THE TRANSFERABLE BELIEF MODEL.

Philippe SMETS and Robert KENNES[1]
IRIDIA
Université Libre de Bruxelles
50 av. F. Roosevelt, CP194/6
1050 Brussels, Belgium.
email: psmets@ulb.ac.be

Abstract: We describe the Transferable Belief Model, a model for representing quantified beliefs based on *belief functions*. Beliefs can be held at two levels: 1) a credal level where beliefs are entertained and quantified by belief functions, 2) a pignistic level where beliefs can be used to make decisions and are quantified by probability functions. The relation between the belief function and the probability function when decisions must be made is derived and justified. Four paradigms are analysed in order to compare Bayesian, upper and lower probability and the transferable belief approaches.

Keywords: Belief Function, Dempster-Shafer Theory, Quantified Beliefs.

1. Introduction.

1) The aim of this paper is to present the transferable belief model (TBM) i.e. our interpretation of the Dempster-Shafer model. The TBM is a model for representing the quantified beliefs held by an agent at a given time on a given frame of discernment. It concerns the same concepts as considered by the Bayesian model, except it does not rely on probabilistic quantification, but on a more general system based on belief functions.

Since Shafer introduced his model based on belief functions in his book (Shafer 1976), many interpretations of it have been proposed. Three main interpretations have been developed: the random set, the generalized Bayesian and the upper and lower probability interpretations. However, great confusion and even blatant errors pervade the literature about the meaning and applicability of these models (Pearl (1990), Smets (1992c)). We

[1]1 This work has been partially funded by the CEC-ESPRIT II Basic Research Project 3085 (DRUMS), by the CEC-ESPRIT II Project 2256 (ARCHON) and the Belgian National Incentive-Program for Fundamental Research in Artificial Intelligence.

personally develop a model for point-wise quantified beliefs - the Transferable Belief Model - and show how belief functions can be used for such a quantification. Bayesian probability is the most classical model for quantified beliefs. So our presentation focuses on comparing the TBM with its real contender: the Bayesian model. In particular we will discuss the problem of decision making within the TBM because it is necessary to explain how the model is used in real situations where decisions must be made, and because it is central to any Bayesian presentation. We even argue that Dutch Books - a betting strategy that would lead to a sure loss - cannot be raised against TBM users. In fact when decisions must be made, we require that beliefs be quantified by probability functions in order to avoid Dutch Books .

Several paradigms are analysed in order to provide some insight into the nature of the TBM. These paradigms are used to contrast the TBM solution with the Bayesian, upper and lower probabilities, likelihood and fiducial solutions. The TBM is compared with random sets in Smets (1992b), with possibility functions in Smets (1990b), and with upper and lower probabilities in Smets (1987) and Halpern and Fagin (1992). The major differences between these models can be found in the way updating/conditioning must be performed. Axiomatic justifications are not given here but are developed in Smets (1991b) and Wong et al. (1990).

We also argue in this paper that the TBM should not be considered as just a generalized probability model: indeed there are no *necessary* links between the TBM and any underlying probability model. Hence we dissociate ourselves from Dempster's model where some underlying probability is essential. Any decisions as to the nature of Shafer's model are left to Shafer himself (see Shafer 1990), but in our opinion, the TBM is very close to what Shafer described in his book (Shafer 1976). In later work, Shafer creates confusion by speaking about random sets and upper and lower probabilities interpretations. Recently Shafer (1990) clarified his position, rejected these interpretations and defended essentially the Dempster interpretation based on the random codes (a one-to-many mapping with an underlying probability distribution). We depart from this interpretation in that we do not *require* any underlying probability distribution, even though they *may* exist.

Not all interpretations of Dempster-Shafer theory are analysed here (see Smets 1990d). We do not discuss the interpretation of a belief as being the probability of a modal proposition (Ruspini 1986) or the probability of provability (Pearl 1988).

2) The **transferable belief model** is based on:

- a two-level model: there is a *credal level* where beliefs are entertained and a *pignistic level* where beliefs are used to make decisions (from *pignus* = a bet in Latin, Smith 1961).
- at the credal level beliefs are quantified by belief functions.
- the credal level precedes the pignistic level in that at any time, beliefs are entertained (and updated) at the credal level. The pignistic level appears only when a decision needs to be made.
- when a decision must be made, beliefs at the credal level induce a probability measure at the pignistic level, i.e. there is a *pignistic transformation* from belief functions to probability functions.

Bayesians do not consider an autonomous credal level. The introduction of a two-level model would be useless if decisions were the same as those derived within the Bayesian model. We will show in the “Mr. Jones” paradigm (section 4) that thit is not the case. The introduction of a credal level therefore is not merely an academic subtlety.

The TBM essentially fits with the model developed in Shafer's book (1976) except for some differences and explanations such as:
1) the complete dissociation from any *necessary* underlying probability model that precedes the construction of the belief functions at the credal level, as encountered in Dempster’s approach (we do not mean the pignistic probabilities used at the pignistic level and that are derived from the belief functions).
2) the fundamental concept of transferable 'parts of belief'.
3) the two-level model and the pignistic transformation.
4) the “open-world” and “closed-world” assumptions and the introduction of the unnormalized belief functions (Smets 1988).
5) the precedence of the conditioning process over the combination process.
6) the justification of Dempster's rule of combination as the unique compositional rule to combine two belief functions (Smets 1990a, Klawonn and Schwecke 1992, Klawonn and Smets 1992, Nguyen and Smets 1991).

3) **the TBM is unrelated to a probability model.** The TBM is intended to model subjective, personal beliefs, i.e. what the Bayesians claim to be their domain of application. The major point of the TBM is its complete dissociation from any model based on probability functions. This contrasts with what has been done in some of Shafer's more recent publications that favor the random set interpretation (Nguyen 1978, Shafer 1987), and most publications on Dempster-Shafer's model (Kyburg, 1987b, Black 1987). The TBM is neither a random sets model (Smets 1992b) nor a generalization of the Bayesian model nor of some upper and lower probability (ULP) models (Halpern and Fagin 1992). It is another model whose aim is to quantify someone's degree of belief. The

model is normative, supposedly simulating the behavior of a reasonable and consistent agent, the "stat rat" of Barnard (see discussion in Smith 1961, pg 26).

To support our case that the TBM is different from the Bayesian model, we present an example, the "Mr. Jones" case, that leads to different results according to which model is used to analyse it. Such an example might provide a tool for discriminating between the two models: according to which result fits your requirements, you can select the model.

Other examples have already been provided to show the difference between the Bayesian model and the TBM. But their power of persuasion as a discriminating tool is weak as the TBM answer can usually also be derived from a Bayesian analysis. The interest of the “Mr. Jones” example is that the TBM solution can only be obtained by a Bayesian analysis by introducing some unpalatable assumptions.

4) **Summary of content.** The TBM is presented in §2. We then present a theory for decision making (§3). That decisions are based on probability functions (and expected utilities) is not disputed. Whenever a decision has to be made by an agent, he/she constructs a probability distribution derived from the belief function that describes his/her credal state. Bear in mind that the existence of such a probability distribution when decisions are made does not imply that this probability function quantifies our belief at the credal level (i.e. outside of any decision context).

We show 1) the impact of a *betting frame* on bet, 2) how someone's betting behavior could be used to assess a belief function, 3) how conditioning acts on the betting behavior and 4) how Dutch Books are avoided.

We then proceed by analysing several paradigms in detail. We know from experience that these paradigms are very useful in appreciating the particularity of the TBM, especially when compared with other approaches. Each paradigm enables the difference to be shown between the TBM and some of its contenders.

In §4 we present "Mr. Jones" example, a very pointed example that shows the difference between the TBM approach and the Baeysian approach.

In §5, we present the “guards and posts” paradigm. It clarifies the nature of the conditioning process used in the TBM.

In §6 and §7, we present two others paradigms to illustrate situations where the TBM leads to results different from those of its contenders: the Bayesian model, the ULP model, the likelihood model, the fiducial model. Some of these comparisons have also been attempted by Hunter (1987), Laskey (1987). Other comparisons are presented in Smets (1990d).

In §8, we discuss the origin of the basic belief assignment used in our paradigms.

In §9, we show the difference between the credal level where someone's beliefs are quantified by belief functions and the pignistic level where "pignistic" probabilities must be constructed. Revision of beliefs is performed at the credal level by Dempster's rule of conditioning, not at the pignistic level by probability conditioning.

In §10, we conclude by answering some potential criticisms of the TBM.

2. The Transferable Belief Model.

2.1. The model.

1) The necessary background information on belief functions is summarized hereafter. A full description can be found in Shafer's book (1976). A somehow revised version appears in Smets (1988). Further results on Bayes theorem and the disjunctive rule of combination appear in Smets (1978, 1991c).

Let L be a finite **propositional language**, and $\Omega = \{\omega_1, \omega_2, ...\omega_n\}$ be the set of **worlds** that correspond to the interpretations of L. **Propositions** identify subsets of Ω. Let $\top$ be the tautology and $\bot$ be the contradiction. For any proposition X, let $[\![X]\!] \subseteq \Omega$ be the set of worlds identified by X. Let A be a subset of Ω, then f_A is any proposition that identifies A. So $A=[\![f_A]\!]$, $\emptyset=[\![\bot]\!]$ and $\Omega=[\![\top]\!]$. By definition there is an **actual world** ϖ and it is an element of Ω. In L, two propositions A and B are **logically equivalent**, denoted $A \equiv B$, iff $[\![A]\!]=[\![B]\!]$.

Let Π be a partition of Ω. Given the elements of the partition Π, we build $\Re$, the **Boolean algebra** of the subsets of Ω generated by Π. We call Ω the **frame of discernment** (the frame for short). The elements of the partition Π are called the **atoms** of $\Re$. Given $\Re$, the number of atoms in a set $A \in \Re$ is the number of atoms of $\Re$ that are included in A. We call the pair $(\Omega, \Re)$ a **propositional space**.

By abuse of language but for the sake of simplicity, we do not distinguish between the subsets of Ω and the propositions that denote them. We use the same notation for both of them. So the same symbol (like A, B, C..) is used for a subset of Ω and for any proposition that denote that subset. The standard Boolean notation is used. Let $A, B \in \Re$.

$\overline{A}$ stands for the complement of A relative to Ω. A∪B, A∩B denote the set-theoretic union and intersection of the (subsets denoted by the) propositions A and B. A⊆B means that all the elements of A (the subset denoted by A) are elements of B (the subset denoted by B) (or equivalently, the proposition A implies the proposition B). Any algebra ℜ defined on Ω contains two special propositions: ⊤ and ⊥ denoted by their corresponding sets Ω and Ø.

All beliefs entertained by You[1] 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 Your evidential corpus is deductively closed. The **credal state** on a propositional space (Ω, ℜ) describes Your subjective, personal judgment that ϖ∈A for every proposition A∈ℜ. 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 ϖ∈A. Your credal state results from EC_t^Y that induces in You some partial beliefs on the propositions of ℜ (note that we did not say Ω). These partial beliefs quantify the strength of Your belief that ϖ∈A, ∀A∈ℜ. It is an epistemic construct as it is relative to Your knowledge included in Your evidential corpus EC_t^Y.

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 propositions of ℜ. For A∈ℜ, 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* sub*proposition of A.*

The m(A) values, A∈ℜ, are called the **basic belief masses** (bbm) and the m function is called the **basic belief assignment** (bba)[2].

Let m: ℜ→[0,1] with

$$\sum_{A\in\Re} m(A) = 1 \qquad\qquad m(\emptyset) = 0$$

Every A∈ℜ such that m(A)>0 is called a focal proposition. The difference with probability models is that masses can be given to any proposition of ℜ instead of only to the atoms of ℜ.

---

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

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

As an example, let us consider a somehow reliable witness in a murder case who testifies to You that the killer 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 killer 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) plus the probability of M given the witness is not reliable (.5) weighted by the probability that the witness is not reliable (.3)). The TBM analysis will give a belief .7 to M. The .7 can be viewed as the *justified* component of the probability given to M (called the belief or the support) whereas the .15 can be viewed as the *aleatory* component of that probability. The TBM deals only with the justified components. (Note: the Evidentiary Value Model (Ekelof 1982, Gardenfors et al. 1983) describes the same belief component, but within a strict probability framework, and differs thus from the TBM once conditioning is introduced.)

2) If some further evidence becomes available to You and implies that B is true, then the mass m(A) initially allocated to A is transferred to $A \cap B$. Hence the name TBM.

Continuing with the murder case, suppose there are only two potential male suspects: Phil and Tom. Then You learn that Phil is not the killer. The testimony now supports that the killer is Tom. The reliability .7 You gave to the testimony initially supported 'the killer is Phil or Tom'. The new information about Phil implies that .7 now supports 'the killer is Tom'.

The transfer of belief described in the TBM satisfies the so-called **Dempster rule of conditioning**. Let m be a basic belief assignment on the propositional space $(\Omega, \Re)$ and suppose the conditioning evidence tells You that the truth is in $B \in \Re$, the basic belief assignment m are transformed into $m_B: \Re \rightarrow [0,1]$ with:

$$m_B(A) = c \sum_{X \subseteq \overline{B}} m(A \cup X) \qquad \text{for } A \subseteq B \qquad (2.1)$$

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

and $m_B(\emptyset) = 0$

with $c = \dfrac{1}{1 - \sum_{X \subseteq \overline{B}} m(X)}$

In this presentation we have asumed that one and only one element of Ω is true (closed-world assumption). In Smets (1988) we generalized the model and accepted that none of the elements could be true (open-world assumption). In that last case, positive basic belief masses could be given to Ø and the normalization coefficient c in (2.1) is 1. The closed-world assumption is accepted hereafter. The meaning of the basic belief mass given to Ø is analysed in Smets (1992a)

3) Given (Ω, $\Re$), the **degree of belief** of A, 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 proposition $X \in \Re$ with $X \subseteq A$ (and $X \neq \emptyset$)

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

We say *justified* because we include in bel(A) *only* the basic belief masses given to subsets of A. For instance, consider two distinct atoms x and y of $\Re$. The basic belief mass m({x,y}) given to {x,y} could support x if further information indicates this. However given the available information the basic belief mass can only be given to {x,y}. (Note that under open-world assumption, m(Ø) might be positive. The basic belief mass m(Ø) should not be included in bel(A) nor in pl(A), as it 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:$\Re \rightarrow [0,1]$ is called a belief function. The triple (Ω, $\Re$,bel) is called a credibility space. Belief functions satisfy the following inequalities (Shafer 1976):
$\forall n \geq 1$, $A_1, A_2, \ldots A_n \in \Re$,

$$\text{bel}(A_1 \cup A_2 \cup \ldots A_n) \geq \sum_i \text{bel}(A_i) - \sum_{i>j} \text{bel}(A_i \cap A_j) \ldots. -(-1)^n \text{bel}(A_1 \cap A_2 \cap \ldots A_n) \qquad (2.2)$$

The **degree of plausibility** of A, pl(A), quantifies the maximum amount of *potential specific support* that could be given to $A \in \Re$. It is obtained by adding all those basic belief masses given to propositions X compatible with A, i.e. such that $X \cap A \neq \emptyset$:

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

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 corespondence with belief functions. 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.

Dempster's rule of conditioning expressed with bel and pl is:

$$\mathrm{bel}(A|B) = \frac{\mathrm{bel}(A\cup\overline{B}) - \mathrm{bel}(\overline{B})}{1-\mathrm{bel}(\overline{B})} \qquad \mathrm{pl}(A|B) = \frac{\mathrm{pl}(A\cap B)}{\mathrm{pl}(B)}$$

4) Beside the logical equivalence already mentioned, there is another concept of equivalence related to Your evidential corpus EC_t^Y. Let $[\![EC_t^Y]\!]$ represents the set of worlds in Ω where all propositions deduced on Ω from EC_t^Y are true. All the worlds in Ω that are not in $[\![EC_t^Y]\!]$ are accepted as 'impossible' by You at time t. Two propositions A and B are said to be **doxastically equivalent** for You at time t, denoted $A\cong B$, iff $[\![EC_t^Y]\!]\cap[\![A]\!] = [\![EC_t^Y]\!]\cap[\![B]\!]$. Logical equivalence implies doxastic equivalence. This is important as it implies that A and B should get the same support, the same degree of belief. Hence the Consistency Axiom that expresses the equi-credibility of doxastically equivalent propositions.

Consistency axiom: Let us consider two credibility spaces $(\Omega, \Re_i, \mathrm{bel}_i)$, i=1,2 that represent Your beliefs on two algebras $\Re_1$ and $\Re_2$ as induced by Your EC_t^Y. Let $A_1\in\Re_1$, $A_2\in\Re_2$. If $A_1\cong A_2$, then $\mathrm{bel}_1(A_1) = \mathrm{bel}_2(A_2)$.

This consistency axiom means that doxastically equivalent propositions share the same degree of belief, which is required since they share the same truth value. It also means that the belief given to a proposition does not depend on the structure of the algebra to which the proposition belongs. This consistency axiom is usually postulated for probability distributions, when they quantify degrees of belief (Kyburg, 1987a). Here it is postulated only for those functions that quantify beliefs at the credal level.

5) **Total ignorance** is represented by a vacuous belief function, i.e. a belief function such that $m(\Omega) = 1$, hence $\mathrm{bel}(A) = 0\ \forall A\in\Re,\ A\neq\Omega$, and $\mathrm{bel}(\Omega) = 1$.The origin of this particular quantification for representing a state of total ignorance can be justified. Suppose that there are three propositions labeled A, B and C, and You are in a state of total ignorance about which is true. You only know that one and only one of them is true but even their content is unknown to You. You only know their number and their label. Then You have no reason to believe any one more than any other, hence, Your beliefs about their truth are equal: $\mathrm{bel}(A) = \mathrm{bel}(B) = \mathrm{bel}(C) = \alpha$ for some $\alpha\in[0,1]$. Furthermore, You have no reason to put more or less belief in $A\cup B$ than in C: $\mathrm{bel}(A\cup B) = \mathrm{bel}(C) = \alpha$ (and similarly $\mathrm{bel}(A\cup C) = \mathrm{bel}(B\cup C) = \alpha$). The vacuous belief function is the only belief function that satisfies equalities like: $\mathrm{bel}(A\cup B) = \mathrm{bel}(A) = \mathrm{bel}(B) = \alpha$. Indeed the inequalities (2.2) are such that $\mathrm{bel}(A\cup B) \geq \mathrm{bel}(A) + \mathrm{bel}(B) - \mathrm{bel}(A\cap B)$. As $A\cap B=\emptyset$, $\mathrm{bel}(A\cap B) = 0$. The inequality becomes $\alpha \geq 2\alpha$ where $\alpha\in[0,1]$, hence $\alpha = 0$.

6) The TBM also includes a description of **Dempster's rule of combination** - a rule for the conjunctive combination of two belief functions, that somehow generalizes Dempster's rule of conditioning. This rule is not used in this presentation. An axiomatic justification of Dempster's rule of combination within the TBM is given in Smets (1990a), Klawonn and Schwecke (1992), Hajek (1992), Klawonn and Smets (1992). There also exists a disjunctive rule of combination of two belief functions described in Smets (1991c).

7) It is important to note that the TBM includes **two components**: one **static**, the basic belief assignment, and one **dynamic**, the transfer process. 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, i.e. how beliefs are updated. Their comparisons are therefore incomplete, if not misleading.

2.2. Refinements and consistent beliefs.

1) Let us consider two propositional languages L_1 and L_2. It is always possible to build a common underlying propositional language L such that each proposition of L_1 (and of L_2) is a proposition of L. Let Ω_1, Ω_2 and Ω be the sets of worlds that correspond to the interpretations of L_1, L_2 and L, respectively. Each world of Ω_1 (and of Ω_2) corresponds to a set of worlds of Ω, and the images of the worlds of Ω_1 (and of Ω_2) constitute a partition of Ω. Hence whenever we describe two propositional spaces, we can always use a common underlying Ω without loss of generality. In fact the concept in a propositional space $(\Omega, \Re)$ that is important for this presentation is the algebra $\Re$, not the set of worlds Ω. All beliefs are build on the algebras $\Re$, not on Ω. The granularity of Ω is irrelevant once Ω is fine enough to allow for a definition of the atoms of $\Re$ (i.e., each atom of $\Re$ contains at least one element of Ω). Therefore, the definition of two propositional spaces $(\Omega_i, \Re_i)$, i=1,2, with different sets Ω_i is equivalent to a definition of two propositional spaces $(\Omega, \Re_i)$, i=1,2 sharing the same Ω. From now on, we will not worry about the Ω, they will be adapted such that each $\Re$ is non-ambiguously defined.

2) Consider two propositional spaces $(\Omega, \Re_1)$ and $(\Omega, \Re)$ (see figure 1, left half, where the elements of Ω are four individuals characterized by name and age, the atoms of $\Re_1$ are male and female and $\Re$ is the power set of Ω.) Let Λ_1 be a one-to-many mapping from $\Re_1$ to $\Re$ such that each atom of $\Re_1$ is mapped on a proposition of $\Re$, the images of the atoms of $\Re_1$ constitute a partition of Ω, and this mapping is additive. Λ_1 is called a **refining** from $\Re_1$ to $\Re$. $\Re$ is called a **refinement** of $\Re_1$. $\Re_1$ is called a **coarsening** of $\Re$ (see Shafer, 1976, p115). For $B \in \Re$, let $\overline{\Lambda}_1^{-1}(B) = \cup\{A: A \in \Re_1, \Lambda_1(A) \cap B \neq \emptyset\}$.

3) Let us consider two propositional spaces $(\Omega, \Re_i)$, i=1,2, and two refinings Λ_i to a common refinement $\Re$. By construction, if $B\in\Re$ is true (the actual world $\varpi\in B$), then $\overline{\Lambda}_1^{-1}(B)$ and $\overline{\Lambda}_2^{-1}(B)$ are true. The two credibility spaces $(\Omega, \Re_i, bel_i)$, i=1,2 are said to be **consistent** if there exists a belief function bel on $\Re$ such that $bel_i(\overline{\Lambda}_i^{-1}(B)) = bel(B)$ for all $B\in\Re$, i=1,2.

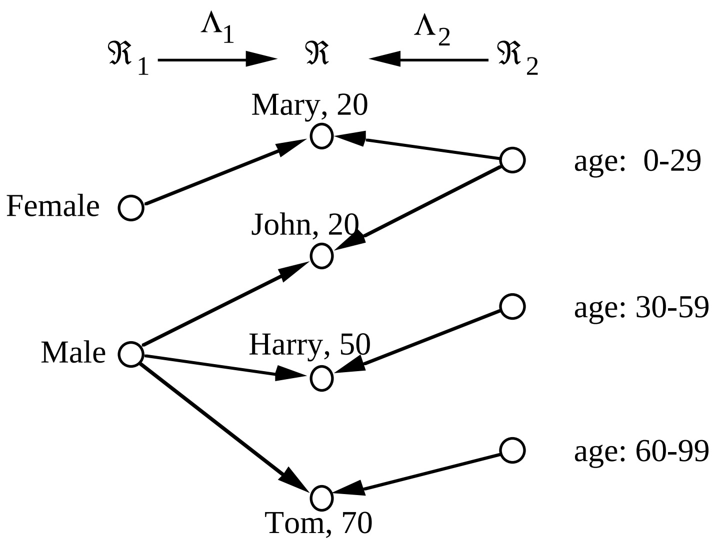

Figure 1: Example of two propositional spaces $(\Omega, \Re_i)$, i = 1,2. Λ_i are the refinings from $\Re_i$ to $\Re$. Each circle is an atom. The atoms of $\Re$ are those of a refinement common to $\Re_1$ and $\Re_2$.

2.3. Least committed belief on $\Re_2$ induced by a belief on $\Re_1$.

Let us suppose a credibility spaces $(\Omega, \Re_1, bel_1)$ induced by Your EC_t^Y. Let Λ be a relation between $\Re_1$ and a new algebra $\Re_2$ defined on Ω such that $\Lambda(\omega)\neq\emptyset$ for every atom ω of $\Re_1$ and $\Lambda(A\cup B) = \Lambda(A)\cup\Lambda(B)$, $\forall A, B\in\Re_1$. The question is to construct a belief function bel_2 on $\Re_2$ consistent with bel_1 and that conveys on $\Re_2$ the same "information" as bel_1 does on $\Re_1$. Let $\mathcal{B}$ be the familly of belief functions bel on $\Re_2$ consistent with bel_1. By the consistency axiom, for every pair (A, B) with $A\in\Re_1$ and $B\in\Re_2$ such that $A\cong B$, one has $bel_1(A) = bel(B)$. But this requirement does not provide the value of bel for those $B\in\Re_2$ that are not doxastically equivalent to some $A\in\Re_1$. The *Principle of Least Commitment* (Hsia 1991, Smets 1991c) allows us to select the belief function $bel^*\in\mathcal{B}$ such that $bel^*(B)\leq bel(B)$, $\forall B\in\Re_2$, $\forall bel\in\mathcal{B}$. This Principle says that

You should not give more support to a proposition than justified by bel_1. It implies that bel_2 (and its bba m_2) is related to bel_1 (and its bba m_1) by:

$$\forall B \in \Re_2 \qquad m_2(B) = \sum_{A \in \Re_1 : \Lambda(A)=B} m_1(A)$$

where the sum is 0 when no A satisfies the constraint. bel_2 is called the *vacuous extension* of bel_1 on $\Re_2$ (Shafer 1976).

3. The pignistic probability derived from a belief function.

3.1. The Generalized Insufficient Reason Principle.

Let us give a context. Given the evidence available to You, the TBM claims the existence of a belief function that describes Your credal state on the frame of discernment.

Suppose a **decision** must be made based on this credal state. As is well known, decisions will be coherent if the underlying uncertainties can be described by a probability distribution defined on 2^Ω (DeGroot 1970). Therefore, one must find a rule that allows for the construction of a probability distribution from a belief function in case of forced decision. We only consider forced bets, forced decisions, as done classically by Bayesians. (The unforced decisions considered in Smith (1961), Giles (1982), Jaffray (1988) concern ULP contexts). The solution that will satisfy some behavior requirements introduced in section 3.2 happens to be a generalization of the insufficient reason principle. Another justification can be found in Smets (1990c). This solution already appeared in Dubois and Prade (1982) and Williams (1982) but with no justification.

Let us consider a credibility space $(\Omega, \Re, bel)$ that describes Your beliefs on $\Re$. Let $A \in \Re$ and $A = A_1 \cup A_2 ... \cup A_n$ where the A_i's are distinct atoms of $\Re$. The mass m(A) corresponds to that part of Your belief that is restricted to A and, due to lack of further information, that cannot be allocated to a proper subset of A. In order to build the probability distribution needed for decision making (hence qualified as pignistic) on $\Re$, You distribute m(A) equally among the atoms of A. Therefore, m(A)/n is given to each A_i: i = 1,..n. This procedure corresponds to the **insufficient reason principle**: if You must build a probability distribution on n elements, given a lack of information, give a probability 1/n to each element. This procedure is repeated for each mass m. Let BetP be the pignistic probability distribution so derived. For all atom $x \in \Re$:

$$BetP(x) = \sum_{x \subseteq A \in \Re} \frac{m(A)}{|A|} = \sum_{A \in \Re} m(A) \frac{|x \cap A|}{|A|}$$

where |A| is the number of atoms of $\Re$ in A, and for $B\in\Re$,

$$\mathrm{BetP}(B) = \sum_{A\in\Re} m(A)\frac{|B\cap A|}{|A|}$$

Of course, BetP is a probability function, but we call it a pignistic probability function to stress the fact that it is the probability function in a decision context. The principle underlying this procedure is called the Generalized Insufficient Reason Principle as the Insufficient Reason Principle has been used at the level of each focal proposition of the belief function. As described up to here it is only an *ad hoc* principle, but it can be justified by natural behavior requirements.

3.2. Derivation of the Generalized Insufficient Reason Principle.

1) Let us give a credibility space (Ω, $\Re$, bel). Let m be the basic belief assignment corresponding to bel. Let BetP(.,m) be the pignistic probability to be defined on $\Re$. The ‘m’ parameter is added in order to enhance the basic belief assignment from which BetP is derived.

Assumption A1: $\forall$x atom of $\Re$, BetP(x,m) depends only on m(X) for $x\subseteq X\in\Re$.

Assumption A2: BetP(x,m) is continuous (or bounded) for each m(X): $x\subseteq X\in\Re$.

Assumption A3: Let G be a permutation defined on Ω. For $X\subseteq\Omega$, let G(X) = $\{G(x):x\in X\}$. Let m' = G(m) be the basic belief assignment given to the propositions of Ω after the permutation has been performed, that is for $X\in\Re$, m'(G(X)) = m(X). Then for any atom x of $\Re$, BetP(x,m) = BetP(G(x),G(m)).

In other terms, BetP is invariant by permutations of Ω.

Assumption A4: Let (Ω, $\Re$, bel) be the credibility space that describes Your beliefs on $\Re$, such that it is known by You that ϖ is not an element of the atom $X\in\Re$ (so $\forall A\in\Re$, $A\cong A\cup X$ and bel(A) = bel($A\cup X$) by the consistency axiom). Let us consider the credibility space (Ω', $\Re'$, bel') where $\Omega' = \Omega$-X, $\Re'$ is the boolean algebra build from the atoms of $\Re$ not subset of X (so every element A of $\Re'$ is also an element of $\Re$, and $\forall A\in\Re'$, bel'(A) = bel(A) by the consistency axiom). Let BetP(x,m) and BetP'(x,m') be the pignistic probabilities derived from bel (m) and bel' (m'), respectively. Then for every atom $x\in\Re'$,

$$BetP(x,m) = BetP'(x,m')$$

and $$BetP(X,m) = 0.$$

The major assumption A1 says that BetP(x,m) may depend on bel($\bar{x}$) but not on the way the basic belief masses used to compute bel($\bar{x}$) are distributed among themselves.

The three other assumptions are classical requirements. Assumption A2 could be weakened as one only needs that, for each m(X): $x \in X$, BetP(x,m) is continuous in a point, or bounded, or measurable, or majorizable by a measurable function on a set of positive measure (see Aczel 1966, pg 142).

Assumption A3 is the classical **anonymity requirement**: renaming the elements of Ω does not modify the probabilities. That m'(G(X)) = m(X) results from the consistency axiom as $G(X) \cong X$.

Assumption A4 only states that impossible atoms do not change the pignistic probabilities.

Theorem 3.1: Let (Ω, $\Re$) be a propositional space and m be a basic belief assignment on $\Re$. Let |A| be the number of atoms of $\Re$ in A. Under assumptions A1 to A4, for any atom x of $\Re$

$$BetP(x,m) = \sum_{x \subseteq A \in \Re} \frac{m(A)}{|A|} \qquad (3.1)$$

Proof: given in appendix 1.

The transformation defined by equation 3.1 is called the pignistic transformation.

Corollary: If bel is a probability distribution P, then BetP is equal to P.

2) The same pignistic transformation was derived in Smets (1990c) by assuming different requirements whose overall idea follows the next scenario. Let us consider two friends of Yours, G and J. You know they will toss a fair coin and the winner will visit You tonight. You want to buy the drink Your friend would like to receive tonight: coke, wine or beer. You can only buy one drink. Let D = {coke, wine, beer}.

Let us suppose that $bel_G(d)$, $\forall d \subseteq D$, quantifies Your belief about the drink G will ask for, should G come. Given bel_G, You build Your pignistic probability $BetP_G$ about the drink G will ask by applying the (still to be deduced) pignistic transformation. You identically build the pignistic probability $BetP_J$ based on bel_J, Your belief about the drink J will ask, should J come. The two pignistic probability distributions $BetP_G$ and $BetP_J$ are in fact the conditional probability distributions about the drink that will be drunk given G or J comes. The pignistic probability distributions $BetP_{GJ}$ about the drink that Your visitor will ask is then:

$$BetP_{GJ}(d) = .5\ BetP_G(d) + .5\ BetP_J(d) \qquad \forall d \subseteq D$$

You will use these pignistic probabilities $BetP_{GJ}(d)$ to decide which drink to buy.

But You could as well reconsider the whole problem and compute first Your belief about the drink Your visitor (V) would like to receive. In Smets (1990c), we show that such a belief is the average of bel_G and bel_J:

$$bel_V(d) = .5\ bel_G(d) + .5\ bel_J(d) \qquad \forall d \subseteq D$$

Given bel_V, You could then build the pignistic probability $BetP_V$ You should use to decide which drink to buy. We proved that if $BetP_V$ and $BetP_{GJ}$ must be equal, then the pignistic transformation must be the one given by the Generalized Insufficient Reason Principle (relation 3.1).

3.3. Betting frames.

1) Let us consider a credibility space $(\Omega, \Re_0, bel_0)$. Before betting, one must define a betting frame $\Re$ on Ω, i.e. the set of atoms on which stakes will be allocated. The **granularity** of this frame $\Re$ is defined so that a stake could be given to each atom independently of the stakes given to the other atoms. For instance, if the stakes given to atoms A and B of $\Re_0$ must necessarily be always equal, both A and B belong to the same granule of $\Re$. The **betting frame** $\Re$ is organized so that the granules are the atoms of $\Re$, and $\Re$ is the result obtained by applying a sequence of coarsenings and/or refinments on $\Re_0$. Let us suppose the initial belief bel_0 is defined on $\Re_0$.Then the belief function bel induced by bel_0 on $\Re$ is (see section 2.3):

$$\forall A \in \Re,\ bel(A) = bel_0(\Lambda^{-1}(A))$$

where Λ is the transformation from $\Re_0$ to $\Re$, and $\forall A \in \Re, \Lambda^{-1}(A) = \cup\{X: X \in \Re_0, \Lambda(X) \subseteq A\}$. (See also Shafer, 1976, pg 146-7).

The pignistic probability BetP is then built from the belief function bel so derived on $\Re$.

2) **Betting under total ignorance.** To show the power of the TBM approach, let us consider one of those disturbing examples based on total ignorance.

Let us consider two propositions denoted A_1 and A_2. You know that one and only one proposition is true. But You don't know what the two propositions are. You just know their number and their labels. You must bet on A_1 versus A_2. In the TBM, Your belief about the truth of A_1 and A_2 is described by a vacuous belief function and the pignistic probabilities on the betting frame $\{A_1, A_2\}$ are

$$BetP(A_1) = BetP(A_2) = 1/2.$$

Let us now consider three propositions denoted B_1, B_2 and B_3. You know that one and only one proposition is true. But You don't know what the three propositions are. You just know their number and their labels. You must bet on B_1 versus B_2 versus B_3. In the TBM, Your belief about the truth of B_1, B_2 and B_3 is described by a vacuous belief function and the pignistic probabilities on the betting frame $\{B_1, B_2, B_3\}$ are

$$BetP'(B_1) = BetP'(B_2) = BetP'(B_3) = 1/3.$$

Now You learn that A_1 is logically (hence doxastically) equivalent to B_1 and A_2 is logically (doxastically) equivalent to $B_2 \cup B_3$. Within the TBM, this information will not modify Your beliefs and Your pignistic probabilities. If You were a Bayesian, You must adapt Your probabilities as they must give the same probabilities to A_1 and B_1. Which set of probabilities are You going to update, and why, especially since it must be remembered that You have no knowledge whatsoever about what the propositions are.

In a Bayesian approach, the problem raised by this type of example results from the requirement that doxastically equivalent propositions should receive identical beliefs, and therefore identical probabilities. Within the TBM, the only requirement is that doxastically equivalent propositions should receive equal beliefs (it is satisfied as $bel(A_1) = bel'(B_1) = 0$). Pignistic probabilities depend not only on these beliefs but also on the structure of the betting frame, hence $BetP(A_1) \neq BetP'(B_1)$ is acceptable as the two betting frames are different.

In a betting context, the set of alternatives and their degrees of refinement is relevant to the way Your bets are organized. Of course, if $BetP(A_1) = 1/2$ had been a well-justified probability, then $BetP'(B_1)$ would also have had to be 1/2. But here $BetP(A_1) = 1/2$ is based only on the knowledge of the number of alternatives on which You can bet and NOTHING ELSE. The difference between $BetP(A_1)$ and $BetP'(B_1)$ reflects the difference between the two betting contexts. Of course, as required, both A_1 and B_1 share the same degrees of belief.

3.4. Assessing degrees of belief.

Given a propositional space $(\Omega, \Re_0)$, the assessment of $bel_0(A)$ $\forall A \in \Re_0$ can be obtained from the betting behavior established on other algebras $\Re$ defined on Ω (or any refinment of Ω). Given such a betting frame $\Re$ and its corresponding pignistic probability BetP on

$\Re$, one can determine the set of belief functions S on 2^Ω that would lead to BetP through (3.1) when the betting frame is $\Re$. Construct various betting frames $\Re_i$ on Ω and assess the corresponding $BetP_i$ and S_i. Note that the same evidence underlies all bets and that the difference between the $BetP_i$ results only from the difference between the structure of the betting frames $\Re_i$. Let us consider a refining Λ from $\Re$ to $\Re'$. Then, given the consistency axiom, the relation between bel defined on $\Re$ and bel' defined on $\Re' = 2^{\Omega'}$ is such that:

$$m'(\Lambda(A)) = m(A) \qquad \forall A \in \Re$$

$$m'(B) = 0 \qquad \text{otherwise}$$

where $\Lambda(A) = \{\Lambda(x) : x \in A\}$.

bel' is the vacuous extension of bel from $\Re_0$ to $2^{\Omega'}$ (Shafer, 1976, pg 146). The strategy for defining various betting frames $\Re_i$ allows for the construction of a family of S_i whose intersection contains only one element bel_0. An empty intersection would imply inconsistency between the pignistic probabilities. The number of potential betting frames is large enough to guarantee that a unique solution be obtained.

Example: let us suppose that $\Omega_0 = \{a, b\}$ where {a} = 'John will come tonight' and {b} = 'John will not come tonight'. Let us consider the betting frame $\Re$ with atoms {a} and {b}, and Your pignistic probabilities on frame $\Re$:

$$BetP(\{a\}) = 4/9 \qquad BetP(\{b\}) = 5/9.$$

Suppose ψ and $\overline{\psi}$ are two complementary but otherwise unknown propositions. $\{a\}\cap\psi$ will occur if John comes tonight and proposition ψ is true. $\{a\}\cap\overline{\psi}$ will occur if John comes tonight and proposition $\overline{\psi}$ is true. Let us consider the betting frame $\Re'$ with atoms $\{a\}\cap\psi$, $\{a\}\cap\overline{\psi}$, {b}, and Your pignistic probabilities on it:

$$BetP'(\{a\}\cap\psi) = BetP'(\{a\}\cap\overline{\psi}) = 7/27 \qquad BetP'(\{b\}) = 13/27.$$

Then the unique solution for m_0 is: $m_0(\{a\}) = 2/9$, $m_0(\{b\}) = 3/9$ and $m_0(\{a,b\}) = 4/9$.
It solves the two linear equations derived from (3.1).

$$4/9 = m_0(\{a\}) + m_0(\{a,b\})/2$$

$$7/27 = m_0(\{a\})/2 + m_0(\{a,b\})/3.$$

It might seem odd that {b} receives pignistic probabilities of 5/9 and 13/27 according to the betting context. It reflects the fact that a large amount (4/9) of Your initial belief was left unassigned (i.e. given to {a,b}). This example corresponds to a state in which You have very weak support for {a} and for {b}. You are not totally ignorant as in section 3.3 (2), but still in a state of 'strong' ignorance. Part of BetP({b}) = 5/9 is due to justified beliefs (3/9) but the remainder results from a completely unassigned part of belief that You distribute equally among the alternatives of Your betting frame.

Wilson (1991) showed that the set of pignistic probabilities that can be obtained from a given belief function bel on a frame $\Re$ is equal to the set of probability functions 'compatible' with bel and its associated plausibilty function pl, i.e. the set of probability functions P on $\Re$ such that $bel(A) \leq P(A) \leq pl(A)\ \forall A \in \Re$. So whatever the betting frame, $BetP(A) \geq bel(A)\ \forall A \in \Re$. Suppose You ignore what the appropriate betting frame is, You nevertheless know that, $\forall A \in \Re$, the lowest bound of BetP(A) is bel(A). Therefore bel(A) can then be understood as the lowest pignisitic probability one could give to A when the betting frame is not fixed (Giles 1982).

3.5. Conditional betting behaviors.

Let us consider a credibility space $(\Omega, \Re, \text{bel})$ and let us suppose You learn that proposition $A \in \Re$ is true. Then bel must be conditioned on A by Dempster's rule of conditioning and BetP is built from this conditional belief function.

But a distinction must be made between the following two cases:

- Suppose You know that A is true, then You condition bel on A before deriving BetP.
- Suppose You know that the bet will be cancelled if A is false, then You derive BetP from the unconditioned bel and condition BetP on A using the classic probabilistic conditioning.

The first case corresponds to 'factual' conditioning, the second to 'hypothetical' conditioning. In the factual case, A is true for every bet that can be created. In the hypothetical case, A can be true in some bets, and false in others. Pignistic probabilities obtained in these two contexts usually reflect the difference between the contexts: (A always true, A possibly false). This distinction was already considered in Ramsey (1931, page 79) who noted that :

> "the degree of belief in P given Q…roughly expresses the odds at which the subject would now bet on P, the bet only to be valid if Q is true.…This is not the same as the degree to which he would believe P, if he believed Q for certain: for knowledge of Q might for psychological reasons profoundly alter his whole system of beliefs."

Ramsey distinguished between a bet on $P \cap Q$ versus $\overline{P} \cap Q$ in a context $\{P \cap Q, \overline{P} \cap Q, \overline{Q}\}$ and a bet on P versus $\overline{P}$ when Q is known to be true, hence in the context $\{P \cap Q, \overline{P} \cap Q\}$. In the TBM, Ramsey's allusion to "psychological reasons" applies at the credal level: learning Q modifies our credal state, hence of course our pignistic probabilities.

Note: Recent work by Dubois and Prade (1991) has shown the difference between two forms of conditioning: focusing and updating (that might better be called a revision). Our factual conditioning seems to corresponds to their updating. Focusing is not considered here.

3.6. The avoidance of Dutch Books.

1) The pignistic probability we build on the betting frame from the underlying belief guarantees that no static Dutch Book can be constructed. To construct a Dutch Book, one implicitly defines the betting frame, i.e. the set of atoms of the Boolean algebra build from all the options. The pignistic probabilities are built using this betting frame and no Dutch Book can be constructed as far as the bets are established according to a probability measure.

In order to show how Dutch Books are avoided, we reconsider the two bets under total ignorance considered in 3.3.(2). One could think of building the following Dutch Book[3].

" Before knowing $A_1 \cong B_1$, You would accept to pay \$.45 for winning \$1 if A_1 were true (as $BetP(A_1)=.5$). (For any bet, You would accept to pay up to \$x with x = BetP(X) if You won \$1 when X is true). You would also accept to pay \$.60 for winning \$1 if $(B_2 \cup B_3)$ were true (as $BetP'(B_2 \cup B_3)=.66$). Given that You don't know what A_1, A_2, B_1, B_2 and B_3 say, the two bets are acceptable together. Now You learn that $(B_2 \cup B_3)$ is true iff A_2 is true. Therefore, by accepting the two bets together, You commit Yourself to pay \$(.45+.60) = \$1.05 for the certainty of winning \$1. Hence a Dutch Book has been build against You, as You will surely loose \$.05."

The argument is wrong because it does not take into due consideration the problem of the betting frame. Once $A_1 \cong B_1$ is known, You will not accept both bets simultaneously. Before accepting a bet, You must always build the betting frame i.e. You must establish the granularity i.e. the list of elements on which stakes can freely be allocated.

In the present case, once You know $A_1 \cong B_1$, You must decide if stakes on B_2 and B_3 will always be the same or might vary. If they must always be the same, then You use the betting frame $\{A_1, A_2\}$ and reject the second bet. If they might be different, then You use the betting frame $\{B_1, B_2, B_3\}$ and reject the first bet. Dutch Books are thus avoided.

2) The existence of two types of conditioning does not permit the construction of a dynamic Dutch Book. If bets are based on 'hypothetical' facts, conditioning must then be performed by applying classical probability conditioning. If bets are based on 'factual' facts, then ***every*** bets must be organized accordingly. Some atoms are definitively

[3] This example was suggested by P. Garbolino in Clarke et al. (1991).

eliminated as they are impossible, conditioning is performed at the level of the belief function by applying Dempster's rule of conditioning and BetP is derived from the conditional belief function. Dutch Books can still be avoided as one can not build a set of bets where the 'factual' fact is treated as unknown in some cases, and accepted in others. A 'factual' fact is either known or unknown, but once it is known, it is known for every bets. The difference between 'hypothetical' facts and 'factual' facts is to be found in the fact that 'factual' facts are true for every bets, whereas hypothetical facts can be denied in some bets.

4. The murder of Mr. Jones.

4.1. The problem.

Big Boss has decided that Mr. Jones must be murdered by one of the three people present in his waiting room and whose names are Peter, Paul and Mary. Big Boss has decided that the killer on duty will be selected by a throw of a dice: if it is an even number, the killer will be female, if it is an odd number, the killer will be male. You, the judge, know that Mr. Jones has been murdered, who was in the waiting room and know about the dice throwing, but You do not know what the outcome was and who was selected. *You are also ignorant as to how Big Boss would have decided between Peter and Paul in the case of an odd number being observed.* Given the available information, Your odds for betting on the sex of the killer would be 1 to 1 for male versus female.

You then learn that should Peter not be selected by Big Boss, he would necessarily have gone to the police station at the time of the killing in order to have a perfect alibi. Peter indeed went to the police station, so he is not the killer. The question is how You would bet now on male versus female: should Your odds be 1 to 1 (as in the TBM) or 1 to 2 (as in the Bayesian model)

Note that the alibi evidence makes 'Peter is not the killer' and 'Peter has a perfect alibi' equivalent. The more classical evidence 'Peter has a perfect alibi' would only imply P('Peter is not the killer' | 'Peter has a perfect alibi') = 1. But P('Peter has a perfect alibi' | 'Peter is not the killer') would stay undefined and would then give rise to further discussion, which for our purpose would be useless. In this presentation, the latter probability is also 1.

4.2. The TBM approach

Let k be the killer. The waiting room evidence E_0 and its resulting basic belief assignment m_0 are:

E_0: $k \in \Omega = \{Peter, Paul, Mary\}$, $\Re_0 = 2^\Omega$
$m_0(\{Peter, Paul, Mary\}) = 1$.

The dice throwing pattern (evidence E_1) induces the following basic belief assignment :

E_1 : dice experiment, $\Re_1 = \{Male, Female\}$
$m_1(Female) = .5$
$m_1(Male) = .5$

Conditioning m_0 on E_1 by Dempster's rule of conditioning induces the basic belief assignment m_{01}:

E_{01}: E_0 and E_1, $\Re_{01} = 2^\Omega$
$m_{01}(\{Mary\}) = .5$
$m_{01}(\{Peter, Paul\}) = .5$

The .5 belief mass given to {Peter, Paul} corresponds to that part of belief that supports "Peter or Paul", could possibly support each of them, but given the lack of further information, cannot be divided more specifically between Peter and Paul.

Suppose You had to bet on the killer's sex. You would obviously bet on Male = {Peter, Paul} versus Female = {Mary} at odds 1 to 1.

Peter's alibi pattern (evidence E_2) induces the basic belief assignment m_2.

E_2 : A = "Peter went to the police station" = "Peter has a perfect alibi"
E_2 : $k \in \{Paul, Mary\}$, $\Re_2 = 2^\Omega$
$m_2(\{Paul, Mary\}) = 1$

Conditioning m_{01} on E_2 by Dempster's rule of conditioning leads to m_{012}

E_{012} : E_{01} and E_2, $\Re_{012} = 2^\Omega$
$m_{012}(\{Mary\}) = m_{012}(\{Paul\}) = .5$

The basic belief mass that was given to "Peter or Paul" is transferred to Paul. Your odds for betting on Male versus Female would now still be 1 to 1, as before.

4.3. The Bayesian solution.

Suppose You were a Bayesian. Therefore Your degrees of belief are quantified by probability distributions and all pieces of evidence are taken in consideration by adequately revising Your probability distributions through the Bayesian conditioning processes.

Given E_1 , You build a probability distribution P_1 on Ω = {Peter, Paul, Mary}:

$$P_1(k \in \{Mary\}) = .5 \qquad P_1(k \in \{Peter, Paul\}) = .5.$$

You would also bet on male versus female the odds being 1 to 1.

When You learn E_2, i.e. that Peter went to the police station, You condition P_1 on {Paul, Mary} in order to compute P_{12} where :

$$P_{12}(k \in \{Mary\}) = P_1(k \in \{Mary\} \mid k \in \{Mary, Paul\}) =$$

$$\frac{P_1(k \in \{Mary\})}{P_1(k \in \{Mary\}) + P_1(k \in \{Paul\})} = \frac{.5}{.5 + x}$$

with $x = P_1(k \in \{Paul\})$.

But x is unknown. No evidence whatsoever has been given about x.

Usually Bayesians encountering this problem will assume that x = .25 leading to a 1 to 2 odds. They obtain x = .25 by either applying the insufficient reason principle or a symmetry argument or a minimum entropy argument on P_1 to evaluate $P_1(k \in \{Paul\})$. It is of course the most natural assumption…but it is still an assumption extraneous to the available evidence, and any other value in [0, .5] could as well be assumed. Any such value would correspond to some *a priori* probability on Peter versus Paul, which is not justified by any of the available pieces of evidence. All that is known to You is that there were two men whose names were Peter and Paul… and NOTHING else.

Another justification for x = .25 could be obtained as follows. Suppose evidence E'_2 : “if Paul were not the killer, he would go to the police station to have a perfect alibi and Paul went to the police station”. E'_2 is E_2 where Peter and Paul interchange their role. A bet on male versus female should be the same after evidence E_2 and after evidence E'_2. This symmetry requirement is satisfied only with x =.25. Therefore Bayesians can hardly avoid the 1 to 2 odds. In the TBM, the requirement that bets after E_2 and after E'_2 should be the same is automatically satisfied: the .5 mass that was given by m_{01} to ‘Peter or

Paul’ is transferred to ‘Paul’ under E_2 and to ‘Peter’ under E'_2 and the bets of male versus female remains unchanged.

Our analysis of Mr. Jones's case could be rephrased by saying that Big Boss used a deck of 52 cards instead of a dice. Mary is the killer if the card is red, male is the killer if the card is black. Peter is not the killer. In how many ways could Big Boss have selected a card so that Paul is the killer? The answer is not "any number between 0 and 26" as none of the cards had Paul written on them. All black cards are identical, they all mean "male". To introduce the idea that some black cards could point to Paul, the others to Peter, would lead to an ULP analysis as we would be in a context in which there is a probability that Paul is the killer (the proportion of cards marked Paul) but we do not know the value of such a proportion. This is another problem, different from the one we have analysed. The two problems should not be confused. The difference between such ULP approach and the TBM is detailed in Dubois et al. (1991), Smets (1987, 1990d).

4.4. Conditional bets.

The example can be used to illustrate the difference between bets according to the context in which the conditioning information E_2 is taken into account (see §3.5). Before learning evidence E_2, if You want to bet on Paul versus Mary, the betting frame is {Paul, Mary, Peter} and BetP(Paul) = BetP(Peter) = .25, BetP(Mary) = .5. To bet on Paul versus Mary corresponds then to conditioning the pignistic probabilities BetP on ¬Peter, hence the resulting pignistic probabilities BetP'(Paul)=1/3 and BetP'(Mary)=2/3 and the 1 to 2 odds. After learning evidence E_2, the betting framework is {Paul, Mary}, You condition bel_{01} on ¬Peter from which You derive bel_{012}, the pignistic probabilities BetP(Paul) = .5 and BetP(Mary) = .5 and the 1 to 1 odds.

The difference results from Your openness to the fact that Peter might be the killer before learning E_2 and Your knowledge that Peter is not the killer after learning E_2.

5. The guards and posts paradigm.

1) We will present **three paradigms**: the guards and posts, the translators and the unreliable sensor paradigms. The first paradigm helps to explain the conditioning process. The second paradigm shows that the TBM solution is fundamentally different from the ULP solution, but might lead to the mistaken idea that the TBM is somehow related to likelihood theory. The third paradigm shows that the TBM leads to a solution different from the Bayesian, the likelihood and the fiducial solutions.

In each paradigm, the Boolean algebra $\Re$ on which beliefs are defined is the power set of Ω, the frame of discernment.

2) **The paradigm**[4]**.** Suppose a military officer must organise guard duty in his camp. There are three possible posts (π_1, π_2 and π_3) but only one is to be occupied. There are three soldiers who could be appointed for guard duty (S_1, S_2 and S_3). The officer will randomly select one of the two soldiers by tossing a dice. Soldier S_1 is selected if the dice outcome is 1 or 2, soldier S_2 is selected if the dice outcome is 3 or 4, and soldier S_3 is selected if the dice outcome is 5 or 6. Each soldier has a habit in that

- if selected, soldier S_1 will always go to post π_1 or π_2,
- if selected, soldier S_2 will always go to post π_1 or π_2 or π_3.
- if selected, soldier S_3 will always go to post π_2,

Before the officer selects the soldier, each of them writes down on a piece of paper where he will go if he is selected. As a result, there are six possible worlds $w_{i,}$ i=1,…6, where each world corresponds to one particular selection of the posts (see left-hand column of table 1). After the officer selects the guard, there are 18 worlds (referred as worlds w_{ij} if soldier S_j is selected in world w_i). You are about to attack the camp and You want to know which post is occupied in order to avoid it. You know all the facts described up to here, but You do not know which soldier was selected. What is Your belief about which post is occupied. The frame of discernement $\Omega = \{\pi_1, \pi_2, \pi_3\}$ and $\Re = 2^\Omega$. Table 2 presents the degrees of belief for each sets of posts. Initially the basic belief assignment on $\Re$ is such that $m(\{\pi_1, \pi_2\}) = m(\{\pi_1, \pi_2, \pi_3\}) = m(\{\pi_2\}) = 1/3$.

[4] This paradigm was suggested by Yen-Teh Hsia.

world	post selected by each soldier			occupied post according to the selected soldier			remaining worlds after case 1 conditioning			remaining worlds after case 2 conditioning		
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
w_1	π_1	π_1	π_2	π_1	π_1	π_2	π_1	π_1	π_2	π_1	π_1	
w_2	π_1	π_2	π_2	π_1	π_2	π_2				π_1		
w_3	π_1	π_3	π_2	π_1	π_3	π_2	π_1	π_3	π_2	π_1	π_3	
w_4	π_2	π_1	π_2	π_2	π_1	π_2					π_1	
w_5	π_2	π_2	π_2	π_2	π_2	π_2						
w_6	π_2	π_3	π_2	π_2	π_3	π_2					π_3	
worlds w_i	1/3	1/3	1/3	18 worlds w_{ij} where soldier S_j is selected in w_i								
	probabilities of being selected											

Table 1: The set of six worlds that represent the six possible ways posts could be selected by each soldier, and the post occupied in the eighteen possible worlds after the soldier has been selected by the officer.

posts	initial state		case 1		case 2	
	m	bel	m	bel	m	bel
π_1	0	0	.33	.33	.5	.5
π_2	.33	.33	.33	.33	0	0
π_3	0	0	0	0	0	0
$\{\pi_1, \pi_2\}$	.33	.66	0	.66	0	.5
$\{\pi_1, \pi_3\}$	0	0	.33	.66	.5	1.0
$\{\pi_2, \pi_3\}$	0	.33	0	.33	0	0
$\{\pi_1, \pi_2, \pi_3\}$	.33	1.0	0	1.0	0	1.0

Table 2: Degrees of belief and their related basic belief masses for the paradigm of table 1. Before conditioning and after cases 1 and 2 conditionings.

3) Two cases of conditioning can then be considered.

Case 1: The soldiers and You learn that post π_2 is so unpleasant to occupy that the soldiers will not select it if they can go elsewhere (it applies thus to soldiers S_1 and S_2, but not S_3). Hence the worlds w_2, w_4, w_5 and w_6 become impossible (table 1). Table 2 presents Your beliefs about which post is occupied. The .33 basic belief masses given initially to $\{\pi_2\}$, $\{\pi_1, \pi_2\}$ and $\{\pi_1, \pi_2, \pi_3\}$ are tranferred to $\{\pi_2\}$, $\{\pi_1\}$ and $\{\pi_1, \pi_3\}$, respectively. Suppose You decide to bet on which post is occupied. The betting frame is $\{\pi_1, \pi_2, \pi_3\}$. The pignistic probability is $BetP(\pi_1) = 3/6$, $BetP(\pi_2) = 2/6$ and $BetP(\pi_3) = 1/6$.

Case 2: You are able to observe post π_2 and it is empty. Hence, the selected soldier had not selected π_2 before being assigned guard duty. Thus the actual world is not one of the worlds $w_{2,2}$, $w_{4,1}$, $w_{5,1}$, $w_{5,2}$ $w_{6,1}$, $w_{1,3}$, $w_{2,3}$, $w_{3,3}$, $w_{4,3}$, $w_{5,3}$, and $w_{6,3}$. After renormalization, basic belief masses .5 are given to $\{\pi_1\}$ and $\{\pi_1, \pi_3\}$. The betting frame is $\{\pi_1, \pi_3\}$. The pignistic probability is $BetP(\pi_1) = 3/4$ and $BetP(\pi_3) = 1/4$.

4) A probability solution could be derived if You accept that the .33 masses given to each soldier are somehow distributed among the six possible worlds. Suppose You accept an equal distribution. So each of the eighteen worlds receives a probability of 1/18. Case 1 conditioning would leave a probability 1/6 to each of the six remaining worlds. The derived probabilities on $\{\pi_1, \pi_2, \pi_3\}$ are $P(\pi_1) = 3/6$, $P(\pi_2) = 2/6$ and $P(\pi_3) = 1/6$, as in the TBM analysis. In case 2 conditioning, the solutions differ: $P(\pi_1) = 5/7$ and $P(\pi_3) = 2/7$.

5) Even without postulating the equi-distribution of the .33 basic belief masses among the eighteen worlds w_{ij}, probabilists might be tempted to defend the idea that the probabilities .33 used for the soldier selection do not apply once the conditioning information of case 2 is known. Indeed, they could defend that the fact that π_2 is not occupied somehow supports the hypothesis that soldier S_2 was selected. Hence, an updated probabilities P' should be such that $P'(S_2)>P'(S_1)$. This is the basis of Levi's criticisms (see section 8). The answer from the TBM point of view is that no probability whatsoever is built on the w_{ij} space, only on the S_j space. So the fact that, in case 2, there are fewer remaining possible worlds for S_1 than for S_2 (3 versus 4) is irrelevant. Case 2 really is the case in which the TBM acquires its originality when compared with the probability approach.

6. The translator paradigm.

1) Shafer and Tversky (1985) have described a **translator experiment** to explain Shafer's theory. Let $T = \{t_i: i = 1, 2 \dots n\}$ be a set of translators and $\Omega = \{c_j: j=1, 2, 3\}$ be a set of messages that can be generated by a given device. For each message $c_j \in \Omega$, the translator t_i translates it into an element of some given space Θ. Let $f_i(c_j)$ denotes the element of Θ obtained by the translation performed by translator t_i of the message c_j. Table 3 presents an example where $\Omega = \{c_1, c_2, c_3\}$, $\Theta = \{\theta, \overline{\theta}\}$ and $T = \{t_0, \dots t_7\}$. The

crosses in the right half indicate the elements of Ω that are translated into θ for each translator. So translator t_1 translates c_1 into θ and c_2 and c_3 into $\overline{\theta}$: $f_1(c_1) = \theta$, $f_1(c_2) = f_1(c_3) = \overline{\theta}$. Given $\theta \in \Theta$, let $A_i \subseteq \Omega$ be the set of messages that are translated as θ by translator t_i. In table 3, $A_1 = \{c_1\}$, $A_4 = \{c_1, c_2\}$… Note that it was not said that $A_i \cap A_j = \emptyset$ for $i \neq j$. Suppose that a message is selected in Ω (we do not say randomly selected, see section 8). Suppose that a translator is selected **by a chance process** among the set T of translators and independently of the selected message. Let p_i be the probability that translator t_i is selected. You observe only θ, the result of the transformation of the unknown message, but You ignore which translator was selected. Furthermore You are totally ignorant of how the message was selected.

What can be said about Your beliefs bel(c) for $c \subseteq \Omega$ given θ was observed?

		$f_i(c_j)$			given θ		
			Ω			Ω	
T	$P(t_i)$	c_1	c_2	c_3	c_1	c_2	c_3
t_0	p_0	$\overline{\theta}$	$\overline{\theta}$	$\overline{\theta}$	.	.	.
t_1	p_1	θ	$\overline{\theta}$	$\overline{\theta}$	x	.	.
t_2	p_2	$\overline{\theta}$	θ	$\overline{\theta}$	.	x	.
t_3	p_3	$\overline{\theta}$	$\overline{\theta}$	θ	.	.	x
t_4	p_4	θ	θ	$\overline{\theta}$	x	x	.
t_5	p_5	θ	$\overline{\theta}$	θ	x	.	x
t_6	p_6	$\overline{\theta}$	θ	θ	.	x	x
t_7	p_7	θ	θ	θ	x	x	x

Table 3: Translator paradigm with 8 translators t_i, 3 messages c_j and 2 observations θ and $\overline{\theta}$. The last part presents the values of $f_i(c_j)$. The right part presents the elements of Ω that are translated into θ for each translator.

2) With the **TBM**, the following basic belief masses are assumed on TxΩx Θ:

$$m\left(\cup_{\tau \in \Theta} \cup_{j \in J_\tau} \{(t_i, \tau, c_j)\} \right) = p_i \quad \text{where } J_\tau = \{j: f_i(c_j) = \tau\} \tag{5.1}$$

So p_2 is allocated to $\{(t_2, \overline{\theta}, c_1), (t_2, \theta, c_2), (t_2, \overline{\theta}, c_3)\}$, p_5 is allocated to $\{(t_5, \theta, c_1), (t_5, \overline{\theta}, c_2), (t_5, \theta, c_3)\}$... The origin of such an assignment is to be found in section 8.

Learning that θ is true, the transfer of the basic belief masses m by Dempster's rule of conditioning leads to the basic belief masses m* on TxΩx{θ}:

$$m^*\left(\cup_{j \in J_\theta} \{(t_i, \theta, c_j)\} \right) = p_i$$

and $\quad m'(\cup_{j\in J_\theta} \{(t_i, \theta, c_j)\}) = \frac{p_i}{1-p_0} = p'_i$

i.e. the weight p_i is given to those messages indicated by the crosses on the line t_i in the column 'given θ' (table 3, right part). The knowledge that there are only eight translators (closed-world assumption) justifies the normalization of the basic belief assignment m* into m' by dividing m* by $1-p_0$.

By marginalization of m' (i.e. bel') on Ω, one computes for $c\subseteq\Omega$

$$bel_\theta(c) = bel'(\cup_i \cup_{c_j\in c} \{(t_i, \theta, c_j)\}) = \sum_{i\in I} p'_i$$

with $I = \{i: f_i^{-1}(\theta)\subseteq c\}$ and where $p'_0 = 0$.

For example: $m_\theta(\{c_1\}) = p'_1$, $m_\theta(\{c_1, c_2\}) = p'_4$.

Table 4 presents the value of $bel_\theta(c)$ and $pl_\theta(c)$ for some $c\subseteq\Omega$.

Ω	$bel_\theta(c)$	$pl_\theta(c)$
$\{c_1\}$	p'_1	$p'_1+p'_4+p'_5+p'_7$
$\{c_1,c_2\}$	$p'_1+p'_2+p'_4$	$p'_1+p'_2+p'_4+p'_5+p'_6+p'_7$
$\{c_1,c_2,c_3\}$	$p'_1+p'_2+p'_3+p'_4+p'_5+p'_6+p'_7$	$p'_1+p'_2+p'_3+p'_4+p'_5+p'_6+p'_7$

Table 4: Translator paradigm: TBM analysis.

3) For a **Bayesian analysis**, one claims the existence of some $P(c_j)$: j = 1, 2, 3, but their values are missing. One must compute:

$$P(c_j \mid \theta) = \frac{P(\theta \mid c_j)\, P(c_j)}{P(\theta)}.$$

One has:

$$P(\theta \mid c_j) = \sum_i P(\theta \mid t_i, c_j)\, P(t_i \mid c_j)$$

$$= \sum_{i\in I} p_i \qquad \text{where } I = \{i: f_i(c_j)=\theta\}$$

because $\quad P(\theta \mid t_i, c_j) = 0$ if $f_i(c_j) = \bar{\theta}$

$\qquad 1$ if $f_i(c_j) = \theta$

and $\quad P(t_i \mid c_j) = p_i$

as the translators are selected independently of the message.

The ULP for $P(c_j \mid \theta)$ are computed by evaluating its extremes where the $P(c_j)$ are allowed to vary on their domain. Let Δ be the set of all vectors (x, y, z) where $x, y, z \in [0,1]$ and $x+y+z=1$. The vectors $(P(c_1), P(c_2), P(c_3))$ are the elements of Δ. The upper and lower conditional probabilities for c_j given θ are

$$P^*(c_j \mid \theta) = \sup_\Delta P(c_j \mid \theta) = \sup_\Delta \frac{P(\theta \mid c_j)\, P(c_j)}{\sum_\nu P(\theta \mid c_\nu)\, P(c_\nu)} = 1$$

$$P_*(c_j \mid \theta) = \inf_\Delta P(c_j \mid \theta) = \inf_\Delta \frac{P(\theta \mid c_j)\, P(c_j)}{\sum_\nu P(\theta \mid c_\nu)\, P(c_\nu)} = 0$$

This ULP solution provides no information and is different from the TBM solution.

4) The **Dempster-Shafer** analysis of the paradigm leads to the same solution as the TBM. But the origin of this solution is connected to probability theory and open to criticisms about the appropriateness of the use of Dempster's rule of conditioning.

In Dempster-Shafer theory, it is postulated that there is a mapping M from T to $T\text{x}\Omega\text{x}\Theta$, that there is a probability distribution on T and $bel(\omega)$ for $\omega \subseteq T\text{x}\Omega\text{x}\Theta$ is defined as the probability of $M_*^{-1}(\omega) = \{t_i : M(t_i) \subseteq \omega\}$:

$$bel(\omega) = P(M_*^{-1}(\omega))$$

The knowledge that θ is true induces the adaptation of M into $M_\theta = M \cap (T\text{x}\Omega\text{x}\{\theta\})$ and bel is updated into:

$$bel_\theta(\omega) = P(M_{\theta *}^{-1}(\omega)) = P(\{t_i : M_\theta(t_i) \subseteq \omega\})$$

In the present paradigm, one has a.o. $M(t_4) = \{(t_4, c_1, \theta), (t_4, c_2, \theta), (t_4, c_3, \overline{\theta})\}$ and $P(M(t_4)) = p_4$ is correct. Once θ is known, the mapping M becomes $M_\theta = M \cap (T\text{x}\Omega\text{x}\{\theta\})$: a.o. $M_\theta(t_4) = \{(t_4, c_1, \theta),(t_4, c_2, \theta)\}\})$ and $bel(M_\theta(t_4)) = p'_4$ where the denominator is the normalization factor related to the closed-world assumption (given θ, we know that t_0 was not the translator).

The problem with such a model is: why do we use p'_i and not $P(t_i|\theta)$ as it should according to probability theory (Levi 1983)?

One has $P(t_i|\theta) \approx P(\theta|t_i)\, p_i = P(\{c_j: f_i(c_j)=\theta\}|t_i)\, p_i$

e.g. $P(t_1|\theta) \approx P(c_1)\, p_1$, $P(t_4|\theta) \approx P(\{c_1, c_2\})\, p_4$… It is impossible to get $P(t_i|\theta) = P(t_i)$ for all i. So the Dempster-Shafer solution can not be obtained.

The difficulty with the paradigm lies in the fact that the values of $P(t_i|\theta)$ are unknown as the P(c)'s are unknown. Note that if one could assess $P(t_i|\theta)$, then P(c) could be deduced. In that case all problems would disappear. Each analysis, be it the TBM, the Dempster-Shafer or any upper and lower probabilities analysis, would lead to the Bayesian solution. But we are considering the case where $P(t_i|\theta)$ cannot be assessed as P(c) is completely unknown. In the TBM, such a probability measure on $T \times \Omega \times \Theta$ and the concept of $P(t_i|\theta)$ are neither assumed nor even defined. Once θ is known, the TBM conditions the initial belief on $T \times \Omega \times \{\theta\}$ by transferring the basic belief masses.

Levi's criticism of the Dempster-Shafer analysis is based on the assumption that this last analysis represents a generalized Bayesian analysis, in which case the concept of a probability on $T \times \Omega \times \Theta$ is claimed. Once all relations with probability theory are set aside, such grounds for criticism disappear (see section 8(5)).

The **difference between the TBM and the Dempster-Shafer solutions** resides in the fact the TBM is free from any underlying probability theory. The probability information relative to the translators explained the origin of the basic belief masses at the credal level. But apart from that, any concept of probability is useless. A statement like

$$bel_\theta(c) \leq P(c|\theta) \leq pl_\theta(c)$$

is meaningless as we never build any probability measure on the frame of discernment Ω at the credal level, so the symbol P is undefined. ($P(c|\theta)$ should not be confused with $BetP_\theta(c)$ derived at the pignistic level.). Note that given any belief function, one can build a set of compatible probability functions such that

$$bel(A) \leq P(A) \leq pl(A) \qquad \forall A \in \Re$$

This is just a mathematical property without any interpretation relevant to the model (except for the comments at the end of section 3.4).

The TBM could be viewed as a 'purified' Dempster-Shafer model, i.e. purified from any probabilistic connotation. Hence, it forestalls criticisms aimed at the strange conditioning process encountered in the Dempster-Shafer solution which is at odds with plain probability approaches.

5) It is interesting to note that

$$pl_\theta(c_j) = P(\theta \mid c_j) = l(c_j \mid \theta)$$

where $l(c_j \mid \theta)$ is the likelihood of c_j given θ. There are some analogies between the TBM and **likelihood theory**. On the singletons of Ω, the two solutions are equivalent here. The analogy between the likelihood solution and the TBM solution is not always present as will be seen in the unreliable sensor paradigm.

7. The unreliable sensor paradigm.

1) **The paradigm.** Let us consider a sensor with which You must check the temperature of a preparation: either the temperature is 'cold' (Cold) or 'hot' (Hot). Under correct working conditions, the sensor answers are given by a lamp that is "blue" (B) if the temperature is cold, and "red" (R) if the temperature is hot. Unfortunately, the sensor is not reliable as its thermometer is sometimes broken, in which case the sensor status can be B or R. In such a context, the sensor answer (B or R) is unrelated to the real temperature (Cold or Hot).

The only information known by You is what is indicated on the box containing the sensor: "Warning: the thermometer included in this sensor can be broken. **The probability that it is broken is 20%**. When the thermometer is **not broken**, the sensor is a perfectly reliable detector of the temperature situation. When the thermometer is not broken: a blue light means the temperature is cold, a red light means that the temperature is hot. When the thermometer is **broken**, the sensor answer is unrelated to the temperature".

You use the sensor and the light is red. What is Your degree of belief bel(Hot|R) that the temperature is hot given the red light is on?

Let Θ = {R, B}, Ω = {Cold, Hot}, T = {ThW, ThB} where ThW and ThB mean 'thermometer-sensor in working conditions' and 'thermometer-sensor broken'.

2) The **TBM solution** consists in assuming that the masses .8 and .2 are allocated on ΘxΩxT such that (see figure 2)

m({(R, Hot, ThW) , (B, Cold, ThW)}) = .8

m({(R, Cold, ThB) , (R, Hot, ThB) , (B, Cold, ThB) , (B, Hot, ThB)}) = .2

When You learn that the light is red (R), the masses are transferred such that

m'({(R, Hot, ThW)}) = .8

m'({(R, Cold, ThB) , (R, Hot, ThB)}) = .2

Marginalization on Ω provides:

$bel_R(Hot) = .8$ $bel_R(Cold) = .0$

$pl_R(Hot) = 1.$ $pl_R(Cold) = .2$

Should You have any *a priori* knowledge about the risk that the temperature is hot or cold, the credal equivalent of P(Hot) and P(Cold), it should be combined with the present results by Dempster's rule of combination.

3) The **Bayesian solution** assumes P(ThW) = .8, P(ThB) = .2. One also has P(Hot|R, ThW) = 1, P(Cold|B, ThW) = 1. Note that we do not have P(R|ThB) = 0, the light can be red when the thermometer is broken. Furthermore, when the thermometer is broken, the probability that the system is Hot is independent from the sensor answer, and the thermometer is broken independently of the status of the system: P(Hot|R, ThB) = P(Hot). Also when the thermometer is in working condition (ThW), the probability that the light is red is the probability that the temperature is hot (Hot): P(R|ThW) = P(Hot). Then:

$$\begin{aligned} P(Hot|R) &= P(Hot|R, ThW)\, P(ThW|R) + P(Hot|R, ThB)\, P(ThB|R) \\ &= \frac{1 \,.\, P(R|ThW)\, P(ThW)}{P(R)} + \frac{P(Hot)\, P(R|ThB)\, P(ThB)}{P(R)} \\ &= \frac{.8\, P(Hot) + .2\, P(Hot)\, P(R|ThB)}{P(R|ThW)\, P(ThW) + P(R|ThB)\, P(ThB)} \\ &= \frac{P(Hot)\, (\, .8 + .2\, P(R|ThB)\,)}{8\, P(Hot) + .2\, P(R|ThB)} \end{aligned}$$

The real problem encountered by the Bayesians is not so much in assessing P(Hot), which could be known in practice but P(R|ThB), i.e. the probability that the light is red when the thermometer is broken. It is hard to imagine a serious hypothesis for such an ill-defined probability. There are so many ways for a thermometer to be broken that any particular value seems hardly justified. Bayesians could go on by assuming such a value … but of course the quality of their conclusions is strictly related to the quality of their assumptions.

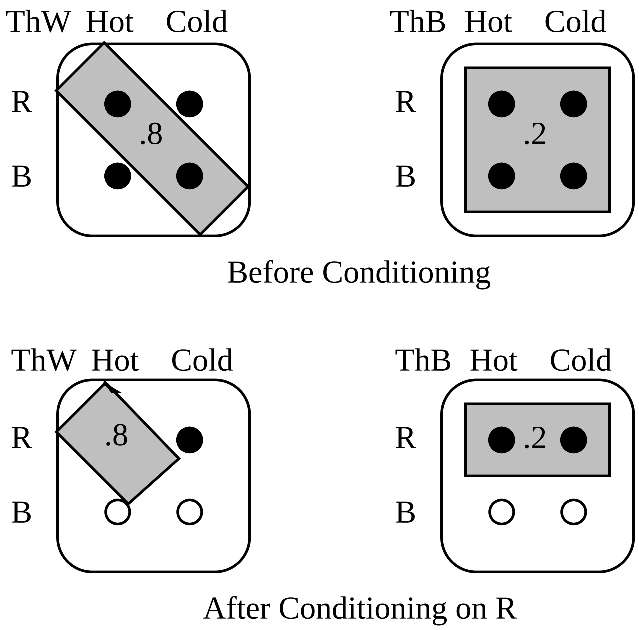

Figure 2: The unreliable sensor paradigm, basic belief assignment and impact of conditioning on R.

4) The **likelihood solution** l(Hot|R) = P(R|Hot) cannot be derived in this example as we cannot assess P(R|Hot)

P(R|Hot) = P(R|ThW, Hot) P(ThW|Hot) + P(R|ThB, Hot) P(ThB|Hot)
= 1x.8 + .2 P(R|ThB)

and we are faced with the same problem as the Bayesians: what is P(R|ThB)? In this example pl_R(Hot) = 1: it is different from l(Hot|R) (except if you can defend P(R|ThB) = 1). Hence the TBM solution is not the likelihood solution.

5) The TBM solution is also different from the fiducial solution. A **fiducial analysis** might consist in assuming:

P(R|Hot) = P(B|Cold) = .8
P(B|Hot) = P(R|Cold) = .2

in which case, whatever P(Hot),

P({(R, Hot) , (B, Cold)}) = .8

and P({(B, Hot) , (R, Cold)}) = .2

As we know that R is true, the .8 mass is transferred to (R, Hot) and the mass .2 to (R, Cold). Marginalization on Ω gives P(Hot|R) = .8 and P(Cold|R) = .2. The solution is similar to the TBM as far as bel(Hot|R) is concerned, but not as far as bel(Cold|R) is concerned. The TBM does not provide any support to Cold, whereas the fiducial model gives it a .2 support, hence the difference between them.

8. Origin of the basic belief masses.

1) In each paradigm, a certain probability is associated with a certain basic belief mass. Such underlying probabilities are not necessary, as shown in the example of section 3.2 (2), but they simplify our presentation. Should they had been omitted, the origin of the basic belief masses might have been felt as somehow misterious. We explain now the link between the basic belief masses and the probabilities when they exist.

Let $(\Omega, \Re)$ be a propositional space. Let us suppose that Your evidential corpus EC_t^Y induces a belief on $\Re$ such that there is a coarsening $\Re'$ of $\Re$ on which Your beliefs are described by a probability distribution P'. To obtain Your degrees of belief bel' on $\Re'$, only the frame $\Re'$ needs to be considered. One gets: $\forall A \in \Re'$, bel'(A) = P'(A), and

$m'(x) = P'(x)$ for all atoms x of $\Re'$

$m'(A) = 0$ for non atomic A of $\Re'$.

The numerical equality between bel' and P' can be justified by generalizing Hacking’s Frequency Principle (Hacking 1965) to belief functions. The original Principle is: when the objective probability of an event A is p, then the subjective probability of A is p. We just generalize it by requiring that the belief in A is p when the probability of A is p (whatever the nature of the probability).

In our paradigms, the atoms of the coarsenings $\Re'$ are:

-in the murder of Mr. Jones: {Mary} and {Peter, Paul}
-in the guards and posts paradigm: the three soldiers
-in the translator paradigm: the eight translators
-in the unreliable sensor: the states ThW and ThB.

Problems appear once $\Re$ is considered. Probabilists claim that the probability P'(x) given to atom $x \in \Re'$ is the sum of the probabilities P(y) given to the atoms y of $\Re$ that belong to $\Lambda(x)$, where Λ is the refining from $\Re'$ to $\Re$ corresponding to the coarsening from $\Re$ to $\Re'$:

$$P'(x) = \sum_{y \in \Lambda(x)} P(y)$$

Given EC_t^Y that tells nothing about the value of P, You can only compute ULP for the P(A)'s for $A \in \Re$ or create P by using some general principles (like the insufficient reason principle or maximum entropy principle…). The major point about probability analyses is that the probabilists postulate the existence of a probability distribution P on $\Re$, an item of information in fact not included in EC_t^Y.

The TBM considers only the information in EC_t^Y, hence it **does not postulate any probability distribution P on** $\Re$. Nowhere in the paradigms are such functions P on $\Re$ claimed. In practice to build bel on $\Re$, we only allocate the masses m'(x) to $\Lambda(x) \in \Re$

$$m(\Lambda(x)) = m'(x) \quad \text{for all atoms } x \in \Re'$$
$$= 0 \quad \text{otherwise}$$

Such allocation is justified by the Principle of Least Commitment (Smets 1991c). This principle translates the idea that You should not give more support to a proposition than justified. The least committed belief function bel induced by bel' on $\Re$ is the vacuous extension of bel' on $\Re$ (Shafer 1976 pg 146). Let $\mathfrak{B}$ be the set of belief functions bel* defined on $\Re$ such that $bel^*(X) = bel'(X)\ \forall X \in \Re'$. The vacuous extension bel of bel' is the minimal element of $\mathfrak{B}$ such that $bel(A) \le bel^*(A)\ \forall A \in \Re, \forall bel^* \in \mathfrak{B}$.

2) We selected paradigms for which there exists a probability distribution on a coarsening because at least the numerical values given initially to the basic belief masses can be explained. The evaluation of the basic belief masses when there is no coarsening $\Re'$ on which a probability distribution can be defined is discussed in 3.4.

3) The major difference between the TBM and probabilistic approaches obviously lies in the way we create the beliefs on $\Re$ knowing the belief on $\Re'$. The TBM is based on what is available and nothing else, whereas the probability analysis requires the existence of a probability distribution on $\Re$. Consider the murder of Mr. Jones: in the case of a male killer (odd number thrown) the TBM accepts that Peter is **arbitrarily** selected by Big Boss whereas the probabilists claim that Peter is **randomly** selected by Big Boss.

4) The non existence of a probability distribution on $\Re$ resolves the problem raised by Levi (1983). Let us consider the translator paradigm. Once θ is learnt, why don't we condition the $p_i = p(t_i)$ on θ and thus use $p(t_i|\theta)$ as should be the case in a bona fide probability analysis? The concept of $p(t_i|\theta)$ is valid iff one can describe a probability distribution at least on the space TxΘ, which is not claimed in the TBM analysis. Hence, Levi's criticism does not apply to our model, but it does apply to some interpretations of the Dempster-Shafer model (those with an ULP connotation).

9. Handling evidence at the credal level.

We shall now detail how evidence is handled in the TBM and Bayesian models.

The TBM is based on the following underlying model:

- credal states that represent the impact of the evidential corpus EC_t^Y on a boolean algebra $\Re$ are described by belief functions on $\Re$,
- pieces of evidence are to be taken into consideration at the credal level, conditioning being obtained via Dempster's rule of conditioning, the rule that underlies the TBM.
- whenever a bet has to be established, the betting frame must be defined and the credal state prevailing at that time t induces a pignistic probability distribution on the elements of the bet.

This schema becomes in the case of the murder of Mr. Jones:

		E_1		E_2	
credal state	m_0	-------------> throwing the dice	m_{01}	-------------> Peter's alibi	m_{012}
			| | ↓		| | ↓
pignistic probability			P_{01}		P_{012}

In the Bayesian model, the same pattern reduces to:

		E_1		E_2	
probability	P_0	-------------> throwing the dice	P_1	-------------> Peter's alibi	P_2

The difference between the credal and pignistic levels is reminiscent of the difference between thought and action, between "inference" (how belief is affected by evidence) and "action" (which of several possible courses of action seems best) (Smith 1961, pg 1).

Which model fits "reality"? Justifications of the Bayesian model are based on betting and decision arguments through the introduction of some requirements that lead to additive measures. But at the pignistic level, we also represent Your beliefs by a probability distribution, therefore we satisfy those requirements. This does not mean that additivity also pervades the credal level. No justifications are given by Bayesians for such requirements except that they just do not distinguish between a credal and a pignistic level. Their axioms always centre on forced decisions (or preferences) but not on belief itself. They relate to observable behaviors that reflect an underlying credal state, not to the credal state itself.

To understand the difference between the two models, let us re-analyse in detail the Bayesian solution for the case of Mr. Jones. Re-consider the formula: P(killer is Mary|Peter's alibi) = $\frac{1}{1+x}$ where x = P(killer is Paul|dice = odd).

The only way to arrive at a Bayesian solution which is identical to the TBM solution is by postulating x = 1, i.e. that Big Boss will select Paul if an odd number is thrown. This case does not fit in with the available evidence (those in EC_t^Y). The case x = 1 fits in with the case 'Peter was not and could not have been the killer' whereas the case x < 1 fits with the available information: 'Peter was not the killer but could have been'.

Intuitively, the real conceptual problem is to decide if, given Peter was not the killer, the knowledge that Peter might have been the killer or not is relevant to the bet 'male versus female'. Bayesians say it is, the TBM says it is not.

Mathematically, the difference between the 2 solutions results from the necessity, in the context of probability, to split the .5 probability given to the males by the throwing of dice among the two males. Then later, the mass given to Peter cannot be given back to Paul once the alibi for Peter becomes available. Instead, in the TBM, the mass is not split, and is later transferred as a whole to Paul.

10. Conclusions.

1) An argument we encountered when comparing the respective merits of the Bayesian model and the TBM runs as follow.

Let us consider the case of Mr. Jones.
Let M = "the male chosen is always Paul"
B = "the TBM is true"
P = "the Bayesian model is true"
C = the odds on male versus female are 1 to 1 once Peter's alibi is available"

One has to be careful not to use the following deceptive reasoning:
B implies C
assumption M is necessary in order to get C
I dislike assumption M
therefore I dislike B.

The second proposition is wrong. The correct reasoning is:

B implies C

If P then assumption M is necessary in order to get C

I dislike assumption M

therefore If P then I dislike B.

which is not sufficient to conclude that "I dislike B".

2) For the case of Mr. Jones, the Bayesian approach leads to a bet on Male versus Female with the odds at 1 to 2 whereas the belief functions approach leads to a bet with the odds at 1 to 1. Which of the two is adequate is a matter of personal opinion? We feel that 1 to 1 is adequate. Others might prefer 1 to 2.

The argument that the Bayesian approach is correct because it complies with the probability theory is circular, hence useless. Description of credal states can be done by at least two normative models: the classical Bayesian and the TBM. Which of the two is correct cannot be established. It can only be tested. As Smith (1961 pg 1) stated: "beliefs are insubstantial mental processes and it is not easy to lay down generally acceptable principles according to which our belief is 'better' than another".

The interest of the "Mr. Jones" example lies in the fact that there is a case where both theories lead to different results. As a result, this particular example can be used as a test, a discriminating tool to distinguish between the two models. That it would convince the Bayesians is not sure but we hope here to have suggested some answers to Lindley's challenge (1987).

3) **In summary,** we have presented the TBM through the analysis of some paradigms and the comparison of the TBM solutions with the classical Bayesian solutions. The TBM aims at quantifying our degree of belief that a given proposition is true (where 'belief' could be renamed 'support', 'assurance', 'commitment'...). We use belief as it is the most natural word even though one could argue about the value of such a choice.

The principal assumption on which the TBM depends is the concept of 'parts of belief' supporting propositions and that due to a lack of further information cannot support a more specific proposition. We show how betting behaviors can be established by constructing pignistic probabilities, and explained why Dutch Books cannot be constructed to disprove the TBM. The semantics of our model are provided by its betting behavior. It is essentially identical to the ‘exchangeable bets’ semamtics of the Bayesians. The difference lies in the way bets are adapted when the betting frames are changed. The paradigms illustrate the model and allow us to enhance its originality in comparison to the classical probability models. The two-level structure of our belief (credal and pignistic) is detailed. The missing element of our presentation, a clear axiomatic justification of the TBM, is presented in a forthcoming paper (see also Wong 1990). The present paper concentrates on the presentation of the model rather than its foundations.

4) Uncertainty is a polymorphous phenomenon (Smets 1991a). There is a different mathematical model for each of its varieties. No single model fits all cases. The real problems when quantifying uncertainty is to recognize its nature and to select the appropriate model. The Bayesian model is only one of them. The TBM is also only one of them. Each has its own field of applicability. Neither is always better than the other (Saffiotti 1988). As Fisher once put it (1948, pg 193): '…La seule direction pratique qui nous est ouverte, est de concevoir clairement le processus intellectuel exact d'une méthode et ensuite de peser, considérer, critiquer et finalement décider si la méthode est ou non acceptable, si j'ose dire, pour notre conscience scientifique…'[5].

Appendix 1.

Proof of Theorem 3.1: Let $(\Omega, \Re, bel)$ be a credibility space and m be the bba associated to bel. Let $\mathcal{A}$ denotes the set of atoms of $\Re$. Given Assumption A1, there is a f function such that, for $x \in \mathcal{A}$, $BetP(x,m) = f(\{m(X): x \subseteq X\})$.

Suppose bel is such that there is an $A \in \Re$ with $m(A)>0$ and a pair of atoms y and z of $\Re$ with $y \neq z$ and $y \cap A = \emptyset$, $z \cap A \neq \emptyset$. Let $0 \leq \delta \leq \varepsilon \leq m(A)$.

Let m' be the bba on $\Re$ such that $m'(A) = m(A)-\varepsilon$, $m'(y) = m(y)+\varepsilon-\delta$, $m'(z) = m(z)+\delta$ and $m'(B) = m(B)$ for all $B \in \Re$, $B \neq A$, $B \neq y$, $B \neq z$.

$$
\begin{array}{llll}
\text{Let} & g(x,A,\varepsilon) & = f(m(x),...m(A)-\varepsilon,...) - f(m(x),...m(A),...) & \text{if } x \subseteq A \\
 & & = 0 & \text{otherwise.} \\
\text{Let} & h(x,y,\varepsilon-\delta) & = - f(m(y)+\varepsilon-\delta,\ldots) + f(m(y),\ldots) & \text{if } x=y \\
 & & = 0 & \text{if } x \neq y. \\
\text{Let} & h(x,z,\delta) & = - f(m(z)+\delta,\ldots) + f(m(z),\ldots) & \text{if } x=z \\
 & & = 0 & \text{if } x \neq z.
\end{array}
$$

As
$$\sum_{x \in \mathcal{A}} BetP(x,m) = \sum_{x \in \mathcal{A}} BetP(x,m') = 1,$$

[5] '...The only practical direction open to us is to conceive clearly the exact intellectual process of a method and then to weight, consider, crticize and finally decide whether or not the method is acceptable, if I dare say it, to our scientific conscience...'

then, $\sum_{x\in \mathcal{A}} (\ \mathrm{BetP}(x,m') - \mathrm{BetP}(x,m)\ ) = 0.$

Therefore, for the given $A\in\Re$,

\I\su(x∈ $\mathcal{A}$, x\S\UP1(⊆) A,, $g(x,A,\varepsilon)) = h(y,y,\varepsilon-\delta) + h(z,z,\delta)$ (3.2)

As $g(x,A,\varepsilon)$ is independent of δ for all $x\subseteq A$, $h(y,y,\varepsilon-\delta) + h(z,z,\delta)$ is also independent of δ.

Let $h(y,y,\varepsilon-\delta) + h(z,z,\delta) - f(m(y),\ldots) - f(m(z),\ldots) = H(\varepsilon)$

$K(\varepsilon-\delta) = - f(m(y)+\varepsilon-\delta,\ldots)$

$L(\delta) = - f(m(z)+\delta,\ldots)$

The relation

$$h(y,y,\varepsilon-\delta) + h(z,z,\delta) = - f(m(y)+\varepsilon-\delta,\ldots) + f(m(y),\ldots) - f(m(z)+\delta,\ldots) + f(m(z),\ldots)$$

can be written as:

$$H(\varepsilon) = K(\varepsilon-\delta) + L(\delta).$$

It is a Pexider's equation whose solutions for H, K and L, given Assumption A2, are linear in their argument. (Aczel, 1966, theorem 1, pg 142).

Hence $f(m(z), ...) = \alpha + \beta\, m(z)$

where α and β may depend on the bbm given to the strict supersets of z.

The important point up to here is that both $h(y,y,\varepsilon-\delta) + h(z,z,\delta)$ is linear in ε and does not depend on δ. Let $h(y,y,\varepsilon-\delta) + h(z,z,\delta) = c\varepsilon + d$

The proof that $g(x,A,\varepsilon)$ in (3.2) is linear in all its arguments m(.) is based on the following procedure given in case A is the union of four atoms y_1, y_2, y_3, y_4.

Let m(A) =a.

For i, j, k = 1, 2, 3, 4, $i\neq j\neq k\neq i$, let $x_i = \{m(y_i\cup B)\cup B: B\subseteq\overline{A}\}$,

$x_{ij} = \{m(y_i\cup y_j\cup B)\cup B: B\subseteq\overline{A}\}$, $x_{ijk} = \{m(y_i\cup y_j\cup y_k\cup B)\cup B: B\subseteq\overline{A}\}$, and

$x_{1234} = \{m(y_1\cup y_2\cup y_3\cup y_4\cup B): B\subseteq\overline{A}, B\neq\emptyset\}$.

Then (3.2) becomes: (the m(A) term is put as first element of f, and is not included in x_{1234})

$$f(a-\varepsilon, x_1, x_{12}, x_{13}, x_{14}, x_{123}, x_{124}, x_{134}, x_{1234})$$
$$- f(a, x_1, x_{12}, x_{13}, x_{14}, x_{123}, x_{124}, x_{134}, x_{1234}) +$$
$$f(a-\varepsilon\ x_2, x_{12}, x_{23}, x_{24}, x_{123}, x_{124}, x_{234}, x_{1234})$$
$$- f(a, x_2, x_{12}, x_{23}, x_{24}, x_{123}, x_{124}, x_{234}, x_{1234}) +$$
$$f(a-\varepsilon, x_3, x_{13}, x_{23}, x_{34}, x_{123}, x_{134}, x_{234}, x_{1234})$$
$$- f(a, x_3, x_{13}, x_{23}, x_{34}, x_{123}, x_{134}, x_{234}, x_{1234}) +$$
$$f(a-\varepsilon, x_4, x_{14}, x_{24}, x_{34}, x_{124}, x_{134}, x_{234}, x_{1234})$$
$$- f(a, x_4, x_{14}, x_{24}, x_{34}, x_{124}, x_{134}, x_{234}, x_{1234}) = c\varepsilon + d$$

Let $x_i = u$, $x_{ij} = v$, $x_{ijk} = w$, $x_{1234} = t$ for all i, j, k = 1, 2, 3, 4. One gets:
$4 (f(a-\varepsilon, u, v, v, v, w, w, w, t) - f(a, u, v, v, v, w, w, w, t)) = c\varepsilon + d$
hence f(a, u, v, v, v, w, w, w, t) is linear in a.

Keep all equalities as before except for $x_{123} \neq w$, then:
$$3 (f(a-\varepsilon, u, v, v, v, x_{123}, w, w, t) - f(a, u, v, v, v, x_{123}, w, w, t))$$
$$+ (f(a-\varepsilon, u, v, v, v, w, w, w, t) - f(a, u, v, v, v, w, w, w, t)) = c\varepsilon + d$$
The second term in the LHS is linear in a, so the first term in the LHS is also linear in a.

Suppose now $x_{124} \neq w$, then
$$2 (f(a-\varepsilon, u, v, v, v, x_{123}, x_{124}, w, t) - f(a, u, v, v, v, x_{123}, x_{124}, w, t))$$
$$+ (f(a-\varepsilon, u, v, v, v, x_{124}, w, w, t) - f(a, u, v, v, v, x_{124}, w, w, t))$$
$$+ (f(a-\varepsilon, u, v, v, v, w, w, w, t) - f(a, u, v, v, v, w, w, w, t)) = c\varepsilon + d$$
The second and third terms in the LHS are linear in a, hence so is the first. Therefore f is linear in a whatever the x_{ijk} terms and x_{1234}. We drop them in the following relations about f.

Suppose $x_{12} \neq v$, then
$$2 (f(a-\varepsilon, u, x_{12}, v, v) - f(a, u, x_{12}, v, v))$$
$$+ 2 (f(a-\varepsilon, u, v, v, v) - f(a, u, v, v, v)) = c\varepsilon + d$$
The second term in the LHS is linear in a, hence so is the first.

Suppose $x_{13} \neq v$, then
$$(f(a-\varepsilon, u, x_{12}, x_{13}, v) - f(a, u, x_{12}, x_{13}, v))$$
$$+ (f(a-\varepsilon, u, x_{12}, v, v) - f(a, u, x_{12}, v, v))$$
$$+ (f(a-\varepsilon, u, x_{13}, v, v) - f(a, u, x_{13}, v, v))$$
$$+ (f(a-\varepsilon, u, v, v, v) - f(a, u, v, v, v)) \quad = c\varepsilon + d$$
The second, third and fourth terms in the LHS are linear in a, hence so is the first.
Therefore f is linear in its first argument a whatever its other arguments.

The general proof of the linearity of f in its arguments m(.) is obtained by tediously generalizing this reasoning for any A. Let n_X be the number of atoms of $\mathcal{A}$ in $X \in \Re$. The proof is valid if $n_A < n_\Omega - 1$, as we need at least two atoms of $\Re$ not in the set A used in the derivation.

Let $F = \{ X: x \subseteq X \}$. The general solution can be written as:
$$f(\{m(X): x \subseteq X\}) = \sum_{G \subseteq F} \beta(G) \prod_{Y \in G} m(Y) \tag{3.3}$$

where the $\beta(G)$ might depend on the $m(Y)$ with $n_Y \geq n_\Omega - 1$.

Suppose a belief function bel with $m(X) = 1\text{-}\omega$ and $m(\Omega) = \omega$ for $X \in \Re$ and $n_X < n_\Omega - 1$. Then for all atoms x in X,

$$BetP(x,m) = \beta(\{\}) + \beta(\{X\})\,(1\text{-}\omega) + \beta(\{\Omega\})\,\omega + \beta(\{X,\Omega\})\,\omega\,(1\text{-}\omega)$$

and for the y not atom of X

$$BetP(y,m) = \beta(\{\}) + \beta(\{\Omega\})\,\omega$$

By adding these terms on the atoms of $\Re$, one gets:

$$1 = n_\Omega\,\beta\{\}) + n_X\,\beta(\{X\})\,(1\text{-}\omega) + n_\Omega\,\beta(\{\Omega\})\,\omega + n_X\,\beta(\{X,\Omega\})\,\omega\,(1\text{-}\omega)$$

This being true for all ω in [0,1], the coefficients of the terms in ω and ω^2 must be nul. So $\beta(\{X,\Omega\}) = 0$, $\quad n_\Omega\,\beta(\{\Omega\}) = n_X\,\beta(\{X\})$, $\quad$ and $\quad 1 = n_\Omega\,\beta(\{\}) + n_X\,\beta(\{X\})$.

The same argument can be repeated in order to show that every coefficients $\beta(G) = 0$ whenever there are more than one elements in G. Relation (3.3) becomes:

$$BetP(x,m) = \beta(\{\}) + \sum_{x \subseteq X} \beta(\{X\})\,m(X)$$

where the β may depend on the bbm given to those elements of $\Re$ with n_Ω-1 or n_Ω atoms. We show that β does not depend on those bbm.

Let $(\Omega, \Re, bel)$ be a credibility space. Suppose it is known (by You) that the actual world ϖ is not an element of the set B that contains two atoms b_1 and b_2 of $\Re$. So $\forall A \in \Re$, $A \cong A \cup X$ where $X = b_1$, b_2 or $b_1 \cup b_2$, and $bel(A) = bel(A \cup X)$ by the consistency axiom. Let $(\Omega', \Re', bel')$ be the credibility space where $\Omega' = \Omega\text{-}B$, the set $\mathcal{A}'$ of atoms of $\Re'$ equal to $\mathcal{A}\text{-}(b_1 \cup b_2)$, and $bel'(A) = bel(A)$ for all $A \in \Re'$. By construction, for all $Y \in \Re$, $n_Y \geq n_\Omega\text{-}1$, $m(Y)=0$. Let $BetP(x,m)$ and $BetP'(x,m')$, x atom of $\Re'$, be the pignistic probabilities derived from bel (m) and bel' (m'). The bbm involved in the coefficients β are explicitely written. One has:

$$\begin{aligned} BetP(x,m) = {} & \beta(\{\}, \{m(Y): n_Y = n_\Omega\text{-}1\}, m(\Omega)) + \\ & \sum_{x \subseteq X \in \Re} \beta(\{X\}, \{m(Y): n_Y = n_\Omega\text{-}1\}, m(\Omega))\, m(X) \\ = {} & \beta(\{\}, \{0: n_Y = n_\Omega\text{-}1\}, 0) + \\ & \sum_{x \subseteq X \in \Re'} \beta(\{X\}, \{0: n_Y = n_\Omega\text{-}1\}, 0)\, m(X) \end{aligned}$$

where all terms $m(X) = 0$ if $n_X \geq n_\Omega\text{-}1$, or equivalently $m(X) = 0$ if $X \notin \Re'$.

One has also:

$$BetP'(x,m') = \beta(\{\}, \{m(Z): n_Z = n_{\Omega'}\text{-}1\}, m(\Omega')) +$$

$$\sum_{x \subseteq X \in \Re'} \beta(\{X\}, \{m(Z): n_Z = n_{\Omega'}-1\}, m(\Omega')) \, m(X)$$

By Assumption A4, BetP(x,m) = BetP' (x,m') for all x atom of $\Re'$. Hence:

$$\beta(\{X\}, \{m(Z): n_Z = n_{\Omega'}-1\}, m(\Omega')) = \beta(\{X\}, \{0: n_Y = n_{\Omega}-1\}, 0)$$

so β does not depend on the bbm $m(Z): n_Z \geq n_{\Omega'}-1$.

Furthermore, by Assumption A3, the coefficients β depends only on the number of atoms in their arguments. Hence:

$$BetP(x,m) = 1/n_{\Omega} - \beta(\{\Omega\}) + n_{\Omega}\, \beta(\{\Omega\}) \sum_{x \subseteq X} m(X) / n_X$$

Let $m(A) = 1$ for $A \in \Re$ and $n_A < n_{\Omega}-1$. By Assumption A3, $BetP(x,m) = 1/n_A$ for every atom x subset of A. It implies: $\beta(\{\Omega\}) = 1/n_{\Omega}$.

Hence: $$BetP(x,m) = \sum_{x \subseteq X} m(X) / n_X$$ QED

Bibliography.

ACZEL J. (1966) Lectures on functional equations and their applications. Academic Press, New York, NY.

BLACK P.K. (1987) Is Shafer general Bayes? in AAAI Third Worskshop on Uncertainty in Artificial Intelligence, Seattle, 2-9.

CLARKE M.R.B., FROIDEVAUX C. GREGOIRE E. and SMETS Ph. (1991) Guest Editors of the Special Issue on 'Uncertainty, Conditional and Non Monotonicity. Positions and Debates in Non-Standard Logics.'. J. Applied Non-Classical Logics. Vol.1, N° 2.pg. 103-310.

COX R.T. (1946) Probability, frequency and reasonable expectation. Amer.J.Phys. 14:1-13.

DEGROOT M.H. (1970) Optimal statistical decisions. McGraw-Hill, New York.

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. (1982) On several representations of an uncertain body of evidence. in GUPTA M.M. and SANCHEZ E. eds, Fuzzy information and decision processes. North Holland, Amsterdam. p 167-181.

DUBOIS D. and PRADE H. (1991) Focusing versus updating in belief function theory. In : Internal Report IRIT/91-94/R, IRIT, Univ. P. Sabatier, Toulouse, France.

EKELOF P.O. (1982) Rättegång IV. Fifth edition, Stockholm.

FISHER R.A. (1948) Conclusions fiduciaires. Annales de l'Institut Henri Poincarré, 10:191-213.

GARDENFORS P., HANSSON B. and SAHLIN N.E. (1983) Evidentiary value: philosophical, judicial and psychological aspects of a theory. C.W.K. Gleerups, Lund, Sweden.

GILES R. (1982) Foundation for a possibility theory. in Fuzzy Information and Decision Processes, GUPTA M.M. and SANCHEZ E. eds. North Holland, Amsterdam, 183-195.

HACKING I. (1965) Logic of statistical inference. Cambridge University Press, Cambridge, U.K.

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

HALPERN J.Y. and FAGIN R. (1992) Two views of belief: belief as generalized probability and belief as evidence. Artificial Intelligence 54: 275-318.

HSIA Y.-T. (1991) Characterizing Belief with Minimum Commitment. IJCAI-91:1184-1189.

HUNTER D. (1987) Dempster-Shafer versus Probabilistic Logic. in AAAI Third Worskshop on Uncertainty in Artificial Intelligence, Seattle, 22-29.

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

KLAWONN F. and SCHWECKE E. (1990) On the axiomatic justification of Dempster's rule of combination. Int. J. Intel. Systems 7:469-478.
KLAWONN F. and SMETS P. (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.
KYBURG H.E.Jr. (1987a) Objectives probabilities. Intern.Joint Confer.Artificial Intelligence, IJCAI-87, 902-904.
KYBURG H.E.Jr. (1987b) Bayesian and non-Bayesian evidential updating. Artificial Intelligence, 31:271-294.
LASKEY K.B. (1987) Beliefs in belief functions: an examination of Shafer's canonical examples. in AAAI Third Worskshop on Uncertainty in Artificial Intelligence, Seattle, 39-46.
LEVI I. (1983) Consonance, dissonance end evidentiary mechanisms. in GARDENFORS P., HANSSON B. and SAHLIN N.E. (1983) p. 27-43.
LINDLEY D.V. (1987) The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. Statistical Sc. 2, 17-24.
NGUYEN H.T. (1978) On random sets and belief functions J. Math.Anal. Applic. 65:531-542.
NGUYEN H.T. and SMETS P. (1991) On Dynamics of Cautious Belief and Conditional Objects. Intern. J. Approx. Reasoning. Forthcoming.
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.
RAMSEY F.P. (1931) Truth and probability. in Studies in subjective probability, eds. KYBURG H.E. and SMOKLER H.E., p. 61-92. Wiley, New York.
RUSPINI E.H. (1986) The logical foundations of evidential reasoning. Technical note 408, SRI International, Menlo Park, Ca.
SAFFIOTTI A. (1988) An AI view of the treatment of uncertainty. Knowledge Engineering Revue 2:75-98.
SHAFER G. (1976) A mathematical theory of evidence. Princeton Univ. Press. Princeton, NJ.
SHAFER G. and TVERSKY A. (1985) Languages and designs for probability. Cognitive Sc. 9:309-339.
SHAFER G., SHENOY P.P. and MELLOULI K. (1987) Propagating belief functions in qualitative Markov trees. Int. J. Approx. Reasoning, 1:349-400.
SHAFER G. (1990) Perspectives in the theory and practice of belief functions. Intern. J. Approx. Reasoning, 4:323-362.
SMETS Ph. (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 Ph. (1987) Upper and lower probability functions versus belief functions. Proc. International Symposium on Fuzzy Systems and Knowledge Engineering, Guangzhou, China, July 10-16, pg 17-21.
SMETS Ph. (1988) Belief functions. in SMETS Ph, MAMDANI A., DUBOIS D. and PRADE H. ed. Non standard logics for automated reasoning. Academic Press, London p 253-286.
SMETS Ph. (1990a) The combination of evidence in the transferable belief model. IEEE-Pattern analysis and Machine Intelligence, 12:447-458.
SMETS Ph. (1990b) The transferable belief model and possibility theory. Proc. NAFIPS-90, pg. 215-218.
SMETS Ph. (1990c) Construucting 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. (1990d) The Transferable Belief Model and Other Interpretations of Dempster-Shafer's Model. Procs of the 6th Conf.on Uncertainty in AI, Cambridge, MA.
SMETS Ph. (1991a) Varieties of ignorance. Information Sciences. 57-58:135-144.
SMETS Ph. (1991b) Axioms for belief functions. under preparation.
SMETS P. (1991c) Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. To be published in Int. J. Approximate Reasoning.
SMETS Ph. (1992a) 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 Ph. (1992b) The transferable belief model and random sets. Int. J. Intell. Systems 7:37-46.

SMETS Ph. (1992c) Resolving misunderstandings about belief functions: A response to the many criticisms raised by J. Pearl. Intern. J. Approx. Reasoning 6: 321-344.
SMITH C.A.B. (1961) Consistency in statistical inference and decision. J. Roy. Statist. Soc. B23:1-37.
WILLIAMS P.M. (1982) discussion in SHAFER G. Belief functions and parametric models. J. Roy. Statist. Soc. B44: 322-352.
WILSON N. (1991) The Representation of Prior Knowledge in a Dempster-Shafer Approach. TR/Drums Conference, Blanes
WONG S.K.M., YAO Y.Y., BOLLMANN P. and BÜRGER H.C. (1990) Axiomatization of qualitative belief structure. submitted for publication.

Acknowledgements.

The authors are indebted to Ph. Besnard, M. Clarke, D. Dubois, P. Garbolino, H. Kyburg, B. Marchal, V. Poznanski, H. Prade, E. Ruspini, T. Seidenfelt, D. Spiegelhalter, L. Simar and N. Wilson for their useful and most stimulating comments, especially when they disagree.