What is Dempster-Shafer's model?

Philippe SMETS[1]
IRIDIA, Université Libre de Bruxelles
50 av. F. Roosevelt, CP194/6,
1050 Brussels, Belgium.

Abstract: Several mathematical models have been proposed for the modelling of someone's degrees of belief. The oldest is the Bayesian model that uses probability functions. The upper and lower probabilities (ULP) model, Dempster's model, the evidentiary value model (EVM) and the probability of modal propositions somehow generalize the Bayesian approach. The transferable belief model (TBM) is based on other premises and uses belief functions. None of these models is THE best: each has its own domain of application. We spell out through examples what are the underlying hypotheses that lead to the selection of an adequate model for a given problem. We give indications on how to choose the appropriate model. The major discriminating criterion is: if there exists a probability measure with known values, use the Bayesian model, if there exists a probability measure but with some unknown values, use the ULP models, if the existence of a probability measure is not known, use the TBM. Dempster's model is essentially a special case of ULP model. The EVM and the probability of modal propositions (provability, necessity...) corresponds to a special use of the Bayesian model.

Keywords: belief functions, Dempster-Shafer theory, belief modelling, upper and lower probabilities, transferable belief model.

1. Introduction

1.1. Modelling beliefs

I do not know what is Dempster-Shafer's model[2], except that it is a model that uses the mathematical object called 'belief function'. Usually its aim is in the *modelling of*

---

[1] The following text presents research results of the Belgian National incentive-program for fundamental research in artificial intelligence initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. Scientific responsibility is assumed by its author.
A preliminary and much shorter version appeared in Smets (1990).

[2] By model, we simply mean a mathematical representation (idealization) of some process, the subjective, personal beliefs here. It should not be confused with the concept of model encountered in logic.

someone's degrees of belief. But there are so many interpretations of Dempster-Shafer's theory in the literature that it seems useful to present the various contenders in order to clarify their respective meanings and domains of applicability.

Beliefs results from *uncertainty* and the interpretations we present correspond to different forms of uncertainty. The uncertainty encountered here concerns the one usually quantified by probability functions. We do not tackle the problem of vagueness and ambiguity studied in fuzzy sets theory and possibility theory. Here, the uncertainty sometimes results from a random process (the objective probability case), it sometimes results only from the lack of information that induces some 'belief' (instead of some 'knowledge'). The domain of application of the various models we study here is the same as the one studied by the Bayesian, but in some cases like the transferable belief model we propose solutions different from those defended by the Bayesians.

We will consider successively the classical probability model, the upper and lower probabilities (ULP) model, Dempster's model, the transferable belief model (TBM), the evidentiary value model (EVM), the provability / necessity model. Each model corresponds to a different form of uncertainty or to the introduction of some particular extra assumptions.

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

Both Dempster and Shafer introduced models but I do not think they would recognise their model as a 'Dempster-Shafer' model. Dempster's seminal work was not orientated toward the modelization of someone's beliefs. The idea of Shafer was to use the mathematics introduced by Dempster in order to modelize someone's belief (Shafer 1976a). We share this idea, but consider that Shafer did not justify his model in his book, and later justifications were too much based on random sets, one-to-many mappings and upper and lower probabilities - all approaches open to criticism, as shown in this paper. So we introduced the concept of the transferable belief model (TBM) in order to justify the use of belief functions (including Dempster's rule of conditioning and Dempster's rule of combination) to modelize someone's belief (Smets 1988). This papers focuses more on a presentation of the TBM and a criticism of most of the justifications provided in the literature. The TBM was in fact developed to meet these criticisms. Justifications of the use of belief functions is in Smets and Kennes (1990). Justification of Dempster's rule of conditioning is in Smets (1991a). Justification of Dempster's rule of combination is in Smets (1990a), Klawonn and Schwecke (1990).

The problem with Dempster's model lies in the fact that it sets out to assess 'beliefs' on some space Y on which the existence of a *probability measure is acknowledged, but not precisely known* in that the probability is known for some of its subsets, not for all of them. The updating of this 'partially defined' probability measure on Y by conditionalisation cannot always be performed, but the solution proposed in Dempster's model (the D-conditioning in section 5) has been criticized for violating the basic rules of probability theory (Levi, 1983).

In the TBM, such probability measure on Y is not claimed. If there is such a probability measure, the TBM is reduced to the classic probabilistic model. *But the TBM can be described without assuming such a probability measure on Y.* And if such a probability measure on Y is not assumed, the constraints of the probability theory - because of its irrelevance - can no longer be brought into play to criticize the updated beliefs derived on Y after applying Dempster's rule of conditioning.

This explains why we described the TBM as a *'purified' model*, 'purified' inasmuch as a connection with probability concepts is not necessary. In the TBM, we use the available information but no more. We do not assume the existence of a probability measure on every space as do the Bayesians. We keep only those probabilities that are justified. The obvious interest of the TBM is that it can be applied even in contexts where every probability concept is absent. And when probabilities are defined everywhere, the TBM is reduced to the Bayesian model.

We study a paradigm, the unreliable sensor. We apply each model to it and show what the underlying hypotheses are that justify the use of each model. For each model, we successively present a short summary of the underlying theory followed by an analysis of the paradigm. For a real problem, the user will have to select the appropriate model, and that selection should be governed by the explicit acknowledgment of the underlying hypothesis. So in each context, there will be only one satisfactory analysis. There is no satisfactory general method. Each case requires its own method. What that method might be, depends on the underlying hypothesis.

1.2. The static and dynamic components

Any model for belief has, at least, two components: one, static, that describes our state of belief and the other, dynamic, that explains how to update our belief given new pieces of information. We insist on the fact that both components must be considered in order to compare these models. Unfortunately, too many publications are restricted to the static component and fail to detect the differences between these models. In fact, the originality of the models based on belief functions lies essentially in their dynamic component.

The difference between the various models we have studied can be described through the Moebius transform of the various measure of belief presented here. Let Ω be a finite set Ω with its power set denoted by 2^{Ω}. Given a function F: $2^{\Omega} \rightarrow [0, 1]$, its **Möbius transform** m: $2^{\Omega} \rightarrow [0, 1]$ is:

$$m(A) = \sum_{X \subseteq A} (-1)^{|A|-|X|} F(X)$$

and

$$F(A) = \sum_{X \subseteq A} m(X)$$

where $|X|$ is the number of elements in X. When F is a probability function P, m(A) is null whenever $|A| \neq 1$ and $m(\omega) = P(\{\omega\}), \forall \omega \in \Omega$. When F is a belief function bel (or a lower probability function P_* in Dempster's model), then $m(A) \geq 0$ for all $A \subseteq \Omega$, $m(\emptyset) = 0$[1] and m is called a basic belief assignment, and the m's values are called the basic belief masses.

2. The breakable sensor

Suppose I am a new technician and must check the temperature of a process. To do this, I have a sensor that can check the temperature of the process - a temperature that can be only hot or cold. If the temperature is hot (TH), the sensor light is red (R) and if the temperature is cold (TC), the sensor light is blue (B). The sensor is made of a thermometer and a device that turns on the blue-red light according to the temperature reading. Unfortunately, the thermometer may well be broken.

The only known information is what is written on the box containing the sensor. "Warning: the thermometer included in this sensor may 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, red light means the temperature is hot, blue light means the temperature is cold. When the thermometer is broken, the sensor answer is unrelated to the temperature".

I am a new technician and have never seen this sensor before. I know nothing about it except the warning written on the box. I use it and the blue light goes on. How do I assess the temperature status? What is my opinion (belief) that the temperature status is hot or cold?

Before constructing any quantified description of our personal belief with respect to the temperature status, one must build a frame of discernment Ω (also called the universe of discourse or state space) on which beliefs will be allocated and updated. Its structure is

---

[1] We accept the closed world assumption (as in Shafer's work). Under closed world assumption $m(\emptyset)=0$. Under open world assumption, $m(\emptyset)$ may be positivie, and bel(A) = \i\su(Ø≠X\S\UP1(⊆) A;; m(X)), in which case $bel(\Omega) = 1 - m(\emptyset)$. For the difference betwen the open and closed world assumptions, see Smets (1988)

a finite boolean algebra of propositions or of sets, which here is equivalent. $\Omega = S x T x \Theta$, the cartesian product of spaces S, T, Θ where:
- S = {B, R}, the sensor status, Blue or Red,
- T = {TH, TC}, the temperature status, Hot or Cold
- Θ = {ThW, ThB}, the thermometer status, Working or Broken.

The eight elements of the space Ω are detailed in table 1.

Table 1. The labels of the eight elements of Ω.

	B		R	
	TH	TC	TH	TC
ThW	a	b	c	d
ThB	e	f	g	h

3. The probability model

3.1. The model

In probability theory, the static component consists of the assessment of a probability density p on the elements of Ω such that p: $\Omega \to [0, 1]$, $\sum_{\omega \in \Omega} p(\omega) = 1$.

Degrees of belief on subsets of Ω are quantified by a probability distribution P: $2^{\Omega} \to [0, 1]$ such that $\forall \omega \in \Omega$, $P(\{\omega\}) = p(\omega)$ and $\forall A, B \subseteq \Omega$ with $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$ and $P(A) = \sum_{\omega \in A} p(\omega)$.

The only dynamic component is the conditioning rule: when you learn that $B \subseteq \Omega$ is true (and if $P(B) \neq 0$), P is updated into the conditional probability distribution P(.|B) defined on 2^{Ω} such that $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

3.2 The probabilist analysis

A probabilist will build a probability distribution P on 2^{Ω}. The information on the box induces the constraints:

$$P(ThW) = P(\{a, b, c, d\}) = p(a) + p(b) + p(c) + p(d) = 0.8$$
$$P(ThB) = P(\{e, f, g, h\}) = p(e) + p(f) + p(g) + p(h) = 0.2$$

When the sensor is working (ThW), the sensor is red (R) when the temperature is hot (TH) and blue (B) otherwise (TC). So

$$p(a) = p(d) = 0$$

When the sensor is broken (ThB), the sensor status (B or R) is unrelated to the temperature status. It translates as P(B|ThB, TH) = P(B|ThB, TC), i.e.:

$$\frac{p(e)}{p(e) + p(g)} = \frac{p(f)}{p(f) + p(h)}.$$

Let $x = P(B|ThB) = \frac{p(e) + p(f)}{p(e) + p(f) + p(g) + p(h)}$ denotes the probability that the sensor is blue when the thermometer is broken and $\pi = P(TC) = p(b) + p(f) + p(h)$ denotes the a priori probability that the temperature is cold i.e. before we perform our measurement. Finally the status of the thermometer (ThB or ThW) is unrelated to the temperature (TH or TC): P(ThW|TH) = P(ThW), i.e. $\frac{p(a) + p(c)}{p(a) + p(c) + p(e) + p(g)} = 0.8$.

Table 2: Probability distribution on Ω = SxTxΘ

	B		R	
	TH	TC	TH	TC
ThW	0	.8 π	.8 (1-π)	0
ThB	.2 (1-π) x	.2 π x	.2 (1-π) (1-x)	.2 π (1-x)

Table 2 presents the probability distribution p on Ω. The set of constraints is not sufficient to define uniquely π and x. There are several ways to describe the missing information, those used simplifies the discussion as π and x have natural connotations. We must compute

$$P(TC|B) = \frac{p(b) + p(f)}{p(a) + p(b) + p(e) + p(f)} = \frac{(.8 + 0.2\ x\ )\ \pi}{.8\pi + 0.2x}$$

Even if we knew π, what value should be given to x, the probability that the sensor status is blue when the thermometer is broken? Nothing in the available evidence tells us what value to give to x. A probabilist facing such a problem can follow several approaches.

- He can try to collect data about x … If he could, then his results would be the same as those obtained with the various alternative models analysed hereafter. The problem we study is the one where no further data can be collected about x.
- He can propose extraneous assumptions like:
  - the principle of insufficient reason: when the thermometer is broken, the sensor can only be blue or red. Knowing nothing more, I postulate that both options have the same probability, hence x = .5
  - a maximum entropy argument that leads to the same result

- a meta-probability that describes his belief about the possible values of x … but how can one justify where the meta-probability comes from? Remember all the technician knows is the warning on the box containing the sensor.

So to perform a strict probability analysis, you must introduce some extraneous assumption. The value of your results will only reflect the value of this assumption. The classical probability approach therefore fails in such states of ignorance. This failure explains why other models have been proposed.

4. Upper and lower probabilities models

4.1. The models

The upper and lower probabilities model is identical to the probability model except that it acknowledges that some probabilities might be unknown. Let the set Π of all those probability distributions compatible with the available information. Instead of building a meta-probability distribution on Π as strict bayesians would recommend, one considers critical values - usually the extremes - of the various probabilities one is interested in. Various forms of partly known probability models can be described. Sometime Π is uniquely defined through its so-called upper and lower probabilities functions P^* and P_* where

$$\forall A \subseteq \Omega \qquad P^*(A) = \sup_{P \in \Pi} P(A) \qquad P_*(A) = \inf_{P \in \Pi} P(A)$$

In other cases, Π is a convex set of probability distributions (Kyburg 1987). Sometimes Π is just a set of probability distributions, as in the case where we have a double head or double tail coin, but we do not know which, hence P(Head) is either 0 or 1 and Π has only two elements. When Π contains only one element, the model is reduced to the classic probability model.

In the usual upper and lower probabilities model, the static part consists of defining the upper probability distribution P^* or the lower probability distribution P_*, both from 2^Ω to [0, 1]. For each P in Π, one has:

$$\forall A \subseteq \Omega, \quad P_*(A) \leq P(A) \leq P^*(A)$$

By construction, $P^*(A) = 1 - P_*(\overline{A})$

Conditioning on $B \subseteq \Omega$ is obtained by considering each probability distribution P in Π, and conditioning them on B. Let Π_B be the resulting set of conditional probability distributions:

$$\Pi_B = \{ P_B : \forall A \subseteq \Omega \; P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, P \in \Pi \}$$

The upper and lower conditional probabilities functions are the upper and lower limits of these conditional probabilities.

$\forall A \subseteq \Omega,$ $\quad P_*(A|B) = \inf_{P_B \in \Pi_B} P_B(A) = \inf_{P \in \Pi} P(A|B)$

$$P^*(A|B) = \sup_{P_B \in \Pi_B} P_B(A) = \sup_{P \in \Pi} P(A|B)$$

4.2. The upper and lower probabilities analysis.

In the probability approach, the missing information is π and x. They are only known to be in $[0, 1]^2$. For each pair (π, x) corresponds an element P of Π. We will also consider the case where π is known, in which case the set of compatible probability distributions is Π' where each value of x in [0,1] defines an element P of Π'. (The ' notation will distinguish the two cases). These constraints describe the static part of the model.

The dynamic part is applied once the blue light (B) is observed. Suppose first that π and x are both unknown. Conditioning on B results in:

$$P_*(TC|B) = \inf_{P \in \Pi} P(TC|B) = \inf_{(\pi, x) \in [0, 1]x[0, 1]} P(TC|B) = 0$$

$$P^*(TC|B) = \sup_{P \in \Pi} P(TC|B) = \sup_{(\pi, x) \in [0, 1]x[0, 1]} P(TC|B) = 1$$

Should π be known, then

$$P'_*(TC|B) = \inf_{P \in \Pi'} P(TC|B) = \inf_{x \in [0, 1]} P(TC|B) = \frac{\pi}{.8\,\pi + 0.2}$$

$$P'^*(TC|B) = \sup_{P \in \Pi'} P(TC|B) = \sup_{x \in [0, 1]} P(TC|B) = 1$$

So if the technician is ignorant of π, he is left in a state of total ignorance even after observing B. If he knows π, he can compute informative limits and try to make a decision given this partial knowledge about P(TC|B), a problem still open to discussion.

5. Dempster's model.

5.1. The model.

Dempster (1967) introduced a special form of upper and lower probabilities model. For the static part of the model, he considers a space X endowed with a probability measure P_X and a mapping M from space X to space 2^Y. Let M(x) denote the image of x under M for $x \in X$. He defines upper and lower probabilities measures P^* and P_* on Y such that $\forall A \subseteq Y$:

$$P_*(A) = P_X(M_*(A))$$

and

$$P^*(A) = P_X(M^*(A))$$

where

$$M_*(A) = \{x: x \in X, M(x) \subseteq A, M(x) \neq \emptyset\}$$

and

$$M^*(A) = \{x: x \in X, M(x) \cap A \neq \emptyset\}.$$

The functions P_* and P^* are a belief function and plausibility function respectively (see section 6).

The upper and lower probabilities nature of Dempster's solution can be shown. One has:

$$\forall A \subseteq Y \qquad \inf P_Y(A) = \inf \sum_{x \in X} P_Y(A|x).P_X(x)$$

where the inf is taken over all possible values of the $P_Y(A|x)$. Index of the P functions indicate the domain of P. One has $P_Y(A|x) = 1$ if $M(x) \subseteq A$ and anything in [0,1] otherwise. The minimum is obtained by taking $P_Y(A|x) = 0$ whenever possible. Hence:

$$\inf P_Y(A) = \sum_{x:M(x) \subseteq A} P_X(x) = P_X(M_*(A)) = P_*(A).$$

The Möbius transform m of the lower probability distribution P_* is non negative: for each $x \in X$, $m(M(x)) = P_X(\{x\})$, and all other m on 2^Y are null.

For the dynamic part of the model, two types of conditioning can be considered.

$$P_*(A|B) = \inf P_Y(A|B) = \inf \frac{\sum_{x \in X} P_Y(A \cap B|x).P_X(x)}{\sum_{x \in X} P_Y(B|x).P_X(x)}$$

$$= \frac{P_*(A \cap B)}{P_*(A \cap B) + P^*(\overline{A} \cap B)}$$

where the inf is taken over all possible values of the $P_Y(A|x)$. This conditioning, called hereafter the G-conditioning, corresponds to the solution described in the upper and lower probabilities model in section 4. It is not the one considered by Dempster.

Instead he considers that conditioning on $B \subseteq Y$ means that the mapping $M:X \to 2^Y$ has been transformed into mapping $M_B:X \to 2^Y$ with:

$$M_B(x) = M(x) \cap B.$$

It corresponds to the idea that the conditioning information indicates that the set of images of each $x \in X$ is $M(x) \cap B$. Dempster postulates also that the knowledge of the conditioning event B does not modify P_X, i.e. $P_X(x|B) = P_X(x)$, a requirement open to criticism (see section 5.3). In that case,

$$P_*(A|B) = \inf \sum_{x \in X} P_{Y|B}(A|x).P_X(x)$$

$$P^*(A|B) = \sup \sum_{x \in X} P_{Y|B}(A|x).P_X(x)$$

where $P_{Y|B}(A|x) = 1$ if $M(x) \cap B \subseteq A$ and anything in [0, 1] otherwise. Then

$$P_*(A|B) = \frac{P_*(A \cup \overline{B}) - P_*(\overline{B})}{1 - P_*(\overline{B})} \qquad P^*(A|B) = \frac{P^*(A \cap B)}{P^*(B)}$$

what we call the D-conditioning. Dempster's model is defined as the one endowed with the D-conditioning rule.

Apart from the D-conditioning, Dempster also defines the so-called Dempster's rule of combination, a rule to combine the beliefs induced by two distinct pieces of information. It does not have an immediate counterpart in probability theory. It is not necessary for our presentation, so we shall not insist on it (see Shafer 1986 , Smets 1990a).

5.2. Dempster's analysis

The static part of Dempster's model is perfectly acceptable by any probabilist. The constraints induce the same solutions as described with the upper and lower probabilities analysis in section 4. The X and Y spaces correspond to the Θ and the Ω spaces respectively. The mapping M is such that M(ThW) = {a, b, c ,d} and M(ThB) = {e, f, g, h}. Finally, the probabilities defined on Θ induce the basic belief masses m on 2^{Ω}:

$$m(\{b, c\}) = 0.8 \qquad \text{and} \qquad m(\{e, f, g, h\}) = 0.2$$

Suppose π is unknown. Once B is known to have occurred (i.e. {c, d, g, h} is impossible) one adapts the mapping M into M_B, where M_B(ThW) = {a, b} and M_B(ThB) = {e, f}. Then

$$P_*(TC|B) = P_*(\{b, d, f, h\}|\{a, b, e, f\}) =$$

$$= \frac{P_*(\{b, c, d, f, g, h\}) - P_*(\{c, d, g, h\})}{1 - P_*(\{c, d, g, h\})} = \frac{0.8 - 0}{1 - 0} = 0.8$$

and

$$P^*(TC|B) = \frac{P^*(\{b, f\})}{P^*(\{a, b, e, f\})} = \frac{1}{1} = 1$$

Suppose π is known. The X domain is now the space $T \times \Theta$ and the Y domain is still Ω. The induced basic belief masses m on 2^{Ω} are given in table 3.

Then $$P'_*(TC|B) = \frac{P'_*(\{b, c, d, f, g, h\}) - P'_*(\{c, d, g, h\})}{1 - P'_*(\{c, d, g, h\})} = \frac{\pi}{0.8\,\pi + 0.2}$$

$$P'^*(TC|B) = \frac{P'^*(\{b, f\})}{P'^*(\{a, b, e, f\})} = \frac{\pi}{0.8\,\pi + 0.2}$$

The fact $P'^* = P'_*$ reflects the Bayesian nature of the solution. As shown in Shafer (1976a), once a belief is quantified by a probability distribution, all updating will result in a probability distribution.

Table 3: List of elements x of the space TxΘ, with their image under M and the value of the basic belief masses m induced by P_X.

elements x of X=T xΘ	M(x)	$m(M(x)) = P_X(\{x\})$
$ThW \cap TH$	{c}	0.8 (1-π)
$ThW \cap TC$	{b}	0.8 π
$ThB \cap TH$	{e, g}	0.2 (1-π)
$ThB \cap TC$	{f, h}	0.2 π

5.3. Remarks.

The adaptation of the mapping is perfectly acceptable. Disagreement appears in the dynamic part when the D-conditioning is used (Levi 1983, Good 1982, Williams 1982, Shafer, 1986). Why does one keep the same probability distribution on Θ once B has been learned. Bona fide probability analysis would require that P_Θ also be updated once B is known to have occurred. One should have

$$P(ThB|B) = \frac{P(ThB)\, P(B|ThB)}{P(B)} = \frac{0.2\ x}{0.2\ x + 0.8\ \pi}$$

$$P(ThW|B) = \frac{0.8\ \pi}{0.2\ x + 0.8\ \pi}$$

To be allowed to use P(ThB|B) = P(ThB) = 0.2, one must have x = π - a constraint unjustified in the context and one that can usually not be justified in more elaborate examples. Of course, to answer Levi's criticisms will take us straight back to the upper and lower probabilities model.

An important point in Dempster's model is that one recognizes the existence of a probability distribution P_Y on Y. The statement $P_*(A) \leq P_Y(A) \leq P^*(A)$ is meaningful as $P_Y(A)$ exists, even though its exact value is unknown. The so-called Dempster-Shafer's model usually corresponds to this interpretation. It is not the case with the transferable belief model (section 6) where no concept of probability distribution on Y is assumed or required.

The important point in the Dempster's model is that one starts with a probability measure P_X on X and assumes a probability measure P_Y on Y.

Levi's criticism of Dempster-Shafer's analysis is based on the assumption that there is a probability distribution P_Y on Y, in which case the Dempster's rule of conditioning is questionable. But if you do not assume a probability distribution on Y, the criticism becomes unfounded. This explains why we try to describe the TBM without assuming a probability distribution on Y, even though there may exist one on X. We reject the

Bayesian postulate that *on any space* there is *always* a probability measure that quantifies our degree of belief. Introduce probabilities on Y, and Levi's criticisms must be resolved, but how? Shafer's answer (see Shafer, 1986) is hardly convincing.

6. The transferable belief model

6.1. The model

The transferable belief model is a mathematical, idealized model that sets out to represent someone's beliefs. For such a representation, it uses belief function where bel(A) is my degree of belief that A is true.

The **transferable belief model** is based on a two-level model:
- a credal level where beliefs are entertained, combined and updated,
- a pignistic level where beliefs are used to take decisions (from pignus, a bet in Latin, cf. Smith, 1961).

At the credal level, each piece of evidence leads to the allocation of parts of some initial finite amount of belief to subsets of the universe of discourse Ω, also called the frame of discernment. If all these parts were allocated only to the singletons of Ω, the resulting model would correspond to the Bayesian model. But, in general, parts may be allocated to subsets. They represent that part of our belief that supports some subset of the frame of discernment, but that does not support any more specific subsets because of lack of information.

Suppose a mass m(A) (called a **basic belief mass**) supports a set A of Ω, and then you learn that subset X of Ω is impossible. The basic belief mass given initially to A now supports $A\cap\overline{X}$. So the basic belief mass m(A) is transferred to $A\cap\overline{X}$, hence the name of the model. This corresponds to Dempster's rule of conditioning.

The **degree of belief** bel(A) given to the set A of Ω is defined as the sum of all masses that support A,

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

The **degree of plausibility** function pl(A) quantifies the total amount of belief that might support A:

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

The transferable belief model claims that beliefs are quantified by *a single number* (bel). In that, it differs from Kyburg's (1987) models that considers that states of belief are described by sets of probability distributions, and therefore do not require that degrees of beliefs be single-valued.

The transferable belief model is not a particular case or a generalization of probability models, nor of any meta-model based on probability distributions. We never assume the existence of a probability distribution on Ω (a contrario, it is assumed in Dempster's model). Nevertheless, one must be careful to avoid the treachorous pitfall. Inasmuch as the quantification of our belief on Ω is represented by a belief function, one can always build an 'artificial' space Θ endowed with a probability measure and with a one-to-many mapping between Θ and Ω, just as in the Dempster's model. But, here, this is just a mathematical property without any practical meaning. This probability measure on Θ does not correspond to a measure of some subjective belief on Θ. It is merely a mathematical construct.

The existence of an underlying probability measure reflecting some beliefs is irrelevant to our approach. We only postulate the existence of basic belief masses assigned to subsets A of Ω, each expressing the support given specifically to A and induced by the evidence, and that could be transferred to more specific subsets of Ω should new pieces of evidence become available (the conditioning process).

In any modelling of uncertainty, it is required that a method should be provided for making decisions. If the TBM is to be more than just an artificial idealized mathematical model of how beliefs could be represented and updated, one must describe - and justify - a tool for decision making. Our solution is detailed in (Smets 1990b, Smets and Kennes 1990).Suppose one must bet on the elements of Ω. One must build on Ω a probability distribution (the pignistic probability distribution BetP) derived from the belief function that describes our credal state. For all elements $x \in \Omega$:

$$\mathrm{BetP}(x) = \sum_{x \in A \subseteq \Omega} m(A) / |A| = \sum_{A \subseteq \Omega} m(A)\ |\{x\} \cap A| / |A|$$

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

6.2. The transferable belief model analysis

When π in unknown, the static part results in the basic belief masses

$m(\{b, c\}) = 0.8$ and $m(\{e, f, g, h\}) = 0.2$.

Conditioning on B implies the transfer of all basic belief masses within the set {a, b, e, f}. The updated basic belief masses m_B are

$m_B(\{b\}) = 0.8$ and $m_B(\{e, f\}) = 0.2$

So $bel(TC|B) = bel_B(\{b, f\}) = 0.8$, and $pl(TC|B) = pl_B(\{b, f\}) = 1$.

When π in known, the static part results in the basic belief masses m' with

$m'(\{b\}) = 0.8\ \pi$ $m'(\{c\}) = 0.8\ (1-\pi)$ $m'(\{e, g\}) = 0.2\ (1-\pi)$ $m'(\{f, h\}) = 0.2\ \pi$

Conditioning on B implies the transfer of all basic belief masses within the set {a, b, e, f}. The updated basic belief masses m'_B are

$$m'_B(\{b\}) = 0.8\ \pi \quad m'_B(\{e\}) = 0.2\ (1-\pi) \quad m'_B(\{f\}) = 0.2\ \pi$$

So $bel'(TC|B) = bel'_B(\{b, f\}) = \dfrac{\pi}{0.8\ \pi + 0.2}$,

and $pl'(TC|B) = pl'_B(\{b, f\}) = \dfrac{\pi}{0.8\ \pi + 0.2}$.

6.3. Remarks

These solutions are numerically the same as those obtained by Dempster's analysis. The difference does not reside in the numerical results, but in their interpretation. Nowhere have we postulated a probability distribution on Ω. But one might argue that the information used to build the initial basic belief masses is probability related. Though true, it is not necessarily so. To understand the exact nature of the origin of these masses and the relation with probability distribution, one must consider space Ω and the various algebra $\Re$ one can build on Ω. If there exists an algebra $\Re_0$ on which our belief happens to be quantified by a probability distribution P_0, whatever the origin of this probability distribution, then $\forall A \in \Re_0$, $bel(A) = P_0(A)$. This is a requirement to be compared with Hacking's frequency principle (Hacking, 1965):

$$\text{Belief}(A|\ \text{chance}(A) = p) = p$$

It here becomes:

If there exists an algebra $\Re \subseteq 2^{\Omega}$ on which a probability distribution $P_0(A)$ is known $\forall A \in \Re$, then $bel_{\Re}(A) = P_0(A)$

In case there are two algebras $\Re_1$ and $\Re_2$ on Ω on which our beliefs are quantified by probability distribution P_1 and P_2, and such that $\Re_1 \subseteq \Re_2$ then use P_2 to define your belief function bel on $\Re_2$. P_1 is neglected as it is less informative then P_2 and P_2 contains the information included in P_1 (and even more in general). The case when neither of the two algebras is a subalgebra of the other is not studied here as being not fundamental to our discussion.

Suppose bel_0 is known on some algebra $\Re_0$ and belief must be expressed on 2^{Ω}, in the absence of any further information, build the vacuous extension bel of bel_0 on 2^{Ω}, with basic belief masses m such that:

$$m(A) = \begin{cases} m_0(A), \forall A \in \Re \\ 0, \quad \text{otherwise} \end{cases}$$

This is the extension used in our example when we wrote:

$$m(\{b, c\}) = 0.8 \qquad \text{and} \qquad m(\{e, f, g, h\}) = 0.2.$$

When no algebra can be founded on which a probability distribution could describe our beliefs, we can nevertheless assess basic belief masses using betting behaviours and varying betting frames as detailed in (Smets and Kennes, 1990).

7. The Evidentiary Value Model

Ekelöf (1982) initially proposed a theory of evidentiary value in the judicial context (see Gardenförs et al. 1983 for a survey of the topic). The model is very close to Dempster-Shafer's model and the transferable belief model.

An evidentiary argument includes three components (Gardenförs, 1983):

- an evidentiary theme that is to be proved,
- evidentiary facts,
- evidentiary mechanisms which stating that an evidentiary fact is caused by an evidentiary theme.

The authors suggest that the probability that the evidentiary mechanism has worked given the evidentiary facts is considered more important judicially than the probability of the evidentiary theme given the evidentiary facts.

This model fits inperfectly with the breakable sensor paradigm. The evidentiary value EV(TC|B) that the temperature is cold given the light is B is the probability that the fact B is caused by the theme TC, i.e. the probability that the thermometer is working:

$$EV(TC|B) = P(ThW) = 0.8$$

Identically, EV(TH|B) = 0 as a blue light B is not caused by a hot temperature. It can of course occur in such cases, but it would not result from a causal link.

Thus EV(TC|B) = bel(TC|B) and 1-EV(TH|B) = pl(TC|B). This equality has led some authors to consider that Dempster's model might find its justification in the evidentiary value model. But the problem with this approach is the same as with Dempster's model. Why do we keep P(ThW) = 0.8 once B is learned, and why don't we use P(ThW|B) as the causal weight? Levi's criticisms apply here also.

Another problem appears once the prior probability π is known. What is then the probability that the fact B is caused by TC. In the context $ThW \cap TC$ (probability 0.8π, see table 3), B is caused by TC. But what about context $ThB \cap TC$ (with probability 0.2π): can we say that B is caused by TC? Hicks (1979, p. 13) requires four elements to assert that "TC causes B".

1) TC existed
2) B existed
3) the hypothetical situation in which TC did not exist, ceteris paribus,

can be constructed

4) if that situation had existed, B would not have occurred.

In context $ThW \cap TC$, the causality requirements are satisfied: TC and B existed, the hypothetical context is $ThW \cap TH$ (the ceteris paribus clause requires that the thermometer status is kept identical) and in that context, B can not occur. Hence $P(ThW \cap TC) = 0.8\pi$ is part of the $EV(TC|B)$.

In the context $ThB \cap TC$, the causality requirements are not satisfied: TC and B existed, the hypothetical context is $ThB \cap TH$ and in that context, B can occur. So $P(ThB \cap TC) = 0.2\pi$ is not included in $EV(TC|B)$.

Hence when π is considered, Dempster's and the transferable belief solutions are different from the evidentiary value solution. So these models are different.

8. Probability of modal propositions

Ruspini (1986) interprets Dempster-Shafer's model as a model to quantify not the probability that a proposition is true, but the probability that a proposition is necessary. Pearl (1988) interprets Dempster-Shafer's model as a model to quantify not the probability that a proposition is true, but the probability that a proposition is provable. Necessity and provability are very similar modal concepts and can be represented by the modal operator $\Box$. So both interpretations of bel is that bel(A) is the probability of $\Box A$. These interpretations are similar to those in the evidentiary value model.

If $bel(A) = P(\Box A)$, then $bel = P\Box$. In that case it is easy to show that $P\Box$ is a capacity of order infinite (i.e. it satisfies all the inequalities that characterize a belief function). But difficulty appears in defining the concept of updating and justifying Dempster's rule of conditioning. What is $bel(A|B)$? Is it $P(\Box A|B)$, $P(\Box(A|B))$ or $P(\Box A|\Box B)$? The problem is still open. If one defines it as $P(\Box A|\Box B)$ (probably the most immediate proposal), one might expect that[1]:

$$bel(A|B) = P(\Box A|\Box B) = \frac{P(\Box A \wedge \Box B)}{P(\Box B)} = \frac{P(\Box(A \wedge B))}{P(\Box B)} = \frac{bel(A \wedge B)}{bel(B)}.$$

This conditioning rule is called the geometrical rule of conditioning (see Suppes and Zanotti, 1977, Shafer, 1976b, Smets, 1991a). It is encountered also in random sets theory (Hestir in Nguyen, 1991).

If one takes Nguyen's interpretation of A|B as a conditional object (Goodman, Nguyen and Walker, 1991) , one can obtain:

---

[1] The context being propositional logic, we use the conjunction operator $\wedge$ instead of the set operator. In modal logic, $\Box A \wedge \Box B \equiv \Box(A \wedge B)$.

$$\mathrm{bel}(A|B) = P(\Box(A|B)) = \frac{P(\Box(A \wedge B))}{P(\Box B)} = \frac{\mathrm{bel}(A \wedge B)}{1+\mathrm{bel}(A \wedge B)-\mathrm{bel}(\neg B \vee A)}.$$

But these two rules do not correspond to Dempster's rule of conditioning defended by Shafer and the TBM. Ruspini (1986) uses Dempster's rule of conditioning to define bel(A|B), but does not fully justify such a choice.

One could be tempted to claim that the TBM and other models based on belief functions are nothing but models for probability of modal propositions. This claim, though tempting, is not yet justified. To be acceptable, one will have to explain the origin of Dempster's rule of conditioning. In the TBM, it is an integral part of the model: it corresponds to the ability for each mass m(A), $A \subseteq \Omega$, to move freely among the subsets of A if further information justifies it. In Dempster's model, it is deduced in the D-conditioning if one accepts the postulates that the knowledge of the conditioning event do not modify the probabilities (except for a possible rescaling) on the X space (see section 5.1).

9. The coin experiment or 'Do we need the TBM?'

To explain the difference between the TBM, the Dempsterian and the Bayesian analysis, we consider a classic example already analysed in Shafer and Tversky (1985).

Suppose there are four men: Mr. Truth, Mr. False, Mr. Head and Mr. Tail. They all look at the visible face of a coin you have put on the table (I did not say that you have tossed the coin before putting it on the table). I know that:

Mr. Truth will say H when the face is Head, T when the face is Tail.
Mr. False will say T when the face is Head, H when the face is Tail.
Mr. Head will always say H.
Mr. Tail will always say T.

One of the four men is selected by a random process. The selection is independent from the visible face of the coin. For each of the four men, the probability hewill be selected is 0.25. I do not know who the randomly selected man is. I hear he says H. I wish to determine what my belief is that the face is Head?

Let Ω be the space {H, T} and the collected information be denoted SH (for the selected person said H).

Both the TBM and the Dempsterian analysis of the scenario lead to a first conclusion that Mr. Tail was certainly not selected. The probabilities allocated to the other options therefore are proportionally rescaled into 0.33 (the closed world assumption is accepted for the TBM analysis). Then:

the 0.33 allocated to Mr. Truth supports the fact that the H is true, the 0.33 allocated to Mr. False supports the fact that T is true,

the 0.33 allocated to Mr. Head supports nothing (hence Ω).

Hence bel(H|SH) = 0.33, bel(T|SH) = 0.33 and bel(H$\cup$T|SH) = 1.

Levi's criticism is that the updating of the three probabilities is unjustified. For a bayesian, the appropriate updated probabilities are the probabilities that Mr. Truth (False, Head, Tail) were indeed selected given I know now that the selected person said H. Of course, the available data do not enable us to unambiguously define these probabilities, except for the one related to Mr. Tail, which is 0 by necessity. The missing information is the a priori probability that the face was H (or T).

Should you have tossed the coin, that probability would have been .5, and the bayesian - as well as any other approach in fact - would have led to updating the initial 0.25 probabilities into:

P(Mr. Truth | SH)	= P(SH | Mr. Truth) P(Mr. Truth) / P(SH)	= 0.5∗0.25/0.5	= 0.25
P(Mr. False | SH)	= P(SH | Mr. False) P(Mr. False) / P(SH)	= 0.5 ∗0.25/0.5	= 0.25
P(Mr. Head | SH)	= P(SH | Mr. Head) P(Mr. Head) / P(SH)	= 1.∗0.25/0.5	= 0.5
P(Mr. Tail | SH)	= P(SH | Mr. Tail) P(Mr. Tail) / P(SH)	= 0.0∗0.25/0.5	= 0.0

where the denominator 0.5 is the normalization factor such that the probabilities add to one as they should. Of course one does not find the 0.33 of the previous analysis. It reflects the impact of the supplementary information about how the coin was dropped on the table.

But this is not the situation that prevails here. We ignore how the visible face of the coin was selected: no randomness, not even a probability measure were assumed on Ω.

For a Dempsterian analysis, one defines the spaces Ω = {H, T}, Π = {Mr. Truth, Mr. False, Mr. Head, Mr. Tail}, Σ = {SH, ¬SH} (the possible observations). The X and Y spaces and the M mappings of section 5.1. correspond here to the spaces Π, $\Pi x\Omega x\Sigma$ and the mapping defined by the four pieces of information on how each person answers once he has seen the face of the coin (e.g. M(Mr.Truth) = {(Mr. Truth, H, SH)$\cup$(Mr. Truth, T, ¬SH)}, M(Mr.Head) = {(Mr. Head, H, SH)$\cup$(Mr. Head, T, SH)},...). Given the independence assumptions between the selection process on Π and the value of Ω, the knowledge of a probability distribution on Ω would be sufficient to build the probability measure on $\Pi x\Omega x\Sigma$. One would then derive the probabilities on Ω once SH is learned, by a direct application of the probability calculus.

But such a probability on Ω is unknown, let alone the probability distribution on $\Pi x\Omega x\Sigma$. In a Dempsterian approach, one nevertheless somehow acknowledges that there exists some probability measure P on $\Pi x\Omega x\Sigma$, that one can speak of P(A) for A in $\Pi x\Omega x\Sigma$. Of course, the value of P(A) can only be claimed to be in some interval, the interval [P∗(A), P*(A)]. Once SH is learned, according to the model, one should update the mapping (by deleting all elements with ¬SH in the space ΠxΩxΣ). But why was the probability measure on Π not updated as it should be in probability theory?

In the TBM, we do not postulate the existence of a probability measure on Ω, nor on ΠxΩxΣ. We acknowledge an initial probability measure on Π. We build a belief function on ΠxΩxΣ by vacuously extending the initial information (according to the minimal commitment principle, see Smets, 1991a). The important point is that we never assumed a probability measure on ΠxΩxΣ. Then we learn SH. The conditioning event is not 'measurable' on Π, where 'non measurable on Π' means that the set of elements of ΠxΩxΣ that become impossible after we learn SH is not a cylindrical extension of a subset of Π on ΠxΩxΣ (is not a subset of ΠxΩxΣ that can be written as πxΩxΣ where $\pi \subseteq \Pi$). In fact the knowledge of SH leave us ignorant as to what should become the probability on Π (except that we know that Mr. Tail was not selected). We can nevertheless update our belief on ΠxΩxΣ (by transfering the basic belief masses as would be given by Dempster's rule of conditioning), and derive the TBM results (that are of course numerically equal to those of the Dempsterian analysis).

A harsh critic could say: you had a probability on Π, and it induces your basic belief masses on ΠxΩxΣ. Once you learn SH, you should have updated your probabilities on Π, and used these to update your basic belief masses on ΠxΩxΣ. This is only a rephrasing of Levi's criticisms. The problem is that the probability theory does not provide a method to update probabilities on non-measurable events. I know the probability on every subset of Π, and I could update these probabilities if I learn that some subset of Π is true by applying the conditioning operator (PA|B) = PA∩B)/P(B) $\forall A, B \subseteq \Pi$). But here I must condition on a event on ΠxΩxΣ that is non-measurable on Π. SH is not equivalent to a subset of Π. So probability theory does not put any constraints on how to update my probabilities on Π.

Now why do I later worry about distinguishing the TBM from the Dempsterian model. Because, in the last instance, one accepts a probability measure on ΠxΩxΣ (with some of its values unknown), hence, the probability measure on Π is the marginal of that 'poorly known' probability measure on ΠxΩxΣ.Once SH is known, the probability distribution on ΠxΩxΣ must be updated by the conditioning rule described in probability theory. And the update probability distribution on Π will be the marginal of the new conditional probability defined on ΠxΩxΣ. SH is no longer a non-measurable event. It is measurable on ΠxΩxΣ and its impact on Π can be derived by the marginalization constraints. In fact, once a probability distribution on ΠxΩxΣ is introduced, it is back to an upper and lower probabilities scenario. But then Dempster's rule of conditioning (and combination) becomes irrelevant. Correct updating must be performed by the G-conditioning.

This argument has shown the danger of accepting a probability distribution on $\Pi x \Omega x \Sigma$. Reject it as in the TBM, and all problems related to Levi's criticisms are resolved as they no longer apply.

10. The TBM versus Dempsterian ULP.

The next problem we wish to tackle is related to the misuses of Dempster's rule of conditioning (not speaking of the combination rule) in contexts where it should not be used. The error results from the erroneous assumption that Dempster's rule of conditioning is THE appropriate rule for updating all beliefs that happen to be quantified by belief functions, regardless of the context. Figure 1 summarizes the differences between the Dempsterian ULP model and the TBM.

10.1. Dempsterian ULP model context.

Suppose a space X on which my belief is quantified by a probability distribution P_X. Updating of P_X on subsets B of X is performed by the probabilistic conditioning rule:

$$P_X(A|B) = \frac{P_X(A \cap B)}{P_X(B)} \qquad \forall AB \subseteq X$$

Suppose a one-to-many mapping M between space X and the power set 2^Y of a space Y. Suppose we postulate there is a probability distribution P_Y on Y. Can we build P_Y such that it satisfies the constraints due to the existence of the mapping M and the probability distribution P_X? Of course not, but we can define for every subsets A of Y limits between which $P_Y(A)$ must be. Let these limits be $P_*(A)$ and $P^*(A)$ (see section 5.1.). These limits characterize a family of probability distributions Π on Y. Each element of Π is compatible with the mapping M and probability distribution P_X.

But suppose you would like to select THE best representative of Π. You might want to select the element of Π that is the 'least informative' on Y. The immediate translation of this requirement is: select the element of Π whose entropy is maximal. Let $P°$ be such element of Π. You can then proceed with this probability distribution as if it was THE probability distribution induced by M and P_X on Y. So updating on $B \subseteq Y$ would be obtained by the classic conditioning rule of probability theory applied to $P°$.

But you could also refuse to reduce Π to one of its elements and consider that all that is known is that P_Y is in Π. Then updating on $B \subseteq Y$ is obtained by considering each element of Π and conditioning it on B, and so deriving the set Π_B of all the conditional probability distributions on Y given B. This process is nothing else but that which led to the G-conditioning in sections 4.1 and 5.1.

10.2. TBM context

A totally different scenario, but perfectly parallel in its form, can be constructed in the TBM context. Suppose a space X on which my belief is quantified by a belief function bel_X. Updating bel_X on subsets $A \subseteq X$ is obtained by Dempster's rule of conditioning.

Suppose then the same mapping M as above. On Y one can then define a family B of belief functions compatible with bel_X and M. One can define the upper and lower belief and plausibility functions bel^* and bel_*, pl^* and pl_* where for instance $\forall A \subseteq Y$, $bel^*(A) = \max bel(A)$ and the max is taken on the elements of B. Mathematically, bel_* is a belief function, pl^* is the dual plausibility function (Dubois and Prade, 1986). The other two functions are neither belief nor plausibility functions.

Dempsterian ULP versus TBM.

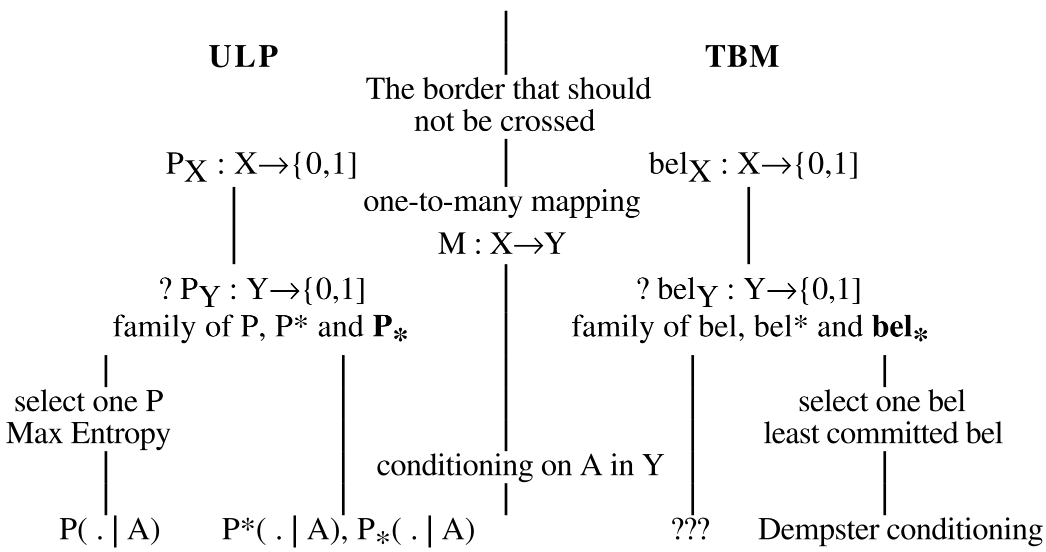

Figure 1: comparing the Dempsterian ULP model and the TBM.

As in the probabilistic approach one could select THE best representative of B. In the TBM context, there is a principle, the least commitment principle, that allows the selection of the least informative belief function in B. For a pair of belief functions defined on a space Ω, one says that bel_1 is less committed than bel_2 if $\forall A \subseteq \Omega$ $bel_1(A) \leq bel_2(A)$ (under closed world assumption only) or if $\forall A \subseteq \Omega$ $pl_1(A) \geq pl_2(A)$ (under closed world assumption, both definitions are identical, under open world assumption, only the plausibility-based definition is valid (Smets 1991b)). This principle expresses the idea that one should never commit more belief than what is justified by the available information. It could of course be criticized, but it is surely much easier to defend than the maximum entropy principle. In the family B induced by bel_X and M, the least committed belief function is bel_*. This is the belief function that contains all and nothing but the information induced on Y by bel_X and M. So the normal attitude of the

TBM analysis is to use this belief function and to proceed with it. Updating for some B in Y is realized by the application of Dempster's rule of conditioning.

One could try to simulate the ULP attitude, keep the family B of belief functions, condition each of them, and derive the set of compatible conditional belief functions on Y given B. This approach seems to have never been pursued, maybe because the approach based on selecting the least committed solution is convincing.

The parallelism between the two scenarios is quite obvious. Nevertheless an error is often encountered. It happens that the lower probability P_* induced by P_X and M on Y is a belief function. Discovering that property, users come to think that the appropriate rule for updating this lower probability is Dempster's rule of conditioning. It is totally unjustified (see Fagin and Halpern, 1990). The nature of the two scenarios is completely different: in one case one has a family of probability distributions on Y, in the other a family of belief functions on Y. The origin and nature of these sets must be recognized, the fact that P_* is a belief function does not entitle the user to cross the line between the two parallel scenarios and assimilate the lower probability to a belief function on which Dempster's rule of conditioning can be applied.

10.3. Final remarks on the difference

For every problem, the solutions obtained by the Dempsterian ULP analysis and the TBM analysis are numerically the same. So what is the difference between the two models? The difference laie essentially in the aim and nature of the two models.

In the TBM, the aim is to model degrees of belief. Any underlying concept of probability is irrelevant. One starts with the bbm allocated to subsets of Ω that cannot be allocated to more specific subsets because of a lack of information, but that will be transferred to more specific subsets of Ω in the light of new pieces of evidence. The dynamic nature of the model is part of its definition.

In the Dempsterian ULP model, the essential aim is not in modelling degrees of belief. The problem tackled is the evaluation of the boundaries between which lie some probabilities that are constrained by some underlying probabilities and a one-to-many mappings. The bbm are just incidental mathematical niceties and the conditioning rule is open to criticism like those of Levi.

Shafer's (1976a) initial contribution was to introduce the use of the Dempsterian ULP model to model degrees of belief, but his recent presentations are very much in line with the underlying probability measure and the one-to-many mapping. His modelisation further acknowledges the presence of some probability measure on the frame of discernment and, as a result, it is open to Levi's criticisms.

11. Conclusions

Consider the mapping M between the X and 2^Y as presented in section 5.1.The major difference between the transferable belief and the probability approaches is of course in the way we create our beliefs on Y knowing the belief on X. The transferable belief model is based on what is available and nothing else whereas the probability analysis postulates the existence of a probability distribution on Y. Bayesians assume that whenever a probability distribution P_X is defined on X, then one can describe a probability distribution P_Y on Y where P_Y satisfies the constraints induced by P_X and the mapping M.

Reconsider the $x = P(B|ThB)$ parameter in the breakable sensor paradigm. Bayesians claim the existence of such a probability, i.e. they claim that in context ThB, the B or R light data is governed by a random process, not by an arbitrary process as acknowledged in the transferable belief model analysis.

What is the **difference between a random and an arbitrary selection procedure**?

The **frequentists** suppose that an experience is repeated n times and a particular event X is observed r times. They claim there exists a limit for the ratio r/n when n tends to infinity (the limit being the probability of X). If the events are generated arbitrarily, the existence of such a limit is not claimed. We know NOTHING about an arbitrary process, whereas we know at least something about a random process: the existence of a limit.

In a classical frequentist analysis one knows the existence of a limit and its value, in an ULP analysis one knows the existence of a limit but not its value, in the transferable belief model one does not even know the existence of a limit (let alone its value).

The **Bayesian** can justify his probabilities through betting argument. The pignistic probabilites are such that Dutch books are avoided at the synchronic level. At the diachronic level, the argument is based on an assumption of temporal coherence that can be disposed of without falling into the Dutch Book trap (Smets and Kennes, 1990). The TBM and its related decision making rules are so that there exists no a priori startegy that lead to a sure loss. In the TBM, contingent conditioning and updating are different concepts (Walley, 1991).

The requirement about the existence of a limit or of some coherent behaviour can be avoided in a logical approach like the one based on **Cox's axioms** (Cox, 1946), but at the cost of introducing the axiom : 'the belief given to the complement of a proposition A is a function of the belief given to A'. This postulate is not required in the transferable belief model (Smets et al., 1991).

All alternatives to the transferable belief model explicitly or implicitly accept the Bayesian assumption: the existence of probability distributions on all relevant spaces. The genuine difference between the transferable belief model and all its contenders lies in this assumption. Accept it and Levi's remarks are relevant. In the transferable belief model, one never postulates the existence of these probability distributions. One only recognizes that if a probability distribution can be defined on some algebra, it should induce coherence constraints on the way beliefs are allocated. But never infer that a probability distribution exists on those spaces on which we extend vacuously the belief function derived from the initial probability constraints.

At the credal level, nothing requires that beliefs be quantified by probability distributions, even though it may happen sometimes. All arguments in favour of representing degrees of belief by probability distributions and based on betting behaviour or decisions theory are also satisfied in our model at the pignistic level.

Claiming the existence of a probability distribution on the space on which our beliefs are assessed is already an information in itself. Should you accept it, then the upper and lower probabilities model should be applied.

We strongly reject the following interpretation where belief functions are used instead of upper and lower probabilities. Some authors consider that Dempster-Shafer's model (i.e. belief and plausibility functions) can be used to handle cases of ill-defined probabilities, cases where there exists a probability function on Ω but we only know that its values for each $A \subseteq \Omega$ is contained between two limits. They claim that all that is known is a belief function bel (or equivalently a plausibility function pl as $pl(A) = 1-bel(\overline{A})$) such that:

$$\forall A \subseteq \Omega, \qquad bel(A) \leq P(A) \leq pl(A)$$

This might be the case, but then they should justify why the lower limits are quantified by a belief function, why the lower limits conform to the inequalities that characterize belief functions, why the Moebius transform of the lower limits are non negative? Furthermore when conditioning is involved, how do they justify the use of the D-conditioning and not that of the G-conditioning? These questions have to be answered before using belief functions instead of lower probabilities functions as lower limits for the intervals and before using Dempster's rule of conditioning (and Dempster's rule of combination). All too often authors simply confuse the two theories. This explains why we felt justified in writing this paper. We hope we have succeeded in somewhat clarifying the matter.

Acknowledgments. The author wishes to thank Y.T. Hsia, R. Kennes, H.T. Nguyen, A. Saffiotti for their help and contribution. His thanks extend to all his friends who so often disagree with him, thereby forcing him to refine his arguments.

Bibliography:

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

DEMPSTER A.P. (1967) Upper and lower probabilities induced by a multivalued mapping. Annals of Math. Statistics 38:325-339.

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

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

FAGIN R. and HALPERN J. (1990) A new approach to uptdating beliefs. 6th Conf. on Uncertainty in AI.

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.

GARDENFORS P. (1983) Probabilistic reasoning and evidentiary value. in GARDENFORS P., HANSSON B. and SAHLIN N.E. (eds) Evidentiary value: philosophical, judicial and psychological aspects of a theory. C.W.K. Gleerups, Lund, Sweden. pg 44-57.

GOOD I.J. (1982) Comments on Lindley's paradox, by G. Shafer. J. Amer. Statist. Ass. 77: 342-344.

GOODMAN I.R., NGUYEN H.T. and WALKER E.A. (1991) Conditional inference and logic for intelligent systems. Elsevier, Amsterdam.

GORDON J. and SHORTLIFFE E.H. The Dempster-Shafer theory of evidence, in BUCHANAN B.G. and SHORTLIFFE E.H. (Eds), Rule-Based Expert Systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading, MA (1984) 272-292.

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

HESTIR K, NGUYEN H.T. and ROGERS G.S. (1991) A random set formalism for evidence reasoning. In "Conditional Logic in Expert Systems" , Goodman I.R. Gupta M.M., Nguyen H.T. and Rogers G.S. eds. Elsevier, Amsterdam.

HICKS J. (1979) Causality in economics. Basil Blackwell, Oxford, UK.

KLAWONN F. and SCHWECKE E. (1990) On the axiomatic justification of Dempster's rule of combination. submitted for publication.

KYBURG H.E.Jr. (1987) Bayesian and non-Bayesian evidential updating. Artificial Intelligence, 31:271-294.

LEVI I. (1983) Consonance, dissonance end evidentiary mechanisms. in GARDENFORS P., HANSSON B. and SAHLIN N.E. (eds) Evidentiary value: philosophical, judicial and psychological aspects of a theory. C.W.K. Gleerups, Lund, Sweden. pg. 27-43.

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.

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

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

SHAFER G. (1976b) A theory of statistical evidence. in Foundations of probability theory, statistical inference, and statistical theories of science. Harper and Hooker ed. Reidel, Doordrecht-Holland.

SHAFER G. (1986) The combination ofd evidence. Int. J. Intell. Systems 1:155-180.

SHAFER G. and TVERSKY A. (1985) Languages and designs for probability. Cognitive Sc. 9:309-339.

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 pg 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) 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. (1990c) 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) Axioms for belief functions. under preparation.

SMETS Ph. (1991b) Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. (submitted for publication)

SMETS Ph. and KENNES (1990) The transferable belief model. Technical Report: TR-IRIDIA-90-14.

SMETS Ph., CLARKE M.R.B. and FROIDEVAUX Ch. (1991) Uncertainty, Conditional and Non-Monotonicity. Positions and debate in non standard logics. Submitted for publication.

SUPPES P. and ZANOTTI M. (1977) On using random relations to generate upper and lower probabilities. Synthesis 36:427-440.

WALLEY P. (1991) Statistical reasoning with imprecise probabilities. Chapman and Hall, London.

WILLIAMS P.M. (1982) discussion in SHAFER G. Belief functions and parametric models. J. Roy. Statist. Soc. B44: 322-352.