Probability of Provability and Belief Functions.

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

Abstract: We present an interpretation of Dempster-Shafer theory based on the probability of provability. We present two forms of revision (conditioning) that lead to the geometrical rule of conditioning and to Dempster rule of conditioning, respectively.

Keywords: conditioning, belief revision, belief functions, probability of provability.

1. Introduction.

Dempster-Shafer theory has received much attention recently in AI, both in favorable and unfavorable ways (see the special issues of the Intern. J. Approx. Reasoning, vol. 4, 1990 and vol. 6, 1992). Most criticisms are based on confusion resulting from an inappropriate mixing of several interpretations of the theory. No specific interpretation is 'better' than any other, each one fits a specific domain. Ruspini (1986) and Pearl (1988) have considered logical foundations of the model based on the concepts of the probability of a modal proposition (knowing) or of the probability of provability. We analyze in detail the probability of provability interpretation and show how the conditioning rules can be derived. We first present the two conditioning rules that have been distinguished recently by Dubois and Prade (1991), the focusing and the revision. We proceed by detailing the concept of the probability of the provability of some propositions. We show the relation of these particular probability functions with the belief functions. Finally, we study two conditioning processes that correspond to the geometrical rule of conditioning and to the unnormalized (Dempster) rule of conditioning, respectively.

[1] This work has been partially funded by the CEC-ESPRIT III Basic Research Project 6156 (DRUMS II), and the Communauté Française de Belgique, ARC 92/97-160 (BELON).

2. Revision versus focusing :

Dubois and Prade (1991) have introduced beautifully the difference between two types of conditioning :

Case 1. A die has been tossed. You[2] assess the probability that the outcome is ‘Six’. Then a reliable witness says that the outcome is an even number. How do You update the probability that the outcome is ‘six’ taking in due consideration the new piece of information.

Case 2. Hundred dice have been tossed. You assess the proportion of ‘six’. Then You decide to focus Your interest on the dice with an even outcome. How do You compute the proportion of ‘six’ among the dice with an even outcome.

Case 1 corresponds to a **revision**[3] as the probability is modified to take in account a new piece of information.
Case 2 corresponds to a **focusing** : no new piece of information is introduced, we just consider another reference class by focusing our attention on a given subset of the original set.

In probability theory, the distinction is more conceptual than practical as both cases are solved by Bayes’ rule of conditioning. It might explain the lack of interest for such a distinction. Distinction becomes important when degrees of beliefs are quantified by belief functions like in Dempster model (Dempster 1967), in Shafer model (Shafer, 1976a), in the Hints Model (Kohlas and Monney, 1990) and in the Transferable Belief Model (Smets, 1988, Smets and Kennes, 1990).

In the present study of the probability of the provability, we restrict ourselves to the revision case.

3. Probability of Provability.

3.1. Introduction.

Pearl (1988) presents the ‘probability of provability’ interpretation of Dempster-Shafer theory. Ruspini (1986) studied a similar problem by using the modal approach, considering the ‘probability of knowing’. Both approaches fit essentially with the

---

[2] 'You' is the agent that entertains the beliefs considered in this presentation.
[3] In Dubois and Prade (1991), the authors called it an ‘updating’ but today they prefer to call it (Dubois and Prade 1992) a revision in harmony with the Alchouroron, Gärdenfors and Makinson approach (Gärdenfors, 1988) where revision concerns the beliefs held by an agent. They reserve ‘updating’ for the case considered by Katsuno and Mendelzon (1991) that concerns the update of an evolving world.

same ideas. Conceptually, the probability of provability approach is not different from the original Dempsterian approach (Dempster 1967) but it provides a nice framework and explains the origin of the conditioning rules. It can be described as follows.

Let $\mathcal{H}$ be a finite Boolean algebra of propositions. These propositions are called the hypotheses. Let $\mathcal{L}$ be another finite Boolean algebra of propositions. We define a (multi-) mapping from $\mathcal{H}$ to $\mathcal{L}$, where satisfies both:

If HL then HL $\vee$L’ (right weakening)

If HL and HL’ then H L$\wedge$L’ (and)

can be seen as a form of provability or consequence relation. Hereafter we will consider it as the provability operator. One property not satisfied by is:

If HL$\vee$L’ then H L or H L’

The impact of this extra requirement will be studied in section 4 because of its links with probability theory.

Related to , we define the M mapping from $\mathcal{H}$ to $\mathcal{L}$ such that:

$\forall H \in \mathcal{H}$, $M(H) = \wedge L_i : HL_i$, $L_i \in \mathcal{L}$

in which case: HL iff M(H) $\wedge$L = M(H).

M(H) is the most specific proposition in $\mathcal{L}$ that can be deduced from H. In particular, if M(H) = $\bot$, then H = $\bot$, but not the converse.

Suppose there is a **probability measure** $P_{\mathcal{H}} : 2^{\mathcal{H}} \rightarrow [0,1]$ on $2^{\mathcal{H}}$ and let $p_{\mathcal{H}}$: $\mathcal{H} \rightarrow [0,1]$ be the related probability function on $\mathcal{H}$ with $p_{\mathcal{H}}(H) = P_{\mathcal{H}}(\{H\})$ for every $H \in \mathcal{H}$. Note that $\mathcal{H}$ is already an algebra, usually the power set of some set. As $\bot \in \mathcal{H}$, $p_{\mathcal{H}}(\bot)$ may be positive[4].

Given the function M : $\mathcal{H} \rightarrow \mathcal{L}$, we can define on $\mathcal{L}$ the probability that $L \in \mathcal{L}$ is provable and ¬L is not provable. It is denoted by $P_{\mathcal{L}}$(>L). We use the symbol > in $P_{\mathcal{L}}$ to enhance the fact that those H that would also prove ¬L are eliminated. $P_{\mathcal{L}}$(>L) is the probability that an hypothesis selected randomly in $\mathcal{H}$ (according to the probability measure $P_{\mathcal{H}}$) proves L and does not prove ¬L:

$$P_{\mathcal{L}}(\ >L) =_{def} P_{\mathcal{H}}(\{H : H \in \mathcal{H}, HL, H \ \nvdash \neg L\}) = \sum_{H: H \in \mathcal{H}, H \vdash L, H \nvdash \neg L} p_{\mathcal{H}}(H)$$

Example 1: (This didactic example will be studied throughout the paper). To illustrate the meaning of the various components of the model, let the propositions:

R = ‘It rains’

W = ‘It is windy’

C = ‘Paul spend the evening with Carol’

J = ‘Paul spend the evening with John’

---

[4] In Smets (1992) a similar problem is analyzed, i.e., belief functions where m($\bot$) might be positive.

Let $\mathcal{H} = \mathcal{P}$({R, ¬R} x {W, ¬W}) where $\mathcal{P}$ denotes the power set and x denotes the Cartesian product. Let $\mathcal{L} = \mathcal{P}$({C, ¬C} x {J, ¬J}) . Let T denotes the tautology.

$H_i \in \mathcal{H}$		$p_{\mathcal{H}}$	$M(H_i) \in \mathcal{L}$	
H_1	T	.10	L_1	J
H_2	R	.15	L_2	¬C∧J
H_3	W	.18	L_3	C
H_4	R∧W	.24	L_4	C∧¬J
H_5	R∧¬W	.30	L_5	¬C∨¬J
H_6	R∧¬R	.03	L_6	⊥

Table 1: Elements of $\mathcal{H}$, $\mathcal{L}$ and the probability distribution $p_{\mathcal{H}}$ on $\mathcal{H}$.

Table 1 presents a set of hypotheses (H_1 to H_6), their M values (L_1 to L_6) and the probability distribution $p_{\mathcal{H}}$. The origin of such data could be that You are trying to guess with whom Paul will spend his evening (an element of $\mathcal{L}$), and Paul's intentions depend on the 'exact information' that Paul will have about the weather (an element of $\mathcal{H}$). By 'exact information' we mean 'all what is known'. So Paul knows exactly H_4 does not implies that Paul knows exactly H_3. Table 1 says:
1) There are 6 weather reports (H_i : i=1, ..6) available about tonight weather (H_i),
2) Paul will obtain only one of them and the probability that Paul obtains report H_i is $p_{\mathcal{H}}(H_i)$, and
3) the M values are the most specific information You know about Paul's intentions ($M(H_i)$) according to what Paul exactly knows about the weather (H_i).

In particular, if the weather report available to Paul says 'it will be windy' (H_3), then Paul will spend the evening with Carol. For what concerns John, he may or not be present. In case H_6, the report says 'R ∧ ¬R' (i.e. ⊥). In that case, every proposition in L is provable, hence M(⊥) = ⊥.

Then for instance:

$P_{\mathcal{L}}$(>T) = $p_{\mathcal{H}}(H_1) + p_{\mathcal{H}}(H_2) + p_{\mathcal{H}}(H_3) + p_{\mathcal{H}}(H_4) + p_{\mathcal{H}}(H_5)$ = .97
$P_{\mathcal{L}}$(>J) = $p_{\mathcal{H}}(H_1) + p_{\mathcal{H}}(H_2)$ = .25
$P_{\mathcal{L}}$(>¬J) = $p_{\mathcal{H}}(H_4)$ = .24
$P_{\mathcal{L}}$(>C∧J) = 0.00
$P_{\mathcal{L}}$(>¬C∧J) = $p_{\mathcal{H}}(H_2)$ = 0.15.

So one has, among others,

$P_{\mathcal{L}}$(>T) = .97 < 1.00
$P_{\mathcal{L}}$(>J) = .25 < 1 - $P_{\mathcal{L}}$(>¬J) = 1 - .24 = .76
and $P_{\mathcal{L}}$(>J) = .25 > $P_{\mathcal{L}}$(>C∧J) + $P_{\mathcal{L}}$(>¬C∧J) = 0.00 + 0.15

Those inequalities are typical of the unnormalized belief functions and indeed $P_{\mathcal{L}}(\ >L)$ considered as a function of L happens to be a belief function on $\mathcal{L}$ as shown below. ∇

3.2. Belief Functions.

The use of belief functions has been advocated in order to represent quantified beliefs. It is based on the idea that the impact of a piece of evidence on an agent's beliefs over a propositional algebra $\mathcal{L}$ can be expressed by an allocation of parts of an initial unitary amount of belief among the propositions of $\mathcal{L}$. These parts of beliefs are quantified by a basic belief assignment m: $\mathcal{L} \rightarrow [0,1]$ where m(A) for $A \in \mathcal{L}$ is that part of belief that supports A without supporting any proposition B≠A that implies A because of a lack of information justifying this more specific allocation. The sum of those masses that support any proposition B that implies A without implying ¬A is called the degree of belief given to A, denoted bel(A), where bel: $\mathcal{L} \rightarrow [0,1]$ is a belief function. The sum of those masses that support any proposition B that is not contradictory to A is called the degree of plausibility given to A, denoted pl(A), where pl: $\mathcal{L} \rightarrow [0,1]$ is a plausibility function.

Mathematically these functions are defined by:

$$\text{bel}(A) = \sum_{X:X \in \mathcal{L}, X \wedge A = X, X \neq \bot} m(X)$$

$$\text{pl}(A) = \sum_{X:X \in \mathcal{L}, X \wedge A \neq \bot} m(X) = \text{bel}(\top) - \text{bel}(\neg A)$$

Belief functions satisfy the following inequality:

$\forall n \geq 1, A_1, A_2, \ldots A_n \in \mathcal{L}$

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

Probability functions satisfy similar relations except that the inequality is replaced by an equality.

3.3. The probability of provability is a belief function.

Let $\text{bel}_{\mathcal{L}} : \mathcal{L} \rightarrow [0,1]$ be the (unnormalized) **belief function**[5] on $\mathcal{L}$ induced by a given basic belief assignment $m_{\mathcal{L}} : \mathcal{L} \rightarrow [0,1]$. By definition $\text{bel}_{\mathcal{L}}(L)$ is the sum of the basic belief masses $m_{\mathcal{L}}(X)$ given to the propositions X in $\mathcal{L}$ that imply L without

[5] bel is an unnormalized belief function as we do not require m(⊥)=0 (Smets 1988, 1992).

implying ¬L. It can be shown that $P_{\mathcal{L}}(\ >L)$, $L \in \mathcal{L}$ is equal to the belief function $bel_{\mathcal{L}}$ on $\mathcal{L}$ induced by the basic belief assignment $m_{\mathcal{L}} : \mathcal{L} \rightarrow [0,1]$ with:

$$m_{\mathcal{L}}(L) = \sum_{H : H \in \mathcal{H}, M(H) = L} p_{\mathcal{H}}(H)$$

and $m_{\mathcal{L}}(L) = 0$ if the sum is taken over an empty set.

One has:

$$P_{\mathcal{L}}(\ >L) = \sum_{H : H \in \mathcal{H}, H \vdash L, H \nvdash \neg L} p_{\mathcal{H}}(H)$$

$$= \sum_{H : H \in \mathcal{H}, M(H) \vdash L, M(H) \nvdash \neg L} p_{\mathcal{H}}(H)$$

$$= \sum_{L_i : L_i \in \mathcal{L}, L_i \vdash L, L_i \nvdash \neg L} \ \sum_{H : H \in \mathcal{H}, M(H) = L_i} p_{\mathcal{H}}(H)$$

$$= \sum_{L_i : L_i \in \mathcal{L}, L_i \vdash L, L_i \nvdash \neg L} m_{\mathcal{L}}(L_i)$$

So: $P_{\mathcal{L}}(\ >L) = bel_{\mathcal{L}}(L)$.

Similarly the plausibility function $pl_{\mathcal{L}} : \mathcal{L} \rightarrow [0,1]$ is :

$$pl_{\mathcal{L}}(L) = bel_{\mathcal{L}}(\top) - bel_{\mathcal{L}}(\neg L) = \sum_{X : X \in \mathcal{L}, X \wedge L \neq \bot} m_{\mathcal{L}}(X) \qquad \text{for } L \in \mathcal{L}$$

where $\top$ is the maximal element of $\mathcal{L}$. It can be shown that:

$$pl_{\mathcal{L}}(L) = \sum_{H : H \in \mathcal{H}, M(H) \wedge L \neq \bot} p_{\mathcal{H}}(H)$$

Example 2: Given the data of table 1, table 2 presents the belief and plausibility values for several propositions of $\mathcal{L}$. ∇

	bel_L	pl_L
$C \wedge J$	.00	.28
$\neg C \wedge J$	.15	.55
J	.25	.73
$C \vee \neg J$	.42	.82
$\neg C \vee \neg J$	.69	.97
$\neg J$	.24	.72
T	.97	.97

Table 2: Belief and plausibility values for several propositions of L.

This model for the probability of provability is not different from Dempster's model (Dempster 1967) and Shafer's translator model (Shafer and Tversky 1985, Dubois et al. 1991). Both models consider an X domain (the translator-source domain) endowed with a probability measure, an Y domain (the message-data domain) and a one-to-many mapping from X to Y. The H space corresponds to the X domain, the L to the Y, the M mapping to the one-to-many mapping, and P_H to the probability on the X domain.

3.4. Revision.

All probabilities like P_H and beliefs like bel_L considered in this paper are entertained by You at a given time t. They are induced relatively to a given **evidential corpus**, denoted 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. EC_t^Y induces in You at t a credal state representing the partial beliefs You entertain on which propositions of H are true and which propositions of L can be deduced. Let $[EC_t^Y]$ represents the conjunction of the propositions in EC_t^Y. $[EC_t^Y]$ is the background knowledge.

The probability measure $P_H : 2^H \rightarrow [0,1]$ is induced from EC_t^Y. Revision process consists in changing EC_t^Y by adding to it some new piece of information. In this paper, we consider only the case where the added information is compatible with EC_t^Y, i.e. its conjunction with $[EC_t^Y]$ is not contradictory. Therefore the revision processes we are going to described are the result of computing Your beliefs once $[EC_t^Y]$ has been transformed into $[EC_t^Y] \wedge A$ where A is the proposition that represent the new piece of information. In order to simplify the notation, we will not mention EC_t^Y. It is understood in each revision considered here after.

We proceed by considering two revision processes that correspond to some data-conditioning and some source-conditioning, i.e., conditioning on an information relative to the data or to the source (Kruse and Gebhardt, 1993).

3.5. Data-conditioning.

The **data-conditioning** fits the scenario where You learn that a particular proposition L* of $\mathcal{L}$ is true. In that case, the hypothesis H that was proving in the context $[EC_t^Y]$ all those propositions in $\mathcal{L}$ proved by M(H) now proves in the revised context all those propositions in $\mathcal{L}$ proved by M(H)∧L*. The basic belief assignment $m_{\mathcal{L}}$ is revised into $m_{\mathcal{L}}^*$ with:

$$m_{\mathcal{L}}^*(L) = \sum_{H:H\in \mathcal{H}, M(H)\wedge L^*=L} p_{\mathcal{H}}(H) \qquad \text{if } L L^*$$

and $$m_{\mathcal{L}}^*(L) = 0 \qquad \text{otherwise}$$

Note that for all $L\in \mathcal{L}$ such that $L L^*$, one has $L\wedge L^*\equiv L$. The revised plausibility $pl_{\mathcal{L}}^*$ on $\mathcal{L}$ for all $L\in \mathcal{L}$ is:

$$pl_{\mathcal{L}}^*(L) = \sum_{X:X\in \mathcal{L}, X\wedge L\neq\perp} m_{\mathcal{L}}^*(X)$$

$$= \sum_{X:X\in \mathcal{L}, X\wedge L\neq\perp} \; \sum_{H:H\in \mathcal{H}, M(H)\wedge L^*=X} p_{\mathcal{H}}(H)$$

$$= \sum_{H:H\in \mathcal{H}, M(H)\wedge L\wedge L^*\neq\perp} p_{\mathcal{H}}(H) = pl_{\mathcal{L}}(L\wedge L^*)$$

One has for all $L\in \mathcal{L}$, $L\neq\perp$,

$$bel_{\mathcal{L}}^*(L) = pl_{\mathcal{L}}^*(\top) - pl_{\mathcal{L}}^*(\neg L) = pl_{\mathcal{L}}(L^*) - pl_{\mathcal{L}}(\neg L\wedge L^*) =$$

$$= bel_{\mathcal{L}}(\top) - bel_{\mathcal{L}}(\neg L^*) - bel_{\mathcal{L}}(\top) + bel_{\mathcal{L}}(L\vee\neg L^*)$$

$$= bel_{\mathcal{L}}(L\vee\neg L^*) - bel_{\mathcal{L}}(\neg L^*).$$

These relations for $bel_{\mathcal{L}}^*$ and $pl_{\mathcal{L}}^*$ correspond to the **unnormalized rule of conditioning** (Smets 1993b) Normalization is achieved by further conditioning on L being not equivalent to a contradiction. After normalization, the resulting relation is Dempster's rule of conditioning.

Example 3: Suppose You learn the extra information that Paul has decided without regard to the weather report to spend the evening with John. So L* = J. What can You say about Carol. Table 3 presents the result of the revision of the data of table 2 by this information. The data are obtained by replacing the L_i by $L_i\wedge J$ in table 1. In particular under H_3 (windy) You knew that Paul would spend his evening with Carol. You know that John will also be there. So You know now that under H_3, Paul will spend his evening with Carol and John. ∇

	bel_L *	pl_L *
$C \wedge J$	.18	.28
$\neg C \wedge J$	.45	.55
J	.73	.73

Table 3: Belief and plausibility values for several propositions of L after data-conditioning on L* = J.

3.6. Source-conditioning.

The **source-conditioning** fits to the following scenario. Given $L^{**} \in L$, You consider only those hypotheses H that prove L** (without proving ¬L**) and ask then what is the probability that L is provable, for $L \in L$. This revision corresponds to the impact on the evidential corpus of the information 'H proves L**', or equivalently 'L** is provable'. Therefore You restrict Your attention to those $H \in H$ such that $H \vdash L^{**}$, $H \nvdash \neg L^{**}$.

Let $L \vdash L^{**}$. So $L \wedge L^{**} = L$. Let $P_L^{**}(>L)$ be the probability that L is provable by one of those hypotheses that prove L** without proving ¬L**. It is:

$$bel_L^{**}(L) = \frac{P_H(\{H: H \vdash L \wedge L^{**}, H \nvdash \neg(L \wedge L^{**})\})}{P_H(\{H: H \vdash L^{**}, H \nvdash \neg L^{**})\}}$$

$$= \frac{bel_L(L \wedge L^{**})}{bel_L(L^{**})}$$

This relation is known as the **geometrical rule of conditioning** (Shafer 1976b, Suppes and Zanotti, 1977).

Example 4: Suppose You learn that H proves L** = J, i.e., You learn that the weather report (the source) was such that John will spend the evening with Paul (not that John just did it). What can You say about Carol? Only H_1 and H_2 remain relevant and :

$$bel_L^{**}(\neg C) = \frac{.15}{.10 + .15} = .60$$

$$bel_L^{**}(C) = .00$$

$$bel_L^{**}(C \vee \neg C) = 1.00.$$

The beliefs computed here are those given to these propositions considered in table 3. Resulting beliefs are different, enhancing the difference between the two forms of revisions. ∇

It might be worth reconsidering the conditioning events. In the first case (data conditioning), You revise Your belief on the extra information 'L* is true'. In the second case (the source conditioning), You revise Your beliefs on the fact 'L** is deducible from the hypotheses in *H*'. The first case seems the most natural form of conditioning, what does not mean it always applies. The second form is the one encountered more naturally in the random set approach to belief functions (Nguyen 1978, Smets 1991a, 1991b). It is also the one encountered by applying Bayes' rule of conditioning blindly. If one writes:

$$\text{bel}(A|B) = P(\ \ >A|\ \ >B),$$

then it becomes very tempting to write:

$$P(\ \ >A|\ \ >B) = \frac{P(\ \ >A \wedge\ \ \ >B)}{P(\ \ >B)} = \frac{P(\ \ >A\wedge B)}{P(\ \ >B)}$$

in which case,

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

The adequate choice between the two rules is straightforward when the conditioning event is well understood. The important point is not to apply Bayes' rule blindly, as the geometrical rule would automatically result, even when Dempster's rule would be the adequate rule.

The data conditioning (and its derived Dempster's rule of conditioning) is the case classically considered in Dempster-Shafer theory and in the transferable belief model, where conditioning is performed on a new piece of evidence related to the data domain *L*.

The normalized Dempster's rule of conditioning as described in Dempster-Shafer theory can be explained as the result of a combination of both the source and the data conditioning. The conditioning event L* concerns the data domain *L*, and induces a data conditioning, i.e., the unnormalized rule of conditioning is the appropriate rule. But atop of such conditioning, there is a supplementary hidden permanent assumption held by You in Your evidential corpus EC_t^Y: every hypothesis H in *H* should be compatible with some proposition of *L* at any time, i.e., after conditioning on any L* one may keep only those H in *H* such that $M(H)\wedge L^* \neq \perp$. The other hypothesis are 'eliminated' through a source-conditioning, i.e., they get a zero probability, or equivalently are considered as impossible under the revised EC_t^Y.

So given an initial probability measure P_{H} on *H* (such that $P_{H}(\{H: M(H) = \perp\}) = 0$ as the initial probability measure must also satisfy the hidden assumption), the impact of a data conditioning on L* is obtained as follows:

1° determine $H^* = \{H : M(H)\wedge L^* = \perp\}$

2° compute P_{H}^* on H* by conditioning P_{H} on H* (by Bayes' rule)

3° condition on L* with Dempster's rule of conditioning, using P_{H}^* as the probability measure on the hypothesis space.

The results are normalized by construction. The order of the conditioning is irrelevant as the same results are obtained by first applying the unnormalized rule of conditioning (a data-conditioning) on L* and then applying the hidden assumption, i.e. applying the geometrical rule of conditioning (a source-conditioning) on H*. This explanation might help in understanding the origin of the normalization factor after conditioning as always advocated by Shafer. It also enhance the presence of the hidden assumption in EC_t^Y that could be relaxed, as done in the transferable belief model where we do not apply the normalization automatically, but only when explicitly justified and required.

4. The probability case.

There are cases where the probability of provability happens to be a probability measure over $\mathcal{L}$ (i.e. $bel_{\mathcal{L}}$ happens to be a probability function).

Case 1: Suppose there is a partition Π of $\mathcal{L}$ such that for all H, {M(H)} is an element of Π iff $p_{\mathcal{H}}(H)>0$. In that case $bel_{\mathcal{L}}$ becomes an additive measure on 2^{Π} (a probability measure if it were normalized). When all belief functions are normalized, the whole model collapses into a classical probability model and both revision cases degenerate into the classical Bayes' rule of conditioning.

Case 2: Going back to the definition of $P_{\mathcal{L}}(\ >L)$, suppose we add the following requirement on the provability relation :

$$\text{If } HL\vee L' \text{ then } H\ \ L \text{ or } H\ \ L'. \qquad (4.1)$$

This relation is the disjunctive property of intuitionistic logic. Here we require that the provability relation be so restricted.

Under the requirement (4.1),

$$\begin{aligned}
P_{\mathcal{L}}(\ >L\vee L') &= P_{\mathcal{H}}(\{H: H\in \mathcal{H}, HL\vee\ \ L', H\nvdash\neg L\vee L'\}) = \\
&= P_{\mathcal{H}}(\{H: H\in \mathcal{H}, HL \text{ or } H\ \ L', H\ \ \nvdash\bot\}) \\
&= P_{\mathcal{H}}(\{H: H\in \mathcal{H}, HL, H\ \ \nvdash\bot\}\cup\{H: H\in \mathcal{H}, HL', H\ \ \nvdash\bot\}) \\
&= P_{\mathcal{H}}(\{H: H\in \mathcal{H}, HL, H\ \ \nvdash\bot\}) + P(\{H: H\in \mathcal{H}, HL', H\ \ \nvdash\bot\}) \\
&\qquad - P_{\mathcal{H}}(\{H: H\in \mathcal{H}, HL\wedge\ \ L', H\nvdash\bot\}) \\
&= P_{\mathcal{L}}(\ >L) + P_{\mathcal{L}}(\ >L') - P_{\mathcal{L}}(\ >L\wedge L')
\end{aligned}$$

Therefore,

$$bel_{\mathcal{L}}(L\vee L') = bel_{\mathcal{L}}(L) + bel_{\mathcal{L}}(L') - bel_{\mathcal{L}}(L\wedge L').$$

This relation means that bel is an additive measure. It is not yet a probability function as $bel_{\mathcal{L}}(\top) = 1$ is not required. $bel_{\mathcal{L}}(\top) = 1$ corresponds to the extra requirement that $P_{\mathcal{L}}(>\top) = 1$, i.e. for every $H \in \mathcal{H}$ such that $p_{\mathcal{H}}(H)>0$ one has $M(H) \neq \perp$, or equivalently $p_{\mathcal{H}}(\perp)=0$. This property might easily be assumed, but is not fundamentally required.

The property (4.1) can be shown to be equivalent to the requirement that M(H) is an atom of $\mathcal{L}$ (where $\mathcal{L}$ is viewed as a lattice, i.e., an atom of $\mathcal{L}$ is a proposition that corresponds to an interpretation of $\mathcal{L}$). Indeed let $L = l_1 \vee l_2 \vee .. \vee l_n$ where the l_i 's are atoms of $\mathcal{L}$. Then if HL, then by 4.1 there is a l_i such that $M(H) = l_i$. It is interesting to note that the requirement (4.1) is equivalent to both the additivity of $bel_{\mathcal{L}}$ and the fact that for every $H \in \mathcal{H}$, M(H) is an atom of $\mathcal{L}$.

5. Conditioning by adding a new hypothesis.

One might be tempted to find a revision rule when the conditioning event results in adding a new hypothesis H* to EC_t^Y that must be accepted. So the probability initially given under EC_t^Y to the hypothesis H would be given to $H \wedge H^*$ after revision of EC_t^Y. Unfortunately, the revised beliefs induced on $\mathcal{L}$ cannot be derived from the beliefs held on $\mathcal{L}$ before learning about H*. Computation must be completely redone, starting from the initial $P_{\mathcal{H}}$.

Example 5: As an example, consider example 1 and let H* = W (You know and Paul knows that it will be windy). We do not say that the weather report will say ‘it will be windy’, what would result in a conditioning of $p_{\mathcal{H}}$ on $H_3 \vee H_4$ by the classical Bayesian rule. Table 4 presents the result of transforming the data of table 1 by adding the hypothesis H*=W to every $H_i \in H$. The difficulty in computing the revised $bel_{\mathcal{L}}$ from the initial $bel_{\mathcal{L}}$ is due to the fact that in general $M(H_i \wedge H^*)$ is not a function of $M(H_i)$ and $M(H^*)$. ∇

$H_i \in \mathcal{H}$		$H_i \wedge H^*$	$M(H_i \wedge H^*)$		$M(H_i) \wedge M(H^*)$
H_1	$\top$	W	L_3	C	$C \wedge J$
H_2	R	$R \wedge W$	L_4	$C \wedge \neg J$	$\perp$
H_3	W	W	L_3	C	C
H_4	$R \wedge W$	$R \wedge W$	L_4	$C \wedge \neg J$	$C \wedge \neg J$
H_5	$R \wedge \neg W$	$\perp$	L_6	$\perp$	$C \wedge \neg J$
H_6	$R \wedge \neg R$	$\perp$	L_6	$\perp$	$\perp$

Table 4: Transformation of table 1 data by the added hypothesis H*=W. Elements of $\mathcal{H}$, their transformation by H*, the correct related L value with $L = M(H_i \wedge H^*)$ and the value of $M(H_i) \wedge M(H^*)$.

Suppose the special case where M satisfies $M(H_1 \wedge H_2) = M(H_1) \wedge M(H_2)$ for all H_1 and $H_2 \in \mathcal{H}$ (a constraint not satisfied in the example of table1: e.g., $L_4 = C \wedge \neg J = M(R \wedge W) \neq M(R) \wedge M(W) = L_2 \wedge L_3 = \bot$). Under such a condition (modulo the 'and' and 'right weakening' of section 3.1), one has:

If $H_1 L_1$ and $H_2 L_2$, then $H_1 \wedge H_2 L_1 \wedge L_2$.

that would be obtained if the operator satisfies also 'monotonicity' (if $H_1 L$ then $H_1 \wedge H_2 L$). In that case M satisfies $M(H) = M(H) \wedge M(\top)$ for every $H \in \mathcal{H}$, and $M(H) \wedge M(H') = \bot$ only if $H \wedge H' = \bot$. From these properties, one has $bel_{\mathcal{L}}(L) = P_{\mathcal{H}}(\{H': H' \wedge H = H', H' \neq \bot\})$ where $L = M(H)$. The revision $pl_{\mathcal{L}}*$ of $pl_{\mathcal{L}}$ induced by the acceptation of a new hypothesis H* can now be obtained directly from $pl_{\mathcal{L}}$

$pl_{\mathcal{L}}*(L) = pl_{\mathcal{L}}(L)$ for $L \wedge M(H^*) = L$,

i.e., the result is the one obtained by conditioning the initial plausibility function $pl_{\mathcal{L}}$ on $L^* = M(H^*)$ with the unnormalized rule of data-conditioning. This relation does not seem very useful as the constraints on M are quite artificial. Nevertheless this example may help to understand the difference between the two types of conditioning analyzed in this paper.

When the M mapping satisfies the above constraints ('and' and 'monotonicity'), both source-conditioning and data-conditioning can be performed by appropriately transferring the probabilities $p_{\mathcal{H}}$ among the hypotheses. The data-conditioning on $L^* \in \mathcal{L}$ can be realized by accepting the hypothesis H* where $M(H^*) = L^*$. The source-conditioning on L** can be realized by keeping only those hypotheses H such that $M(H) \wedge L^{**} = M(H)$. If one defines the (maybe new) hypothesis H** such that $M(H^{**}) = L^{**}$, then source-conditioning is obtained by keeping only those hypotheses H such that $H \wedge H^{**} = H$ (and normalizing). The data-conditioning is obtained by allocating $p_{\mathcal{H}}(H)$ to $H \wedge H^*$ whereas the source-conditioning is obtained by keeping $p_{\mathcal{H}}(H)$ to H iff $H \wedge H^{**} = H$, otherwise $p_{\mathcal{H}}(H)$ is transferred to $\bot$ (and normalization will take care of it). The essential difference between the two conditioning rules is found for those hypotheses compatible with both the conditioning hypothesis and its negation: they are kept in the data-conditioning and eliminated in the source-conditioning. Source-conditioning results in a much more radical revision than the data-conditioning.

6. Conclusions.

We have presented the probability of provability interpretation of Dempster-Shafer theory. We have derived the revision rules that can be described under that interpretation. Both Dempster's rule of conditioning and the geometric rule of conditioning are obtained. These derivations cover the cases considered by Dempster-Shafer theory (Dempster 1967), but not all those considered by the transferable belief model (Smets and Kennes 1990) where we do not necessarily assumed a probability measure on some hypothesis space.

Dempster-Shafer theory has been criticized by some Bayesians as inappropriate: they claim that conditioning by Dempster's rule of conditioning is inadequate. A strict Bayesian would claim the existence of a probability measure $P_{\mathcal{H} \times \mathcal{L}}$ on the product space $\mathcal{H} \times \mathcal{L}$ that represents the agent's belief on that space[6]. They ask for the application of Bayes' rule of conditioning on $P_{\mathcal{H} \times \mathcal{L}}$, and the marginalization of the result on $\mathcal{L}$. Of course, the available information consists only of the marginalization of $P_{\mathcal{H} \times \mathcal{L}}$ on $\mathcal{H}$. The conditioning process cannot be achieved in general for lack of appropriate information. Only upper and lower conditional probabilities can be computed (Fagin and Halpern, 1990, Jaffray, 1992). Bayesians conclude that Dempster's rule of conditioning is inappropriate (Levi, 1983) what is exact once their preliminary assumption (the probability measure on $\mathcal{H} \times \mathcal{L}$ that represents the agent's belief on that space) is accepted.

To get out of the Bayesian criticisms, it is sufficient to reject the probability measure on the product space $\mathcal{H} \times \mathcal{L}$ i.e. to reject the Bayesian dogma that there exists a probability measure representing the agent's beliefs on ANY and EVERY space. Not to assume such a probability measure on $\mathcal{H} \times \mathcal{L}$ represents the agent's belief is what is done explicitly in the Hints' model of Kohlas (and implicitly in Dempster-Shafer theory). In the transferable belief model we even go further by not requiring the existence of any hypothesis space $\mathcal{H}$ and considering ONLY the $\mathcal{L}$ space by itself, cutting therefore all links with the Dempster-Shafer model.

In the transferable belief model, we consider only the basic belief assignment $m_{\mathcal{L}}$ and its related belief function $bel_{\mathcal{L}}$ and plausibility function $pl_{\mathcal{L}}$. No concept of some $\mathcal{H}$ space endowed with a probability measure is needed. The meaning of $m_{\mathcal{L}}(L)$ for $L \in \mathcal{L}$ is that $m_{\mathcal{L}}(L)$ is the part of belief allocated to L and that could be allocated to any proposition L' that proves L if further information justifies such transfer. Dempster's rule of conditioning is directly introduced as it is part of the overall description of the transferable belief model. The geometrical rule is derived if one asks for the proportion of the belief that supports $L \in \mathcal{L}$ given it supports $L^{**} \in \mathcal{L}$. Both rules have also been derived axiomatically in Smets (1993a) while looking for quantitative representations of credibility in general.

Acknowledgments.
The author is indebted to Ph. Besnard and A. Saffiotti for many suggestions about this work.

---

[6] Of course, mathematically, one can always construct many probability measures on $\mathcal{H} \times \mathcal{L}$. The question is: does one of these probability measures represent an agent's belief or not? Bayesians claim it does.

References.

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

DUBOIS D. and PRADE H. (1991) Focusing versus Updating in Belief Function Theory. Techn. Rep. IRIT. To appear in Dempster-Shafer theory of evidence. Fedrizzi M., Kaczprzyk J. and Yager R. eds, Wiley.

DUBOIS D. and PRADE H. (1992) Belief revision and updates in numerical formalism. Tech. Rep. IRIT.

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

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

GARDENFORS P. (1988) Knowledge in flux. Modelling the dynamics of epistemic states. MIT Press, Cambridge, Mass.

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

JEFFREY R.C. (1965) The logic of decision. McGraw-Hill, (1983) 2nd Ed. Univ. Chicago Press, Chicago.

KATSUNO H. and MENDELZON A.O. (1991) On the difference between updating a knowledge base and revising it. Proc. 2nd. Inter. Conf. Principle of Knowledge Representation and Reasoning (KR-91), J.Allen, R. Fikes and E. Sandewall eds). Morgan-Kaufman, Cambridge, Mass. 387-394.

KOHLAS J. and MONNEY P. A. (1990) Modeling and reasoning with hints. Technical Report. Inst. Automation and OR. Univ. Fribourg.

KRUSE R. and GEBHARDT J. (1993) Updating mechanisms for imprecise data. DRUMS II workshop on Belief Change. Dubois D. and Prade H. Eds., IRIT, Univ. Paul Sabatier, Toulouse, pg. 8-22.

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. p. 27-43.

NGUYEN H.T. (1978) On random sets and belief functions J. Math.Anal. Applic. 65:531-542.

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

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. (1981) Jeffrey's rule of conditioning. Philosophy of Sciences, 48:337-362.

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

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

SMETS P. (1991a) About updating. in D'Ambrosio B., Smets P., and Bonissone P.P. eds, Uncertainty in AI 91, Morgan Kaufmann, San Mateo, Ca, USA, 1991, 378-385.

SMETS P. (1991b) What is Dempster-Shafer's model? To appear in Dempster-Shafer theory of evidence. Fedrizzi M., Kacprzyk J. and Yager R. eds, Wiley.

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

SMETS P. (1993a) An axiomatic justification for the use of belief function to quantify beliefs. IJCAI-93. Chambery. pg. 598-603.

SMETS P. (1993b) Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approximate Reasoning.

SMETS P. and KENNES R. (1990) The transferable belief model. Technical Report: IRIDIA-TR-90-14. To appear in Artificial Intelligence.

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