On the mathematical theory of evidence and Dempster's rule of combination

Jean Dezert, Albena Tchamova, Frederic Dambreville

On the mathematical theory of evidence and Dempster's rule of combination

Jean Dezert[1], Albena Tchamova[2], Frédéric Dambreville[3]
1. Onera - The French Aerospace Lab, F-91761 Palaiseau, France,
Email: jean.dezert@onera.fr
2. Institute for Information and Communication Technologies, Bulgarian Academy of Sciences,
"Acad.G.Bonchev" Str., bl.25A, 1113 Sofia, Bulgaria,
Email: tchamova@bas.bg
3. DGA/Ingénierie de Projet/EORD,7–9 rue des Mathurins, 92221 Bagneux Cedex, France,
Email: general@fredericdambreville.com

***Abstract*—In this paper we present an analysis of the use of Dempster's rule of combination, its consistency with the probability calculus and its usefulness for combining sources of evidence expressed by belief functions in the framework of the Mathematical Theory of Evidence, known also as Dempster-Shafer Theory (DST), or as the classical theory of belief functions. We show that the direct combination of original basic belief assignments of sources of evidence with Dempster's rule is inconsistent with probability calculus and we explain this from a simple experimental protocol. We also show how Dempster's rule can be reconciled with probability calculus for such examples if the combination of sources is done differently. In that case the notion of conflict between sources of evidence becomes meaningless (it just vanishes) because Dempster's rule coincides with the pure conjunctive operator, and the entire fusion process becomes comparable to simple application of Total Probability Theorem (i.e. the weighted average fusion rule). While a direct application of Dempster's rule becomes questionable in the most general case, then arises the need of a methodoloy for organizing and implementing the combination rules with respect to the applications.**

Keywords: Dempster-Shafer Theory, Mathematical Theory of Evidence, Generalized Bayesian Inference, belief functions.

I. Introduction

The Mathematical Theory of Evidence developed by Shafer in [9] from Dempster's works [1]–[3] is considered as one of the main paradigms for reasoning under uncertainty thanks to the elegant introduction of belief functions. This theory however has been a source of many debates since its development both on the interpretation (semantics) of belief functions and on its efficiency for combining or conditioning sources of evidence in order to provide pertinent results for real applications. Here we show through very simple examples and experimental protocol why there is problem when applying directly Dempster's rule of combination as proposed originally in Shafer's milestone book [9]. We want also to clarify some important points about its usefulness with respect to what can be obtained from basic probability calculus. We assume the reader already familiar with DST and we just recall the minimal basics required for having a self-contained presentation. Readers not familiar with DST must refer to [9] for more details.

Let's consider a frame of discernment $\Theta = \{\theta_1, \ldots, \theta_n\}$ which corresponds to a finite set of n exhaustive and exclusive elements. A basic belief assignment (bba) associated to a given source/body of evidence $\mathcal{B}$ is defined [9] as a mapping $m_{\mathcal{B}}(.) : 2^{\Theta} \rightarrow [0,1]$:

$$m_{\mathcal{B}}(\emptyset) = 0 \qquad \text{and} \qquad \sum_{X \in 2^{\Theta}} m_{\mathcal{B}}(X) = 1 \tag{1}$$

The elements of the power set 2^{Θ} having a strict positive mass of belief are called focal elements of $m_{\mathcal{B}}(.)$. The set of all focal elements is called the core of $m_{\mathcal{B}}(.)$ and is denoted $\mathcal{K}_{\mathcal{B}}$. Among all possible bba's one can define on 2^{Θ}, there exists a very particular one called in DST the vacuous bba denoted $m_{\mathcal{V}}(.)$ that plays an essential role in DST since from DS point of view it represents adequately a total ignorant source of evidence. $m_{\mathcal{V}}(.)$ commits all the mass of belief to the whole frame Θ, i.e. $m_{\mathcal{V}}(\Theta) = 1$. In DST, several other belief functions (BF) have also been introduced, mainly the credibility function (also called sometimes support or balance function) $Bel_{\mathcal{B}}(.)$ and plausibility function $Pl_{\mathcal{B}}(.)$ which are in one-to-one correspondance with the bba $m_{\mathcal{B}}(.)$. $Bel_{\mathcal{B}}(.)$ and $Pl_{\mathcal{B}}(.)$ are respectively interpreted as lower and upper bounds of unknown probability measures defined on Θ and compatible with the bba $m_{\mathcal{B}}(.)$. They are mathematically defined by:

$$\text{Bel}_{\mathcal{B}}(X) \triangleq \sum_{\substack{Y \subseteq X \\ Y \in 2^{\Theta}}} m_{\mathcal{B}}(Y) \quad \text{and} \quad \text{Pl}_{\mathcal{B}}(X) \triangleq \sum_{\substack{Y \cap X \neq \emptyset \\ Y \in 2^{\Theta}}} m_{\mathcal{B}}(Y) \tag{2}$$

Dempster's rule of combination, also called Dempster-Shafer rule by some authors (or just "DS rule" for short), denoted with $\oplus$ symbol, is the key operator in DST proposed for combining (or conditioning) several bba's associated to distinct bodies of evidences. Dempster's rule corresponds to a normalized conjunctive operation. The combination of two distinct (independent) sources of evidences characterized by the bba's $m_{\mathcal{B}_1}(.)$ and $m_{\mathcal{B}_2}(.)$ defined on the same common frame of discernment Θ is done, according to DS reasoning, with Dempster's rule as follows: for the empty set $m^{DS}_{(\mathcal{B}_1\mathcal{B}_2)}(\emptyset) = 0$, and $\forall X \in 2^{\Theta} \setminus \{\emptyset\}$ one has

$$m^{DS}_{(\mathcal{B}_1\mathcal{B}_2)}(X) = [m_{\mathcal{B}_1} \oplus m_{\mathcal{B}_2}](X) = \frac{m_{12}(X)}{1 - m_{12}(\emptyset)} \quad (3)$$

where

$$m_{12}(X) \triangleq \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_{\mathcal{B}_1}(X_1) m_{\mathcal{B}_2}(X_2) \quad (4)$$

corresponds to the conjunctive consensus on X between the two sources of evidence $\mathcal{B}_1$ and $\mathcal{B}_2$. Shafer defines the *degree of conflict* between the sources of evidence by

$$K_{12} \triangleq m_{12}(\emptyset) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = \emptyset}} m_{\mathcal{B}_1}(X_1) m_{\mathcal{B}_2}(X_2) \quad (5)$$

When $m_{12}(\emptyset) = 1$ the two sources are said in total conflict and they cannot be combined by Dempster's rule because of 0/0 indeterminacy [9]. Dempter's rule is commutative and associative which makes it very attractive since the combinations of sources can be done sequentially instead globally and the order doesn't matter. For example with three sources, all the following combinations provide same results with DS rule for any $X \in 2^\Theta$

$$\begin{aligned}[m_{\mathcal{B}_1} \oplus m_{\mathcal{B}_2} \oplus m_{\mathcal{B}_3}](X) &= [(m_{\mathcal{B}_1} \oplus m_{\mathcal{B}_2}) \oplus m_{\mathcal{B}_3}](X) \\ &= [m_{\mathcal{B}_1} \oplus (m_{\mathcal{B}_2} \oplus m_{\mathcal{B}_3})](X) \\ &= [m_{\mathcal{B}_2} \oplus (m_{\mathcal{B}_1} \oplus m_{\mathcal{B}_3})](X) \quad (6)\end{aligned}$$

The conditioning of a bba $m_{\mathcal{B}}(.)$ by a conditional element $Z \in 2^\Theta \setminus \{\emptyset\}$, denoted $m_{\mathcal{B}}(.|Z)$ and proposed in DST, is obtained by Dempster-Shafer rule of combination of $m_{\mathcal{B}}(.)$ with the bba $m_Z(.)$ that is focused only on Z, i.e. such that $m_Z(Z) = 1$. For any element X of the power set 2^Θ this is mathematically expressed by

$$m_{\mathcal{B}}(X|Z) = [m_{\mathcal{B}} \oplus m_Z](X) = [m_Z \oplus m_{\mathcal{B}}](X) \quad (7)$$

It has been proved by Shafer [9] that this rule of conditioning expressed in terms of plausibility functions yields to the following formula:

$$Pl_{\mathcal{B}}(X|Z) = \frac{Pl_{\mathcal{B}}(X \cap Z)}{Pl_{\mathcal{B}}(Z)} \quad (8)$$

The latter formula is very similar to Bayes formula $P(X|Z) = P(X \cap Z)/P(Z)$.

In summary, the elegant way of representing uncertainty thanks to belief functions, the nice "appealing" properties of Dempster's rule and its apparent consistency with Bayes formula for conditioning include all the ingredients to make it very attractive for a wide community of researchers and engineers faced to problems where not only randomness is involved but also where epistemic uncertainty can be present. But does this suffice to state that we can use DS reasoning without problems as originally proposed in Shafer's book? We can't unfortunately answer positively to this question as proved in the following section. The first researcher who has expressed his doubts on the validity of Dempster's rule as originally proposed was Lotfi Zadeh in [11]–[13]. Since Zadeh's comments, many efforts have been done by researchers either to try to give a clear semantics to BF in DS reasoning, to justify Dempster's rule, or to start to develop "better" solutions to combine BF in order to circumvent counter-intuitive results put in light by Zadeh. Surprisingly, few research works seem have been done to verify the consistency of DS reasoning with the basic probability calculus from a strict application of the combination principle based on Dempster's rule. In nineties Pearl pointed out the question on the compatibility of DST with probability calculus in [7], [8], and more recently Gelman have highlighted also this question in [5]. Our recent research works done independently supports Pearl's and Gelman's analysis but are based on different approach with examples drawn from an experimental protocol in order to validate our conclusions. In the next section we show why there is an inconsistency of DS reasoning with probability calculus if one applies directly the combination of sources of evidence with Dempster's rule as originally proposed by Shafer in [9]. Therefore DS reasoning cannot find its legitimacy from the the probabiliy calculus if it is applied in the classical/original way. This is the origin of most of debates about the validity of Dempster's rule. The next section will explain what is wrong when applying Dempster's rule as suggested by Shafer for combining directly sources of evidence, and we will explain the consequences of such improper use of Dempster's rule. The section III will explain how and what to do to reconcile Dempster's rule with basic probability calculus in order to avoid to fall in the trap of inconsistent reasoning under uncertainty. The purpose of this paper is to put in light and alerting the scientific community of this problem in order to prevent miss-use of belief functions.

II. On the inconsistency of Dempster's rule

In this section we put in light the inconsistency of Dempster's rule of combination with the probability calculus if one applies Dempster's rule to combine directly (as usually done) the original bba's provided by the sources of evidences. We will prove the following statements on the basis of an experimental protocol:

- S1: The "conflict" between two sources is not properly managed by Dempster's rule.
- S2: Dempster's rule yields to inconsistent results with classical probability calculus.
- S3: The associativity property of Dempster's rule is inconsistent with the probability calculus.
- S4: The conjunctive operator of bba's is incompatible with probability calculus.
- S5: The "vacuous" belief function does count in the consistent fusion of bba's.

A. A simple generic example

Let's start with a simple generic exemple. Consider a frame of discernment with three elements only, say $\Theta = \{A, B, C\}$ satisfying Shafer's model, i.e. the elements of the frame

are truly exhaustive and exclusive. Let's consider two non-Bayesian bba's (i.e. bba's having not only singletons as focal elements) to combine as given in Table I where $a \in [0,1]$ and $b_1, b_2 > 0$ such that $b_1 + b_2 \in [0,1[$.

Focal elem. \ bba's	$m_{\mathcal{B}_1}(.)$	$m_{\mathcal{B}_2}(.)$
A	a	0
$A \cup B$	$1-a$	b_1
C	0	$1-b_1-b_2$
$A \cup B \cup C$	0	b_2

Table I
INPUT BBA'S $m_1(.)$ AND $m_2(.)$.

Naturally, one has assumed that these two bba's $m_{\mathcal{B}_1}(.)$ and $m_{\mathcal{B}_2}(.)$ satisfy Dempster's condition of independence of sources of evidence in order to apply Dempster's rule of combination as proposed by Shafer. Note that in this simple generic and parametric example the focal elements of $m_{\mathcal{B}_2}(.)$ are not nested (consonant) and there really does exist a non null apparent "conflict"[1] between two sources as it will be shown more explicitly in the derivations. When applying Dempster's rule of combination, one gets:

1) using at first the conjunctive operator:

$$m_{12}(A) = a(b_1 + b_2) \quad (9)$$

$$m_{12}(A \cup B) = (1-a)(b_1 + b_2) \quad (10)$$

$$K_{12} = m_{12}(\emptyset) = 1 - b_1 - b_2 \text{ (conflicting mass)} \quad (11)$$

2) and then after the normalization by $1 - K_{12} = b_1 + b_2$, the final result is as follows:

$$\begin{aligned} m^{DS}_{(\mathcal{B}_1\mathcal{B}_2)}(A) &= \frac{m_{12}(A)}{1-K_{12}} \\ &= \frac{a(b_1+b_2)}{b_1+b_2} = a \\ &= m_{\mathcal{B}_1}(A) \end{aligned} \quad (12)$$

$$\begin{aligned} m^{DS}_{(\mathcal{B}_1\mathcal{B}_2)}(A \cup B) &= \frac{m_{12}(A \cup B)}{1-K_{12}} \\ &= \frac{(1-a)(b_1+b_2)}{b_1+b_2} = 1-a \\ &= m_{\mathcal{B}_1}(A \cup B) \end{aligned} \quad (13)$$

Clearly, our statement (S1) is true since in such example Dempster's rule is indifferent to the level of conflict between the sources. Whatever the value of the conflict $K_{12} > 0$ is, one gets $m^{DS}_{(\mathcal{B}_1\mathcal{B}_2)}(.) = m_{\mathcal{B}_1}(.)$. This behavior is counter-intuitive because in this simple example one of the sources doesn't finally count (impact) in the DS result, although none of the sources are vacuous and there does exist a non-null conflict betwen them. Actually $K_{12} > 0$ can be either very low or very high depending of the choice one takes for parameters a, b_1 and b_2 but it doesn't matter because $(1 - K_{12})$ is automatically simplified by the normalization step as we see through our elementary calculus. It can be easily verified that the same problem occurs even if we take $b_1 = 0$, i.e. we consider only C and $A \cup B \cup C$ as focal elements of $m_{\mathcal{B}_2}(.)$ (i.e. when focal elements of $m_{\mathcal{B}_2}(.)$ are nested). Actually it is easy to construct an infinite family of examples generating such irrational behavior as we did already prove in [4]. Dempster's rule behavior goes against the common sense if we combine directly the original bba's of the sources by Dempster's rule.

Is it a serious flaw or not of Dempster-Shafer Theory? At least, the intuition usually behind the notion of conflict may be questionned. Whatever, from a such generic and abstract example like the previous one, it may be not obvious for the reader to be convinced by the positive answer we give to this question if one adopts the combination of sources as proposed originally in DST. We will show however in section III what solution we propose for combining bba's in consistent manner with Dempster's rule. So, let's make the previous example a bit more concrete now and let's go deeper in the analysis with the following basic experimental protocol.

B. Experimental protocol and concrete example

The context is as follows. We consider a future president election for a country with only three distinct candidates (to keep the example simple). Each candidate represents a political party. So we assume a candidate for the Left party (called L), for the Moderate party (called M) and for the Right party (called R), so that $\Theta = \{L, M, R\}$. Let consider also two distinct sets of people allowed to vote in such election: $\mathcal{W}$ is the set of women and $\mathcal{M}$ is the set of men and we assume $|\mathcal{W}| = |\mathcal{M}|$, i.e. there is the same number of women and men. $\mathcal{M}$ and $\mathcal{W}$ will in fact play the roles of two distinct bodies of evidence $\mathcal{B}_1$ and $\mathcal{B}_2$ in the sequel.

B.1 Experimental protocol

The experiment we propose is elementary and can be done easily in practice. It consists in two steps: step 1) one computes the bba's of the sources of evidences; and step 2) one uses the real consistent/compatible precisiated probability measures expressed by the sources to check if the original DS reasoning is consistent with probability calculus. Let's explain this in details. We call this protocol the "voting-intention & voting experiment" (VIVE for short).

- Step 1 (construction of bba's):
  Before the election day (say 6 months before by example), all people (men and women belonging to $\mathcal{M}$ and $\mathcal{W}$ respectively) are asked to give their preferences/vote intentions[2] among the three given candidates. We assume that they provide their answer independently of what other people do, say or think. What is important in step 1 is that men and women are allowed to provide not only one name of a candidate,

[1]Here we use the notion of "Conflict" with the same interpretation and definition given by Shafer in [9].

[2]i.e. their belief in who they think is the best candidate to become president for their country.

but they are free to give one, two or eventually three names depending on their own preferences. For example if a man or a woman declares $L \cup M$ as his/her voting intention it means that he/she will vote either for L or M. $L \cup M$ voting intention indicates that he/she has not yet taken his final choice 6 months before the election day but it means that he/she is sure that he will not vote for R in order to maintain a coherence between what he/she prefers and what he/she will finally decide at election day. So all men and women are asked to write on a sheet of paper only one element of $2^\Theta \setminus \{\emptyset\}$. The vote intentions for men and women are written on sheets of paper and are put into two separate boxes denoted $B_\mathcal{M}$ and $B_\mathcal{W}$. By opening the boxes and counting the ratio $n_\mathcal{M}(X)/|\mathcal{M}|$ where $n_\mathcal{M}(X)$ is the number of voting intentions committed to X, one can compute the bba assignment $m_\mathcal{M}(X)$ for any X in $2^\Theta \setminus \{\emptyset\}$. A similar procedure is done to get $m_\mathcal{W}(X)$ for any X in $2^\Theta \setminus \{\emptyset\}$ too. From $m_\mathcal{M}(.)$ and $m_\mathcal{W}(.)$ we are now able to compute the lower and upper bounds $[Bel_\mathcal{M}(.); Pl_\mathcal{M}(.)]$ and $[Bel_\mathcal{W}(.); Pl_\mathcal{W}(.)]$ of unknown probabilities $P_\mathcal{M}(.)$ and $P_\mathcal{W}(.)$ based only on voting intentions in boxes $B_\mathcal{M}$ and $B_\mathcal{W}$. The boxes $B_\mathcal{M}$ and $B_\mathcal{W}$ can be seen as two independent well-defined sources/bodies of evidence from which one can compute the combined bba's denoted[3] $m^{DS}_{\mathcal{MW}}(.)$ with Dempster's rule, i.e. $m^{DS}_{\mathcal{MW}}(.) = [m_\mathcal{M} \oplus m_\mathcal{W}](.)$ and estimate lower and upper probability bounds $[Bel^{DS}_\mathcal{M}(.); Pl^{DS}_\mathcal{M}(.)]$ given by considering the sources of evidence altogether.

- Step 2 (Precisiation of probabilities of sources of evidences and verification of DS reasoning consistency):
  At the election day, the same people (men and women) are asked to vote for only one candidate among the three candidates. No blank vote is allowed. The vote of each man and woman must be coherent, i.e. the name of the candidate he/she votes for has to be indeed included in the list he/she wrote in their voting intention. The men and women ballots are put into two separate boxes $V_\mathcal{M}$ and $V_\mathcal{W}$. From the ballots in each box and based on classical frequentist calculus of probabilities, we can now precisiate the original imprecise probabilties $P_\mathcal{M}(.)$ and $P_\mathcal{W}(.)$ of each sources of evidence that were known to satisfy $P_\mathcal{M}(.) \in [Bel_\mathcal{M}(.); Pl_\mathcal{M}(.)]$ and $P_\mathcal{W}(.) \in [Bel_\mathcal{W}(.); Pl_\mathcal{W}(.)]$ according to the voting intentions in boxes $B_\mathcal{M}$ and $B_\mathcal{M}$. Clearly if one draws randomly a ballot in the box $V_\mathcal{M}$, the probability of the outcome will follow $P_\mathcal{M}(.)$ and if one draws randomly a ballot in the box $V_\mathcal{W}$ the probability of the outcome will follow $P_\mathcal{W}(.)$. To test the consistency of Dempster's rule[4] with the probability calculus it suffices to put all the ballots of boxes $V_\mathcal{M}$ and $V_\mathcal{M}$ altogether into a single empty box $V_\mathcal{MW}$ and estimate the resulting/combined real probability $P^{real}_\mathcal{MW}(.)$ computed by the frequentist approach of probability calculus. So if one draws randomly a ballot in the box $V_\mathcal{MW}$ the probability of the outcome will be governed by $P_\mathcal{MW}(.)$. The consistency of original DS reasoning will be satisfied if and only if for any election result compatible with the expressed voting intentions, the following inequality is always verified

$$Bel^{DS}_\mathcal{MW}(.) \leq P^{real}_\mathcal{MW}(.) \leq Pl^{DS}_\mathcal{MW}(.) \qquad (14)$$

  If there exists at least one example where inequality (14) fails, then consequently Dempster's rule of combination as used originally in DST to combine sources of evidences is proved to be inconsistent with the probability calculus.

Of course, the VIVE protocol is not limited to the specific case that we took here with only 2 sources of evidence and with $|\Theta| = 3$. It can be applied in general for any number $k \geq 2$ of sources and any size of frame of discernment greater or equal to 2. The VIVE protocol is very general and can be used to construct any bba's and test consistency of reasonings with belief functions with probability calculus.

B.2 A simple concrete example

Let's take a simple concrete example, apply VIVE protocol and go a bit further in details in the analysis to prove the inconsistency of Dempster's rule result with probability calculus as proposed in DST. We take $\Theta = \{L, M, R\}$, and for simplicity[5] of derivations we take $|\mathcal{M}| = |\mathcal{W}| = 100$ and we assume that from voting intentions we have obtained the bba's in Table II to combine. The table II means that 45 men have declared to prefer to vote for L, 55 men have declare to prefer to vote either for L or for M, 98 women have declared to prefer to vote for R and only 2 women were fully uncertain on their preference. From these bba's one can compute easily credibilities and plausibilities of each candidates, i.e. the lower and upper bounds of the unknown underlying probabilities that will govern the election result. So one gets for each source considered separately the values given in Tables III and IV. The Table III indicates $P_\mathcal{M}(R) = 0$, i.e. R can never occur based only of men voting intentions.

When considering the sources of evidences altogether (based on a mixed population with $\mathcal{MW} = \mathcal{M} \cup \mathcal{W}$) and because $m^{DS}_\mathcal{MW}(.) = m_\mathcal{M}(.)$ in this example, more precisely $m^{DS}_\mathcal{MW}(L) = 0.45$ and $mDS_\mathcal{MW}(L \cup M) = 0.55$, one gets the following lower and upper bounds of unknown probability that is supposed to really govern the election outcome in VIVE protocol.

[3]One uses the index $\mathcal{MW}$ to denote the fact that the results take into account both $\mathcal{M}$ and $\mathcal{W}$.

[4]as done in DST; that is from the direct Dempster's combination of original bba's of each source.

[5]We take small sizes for men and women voting populations to simplify the presentation but the derivations can also be done with bigger population sizes without changing the conclusions of our analysis.

Focal elem. \ bba's	$m_{\mathcal{M}}(.)$	$m_{\mathcal{W}}(.)$
L	0.45	0
$L \cup M$	0.55	0
R	0	0.98
$L \cup M \cup R$	0	0.02

Table II
INPUT BBA'S OF SOURCES OF EVIDENCE.

Candidates \ Source	Men population $\mathcal{M}$
L	$P_{\mathcal{M}}(L) \in [Bel_{\mathcal{M}}(L); Pl_{\mathcal{M}}(L)] = [0.45; 1]$
M	$P_{\mathcal{M}}(M) \in [Bel_{\mathcal{M}}(M); Pl_{\mathcal{M}}(M)] = [0; 0.55]$
R	$P_{\mathcal{M}}(R) \in [Bel_{\mathcal{M}}(R); Pl_{\mathcal{M}}(R)] = [0; 0]$

Table III
LOWER AND UPPER BOUNDS OF PROBABILTITY FROM THE SOURCE $\mathcal{M}$.

Candidates \ Source	Women population $\mathcal{W}$
L	$P_{\mathcal{W}}(L) \in [Bel_{\mathcal{W}}(L); Pl_{\mathcal{W}}(L)] = [0; 0.02]$
M	$P_{\mathcal{W}}(M) \in [Bel_{\mathcal{W}}(M); Pl_{\mathcal{W}}(M)] = [0; 0.02]$
R	$P_{\mathcal{W}}(R) \in [Bel_{\mathcal{W}}(R); Pl_{\mathcal{W}}(R)] = [0.98; 1]$

Table IV
LOWER AND UPPER BOUNDS OF PROBABILTITY FROM THE SOURCE $\mathcal{W}$.

Lower & upper bounds \ Source	Mixed population $\mathcal{MW}$
$[Bel^{DS}_{\mathcal{MW}}(L); Pl^{DS}_{\mathcal{MW}}(L)]$	$[0.45; 1]$
$[Bel^{DS}_{\mathcal{MW}}(M); Pl^{DS}_{\mathcal{MW}}(M)]$	$[0; 0.55]$
$[Bel^{DS}_{\mathcal{MW}}(R); Pl^{DS}_{\mathcal{MW}}(R)]$	$[0; 0]$

Table V
LOWER AND UPPER BOUNDS OF PROBABILITIES OBTAINED FROM DEMPSTER'S RULE.

Therefore, according to DS reasoning, one should have

$$\begin{cases} P^{DS}_{\mathcal{MW}}(L) \in [0.45; 1] \\ P^{DS}_{\mathcal{MW}}(M) \in [0; 0.55] \\ P^{DS}_{\mathcal{MW}}(R) \in [0; 0] \end{cases} \tag{15}$$

Clearly from (15), one obtains $P^{DS}_{\mathcal{MW}}(R) = 0$, meaning that the outcome R can never occur (according to Dempster-Shafer interpretation) which is intuitively absurd because at least 98% of the women have declared their intentions to vote for R. Stated differently, women opinions do not count here in Dempster-Shafer reasoning since we get finally $m^{DS}_{\mathcal{MW}}(.) = m_{\mathcal{M}}(.)$. How such very counter-intuitive behavior of Dempster's rule can be justified?

To prove the mathematical inconsistency of DS reasoning with probability calculus let's examine what can happen in reality at the end of the election day when every man and woman has finally voted only for one candidate in accordance with their vote intentions so that the imprecise probability has been finally precisiated by the sources of evidences. We assume that blank votes are not allowed and all votes do count in the election result. So take by example the following possible outcome of the election:

• Assume that in men population $\mathcal{M}$, forty five (45) men have voted for L and among the fifty five (55) men who provided imprecise voting intentions, five (5) men did finally vote for L and the other fifty (50) men did vote for M. So we get using frequentist approach of probability calculus

$$\begin{cases} P^{real}_{\mathcal{M}}(L) = (45+5)/100 = 0.5 \\ P^{real}_{\mathcal{M}}(M) = 50/100 = 0.5 \\ P^{real}_{\mathcal{M}}(R) = 0/100 = 0 \end{cases} \tag{16}$$

We see that this probability measure $P^{real}_{\mathcal{M}}(.)$ is fully compatible/consistent with the bba $m_{\mathcal{M}}(.)$ because as expected from the Table III one has

$$\begin{cases} P^{real}_{\mathcal{M}}(L) = 0.5 \in [0.45; 1] \\ P^{real}_{\mathcal{M}}(M) = 0.5 \in [0; 0.55] \\ P^{real}_{\mathcal{M}}(R) = 0 \in [0; 0] \end{cases} \tag{17}$$

• Assume that in women population $\mathcal{W}$, ninety eight (98) women did vote for R, and among the two (2) other women who provided imprecise voting intention, one woman did really vote for L and the other did vote for M. So we get using basic probability calculus

$$\begin{cases} P^{real}_{\mathcal{W}}(L) = 1/100 = 0.01 \\ P^{real}_{\mathcal{W}}(M) = 1/100 = 0.01 \\ P^{real}_{\mathcal{W}}(R) = 98/100 = 0.98 \end{cases} \tag{18}$$

We see that this probability measure $P^{real}_{\mathcal{W}}(.)$ is also fully compatible/consistent with the bba $m_{\mathcal{W}}(.)$ because as expected from the Table IV one has

$$\begin{cases} P^{real}_{\mathcal{W}}(L) = 0.01 \in [0; 0.02] \\ P^{real}_{\mathcal{W}}(M) = 0.01 \in [0; 0.02] \\ P^{real}_{\mathcal{W}}(R) = 0.98 \in [0.98; 1] \end{cases} \tag{19}$$

• Let's now put all the ballots of $V_{\mathcal{M}}$ and $V_{\mathcal{W}}$ boxes together in a new empty box $V_{\mathcal{MW}}$ and compute $P^{real}_{\mathcal{MW}}(.)$ using frequentist approach. One obtains:

$$\begin{cases} P^{real}_{\mathcal{MW}}(L) = (45+5+1)/(100+100) = 0.255 \\ P^{real}_{\mathcal{MW}}(M) = (50+1)/(100+100) = 0.255 \\ P^{real}_{\mathcal{MW}}(R) = 98/(100+100) = 0.490 \end{cases} \tag{20}$$

These results make intuitively perfect sense because one knows that at least 98% of women wanted to vote for R based on the evidence drawn from $\mathcal{W}$, collected in $B_{\mathcal{W}}$ box, and expressed in $m_{\mathcal{W}}(.)$. Clearly Dempster-Shafer's upper and lower bounds $[Bel^{DS}_{\mathcal{MW}}(R); Pl^{DS}_{\mathcal{MW}}(R)] = [0; 0]$ do not include $P^{real}_{\mathcal{MW}}(R) = 0.49$. This proves the inconsistency of Dempster's rule with the probability calculus when combining directly the original bba's

$m_{\mathcal{M}}(.)$ and $m_{\mathcal{W}}(.)$ as proposed in DST. Note also that $P^{real}_{\mathcal{MW}}(L) = 0.255 \notin [Bel^{DS}_{\mathcal{MW}}(L); Pl^{DS}_{\mathcal{MW}}(L)] = [0.45; 1]$.

From such of very simple concrete example, we have really put in light a flaw of Dempster's rule when combining directly the original bba's of sources of evidences because Dempster's rule is not directly compatible with the calculus of a bba (or a probability) obtained when putting the pieces of evidences altogether. If we apply Demspter's rule to combine directly the original bba's of sources of evidences, we get inconsistent lower and upper bounds of imprecise probability as shown through VIVE. Therefore, our statement S2 holds and supports Pearl's doubts expressed in [7], [8].

B.3 Zadeh's example revisited

Let's revisit now Zadeh's example [12] but transposed with VIVE protocol. We take $\Theta = \{L, M, R\}$ and the following bba's drawn from men and women populations with $|\mathcal{M}| = |\mathcal{W}| = 100$:

Focal elem. \ bba's	$m_{\mathcal{M}}(.)$	$m_{\mathcal{W}}(.)$
L	0.9	0
M	0	0.9
R	0.1	0.1

Table VI
INPUT BBA'S FOR ZADEH'S EXAMPLE.

Following similar derivations as previously based on the basic probability calculus, one gets as consistent result

$$\begin{cases} P^{real}_{\mathcal{MW}}(L) = 90/200 = 0.45 \\ P^{real}_{\mathcal{MW}}(M) = 90/200 = 0.45 \\ P^{real}_{\mathcal{MW}}(R) = (10+10)/200 = 0.10 \end{cases} \quad (21)$$

whereas the direct Dempster's combination of $m_{\mathcal{M}}(.)$ with $m_{\mathcal{W}}(.)$ provides in such particular "highly conflicting" example $m^{DS}_{\mathcal{MW}}(R) = 1$, or equivalently $P^{DS}_{\mathcal{MW}}(R) \in [1; 1]$, which is against the common sense as pointed out by Zadeh three decades ago. The "paradox" comes from the use of Dempster's rule for combining directly the original bba's of the sources. Zadeh's paradox disappears when applying the probability calculus as explained in VIVE. Actually, the notion of "conflict" between sources appears totally meaningless when using the probability calculus for combining the sources of evidence.

Of course, it may be argued that our VIVE analysis is done with finite populations of small size so that the estimation of probabilities based on frequentist approach of probability estimation is not precise enough, but fundamentally we can make the same analysis with bigger sizes of populations and still our analysis remains valid. We just take here $|\mathcal{M}| = |\mathcal{W}| = 100$ to make the presentation and derivations simpler. Based on this analysis, Zadeh's doubts on the validity of Dempster's rule to combine directly original bba's of the sources are justified since the inconsistency of Dempster's rule with the probability calculus is proved.

C. How to combine bba's in a consistent manner?

From our previous analyses, one sees that the consistent "combination/fusion" within VIVE protocol of k equi-weighted[6] bba's[7] $m_1(.) = P_1(.)$, ..., $m_k(.) = P_k(.)$ must be done by the arithmetic mean of $P_i(.)$, for $i = 1, \ldots, k$. More precisely, the "fusionned" probability $P_{(123\ldots k)}(.)$ is obtained by

$$P_{(123\ldots k)}(.) = (1/k)P_1(.) + \ldots + (1/k)P_k(.) \quad (22)$$

Formula (22) corresponds to the application of Total Probability Theorem using all equal priors. It is easy to prove that the consistency of the results when combining sequentially (or by any grouping/clustering choice) the Bayesian sources of evidences is achieved with the proper weighted average of sources. The weighting factors are computed as the number of sources involved in a given group divided by the total number of sources. For example, if one considers 3 equi-weighted Bayesian bba's to combine, one will get when combining these probability measures altogether

$$P_{(123)}(.) = (1/3)P_1(.) + (1/3)P_2(.) + (1/3)P_3(.) \quad (23)$$

Now if we combine first $P_1(.)$ with $P_2(.)$ (assuming that one has only these two sources of evidence in hands), one gets

$$P_{(12)}(.) = (1/2)P_1(.) + (1/2)P_2(.) \quad (24)$$

Then if a third distinct source of evidence comes in and one wants to combine it with $P_{(12)}(.)$, the consistent probability calculus must be done as follows

$$P_{(12)3}(.) = (2/3)P_{(12)}(.) + (1/3)P_3(.) \quad (25)$$

in order to be consistent with $P_{(123)}(.)$ obtained when putting all the pieces of evidences together. Clearly the fusion of the 2 sources of evidence $P_{(12)}$ must doubly count with respect to the single third source.

Because non-Bayesian bba's $m_i(.)$, $i = 1, 2, \ldots, k$ can be interpreted as classical probabilities defined on the power-set of Θ (this is how one can obtain them from voting intentions in Step 1 of VIVE protocol), the consistent fusion rule of probability calculus (22), i.e. the Total Probability Theorem, can be applied also. This yields to the following simple consistent fusion rule for bba's:

$$m^{Aver}_{(123\ldots k)}(.) = (1/k)m_1(.) + \ldots + (1/k)m_k(.) \quad (26)$$

One sees that this simple fusion rule (arithmetic mean/average) has a low complexity with respect to most of previous combination rules developed so far (i.e. Dempster's rule and its alternatives) to combine directly the original bba's of the sources of evidence. The averaging technique had been discussed by Murphy in [6] but without clear conclusions,

[6]Equi-weighting of bba's means that sources of evidences are based on the same population size when transposing bba's in VIVE protocol

[7]For notation concision in formulas, we now use $m_i(.)$ instead of $m_{\mathcal{B}_i}(.)$ to denote the bba associated with a source of evidence $\mathcal{B}_i$ defined with respect to a given frame Θ, for $i = 1, 2, \ldots$

nor justifications of its real usefulness. In our analysis we give a strong and clear justification of arithmetic mean within the VIVE paradigm that is based on the consistency of the fusion rule with the probability calculus drawn form a direct application of the Total Probability Theorem.

It is easy to verify that the average fusion rule (26) provides consistent results for the concrete example given in B.2. Indeed from values of bba's given in Table II, one will get from (26):

$$\begin{cases} m^{Aver}_{(\mathcal{MW})}(L) = 0.45/2 = 0.225 \\ m^{Aver}_{(\mathcal{MW})}(L \cup M) = 0.55/2 = 0.275 \\ m^{Aver}_{(\mathcal{MW})}(R) = 0.98/2 = 0.49 \\ m^{Aver}_{(\mathcal{MW})}(L \cup M \cup R) = 0.02/2 = 0.01 \end{cases} \quad (27)$$

Therefore the lower and upper bounds of imprecise probabilities will be given by

$$\begin{cases} [Bel^{Aver}_{(\mathcal{MW})}(L); Pl^{Aver}_{(\mathcal{MW})}(L)] = [0.225; 0.51] \\ [Bel^{Aver}_{(\mathcal{MW})}(M); Pl^{Aver}_{(\mathcal{MW})}(M)] = [0; 0.276] \\ [Bel^{Aver}_{(\mathcal{MW})}(R); Pl^{Aver}_{(\mathcal{MW})}(R)] = [0.49; 0.50] \end{cases} \quad (28)$$

These (consistent) lower and upper bounds do now include the precisiated probability $P^{real}_{\mathcal{MW}}(.)$ given in (20) and drawn from step 2 of VIVE protocol, contrariwise to Dempster-Shafer's lower and upper bounds given in Table V.

D. On the associativity of a fusion rule

From our previous analysis and the Total Probability Theorem, it becomes clear that the appealing requirement of the associativity of a fusion rule cannot be obtained/consistent with probability calculus as shown through (25) when sources to combine express different opinions. To explain this easily, what is going wrong with associativity property requirement in the fusion is that the sources are not combined with same equal weight depending on how we do their combination.

For example, let's consider 3 independent sources of evidence to combine according to a VIVE interpretation, and defined on the same frame Θ with bba's $m_1(.)$, $m_2(.)$ and $m_3(.)$. We assume that sources have same weights. If we combine them altogether by Dempster's rules we get $m_{(123)}(.) = [m_1 \oplus m_2 \oplus m_3](.)$. Now if we combine $m_1(.)$ and $m_2(.)$ by Dempster's rule (assuming one has only these two sources of evidence in hands at first), one gets $m_{(12)}(.) = [m_1 \oplus m_2](.)$. When fusioning $m_{(12)}(.)$ with the third source m_3 through Dempster'rule both sources $m_{(12)}(.)$ (the previous combined bba) and $m_3(.)$ are still considered to have the same weight which is obviously wrong because $m_{(12)}(.)$ is based on more evidences than $m_3(.)$ in fact. As shown already in (25), if we restrict this example to Bayesian bba's case, i.e. with $m_1(.) = P_1(.)$, $m_2(.) = P_2(.)$ and $m_3(.) = P_3(.)$, $P_{(12)3}(.)$ cannot be computed just as $(1/2)P_{12} + (1/2)P_3$ because otherwise it becomes inconsistent with $P_{(123)}(.)$.

In summary, the associativity requirement of a fusion rule is incompatible with probability calculus and consequently any associative rule of combination will yield to inconsistent results in at least a particular case. Therefore the associativity property of Dempster's rule is inconsistent with the probability calculus when combining directly bba's of sources of evidence and our statement S3 holds.

E. On the conjunctive rule of combination

Here we show that the conjunctive operator is inconsistent with probability calculus when combining directly original bba's of sources of evidence. To prove this, let's take $\Theta = \{A, B\}$, Shafer's model and the two following bba's given in the Table VII that could be obtained from Step 1 of VIVE protocol.

Focal elem. \ bba's	$m_1(.)$	$m_2(.)$
A	0.7	0.2
$A \cup B$	0.3	0.8

Table VII
INPUT BBA'S TO COMBINE.

From the Table VII, the lower and upper bounds of unknown $P_1(.)$ and $P_2(.)$ probability measures are given by (thanks to credibility and plausibility functions):

$$\begin{cases} 0.7 \leq P_1(A) \leq 1 \\ 0 \leq P_1(B) \leq 0.3 \end{cases} \text{ and } \begin{cases} 0.2 \leq P_2(A) \leq 1 \\ 0 \leq P_2(B) \leq 0.8 \end{cases} \quad (29)$$

In this very simple example, there is no "conflict" between these two sources according to Dempster-Shafer reasoning. Dempster's rule coincides here with the pure conjunctive rule because there is no need of normalization since no "conflict" occurs. So one gets from the conjunctive fusion rule

$$\begin{cases} m^{DS}_{(12)}(A) = 0.7 \cdot 0.2 + 0.7 \cdot 0.8 + 0.2 \cdot 0.3 = 0.76 \\ m^{DS}_{(12)}(A \cup B) = 0.3 \cdot 0.8 = 0.24 \end{cases} \quad (30)$$

and the following lower and upper bounds

$$\begin{cases} 0.76 \leq P^{DS}_{(12)}(A) \leq 1 \\ 0 \leq P^{DS}_{(12)}(B) \leq 0.24 \end{cases} \quad (31)$$

With the consistent (average) fusion rule, one gets

$$\begin{cases} m^{Aver}_{(12)}(A) = (0.7 + 0.2)/2 = 0.45 \\ m^{Aver}_{(12)}(A \cup B) = (0.3 + 0.8)/2 = 0.55 \end{cases} \quad (32)$$

and the following lower and upper bounds

$$\begin{cases} 0.45 \leq P^{real}_{(12)}(A) \leq 1 \\ 0 \leq P^{real}_{(12)}(B) \leq 0.55 \end{cases} \quad (33)$$

It is not too difficult with VIVE protocol to get two probabilties $P_1(.)$ and $P_2(.)$ compatible with $m_1(.)$ and $m_2(.)$ and to apply basic probability calculus to see that indeed the real probability governing the experiment can be out of the lower and upper bounds given by the pure conjunctive rule. For example, if one considers the compatible probabilities $P_1(A) = 0.7$, $P_1(B) = 0.3$ and $P_2(A) = 0.2$, $P_2(B) = 0.8$ then the real "combined" probability obtained from $P_1(.)$ and $P_2(.)$ by probability calculus will be $P^{real}_{(12)}(A) = (0.7 + 0.2)/2 = 0.45$

and $P^{real}_{(12)}(B) = (0.3+0.8)/2 = 0.55 = 1 - P^{real}_{(12)}(A)$. These probabilities are out of the lower and upper bounds obtained with conjunctive rule. In summary, even if the sources to combine are *not* in conflict (in Dempster-Shafer sense), then Dempster's rule coinciding with the conjunctive rule remains inconsistent with probability calculus if we apply Dempster's rule to combine directly the bba's of the sources of evidence. Hence our statement S4 holds.

F. On the vacuous bba

As clearly shown already through VIVE protocol, all bba's do count (even the "vacuous" bba of course) if one uses the consistent fusion rule (average), hence statement S5 holds. An interesting property is that the consistent fusion rule is an idempotent rule because if $m_1(.) = m_2(.) = \ldots = m_k(.) = m(.)$ then $m^{Aver}_{(12\ldots k)}(.) = \frac{1}{k}\sum_{i=1}^{k} m_i(.) = \frac{1}{k} k \cdot m(.) = m(.)$. Stated otherwise in VIVE framework, even if all women (or men) express their total indifference in their voting intentions, their final votes (assumed to be valid and consistent with the intentions) will obviously count in the final result of the election and in derivation of lower and upper bounds of imprecise probabilities. In case of total indifference in voting intentions for both men and women populations, i.e. $m_{\mathcal{M}}(L \cup M \cup R) = m_{\mathcal{W}}(L \cup M \cup R) = 1$ any election result (outcome) is possible because all unknown underlying probabilities are totally imprecise. It is clear that only in this case Dempster's result coincides with consistent average fusion rule if it is used to combine directly original bba's of the sources. This remark is however of little interest because no information can be drawn from probability calculus, or from DS reasoning in such "pathological" case. Indeed, both rules provide $m^{DS}_{\mathcal{MW}}(L \cup M \cup R) = m^{Aver}_{\mathcal{MW}}(L \cup M \cup R) = 1$ as a result. Thus, S5 holds, and we do think that the definition and the interpretation of the "vacuous" bba proposed by Shafer should be rethinked in accordance to such result, and is related to the way the fusion process is organized.

III. On the reconcilement of Dempster's rule with probability calculus

A. General principle for reconcilement

In this section we show how the reconcilement of Dempster's rule with probability calculus can be achieved on VIVE examples if one combines the sources not directly from Dempster's combination of original bba's provided by the sources, but in a different manner to make it consistent with the probabilistic approach. It is inspired from the Total Probability approach for solving the VIVE problems. The principle of the reconcilement consists in the following steps:

- **Step 1** (Inputs of the fusion problem): We assume a given finite and discrete frame of discernment Θ satisfying Shafer's model, and $k \geq 2$ distinct bodies of evidence $\mathcal{B}_k$ providing their basic belief assignments $m_{\mathcal{B}_k}(.)$ on 2^{Θ}. The joint evidence space will be denoted $\mathcal{B} \triangleq \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$.
- **Step 2** (ballooning extension of bba's): In order to combine the sources of evidence in a consistent manner, one needs first to work in the Cartesian product (joint) space $\Theta_{\mathcal{B}} \triangleq \Theta \times \mathcal{B} = \Theta \times \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$ to combine the sources of evidences. Since each bba $m_{\mathcal{B}_i}(.)$, $i = 1, 2, \ldots, k$ does in fact only express the opinion/evidence in the marginal space $\Theta_{\mathcal{B}_i} \triangleq \Theta \times \{\mathcal{B}_i\}$, one needs to extend these bba's into the joint space $\Theta_{\mathcal{B}}$ in such a way that no extra information is brought and the subsets of $\Theta_{\mathcal{B}}$ do not receive more support than justified. The least committed bba [10] on $\Theta_{\mathcal{B}}$, such that its conditioning on $\Theta_{\mathcal{B}_i}$ is $m_{\mathcal{B}_i}(.)$, is given by the so-called "ballooning" extension, denoted $m_{\mathcal{B}_i}^{\Theta_{\mathcal{B}_i} \Uparrow \Theta_{\mathcal{B}}}(.)$ and defined as

$$m_{\mathcal{B}_i}^{\Theta_{\mathcal{B}_i} \Uparrow \Theta_{\mathcal{B}}}(X) = \begin{cases} m_{\mathcal{B}_i}^{\Theta_{\mathcal{B}_i}}(Y) \\ \text{if } Y \subseteq \Theta_{\mathcal{B}_i} \text{ s.t. } X = Y \cup \bar{\Theta}_{\mathcal{B}_i} \\ 0 \qquad \text{otherwise} \end{cases} \tag{34}$$

  where $\bar{\Theta}_{\mathcal{B}_i}$ is the complement of $\Theta_{\mathcal{B}_i}$ relative to the frame $\Theta_{\mathcal{B}}$.
- **Step 3** (combination): The "fusion" of ballooned bba's $m_{\mathcal{B}_i}^{\Theta_{\mathcal{B}_i} \Uparrow \Theta_{\mathcal{B}}}(.)$ given in (34) can now be done with Dempster's rule and one has

$$m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}(.) = [m_{\mathcal{B}_1}^{\Theta_{\mathcal{B}_1} \Uparrow \Theta_{\mathcal{B}}} \oplus \ldots \oplus m_{\mathcal{B}_k}^{\Theta_{\mathcal{B}_k} \Uparrow \Theta_{\mathcal{B}}}](.) \tag{35}$$

  Because of the ballooning extension principle, Dempster's rule coincides exactly with the conjunctive rule since no "conflict" occurs between the focal elements of these ballooned bba's. Consequently, the original interest of Dempster's rule to manage efficiently "conflicting" information appears quite limited.
- **Step 4** (optional): If one needs, one can also take into account (include) in the previous fusion formula (40) our particular prior belief on the sources represented by a given bba $m_0^{\Theta_{\mathcal{B}}}(.)$ expressing no specific information on Θ.
- **Step 5** (marginalization of resulting bba): To express the final resulting bba $m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}(.)$, compatible with the probability calculus, we project this bba into the original frame Θ. This is done by keeping only the components of focal elements of $m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}(.)$ belonging to Θ, and it will be denoted $m_{\mathcal{B}}^{\Theta_{\mathcal{B}} \downarrow \Theta}(.)$

B. Theoretical interpretation

The previously defined process should be compared to what is obtained from the theory of probability. The protocol VIVE is solved at a Bayesian level as follows:

$$P(y) = \sum_{x \in \{\mathcal{M}, \mathcal{W}\}} P(x)P(y|x) \quad \text{for } y \in \{L, M, R\}, \tag{36}$$

which may be depicted along the following steps:

- **Step a** The inputs of the fusion problem are given by $P(y|x)$ for $x \in \{\mathcal{M}, \mathcal{W}\}$ and $y \in \{L, M, R\}$,
- **Step b** The prior information on the voters are given by $P(x)$,

- **Step c** The fusion is done "conjunctively" on the joint space:
$$P(x,y) = P(x)P(y|x)\,, \tag{37}$$
- **Step d** The law is marginalized to y:
$$P(y) = \sum_{x\in\{\mathcal{M},\mathcal{W}\}} P(x,y)\,. \tag{38}$$

Actually, *Step a* should be compared to *Step 1* and *Step 2*, *Step b* to *Step 4*, *Step c* to *Step 3*, and *Step d* to *Step 5*. Now, let us apply these steps in details:

- **Step a** It is given the inputs, which are actually related to conditional informations, $m_{\mathcal{B}_i}(.) = m(.|\mathcal{B}_i)$ defined on $\Theta_{\mathcal{B}_i}$,
- **Step a'** For any i, the input $m_{\mathcal{B}_i}(Y)$ is prepared for fusion by means of *information conservative* mapping into the joint space $\Theta\times\mathcal{B}$. This information conservative mapping is obtained actually by the ballooning extensions $m_{\mathcal{B}_i}^{\Theta_{\mathcal{B}_i}\Uparrow\Theta_{\mathcal{B}}}(.)$. Indeed, the ballooning keeps the information conditionned by $\mathcal{B}_i$ but keeps also the vacuous information related to $\overline{\mathcal{B}_i}$,
- **Step b** It is given the prior $m_0(X)$ defined on $2^{\mathcal{B}}$,
- **Step b'** The prior m_0 is also prepared for fusion by means of *information conservative* mapping into the joint space $\Theta\times\mathcal{B}$. This mapping is obviously obtained as follows:
$$\left\{\begin{array}{l} m_0^{\Theta_{\mathcal{B}}}(\Theta\times X) = m_0(X) \quad \text{for } X\in 2^{\mathcal{B}}\,, \\ m_0^{\Theta_{\mathcal{B}}}(Y) = 0 \quad \text{otherwise} \end{array}\right. \tag{39}$$
- **Step c** All information are fused in the joint space $\Theta\times\mathcal{B}$:
$$m^{\Theta_{\mathcal{B}}}(.) = m_0^{\Theta_{\mathcal{B}}} \oplus [m_{\mathcal{B}_1}^{\Theta_{\mathcal{B}_1}\Uparrow\Theta_{\mathcal{B}}} \oplus \ldots \oplus m_{\mathcal{B}_k}^{\Theta_{\mathcal{B}_k}\Uparrow\Theta_{\mathcal{B}}}](.) \tag{40}$$
- **Step d** At last, the bba $m^{\Theta_{\mathcal{B}}}$ is marginalized resulting in a final answer:
$$m(X) = \sum_{Z\subset\Theta\times\mathcal{B}:\pi(Z)=X} m^{\Theta_{\mathcal{B}}}(Z), \tag{41}$$
where: $\pi(Z) = \{x\in\Theta,\ \exists y\in\mathcal{B},\ (x,y)\in Z\}$,
the function π being a simple projection from $2^{\Theta\times\mathcal{B}}$ to 2^{Θ}.

Now, we will show that this solution to the VIVE protocol is coherent with the mean rule derived from the Bayesian approach.

a) Property: Assume that m_0 is probabilistic, that is $\sum_{i=1}^k m_0(\{\mathcal{B}_i\}) = 1$. Then:
$$m(.) = \sum_{i=1}^k m_0(\{\mathcal{B}_i\}) m_{\mathcal{B}_i}(.) \tag{42}$$

Proof

First at all, it noticed thanks to the ballooning extension, that $m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}(.) = [m_{\mathcal{B}_1}^{\Theta_{\mathcal{B}_1}\Uparrow\Theta_{\mathcal{B}}} \oplus \ldots \oplus m_{\mathcal{B}_k}^{\Theta_{\mathcal{B}_k}\Uparrow\Theta_{\mathcal{B}}}](.)$ is such that:
$$\left\{\begin{array}{l} m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}\left(\bigcup_{i=1}^k (Y_i\times\{\mathcal{B}_i\})\right) = \prod_{i=1}^k m_{\mathcal{B}_i}(Y_i) \\ m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}(Z) = 0 \quad \text{otherwise} \end{array}\right. \tag{43}$$
for $Y_{1:k}\in(2^{\Theta})^k$, and $Z\in 2^{\Theta\times\mathcal{B}}$.
Then, by fusing with "probabilistic" $m_0^{\Theta_{\mathcal{B}}}$, it is obtained:
$$\left\{\begin{array}{l} m^{\Theta_{\mathcal{B}}}(Y_i\times\{\mathcal{B}_i\}) = m_0(\{\mathcal{B}_i\})\times \\ \qquad\qquad \sum_{Y_{1:i-1}} \sum_{Y_{i+1:k}} \prod_{j=1}^k m_{\mathcal{B}_j}(Y_j) \\ m^{\Theta_{\mathcal{B}}}(Z) = 0 \quad \text{otherwise.} \end{array}\right. \tag{44}$$
This is equivalently rewritten:
$$m^{\Theta_{\mathcal{B}}}(Y_i\times\{\mathcal{B}_i\}) = m_0(\{\mathcal{B}_i\}) m_{\mathcal{B}_i}(Y_i)\,. \tag{45}$$
At last, the marginalization implies for any $Y\in 2^{\Theta}$:
$$m(Y) = \sum_{i=1}^k m_0(\{\mathcal{B}_i\}) m_{\mathcal{B}_i}(Y)\,. \tag{46}$$

□□□

This property implies that the principle of reconcilement actually applies to the VIVE protocol.

C. Concrete example revisited

Let's apply the general principle described in the previous section to the concrete example given in section II-B where only two bodies of evidence are involved, i.e. $\mathcal{B}_1 = \mathcal{M}$ and $\mathcal{B}_2 = \mathcal{W}$. We follow the five aforementioned steps to make the presentation as simple as possible.

- **Step 1** (Inputs of the fusion problem): When expressing explicitly the frames $\Theta_{\mathcal{B}_1}$ and $\Theta_{\mathcal{B}_2}$, the original bba's given in Table II that we need to combine are given as in Tables VIII.

Focal elem. \ bba's	$m_{\mathcal{B}_1}^{\Theta_{\mathcal{B}_1}}(.)$	$m_{\mathcal{B}_2}^{\Theta_{\mathcal{B}_2}}(.)$
$(L,\mathcal{B}_1)$	0.45	0
$(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)$	0.55	0
$(R,\mathcal{B}_2)$	0	0.98
$(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	0	0.02

Table VIII
INPUT BBA'S OF THE SOURCES $\mathcal{B}_1$ AND $\mathcal{B}_2$.

- **Step 2** (ballooning extension of bba's): The ballooned bba's obtained from (34) are given in Tables IX and X.

Focal elem. \ bba's	$m_{\mathcal{B}_1}^{\Theta_{\mathcal{B}_1}\Uparrow\Theta_{\mathcal{B}}}(.)$
$(L,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	0.45
$(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	0.55

Table IX
BALLOONING OF THE BBA OF THE SOURCE $\mathcal{B}_1$.

- **Step 3** (combination): Applying the conjunctive rule (40) to combine $m_{\mathcal{B}_1}^{\Theta_{\mathcal{B}_1}\Uparrow\Theta_{\mathcal{B}}}(.)$ and $m_{\mathcal{B}_2}^{\Theta_{\mathcal{B}_2}\Uparrow\Theta_{\mathcal{B}}}(.)$, one gets the result of the Table XI.
- **Step 4**: Two simple choices are possible for using $m_0^{\Theta_{\mathcal{B}}}$

Focal elem. \ bba's	$m_{\mathcal{B}_2}^{\Theta_{\mathcal{B}_2}\Uparrow\Theta_{\mathcal{B}}}(.)$
$(R,\mathcal{B}_2)\cup(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(R,\mathcal{B}_1)$	0.98
$(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)\cup(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(R,\mathcal{B}_1)$	0.02

Table X
BALLOONING OF THE BBA OF THE SOURCE $\mathcal{B}_2$.

Focal elem. \ bba's	$m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}(.)$
$(L,\mathcal{B}_1)\cup(R,\mathcal{B}_2)$	$0.98\cdot 0.45=0.441$
$(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(R,\mathcal{B}_2)$	$0.98\cdot 0.55=0.539$
$(L,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	$0.02\cdot 0.45=0.009$
$(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	$0.02\cdot 0.55=0.011$

Table XI
CONJUNCTIVE COMBINATION OF BALLOONED BBA'S.

– If one doesn't assume specific prior information on the sources, one can take the uninformative[8] bba defined as

$$m_0^{\Theta_{\mathcal{B}}}((L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(R,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2))=1 \quad (47)$$

If one combines with the conjunctive rule this bba $m_0^{\Theta_{\mathcal{B}}}(.)$ with the resulting bba of Table XI, one still gets the same results as given in Table XI because all focal elements of $m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}(.)$ are included in the focal element $(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(R,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$ of $m_0^{\Theta_{\mathcal{B}}}(.)$. So one will have

Focal elem. \ bba's	$[m_0^{\Theta_{\mathcal{B}}}\oplus m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}](.)$
$(L,\mathcal{B}_1)\cup(R,\mathcal{B}_2)$	0.441
$(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(R,\mathcal{B}_2)$	0.539
$(L,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	0.009
$(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	0.011

Table XII
CONJUNCTIVE COMBINATION OF BALLOONED BBA'S AND $m_0^{\Theta_{\mathcal{B}}}(.)$

– If one assumes (as done in the probability calculus) that both sources of evidences have same weights, then one has to choose as prior bba

$$\begin{cases} m_0^{\Theta_{\mathcal{B}}}((L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)\cup(R,\mathcal{B}_1))=1/2 \\ m_0^{\Theta_{\mathcal{B}}}((L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2))=1/2 \end{cases} \quad (48)$$

If one combines with the conjunctive rule this bba $m_0^{\Theta_{\mathcal{B}}}(.)$ with the resulting bba of Table XI, one finally gets the results given in Table XIII.

Focal elem. \ bba's	$[m_0^{\Theta_{\mathcal{B}}}\oplus m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}](.)$
$(L,\mathcal{B}_1)$	$0.441/2+0.009/2=0.225$
$(L,\mathcal{B}_1)\cup(M,\mathcal{B}_1)$	$0.539/2+0.011/2=0.275$
$(R,\mathcal{B}_2)$	$0.441/2+0.539/2=0.49$
$(L,\mathcal{B}_2)\cup(M,\mathcal{B}_2)\cup(R,\mathcal{B}_2)$	$0.009/2+0.011/2=0.01$

Table XIII
CONJUNCTIVE COMBINATION OF BALLOONED BBA'S AND $m_0^{\Theta_{\mathcal{B}}}(.)$

[8]called vacuous in DST terminology.

- **Step 5** (marginalization of resulting bba): Here are the final marginalized results based on two possible choices for prior bba's $m_0^{\Theta_{\mathcal{B}}}(.)$ of the sources
  – If one takes the uninformative bba (47), one finally obtains

Focal elem. \ bba's	$[m_0^{\Theta_{\mathcal{B}}}\oplus m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}]^{\downarrow\Theta}(.)$
$L\cup R$	0.441
$L\cup M\cup R$	$0.539+0.009+0.011=0.559$

Table XIV
FINAL FUSION RESULT MARGINALIZED ON Θ

From the Table XIV, one finally gets the following lower and upper bounds of the underlying unknown probabilities

$$\begin{cases} [Bel_{(\mathcal{B}_1\mathcal{B}_2)}(L);Pl_{(\mathcal{B}_1\mathcal{B}_2)}(L)]=[0;1] \\ [Bel_{(\mathcal{B}_1\mathcal{B}_2)}(M);Pl_{(\mathcal{B}_1\mathcal{B}_2)}(M)]=[0;0.559] \\ [Bel_{(\mathcal{B}_1\mathcal{B}_2)}(R);Pl_{(\mathcal{B}_1\mathcal{B}_2)}(R)]=[0;1] \end{cases} \quad (49)$$

These bounds do now include the precisiate probability (20) but are very wide due to uninformative prior $m_0^{\Theta_{\mathcal{B}}}(.)$ taken here for the sources, which is normal in such case.
  – If one takes the bba $m_0^{\Theta_{\mathcal{B}}}(.)$ as in (48) to be consistent with the probability calculus, one finally obtains from the Table XIII, the following marginalized result on Θ:

Focal elem. \ bba's	$[m_0^{\Theta_{\mathcal{B}}}\oplus m_{\mathcal{B}}^{\Theta_{\mathcal{B}}}]^{\downarrow\Theta}(.)$
L	0.225
$L\cup M$	0.275
R	0.49
$L\cup M\cup R$	0.01

Table XV
FINAL FUSION RESULT MARGINALIZED ON Θ

The result given in the Table XV is now perfectly consistent with the result obtained directly from the probability calculus shown in (27). Therefore the lower and upper bounds of unknown underlying probability governing the experiment will be

$$\begin{cases} [Bel_{(\mathcal{B}_1\mathcal{B}_2)}(L);Pl_{(\mathcal{B}_1\mathcal{B}_2)}(L)]=[0.225;0.51] \\ [Bel_{(\mathcal{B}_1\mathcal{B}_2)}(M);Pl_{(\mathcal{B}_1\mathcal{B}_2)}(M)]=[0;0.276] \\ [Bel_{(\mathcal{B}_1\mathcal{B}_2)}(R);Pl_{(\mathcal{B}_1\mathcal{B}_2)}(R)]=[0.49;0.50] \end{cases} \quad (50)$$

These bounds coincide with the bounds computed in (28) and obtained directly from the simple averaging of bba's.

One sees that the results obtained from the probability calculus can be entirely recovered in using the simple conjunctive rule of combination applied directly on the ballooned extended bba's on the joint fusion space. There is no problem at all to manipulate (combine) sources of evidences following such very general principle of reconcilement and the notion

of conflict just vanishes on such case. Of course, a similar analysis and conclusions can be done and drawn for Zadeh's example and this is left to readers verification.

IV. Conclusions

In this paper, we have shown through a very simple examples and through an experimental protocol of verification, the inconsistency of the Mathematical Theory of Evidence with the probability calculus as soon as we apply directly Dempster's rule of combination to the original bba's provided by the distinct and independent sources of evidences, and even if they are not conflicting. Our conclusions corroborate the conclusions drawn by some researchers. We have also shown that the doubts on the validity of Dempster's rule of combination pointed out by Zadeh were well justified. Of course, this flaw and unsatisfactory behavior of a direct use of Dempster's rule of combination doesn't mean that belief functions and combinations rules are useless in practice for modeling uncertainty and fuse them. We just have to accept the fact that we need to manage them differently (for combination or conditioning) and in a better consistent way; otherwise we must be ready to accept inconsistent results and their consequences. In the last part of the paper, we have shown how it is possible to reconcile the conjunctive rule of combination with the probability calculus but the mechanism for dealing with sources of evidences differs obviously from the original approach of the Mathematical Theory of Evidence. In our examples, one needed first to extend the bba's into the joint fusion space before combining them with the simple conjunctive rule of combination; this was equivalent to combining conditionnal bba's in a way similar to the Total Probability approach, which was naturally leaded by the nature of the problems. In the VIVE paradigm, the notion of conflict between sources becomes just useless and meaningless. Using our new conjunctive-based approach, the interest in Dempster's rule appears very limited for such problems because it can be replaced in fact by the simple average of the bba's to combine which makes derivations simpler. Whatever, this approach combining a conditional formalization of the sources of information with the combination rules appears as a way for constructing a coherent framework of implementation of the combination rules. On this point, we are going closer to the Bayesian reasoning. A generalization of the approach, involving inversion similar to the Bayesian inversion, should provide a complete solution for a coherent implementation of the combination rules. At last, we don't claim that the Mathematical Theory of Evidence as originally developed, and its extensions (based on other rules of combinations) are definitely useless in practice, but it seems that the original Dempster-Shafer theory of belief functions belongs to a family of Approximate Reasoning Theories that appear perfectible. If they aren't, they need to be proved consistent in some well defined domains of applicability. The imperfections are due to the inconsistency of the results drawn from the reasoning with respect to the results obtained with the probability calculus validated by our experimental protocol. Since we have no proof that human reasoning is fully consistent with probability calculus, we do easily admit that Approximate Reasoning Theories present some interests of course, but the readers and users must be aware of this inconsistency and should use these theories always with extreme caution in applications.

References

[1] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. Math. Statist. 38, pp. 325–339, 1967.

[2] A.P. Dempster, *A generalization of Bayesian inference*, J. R. Statist. Soc. B 30, pp. 205–247, 1968.

[3] A.P. Dempster, *The Dempster-Shafer calculus for statisticians*, International Journal of Approximate Reasoning, 2008.

[4] J. Dezert, A. Tchamova, *On the behavior of Dempster's rule of combination*, presented during poster session at BFTA 2011, Autrans, France, April 4–8, 2011, and available on http://hal.archives-ouvertes.fr/hal-00577983/

[5] A. Gelman, *The boxer, the wrestler, and the coin flip: a paradox of robust Bayesian inference and belief functions*, American Statistician, Vol. 60, No. 2, pp. 146–150, 2006.

[6] K. Murphy, *Combining belief functions when evidence conflicts*, Decision Support Systems, Vol. 29, pp. 1–9, 2000.

[7] J. Pearl, *Reasoning with belief functions: An analysis of compatibility*, International Journal of Approximate Reasoning, Vol. 4, pp. 363–389, 1990.

[8] J. Pearl, *Rejoinder of comments on "Reasoning with belief functions: An analysis of compatibility"*, International Journal of Approximate Reasoning, Vol. 6, pp. 425–443, 1992.

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

[10] P. Smets, *The transferable belief model for quantified belief representation*, In D.M. Gabbay and P. Smets (Eds.) Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol.1, Doortrecht, The Netherlands, Kluwer, pp. 267–301, 1998.

[11] L.A. Zadeh, *On the validity of Dempster's rule of combination*, Memo M79/24, Univ. of California, Berkeley, USA, 1979.

[12] L.A. Zadeh, *Book review: A mathematical theory of evidence*, The Al Magazine, Vol. 5, No. 3, pp. 81-83, 1984.

[13] L.A. Zadeh, *A simple view of the Dempster-Shafer Theory of Evidence and its implication for the rule of combination*, The Al Magazine, Vol. 7, No. 2, pp. 85-90, Summer 1986.