	m([40, 50]) = .2	
Henri	m([30-50]) = 1	Paul

Table 5.1: relation AGE.

You, the user of the DB might have some opinions about the reliability of the sources, as given in table 5.2. Reliability can be easily iterated by introducing reliability about the reliability of the sources.

Source	Reliability
Paul	m(reliable) = .7
Peter	m(reliable) = .8

Table 5.2: Sources reliability according to You.

Suppose You want to assess Your beliefs about Henri's age. Peter is the only source about Henri's age. You use the belief presented in table 5.1 discounted by the factor .2. (given in table 5.2). Computation is done as detailed in section 3.7. You end up with a bbm m([45]) = .8 x .6 = .48 and m([0, 120]) = 1 - .48 = .52.

Name	Age
John	m([45] = .7
	m([0,120]) = .3
Henri	m([45])=.48
	m([0,120]) = .52
Jack	m([30-40]) = .8
	m([0,120]) = .2
Jim	m([32]) = .8
	m([0,120]) = .2
Phil	m([54] = .35
	m([52, 55] = .21
	m([40, 50]) = .14
	m([0,120]) = .3
Henri	m([30-50]) = .7
	m([0-120]) = .3

Table 5.3: Your relation AGE (after applying Your discounting factors on the Sources).

Table 5.3 presents Your relation AGE after You have applied the discounting factors of table 5.1. One notice that Henri's age is provide by two sources (Paul and Peter). One could have combined the two bbm presented in table 5.1, but as far as Peter and Paul are two different sources of information, discounting must be performed before combination. The result obtained by combining the two belief functions over Henri's age given in table 5.3 is:

m([45])	= .48 x (.7 + .3)	= .480
m([30-50])	= .52 x .7	= .364
m([0-120])	= .52 x .3	= .156

Your final AGE relation is the one given in table 5.3, after deleting all data related to Henri and replacing them by the data just computed. Table 5.4 presents the list of cases with age in [44, 56]. As far as Your knowledge about the ages is never certain, You can only assess Your belief and plausibility for each case that it belongs to the set of cases with age in [44, 56]. If You want to select only those cases for which Your belief is larger than .5, You would select John and Phil.

Name	bel	pl
John	.70	1.00
Henri	.48	1.00
Jack	.00	.20
Jim	.00	.20
Phil	.56	.86

Table 5.4. Relation AGE $\in$ [44-56].

6. Application for Diagnosis.

The major advantage of the TBM approach for diagnostic purpose resides in the fact the Bayesian Theorem can be applied in cases where there is no prior beliefs on the set of diagnosis (Smets 1981). This solves the major criticism addressed to the Bayesian approach of the diagnostic process: where comes the *a priori* beliefs on the set of diagnosis from ? Beside, the TBM approach profits from the already explained advantage that one must provide only the available information on the set of symptoms. No forced probabilization is required. Within each diagnostic class, one must provide only what symptoms are supported and how much, not a full probability function on the set of symptoms. One can even introduce a diagnostic class made of the 'still unknown diseases' in which case the beliefs over the symptoms is of course vacuous. Furthermore one may compute the support given to the fact that the patient belongs to the class of the 'still unknown diseases'. This cannot be achieved in probability theory, as the state of complete ignorance on the symptom domain as required in the 'still unknown disease' class cannot be adequately represented in probability theory.

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

We consider two sets X and Y of symptoms with $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2\}$. Tables 6.1 and 6.2 present the beliefs over X and Y when each of the individual diseases holds. The beliefs translate essentially the facts that θ_1 'causes' (supports) x_3 and y_2, and $\theta2$ 'causes' x_1 or x_2 (without preference) and y_1. When we only know that θ_1 or θ_2 holds, then we have a balanced support over X, and some support in favor of y_1.

X	$\{\theta_1\}$ m	bel	$\{\theta_2\}$ m	bel	$\{\theta_\omega\}$ m	bel
$\{x_1\}$	.0	.0	.0	.0	.0	.0
$\{x_2\}$	.0	.0	.0	.0	.0	.0
$\{x_3\}$	.5	.5	.2	.2	.0	.0
$\{x_1, x_2\}$	.2	.2	.6	.6	.0	.0
$\{x_1, x_3\}$	.0	.5	.1	.3	.0	.0
$\{x_2, x_3\}$	.0	.5	.1	.3	.0	.0
$\{x_1, x_2, x_3\}$	.3	1.0	.0	1.0	1.0	1.0

Table 6.1: Conditional beliefs (bel) and bbm (m) on the symptoms $x \subseteq X$ within each of the mutually exclusive and exhaustive diagnosis θ_1, θ_2 and $\theta_\omega \in \Theta$.

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

Table 6.2: Conditional beliefs (bel) and bbm (m) on the symptoms $y \subseteq Y$ within each of the mutually exclusive and exhaustive diagnosis θ_1, θ_2 and $\theta_\omega \in \Theta$.

Table 6.3 presents the beliefs induced on Θ by the individual observation of symptom x_3 or of symptom y_2, respectively. We assume that the symptoms are independent within each disease, hence the GBT can be applied. The independence assumption means that if we knew which disease holds, the observation of one of the symptoms would not change our belief about the status of the other symptom. The right half of table 6.3 presents the beliefs induced on Θ by the joint observation of symptoms x_3 and y_2. The beliefs are computed by the application of relation in section 3.5. The symptoms individually and jointly support essentially $\{\theta_1, \theta_\omega\}$. The meaning of $bel(\theta_\omega|x_3, y_2) = 0.27$ merits some consideration. It quantifies our belief that the joint symptoms x_3 and y_2 are neither 'caused' by θ_1 nor by θ_2. It supports the fact that the joint observation is 'caused' by another disease or by some still unknown disease. A large value for $bel(\theta_\omega|x_3, y_2)$ somehow supports the fact that we might be facing a new disease. In any case it should induce us in looking for other potential causes to explain the observations.

	$\vert x_3$	$\vert y_2$		$\vert x_3, y_2$	
Θ	m	m	m	bel	pl
$\{\theta_1\}$	.00	.00	.00	.00	.64
$\{\theta_2\}$	.00	.00	.00	.00	.24
$\{\theta_\omega\}$	.12	.08	.27	1.27	1.00
$\{\theta_1, \theta_2\}$	.00	.00	.00	.00	.73
$\{\theta_1, \theta_\omega\}$	.48	.32	.49	.76	1.00
$\{\theta_2, \theta_\omega\}$	.08	.12	.09	.36	1.00
$\{\theta_1, \theta_2, \theta_\omega\}$	.32	.48	.15	1.00	1.00

Table 6.3: Left part: the basic belief masses (m) induced on Θ by the observation of symptom x_3 or of symptom y_2, ac computed from the Generalized Bayesian Theorem (section 3.5). Right part, the basic belief masses (m) and the related belief function (bel) and plausibility function (pl) induced on Θ by the joint observation of x_3 and y_2, as computed by the application of Dempster's rule of combination (section 3.5) on the bbm obtained after observing x_3 and y_2, respectively.

Table 6.4 presents the beliefs induced on $\{\theta_1, \theta_2\}$ when we condition our beliefs on Θ on the fact$\{\theta_1, \theta_2\}$, or when we have some *a priori* belief on Θ. The results are obtained by the application of the conjunctive rule of combination applied to the *a priori* belief on Θ and the belief induced by the joint observations. The belief functions presented are normalized.

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

Table 6.4: The basic belief masses (m) and the related (normalized) belief function (bel_n) induced on Θ by the joint observation of x_3 and y_2, and based on three different *a priori* beliefs on Θ: an *a priori* that reject θ_ω, a probabilistic *a priori* on $\{\theta_1, \theta_2\}$ and a simple support function on $\{\theta_1, \theta_2\}$.

Acknowledgements.

The authors would like to thank E. Zimanyi for helpful feedback.

References.

BOSC P. and PRADE H. (1995) An introduction to fuzzy set and possibility theory-based approaches to the treatment of uncertainty and imprecision in data base management systems. This volume.

DEMPSTER A.P. Upper and lower probabilities induced by a multplevalued mapping. Ann. Math. Statistics 38 (1967) 325-339.

DUBOIS D., GARBOLINO P., KYBURG H.E., PRADE H. and SMETS Ph. (1991) Quantified Uncertainty. J. Applied Non-Classical Logics 1:105-197.

DUBOIS D. and PRADE H. (1986a) On the unicity of Dempster rule of combination. Int. J. Intelligent Systems, 1:133-142.

DUBOIS D. and PRADE H. (1986b) A set theoretical view of belief functions. Int. J. Gen. Systems, 12:193-226.

GEBHARDT J. and KRUSE R. (1993) The Context Model : An Integrating View of Vagueness and Uncertainty, Int. J. of Approximate Reasoning 9, 283-314.

GOOD I.J. (1962)Subjective probility as a measure of a non-measurable set. Logic, Methodology and Philisophy of Science, ed. by P. Suppes, E. Nagel and A. Tarski. Stanford Univ. Press, 319-329.

HAJEK P. (1992) Deriving Dempster's rule. in IPMU '92 proceedings, pg. 73-75.

JAFFRAY J.Y. (1988) Application of linear utility theory for belief functions. in BOUCHON B., SAITTA L. and YAGER R.R. eds. Uncertainty and Intelligent Systems, Springer-Verlag, Berlin. p. 1-8

JAFFRAY J.Y. (1992) Bayesian updating and belief functions. IEEE Trans. SMC, 22:1144-1152.

KENNES R.(1991) Evidential Reasoning in a Categorial Perspective : Conjunction and Disjunction of Belief Functions. in D'Ambrosio B., Smets P., and Bonissone P.P. eds, Uncertainty in AI 91,Morgan Kaufmann, San Mateo, Ca, USA, 1991, pg.174-181

KLAWONN, F.and SCHWECKE, E. (1992) On the axiomatic justification of Dempster's rule of combination. Int. J. of Intelligent Systems 7, 469-478.

KLAWONN F. and SMETS Ph. (1992) The dynammic of belief in the transferable belief model and specialization-generalization matrices. in Dubois D., Wellman M.P., d'Ambrosio B. and Smets P. Uncertainty in AI 92. Morgan Kaufmann, San Mateo, Ca, USA, 1992, pg.130-137.

KOHLAS J. (1993) Support and Plausibility Functions Induced by Filter-Valued Mappings. Int. J. General Systems, 21:343-363.

KOHLAS J. and MONNEY P.A. (1994a) Theory of Evidence: A Survey of its Mathematical Foundations, Applications and Computational Aspects. ZOR - Mathematical Methods of Operations Research, 39:35-68

KOHLAS J. and MONNEY P.A. (1994b) Representation of Evidence by Hints. In: R.R. Yager, M. Fedrizzi, J. Kacprzyk (eds.) Advances in the Dempster-Shafer theory of evidence. Wiley, pp. 473-492

KOHLAS J. and MONNEY P.A. (1995) A Mathematical Theory of Hints. An Approach to Dempster-Shafer Theory of Evidence. Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag.

KOOPMAN B.O. (1940) The bases of probability. Bull. Amer. Math. Soc. 46:763-774.

KRUSE R., NAUCK D. and KLAWONN F. (1991) Reasoning with Mass. In: B.D. D'Ambrosio, P. Smets, P.P. Bonissone (eds.): Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, San Mateo, California (1991), 182-187

KRUSE R. and SCHWECKE E. (1991) Specialization: a new concept for uncertainty handling with belief functions. Int. J. Gen. Systems 18:49-60.

KRUSE R., E. SCHWECKE E. and F. KLAWONN F. (1991) On a Tool for Reasoning with Mass Distributions. Proc. 12th International Joint Conference on Artificial Intelligence (IJCAI--91) Vol. 2. Morgan Kaufmann Publishers, San Mateo, California (1991), 1190-1195

KYBURG H.E.Jr. (1961) Probability and the logic of rational belief. Wesleyan Univ. Press.

LOHMANN G. (1991) An evidential reasoning approach to the classification of satellite images. in KRUSE R. and SIEGEL P (eds) Symbolic and Quantitative Approaches to Uncertainty. Springer-Verlag, Berlin, pg. 227-231

LING X.N. and RUDD W.G. (1989) Combining opinions from several experts. Applied Artificial Inteligence, 3:439-452.

NGUYEN T. H. and SMETS Ph. (1993) On Dynamics of Cautious Belief and Conditional Objects. Int. J. Approximate Reas. 8:89-104.

PAL N., BEZDEK J. and HEMASINHA R. (1992) Uncertainty measures for evidential reasoning I: a review. Int. J. Approx. Reas. 7:165-183.

PAL N., BEZDEK J. and HEMASINHA R. (1993) Uncertainty measures for evidential reasoning II: new measure of total uncertainty. Int. J. Approx. Reas. 8:1-16.

PEARL J. (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Pub. San Mateo, Ca, USA.

PEARL J. (1990) Reasoning with Belief Functions: an Analysis of Compatibility. Intern. J. Approx. Reasoning, 4:363-390.

PLANCHET B. (1989) Credibility and conditioning. J. Theor. Probabil. 2:289-299.

RUSPINI E.H. (1986) The logical foundations of evidential reasoning. Technical note 408, SRI International, Menlo Park, Ca.

SAVAGE L.J. (1954) Foundations of Statistics. Wiley, New York.

SHAFER G. (1976) A mathematical theory of evidence. Princeton Univ. Press. Princeton, NJ.

SHAFER G. (1990) Perspectives in the theory and practice of belief functions. Intern. J. Approx. Reasoning, 4:323-362.

SMETS P. (1978) Un modèle mathématico-statistique simulant le processus du diagnostic médical. Doctoral dissertation, Université Libre de Bruxelles, Bruxelles, (Available through University Microfilm International, 30-32 Mortimer Street, London W1N 7RA, thesis 80-70,003)

SMETS P. (1981) Medical diagnosis: fuzzy sets and degree of belief. Fuzzy Sets and Systems 5:259-266.

SMETS P. (1986) Combining non distinct pieces of evidence. Proc.NAFIP86, New Orleans, 544-548.

SMETS P. (1988) Belief functions. in SMETS Ph, MAMDANI A., DUBOIS D. and PRADE H. eds. Non standard logics for automated reasoning. Academic Press, London p 253-286.

SMETS P. (1990a) The combination of evidence in the transferable belief model. IEEE-Pattern analysis and Machine Intelligence, 12:447-458.

SMETS P. (1990b) Constructing the pignistic probability function in a context of uncertainty. Uncertainty in Artificial Intelligence 5, Henrion M., Shachter R.D., Kanal L.N. and Lemmer J.F. eds, North Holland, Amsterdam, , 29-40.

SMETS Ph. (1991) Probability of provability and belief functions. Logique et Analyse, 133-134:177-195.

SMETS P. (1992a) Resolving misunderstandings about belief functions: A response to the many criticisms raised by J. Pearl. Int. J. Approximate Reasoning.6: 321-344.

SMETS P. (1992b) The nature of the unnormalized beliefs encountered in the transferable belief model. in Dubois D., Wellman M.P., d'Ambrosio B. and Smets P. Uncertainty in AI 92. Morgan Kaufmann, San Mateo, Ca, USA, 1992, pg.292-297.

SMETS P. (1992c) The concept of distinct evidence. IPMU 92 Proceedings.pg. 789-794.

SMETS P. (1993a) Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approximate Reasoning 9:1-35.

SMETS P. (1993b) Probability of Deductibility and Belief Functions. ESQARU 93. Springer Verlag.

SMETS P. (1993c) An axiomatic justifiaction for the use of belief function to quantify beliefs. IJCAI'93 (Inter. Joint Conf. on AI), Chambery. pg. 598-603.

SMETS P. (1993d) Jeffrey's rule of conditioning generalized to belief functions. Uncertainty in AI 93. Heckerman D. and Mamdani A. eds. Morgan Kaufmann, San Mateo, Ca, USA, 500-505.

SMETS Ph. (1993e) No Dutch Book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. SIS, Workshop on Probabilisitic Expert Systems, (ed. R. Scozzafava), Roma, pg. 181-204.
SMETS P. (1995) Imperfect Information: Imprecision - Uncertainty. This volume.
SMETS Ph. and KENNES R. (1994) The transferable belief model. Artificial Intelligence 66:191-234.
SMITH C.A.B. (1961) Consistency in statistical inference and decision. J. Roy. Statist. Soc. B23:1-37.
SMITH P. and JONES O.R. (1986) The philosophy of mind, an introduction. Cambridge University Press, Cambridge.
STRAT T.M. (1989) Making decisions with belief functions. Proc. 5th Workshop on Uncertainty in IA. Windsor, 351-360.
STRAT T.M. (1990) Decision analysis using belief functions. Int. J. Approx. Reasoning 4: 391-418
WALLEY P. (1991) Statistical reasoning with imprecise probabilities. Chapman and Hall, London.

Focusing: adapting beliefs when the reference class is hypothetically changed. Must be coherent with other hypothetical changes.
Revision: adapting beliefs when a new pieces of evidence is added, in which case the reference class is definitively and uniquely changed.

Rudolf: help!!!

How to treat missing values:
D = value domain of a given attribute.
e = attribute not applicable
D+ = $D \cup \{e\}$ see Bosc and Prade.

$m(D) = 1$	attribute applicable but total ignorance about its value.
$m(\{e\}) = 1$	attribute non applicable
$m(D\cup\{e\})=1$	I donot even know if attribute is applicable
$m(\{\}) = 1$	(strongest) contradiction (2 sources that supports incompatible subsets)
	actual value is not in D+: applicable but out of range, bad definition of D.

I cannot make the difference between the last two, except by creating D such that $\neg e \supset bel(D)=1$
or better create domain D ++ = $D\cup\{\delta\}\cup\{e\}$ where δ = any other value, those not in D, and $\neg e \supset bel(D\cup\{\delta\})=1$

$m(D) = 1$	attribute applicable, total ignorance about its value,
	but value among those considered in D.
$m(\{\delta\}) = 1$	attribute applicable, value not in those considered in D
$m(\{e\}) = 1$	attribute non applicable
$m(D\cup\{\delta\})=1$	attribute applicable
$m(D\cup\{e\})=1$	either attribute non applicable or value among those considered in D.
$m(\{e\}\cup\{\delta\})=1$	either attribute non applicable or value not among those considered in D.
m(D++)	I am in a state of total ignorance
	(like when I donot even understand what the attribute is about)
$m(\{\}) = 1$	(strongest) contradiction (2 sources that supports incompatible subsets)

comments by Rudolf

m(D)=1 : attribute applicable but total ignorance about its value -->OK
m({e})=1 : attribute not applicable -->OK
m(D u {e})=1 : I do not even know if attribute is applicable -->OK
m({})=1 : (strongest) contradiction (2 sources that support
incompatible subsets --> NOT OK

The first three interpretations are right, but the last one is conceptual

doubtful.
The transferable belief model is not capable of handling this case.
The reason for it refers to the fact that you have no clear formal distinction between information sources and data in mass distributions. If you consider "two sources that support incompatible subsets", then, in the basic modelling, this means that you have two different contexts (sources) c1, c2, delivering context-dependent data sets A1, A2, where the intersection of A1 and A2 is empty. A1 and A2 are themselves, of course, not empty. If there are no more contexts (sources) to be considered, and the weigthing of the contexts is w1 > 0, w2 = 1 - w1 > 0, respectively, then we obtain the induced mass distribution m, defined by m(A1) = w1, m(A2) + w2, but m({}) = 0.
To summarize, m({}) = 1 reflects the fact that a priori (i.e. without combination, conditioning, data revision and so on) the involved contexts themselves lead to context-dependent data represented by the empty set (A1 and A2 both empty). This seems not to be reasonable at all, when the closed world assumption is accepted.
Taking the open world assumption, the best way is to add a single element e (not contained in D) as a representation for elements except those of D. Doing it (dangerous!), the resulting set D+ = D u {e} can be viewed as a new domain, where - on a formal level -you can again turn back to the closed world assumption, that is now related to D+.
Since on a formal as well as on the semantical level it is very dangerous, we have problems with accepting open world assumptions.
One more hint: Do not apply your decision making process on the pignistic level, when the extended domain D+ is considered, because the generalized insufficient reason principle does not hold for this case, and you will come to wrong decisions.

Proposal:
Since your paper refers to the transferable belief model and the problem of contradicting sources is not covered by this model, avoid to involve such considerations. A more general approach that solves such problems is, for example, the application of the context model.
Hence, ignore the case m({})=1 and restrict your presentation to data given by a single source (like, f.e., in section 5, where the ages of John, Henri, Jack, Jim, and Phil are chareacterized by single sources).

6. Application for Diagnosis.

note by Rudolf
It might be reasonable to introduce an element theta_omega in order to sustain the closed world assumption. Problems might arise, when theta_omega actually represents a set of elements and we come to decision making as it is described in Section 3.9. BetP(A) changes if we refine theta_omega by replacing it by a finite number of diseases. Of course, the refinement does not alter the order of the values BetP(A) for A intersection `theta_omega' = empty. Some explaining remarks in this direction could be helpful.