Why Dempster's fusion rule is not a generalization of Bayes fusion rule

Jean Dezert
The French Aerospace Lab
Chemin de la Hunière
F-91761 Palaiseau, France.
jean.dezert@onera.fr

Albena Tchamova
Institute of I&C Tech.
Bulgarian Acad. of Sci.
Sofia, Bulgaria.
tchamova@bas.bg

Deqiang Han
Inst. of Integrated Automation
Xi'an Jiaotong University
Xi'an, China 710049.
deqhan@gmail.com

Jean-Marc Tacnet
Irstea, UR ETGR
2 Rue de la Papéterie - BP 76
F-38402 St-Martin-d'Hères, France.
jean-marc.tacnet@irstea.fr

***Abstract*—In this paper, we analyze Bayes fusion rule in details from a fusion standpoint, as well as the emblematic Dempster's rule of combination introduced by Shafer in his Mathematical Theory of evidence based on belief functions. We propose a new interesting formulation of Bayes rule and point out some of its properties. A deep analysis of the compatibility of Dempster's fusion rule with Bayes fusion rule is done. We show that Dempster's rule is compatible with Bayes fusion rule only in the very particular case where the basic belief assignments (bba's) to combine are Bayesian, and when the prior information is modeled either by a uniform probability measure, or by a vacuous bba. We show clearly that Dempster's rule becomes incompatible with Bayes rule in the more general case where the prior is truly informative (not uniform, nor vacuous). Consequently, this paper proves that Dempster's rule is not a generalization of Bayes fusion rule.**

***Keywords*—Information fusion, Probability theory, Bayes fusion rule, Dempster's fusion rule.**

I. Introduction

In 1979, Lotfi Zadeh questioned in [1] the validity of the Dempster's rule of combination [2], [3] proposed by Shafer in Dempster-Shafer Theory (DST) of evidence [4]. Since more than 30 years many strong debates [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15] on the validity of foundations of DST and Dempster's rule have bloomed. The purpose of this paper is not to discuss the validity of Dempster's rule, nor the foundations of DST which have been already addressed in previous papers [16], [17], [18]. In this paper, we just focus on the deep analysis of the real incompatibility of Dempster's rule with Bayes fusion rule. Our analysis supports Mahler's one briefly presented in [19].

This paper is organized as follows. In section II, we recall basics of conditional probabilities and Bayes fusion rule with its main properties. In section III, we recall the basics of belief functions and Dempster's rule. In section IV, we analyze in details the incompatibility of Dempster's rule with Bayes rule in general and its partial compatibility for the very particular case when prior information is modeled by a Bayesian uniform basic belief assignment (bba). Section V concludes this paper.

II. Conditional probabilities and Bayes fusion

In this section, we recall the definition of conditional probability [20], [21] and present the principle and the properties of Bayes fusion rule. We present the structure of this rule derived from the classical definition of the conditional probability in a new uncommon interesting form that will help us to analyze its partial similarity with Dempster's rule proposed by Shafer in his mathematical theory of evidence [4]. We will show clearly why Dempster's rule fails to be compatible with Bayes rule in general.

A. Conditional probabilities

Let us consider two random events X and Z. The conditional probability mass functions (pmfs) $P(X|Z)$ and $P(Z|X)$ are defined[1] (assuming $P(X)>0$ and $P(Z)>0$) by [20]:

$$P(X|Z) \triangleq \frac{P(X\cap Z)}{P(Z)} \quad \text{and} \quad P(Z|X) \triangleq \frac{P(X\cap Z)}{P(X)} \tag{1}$$

From Eq. (1), one gets $P(X\cap Z) = P(X|Z)P(Z) = P(Z|X)P(X)$, which yields to Bayes Theorem:

$$P(X|Z) = \frac{P(Z|X)P(X)}{P(Z)} \text{ and } P(Z|X) = \frac{P(X|Z)P(Z)}{P(X)} \tag{2}$$

where $P(X)$ is called the *a priori probability* of X, and $P(Z|X)$ is called the *likelihood of* X. The denominator $P(Z)$ plays the role of a normalization constant warranting that $\sum_{i=1}^{N} P(X=x_i|Z)=1$. In fact $P(Z)$ can be rewritten as

$$P(Z) = \sum_{i=1}^{N} P(Z|X=x_i)P(X=x_i) \tag{3}$$

The set of the N possible exclusive and exhaustive outcomes of X is denoted $\Theta(X) \triangleq \{x_i, i=1,2,\ldots,N\}$.

B. Bayes parallel fusion rule

In fusion applications, we are often interested in computing the probability of an event X given two events Z_1 and Z_2 that have occurred. More precisely, one wants to compute $P(X|Z_1\cap Z_2)$ knowing $P(X|Z_1)$ and $P(X|Z_2)$, where X can take N distinct exhaustive and exclusive states x_i, $i=1,2,\ldots,N$. Such type of problem is traditionally called a *fusion problem*. The computation of $P(X|Z_1\cap Z_2)$ from

[1] For convenience and simplicity, we use the notation $P(X|Z)$ instead of $P(X=x|Z=z)$, and $P(Z|X)$ instead of $P(Z=z|X=x)$ where x and z would represent precisely particular outcomes of the random variables X and Z.

$P(X|Z_1)$ and $P(X|Z_2)$ cannot be done in general without the knowledge of the probabilities $P(X)$ and $P(X|Z_1 \cup Z_2)$ which are rarely given. However, $P(X|Z_1 \cap Z_2)$ becomes easily computable by assuming the following conditional statistical independence condition expressed mathematically by:

$$(A1): \qquad P(Z_1 \cap Z_2|X) = P(Z_1|X)P(Z_2|X) \tag{4}$$

With such conditional independence condition (A1), then from Eq. (1) and Bayes Theorem one gets:

$$\begin{aligned} P(X|Z_1 \cap Z_2) &= \frac{P(Z_1 \cap Z_2 \cap X)}{P(Z_1 \cap Z_2)} = \frac{P(Z_1 \cap Z_2|X)P(X)}{P(Z_1 \cap Z_2)} \\ &= \frac{P(Z_1|X)P(Z_2|X)P(X)}{\sum_{i=1}^{N} P(Z_1|X=x_i)P(Z_2|X=x_i)P(X=x_i)} \end{aligned}$$

Using again Eq. (2), we have:

$$P(Z_1|X) = \frac{P(X|Z_1)P(Z_1)}{P(X)} \text{ and } P(Z_2|X) = \frac{P(X|Z_2)P(Z_2)}{P(X)}$$

and the previous formula of conditional probability $P(X|Z_1 \cap Z_2)$ can be rewritten as:

$$P(X|Z_1 \cap Z_2) = \frac{\frac{P(X|Z_1)P(X|Z_2)}{P(X)}}{\sum_{i=1}^{N} \frac{P(X=x_i|Z_1)P(X=x_i|Z_2)}{P(X=x_i)}} \tag{5}$$

The rule of combination given by Eq. (5) is known as *Bayes parallel (or product) rule* and dates back to Bernoulli [22]. In the classification framework, this formula is also called the *Naive Bayesian Classifier* because it uses the assumption (A1) which is often considered as very unrealistic and too simplistic, and that is why it is called a *naive* assumption. The Eq. (5) can be rewritten as:

$$P(X|Z_1 \cap Z_2) = \frac{1}{K(X, Z_1, Z_2)} \cdot P(X|Z_1) \cdot P(X|Z_2) \tag{6}$$

where the coefficient $K(X, Z_1, Z_2)$ is defined by:

$$K(X, Z_1, Z_2) \triangleq P(X) \cdot \sum_{i=1}^{N} \frac{P(X=x_i|Z_1)P(X=x_i|Z_2)}{P(X=x_i)} \tag{7}$$

C. Symmetrization of Bayes fusion rule

The expression of Bayes fusion rule given by Eq. (5) can also be symmetrized in the following form that, quite surprisingly, rarely appears in the literature:

$$P(X|Z_1 \cap Z_2) = \frac{\frac{P(X|Z_1)}{\sqrt{P(X)}} \cdot \frac{P(X|Z_2)}{\sqrt{P(X)}}}{\sum_{i=1}^{N} \frac{P(X=x_i|Z_1)}{\sqrt{P(X=x_i)}} \cdot \frac{P(X=x_i|Z_2)}{\sqrt{P(X=x_i)}}} \tag{8}$$

or in an equivalent manner:

$$P(X|Z_1 \cap Z_2) = \frac{1}{K'(Z_1, Z_2)} \cdot \frac{P(X|Z_1)}{\sqrt{P(X)}} \cdot \frac{P(X|Z_2)}{\sqrt{P(X)}} \tag{9}$$

where the normalization constant $K'(Z_1, Z_2)$ is given by:

$$K'(Z_1, Z_2) \triangleq \sum_{i=1}^{N} \frac{P(X=x_i|Z_1)}{\sqrt{P(X=x_i)}} \cdot \frac{P(X=x_i|Z_2)}{\sqrt{P(X=x_i)}} \tag{10}$$

We call the quantity $A_2(X = x_i) \triangleq \frac{P(X=x_i|Z_1)}{\sqrt{P(X=x_i)}} \cdot \frac{P(X=x_i|Z_2)}{\sqrt{P(X=x_i)}}$ entering in Eq. (10) the *Agreement Factor* on $X = x_i$ of order 2, because only two posterior pmfs are used in the derivation. $A_2(X = x_i)$ corresponds to the posterior conjunctive consensus on the event $X = x_i$ taking into account the prior pmf of X. The denominator of Eq. (8) measures the level of the Global Agreement (GA) of the conjunctive consensus taking into account the prior pmf of X. It is denoted[2] GA_2.

$$\begin{aligned} GA_2 &\triangleq \sum_{i_1, i_2=1|i_1=i_2}^{N} \frac{P(X=x_{i_1}|Z_1)}{\sqrt{P(X=x_{i_1})}} \cdot \frac{P(X=x_{i_2}|Z_2)}{\sqrt{P(X=x_{i_2})}} \\ &= \sum_{i=1}^{N} \frac{P(X=x_i|Z_1)}{\sqrt{P(X=x_i)}} \cdot \frac{P(X=x_i|Z_2)}{\sqrt{P(X=x_i)}} = K'(Z_1, Z_2) \end{aligned} \tag{11}$$

In fact, with assumption (A1), the probability $P(X|Z_1 \cap Z_2)$ given in Eq. (9) is nothing but the simple ratio of the agreement factor $A_2(X)$ (conjunctive consensus) on X over the global agreement $GA_2 = \sum_{i=1}^{N} A_2(X = x_i)$, that is:

$$P(X|Z_1 \cap Z_2) = \frac{A_2(X)}{GA_2} \tag{12}$$

The quantity GC_2 given in Eq. (13) measures the global conflict (i.e. the total conjunctive disagreement) taking into account the prior pmf of X.

$$GC_2 \triangleq \sum_{i_1, i_2=1|i_1 \neq i_2}^{N} \frac{P(X=x_{i_1}|Z_1)}{\sqrt{P(X=x_{i_1})}} \cdot \frac{P(X=x_{i_2}|Z_2)}{\sqrt{P(X=x_{i_2})}} \tag{13}$$

- **Generalization to** $P(X|Z_1 \cap Z_2 \cap \ldots \cap Z_s)$

It can be proved that, when assuming conditional independence conditions, Bayes parallel combination rule can be generalized for combining $s > 2$ posterior pmfs as:

$$P(X|Z_1 \cap \ldots \cap Z_s) = \frac{1}{K(X, Z_1, \ldots, Z_s)} \cdot \prod_{k=1}^{s} P(X|Z_k) \tag{14}$$

where the coefficient $K(X, Z_1, \ldots, Z_s)$ is defined by:

$$K(X, Z_1, \ldots, Z_s) \triangleq P(X) \sum_{i=1}^{N} \frac{(\prod_{k=1}^{s} P(X=x_i|Z_k))}{P(X=x_i)} \tag{15}$$

The symmetrized form of Eq. (14) is:

$$P(X|Z_1 \cap \ldots \cap Z_s) = \frac{1}{K'(Z_1, \ldots, Z_s)} \cdot \prod_{k=1}^{s} \frac{P(X|Z_k)}{\sqrt[s]{P(X)}} \tag{16}$$

with the normalization constant $K'(Z_1, \ldots, Z_s)$ given by:

$$K'(Z_1, \ldots, Z_s) \triangleq \sum_{i=1}^{N} \prod_{k=1}^{s} \frac{P(X=x_i|Z_k)}{\sqrt[s]{P(X=x_i)}} \tag{17}$$

[2]The index 2 is introduced explicitly in the notations because we consider only the fusion of two posterior pmfs.

The generalization of $A_2(X)$, GA_2, and GC_2 provides the agreement $A_s(X)$ of order s, the global agreement GA_s and the global conflict GC_s for s sources as follows:

$$A_s(X = x_i) \triangleq \prod_{k=1}^{s} \frac{P(X = x_i|Z_k)}{\sqrt[s]{P(X = x_i)}}$$

$$GA_s \triangleq \sum_{i_1,\ldots,i_s=1|i_1=\ldots=i_s}^{N} \frac{P(X = x_{i_1}|Z_1)}{\sqrt[s]{P(X = x_{i_1})}} \cdots \frac{P(X = x_{i_s}|Z_s)}{\sqrt[s]{P(X = x_{i_s})}}$$

$$GC_s \triangleq \sum_{i_1,\ldots,i_s=1}^{N} \frac{P(X = x_{i_1}|Z_1)}{\sqrt[s]{P(X = x_{i_1})}} \cdots \frac{P(X = x_{i_s}|Z_s)}{\sqrt[s]{P(X = x_{i_s})}} - GA_s$$

• **Symbolic representation of Bayes fusion rule**

The (symmetrized form of) Bayes fusion rule of two posterior probability measures $P(X|Z_1)$ and $P(X|Z_2)$, given in Eq. (9), requires an extra knowledge of the prior probability of X. For convenience, we denote symbolically this fusion rule as:

$$P(X|Z_1 \cap Z_2) = Bayes(P(X|Z_1), P(X|Z_2); P(X)) \quad (18)$$

Similarly, the (symmetrized) Bayes fusion rule of $s \geq 2$ probability measures $P(X|Z_k)$, $k = 1, 2, \ldots, s$ given by Eq. (16), which requires also the knowledge of $P(X)$, will be denoted as:

$$P(X|Z_1 \cap \ldots \cap Z_s) = Bayes(P(X|Z_1), \ldots, P(X|Z_s); P(X))$$

• **Particular case: Uniform a priori pmf**

If the random variable X is assumed as a priori uniformly distributed over the space of its N possible outcomes, then the probability of X is equal to $P(X = x_i) = 1/N$ for $i = 1, 2, \ldots, N$. In such particular case, all the prior probabilities values $\sqrt{P(X = x_i)} = \sqrt{1/N}$ and $\sqrt[s]{P(X = x_i)} = \sqrt[s]{1/N}$ can be simplified in Bayes fusion formulas Eq. (9) and Eq. (10). Therefore, Bayes fusion formula (9) reduces to:

$$P(X|Z_1 \cap Z_2) = \frac{P(X|Z_1)P(X|Z_2)}{\sum_{i=1}^{N} P(X = x_i|Z_1)P(X = x_i|Z_2)} \quad (19)$$

By convention, Eq. (19) is denoted symbolically as:

$$P(X|Z_1 \cap Z_2) = Bayes(P(X|Z_1), P(X|Z_2)) \quad (20)$$

Similarly, $Bayes(P(X|Z_1), \ldots, P(X|Z_s))$ rule defined with an uniform a priori pmf of X will be given by:

$$P(X|Z_1 \cap \ldots \cap Z_s) = \frac{\prod_{k=1}^{s} P(X|Z_k)}{\sum_{i=1}^{N} \prod_{k=1}^{s} P(X = x_i|Z_k)} \quad (21)$$

When $P(X)$ is uniform and from Eq. (19), one can redefine the global agreement and the global conflict as:

$$GA_2^{unif} \triangleq \sum_{i,j=1|i=j}^{N} P(X = x_i|Z_1)P(X = x_j|Z_2) \quad (22)$$

$$GC_2^{unif} \triangleq \sum_{i,j=1|i\neq j}^{N} P(X = x_i|Z_1)P(X = x_j|Z_2) \quad (23)$$

Because $\sum_{i=1}^{N} P(X = x_i|Z_1) = 1$ and $\sum_{j=1}^{N} P(X = x_j|Z_2) = 1$, then

$$\begin{aligned}
1 &= (\sum_{i=1}^{N} P(X = x_i|Z_1))(\sum_{j=1}^{N} P(X = x_j|Z_2)) \\
&= \sum_{i,j=1}^{N} P(X = x_i|Z_1)P(X = x_j|Z_2) \\
&= \sum_{i,j=1|i=j}^{N} P(X = x_i|Z_1)P(X = x_j|Z_2) \\
&\quad + \sum_{i,j=1|i\neq j}^{N} P(X = x_i|Z_1)P(X = x_j|Z_2)
\end{aligned}$$

Therefore, one has always $GA_2^{unif} + GC_2^{unif} = 1$ when $P(X)$ is uniform, and Eq. (19) can be expressed as:

$$P(X|Z_1 \cap Z_2) = \frac{P(X|Z_1)P(X|Z_2)}{GA_2^{unif}} = \frac{P(X|Z_1)P(X|Z_2)}{1 - GC_2^{unif}} \quad (24)$$

By a direct extension, one will have:

$$P(X|Z_1 \cap \ldots \cap Z_s) = \frac{\prod_{k=1}^{s} P(X|Z_k)}{GA_s^{unif}} = \frac{\prod_{k=1}^{s} P(X|Z_k)}{1 - GC_s^{unif}} \quad (25)$$

$$GA_s^{unif} = \sum_{i_1,\ldots,i_s=1|i_1=\ldots=i_s}^{N} P(X = x_{i_1}|Z_1) \ldots P(X = x_{i_s}|Z_s)$$

$$GC_s^{unif} = 1 - GA_s^{unif}$$

Remark 1: The normalization coefficient corresponding to the global conjunctive agreement GA_s^{unif} can also be expressed using belief function notations [4] as:

$$GA_s^{unif} = \sum_{\substack{x_{i_1},\ldots,x_{i_s} \in \Theta(X) \\ x_{i_1} \cap \ldots \cap x_{i_s} \neq \emptyset}} P(X = x_{i_1}|Z_1) \ldots P(X = x_{i_s}|Z_s)$$

and the global disagreement, or total conflict level, is given by:

$$GC_s^{unif} = \sum_{\substack{x_{i_1},\ldots,x_{i_s} \in \Theta(X) \\ x_{i_1} \cap \ldots \cap x_{i_s} = \emptyset}} P(X = x_{i_1}|Z_1) \ldots P(X = x_{i_s}|Z_s)$$

D. *Properties of Bayes fusion rule*

In this subsection, we analyze Bayes fusion rule (assuming condition (A1) holds) from a pure algebraic standpoint. In fusion jargon, the quantities to combine come from sources of information which provide inputs that feed the fusion rule. In the probabilistic framework, a source s to combine corresponds to the posterior pmf $P(X|Z_s)$. In this subsection, we establish five interesting properties of Bayes rule. Contrary to Dempster's rule, we prove that Bayes rule is not associative in general.

• **(P1) : The pmf $P(X)$ is a neutral element of Bayes fusion rule when combining only two sources.**

Proof: A source is called a neutral element of a fusion rule if and only if it has no influence on the fusion result. $P(X)$ is a neutral element of Bayes rule if and only if

$Bayes(P(X|Z_1), P(X); P(X)) = P(X|Z_1)$. It can be easily verified that this equality holds by replacing $P(X|Z_2)$ by $P(X)$ and $P(X = x_i|Z_2)$ by $P(X = x_i)$ (as if the conditioning term Z_2 vanishes) in Eq. (5). One can also verify that $Bayes(P(X), P(X|Z_2); P(X)) = P(X|Z_2)$, which completes the proof.

Remark 2: When considering Bayes fusion of more than two sources, $P(X)$ doesn't play the role of a neutral element in general, except if $P(X)$ is uniform. For example, let us consider 3 pmfs $P(X|Z_1)$, $P(X|Z_2)$ and $P(X|Z_3)$ to combine with formula (14) with $P(X)$ not uniform. When Z_3 vanishes so that $P(X|Z_3) = P(X)$, we can easily check that:

$$\begin{aligned} &Bayes(P(X|Z_1), P(X|Z_2), P(X); P(X)) \\ &\qquad \neq Bayes(P(X|Z_1), P(X|Z_2); P(X)) \end{aligned} \tag{26}$$

• **(P2) : Bayes fusion rule is in general not idempotent.**

Proof: A fusion rule is idempotent if the combination of all same inputs is equal to the inputs. To prove that Bayes rule is not idempotent it suffices to prove that in general:

$$Bayes(P(X|Z_1), P(X|Z_1); P(X)) \neq P(X|Z_1)$$

From Bayes rule (5), when $P(X|Z_2) = P(X|Z_1)$ we clearly get in general

$$\frac{1}{P(X)} \frac{P(X|Z_1)P(X|Z_1)}{\sum_{i=1}^{N} \frac{P(X=x_i|Z_1)P(X=x_i|Z_1)}{P(X=x_i)}} \neq P(X|Z_1) \tag{27}$$

but when Z_1 and Z_2 vanish, because in such case Eq. (27) reduces to $P(X)$ on its left and right sides.

Remark 3: In the particular (two sources) degenerate case where Z_1 and Z_2 vanish, one has always: $Bayes(P(X), P(X); P(X)) = P(X)$. However, in the more general degenerate case (when considering more than 2 sources), one will have in general: $Bayes(P(X), P(X), \ldots, P(X); P(X)) \neq P(X)$, but when $P(X)$ is uniform, or when $P(X)$ is a "deterministic" probability measure such that $P(X = x_i) = 1$ for a given $x_i \in \Theta(X)$ and $P(X = x_j) = 0$ for all $x_j \neq x_i$.

• **(P3) : Bayes fusion rule is in general not associative.**

Proof: A fusion rule f is called associative if and only if it satisfies the associative law: $f(f(x, y), z) = f(x, f(y, z)) = f(y, f(x, z)) = f(x, y, z)$ for all possible inputs x, y and z. Let us prove that Bayes rule is not associative from a very simple example.

Example 1: Let us consider the simplest set of outcomes $\{x_1, x_2\}$ for X, with prior pmf:

$$P(X = x_1) = 0.2 \text{ and } P(X = x_2) = 0.8$$

and let us consider the three given sets of posterior pmfs:

$$\begin{cases} P(X = x_1|Z_1) = 0.1 \text{ and } P(X = x_2|Z_1) = 0.9 \\ P(X = x_1|Z_2) = 0.5 \text{ and } P(X = x_2|Z_2) = 0.5 \\ P(X = x_1|Z_3) = 0.6 \text{ and } P(X = x_2|Z_3) = 0.4 \end{cases}$$

Bayes fusion $Bayes(P(X|Z_1),)P(X|Z_2), P(X|Z_3); P(X))$ of the three sources altogether according to Eq. (16) provides:

$$\begin{cases} P(X = x_1|Z_1 \cap Z_2 \cap Z_3) = \frac{1}{K_{123}} \frac{0.1}{\sqrt[3]{0.2}} \frac{0.5}{\sqrt[3]{0.2}} \frac{0.6}{\sqrt[3]{0.2}} = 0.40 \\ P(X = x_2|Z_1 \cap Z_2 \cap Z_3) = \frac{1}{K_{123}} \frac{0.9}{\sqrt[3]{0.8}} \frac{0.5}{\sqrt[3]{0.8}} \frac{0.4}{\sqrt[3]{0.8}} = 0.60 \end{cases}$$

where the normalization constant K_{123} is given by:

$$K_{123} = \frac{0.1}{\sqrt[3]{0.2}} \frac{0.5}{\sqrt[3]{0.2}} \frac{0.6}{\sqrt[3]{0.2}} + \frac{0.9}{\sqrt[3]{0.8}} \frac{0.5}{\sqrt[3]{0.8}} \frac{0.4}{\sqrt[3]{0.8}} = 0.3750$$

Let us compute the fusion of $P(X|Z_1)$ with $P(X|Z_2)$ using $Bayes(P(X|Z_1), P(X|Z_2); P(X))$. One has:

$$\begin{cases} P(X = x_1|Z_1 \cap Z_2) = \frac{1}{K_{12}} \frac{0.1}{\sqrt{0.2}} \frac{0.5}{\sqrt{0.2}} \approx 0.3077 \\ P(X = x_2|Z_1 \cap Z_2) = \frac{1}{K_{12}} \frac{0.9}{\sqrt{0.8}} \frac{0.5}{\sqrt{0.8}} \approx 0.6923 \end{cases}$$

where the normalization constant K_{12} is given by:

$$K_{12} = \frac{0.1}{\sqrt{0.2}} \frac{0.5}{\sqrt{0.2}} + \frac{0.9}{\sqrt{0.8}} \frac{0.5}{\sqrt{0.8}} = 0.8125$$

Let us compute the fusion of $P(X|Z_2)$ with $P(X|Z_3)$ using $Bayes(P(X|Z_2), P(X|Z_3); P(X))$. One has

$$\begin{cases} P(X = x_1|Z_2 \cap Z_3) = \frac{1}{K_{23}} \frac{0.5}{\sqrt{0.2}} \frac{0.6}{\sqrt{0.2}} \approx 0.8571 \\ P(X = x_2|Z_2 \cap Z_3) = \frac{1}{K_{23}} \frac{0.5}{\sqrt{0.8}} \frac{0.4}{\sqrt{0.8}} \approx 0.1429 \end{cases}$$

where the normalization constant K_{23} is given by:

$$K_{23} = \frac{0.5}{\sqrt{0.2}} \frac{0.6}{\sqrt{0.2}} + \frac{0.5}{\sqrt{0.8}} \frac{0.4}{\sqrt{0.8}} = 1.75$$

Let us compute the fusion of $P(X|Z_1)$ with $P(X|Z_3)$ using $Bayes(P(X|Z_1), P(X|Z_3); P(X))$. One has:

$$\begin{cases} P(X = x_1|Z_1 \cap Z_3) = \frac{1}{K_{13}} \frac{0.1}{\sqrt{0.2}} \frac{0.6}{\sqrt{0.2}} = 0.4 \\ P(X = x_2|Z_1 \cap Z_3) = \frac{1}{K_{13}} \frac{0.9}{\sqrt{0.8}} \frac{0.4}{\sqrt{0.8}} = 0.6 \end{cases}$$

where the normalization constant K_{13} is given by:

$$K_{13} = \frac{0.1}{\sqrt{0.2}} \frac{0.6}{\sqrt{0.2}} + \frac{0.9}{\sqrt{0.8}} \frac{0.4}{\sqrt{0.8}} = 0.75$$

Let us compute the fusion of $P(X|Z_1 \cap Z_2)$ with $P(X|Z_3)$ using $Bayes(P(X|Z_1 \cap Z_2), P(X|Z_3); P(X))$. One has

$$\begin{cases} P(X = x_1|(Z_1 \cap Z_2) \cap Z_3) = \frac{1}{K_{(12)3}} \frac{0.3077}{\sqrt{0.2}} \frac{0.6}{\sqrt{0.2}} \approx 0.7273 \\ P(X = x_2|(Z_1 \cap Z_2) \cap Z_3) = \frac{1}{K_{(12)3}} \frac{0.6923}{\sqrt{0.8}} \frac{0.4}{\sqrt{0.8}} \approx 0.2727 \end{cases}$$

where the normalization constant $K_{(12)3}$ is given by

$$K_{(12)3} = \frac{0.3077}{\sqrt{0.2}} \frac{0.6}{\sqrt{0.2}} + \frac{0.6923}{\sqrt{0.8}} \frac{0.4}{\sqrt{0.8}} \approx 1.26925$$

Let us compute the fusion of $P(X|Z_1)$ with $P(X|Z_2 \cap Z_3)$ using $Bayes(P(X|Z_1), P(X|Z_2 \cap Z_3); P(X))$. One has

$$\begin{cases} P(X = x_1|Z_1 \cap (Z_2 \cap Z_3)) = \frac{1}{K_{1(23)}} \frac{0.1}{\sqrt{0.2}} \frac{0.8571}{\sqrt{0.2}} \approx 0.7273 \\ P(X = x_2|Z_1 \cap (Z_2 \cap Z_3)) = \frac{1}{K_{1(23)}} \frac{0.9}{\sqrt{0.8}} \frac{0.1429}{\sqrt{0.8}} \approx 0.2727 \end{cases}$$

where the normalization constant $K_{1(23)}$ is given by

$$K_{1(23)} = \frac{0.1}{\sqrt{0.2}} \frac{0.8571}{\sqrt{0.2}} + \frac{0.9}{\sqrt{0.8}} \frac{0.1429}{\sqrt{0.8}} \approx 0.58931$$

Let us compute the fusion of $P(X|Z_1 \cap Z_3)$ with $P(X|Z_2)$ using $Bayes(P(X|Z_1 \cap Z_3), P(X|Z_2); P(X))$. One has

$$\begin{cases} P(X = x_1|(Z_1 \cap Z_3) \cap Z_2) = \frac{1}{K_{(13)2}} \frac{0.4}{\sqrt{0.2}} \frac{0.5}{\sqrt{0.2}} \approx 0.7273 \\ P(X = x_2|(Z_1 \cap Z_3) \cap Z_2) = \frac{1}{K_{(13)2}} \frac{0.6}{\sqrt{0.8}} \frac{0.5}{\sqrt{0.8}} \approx 0.2727 \end{cases}$$

where the normalization constant $K_{(13)2}$ is given by

$$K_{(13)2} = \frac{0.4}{\sqrt{0.2}}\frac{0.5}{\sqrt{0.2}} + \frac{0.6}{\sqrt{0.8}}\frac{0.5}{\sqrt{0.8}} = 1.375$$

Therefore, one sees that even if in our example one has $f(x, f(y, z)) = f(f(x, y), z) = f(y, f(x, z))$ because $P(X|(Z_1 \cap Z_2) \cap Z_3) = P(X|Z_1 \cap (Z_2 \cap Z_3)) = P(X|Z_2 \cap (Z_1 \cap Z_3))$, Bayes fusion rule is not associative since:

$$\begin{cases} P(X|(Z_1 \cap Z_2) \cap Z_3) \neq P(X|Z_1 \cap Z_2 \cap Z_3) \\ P(X|Z_1 \cap (Z_2 \cap Z_3)) \neq P(X|Z_1 \cap Z_2 \cap Z_3) \\ P(X|Z_2 \cap (Z_1 \cap Z_3)) \neq P(X|Z_1 \cap Z_2 \cap Z_3) \end{cases}$$

• **(P4) : Bayes fusion rule is associative if and only if $P(X)$ is uniform.**

Proof: If $P(X)$ is uniform, Bayes fusion rule is given by Eq. (21) which can be rewritten as:

$$P(X|Z_1 \cap \ldots \cap Z_s) = \frac{P(X|Z_s)\prod_{k=1}^{s-1} P(X|Z_k)}{\sum_{i=1}^{N} P(X = x_i|Z_s)\prod_{k=1}^{s-1} P(X = x_i|Z_k)}$$

By introducing the term $1/\sum_{i=1}^{N}\prod_{k=1}^{s-1} P(X = x_i|Z_k)$ in numerator and denominator of the previous formula, it comes:

$$P(X|Z_1 \cap \ldots \cap Z_s) = \frac{\frac{\prod_{k=1}^{s-1} P(X|Z_k)}{\sum_{i=1}^{N}\prod_{k=1}^{s-1} P(X=x_i|Z_k)} P(X|Z_s)}{\sum_{i=1}^{N} \frac{\prod_{k=1}^{s-1} P(X=x_i|Z_k)}{\sum_{i=1}^{N}\prod_{k=1}^{s-1} P(X=x_i|Z_k)} P(X = x_i|Z_s)}$$

which can be simply rewritten as:

$$P(X|Z_1 \cap \ldots \cap Z_s) = \frac{P(X|Z_1 \cap \ldots \cap Z_{s-1})P(X|Z_s)}{\sum_{i=1}^{N} P(X = x_i|Z_1 \cap \ldots \cap Z_{s-1})P(X = x_i|Z_s)}$$

Therefore when $P(X)$ is uniform, one has:

$$\begin{aligned} &Bayes(P(X|Z_1), \ldots, P(X|Z_s)) = \\ &\quad Bayes(Bayes(P(X|Z_1), \ldots, P(X|Z_{s-1})), P(X|Z_s)) \end{aligned}$$

The previous relation was based on the decomposition of $\prod_{k=1}^{s} P(X|Z_k)$ as $P(X|Z_s)\prod_{k=1}^{s-1} P(X|Z_k)$. This choice of decomposition was arbitrary and chosen only for convenience. In fact $\prod_{k=1}^{s} P(X|Z_k)$ can be decomposed in s different manners, as $P(X|Z_j)\prod_{k=1|k\neq j}^{s} P(X|Z_k)$, $j = 1, 2, \ldots s$ and the similar analysis can be done. In particular, when $s = 3$, we will have:

$$\begin{aligned} &Bayes(P(X|Z_1), P(X|Z_2), P(X|Z_3)) = \\ &\quad Bayes(Bayes(P(X|Z_1), P(X|Z_2)), P(X|Z_3)) \\ &\quad = Bayes(P(X|Z_1), Bayes(P(X|Z_2), P(X|Z_3))) \end{aligned}$$

which completes the proof.

• **(P5) : The levels of global agreement and global conflict between the sources do not matter in Bayes fusion rule.**

Proof: This property seems surprising at first glance, but, since the results of Bayes fusion is nothing but the ratio of the agreement on x_i ($i = 1, 2, \ldots, N$) over the global agreement factor, many distinct sources with different global agreements (and thus with different global conflicts) can yield same Bayes fusion result. Indeed, the ratio is kept unchanged when multiplying its numerator and denominator by same non null scalar value. Consequently, the absolute levels of global agreement between the sources (and therefore of global conflict also) do not matter in Bayes fusion result. What really matters is only the proportions of relative agreement factors.

Example 2: To illustrate this property, let us consider Bayes fusion rule applied to two distinct sets[3] of sources represented by $Bayes(P(X|Z_1), P(X|Z_2); P(X))$ and by $Bayes(P'(X|Z_1), P'(X|Z_2); P(X))$ with the following prior and posterior pmfs:

$$P(X = x_1) = 0.2 \text{ and } P(X = x_2) = 0.8$$

$$\begin{cases} P(X = x_1|Z_1) \approx 0.0607 \text{ and } P(X = x_2|Z_1) \approx 0.9393 \\ P(X = x_1|Z_2) \approx 0.6593 \text{ and } P(X = x_2|Z_2) \approx 0.3407 \end{cases}$$

$$\begin{cases} P'(X = x_1|Z_1) \approx 0.8360 \text{ and } P'(X = x_2|Z_1) \approx 0.1640 \\ P'(X = x_1|Z_2) \approx 0.0240 \text{ and } P'(X = x_2|Z_2) \approx 0.9760 \end{cases}$$

Applying Bayes fusion rule given by Eq. (5), one gets for $Bayes(P(X|Z_1), P(X|Z_2); P(X))$:

$$\begin{cases} P(X = x_1|Z_1 \cap Z_2) = \frac{0.2}{0.2+0.4} = 1/3 \\ P(X = x_2|Z_1 \cap Z_2) = \frac{0.4}{0.2+0.4} = 2/3 \end{cases} \tag{28}$$

Similarly, one gets for $Bayes(P'(X|Z_1), P'(X|Z_2); P(X))$

$$\begin{cases} P'(X = x_1|Z_1 \cap Z_2) = \frac{0.1}{0.1+0.2} = 1/3 \\ P'(X = x_2|Z_1 \cap Z_2) = \frac{0.2}{0.1+0.2} = 2/3 \end{cases} \tag{29}$$

Therefore, one sees that $Bayes(P(X|Z_1), P(X|Z_2); P(X)) = Bayes(P'(X|Z_1), P'(X|Z_2); P(X))$ even if the levels of global agreements (and global conflicts) are different. In this particular example, one has:

$$\begin{cases} (GA_2 = 0.60) \neq (GA'_2 = 0.30) \\ (GC_2 = 1.60) \neq (GC'_2 = 2.05) \end{cases} \tag{30}$$

In summary, different sets of sources to combine (with different levels of global agreement and global conflict) can provide exactly the same result once combined with Bayes fusion rule. Hence the different levels of global agreement and global conflict do not really matter in Bayes fusion rule. What really matters in Bayes fusion rule is only the distribution of all the relative agreement factors defined as $A_s(X = x_i)/GA_s$.

III. Belief functions and Dempster's rule

The Belief Functions (BF) have been introduced in 1976 by Glenn Shafer in his mathematical theory of evidence [4], also known as Dempster-Shafer Theory (DST) in order to reason under uncertainty and to model epistemic uncertainties. We will not present in details the foundations of DST, but only the basic mathematical definitions that are necessary for the scope of this paper. The emblematic fusion rule proposed by Shafer to combine sources of evidences characterized by their basic belief assignments (bba) is Dempster's rule that will be analyzed in details in the sequel. In the literature over the years, DST has been widely defended by its proponents in arguing that: 1) Probability measures are particular cases of Belief

[3]The values chosen for $P(X|Z_1)$, $P(X|Z_2)$, $P'(X|Z_1)$, $P'(X|Z_2)$ here have been approximated at the fourth digit. They can be precisely determined such that the expressions for $P(X|Z_1 \cap Z_2)$ and $P'(X|Z_1 \cap Z_2)$ as given in Eqs. (28) and (29) hold. For example, the exact value of $P(x_1|Z_2)$ is obtained by solving a polynomial equation of degree 2 having as a possible solution $P(x_1|Z_2) = \frac{1}{2}(0.72 + \sqrt{0.72^2 - 4 \times 0.04}) = 0.659332590941915 \approx 0.6593$, etc.

functions; and 2) Dempster's fusion rule is a generalization of Bayes fusion rule. Although the statement 1) is correct because Probability measures are indeed particular (additive) Belief functions (called as Bayesian belief functions), we will explain why the second statement about Dempster's rule is incorrect in general.

A. Belief functions

Let Θ be a frame of discernment of a problem under consideration. More precisely, the set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$ consists of a list of N exhaustive and exclusive elements θ_i, $i = 1, 2, \ldots, N$. Each θ_i represents a possible state related to the problem we want to solve. The exhaustivity and exclusivity of elements of Θ is referred as Shafer's model of the frame Θ. A basic belief assignment (bba), also called a belief mass function, $m(.) : 2^\Theta \to [0, 1]$ is a mapping from the power set of Θ (i.e. the set of subsets of Θ), denoted 2^Θ, to $[0, 1]$, that verifies the following conditions [4]:

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{X \in 2^\Theta} m(X) = 1 \tag{31}$$

The quantity $m(X)$ represents the mass of belief exactly committed to X. An element $X \in 2^\Theta$ is called a focal element if and only if $m(X) > 0$. The set $\mathcal{F}(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$ of all focal elements of a bba $m(.)$ is called the core of the bba. A bba $m(.)$ is said Bayesian if its focal elements are singletons of 2^Θ. The vacuous bba characterizing the total ignorance denoted[4] $I_t = \theta_1 \cup \theta_2 \cup \ldots \cup \theta_N$ is defined by $m_v(.) : 2^\Theta \to [0; 1]$ such that $m_v(X) = 0$ if $X \neq \Theta$, and $m_v(I_t) = 1$.

From any bba $m(.)$, the belief function $Bel(.)$ and the plausibility function $Pl(.)$ are defined for $\forall X \in 2^\Theta$ as:

$$\begin{cases} Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y) \\ Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) \end{cases} \tag{32}$$

$Bel(X)$ represents the whole mass of belief that comes from all subsets of Θ included in X. It is interpreted as the lower bound of the probability of X, i.e. $P_{\min}(X)$. $Bel(.)$ is a subadditive measure since $\sum_{\theta_i \in \Theta} Bel(\theta_i) \leq 1$. $Pl(X)$ represents the whole mass of belief that comes from all subsets of Θ compatible with X (i.e., those intersecting X). $Pl(X)$ is interpreted as the upper bound of the probability of X, i.e. $P_{\max}(X)$. $Pl(.)$ is a superadditive measure since $\sum_{\theta_i \in \Theta} Pl(\theta_i) \geq 1$. $Bel(X)$ and $Pl(X)$ are classically seen [4] as lower and upper bounds of an unknown probability $P(.)$, and one has the following inequality satisfied $\forall X \in 2^\Theta$: $Bel(X) \leq P(X) \leq Pl(X)$. The belief function $Bel(.)$ (and the plausibility function $Pl(.)$) built from any Bayesian bba $m(.)$ can be interpreted as a (subjective) conditional probability measure provided by a given source of evidence, because if the bba $m(.)$ is *Bayesian* the following equality always holds [4]: $Bel(X) = Pl(X) = P(X)$.

B. Dempster's rule of combination

Dempster's rule of combination, denoted DS rule[5] is a mathematical operation, represented symbolically by $\oplus$, which corresponds to the normalized conjunctive fusion rule. Based on Shafer's model of Θ, the combination of $s > 1$ independent and distinct sources of evidences characterized by their bba $m_1(.)$, $\ldots$, $m_s(.)$ related to the same frame of discernment Θ is denoted $m_{DS}(.) = [m_1 \oplus \ldots \oplus m_s](.)$. The quantity $m_{DS}(.)$ is defined mathematically as follows: $m_{DS}(\emptyset) \triangleq 0$ and $\forall X \neq \emptyset \in 2^\Theta$

$$m_{DS}(X) \triangleq \frac{m_{12\ldots s}(X)}{1 - K_{12\ldots s}} \tag{33}$$

where the conjunctive agreement on X is given by:

$$m_{12\ldots s}(X) \triangleq \sum_{\substack{X_1, X_2, \ldots, X_s \in 2^\Theta \\ X_1 \cap X_2 \cap \ldots \cap X_s = X}} m_1(X_1) m_2(X_2) \ldots m_s(X_s) \tag{34}$$

and where the global conflict is given by:

$$K_{12\ldots s} \triangleq \sum_{\substack{X_1, X_2, \ldots, X_s \in 2^\Theta \\ X_1 \cap X_2 \cap \ldots \cap X_s = \emptyset}} m_1(X_1) m_2(X_2) \ldots m_s(X_s) \tag{35}$$

When $K_{12\ldots s} = 1$, the s sources are in total conflict and their combination cannot be computed with DS rule because Eq. (33) is mathematically not defined due to 0/0 indeterminacy [4]. DS rule is commutative and associative which makes it very attractive from engineering implementation standpoint.

It has been proved in [4] that the vacuous bba $m_v(.)$ is a neutral element for DS rule because $[m \oplus m_v](.) = [m_v \oplus m](.) = m(.)$ for any bba $m(.)$ defined on 2^Θ. This property looks reasonable since a total ignorant source should not impact the fusion result because it brings no information that can be helpful for the discrimination between the elements of the power set 2^Θ.

IV. Analysis of compatibility of Dempster's rule with Bayes rule

To analyze the compatibility of Dempster's rule with Bayes rule, we need to work in the probabilistic framework because Bayes fusion rule has been developed only in this theoretical framework. So in the sequel, we will manipulate only probability mass functions (pmfs), related with Bayesian bba's in the Belief Function framework. This perfectly justifies the restriction of singleton bba as a prior bba since we want to manipulate prior probabilities to make a fair comparison of results provided by both rules. If Dempster's rule is a true (consistent) generalization of Bayes fusion rule, it must provide same results as Bayes rule when combining Bayesian bba's, otherwise Dempster's rule cannot be fairly claimed to be a generalization of Bayes fusion rule. In this section, we analyze the real (partial or total) compatibility of Dempster's rule with Bayes fusion rule. Two important cases must be analyzed depending on the nature of the prior information $P(X)$ one has in hands for performing the fusion of the sources. These

[4]The set $\{\theta_1, \theta_2, \ldots, \theta_N\}$ and the complete ignorance $\theta_1 \cup \theta_2 \cup \ldots \cup \theta_N$ are both denoted Θ in DST.

[5]We denote it DS rule because it has been proposed historically by Dempster [2], [3], and widely promoted by Shafer in the development of DST [4].

sources to combine will be characterized by the following Bayesian bba's:

$$\begin{cases} m_1(.) \triangleq \{m_1(\theta_i) = P(X = x_i|Z_1), i = 1, 2, \ldots, N\} \\ \vdots \quad \vdots \qquad\qquad\qquad \vdots \\ m_s(.) \triangleq \{m_s(\theta_i) = P(X = x_i|Z_s), i = 1, 2, \ldots, N\} \end{cases} \tag{36}$$

The prior information is characterized by a given bba denoted as $m_0(.)$ that can be defined either on 2^Θ, or only on Θ if we want to deal for the needs of our analysis with a Bayesian prior. In the latter case, if $m_0(.) \triangleq \{m_0(\theta_i) = P(X = x_i), i = 1, 2, \ldots, N\}$ then $m_0(.)$ plays the same role as the prior pmf $P(X)$ in the probabilistic framework.

When considering a non vacuous prior $m_0(.) \neq m_v(.)$, we denote Dempster's combination of s sources symbolically as:

$$m_{DS}(.) = DS(m_1(.), \ldots, m_s(.); m_0(.))$$

When the prior bba is vacuous $m_0(.) = m_v(.)$ then $m_0(.)$ has no impact on Dempster's fusion result, and so we denote symbolically Dempster's rule as:

$$\begin{aligned} m_{DS}(.) &= DS(m_1(.), \ldots, m_s(.); m_v(.)) \\ &= DS(m_1(.), \ldots, m_s(.)) \end{aligned}$$

A. Case 1: Uniform Bayesian prior

It is important to note that Dempster's fusion formula proposed by Shafer in [4] and recalled in Eq. (33) makes no real distinction between the nature of sources to combine (if they are posterior or prior information). In fact, the formula (33) reduces exactly to Bayes rule given in Eq. (25) if the bba's to combine are Bayesian and if the prior information is either uniform or vacuous. Stated otherwise the following functional equality holds

$$\begin{aligned} &DS(m_1(.), \ldots, m_s(.); m_0(.)) \equiv \\ &\qquad Bayes(P(X|Z_1), \ldots, P(X|Z_s); P(X)) \end{aligned} \tag{37}$$

as soon as all bba's $m_i(.)$, $i = 1, 2, \ldots, s$ are Bayesian and coincide with $P(X|Z_i)$, $P(X)$ is uniform, and either the prior bba $m_0(.)$ is vacuous ($m_0(.) = m_v(.)$), or $m_0(.)$ is the uniform Bayesian bba.

Example 3: Let us consider $\Theta(X) = \{x_1, x_2, x_3\}$ with two distinct sources providing the following Bayesian bba's

$$\begin{cases} m_1(x_1) = P(X = x_1|Z_1) = 0.2 \\ m_1(x_2) = P(X = x_2|Z_1) = 0.3 \\ m_1(x_3) = P(X = x_3|Z_1) = 0.5 \end{cases} \text{and} \quad \begin{cases} m_2(x_1) = 0.5 \\ m_2(x_2) = 0.1 \\ m_2(x_3) = 0.4 \end{cases}$$

• If we choose as prior $m_0(.)$ the vacuous bba, that is $m_0(x_1 \cup x_2 \cup x_3) = 1$, then one will get

$$\begin{cases} m_{DS}(x_1) &= \frac{1}{1-K_{12}^{vacuous}} m_1(x_1)m_2(x_1)m_0(x_1 \cup x_2 \cup x_3) \\ &= \frac{1}{1-0.67} 0.2 \cdot 0.5 \cdot 1 = \frac{0.10}{0.33} \approx 0.3030 \\ m_{DS}(x_2) &= \frac{1}{1-K_{12}^{vacuous}} m_1(x_2)m_2(x_2)m_0(x_1 \cup x_2 \cup x_3) \\ &= \frac{1}{1-0.67} 0.3 \cdot 0.1 \cdot 1 = \frac{0.03}{0.33} \approx 0.0909 \\ m_{DS}(x_3) &= \frac{1}{1-K_{12}^{vacuous}} m_1(x_3)m_2(x_3)m_0(x_1 \cup x_2 \cup x_3) \\ &= \frac{1}{1-0.67} 0.5 \cdot 0.4 \cdot 1 = \frac{0.20}{0.33} \approx 0.6061 \end{cases}$$

with

$$\begin{aligned} K_{12}^{vacuous} = 1 &- m_1(x_1)m_2(x_1)m_0(x_1 \cup x_2 \cup x_3) \\ &- m_1(x_2)m_2(x_2)m_0(x_1 \cup x_2 \cup x_3) \\ &- m_1(x_3)m_2(x_3)m_0(x_1 \cup x_2 \cup x_3) = 0.67 \end{aligned}$$

• If we choose as prior $m_0(.)$ the uniform Bayesian bba given by $m_0(x_1) = m_0(x_2) = m_0(x_3) = 1/3$, then we get

$$\begin{cases} m_{DS}(x_1) &= \frac{1}{1-K_{12}^{uniform}} m_1(x_1)m_2(x_1)m_0(x_1) \\ &= \frac{1}{1-0.89} 0.2 \cdot 0.5 \cdot 1/3 = \frac{0.10/3}{0.11} \approx 0.3030 \\ m_{DS}(x_2) &= \frac{1}{1-K_{12}^{uniform}} m_1(x_2)m_2(x_2)m_0(x_2) \\ &= \frac{1}{1-0.89} 0.3 \cdot 0.1 \cdot 1/3 = \frac{0.03/3}{0.11} \approx 0.0909 \\ m_{DS}(x_3) &= \frac{1}{1-K_{12}^{uniform}} m_1(x_3)m_2(x_3)m_0(x_3) \\ &= \frac{1}{1-0.89} 0.5 \cdot 0.4 \cdot 1/3 = \frac{0.20/3}{0.11} \approx 0.6061 \end{cases}$$

where the degree of conflict when $m_0(.)$ is Bayesian and uniform is now given by $K_{12}^{uniform} = 0.89$.

Clearly $K_{12}^{uniform} \neq K_{12}^{vacuous}$, but the fusion results obtained with two distinct priors $m_0(.)$ (vacuous or uniform) are the same because of the algebraic simplification by 1/3 in Dempster's fusion formula when using uniform Bayesian bba. When combining Bayesian bba's $m_1(.)$ and $m_2(.)$, the vacuous prior and uniform prior $m_0(.)$ have therefore no impact on the result. Indeed, they contain no information that may help to prefer one particular state x_i with respect to the other ones, even if the level of conflict is different in both cases. So, the level of conflict doesn't matter at all in such Bayesian case. As already stated, what really matters is only the distribution of relative agreement factors. It can be easily verified that we obtain same results when applying Bayes Eq. (14), or (16).

Only in such very particular cases (i.e. Bayesian bba's, and vacuous or Bayesian uniform priors), Dempster's rule is fully consistent with Bayes fusion rule. So the claim that Dempster's is a generalization of Bayes rule is true in this very particular case only, and that is why such claim has been widely used to defend Dempster's rule and DST thanks to its compatibility with Bayes fusion rule in that very particular case. Unfortunately, such compatibility is only partial and not general because it is not longer valid when considering the more general cases involving non uniform Bayesian prior bba's as shown in the next subsection.

B. Case 2: Non uniform Bayesian prior

Let us consider Dempster's fusion of Bayesian bba's with a Bayesian non uniform prior $m_0(.)$. In such case it is easy to check from the general structures of Bayes fusion rule (16) and Dempster's fusion rule (33) that these two rules are incompatible. Indeed, in Bayes rule one divides each posterior source $m_i(x_j)$ by $\sqrt[s]{m_0(x_j)}$, $i = 1, 2, \ldots s$, whereas the prior source $m_0(.)$ is combined in a pure conjunctive manner by Dempster's rule with the bba's $m_i(.)$, $i = 1, 2, \ldots s$, as if $m_0(.)$ was a simple additional source. This difference of processing prior information between the two approaches explains clearly the incompatibility of Dempster's rule with Bayes rule when Bayesian prior bba is not uniform. This incompatibility is illustrated in the next simple example. Mahler and Fixsen have already proposed in [23], [24], [25] a modification of

Dempster's rule to force it to be compatible with Bayes rule when combining Bayesian bba's. The analysis of such modified Dempster's rule is out of the scope of this paper.

Example 4: Let us consider the same frame $\Theta(X)$, and same bba's $m_1(.)$ and $m_2(.)$ as in the Example 3. Suppose that the prior information is Bayesian and non uniform as follows: $m_0(x_1) = P(X = x_1) = 0.6$, $m_0(x_2) = P(X = x_2) = 0.3$ and $m_0(x_3) = P(X = x_3) = 0.1$. Applying Bayes rule (12) yields:

$$\begin{cases} P(x_1|Z_1 \cap Z_2) & = \frac{A_2(x_1)}{GA_2} = \frac{0.2 \cdot 0.5/0.6}{2.2667} = \frac{0.1667}{2.2667} \approx 0.0735 \\ P(x_2|Z_1 \cap Z_2) & = \frac{A_2(x_2)}{GA_2} = \frac{0.3 \cdot 0.1/0.3}{2.2667} = \frac{0.1000}{2.2667} \approx 0.0441 \\ P(x_3|Z_1 \cap Z_2) & = \frac{A_2(x_3)}{GA_2} = \frac{0.5 \cdot 0.4/0.1}{2.2667} = \frac{2.0000}{2.2667} \approx 0.8824 \end{cases}$$

Applying Dempster's rule yields $m_{DS}(x_i) \neq P(x_i|Z_1 \cap Z_2)$ because:

$$\begin{cases} m_{DS}(x_1) & = \frac{1}{1-0.9110} \cdot 0.2 \cdot 0.5 \cdot 0.6 = \frac{0.060}{0.089} \approx 0.6742 \\ m_{DS}(x_2) & = \frac{1}{1-0.9110} \cdot 0.3 \cdot 0.1 \cdot 0.3 = \frac{0.009}{0.089} \approx 0.1011 \\ m_{DS}(x_3) & = \frac{1}{1-0.9110} \cdot 0.5 \cdot 0.4 \cdot 0.1 = \frac{0.020}{0.089} \approx 0.2247 \end{cases}$$

Therefore, one has in general[6]:

$$DS(m_1(.), \ldots, m_s(.); m_0(.)) \neq Bayes(P(X|Z_1), \ldots, P(X|Z_s); P(X)) \quad (38)$$

V. Conclusions

In this paper, we have analyzed in details the expression and the properties of Bayes rule of combination based on statistical conditional independence assumption, as well as the emblematic Dempster's rule of combination of belief functions introduced by Shafer in his Mathematical Theory of evidence. We have clearly explained from a theoretical standpoint, and also on simple examples, why Dempster's rule is not a generalization of Bayes rule in general. The incompatibility of Dempster's rule with Bayes rule is due to its impossibility to deal with non uniform Bayesian priors in the same manner as Bayes rule does. Dempster's rule turns to be compatible with Bayes rule only in two very particular cases: 1) if all the Bayesian bba's to combine (including the prior) focus on same state (i.e. there is a perfect conjunctive consensus between the sources), or 2) if all the bba's to combine (excluding the prior) are Bayesian, and if the prior bba cannot help to discriminate a particular state of the frame of discernment (i.e. the prior bba is either vacuous, or Bayesian and uniform). Except in these two very particular cases, Dempster's rule is totally incompatible with Bayes rule. Therefore, Dempster's rule cannot be claimed to be a generalization of Bayes fusion rule, even when the bba's to combine are Bayesian.

Acknowledgment

This study was co-supported by Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National NSF of China (No.61104214, No. 61203222), and also partly supported by the project AComIn, grant 316087, funded by the FP7 Capacity Programme.

[6] but in the very degenerate case when manipulating *deterministic* Bayesian bba's, which is of little practical interest from the fusion standpoint.

References

[1] L.A. Zadeh, On the validity of Dempster's rule of combination, Memo M79/24, Univ. of California, Berkeley, CA, U.S.A., 1979.

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

[3] A. Dempster, A generalization of bayesian inference, J. R. Stat. Soc. B 30, pp. 205–247, 1968.

[4] G. Shafer, A Mathematical theory of evidence, Princeton University Press, Princeton, NJ, U.S.A., 1976.

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

[6] 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, 1986.

[7] J. Lemmer, Confidence factors, empiricism and the Dempster-Shafer theory of evidence, Proc. of UAI '85, pp. 160–176, Los Angeles, CA, U.S.A., July 10–12, 1985.

[8] J. Pearl, Do we need higher-order probabilities, and, if so, what do they mean?, Proc. of UAI '87, pp. 47–60, Seattle, WA, U.S.A., July 10–12, 1987.

[9] F. Voorbraak, On the justification of Demspster's rule of combination, Dept. of Phil., Utrecht Univ., Logic Group, Preprint Ser., No. 42, Dec. 1988.

[10] G. Shafer, Perspectives on the theory and practice of belief functions, IJAR, Vol. 4, No. 5–6, pp. 323–362, 1990.

[11] J. Pearl, Reasoning with belief functions: an analysis of compatibility, IJAR, Vol. 4, No. 5–6, pp. 363–390, 1990.

[12] J. Pearl, Rejoinder of comments on "Reasoning with belief functions: An analysis of compatibility", IJAR, Vol. 6, No. 3, pp. 425–443, May 1992.

[13] G.M. Provan, The validity of Dempster-Shafer belief functions, IJAR, Vol. 6., No. 3, pp. 389–399, May 1992.

[14] P. Wang, A defect in Dempster-Shafer theory, Proc. of UAI '94, pp. 560–566, Seattle, WA, U.S.A., July 29–31, 1994.

[15] 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.

[16] F. Smarandache, J. Dezert (Editors), Applications and advances of DSmT for information fusion, Vol. 3, ARP, U.S.A., 2009. http://fs.gallup.unm.edu/DSmT.htm

[17] J. Dezert, P. Wang, A. Tchamova, On the validity of Dempster-Shafer theory, Proc. of Fusion 2012 Int. Conf., Singapore, July 9–12, 2012.

[18] A. Tchamova, J. Dezert, On the Behavior of Dempster's rule of combination and the foundations of Dempster-Shafer theory, (best paper award), Proc. of 6th IEEE Int. Conf. on Intelligent Systems IS '12, Sofia, Bulgaria, Sept. 6–8, 2012.

[19] R.P. Mahler, Statistical multisource-multitarget information fusion, Chapter 4, Artech House, 2007.

[20] X.R. Li, Probability, random signals and statistics, CRC Press, 1999.

[21] P.E. Pfeiffer, Applied probability, Connexions Web site, Aug. 31, 2009. http://cnx.org/content/col10708/1.6/

[22] G. Shafer, Non-additive probabilities in the work of Bernoulli and Lambert, in Archive for History of Exact Sciences, C. Truesdell (Ed.), Springer-Verlag, Berlin, Vol. 19, No. 4, pp. 309–370, 1978.

[23] R.P. Mahler, Using a priori evidence to customize Dempster-Shafer theory, Proc. of 6th Nat. Symp. on Sensor Fusion, Vol. 1, pp. 331–345, Orlando, FL, U.S.A., April 13–15, 1993.

[24] R.P. Mahler, Classification when a priori evidence is ambiguous, Proc. SPIE Conf. on Opt. Eng. in Aerospace Sensing (Automatic Object Recognition IV), pp. 296–304, Orlando, FL, U.S.A., April 4–8, 1994.

[25] D. Fixsen, R.P. Mahler, The modified Dempster-Shafer approach to classification, IEEE Trans. on SMC, Part A, Vol. 27, No. 1, pp. 96–104, 1997.