Combination of paradoxical sources of information within the neutrosophic framework

Dr. Jean Dezert*

Abstract - ***The recent emergence of Smarandache's logic as foundations for a new general unifying theory for uncertain reasoning is becoming both a new philosophical and mathematical research field and could modify deeply our perception and understanding of our outer and inner worlds in coming years. The ability for neutrosophy to include all existing logics as special cases is undoubtedly appealing. Beside of all potential advantages of neutrosophy to handle antinomies and uncertainties, the current mathematical neutrosophic logic, does not deal directly with the important problem of combination of evidences provided by different bodies of evidence. This paper is the first attempt to develop new foundations for the combination of sources of information in a very general framework where information can be both uncertain and paradoxical. We develop a new rule of combination close to the ad-hoc Dempster-Shafer rule of combination where both conjunctions and disjunctions of assertions are explicitly taking into account in the fusion process. Through several simple examples, we show the efficiency of this new theory of plausible and paradoxical reasoning to solve problems where the Dempster-Shafer theory usually fails. Finally a theoretical bridge between the neutrosophic logic and our new theory is presented, in order to solve the delicate problem of the combination of neutrosophic evidences. The neutrosophic logic seems to be an appealing general framework (prerequesite) for dealing with uncertain and paradoxical sources of information through this new theory.***

Keywords: Data Fusion, Dempster-Shafer theory, Bodies of evidence, Neutrosophy, Neutrosophic Logic.
1991 MSC: 03B99, 03E99, 60A99

1 Introduction

The processing of uncertain information has always been a hot topic of research since mainly the 18th century. Up to middle of the 20th century, most theoretical advances have been devoted to the theory of probabilities through the works of eminent mathematicians like J. Bernoulli (1713), A. De Moivre (1718), T. Bayes (1763), P. Laplace (1774), K. Gauss (1823), S. Poisson (1837), I. Todhunter (1873), J. Bertrand (1889), E. Borel (1909), R. Fisher (1930), F. Ramsey (1931), A. Kolmogorov (1933), H. Jeffreys (1939), R. Cox (1946), I. Good (1950), R. Carnap (1950), G. Polya (1954), R. Jeffrey (1957), B. De Finetti (1958), M. Kendall (1963), L. Savage (1967),T. Fine (1973), E. Jaynes (1995) to name just few of them. With the development of the computers, the last half of the 20th century has became very prolific for the development new original theories dealing with uncertainty and imprecise information. Mainly three major theories are available now as alternative of the theory of probabilities for the automatic plausible reasoning in expert systems: the fuzzy set theory developed by L. Zadeh in sixties (1965), the Shafer's theory of evidence in seventies (1976) and the theory of possibilities by D. Dubois and H. Prade in eighties (1985). Only recently a new general and original theory, called neutrosophy, has been developed by F. Smarandache to unify all these existing theories in a common global framework. This paper is focused on the development of a new theory of plausible and paradoxical reasoning within the neutrosophical framework. After a brief presentation of probability and Dempster-Shafer theories in sections 2 and 3, we propose the foundations for a new theory in section 4 and discuss about the justification of our new rule of combination of uncertain and paradoxical sources of evidences. Several examples of our new inference will also been presented. In the last section of this paper, we will show how our theory can serve as theoretical tool for the problem of the combination of neutrosophical evidences.

2 The probability theory

Let $\Theta = \{\theta_i, i = 1, \ldots, n\}$ be a finite discrete set of *exhaustive* and *exclusive* hypotheses or outcomes of a random experiment. The probability $P(A)$ of $A \subseteq \Theta$ has been defined and interpreted differently (mainly through the geometrical approach, the subjective approach and the frequency approach [35]) since the seventieth century. The frequency approach $P\{A\}$ is defined as the ratio of the number of possible outcomes for event A to the total number of possible outcomes for space Θ. This is still now the easiest approach to introduce the notion of probability (chance) at a low mathematical level. The foundations of the probability theory with a new interpretation as the logic of science can be found through the works of E.T. Jaynes in [27, 25, 26].

*Dr. Jean Dezert is partially with Onera, 29 Av. Division Leclerc,92320 Châtillon, France. This personal research work is not supported by Onera and does not reflect the position or policy of Onera. Email: Jean.Dezert@onera.fr

2.1 Axiomatic approach of the theory of probabilities

Within the frequency approach of probability, one implicitly assumes that each elementary element of Θ is equally probable. Hence the definition of probability itself turns to fall actually in a vicious circle definition. Moreover the principle of sufficient reason (the hypothesis of equiprobable repartition for elementary components of Θ in case of no prior information), called also the principle of indifference, has been strongly critized especially for cases involving infinitely many possible outcomes because this can lead to confusing paradoxes. That is why, since the work of A. Kolmogorov in 1933, the axiomatic of the probability theory based on σ-algebras and measure theory has been definitely adopted. We remind now the four axioms of modern theory of probability:

A1: (Nonnegativity)

$$0 \leq P\{A\} \leq 1 \tag{1}$$

A2: (Unity) Any sure event (the sample space) has probability one

$$P\{\Theta\} = 1 \tag{2}$$

A3: (Finite additivity) If $A_1, \ldots, A_n$ are disjoint events, then

$$P\{A_1 \cup \ldots \cup A_n\} = P\{A_1\} + \ldots + P\{A_n\} \tag{3}$$

A4: (countable additivity) If $A_1, A_2, \ldots$ are disjoint events

$$P\{\bigcup_{i=1}^{\infty} A_i\} = \sum_{i=1}^{\infty} P\{A_i\} \tag{4}$$

2.2 Consequences of axioms and bayesian inference

From these axioms, all other probability laws (especially total probability theorem and Bayes's rule) can be derived. In particular,

$$P\{\emptyset\} = 0 \quad \text{and} \quad P\{A^c\} = 1 - P\{A\} \tag{5}$$

$$A \subset B \Rightarrow P\{A\} \leq P\{B\} \tag{6}$$

$$\forall A, B \subset \Theta, \quad P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\} \tag{7}$$

$$\forall A_1, \ldots, A_n \subset \Theta, \quad P\{A_1 \cup \ldots \cup A_n\} \leq \sum_{i=1}^{n} P\{A_i\} \quad \text{(Boole's inequality)} \tag{8}$$

More precisely, in the general case, one has the Poincaré's equality

$$\begin{aligned} P\{A_1 \cup \ldots \cup A_n\} &= \sum_{i=1}^{n} P\{A_i\} - \sum_{i<j} P\{A_i \cap A_j\} + \ldots \\ &+(-1)^{k-1} \sum_{i_1<\ldots<i_k} P\{A_{i_1} \cap \ldots \cap A_{i_k}\} + \ldots + (-1)^n P\{\bigcap_{i=1,n} A_i\} \end{aligned} \tag{9}$$

which can be also written under a more compact form as

$$P\{A_1 \cup \ldots \cup A_n\} = \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} P\{\bigcap_{i \in I} A_i\} \tag{10}$$

The probability of an event A under the condition that event B has occured (with probability $P\{B\} \neq 0$) is called the conditional (or a posteriori) probability of A given B and is defined as

$$P\{A \mid B\} = \frac{P\{A \cap B\}}{P\{B\}} \tag{11}$$

Events A and B are said to be independent if $P\{A \cap B\} = P\{A\}P\{B\}$ or equivalently $P\{A \mid B\} = P\{A\}$ and $P\{B \mid A\} = P\{B\}$.

The probability of any event B can be recovered from any partition (i.e. a set of exhaustive and disjoint events) $A_1, \ldots, A_n$ of sample space Θ by the total probability theorem

$$P\{B\} = \sum_{i=1}^{n} P\{B \mid A_i\}P\{A_i\} \tag{12}$$

Since $P\{A \mid B\}P\{B\} = P\{A \cap B\} = P\{B \mid A\}P\{A\}$, one gets the famous Baye's formula, also called the Bayesian inference [4]

$$P\{A \mid B\} = \frac{P\{B \mid A\}P\{A\}}{P\{B\}} \tag{13}$$

2.3 Bayesian rule of combination

Suppose now that M independent sources of information (bodies of evidence) $\mathcal{B}_1, \ldots, \mathcal{B}_M$ provide M *subjective* probability functions $P_1\{.\}, \ldots, P_M\{.\}$ over the same space Θ, then the optimal bayesian fusion rule is obtained as follows (see [13] for a more general and theoretical justification).

$$P_{1,\ldots,M}\{\theta_i\} \triangleq [P_1 \oplus \ldots \oplus P_M]\{\theta_i\} = \frac{p_i^{1-M} \prod_{m=1,M} P_m\{\theta_i\}}{\sum_{i=1,n} p_i^{1-M} \prod_{m=1,M} P_m\{\theta_i\}} \tag{14}$$

where p_i is the prior probability of θ_i. This bayesian fusion rule of combination is however not defined when the sources are in full contradiction because in such case the normalization constant $\sum_{i=1,n} p_i^{1-M} \prod_{m=1,M} P_m\{\theta_i\}$ is zero. A source $\mathcal{B}_j$ is said to be in full conflict with a source $\mathcal{B}_k$ if, for all θ_i, $P_j\{\theta_i\}P_k\{\theta_i\} = 0$. When the fusion is possible and when all the prior probabilities p_i are unknown, one has then to use the principle of indifference by setting all $p_i = 1/n$ and the bayesian rule of combination reduces to

$$P_{1,\ldots,M}\{\theta_i\} = [P_1 \oplus \ldots \oplus P_M]\{\theta_i\} = \frac{\prod_{m=1,M} P_m\{\theta_i\}}{\sum_{i=1,n} \prod_{m=1,M} P_m\{\theta_i\}} \tag{15}$$

The Bayesian inference (13) can be interpreted as a special case of bayesian rule of combination (15) between two sources of information (prior and posterior information).

3 The Dempster-Shafer theory of evidence

We present now the basis of the Dempster-Shafer theory (DST) or the Mathematical Theory of Evidence (MTE) [48, 11] called also sometimes the theory of probable or evidential reasoning. The DST is usually considered as a generalization of the bayesian theory of subjective probability [52] and offers a simple and direct representation of ignorance. The DST has shown its compatibility with the classical probability theory, with boolean logic and has a feasible computational complexity [46] for problems of small dimension. The DST is a powerful theoretical tool which can be applied for the representation of incomplete knowledge, belief updating, and for combination of evidence [42, 17] through the Demspter-Shafer's rule of combination presented in the following. The Dempster-Shafer model of representation and processing of uncertainty has led to a huge number of practical applications in a wide range of domains (technical and medical diagnosis under unreliable measuring devices, information retrieval, integration of knowledge from heterogeneous sources for object identification and tracking, network reliability computation, multisensor image segmentation, autonomous navigation, safety control in large plants, map construction and maintenance, just to mention a few).

3.1 Basic probability masses

Definition

Let $\Theta = \{\theta_i, i = 1, \ldots, n\}$ be a finite discrete set of *exhaustive* and *exclusive* elements (hypotheses) called elementary elements. Θ has been called the frame of discernment of hypotheses or universe of discourse by G. Shafer. The cardinality (number of elementary elements) of Θ is denoted $|\Theta|$. The power set $\mathcal{P}(\Theta)$ of Θ which is the set of all subsets of Θ is usually noted $\mathcal{P}(\Theta) = 2^\Theta$ because its cardinality is exactly $2^{|\Theta|}$. Any element of 2^Θ is then a composite event (disjunction) of the frame of discernment. The DST starts by defining a map associated to a body of evidence $\mathcal{B}$ (source of information), called basic assignment probability (bpa) or information granule $m(.) : 2^\Theta \to [0, 1]$ such that

$$m(\emptyset) = 0 \tag{16}$$

$$\sum_{A\in 2^{\Theta}} m(A) \equiv \sum_{A\subseteq\Theta} m(A) = 1 \tag{17}$$

$m(.)$ represents the strength of some evidence provided by the source of information under consideration. Condition (16) reflects the fact that no belief ought to be committed to $\emptyset$ and condition (17) reflects the convention that one's total belief has measure one [48]. The quantity $m(A)$ is called A's basic probability number or sometimes A's basic mass. $m(A)$ corresponds to the measure of the partial belief that is committed *exactly* to A (degree of truth supported exactly by A) by the body of evidence $\mathcal{B}$ but not the total belief committed to A. All subsets A for which $m(A) > 0$ are called focal elements of m. The set of all focal elements of $m(.)$ is called the core $\mathcal{K}(m)$ of m. Note that $m(A_1)$ and $m(A_2)$ can both be 0 even if $m(A_1 \cup A_2) \neq 0$. Even more peculiar, note that $A \subset B \nRightarrow m(A) < m(B)$. Hence, the bpa $m(.)$ is in general different from a probability distribution $p(.)$.

Example

Consider $\Theta = \{\theta_1, \theta_2, \theta_3\}$, then $2^{\Theta} = \{\emptyset, \theta_1, \theta_2, \theta_3, \theta_1\cup\theta_2, \theta_1\cup\theta_3, \theta_2\cup\theta_3, \theta_1\cup\theta_2\cup\theta_3\}$. An information granule $m(.)$ on this frame of discernment Θ could be defined as

$$\begin{aligned} m(\emptyset) &\triangleq 0 & m(\theta_1\cup\theta_2\cup\theta_3) &= 0.05\\ m(\theta_1) &= 0.40 & m(\theta_1\cup\theta_2) &= 0.10\\ m(\theta_2) &= 0.20 & m(\theta_2\cup\theta_3) &= 0.10\\ m(\theta_3) &= 0.05 & m(\theta_1\cup\theta_3) &= 0.10 \end{aligned}$$

In this particular example $\mathcal{K}(m) = \{\theta_1, \theta_2, \theta_3, \theta_1\cup\theta_2, \theta_1\cup\theta_3, \theta_2\cup\theta_3, \theta_1\cup\theta_2\cup\theta_3\}$ and note that $\theta_1 \subset \{\theta_1\cup\theta_2\}$ with $m(\theta_1) > m(\theta_1\cup\theta_2)$.

3.2 Belief functions

To obtain the measure of the total belief committed to $A \in 2^{\Theta}$, one must add to $m(A)$ the masses $m(B)$ for all proper subsets $B \subset A$. G. Shafer has defined the belief (credibility) function $\text{Bel}(.) : 2^{\Theta} \to [0, 1]$ associated with bpa $m(.)$ as

$$\text{Bel}(A) = \sum_{B\subseteq A} m(B) \tag{18}$$

$\text{Bel}(A)$ summarizes all our reasons to believe in A (i.e. the lower probability to believe in A). More generally, a belief function $\text{Bel}(.)$ can be characterized without reference to the information granule $m(.)$ if $\text{Bel}(.)$ satisfies the following three conditions

$$\text{Bel}(\Theta) = 1 \tag{19}$$

$$\text{Bel}(\emptyset) = 0 \tag{20}$$

$$\forall n > 0, \forall A_1, \ldots, A_n \subset \Theta, \qquad \text{Bel}(A_1 \cup \ldots \cup A_n) \geq \sum_{\substack{I\subset\{1,\ldots,n\}\\ I\neq\emptyset}} (-1)^{|I|+1}\text{Bel}(\bigcap_{i\in I} A_i) \tag{21}$$

For any given belief function $\text{Bel}(.)$, one can always associate an unique information granule $m(.)$, called the Möbius inverse of belief function [44], and defined by [48]

$$\forall A \subseteq \theta, \qquad m(A) = \sum_{B\subseteq A} (-1)^{|A-B|}\text{Bel}(B) \tag{22}$$

The *vacuous belief function* having $\text{Bel}(\Theta) = 1$ but $\text{Bel}(A) = 0$ for all $A \neq \Theta$ describes the full ignorance on the frame of discernment Θ. The corresponding bpa $m_v(.)$ is such that $m_v(\Theta) = 1$ and $m_v(A) = 0$ for all $A \neq \Theta$.

For any given belief function $\text{Bel}(.)$ defined on frame Θ, the following inequality holds

$$\forall A, B \subseteq \Theta, \qquad \max(0, \text{Bel}(A) + \text{Bel}(B) - 1) \leq \text{Bel}(A \cap B) \leq \min(\text{Bel}(A), \text{Bel}(B)) \tag{23}$$

Bayesian belief functions

Any belief function satisfying $\text{Bel}(\emptyset) = 0$, $\text{Bel}(\Theta) = 1$ and $\text{Bel}(A \cup B) = \text{Bel}(A) + \text{Bel}(B)$ whenever $A, B \subset \Theta$ and $A \cap B = \emptyset$ is called a *Bayesian belief function*. In such case, relation (21) coincides exactly with (10) and a probability function $P(.)$ is only a particular Dempster-Shafer's belief function. In this sense, the Dempster-Shafer theory can be considered as a generalization of the probability theory.

If $\text{Bel}(.)$ is a bayesian belief function, then all focal elements are only single points of $\mathcal{P}(\Theta)$. The basic probability mass assignement $m(.)$ commits a positive number $m(\theta_i)$ only to some elementary $\theta_i \in \Theta$ (possibly all θ_i) and zero to all possible disjunctions of $\theta_1, \ldots, \theta_n$. In other words there exists a bayesian bpa $m(.) : \Theta \rightarrow [0, 1]$ such that

$$\sum_{\theta_i \in \Theta} m(\theta_i) = 1 \qquad \text{and} \qquad \forall A \subseteq \Theta, \quad \text{Bel}(A) = \sum_{\theta_i \in A} m(\theta_i) \tag{24}$$

3.3 Plausibility functions

Since the degree of belief $\text{Bel}(A)$ does not reveal to what extent one believes its negation A^c, Shafer has introduced the degree of doubt of A as the total belief of A^c. The degree of doubt is less useful than the plausibility (credibility) $\text{Pl}(A)$ of A which measures the total probability mass that can move into A. $\text{Pl}(A)$ can be interpreted as the *upper probability* of A. More precisely $\text{Pl}(A)$ is defined by

$$\text{Pl}(A) \triangleq 1 - \text{Dou}(A) = 1 - \text{Bel}(A^c) = \sum_{B \subseteq \Theta} m(B) - \sum_{B \subseteq A^c} m(B) = \sum_{B \cap A \neq \emptyset} m(B) \tag{25}$$

More generally, the dual of (21) implies

$$\forall n > 0, \forall A_1, \ldots, A_n \subset \Theta, \qquad \text{Pl}(A_1 \cap \ldots \cap A_n) \leq \sum_{\substack{I \subset \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Pl}(\bigcup_{i \in I} A_i) \tag{26}$$

The direct comparison of (18) with (25) indicates that

$$\forall A \subseteq \Theta, \qquad \text{Bel}(A) \leq \text{Pl}(A) \tag{27}$$

For any given plausibility function $\text{Pl}(.)$ defined on frame Θ, the following inequality holds

$$\forall A, B \subseteq \Theta, \qquad \max(\text{Pl}(A), \text{Pl}(B)) \leq \text{Pl}(A \cup B) \leq \min(1, \text{Pl}(A) + \text{Pl}(B)) \tag{28}$$

Let Θ be a given frame of discernment and $m(.)$ a general bpa (neither a vacuous bpa, nor a bayesian bpa) provided by a body of evidence, then it is always possible to build the following *pignistic* probability [63] (bayesian belief function) by choosing $\forall \theta_i \in \Theta, P\{\theta_i\} = \sum_{B \subseteq \Theta | \theta_i \in B} \frac{1}{|B|} m(B)$. In such case, one always has

$$\forall A \subseteq \Theta, \qquad \text{Bel}(A) \quad \leq \quad [P(A) = \sum_{\theta_i \in A} P\{\theta_i\}] \quad \leq \quad \text{Pl}(A) \tag{29}$$

Since $\text{Bel}(A)$ summarizes all our reasons to believe in A and $\text{Pl}(A)$ expresses how much we should believe in A if all currently unknown were to support A, the true belief in A is somewhere in the interval $[\text{Bel}(A), \text{Pl}(A)]$. Now suppose that the true value of a parameter under consideration is known with some uncertainty $[\text{Bel}(A), \text{Pl}(A)] \subseteq [0, 1]$, then its corresponding bpa $m(A)$ can always be constructed by choosing

$$m(A) = \text{Bel}(A) \qquad m(A \cup A^c) = \text{Pl}(A) - \text{Bel}(A) \qquad m(A^c) = 1 - \text{Pl}(A) \tag{30}$$

3.4 The Dempster's rule of combination

Glenn Shafer has proposed the ad-hoc Dempster's rule of combination (orthogonal summation), symbolized by the operator $\oplus$, to combine two so-called distinct bodies of evidences $\mathcal{B}_1$ and $\mathcal{B}_2$ over the same frame of discernment Θ. Let $\text{Bel}_1(.)$ and $\text{Bel}_2(.)$ be two belief functions over the same frame of discernment Θ and $m_1(.)$ and $m_2(.)$ their corresponding bpa masses. The combined global belief function $\text{Bel}(.) = \text{Bel}_1(.) \oplus \text{Bel}_2(.)$ is obtained from the combination of the information granules $m_1(.)$ and $m_2(.)$ as follows: $m(\emptyset) = 0$ and for any $C \neq \emptyset$ and $C \subseteq \Theta$,

$$m(C) \triangleq [m_1 \oplus m_2](C) = \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{\sum_{A \cap B \neq \emptyset} m_1(A) m_2(B)} = \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A) m_2(B)} \tag{31}$$

The summation notation $\sum_{A\cap B=C}$ must be interpreted as the sum over all $A, B \subseteq \Theta$ such that $A \cap B = C$ (interpretation for other summation notations follows directly by analogy). The orthogonal sum $m(.)$ is a proper basic probability assignment if $K \triangleq 1 - \sum_{A\cap B=\emptyset} m_1(A)m_2(B) \neq 0$. If $K = 0$,which means $\sum_{A\cap B=\emptyset} m_1(A)m_2(B) = 1$ then orthogonal sum $m(.)$ *does not exist* and the bodies of evidences $\mathcal{B}_1$ and $\mathcal{B}_2$ are said to be totally (flatly) contradictory or in *full contradiction*. Such case arises whenever the cores of $\text{Bel}_1(.)$ and $\text{Bel}_2(.)$ are disjoint or equivalently when there exists $A \subset \Theta$ such that $\text{Bel}_1(A) = 1$ and $\text{Bel}_2(A^c) = 1$. The same problem of existence has already been pointed out previously in the presentation of the optimal Bayesian fusion rule.

The quantity $\log 1/K$ is called the *weight of conflict* between the the bodies of evidences $\mathcal{B}_1$ and $\mathcal{B}_2$. It is easy to show that the Dempster's rule of combination is commutative ($m_1 \oplus m_2 = m_2 \oplus m_1$) and associative ($[m_1 \oplus m_2] \oplus m_3 = m_1 \oplus [m_2 \oplus m_3]$). The vacuous belief function such that $m_v(\Theta) = 1$ and $m_v(A) = 0$ for $A \neq \Theta$ is the identity element for $\oplus$ binary operator, i.e. $m_v \oplus m = m \oplus m_v \equiv m$. If $\text{Bel}_1(.)$ and $\text{Bel}_2(.)$ are two combinable belief functions and if $\text{Bel}_1(.)$ is Bayesian, then $\text{Bel}_1 \oplus \text{Bel}_2$ is a bayesian belief function.

This rule of combination, initially proposed by G. Shafer without a strong theoretical justification (it's only an ''ad-hoc justification''), has been criticized in the past decades by many disparagers of this theory. Nowadays, this rule of combination has however been fully justified by the axiomatic of the transferable belief model developed by Ph. Smets in [59, 60, 61, 63]. We mention the fact that such theoretical justification had been already attempted by Cheng and Kashyap in [7]. Discussions on justifications and interpretations of the DST and the Dempster's rule of combination can be found in [16, 30, 31, 32, 42, 45, 68]. An interesting discussion on the justification of Dempster's rule of combination from the information entropy viewpoint based on the measurement projection and balance principles can be found in [66]. Connection of the DST with the Fuzzy Set Theory can be found in [5, 64]. The relationship between experimental observations and the DST belief functions is currently a hot topic of research. Several models have been developed for fitting belief functions with experimental data. A very recent detailed presentation and discussion on this problem can be found in [69].

We can see a very close similitude between the Dempster's rule and the optimal bayesian fusion rule (15). Actually these two rules coincides exactly when $m_1(.)$ and $m_2(.)$ become bayesian basic probability assignments and if we accept the principle of indifference within the optimal Bayesian fusion rule.

The complexity of DS rule of combination is important in general with large frames of discernment since the computational burden of finding all pairs A and B of subsets of Θ such that $A \cap B = C$ is $o(2^{|\Theta|-|C|} \times 2^{|\Theta|-|C|})$ which is a large number. For example, if $|\Theta| = 10$ and $|C| = 2$, we will have $o(2^{16}) = o(65536)$ tests to do to find $\{A \cap B | A \cap B = C\}$

A simple example of the Dempster's rule of combination

Consider the simple frame of discernment $\Theta = \{S(\text{unny}), R(\text{ainy})\}$ about the true nature of the weather at a given location L for the next day and let consider two independent bodies of evidence $\mathcal{B}_1$ and $\mathcal{B}_2$ providing the following weather forecasts at L

$$m_1(S) = 0.80 \qquad m_1(R) = 0.12 \qquad m_1(S \cup R) = 0.08$$

$$m_2(S) = 0.90 \qquad m_2(R) = 0.02 \qquad m_2(S \cup R) = 0.08$$

Applying Dempster's rule of combination yields the following result

$$\begin{aligned}
m(S) &= (m_1 \oplus m_2)(S) = (0.72 + 0.072 + 0.064)/(1 - 0.108 - 0.016) \approx 0.9772 \\
m(R) &= (m_1 \oplus m_2)(R) = (0.0024 + 0.0096 + 0.0016)/(1 - 0.108 - 0.016) \approx 0.0155 \\
m(S \cup R) &= (m_1 \oplus m_2)(S \cup R) = 0.0064/(1 - 0.108 - 0.016) \approx 0.0073
\end{aligned}$$

Hence, in this example, the fusion of the two sources of evidence reinforces the belief that tomorrow will be a sunny day at location L (assuming that both bodies of evidence are equally reliable).

Another simple but disturbing example

L. Zadeh has given the following example of a use of the Dempster's rule which shows an unexpected result. Two doctors examine a patient and agree that it suffers from either meningitis (M), concussion (C) or brain tumor (T). Thus $\Theta = \{M, C, T\}$. Assume that the doctors agree in their low expectation of a tumor, but disagree in likely cause and provide the following diagnosis

$$m_1(M) = 0.99 \qquad m_1(T) = 0.01$$

$$m_2(C) = 0.99 \qquad m_2(T) = 0.01$$

If we now combine beliefs using Dempster's rule of combination, one gets the unexpected final conclusion $m(T) = [m_1 \oplus m_2](T) = \frac{0.0001}{1-0.0099-0.0099-0.9801} = 1$ which means that the patient suffers with certainty from brain tumor !!!. This unexpected result arises from the fact that the two bodies of evidence (doctors) agree that patient does not suffer from tumor but are in almost full contradiction for the other causes of the disease. This very simple but practical example shows the limitations of practical use of the DS theory for automated reasoning. Some care must always be taken about the degree of conflict between sources before taking final decision from the result of the Dempster's rule of combination. A justification of non effectivness of the Dempster's rule in such kind of example based on an information entropy argument has already been presented in [66].

Conditional Belief functions

Let $m_B(A) = 1$ if $B \subseteq A$ and $m_B(A) = 0$ for if $B \not\subset A$ (the subset B is the only focal element of Bel_B and its basic probability number is one). Then Bel_B is a belief function that focuses all of the belief on B (note that Bel_B is not in general a Bayesian belief function unless $\mid B \mid= 1$). If we now consider another belief function Bel over Θ combinable with Bel_B, then the orthogonal sum of Bel with Bel_B denoted as $\text{Bel}(. \mid B) = \text{Bel} \oplus \text{Bel}_B$ is defined for all $A \subset \Theta$ by [48]

$$\text{Bel}(A \mid B) = \frac{\text{Bel}(A \cup B^c) - \text{Bel}(B^c)}{1 - \text{Bel}(B^c)} \tag{32}$$

and

$$\text{Pl}(A \mid B) = \frac{\text{Pl}(A \cap B)}{\text{Pl}(B)} \tag{33}$$

If Bel is a Bayesian belief function, then

$$\text{Bel}(A \mid B) = \frac{\text{Bel}(A \cap B)}{\text{Bel}(B)} = \text{Pl}(A \mid B) \tag{34}$$

which coincides exactly with the classical conditional probability $P(A \mid B)$ defined in (11).

4 A new theory for plausible and paradoxical reasoning

4.1 Introduction

As seen in the previous Zadeh's troubling example, the use of the DST must be done only with extreme caution if one has to take a final and important decision with the result of the Dempter's rule of combination. In most (if not all) of practical applications based on the DST, some ad-hoc or heuristic recipes are added to the fusion process to correctly manage or reduce the possibility of high degree of conflict between sources. Otherwise, the fusion results lead to unreliable/dangerous conclusion or cannot provide a result at all when the degree of conflict becomes high. Even if the DST provides fruithfull results in many applications (mainly in artificial intelligence and systems expert areas) in past decades, we argue that this theory is still too limited because it is based on the following very restrictive constraints :

- C1- The DST considers a discrete and finite frame of discernment based on a set of exhaustive and exclusive elementary elements.
- C2- The bodies of evidence are assumed independent (each source of information does not take into account the knowledge of other sources) and provide a belief function on the power set 2^Θ.

These two constraints are very strong in many practical problems involving uncertain and probable reasoning and dealing with fusion of uncertain and imprecise information. A discussion about this important remark had already been discussed earlier in [33, 34, 47]. In [47], the author proposed a new partitioning management technique to overcome mainly the C2 constraint. The first constraint is very severe actually since it does not allow paradoxes on elements of the frame of discernment Θ. The DST accepts as foundation the commonly adopted principle of the third exclude. Even if at first glance, it makes sense in the traditional classical thought, we can develop a new theory which does not accept the principle of the third exclude and accepts and deals with paradoxes. This is the main purpose of this paper.

The constraint C1 assumes that each elementary hypothesis of the frame of discernment Θ is finely and precisely well defined and we are able to discriminate between all elementary hypotheses without ambiguity and difficulty. We argue that this constraint is too limited and that it is not always possible in practice to model a frame of discernment satisfying C1 even for some very simple problems where each elementary hypothesis corresponds to a fuzzy concept or attribute. In such cases, the elementary elements of the "frame of discernment" cannot be precisely separated without ambiguity such that no refinement of the frame of discernment satisfying the first constraint is possible. As a simple example, consider an armed robbery situation having a witness and the frame of discernment (associated to the possible size of the thief) having only two elementary imprecise classes $\Theta = \{\theta_1 = \text{small}, \theta_2 = \text{tall}\}$. An investigator asks the witness about the size of the thief and the witness declares that the thief was tall with bpa number $m(\theta_1) = 0.80$, small with bpa number

$m(\theta_2) = 0.15$ and is uncertain (either tall or small) with $m(\theta_1 \cup \theta_2) = 0.05$. The investigator will have to deal only with this information although the smallness and the tallness have not been precisely defined. The use of this testimony by the investigator (having in other side extra-information about the thief from other sources of information) to infer on the true size of the thief is delicate especially with the important missing information about the size of the witness (who could be either a basketball player, a dwarf or most probably has a size on the average as you and me - assuming you are neither a dwarf or a basketball player. These both hypotheses are not incompatible actually since some dwarfs really enjoy to play basketball).

In many situations, we argue that the frame of discernment Θ can only be described in terms of imprecise elements which cannot be clearly separated and which cannot be considered as fully disjoint and that the refinement of initial frame into a new one satisfying C1 is like a graal quest and cannot be accomplished. Our last remark about C1 constraint concerns the universal nature of the frame of discernment. As shown in our previous simple example, it is clear that, in general, the "same" frame of discernment is interpreted differently by the bodies of evidence. Some subjectivity, or at least some fortuitious biases, on the information provided by a source of information is almost unavoidable, otherwise this would assume (as foundation for the DST) that all bodies of evidence have an objective/universal (possibly uncertain) interpretation or measure of the phenomena under consideration. This vision seems to be too excessive because usually independent bodies of evidence provide their beliefs about some hypotheses only with respect to their own worlds of knowledge and experience. We don't go deeper here in the techniques of refinements and coarsenings of compatible frame of discernments which is a prerequesite to the Dempster's rule of combination. This has already been presented in details in chapter 6 of [48]. We just want to emphaze here that these nice appealing techniques cannot be used at all in all cases where C1 cannot be satisfied and we have more generally to accept the idea to deal with paradoxical information. To convince the reader to accept our radical new way of thought, just think about the true nature of a photon? For experts working in particle physics, photons look like particles, for physicists working in electromagnetic field theory, photons are considered as electromagnetic waves. Both interpretations are true, there is no unicity on the true nature of the photon and actually a photon holds both aspects which appears as a paradoxe for most of human minds. This notion has been accepted in modern physics only with great difficulty and many vigourous discussions about this fondamental question have held at the beginning of the 20th century between all eminent physicists at that time.

The constraint C2 hides a strong difficulty already discussed in the previous paragraph. In order to apply the Dempter's rule of combination of two independent bodies of evidence $\mathcal{B}_1$ and $\mathcal{B}_2$), it is necessary that both frames of discernment Θ_1 and Θ_2 (related to each source $\mathcal{B}_1$ and $\mathcal{B}_2$) have to be compatible and to correspond to the same "universal vision" of the possibilities of the answer of the question under consideration. This constraint is itself very difficult to satisfy actually since each source of information has usually only a personal (and maybe biaised) interpretation of elements of frame of discernment. The belief provided by each local source of information mainly depends on the own knowledge frame of the source without reference to the (inaccessible) absolute truth of the space of possibilities. Therefore, C2 is in many cases also a too strong hypothesis to accept for foundations of a general theory of probable and paradoxical reasoning. A general theory should include the possibility to deal with evidences arising from different sources of information which have no access to absolute interpretation of the elements of the frame of discernment Θ under consideration. This yields to accept paradoxical information as basis for a new general theory of probable reasoning. Actually we will show in the first example on section 4.3 that paradoxical information resulting of fusion of several bodies of evidence is very informative and can be used to help us to take legitimous final decision.

In other words, this new theory can be interpreted as a general and direct extension of probability theory and the Dempster-Shafer theory in the following sense. Let $\Theta = \{\theta_1, \theta_2\}$ be the simpliest frame of discernment involving only two elementary hypotheses (with no more additional assumptions on θ_1 and θ_2), then

- the probability theory deals with basic assignment masses $m(.) \in [0, 1]$ such that

$$m(\theta_1) + m(\theta_2) = 1$$

- the Dempster-Shafer theory extends the probability theory by dealing with basic assignment masses $m(.) \in [0, 1]$ such that

$$m(\theta_1) + m(\theta_2) + m(\theta_1 \cup \theta_2) = 1$$

- our general theory extends the two previous theories by accepting the possibility of paradoxical information and deals with new basic assignment masses $m(.) \in [0, 1]$ such that

$$m(\theta_1) + m(\theta_2) + m(\theta_1 \cup \theta_2) + m(\theta_1 \cap \theta_2) = 1$$

4.2 Hyper-power set and general basic probability assignment $m(.)$

4.2.1 Hyper-power set definition

Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ be a set of n elementary elements considered as exhaustive which cannot be precisely defined and separated so that no refinement of Θ in a new larger set Θ_{ref} of disjoint elementary hypotheses is possible and let's consider the classical set operators $\cup$ (disjunction) and $\cap$ (conjunction). The exhaustive hypothesis about Θ is not a strong constraint since when $\theta_i, i = 1, n$ does not constitute an exhaustive set of elementary possibilities, we can always add an extra element θ_0 such that $\theta_i, i = 0, n$ describes now an exhaustive set. We will assume therefore, from now on and in the following, that Θ characterizes an exhaustive frame of discernment. Θ will be called a *general* frame of discernment in the sequel to emphaze the fact that Θ does not satisfy the Dempster-Shafer C1 constraint.

The classical power set $\mathcal{P}(\Theta) = 2^\Theta$ has been defined as the set of all proper subsets of Θ when all elements θ_i are disjoint. We extend here this notion and define now the "hyper-power" set D^Θ as the set of all composite possibilities build from Θ with $\cup$ and $\cap$ operators such that $\forall A \in D^\Theta, B \in D^\Theta, (A \cup B) \in D^\Theta$ and $(A \cap B) \in D^\Theta$. Obviously, one would always have $D^\Theta \subset 2^{\Theta_{ref}}$ if the refined power set $2^{\Theta_{ref}}$ could be defined and accessible which is unfortunately not possible in general as already argued. The cardinality of hyper-power set D^Θ is majored by 2^{2^n} when Card$(\Theta) = \mid \Theta \mid = n$. The generation of hyper-power set D^Θ corresponds to the famous Dedekind's problem on enumerating the set of monotone Boolean functions (i.e., functions expressible using only AND and OR set operators) [10]. This problem is also related with the Sperner systems [65, 37] based on finite poset (called also as antichains in literature) [8]. The number of antichains on the n-set Θ are equal to the number of monotonic increasing Boolean functions of n variables, and also the number of free distributive lattices with n generators [18, 20, 28, 29, 38, 54]. Determining these numbers is exactly the Dedekind's problem. The choice of letter D in our notation D^Θ to represent the hyper-power set of Θ is in honour of the great mathematician R. Dedekind. The general solution of the Dedekind's problem (for $n > 10$) has not been found yet. We just know that the cardinality numbers of D^Θ follow the integers of the Dedekind's sequence minus one when Card$(\Theta) = n$ increases.

Examples

1. If we consider $\Theta = \{\}$ (empty set) then $D^\Theta = \{\emptyset\}$ and $\mid D^\Theta \mid = 1$
2. If we consider $\Theta = \{\theta_1\}$ then $D^\Theta = \{\emptyset, \theta_1\}$ and $\mid D^\Theta \mid = 2$
3. If we consider $\Theta = \{\theta_1, \theta_2\}$ then $D^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_1 \cup \theta_2, \theta_1 \cap \theta_2\}$ and $\mid D^\Theta \mid = 5$
4. If we consider $\Theta = \{\theta_1, \theta_2, \theta_3\}$ then

$$\begin{aligned} D^\Theta = \{&\emptyset, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cap \theta_2, \theta_1 \cap \theta_3, \theta_2 \cap \theta_3, \theta_1 \cup \theta_2 \cup \theta_3, \theta_1 \cap \theta_2 \cap \theta_3, \\ &(\theta_1 \cup \theta_2) \cap \theta_3, (\theta_1 \cup \theta_3) \cap \theta_2, (\theta_2 \cup \theta_3) \cap \theta_1, (\theta_1 \cap \theta_2) \cup \theta_3, (\theta_1 \cap \theta_3) \cup \theta_2, (\theta_2 \cap \theta_3) \cup \theta_1, \\ &(\theta_1 \cup \theta_2) \cap (\theta_1 \cup \theta_3) \cap (\theta_2 \cup \theta_3)\} \end{aligned}$$

   and $\mid D^\Theta \mid = 19$

   It is not difficult, although tedious, to check that $\forall A \in D^\Theta, B \in D^\Theta, (A \cup B) \in D^\Theta$ and $(A \cap B) \in D^\Theta$. The extension to larger frame of discernment is possible but requires a higher computational burden. The general and direct analytic computation of $\mid D^\Theta \mid$ for a n-set Θ with $n > 10$ is not known and is still under investigations in the mathematical community. Cardinality numbers $\mid D^\Theta \mid$ follow the Dedekind's sequence (minus one), i.e. $1, 2, 5, 19, 167, 7580, 7828353, \ldots$ when Card$(\Theta) = n = 0, 1, 2, 3, 4, 5, 6, \ldots$.

4.3 General basic assignment numbers

4.3.1 Definition

Let Θ be a *general* frame of discernment of the problem under consideration. We define a map $m(.) : D^\Theta \to [0, 1]$ associated to a given body of evidence $\mathcal{B}$ which can support paradoxical information, as follows

$$m(\emptyset) = 0 \tag{35}$$

$$\sum_{A \in D^\Theta} m(A) = 1 \tag{36}$$

The quantity $m(A)$ is called A's general basic probability number. As in the DST, all subsets $A \in D^\Theta$ for which $m(A) > 0$ are called focal elements of $m(.)$ and the set of all focal elements of $m(.)$ is also called the core $\mathcal{K}(m)$ of m. The belief and plausibility functions are defined in the same way as in the DST, i.e.

$$\text{Bel}(A) = \sum_{B \in D^\Theta, B \subseteq A} m(B) \tag{37}$$

$$\text{Pl}(A) = \sum_{B \in D^\Theta, B \cap A \neq \emptyset} m(B) \tag{38}$$

Note that, we don't define here explicitly the complementary A^c of a proposition A since $m(A^c)$ cannot be precisely evaluated from $\cup$ and $\cap$ operators on D^Θ since we include the possibility to deal with a complete paradoxical source of information such that $\forall A \in D^\Theta, \forall B \in D^\Theta, m(A \cap B) > 0$. These definitions are compatible with the DST definitions when the sources of information become uncertain but rational (they do not support paradoxical information). We still have $\forall A \in D^\Theta, \text{Bel}(A) \leq \text{Pl}(A)$.

4.3.2 Construction of pignistic probabilities from general basic assignment numbers

The construction of a pignistic probability measure from the general basic numbers $m(.)$ over D^Θ with $|\Theta| = n$ is still possible and is given by the general expression of the form

$$\forall i = 1, \ldots, n \qquad P\{\theta_i\} = \sum_{A \in D^\Theta} \alpha_{\theta_i}(A) m(A) \tag{39}$$

where $\alpha_{\theta_i}(A) \in [0, 1]$ are weighting coefficients which depend on the inclusion or non-inclusion of θ_i with respect to proposition A. No general analytic expression for $\alpha_{\theta_i}(A)$ has been derived yet even if $\alpha_{\theta_i}(A)$ can be obtained explicitely for simple examples. When general bpa $m(.)$ reduces to classical bpa (i.e. the DS bpa without paradoxe), then $\alpha_{\theta_i}(A) = \frac{1}{|A|}$ if $\theta_i \subseteq A$ and therefore one gets

$$\forall i = 1, \ldots, n \qquad P\{\theta_i\} = \sum_{A \subseteq \Theta | \theta_i \in A} \frac{1}{|A|} m(A) \tag{40}$$

We present now two examples of pignistic probabilities reconstruction from a general and non degenerated bpa $m(.)$ (i.e. $\nexists A \in D^\Theta$ with $A \neq \emptyset$ such that $m(A) = 0$) over D^Θ.

- Example 1 : If $\Theta = \{\theta_1, \theta_2\}$ then

$$P\{\theta_1\} = m(\theta_1) + \frac{1}{2} m(\theta_1 \cup \theta_2) + \frac{1}{2} m(\theta_1 \cap \theta_2)$$

$$P\{\theta_2\} = m(\theta_2) + \frac{1}{2} m(\theta_1 \cup \theta_2) + \frac{1}{2} m(\theta_1 \cap \theta_2)$$

- Example 2 : If $\Theta = \{\theta_1, \theta_2, \theta_3\}$ then

$$\begin{aligned}
P\{\theta_1\} = & m(\theta_1) + \frac{1}{2} m(\theta_1 \cup \theta_2) + \frac{1}{2} m(\theta_1 \cup \theta_3) + \frac{1}{2} m(\theta_1 \cap \theta_2) + \frac{1}{2} m(\theta_1 \cap \theta_3) \\
& + \frac{1}{3} m(\theta_1 \cup \theta_2 \cup \theta_3) + \frac{1}{3} m(\theta_1 \cap \theta_2 \cap \theta_3) \\
& + \frac{1/2 + 1/3}{3} m((\theta_1 \cup \theta_2) \cap \theta_3) + \frac{1/2 + 1/3}{3} m((\theta_1 \cup \theta_3) \cap \theta_2) + \frac{1/2 + 1/2 + 1/3}{3} m((\theta_2 \cup \theta_3) \cap \theta_1) \\
& + \frac{1/2 + 1/2 + 1/3}{5} m((\theta_1 \cap \theta_2) \cup \theta_3) + \frac{1/2 + 1/2 + 1/3}{5} m((\theta_1 \cap \theta_3) \cup \theta_2) \\
& + \frac{1 + 1/2 + 1/2 + 1/3}{5} m((\theta_2 \cap \theta_3) \cup \theta_1) + \frac{1/2 + 1/2 + 1/3}{4} m((\theta_1 \cup \theta_2) \cap (\theta_1 \cup \theta_2) \cap (\theta_2 \cup \theta_3))
\end{aligned}$$

$$\begin{aligned}
P\{\theta_2\} = & m(\theta_2) + \frac{1}{2} m(\theta_1 \cup \theta_2) + \frac{1}{2} m(\theta_2 \cup \theta_3) + \frac{1}{2} m(\theta_1 \cap \theta_2) + \frac{1}{2} m(\theta_2 \cap \theta_3) \\
& + \frac{1}{3} m(\theta_1 \cup \theta_2 \cup \theta_3) + \frac{1}{3} m(\theta_1 \cap \theta_2 \cap \theta_3) \\
& + \frac{1/2 + 1/3}{3} m((\theta_1 \cup \theta_2) \cap \theta_3) + \frac{1/2 + 1/2 + 1/3}{3} m((\theta_1 \cup \theta_3) \cap \theta_2) + \frac{1/2 + 1/3}{3} m((\theta_2 \cup \theta_3) \cap \theta_1) \\
& + \frac{1/2 + 1/2 + 1/3}{5} m((\theta_1 \cap \theta_2) \cup \theta_3) + \frac{1 + 1/2 + 1/2 + 1/3}{5} m((\theta_1 \cap \theta_3) \cup \theta_2) \\
& + \frac{1/2 + 1/2 + 1/3}{5} m((\theta_2 \cap \theta_3) \cup \theta_1) + \frac{1/2 + 1/2 + 1/3}{4} m((\theta_1 \cup \theta_2) \cap (\theta_1 \cup \theta_2) \cap (\theta_2 \cup \theta_3))
\end{aligned}$$

$$
\begin{aligned}
P\{\theta_3\} =& m(\theta_3) + \frac{1}{2}m(\theta_1 \cup \theta_3) + \frac{1}{2}m(\theta_2 \cup \theta_3) + \frac{1}{2}m(\theta_1 \cap \theta_3) + \frac{1}{2}m(\theta_2 \cap \theta_3) \\
&+ \frac{1}{3}m(\theta_1 \cup \theta_2 \cup \theta_3) + \frac{1}{3}m(\theta_1 \cap \theta_2 \cap \theta_3) \\
&+ \frac{1/2 + 1/2 + 1/3}{3}m((\theta_1 \cup \theta_2) \cap \theta_3) + \frac{1/2 + 1/3}{3}m((\theta_1 \cup \theta_3) \cap \theta_2) + \frac{1/2 + 1/3}{3}m((\theta_2 \cup \theta_3) \cap \theta_1) \\
&+ \frac{1 + 1/2 + 1/2 + 1/3}{5}m((\theta_1 \cap \theta_2) \cup \theta_3) + \frac{1/2 + 1/2 + 1/3}{5}m((\theta_1 \cap \theta_3) \cup \theta_2) \\
&+ \frac{1/2 + 1/2 + 1/3}{5}m((\theta_2 \cap \theta_3) \cup \theta_1) + \frac{1/2 + 1/2 + 1/3}{4}m((\theta_1 \cup \theta_2) \cap (\theta_1 \cup \theta_2) \cap (\theta_2 \cup \theta_3))
\end{aligned}
$$

The evaluation of weighting coefficients $\alpha_{\theta_i}(A)$ has been obtained from the geometrical interpretation of the relative contribution of the distinct parts of A with proposition θ_i under consideration. For example, consider $A = (\theta_1 \cap \theta_2) \cup \theta_3$ which corresponds to the area $a_1 \cup a_2 \cup a_3 \cup a_4 \cup a_5$ on the following Venn diagram.

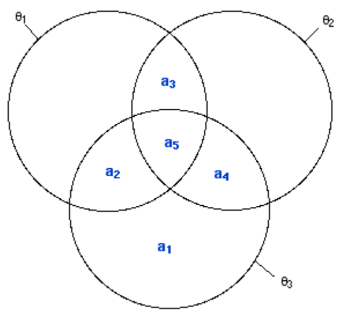

Figure 1: Representation of $A = (\theta_1 \cap \theta_2) \cup \theta_3 \equiv a_1 \cup a_2 \cup a_3 \cup a_4 \cup a_5$

a_1 which is shared only by θ_3 will contribute to θ_3 with weight 1; a_2 which is shared by θ_1 and θ_3 will contribute to θ_3 with weight 1/2; a_3 which is not shared by θ_3 will contribute to θ_3 with weight 0; a_4 which is shared by θ_2 and θ_3 will contribute to θ_3 with weight 1/2; a_5 which is shared by both θ_1,θ_2 and θ_3 will contribute to θ_3 with weight 1/3. Since moreover, one must have $\forall A \in D^\Theta$ with $m(A) \neq 0$, $\sum_{i=1}^{n} \alpha_{\theta_i}(A)m(A) = m(A)$, it is necessary to normalize $\alpha_{\theta_i}(A)$. Therefore $\alpha_{\theta_3}(A)$ is given by

$$
\alpha_{\theta_3}(A) = \frac{1 + 1/2 + 1/2 + 1/3}{5} \quad \text{and similarly} \quad \alpha_{\theta_2}(A) = \frac{1/2 + 1/2 + 1/3}{5}, \quad \alpha_{\theta_1}(A) = \frac{1/2 + 1/2 + 1/3}{5}
$$

All $\alpha_{\theta_i}(A), \forall A \in D^\Theta$ entering in derivation of $P\{\theta_i\}$ can be obtained using similar process.

4.4 General rule of combination of paradoxical sources of evidence

4.4.1 The rule of combination

Let's consider now two distinct (but potentially paradoxical) bodies of evidences $\mathcal{B}_1$ and $\mathcal{B}_2$ over the same frame of discernment Θ with belief functions $\text{Bel}_1(.)$ and $\text{Bel}_2(.)$ associated with information granules $m_1(.)$ and $m_2(.)$. The combined global belief function $\text{Bel}(.) = \text{Bel}_1(.) \oplus \text{Bel}_2(.)$ is obtained through the combination of the granules $m_1(.)$ and $m_2(.)$ by the simple rule

$$
\forall C \in D^\Theta, \qquad m(C) \triangleq [m_1 \oplus m_2](C) = \sum_{A,B \in D^\Theta, A \cap B = C} m_1(A)m_2(B) \tag{41}
$$

Since D^Θ is closed under $\cup$ and $\cap$ operators, this new rule of combination guarantees that $m(.) : D^\Theta \to [0,1]$ is a proper general information granule statisfying (35) and (36). The global belief function $\text{Bel}(.)$ is then obtained from the granule $m(.)$ through (37). This rule of combination is commutative and associative and can always be used for fusion of paradoxical and/or rational sources of information (bodies of evidence). Obviously, the decision process will have to be made with more caution to take final decision based on the general granule $m()$ when internal paradoxical conflicts arise.

It is important to note that any fusion of sources of information generates either uncertainties, paradoxes or in general both. This is intrinsic to the general fusion process itself. For instance, let's consider the frame of discernment $\Theta = \{\theta_1, \theta_2\}$ and the following very simple examples:

- If we consider the two rational information granules

$$m_1(\theta_1) = 0.80 \qquad m_1(\theta_2) = 0.20 \qquad m_1(\theta_1 \cup \theta_2) = 0 \qquad m_1(\theta_1 \cap \theta_2) = 0$$

$$m_2(\theta_1) = 0.90 \qquad m_2(\theta_2) = 0.10 \qquad m_2(\theta_1 \cup \theta_2) = 0 \qquad m_2(\theta_1 \cap \theta_2) = 0$$

  then

$$m(\theta_1) = 0.72 \qquad m(\theta_2) = 0.02 \qquad m(\theta_1 \cup \theta_2) = 0 \qquad m(\theta_1 \cap \theta_2) = 0.26$$

- If we consider the two uncertain information granules

$$m_1(\theta_1) = 0.80 \qquad m_1(\theta_2) = 0.15 \qquad m_1(\theta_1 \cup \theta_2) = 0.05 \qquad m_1(\theta_1 \cap \theta_2) = 0$$

$$m_2(\theta_1) = 0.90 \qquad m_2(\theta_2) = 0.05 \qquad m_2(\theta_1 \cup \theta_2) = 0.05 \qquad m_2(\theta_1 \cap \theta_2) = 0$$

  then

$$m(\theta_1) = 0.805 \qquad m(\theta_2) = 0.0175 \qquad m(\theta_1 \cup \theta_2) = 0.0025 \qquad m(\theta_1 \cap \theta_2) = 0.175$$

- If we consider the two paradoxical information granules

$$m_1(\theta_1) = 0.80 \qquad m_1(\theta_2) = 0.15 \qquad m_1(\theta_1 \cup \theta_2) = 0 \qquad m_1(\theta_1 \cap \theta_2) = 0.05$$

$$m_2(\theta_1) = 0.90 \qquad m_2(\theta_2) = 0.05 \qquad m_2(\theta_1 \cup \theta_2) = 0 \qquad m_2(\theta_1 \cap \theta_2) = 0.05$$

  then

$$m(\theta_1) = 0.72 \qquad m(\theta_2) = 0.0075 \qquad m(\theta_1 \cup \theta_2) = 0 \qquad m(\theta_1 \cap \theta_2) = 0.2725$$

- If we consider the two uncertain and paradoxical information granules

$$m_1(\theta_1) = 0.80 \qquad m_1(\theta_2) = 0.10 \qquad m_1(\theta_1 \cup \theta_2) = 0.05 \qquad m_1(\theta_1 \cap \theta_2) = 0.05$$

$$m_2(\theta_1) = 0.90 \qquad m_2(\theta_2) = 0.05 \qquad m_2(\theta_1 \cup \theta_2) = 0.03 \qquad m_2(\theta_1 \cap \theta_2) = 0.02$$

  then

$$m(\theta_1) = 0.789 \qquad m(\theta_2) = 0.0105 \qquad m(\theta_1 \cup \theta_2) = 0.0015 \qquad m(\theta_1 \cap \theta_2) = 0.199$$

Note that this general fusion rule can also be used with intuitionist logic in which the sum of bpa is allowed to be less than one ($\sum m(A) < 1$) and with the paraconsistent logic in which the sum of bpa is allowed to be greater than one ($\sum m(A) > 1$) as well. In such cases, the fusion result does not provide in general $\sum m(A) = 1$. By example, let's consider the fusion of the paraconsistent source $\mathcal{B}_1$ with $m_1(\theta_1) = 0.60$, $m_1(\theta_2) = 0.30$, $m_1(\theta_1 \cup \theta_2) = 0.20$, $m_1(\theta_1 \cap \theta_2) = 0.10$ with the intuitionist source $\mathcal{B}_2$ with $m_2(\theta_1) = 0.50$, $m_2(\theta_2) = 0.20$, $m_2(\theta_1 \cup \theta_2) = 0.10$, $m_2(\theta_1 \cap \theta_2) = 0.10$. In such case, the fusion result of these two sources of information yields the following global paraconsistent bpa $m(.)$

$$m(\theta_1) = 0.46 \qquad m(\theta_2) = 0.13 \qquad m(\theta_1 \cup \theta_2) = 0.02 \qquad m(\theta_1 \cap \theta_2) = 0.47 \qquad \Rightarrow \sum m = 1.08 > 1$$

In practice, for the sake of fair comparison between several alternatives or choices, it is better and simplier to deal with normalized bpa to take a final important decision for the problem under consideration. A nice property of the new rule of combination of non-normalized bpa is its invariance to the pre- or post-normalization process as we will show right now. In the previous example, the post-normalization of bpa $m(.)$ will yield the new bpa $m'(.)$

$$m'(\theta_1) = \frac{0.46}{1.08} \approx 0.426 \qquad m'(\theta_2) = \frac{0.13}{1.08} \approx 0.12 \qquad m'(\theta_1 \cup \theta_2) = \frac{0.02}{1.08} \approx 0.019 \qquad m'(\theta_1 \cap \theta_2) = \frac{0.47}{1.08} \approx 0.435$$

The fusion of pre-normalization of bpa $m_1(.)$ and $m_2(.)$ will yield the same normalized bpa $m'(.)$ since

$$m_1'(\theta_1) = \frac{0.6}{1.2} = 0.50 \qquad m_1'(\theta_2) = \frac{0.3}{1.2} = 0.25 \qquad m_1'(\theta_1 \cup \theta_2) = \frac{0.2}{1.2} \approx 0.17 \qquad m_1'(\theta_1 \cap \theta_2) = \frac{0.1}{1.2} \approx 0.08$$

$$m_2'(\theta_1) = \frac{0.5}{0.9} \approx 0.56 \qquad m_2'(\theta_2) = \frac{0.2}{0.9} \approx 0.22 \qquad m_2'(\theta_1 \cup \theta_2) = \frac{0.1}{0.9} \approx 0.11 \qquad m_2'(\theta_1 \cap \theta_2) = \frac{0.1}{0.9} \approx 0.11$$

$$m'(\theta_1) \approx 0.426 \qquad m'(\theta_2) \approx 0.12 \qquad m'(\theta_1 \cup \theta_2) \approx 0.019 \qquad m'(\theta_1 \cap \theta_2) \approx 0.435$$

It is easy to verify from the general fusion table that the pre or post normalization step yields always the same global normalized bpa even for the general case (when $|\Theta| = n$) because the post-normalization constant $\sum m(A)$ is always equal to the product of the two pre-normalization constants $\sum m_1(A)$ and $\sum m_2(A)$.

4.4.2 Justification of the new rule of combination from information entropy

Let's consider two bodies of evidence $\mathcal{B}_1$ and $\mathcal{B}_2$ characterized respectively by their bpa $m_1(.)$,$m_2(.)$ and their cores $\mathcal{K}_1 = \mathcal{K}(m_1)$, $\mathcal{K}_2 = \mathcal{K}(m_2)$. Following Sun's notation [66], each source of information will be denoted

$$\mathcal{B}_1 = \begin{bmatrix} \mathcal{K}_1 \\ m_1 \end{bmatrix} = \begin{bmatrix} f_1^{(1)} & f_2^{(1)} & \ldots & f_k^{(1)} \\ m_1(f_1^{(1)}) & m_1(f_2^{(1)}) & \ldots & m_1(f_k^{(1)}) \end{bmatrix} \quad (42)$$

$$\mathcal{B}_2 = \begin{bmatrix} \mathcal{K}_2 \\ m_2 \end{bmatrix} = \begin{bmatrix} f_1^{(2)} & f_2^{(2)} & \ldots & f_l^{(2)} \\ m_2(f_1^{(2)}) & m_2(f_2^{(2)}) & \ldots & m_2(f_l^{(2)}) \end{bmatrix} \quad (43)$$

where $f_i^{(1)}, i = 1, k$ are focal elements of $\mathcal{B}_1$ and $f_j^{(2)}, j = 1, l$ are focal elements of $\mathcal{B}_2$.

Let's consider now the combined information associated with a new body of evidence $\mathcal{B}$ resulting from the fusion of $\mathcal{B}_1$ and $\mathcal{B}_2$ having bpa $m(.)$ with core $\mathcal{K}$. We denote $\mathcal{B}$ as

$$\mathcal{B} \triangleq \mathcal{B}_1 \oplus \mathcal{B}_2 = \begin{bmatrix} \mathcal{K} \\ m \end{bmatrix} = \begin{bmatrix} f_1^{(1)} \cap f_1^{(2)} & f_1^{(1)} \cap f_2^{(2)} & \ldots & f_k^{(1)} \cap f_l^{(2)} \\ m(f_1^{(1)} \cap f_1^{(2)}) & m(f_1^{(1)} \cap f_2^{(2)}) & \ldots & m(f_k^{(1)} \cap f_l^{(2)}) \end{bmatrix} \quad (44)$$

The fusion of 2 informations granules can be represented with the general table fusion as follows

$\oplus$	$m_1(f_1^{(1)})$	$m_1(f_2^{(1)})$	$\ldots$	$m_1(f_i^{(1)})$	$\ldots$	$m_1(f_k^{(1)})$
$m_2(f_1^{(2)})$	$m(f_1^{(1)} \cap f_1^{(2)})$	$m(f_2^{(1)} \cap f_1^{(2)})$	$\ldots$	$m(f_i^{(1)} \cap f_1^{(2)})$	$\ldots$	$m(f_k^{(1)} \cap f_1^{(2)})$
$m_2(f_2^{(2)})$	$m(f_1^{(1)} \cap f_2^{(2)})$	$m(f_2^{(1)} \cap f_2^{(2)})$	$\ldots$	$m(f_i^{(1)} \cap f_2^{(2)})$	$\ldots$	$m(f_k^{(1)} \cap f_2^{(2)})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$m_2(f_j^{(2)})$	$m(f_1^{(1)} \cap f_j^{(2)})$	$m(f_2^{(1)} \cap f_j^{(2)})$	$\ldots$	$m(f_i^{(1)} \cap f_j^{(2)})$	$\ldots$	$m(f_k^{(1)} \cap f_j^{(2)})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$m_2(f_l^{(2)})$	$m(f_1^{(1)} \cap f_l^{(2)})$	$m(f_2^{(1)} \cap f_l^{(2)})$	$\ldots$	$m(f_i^{(1)} \cap f_l^{(2)})$	$\ldots$	$m(f_k^{(1)} \cap f_l^{(2)})$

We look for the optimal rule of combination, i.e. the bpa $m(.) = m_1(.) \oplus m_2(.)$ which maximizes the joint entropy of the two information sources. The justification for the Maxent criteria is discussed in [24, 27]. Thus, one has to find $m(.)$ such that [66, 67].

$$\max_m[H(m)] \equiv \max_m \left[-\sum_{i=1}^{k}\sum_{j=1}^{l} m(f_i^{(1)} \cap f_j^{(2)}) \log[m(f_i^{(1)} \cap f_j^{(2)})] \right] \equiv -\min_m[-H(m)] \quad (45)$$

satisfying both

- the measurement projection principle (marginal bpa), i.e. $\forall i = 1, \ldots, k$ and $\forall j = 1, \ldots, l$

$$m_1(f_i^{(1)}) = \sum_{j=1}^{l} m(f_i^{(1)} \cap f_j^{(2)}) \qquad \text{and} \qquad m_2(f_j^{(2)}) = \sum_{i=1}^{k} m(f_i^{(1)} \cap f_j^{(2)}) \quad (46)$$

  These constraints state that the marginal bpa $m_1(.)$ is obtained by the summation over each column of the fusion table and the marginal bpa $m_2(.)$ is obtained by the summation over each row of the fusion table.

- the measurement balance principle (the sum of all cells of the fusion table must be unity)

$$\sum_{i=1}^{k}\sum_{j=1}^{l} m(f_i^{(1)} \cap f_j^{(2)}) = 1 \quad (47)$$

Using the concise notation $m_{ij} \triangleq m(f_i^{(1)} \cap f_j^{(2)})$, the Lagrangian associated with this optimization problem under equality constraints is given by (we consider here the minimization of $-J(m)$ appearing in r.h.s of (45))

$$\begin{aligned} \mathcal{L}(m, \lambda) = & \sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij} \ln[m_{ij}] \\ & + \sum_{i=1}^{k} \lambda_i [m_1(f_i^{(1)}) - \sum_{j=1}^{l} m_{ij}] + \sum_{j=1}^{l} \gamma_i [m_2(f_j^{(2)}) - \sum_{i=1}^{k} m_{ij}] + \eta[\sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij} - 1] \end{aligned} \quad (48)$$

which can be written more concisely as

$$\mathcal{L}(m,\lambda) = -H(m) + \lambda' \mathrm{g}(m) \tag{49}$$

where $m = [m_{11}\ m_{12}\ \ldots\ m_{kl}]'$ and

$$\lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \gamma_1 \\ \vdots \\ \gamma_l \\ \eta \end{bmatrix} \qquad \text{and} \qquad \mathrm{g}(m) = \begin{bmatrix} m_1(f_1^{(1)}) - \sum_{j=1}^{l} m_{1j} \\ \vdots \\ m_1(f_k^{(1)}) - \sum_{j=1}^{l} m_{kj} \\ m_2(f_1^{(2)}) - \sum_{i=1}^{k} m_{i1} \\ \vdots \\ m_2(f_l^{(2)}) - \sum_{i=1}^{k} m_{il} \\ \sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij} - 1 \end{bmatrix} \tag{50}$$

Following the classical method of Lagrange multipliers, one has to find optimal solution (m^*, λ^*) such that

$$\frac{\partial \mathcal{L}}{\partial m}(m^*,\lambda^*) = \mathbf{0} \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial \lambda}(m^*,\lambda^*) = \mathbf{0} \tag{51}$$

The first $k \times l$ equations express the general solution $m[\lambda]$ and the $k+l+1$ last equations determine λ^* and therefore by substitution into $m[\lambda]$, the optimal solution $m^* = m[\lambda^*]$. One has to solve

$$\frac{\partial \mathcal{L}}{\partial m} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial m_{11}} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial m_{ij}} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial m_{kl}} \end{bmatrix} = \begin{bmatrix} \ln(m_{11}) + 1 + \eta - \lambda_1 - \gamma_1 \\ \vdots \\ \ln(m_{ij}) + 1 + \eta - \lambda_i - \gamma_j \\ \vdots \\ \ln(m_{kl}) + 1 + \eta - \lambda_k - \gamma_l \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \tag{52}$$

which yields $\forall i, j$,

$$m_{ij} = e^{-\eta-1} e^{\lambda_i} e^{\gamma_j} \tag{53}$$

and

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \lambda_1} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial \lambda_k} \\ \frac{\partial \mathcal{L}}{\partial \gamma_1} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial \gamma_l} \\ \frac{\partial \mathcal{L}}{\partial \eta} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} = \mathbf{0} \Leftrightarrow \begin{bmatrix} e^{-\eta-1}\sum_{j=1}^{l} e^{\lambda_1} e^{\gamma_j} \\ \vdots \\ e^{-\eta-1}\sum_{j=1}^{l} e^{\lambda_k} e^{\gamma_j} \\ e^{-\eta-1}\sum_{i=1}^{k} e^{\gamma_1} e^{\lambda_i} \\ \vdots \\ e^{-\eta-1}\sum_{i=1}^{k} e^{\gamma_l} e^{\lambda_i} \\ e^{-\eta-1}\sum_{i=1}^{k}\sum_{j=1}^{l} e^{\gamma_l} e^{\lambda_i} \end{bmatrix} = \begin{bmatrix} m_1(f_1^{(1)}) \\ \vdots \\ m_1(f_k^{(1)}) \\ m_2(f_1^{(2)}) \\ \vdots \\ m_2(f_l^{(2)}) \\ 1 \end{bmatrix} \tag{54}$$

The last constraint in (54) can also be written as

$$e^{-\eta-1}\sum_{i=1}^{k}\sum_{j=1}^{l} e^{\gamma_l} e^{\lambda_i} = e^{-\eta-1}(\sum_{i=1}^{k} e^{\lambda_i})(\sum_{j=1}^{l} e^{\gamma_l}) = 1 \tag{55}$$

Now with basic algebraic manipulation, the optimal global bpa $m_{ij}\ \forall i, j$ we are searching for, can be expressed as

$$\begin{aligned} m_{ij} &= e^{-\eta-1} e^{\lambda_i} e^{\gamma_j} \\ &= e^{-\eta-1} e^{\lambda_i} e^{\gamma_j} \times \overbrace{e^{-\eta-1}(\sum_{i=1}^{k} e^{\lambda_i})(\sum_{j=1}^{l} e^{\gamma_l})}^{1} \\ &= \underbrace{(e^{-\eta-1} e^{\lambda_i} \sum_{j=1}^{l} e^{\gamma_l})}_{m_1(f_i^{(1)})} \underbrace{(e^{-\eta-1} e^{\gamma_j} \sum_{i=1}^{k} e^{\lambda_i})}_{m_2(f_j^{(2)})} \end{aligned}$$

Thus, the solution of the maximisation of the joint entropy is obtained by choosing $\forall i, j$

$$m_{ij} = m(f_i^{(1)} \cap f_j^{(2)}) = m_1(f_i^{(1)})m_2(f_j^{(2)}) \tag{56}$$

Since it may exist several combinations yielding to the same focal element, the bpa of all focal elements equal to $f_i^{(1)} \cap f_j^{(2)}$ over the fusion space is

$$m(f_i^{(1)} \cap f_j^{(2)}) = \sum_{i,j} m_1(f_i^{(1)})m_2(f_j^{(2)}) \tag{57}$$

which coincides exactly with the new rule of combination expressed previously.

4.4.3 Numerical example of entropy calculation

We present here a very simple numerical example of the derivation of entropies of individual sources of informations and the combined (joint) entropy of combined sources. Let's consider the simple frame of discernment $\Theta = \{\theta_1, \theta_2\}$ and the two following (uncertain and paradoxical) information granules

$$m_1(\theta_1) = 0.60 \qquad m_1(\theta_2) = 0.20 \qquad m_1(\theta_1 \cup \theta_2) = 0.10 \qquad m_1(\theta_1 \cap \theta_2) = 0.10$$

$$m_2(\theta_1) = 0.50 \qquad m_2(\theta_2) = 0.20 \qquad m_2(\theta_1 \cup \theta_2) = 0.10 \qquad m_2(\theta_1 \cap \theta_2) = 0.20$$

The fusion rule can be described through the following fusion table

$\oplus$	$m_1(\theta_1) = 0.60$	$m_1(\theta_2) = 0.20$	$m_1(\theta_1 \cup \theta_2) = 0.10$	$m_1(\theta_1 \cap \theta_2) = 0.10$
$m_2(\theta_1) = 0.50$	$0.30 \quad (\theta_1)$	$0.10 \quad (\theta_1 \cap \theta_2)$	$0.05 \quad (\theta_1)$	$0.05 \quad (\theta_1 \cap \theta_2)$
$m_2(\theta_2) = 0.20$	$0.12 \quad (\theta_1 \cap \theta_2)$	$0.04 \quad (\theta_2)$	$0.02 \quad (\theta_2)$	$0.02 \quad (\theta_1 \cap \theta_2)$
$m_2(\theta_1 \cup \theta_2) = 0.10$	$0.06 \quad (\theta_1)$	$0.02 \quad (\theta_2)$	$0.01 \quad (\theta_1 \cup \theta_2)$	$0.01 \quad (\theta_1 \cap \theta_2)$
$m_2(\theta_1 \cap \theta_2) = 0.20$	$0.12 \quad (\theta_1 \cap \theta_2)$	$0.04 \quad (\theta_1 \cap \theta_2)$	$0.02 \quad (\theta_1 \cap \theta_2)$	$0.02 \quad (\theta_1 \cap \theta_2)$

(58)

Each cell of the table provides a part of the global bpa $m(.)$ contribution for the corresponding proposition M indicated between parentheses. The entropies of individual sources are given by

$$H(M_1) = -0.60\ln(0.60) - 0.20\ln(0.20) - 0.10\ln(0.10) - 0.10\ln(0.10) = 1.0889 \; nats$$
$$H(M_2) = -0.50\ln(0.50) - 0.20\ln(0.20) - 0.10\ln(0.10) - 0.20\ln(0.20) = 1.2206 \; nats$$

The conditional entropies $H(M_1|M_2)$ and $H(M_2|M_1)$ are given by [9]

$$\begin{aligned}
H(M_1|M_2) =& m_2(M_2 = \theta_1)H(M_1|M_2 = \theta_1) + m_2(M_2 = \theta_2)H(M_1|M_2 = \theta_2) \\
&+ m_2(M_2 = \theta_1 \cap \theta_2)H(M_1|M_2 = \theta_1 \cap \theta_2) + m_2(M_2 = \theta_1 \cup \theta_2)H(M_1|M_2 = \theta_1 \cup \theta_2) \\
=& 0.5H\left[(\frac{0.30}{0.50}, \frac{0.10}{0.50}, \frac{0.05}{0.50}, \frac{0.05}{0.50})\right] + 0.2H\left[(\frac{0.12}{0.20}, \frac{0.04}{0.20}, \frac{0.02}{0.20}, \frac{0.02}{0.20})\right] \\
&+ 0.1H\left[(\frac{0.06}{0.10}, \frac{0.02}{0.10}, \frac{0.01}{0.10}, \frac{0.01}{0.10})\right] + 0.2H\left[(\frac{0.12}{0.20}, \frac{0.04}{0.20}, \frac{0.02}{0.20}, \frac{0.02}{0.20})\right] \\
=& (0.5 \times 1.0889) + (0.2 \times 1.0889) + (0.1 \times 1.0889) + (0.2 \times 1.0889) \\
=& 1.0889 \; nats
\end{aligned}$$

$$\begin{aligned}
H(M_2|M_1) =& m_1(M_1 = \theta_1)H(M_2|M_1 = \theta_1) + m_1(M_1 = \theta_2)H(M_2|M_1 = \theta_2) \\
&+ m_1(M_1 = \theta_1 \cap \theta_2)H(M_2|M_1 = \theta_1 \cap \theta_2) + m_1(M_1 = \theta_1 \cup \theta_2)H(M_2|M_1 = \theta_1 \cup \theta_2) \\
=& 0.6H\left[(\frac{0.30}{0.60}, \frac{0.12}{0.60}, \frac{0.06}{0.60}, \frac{0.12}{0.60})\right] + 0.2H\left[(\frac{0.10}{0.20}, \frac{0.04}{0.20}, \frac{0.02}{0.20}, \frac{0.04}{0.20})\right] \\
&+ 0.1H\left[(\frac{0.05}{0.10}, \frac{0.02}{0.10}, \frac{0.01}{0.10}, \frac{0.02}{0.10})\right] + 0.1H\left[(\frac{0.05}{0.10}, \frac{0.02}{0.10}, \frac{0.01}{0.10}, \frac{0.02}{0.10})\right] \\
=& (0.6 \times 1.2206) + (0.2 \times 1.2206) + (0.1 \times 1.2206) + (0.1 \times 1.2206) \\
=& 1.2206 \; nats
\end{aligned}$$

Therefore, one has

$$H(M_1) = H(M_1|M_2) = 1.0889 \; nats \qquad \text{and} \qquad H(M_2) = H(M_2|M_1) = 1.2206 \; nats$$

The joint entropy $H(M) = H(M_1, M_2)$ is directly obtained from the cells of the fusion table and one gets

$$H(M) = H[(0.3, 0.1, 0.05, 0.05, 0.12, 0.04, 0.02, 0.02, 0.06, 0.02, 0.01, 0.01, 0.12, 0.04, 0.02, 0.02)] = 2.3095\, nats$$

Hence, one has verified the classical result (chain rule) of the information theory, i.e.

$$H(M) = H(M_1) + H(M_2|M_1) = H(M_2) + H(M_1|M_2)$$

or more specially because of the independence of the two sources of information

$$H(M) = H(M_1) + H(M_2)$$

Note that $H(M)$ must be evaluated using the full description of the fusion table (from all the cells of the table) and not from the global bpa $m(.)$. Otherwise a smaller value for $H(M)$ is deduced, as it can be easily shown. From the fusion table, one gets the final bpa $m(.)$ with

$$m(\theta_1) = 0.41 \qquad m(\theta_2) = 0.08 \qquad m(\theta_1 \cup \theta_2) = 0.01 \qquad m(\theta_1 \cap \theta_2) = 0.50$$

The evaluation of $H(M)$ from bpa $m(.)$ yields the value

$$\tilde{H}(M) = H[(0.41, 0.08, 0.01, 0.50)] = 0.96023\, nats < H(M)$$

Remark

Note that in this example, the combination of the two sources reduces the uncertainty of judgment of each local information sources since $\tilde{H}(M) < H(M_1)$ and $\tilde{H}(M) < H(M_2)$. This is unfortunalely not a valid conclusion in general as many people (wrongly) think. We argue that the fusion of independent sources of information does not necessarly reduces the uncertainty of judgment. To convince the reader, just take the similar example with the following new information granules

$$m_1(\theta_1) = 0.900 \qquad m_1(\theta_2) = 0.090 \qquad m_1(\theta_1 \cup \theta_2) = 0.009 \qquad m_1(\theta_1 \cap \theta_2) = 0.001$$
$$m_2(\theta_1) = 0.090 \qquad m_2(\theta_2) = 0.900 \qquad m_2(\theta_1 \cup \theta_2) = 0.009 \qquad m_2(\theta_1 \cap \theta_2) = 0.001$$

It is not too difficult to check that global bpa $m(.)$ is

$$m(\theta_1) = 0.08991 \qquad m(\theta_2) = 0.08991 \qquad m(\theta_1 \cup \theta_2) = 0.000081 \qquad m(\theta_1 \cap \theta_2) = 0.820099$$

with corresponding entropies

$$H(M_1) = H(M_2) = 0.36084\, nats \qquad \text{and} \qquad (\tilde{H}(M) = 0.59659) < (H(M) = 0.72168)$$

but $\tilde{H}(M) > H(M_1)$ and $\tilde{H}(M) > H(M_2)$. Thus in this case, the fusion increases actually the uncertainty of the final judgment.

4.4.4 Definition for the generalized entropy of a source

The evaluation of the entropy $H(m)$ of a given source from the direct extension of its classical definition, with convention (see [9]) $0\ln(0) = 0$ and with bpa $m(.)$, i.e.

$$H(m) = -\sum_{A \in D^\Theta} m(A)\ln(m(A))$$

seems to not be the best measure for the self-information of a general (uncertain and paradoxical) source of information because it does not catch the intrinsic informational strength (i.i.s. for short) $s(A)$ of the propositions A. An extension of the classical entropy in the DST framework had already been proposed in 1983 by R. Yager based on the weight of conflict between the belief function Bel and the certain support function Bel_A focused on each proposition A (see [70] for details). In the classical definition (based only on probability measure), one always has $s(A) \equiv |A| = 1$. This does not hold in our general theory of plausible and paradoxical reasoning and we propose to generalize the notion of entropy in the following manner to measure correctly the self-information of a general source :

$$H_g(m) = -\sum_{A \in D^\Theta} \frac{1}{s(A)} m(A)\ln(\frac{1}{s(A)} m(A)) \tag{59}$$

$H_g(m)$ will be called the *generalized entropy* of the source associated with bpa $m(.)$. This general definition introduces the cardinality of a general (irreductible) proposition A which can be derived from the two following important rules

$$s\left(\bigcup_{i=1,n} B_i\right) = s\left(B_1 \cup \ldots \cup B_n\right) = \frac{\sum_{i=1,n} 1/s\left(B_i\right)}{\prod_{i=1,n} 1/s\left(B_i\right)} \tag{60}$$

$$s\left(\bigcap_{i=1,n} B_i\right) = s\left(B_1 \cap \ldots \cap B_n\right) = \frac{\prod_{i=1,n} s\left(B_i\right)}{\sum_{i=1,n} s\left(B_i\right)} \tag{61}$$

It is very important to note that these rules apply only on irreductible propositions (logical atoms) A. A proposition A is said to be irreductible (or equivalently has a compact form) if and only if it does not admit other equivalent form with a smaller number of operands and operators. For example $(\theta_1 \cup \theta_3) \cap (\theta_2 \cup \theta_3)$ is not an irreductible proposition since it can be reduced to its equivalent logical atom $(\theta_1 \cap \theta_2) \cup \theta_3$. To compute the i.i.s. $s(A)$ of any proposition A using the rules (60) and (61), the proposition has first to be reduced to its minimal representation (irreductible form).

Examples

Here are few examples of the value of the cardinality for some elementary and composite irreductible propositions A. We recall that θ_i involved in A are singletons such that $|\theta_i| = 1$.

$$\begin{aligned}
&A = \theta_1 \cup \theta_2 \Rightarrow s(A) = 2\\
&A = \theta_1 \cap \theta_2 \Rightarrow s(A) = 1/2\\
&A = \theta_1 \cup \theta_2 \cup \theta_3 = (\theta_1 \cup \theta_2) \cup \theta_3 = \theta_1 \cup (\theta_2 \cup \theta_3) = \theta_2 \cup (\theta_1 \cup \theta_3) \Rightarrow s(A) = 3\\
&A = \theta_1 \cap \theta_2 \cap \theta_3 = (\theta_1 \cap \theta_2) \cap \theta_3 = \theta_1 \cap (\theta_2 \cap \theta_3) = \theta_2 \cap (\theta_1 \cap \theta_3) \Rightarrow s(A) = 1/3\\
&A = (\theta_1 \cap \theta_2) \cup \theta_3 \Rightarrow s(A) = 3/2\\
&A = (\theta_1 \cup \theta_2) \cap \theta_3 \Rightarrow s(A) = 2/3\\
&A = (\theta_1 \cap \theta_2) \cup (\theta_3 \cap \theta_4) \Rightarrow s(A) = 1\\
&A = (\theta_1 \cup \theta_2) \cap (\theta_3 \cup \theta_4) \Rightarrow s(A) = 1\\
&A = (\theta_1 \cap \theta_2) \cup (\theta_3 \cap \theta_4 \cap \theta_5) \Rightarrow s(A) = 5/6\\
&A = (\theta_1 \cup \theta_2) \cap (\theta_3 \cup \theta_4 \cup \theta_5) \Rightarrow s(A) = 6/5
\end{aligned}$$

Thus the evaluation of $s(A)$ for any general irreductible proposition A can always be obtained from the two basic rules (60) and (61). This generalized definition makes sense with the notion of entropy and is coherent with classical definition (i.e. $H_g(m) \equiv H(m)$ when $m(.)$ becomes a bayesian bpa $p(.)$). Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ be a *general* frame of discernment of the problem under consideration and a general body of evidence with information granule $m(.)$ on D^Θ, then the generalized entropy $H_g(m)$ takes its minimal value $-n\ln(n)$ when the source provides the maximum of paradoxe which is obtained when $m(\theta_1 \cap \ldots \cap \theta_n) = 1$. It is important to note that the maximum of uncertainty is not obtained when $m(\theta_1 \cup \ldots \cup \theta_n) = 1$ but rather for a specific $m()$ which distributes some weight of evidence assignment to each proposition $A \in D^\Theta$ because there is less information (from the information theory viewpoint) when there exists several propositions with non nul bpa rather than one. One has also to take into account the intrinsic self-information of the propositions to get a good measure of global information provided by a source. The generalized entropy includes both aspects of the information (the intrinsic and the classical aspect). The uniform distribution for $m(.)$ does not generate the maximum generalized-entropy because of the different intrinsic self-information of each proposition (see next example). We argue that the generalized entropy of any source defined with respect to a frame Θ appears to be a very useful tool to measure the degree of uncertainty and paradoxe of any given source of information.

Example

We give here some values of $H_g(m)$ for different sources of information over the same frame $\Theta = \{\theta_1, \theta_2\}$. The sources have been classified from the most informative one $\mathcal{B}_1$ up to the less informative one $\mathcal{B}_{16}$. $\mathcal{B}_{16}$ corresponds to the source containing minimal information on the hyper-power set of the frame Θ (thus $\mathcal{B}_{16}$ has the minimal discrimination power between all possible propositions). There does not exist a source $\mathcal{B}_k$ such that $H_g^{\mathcal{B}_k}(m) > H_g^{\mathcal{B}_{16}}(m)$ for this simpliest example. Finding $m^*(.)$ such that $H_g(m^*)$ takes its maximal value for a general frame Θ with $|\Theta| = n$ is called the general whitening source problem. No solution for this problem has been obtained so far.

	$m(\theta_1)$	$m(\theta_2)$	$m(\theta_1 \cup \theta_2)$	$m(\theta_1 \cap \theta_2)$	$H_g(m)$
$\mathcal{B}_1$	0	0	0	1	-1.386
$\mathcal{B}_2$	0	0	0.3	0.7	-0.186
$\mathcal{B}_3$	1	0	0	0	0
$\mathcal{B}_4$	0	1	0	0	0
$\mathcal{B}_5$	0.1	0.2	0	0.7	0.081
$\mathcal{B}_6$	0	0	1	0	0.346
$\mathcal{B}_7$	0.8	0.2	0	0	0.500
$\mathcal{B}_8$	0	0	0.7	0.3	0.673
$\mathcal{B}_9$	0.5	0.5	0	0	0.693
$\mathcal{B}_{10}$	0.7	0.2	0.1	0	0.721
$\mathcal{B}_{11}$	0.7	0.2	0	0.1	0.893
$\mathcal{B}_{12}$	0.1	0.2	0.7	1	0.919
$\mathcal{B}_{13}$	0.1	0.2	0.3	0.4	1.015
$\mathcal{B}_{14}$	0.1	0.2	0.4	0.3	1.180
$\mathcal{B}_{15}$	0.25	0.25	0.25	0.25	1.299
$\mathcal{B}_{16}$	0.25	0.25	0.35	0.15	1.359

$\mathcal{B}_1$ is the most informative source because all the weights of evidence about the truth are focused only on the smaller element $\theta_1 \cap \theta_2$ of hyper-powerset D^Θ. $\mathcal{B}_2$ is less informative than $\mathcal{B}_1$ because there exists an ambiguity between the two propositions $\theta_1 \cup \theta_2$ and $\theta_1 \cap \theta_2$. $\mathcal{B}_3$ and $\mathcal{B}_4$ are less informative than $\mathcal{B}_1$ because the weights of evidence about the truth are focused on larger elements (θ_1 or θ_2 respectively) of D^Θ. $\mathcal{B}_6$ is less informative than $\mathcal{B}_3$ or $\mathcal{B}_4$ because the weight of evidence about the truth is focused on a bigger element $\theta_1 \cup \theta_2$ of D^Θ. $\mathcal{B}_7$ is less informative than previous sources since there exists an ambiguity between the two propositions θ_1 and θ_2 but it is more informative than $\mathcal{B}_9$ since the discrimination power (our easiness to decide which proposition supports the truth) is higher with $\mathcal{B}_7$ than with $\mathcal{B}_9$. Note that even if in this very simple example, it is not obvious to see that $\mathcal{B}_{16}$ is the less informative (white) source of information. Most of readers would have probably thought to choose either $\mathcal{B}_6$ or $\mathcal{B}_{15}$. This comes from the confusion between the intrinsic information supported by the proposition itself and the information supported by the whole bpa $m(.)$.

4.4.5 Zadeh's example

Let's take back the disturbing Zadeh's example given in section 3.4. Two doctors examine a patient and agree that it suffers from either meningitis (M), concussion (C) or brain tumor (T). Thus $\Theta = \{M, C, T\}$. Assume that the doctors agree in their low expectation of a tumor, but disagree in likely cause and provide the following diagnosis

$$m_1(M) = 0.99 \qquad m_1(T) = 0.01 \qquad \text{and} \quad \forall A \in D^\Theta, A \neq T, A \neq M, \quad m_1(A) = 0$$

$$m_2(C) = 0.99 \qquad m_2(T) = 0.01 \qquad \text{and} \quad \forall A \in D^\Theta, A \neq T, A \neq C, \quad m_2(A) = 0$$

The new general rule of combination (41), yields the following combined information granule

$$m(M \cap C) = 0.9801 \qquad m(M \cap T) = 0.0099 \qquad m(C \cap T) = 0.0099 \qquad m(T) = 0.0001$$

From this granule, one gets

$$\begin{aligned} \text{Bel}(M) &= m(M \cap C) + m(M \cap T) = 0.99 \\ \text{Bel}(C) &= m(M \cap C) + m(T \cap C) = 0.99 \\ \text{Bel}(T) &= m(T) + m(M \cap T) + m(C \cap T) = 0.0199 \end{aligned}$$

If both doctors can be considered as equally reliable, the combined information granule $m(.)$ mainly focuses weight of evidence on the paradoxical proposition $M \cap C$ which means that patient suffers both meningitis and concussion but almost surely not from brain tumor. This conclusion is coherent with the common sense actually. Then, no therapy for brain tumor (like heavy and ever risky brain surgical intervention) will be chosen in such case. This really helps to take important decision to save the life of the patient in this example. A deeper medical examination adapted to both meningitis and concussion will almost surely be done before applying the best therapy for the patient. Just remember that in this case, the DST had concluded that the patient had brain tumor with certainty . . . .

4.4.6 Mahler's example revisited

Let's consider now the following example excerpt from the R. Mahler's paper [36]. We consider that our classification knowledge base consists of the three (imaginary) new and rare diseases corresponding to following frame of discernment

$$\Theta = \{\theta_1 = \textit{kotosis}, \theta_2 = \textit{phlegaria}, \theta_3 = \textit{pinpox}\}$$

We assume that the three diseases are equally likely to occur in the patient population but there is some evidence that *phlegaria* and *pinpox* are the same disease and there is also a small possibility that *kotosis* and *phlegaria* might be the same disease. Finally, there is a small possibility that all three diseases are the same. This information can be expressed by assigning a priori bpa as follows

$$\begin{array}{lll} m_0(\theta_1) = 0.2 & m_0(\theta_2) = 0.2 & m_0(\theta_3) = 0.2 \\ m_0(\theta_2 \cap \theta_3) = 0.2 & m_0(\theta_1 \cap \theta_2) = 0.1 & m_0(\theta_1 \cap \theta_2 \cap \theta_3) = 0.1 \end{array}$$

Let $\text{Bel}(.)$ the prior belief measure corresponding to this prior bpa $m(.)$. Now assume that Doctor D_1 and Doctor D_2 examine a patient and deliver diagnoses with following reports:

- Report for D_1: $m_1(\theta_1 \cup \theta_2 \cup \theta_3) = 0.05$ $\quad m_1(\theta_2 \cup \theta_3) = 0.95$
- Report for D_2: $m_2(\theta_1 \cup \theta_2 \cup \theta_3) = 0.20$ $\quad m_2(\theta_2) = 0.80$

The combination of the evidences provided by the two doctors $m' = m_1 \oplus m_2$ obtained by the general rule of combination (41) yields the following bpa $m'(.)$

$$m'(\theta_2) = 0.8 \qquad m'(\theta_2 \cup \theta_3) = 0.19 \qquad m'(\theta_1 \cup \theta_2 \cup \theta_3) = 0.01$$

The combination of bpa $m'(.)$ with prior evidence $m_0(.)$ yields the final bpa $m = m_0 \oplus m' = m_0 \oplus [m_1 \oplus m_2]$ with

$$\begin{array}{lll} m(\theta_1) = 0.002 & m(\theta_2) = 0.200 & m(\theta_3) = 0.040 \\ m(\theta_1 \cap \theta_2) = 0.260 & m(\theta_2 \cap \theta_3) = 0.360 & m(\theta_1 \cap \theta_2 \cap \theta_3) = 0.100 \\ m(\theta_1 \cap (\theta_2 \cup \theta_3)) = 0.038 & & \end{array}$$

Therefore the final belief function given by (37) is

$$\begin{array}{l} \text{Bel}(\theta_1) = 0.002 + 0.260 + 0.100 + 0.038 = 0.400 \\ \text{Bel}(\theta_2) = 0.200 + 0.260 + 0.360 + 0.100 = 0.920 \\ \text{Bel}(\theta_3) = 0.040 + 0.360 + 0.100 = 0.500 \\ \text{Bel}(\theta_1 \cap \theta_2) = 0.260 + 0.100 = 0.360 \\ \text{Bel}(\theta_2 \cap \theta_3) = 0.360 + 0.100 = 0.460 \\ \text{Bel}(\theta_1 \cap (\theta_2 \cup \theta_3)) = 0.038 + 0.100 = 0.138 \\ \text{Bel}(\theta_1 \cap \theta_2 \cap \theta_3) = 0.100 \end{array}$$

Thus, on the basis of all the evidences one has, we are able to conclude with high a degree of belief that the patient has phlegaria which is coherent with the Mahler's conclusion based on his Conditioned Dempster-Shafer theory developed from his conditional event algebra although a totally new and simpliest approach has been adopted here.

4.4.7 A thief identification example

Let's revisit a very simple thief identification example. Assume that a 75 years old grandfather is taking a walk with his 9 years old grandson in a park. They saw at 50 meters away, a 45 years old pickpocket robbering the bag of an old lady. A policeman looking for some witnesses of this event asks separately the grandfather and his grandchild if they have seen the thief (they both answer yes) and how was the thief (a young or an old man). The grandfather (source of information B_1 reports that the thief was a young man with high confidence 0.99 and with only a low uncertainty 0.01. His grandson reports that the thief was a old man with high confidence 0.99 and with only a low uncertainty 0.01. These two witnesses provide fair reports (with respect to their own world of knowledge) even if apparently they appear as paradoxical. The policeman then send the two reports with only minimal information about witnesses (saying only their names and that they were a priori fully trustable) to an investigator. The investigator has no possibility to meet or to call back the witnesses in order to get more details.

Under such condition, what would be the best decision to be taken by the investigator about the age of the thief to eventually help to catch him? Such kind of simple examples occur quite frequently in witnesses problems actually. A rational investigator will almost surely suspect a mistake or an error in one or both reports since they appear apparently in full contradiction. The investigator will then try to take his final decision with some other better information (if any).

If the investigator uses our new plausible and paradoxical reasoning, he will defined the following bpa with respect to the frame of discernment $\Theta = \{\theta_1 = \text{young}, \theta_2 = \text{old}\}$ and the available reports $\mathcal{B}_1$ and $\mathcal{B}_2$ with following bpa

$$m_1(\theta_1) = 0.99 \qquad m_1(\theta_2) = 0 \qquad m_1(\theta_1 \cup \theta_2) = 0.01 \qquad m_1(\theta_1 \cap \theta_2) = 0$$

$$m_2(\theta_1) = 0 \qquad m_2(\theta_2) = 0.99 \qquad m_2(\theta_1 \cup \theta_2) = 0.01 \qquad m_2(\theta_1 \cap \theta_2) = 0$$

The fusion of these two sources of information yields the global bpa $m(.)$ with

$$m(\theta_1) = 0.0099 \qquad m(\theta_2) = 0.0099 \qquad m(\theta_1 \cup \theta_2) = 0.0001 \qquad m(\theta_1 \cap \theta_2) = 0.9801$$

Thus, from this global information, the investigator has no better choice but to consider with almost certainty that the thief was both a young and old man. By assuming that the expected life duration is around 80 years, the inspector will deduce that the true age of the thief is around 40 years old which is not too far from the truth. At least, this conclusion could be helpful to interrogate some suspicious individuals.

4.4.8 A model to generate information granules $m(.)$ from intervals

We present here a model to generate information granules $m(.)$ from information represented by intervals. It is very common in practice that uncertain sources of information provide evidence on a given proposition in term of basic intervals $[\epsilon_*, \epsilon^*] \subset [0, 1]$ rather than a direct bpa $m(.)$. In such cases, some preprocessing must be done before applying the general rule of combination between such sources to take the final decision.

In the DST framework, we recall that the simpliest and easiest transformation to convert $[\epsilon_*, \epsilon^*]$ into bpa has already been proposed by A. Appriou in [3]. The basic idea was to interpret ϵ_* as the minimal credibility committed to A and ϵ^* as the plausibility committed to A. In other words, the Appriou's transformation model within the DST is the following one

$$\begin{aligned} \epsilon_* &= m(A) \\ \epsilon^* &= 1 - m(A^c) \\ \epsilon^* - \epsilon_* &= m(A \cup A^c) \end{aligned}$$

This model can be directly extended within our new theory of plausible and paradoxical reasoning by setting now.

$$\begin{aligned} \epsilon_* &= m(A) + \frac{1}{2} m(A \cap A^c) \\ \epsilon^* &= 1 - m(A^c) - \frac{1}{2} m(A \cap A^c) \\ \epsilon^* - \epsilon_* &= m(A \cup A^c) \end{aligned}$$

or equivalently

$$m(A) + \frac{1}{2} m(A \cap A^c) = \epsilon_* \tag{62}$$

$$m(A^c) + \frac{1}{2} m(A \cap A^c) = 1 - \epsilon^* \tag{63}$$

$$m(A \cup A^c) = \epsilon^* - \epsilon_* \tag{64}$$

This appealing model presents nice properties specially when $\epsilon^* = \epsilon_* = 0$ or when $\epsilon^* = \epsilon_* = 1$. This model is moreover coherent with the previous Appriou's model whenever the source becomes rational (i.e $m(A \cap A^c) = 0$). This new model presents however a degree of freedom since one has only two constraints (62) and (63) for three unknowns $m(A)$, $m(A^c)$ and $m(A \cap A^c)$. Thus in general, without an additional constraint, there exists many possible choices for $m(A)$, $m(A^c)$ and $m(A \cap A^c)$ and therefore there exists several bpa $m(.)$ satisfying this transformation model. Without extra prior information, it becomes difficult to justify the choice of a specific bpa versus all other admissible possibilities for $m(.)$.

To solve this important drawback, we propose to add the constraint on the maximization of the generalized-entropy $H_g(m)$. This will allow us to obtain from $[\epsilon_*, \epsilon^*]$ the unique bpa $m(.)$ having the minimum of specificity and admissible with our transformation model. From definition of $H_g(m)$ and previous equations (62)-(64), one gets

$$\begin{aligned} H_g(m) = &- (\epsilon_* - m(A \cap A^c)/2) \ln(\epsilon_* - m(A \cap A^c)/2) - (1 - \epsilon^* - m(A \cap A^c)/2) \ln(1 - \epsilon^* - m(A \cap A^c)/2) \\ &- \frac{1}{2}(\epsilon^* - \epsilon_*) \ln(\frac{1}{2}(\epsilon^* - \epsilon_*)) - 2m(A \cap A^c) \ln(2m(A \cap A^c)) \end{aligned}$$

The maximization of $H_g(m)$ is obtained for the optimal value $m^\star(A \cap A^c)$ such that $\frac{\partial H_g}{\partial m(A\cap A^c)}(m^\star(A \cap A^c)) = 0$ and $\frac{\partial^2 H_g}{\partial m(A\cap A^c)^2}(m^\star(A \cap A^c)) < 0$. The annulation of the first derivative is obtained by the solution of the equation

$$\frac{1}{2}\ln(\epsilon_* - m^\star/2) + \frac{1}{2}\ln(1-\epsilon^* - m^\star/2) - 2m^\star \ln(2m^\star) - 1 = 0$$

or equivalently after basic algebraic manipulations

$$64e^2(m^\star)^4 - (m^\star)^2 + 2(1-\epsilon^* + \epsilon_*)m^\star - 4(1-\epsilon^*)\epsilon_* = 0 \tag{65}$$

The solution of this equation does not admit a simple analytic expression but can be easily found using classical numerical methods. It is also easy to check that the second derivative is always negative and therefore $H_g(m)$ reaches its maximal value when

$$m(A) + \frac{1}{2}m^\star(A\cap A^c) = \epsilon_* \tag{66}$$

$$m(A^c) + \frac{1}{2}m^\star(A\cap A^c) = 1-\epsilon^* \tag{67}$$

$$m(A\cup A^c) = \epsilon^* - \epsilon_* \tag{68}$$

This completes the definition of our new transformation model. Note that $[\epsilon_*, \epsilon^*]$ can also be generated from bpa $m(.)$ through (62)-(64).

Numerical examples

- $[\epsilon_*, \epsilon^*] = [0.0, 0.0]$ $m(A\cap A^c) = 0.000$ $m(A) = 0.000$ $m(A^c) = 1.000$ $m(A\cup A^c) = 0.000$
- $[\epsilon_*, \epsilon^*] = [0.2, 0.2]$ $m(A\cap A^c) \approx 0.164$ $m(A) \approx 0.118$ $m(A^c) \approx 0.718$ $m(A\cup A^c) = 0.000$
- $[\epsilon_*, \epsilon^*] = [0.5, 0.5]$ $m(A\cap A^c) \approx 0.192$ $m(A) \approx 0.404$ $m(A^c) \approx 0.404$ $m(A\cup A^c) = 0.000$
- $[\epsilon_*, \epsilon^*] = [0.8, 0.8]$ $m(A\cap A^c) \approx 0.164$ $m(A) \approx 0.718$ $m(A^c) \approx 0.118$ $m(A\cup A^c) = 0.000$
- $[\epsilon_*, \epsilon^*] = [1.0, 1.0]$ $m(A\cap A^c) = 0.000$ $m(A) = 1.000$ $m(A^c) = 0.000$ $m(A\cup A^c) = 0.000$
- $[\epsilon_*, \epsilon^*] = [0.2, 0.4]$ $m(A\cap A^c) \approx 0.152$ $m(A) \approx 0.124$ $m(A^c) \approx 0.524$ $m(A\cup A^c) = 0.200$
- $[\epsilon_*, \epsilon^*] = [0.6, 0.8]$ $m(A\cap A^c) \approx 0.152$ $m(A) \approx 0.524$ $m(A^c) \approx 0.124$ $m(A\cup A^c) = 0.200$
- $[\epsilon_*, \epsilon^*] = [0.4, 0.6]$ $m(A\cap A^c) \approx 0.170$ $m(A) \approx 0.315$ $m(A^c) \approx 0.315$ $m(A\cup A^c) = 0.200$
- $[\epsilon_*, \epsilon^*] = [0.3, 0.9]$ $m(A\cap A^c) \approx 0.100$ $m(A) \approx 0.250$ $m(A^c) \approx 0.050$ $m(A\cup A^c) = 0.600$
- $[\epsilon_*, \epsilon^*] = [0.0, 1.0]$ $m(A\cap A^c) = 0.000$ $m(A) = 0.000$ $m(A^c) = 0.000$ $m(A\cup A^c) = 1.000$

5 Plausible and paradoxical reasoning in the neutrosophy framework

5.1 Neutrosophy and the neutrosophic logic

The neutrosophy is a new branch of philosophy, introduced by Florentin Smarandache in 1980, which studies the origin, nature and scope of neutralities, as well as their interactions with different ideational spectra. Neutrosophy considers a proposition, theory, event, concept, or entitity, A in relation to its opposite, *anti*-A and that which is not A, *non*-A, and that which is neither A nor *anti*-A, denoted by *neut*-A. Neutrosophy serves as the basis for the neutrosophic logic [22].

The Neutrosophic Logic (NL) or Smarandache's logic is a general framework for the unification of all existing logics [56, 57, 58]). The main idea of NL is to characterize each logical statement in a 3D neutrosophic space where each dimension of the space represents respectively the truth (T), the falsehood (F) and the indeterminacy (I) of the statement under consideration where T, I and F are standard or non-sandard real subsets of $]^-0;1^+[$. Moreover in NL, each statement is allowed to be over or under true, over or under false and over or under inderterminate by using hyper real numbers developed in the non-standard analysis theory [43, 14]. The neutrosophical value $\mathfrak{N}(\mathsf{A}) = (T(\mathsf{A}), I(\mathsf{A}), F(\mathsf{A}))$ in a frame of discernment (world of discourse) Θ of a statement A is then defined as a subset (a volume not necessary connexe; i.e. a set of disjoint volumes) of the neutrosophic space. Any statement A represented by a triplet $\mathfrak{N}(\mathsf{A})$ is called a *neutrosophic event* or $\mathfrak{N}$ – event. The subset $\mathfrak{N}_t \triangleq T(\mathsf{A})$ characterizes the truth part of statement A. $\mathfrak{N}_i \triangleq (\mathsf{A})$ and $\mathfrak{N}_f \triangleq F(\mathsf{A})$ represent the inderterminacy and the falsehood of A. This Smarandache's representation is close to the human reasoning. It characterizes and catches the imprecision of knowledge or linguistic inexactitude received by various observers, uncertainty due to incomplete knowledge of acquisition errors or stochasticity, and vagueness due to lack of clear contours or limits. This approach allows theoretically to consider any kinds of logical statements. For example, the fuzzy set logic or the classical modal logic (which works with statements verifying $T(\mathsf{A})$, $I(\mathsf{A}) \equiv 0$, $F(\mathsf{A}) = 1 - T(\mathsf{A})$, where T is a real number belonging to $[0;1]$) are included in NL. The neutrosophic logic can easily handle also paradoxes. We emphaze the fact that in general the neutrosohic value $\mathfrak{N}(\mathsf{A})$ of a proposition A can also depend on dynamical parameters which can evolve with time, space, etc. For seak of concise notation, we omit to introduce this dependence in our notations in the sequel.

Basic operations on sets

Beside this modelling, F. Smarandache has introduced the following operations on sets. Consider S_1 and S_2 be two (unidimensional) standard or non-standard real subsets. The addition, substraction, multiplication and division (by a non null finite number) of these sets are defined as follows :

- **Addition**

$$S_1 \oplus S_2 = S_2 \oplus S_1 \triangleq \{x \mid x = s_1 + s_2, \forall s_1 \in S_1, \forall s_2 \in S_2\} \tag{69}$$

- **Substraction**

$$S_1 \ominus S_2 = -(S_2 \ominus S_1) \triangleq \{x \mid x = s_1 - s_2, \forall s_1 \in S_1, \forall s_2 \in S_2\} \tag{70}$$

  For *real positive* subsets, the Inf and Sup values of $S_1 \ominus S_2$ are given by

$$\text{Inf}[S_1 \ominus S_2] = \text{Inf}[S_1] - \text{Sup}[S_2] \qquad \text{and} \qquad \text{Sup}[S_1 \ominus S_2] = \text{Sup}[S_1] - \text{Inf}[S_2]$$

- **Multiplication**

$$S_1 \odot S_2 = S_2 \odot S_1 \triangleq \{x \mid x = s_1 \cdot s_2, \forall s_1 \in S_1, \forall s_2 \in S_2\} \tag{71}$$

  For *real positive* subsets, one gets

$$\text{Inf}[S_1 \odot S_2] = \text{Inf}[S_1] \cdot \text{Inf}[S_2] \qquad \text{and} \qquad \text{Sup}[S_1 \odot S_2] = \text{Sup}[S_1] \cdot \text{Sup}[S_2]$$

- **Division of a set by a non null standard number**
  Let $k \in \mathbb{R}^*$, then

$$S_1 \oslash k \triangleq \{x \mid x = s_1/k, \forall s_1 \in S_1\} \tag{72}$$

Neutrosophic topology

Let's construct now a neutrosophic topology (NT) [56] on interval $]^-0;1^+[$, by considering the associated family of standard or non-standard subsets included in $]^-0;1^+[$ and the empty set $\emptyset$, which is closed under set union and finite intersection. The union and intersection of two any propositions A and B (corresponding to either the part of truth, indeterminacy or falsehood of a given assertion defined on $]^-0;1^+[$) are defined as follows

$$A \cup B = (A \oplus B) \ominus (A \odot B) \qquad \text{and} \qquad A \cap B = A \odot B \tag{73}$$

The neutrosophic complement of A is defined as $\bar{A} = \{1^+\} \ominus A$ and the $\mathfrak{N}-$ value of an assertion A is characterized by a mapping function $\mathfrak{N}(.)$ such that

$$\mathfrak{N} : A \quad \mapsto \quad \mathfrak{N}(\mathsf{A}) = (T(\mathsf{A}), I(\mathsf{A}), F(\mathsf{A})) \subset]^-0;1^+[^3 \tag{74}$$

The interval $]^-0;1^+[$, endowed with this topology, forms a *neutrosophic topological space*.

Consider now two statements A_1 and A_2, then one defines the following basic neutrosophic operators:

$$\mathfrak{N}(\mathsf{A}_1) \boxplus \mathfrak{N}(\mathsf{A}_2) = (T_1 \oplus T_2, I_1 \oplus I_2, F_1 \oplus F_2) \tag{75}$$
$$\mathfrak{N}(\mathsf{A}_1) \boxminus \mathfrak{N}(\mathsf{A}_2) = (T_1 \ominus T_2, I_1 \ominus I_2, F_1 \ominus F_2) \tag{76}$$
$$\mathfrak{N}(\mathsf{A}_1) \boxdot \mathfrak{N}(\mathsf{A}_2) = (T_1 \odot T_2, I_1 \odot I_2, F_1 \odot F_2) \tag{77}$$

where $T_i = T(\mathsf{A}_i)$, $I_i = I(\mathsf{A}_i)$, $F_i = F(\mathsf{A}_i)$ for $i = 1, 2$.

Since the truth, falsehood and indeterminacy of any statement must belong to $]^-0;1^+[$, the result of each previous operator $\boxplus$, $\boxminus$ and $\boxdot$ must be in $]^-0;1^+[^3$. Therefore upper and lower bounds of $T_1 \oplus T_2$ must be set respectively to $^-0$ and 1^+ whenever $inf(T_1 \oplus T_2) < 0$ or $sup(T_1 \oplus T_2) > 1$. The same remark applies for $\boxminus$ and $\boxdot$ operators and for falsehood and inderterminacy part of compounded statement.

All classical logical operators and connectors can be extented in the $\mathfrak{N}$ – Logic. For notation convenience, we will identify logical operators with their classical counterpart in set theory as pointed out in [35] (hence the following equivalences will be used $\neg \mathsf{A} \equiv \bar{\mathsf{A}}$, $\mathsf{A}_1 \wedge \mathsf{A}_2 \equiv \mathsf{A}_1 \cap \mathsf{A}_2$ and $\mathsf{A}_1 \vee \mathsf{A}_2 \equiv \mathsf{A}_1 \cup \mathsf{A}_2$ throughout this paper). We recall here only important operators used in the sequel. Additional neutrosophic logical operators like (strong disjunction, implication, equivalence, Sheffer's and Pierce's connectors) and general, physics and philosophical examples of application of neutrosophic operators can be found in [56, 58].

- **Negation**

$$\mathfrak{N}(\bar{\mathsf{A}}) = (\{1\} \ominus T(\mathsf{A}), \{1\} \ominus I(\mathsf{A}), \{1\} \ominus F(\mathsf{A})) \tag{78}$$

- **Conjunction**

$$\mathfrak{N}(\mathsf{A}_1 \cap \mathsf{A}_2) = \mathfrak{N}(\mathsf{A}_1) \boxdot \mathfrak{N}(\mathsf{A}_2) = (T_1 \odot T_2, I_1 \odot I_2, F_1 \odot F_2) \tag{79}$$

- **Disjunction**

$$\begin{aligned}\mathfrak{N}(\mathsf{A}_1 \cup \mathsf{A}_2) &= (T_1 \cup T_2, I_1 \cup I_2, F_1 \cup F_2) \\ &= ((T_1 \oplus T_2) \ominus (T_1 \odot T_2), (I_1 \oplus I_2) \ominus (I_1 \odot I_2), (I_1 \oplus I_2) \ominus (I_1 \odot I_2)) \\ &= [\mathfrak{N}(\mathsf{A}_1) \boxplus \mathfrak{N}(\mathsf{A}_2)] \;\boxminus\; [\mathfrak{N}(\mathsf{A}_1) \boxdot \mathfrak{N}(\mathsf{A}_2)]\end{aligned} \tag{80}$$

$\mathfrak{N}$ – Membership function over a neutrosophic set

Let Θ be a world of discourse (called frame of discernment in the DST). Each $\mathfrak{N}-$element x of Θ is characterized by its own neutrosophical basic assignment ($\mathfrak{N}$ – value) $\mathfrak{N}(x) \triangleq (T(x), I(x), F(x))$ with $T(x)$,$I(x)$ and $F(x) \subset]^{-}0; 1^{+}[$. The $\mathfrak{N}$ – membership function of any neutrosophical element x with any subset $M \subset \Theta$ is defined in similar way by

$$\mathfrak{N}(x \mid M) \triangleq (T_M(x), I_M(x), F_M(x)) \tag{81}$$

with $T_M(x)$,$I_M(x)$ and $F_M(x) \subset]^{-}0; 1^{+}[$. The $\mathfrak{N}$ – value of x over M can be interpreted, by abuse of language, as its membership function to M in the following sense: x is t% true in the set M, i% indeterminate (unknown if it is) in M, and f% false in M, where t varies in T, i varies in I, f varies in F. The standard notation $x \in M$ will be used in the sequel to denote the neutrosophical membership of x to M. One can say actually that any element x of a given frame of discernment supported by a body of evidence neutrosophically belongs to any set, due to the percentages of truth/indeterminacy/falsity involved, which varies between 0 and 1 or even less than 0 or greater than 1. From this definition and previous neutrosophic rules, one gets directly the following basic neutrosophical set operations :

- **Complement of** M

  If $x \in M$ with $\mathfrak{N}(x \mid M) \triangleq (T_M(x), I_M(x), F_M(x))$, then $x \notin M$ with

$$\mathfrak{N}(x \mid \bar{M}) = (\{1\} \ominus T_M(x), \{1\} \ominus I_M(x), \{1\} \ominus F_M(x)) \tag{82}$$

- **Intersection** $M \cap N$

  If $x \in M$ with $\mathfrak{N}(x \mid M) \triangleq (T_M(x), I_M(x), F_M(x))$ and $x \in N$ with $\mathfrak{N}_{|W}(x \mid N) \triangleq (T_N(x), I_N(x), F_N(x))$, then $x \in M \cap N$ with

$$\mathfrak{N}(x \mid M \cap N) = (T_M(x) \odot T_N(x), I_M(x) \odot I_N(x), F_M(x) \odot F_N(x)) \tag{83}$$

- **Union** $M \cup N$

  If $x \in M$ with $\mathfrak{N}(x \mid M) \triangleq (T_M(x), I_M(x), F_M(x))$ and $x \in N$ with $\mathfrak{N}(x \mid N) \triangleq (T_N(x), I_N(x), F_N(x))$, then $x \in M \cup N$ with

$$\mathfrak{N}(x \mid M \cup N) = (T_{M \cup N}(x), I_{M \cup N}(x), F_{M \cup N}(x)) \tag{84}$$

  where

$$T_{M \cup N}(x) \triangleq [T_M(x) \oplus T_N(x)] \;\ominus\; [T_M(x) \odot T_N(x)] \tag{85}$$

$$I_{M \cup N}(x) \triangleq [I_M(x) \oplus I_N(x)] \;\ominus\; [I_M(x) \odot I_N(x)] \tag{86}$$

$$F_{M \cup N}(x) \triangleq [F_M(x) \oplus F_N(x)] \;\ominus\; [F_M(x) \odot F_N(x)] \tag{87}$$

- **Difference** $M - N$

  Since $M - N \triangleq M - \bar{N}$, if $x \in M$ with $\mathfrak{N}(x \mid M) \triangleq (T_M(x), I_M(x), F_M(x))$ and $x \in N$ with $\mathfrak{N}(x \mid N) \triangleq (T_N(x), I_N(x), F_N(x))$, then $x \in M - N$ with

$$\mathfrak{N}(x \mid M - N) = (T_{M-N}(x), I_{M-N}(x), F_{M-N}(x)) \tag{88}$$

  where

$$T_{M-N}(x) \triangleq T_M(x) \ \ominus \ [T_M(x) \odot T_N(x)] \tag{89}$$
$$I_{M-N}(x) \triangleq I_M(x) \ \ominus \ [I_M(x) \odot I_N(x)] \tag{90}$$
$$F_{M-N}(x) \triangleq F_M(x) \ \ominus \ [F_M(x) \odot F_N(x)] \tag{91}$$

- **Inclusion** $M \subset N$

  We will said that $M \subset N$ if for all $x \in M$ with $\mathfrak{N}(x \mid M) \triangleq (T_M(x), I_M(x), F_M(x))$ and $x \in N$ with $\mathfrak{N}(x \mid N) \triangleq (T_N(x), I_N(x), F_N(x))$, one has jointly $T_M(x) \subset T_N(x)$, $I_M(x) \subset I_N(x)$ and $F_M(x) \subset F_N(x)$.

5.2 Combination of neutrosophic evidences

Let's consider a general finite frame of discernment $\Theta = \{\theta_1, \dots, \theta_n\}$ and two bodies of (neutrosophic) evidence $\mathcal{B}_1$ and $\mathcal{B}_2$. In the neutrosophic framework, we assume that each body of evidence provides some report of evidence (i.e. $\mathfrak{N}-$ value) committed to some elements of the hyper-power set D^Θ. In other words, the information one has to deal with is the reports:

- Report for $\mathcal{B}_1$: $R_1 = \{\mathfrak{N}_1(A_1), \dots, \mathfrak{N}_1(A_m)\}$ for $A_1, \dots, A_m \in D^\Theta$
- Report for $\mathcal{B}_2$: $R_2 = \{\mathfrak{N}_2(B_1), \dots, \mathfrak{N}_2(B_n)\}$ for $B_1, \dots, B_n \in D^\Theta$

where each neutrosophic value for a proposition corresponds actually to a given triplet $(T(.), I(.), F(.)) \subset]^-0; 1^+[^3$ Within the neutrosophic logic, one has the full degree of freedom between the $\mathfrak{N}-$ values for a report.

Our major concern now is to solve the difficult question on how to combine such kind of information to get the global and most pertinent information about the problem under consideration. So, is it possible to construct a new global report (and hopefully more informative) R from R_1 and R_2 ? Unfortunatly, the neutrosophic logic which is a new appealing and modelling tool to deal with uncertainties on propositions of same universe of discourse does not provide a clear and direct mathematical mechanism for dealing with combination of such kind of evidences. We propose in this section a possible issue for this important question based on our new generalization of the DST.

The main idea for combining such kind of evidences is to convert the reports into two proper general bpa $m_{R_1}(.)$ and $m_{R_2}(.)$ and then combine them using the general rule of combination (41). The combination of neutrosophic evidences is a two-level process.

Level 1 : the general bpa transformation

The major difficulty is the mapping of the set of neutrosophic values $\{\mathfrak{N}(.)\}$ into a set of corresponding elementary bpa $m(.)$. Several cases are now examined.

- Case 1 (simpliest case) : We assume that each neutrosophic evidence corresponds only to a triplet of real positive or null numbers belonging to $[0; 1]$ (i.e. $T(.)$, $I(.)$ and $F(.)$ are restricted to real numbers $\in [0; 1]$.
  Since in the neutrosophic logic, $T(.)$, $I(.)$ and $F(.)$ have no strong mathematical relationships, the easiest solution within the classical DST would be to use the following transformation

$$m_(A) = T(A)/c \qquad m(A^c) = F(A)/c \qquad m(A \cup A^c) = I(A)/c$$

  where c is a normalization constant such that $m_(A) + m(A^c) + m(A \cap A^c) = 1$.

  In our general theory of plausible and paradoxical reasoning, it seems more judicious to use the following mapping based on our general modelling of information granule described in section 4.4.8. Thus, we are now able to construct from $T(.)$, $I(.)$ and $F(.)$ the three corresponding elementary bpa as follows for any proposition $C \in D^\Theta$ involved in a given report :

$$\begin{array}{lll}
m_1(A) = T(A) - \frac{1}{2}m_1^\star & m_2(A) = 1 - F(A) - \frac{1}{2}m_2^\star & m_3(A) = 0 \\
m_1(A^c) = 1 - T(A) - \frac{1}{2}m_1^\star & m_2(A^c) = F(A) - \frac{1}{2}m_2^\star & m_3(A^c) = 0 \\
m_1(A \cap A^c) = m_1^\star & m_2(A \cap A^c) = m_2^\star & m_3(A \cap A^c) = 1 - I(A) \\
m_1(A \cup A^c) = 0 & m_2(A \cup A^c) = 0 & m_3(A \cup A^c) = I(A)
\end{array}$$

where $m_1^\star$ is given by the solution of equation $64e^2(m_1^\star)^4 - (m_1^\star)^2 + 2m_1^\star - 4(1-T(A))T(A) = 0$ and $m_2^\star$ by the solution of equation $64e^2(m_2^\star)^4 - (m_2^\star)^2 + 2m_2^\star - 4(1-F(A))F(A) = 0$. The mapping $m_3(.)$ comes from the necessity to not assign a prior preference to A rather than to A^c when only indeterminacy is available.

- Case 2 : We assume now that each neutrosophic evidence corresponds only to a triplet of real intervals belonging to $[0;1]$. In this case, the more general mapping is proposed.

$$\begin{array}{lll}
m_1(A) = m_T - \frac{1}{2}m_1^\star & m_2(A) = 1 - M_F - \frac{1}{2}m_2^\star & m_3(A) = (M_I - m_I)/2 \\
m_1(A^c) = 1 - M_T - \frac{1}{2}m_1^\star & m_2(A^c) = m_F - \frac{1}{2}m_2^\star & m_3(A^c) = (M_I - m_I)/2 \\
m_1(A \cap A^c) = m_1^\star & m_2(A \cap A^c) = m_2^\star & m_3(A \cap A^c) = 1 - M_I \\
m_1(A \cup A^c) = M_T - m_T & m_2(A \cup A^c) = M_F - m_F & m_3(A \cup A^c) = m_I
\end{array}$$

where $m_T \triangleq \text{Inf}(T(A))$, $M_T \triangleq \text{Sup}(T(A))$, $m_F \triangleq \text{Inf}(F(A))$, $M_F \triangleq \text{Sup}(F(A))$ and $m_I \triangleq \text{Inf}(I(A))$, $M_I \triangleq \text{Sup}(I(A))$. $m_1^\star$ is given by the solution of equation $64e^2(m_1^\star)^4 - (m_1^\star)^2 + 2(1-M_T+m_T)m_1^\star - 4(1-M_T)m_T = 0$ and $m_2^\star$ by the solution of equation $64e^2(m_2^\star)^4 - (m_2^\star)^2 + 2(1-M_F+m_F)m_2^\star - 4(1-M_I)m_I = 0$.

- Case 3 (general case) : We assume now that each component of neutrosophic value $(T(A) = \bigcup_i T_i(A), I(A) = \bigcup_j I_j(A), F(A) = \bigcup_k F_k(A))$ is actually the union of subintervals of $[0;1]$. In such general case, we propose to construct for each possible combinations of $(T_i(A), I_j(A), F_k(A))$ a corresponding general bpa as for case 2 then combine all bpa using the general rule of combination to get the global bpa relative to the proposition under consideration.

Level 2 : the combination of evidences

We have just shown how general elementary bpa can be evaluated from each neutrosophic values of a report. For the report R_1, we have now in hands a set of bpa $m_1(.), \ldots m_m(.)$ associated to every proposition in this report. Similarly, we get also another set of bpa $m'_1(.), \ldots m'_n(.)$ for report R_2. For each set of bpa, we are now able to compute the global general bpa $m_{R_1}(.)$ and $m_{R_2}(.)$ from the general rule of combination (41) by

$$m_{R_1} = m_1 \oplus m_2 \oplus \ldots \oplus m_m$$

$$m_{R_2} = m'_1 \oplus m'_2 \oplus \ldots \oplus m'_n$$

The next step of the combination is then to combine the bpa m_{R_1} with m_{R_2} by applying for the last time the general rule of combination (41) to finally get the global result we are looking for; i.e.

$$m(.) = m_{R_1} \oplus m_{R_2}$$

From the global bpa $m(.)$ defined on the hyper-power set D^Θ, we will then be able to evaluate the degree of belief of each proposition of D^Θ which will help us to take the most pertinent decision for the problem under consideration.

6 Conclusion

In this paper, the foundations for a new theory of paradoxical and plausible reasoning has been developed which takes into account in the combination process itself the possibility for uncertain and paradoxical information. The basis for the development of this theory is to work with the hyper-power set of the frame of discernment relative to the problem under consideration rather than its classical power set since, in general, the frame of discernment cannot be fully described in terms of an exhaustive and exclusive list of disjoint elementary hypotheses. In such general case, no refinement is possible to apply directly the Dempster-Shafer theory (DST) of evidence. In our new theory, the rule of combination is justified from the maximum entropy principle and there is no mathematical impossibility to combine sources of evidence even if they appear at first glance in contradiction (in the Shafer's sense) since the paradox between sources is fully taken into account in our formalism. We have also shown that in general, the combination of evidence yields unavoidable paradoxes. This theory has shown, through many illustrated examples, that conclusions drawn from it, provides results which agree perfectly with the human reasoning and is useful to take a decision on complex problems where DST usually fails. The last part of this work has been devoted to the development of a theoretical bridge between the neutrosophic logic and this new theory, in order to solve the delicate problem of the combination of neutrosophic evidences. The neutrosophic logic serves here as the most general framework (prerequesite) for dealing with uncertain and paradoxical sources of information through this new theory.

References

[1] Akdag H.,“Une approche logique du raisonnement incertain”, Thèse de Doctorat d’Etat en Informatique, Université Pierre et marie Curie, Paris, Décembre 1992.

[2] Anderson I.,“Combinatorics of Finite Sets”, Oxford University Press, Oxford, England, 1987.

[3] Appriou A.,“Procédure d’aide à la décision multi-informateurs. Application à la classification multi-capteurs de cibles”, AGARD Avionics Panel Symposium on Software Engineering and Its Applications, Turkey, April 25-29, 1988.

[4] Bayes T., “An essay Toward Solving a Problem in the Doctrine of Chances”, Philosophical Trans. of the Royal Society, Vol. 53, pp. 370-418, 1763 (also reprinted in Biometrika, Vol. 45, pp. 293-315, 1958 and in Studies in the History of Statistics and Probability, Pearson and Kendalls eds., Hafner, pp. 131-153, 1970).

[5] Bianaghi E., Madella P.,“Inductive and Deductive Reasoning techniques for Fuzzy Dempster-Shafer Classifiers”, Proc. of 1997 Seventh IFSA World Congress, Prague, pp. 197-302, 1997.

[6] Boole G.,“An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities”, MacMillan, London, 1854 (reprinted by Dover in 1958).

[7] Cheng Y., Kashyap R.L., “Study of the Different Methods for Combining Evidence”, Proceedings of SPIE on Applications of Artificial Intelligence, Vol. 635, pp. 384-393, 1986.

[8] Comtet L., “Sperner Systems”, Advanced Combinatorics: The Art of Finite and Infinite Expansions, Dordrecht, Netherlands:Reidel, pp. 271-273, 1974.

[9] Cover T., Thomas J.A. “Elements of Information Theory”, Wiley Series in Telecommunications, John Wiley and Sons, Inc., 1991.

[10] Dedekind R.,“Über Zerlegungen von Zhalen durch ihre grössten gemeinsammen Teiler”, In Gesammelte Werke, Bd. 1., pp. 103-148, 1897.

[11] Dempster A.P.,“Upper and Lower Probabilities Induced by Multivalued Mapping”, Annals of Mathematical Statistics, Vol. 28, pp. 325-339, 1967.

[12] Dempster A.P.,“A Generalization of Bayesian Inference”, Journal of the Royal Statistical Society, Serie B, Vol. 30, pp. 205-247, 1968.

[13] Dezert J., “Optimal Bayesian Fusion of Multiple Unreliable Classifiers”, Proceedings of 4th Intern. Conf. on Information Fusion (Fusion 2001), Montréal, Aug. 7-10, 2001.

[14] Diener F., Reeb G.,“Analyse Non Standard”, Hermann, 1989.

[15] Dubois D., Prade H.,“Théories des Possibilités. Application à la Représentation des Connaissances en Informatique”, Editions Masson, Paris, 1985.

[16] Dubois D., Prade H.,“On the Unicity of Dempster Rule of Combination”, International Journal of Intelligent Systems, Vol. 1, pp. 133-142, 1986.

[17] Dubois D., Prade H.,“Evidence, Knowledge and Belief Functions”, International Journal of Approximate Reasoning, Vol. 6, pp 295-319, 1992.

[18] Erdös P., Ko C., Rado R.,“Intersection Theorems for Systems of Finite Sets”, Quart. J. Math. Oxford, 12, pp. 313-320, 1961.

[19] Hall M.,“Combinatorial Theory”, Blaisdell, 1967.

[20] Hilton A.J., Milner E.C.,“Some Intersection Theorems of Systems of Finite Sets”, Quart. J. Math. Oxford, 18, pp. 369-384, 1967.

[21] Hintikka J.,“Knowledge and Belief: an Introduction to the Two Notions”, Cornell University press, New york, 1962.

[22] Howe D.,“Dictionary of Computing”, The free Online Dictionary of Computing, Edited by Denis Howe, (http://foldoc.doc.ic.ac.uk), England, 1998.

[23] Jaynes E.T.,“Prior Probabilities”, IEEE Trans. on SSC, Vol. sec-4, no.3, pp. 227-241, 1968.

[24] Jaynes E.T.,“Where do we stand on Maximum Entropy ?”, presented at the Maximum Entropy Formalism Conference, 104 pages, MIT, May 2-4, 1978.

[25] Jaynes E.T.,“Bayesian methods: General Background”, 4th Annual Workshop on Bayesian/Maximum Entropy Methods, Univ. of Calgary, Aug. 1984 (also published in the Proceedings Volume, Maximum Entropy and Bayesian Methods in Applied Statistics, J.H. Editor, Cambridge University Press, pp. 1-25, 1985).

[26] Jaynes E.T.,“Probability Theory as a Logic”, 9th Annual Workshop on Bayesian/Maximum Entropy Methods, Dartmouth College, New Hampshire, Aug. 14, 1989 (also published in the Proceedings Volume, Maximum Entropy and Bayesian Methods in Applied Statistics, P.F. Fougere Editor, Kluwer Academic Publishers, Dordrecht, Holland, 1990).

[27] Jaynes E.T.,“Probability Theory: The Logic of Science”, Fragmentary Edition of March 1996, full edition in preparation by Cambridge Univ. Press, MA, USA.

[28] Kleitman D., “On Dedekind’s Problem: The Number of Monotone Boolean Functions”, Proc. Amer. Math. Soc., Vol. 21, pp. 677-682, 1969.

[29] Kleitman D., Markowsky G., “On Dedekind’s Problem: The Number of Isotone Monotone Boolean Functions II”, Trans. Amer. Math. Soc., Vol. 213, pp. 373-390, 1975.

[30] Klopotek M.A., “Interpretation of belief function in Dempster-Shafer theory”, Foundations of Computing and Decison Sciences, Vol. 20, no. 4, pp. 287-306, 1995.

[31] Klopotek M.A., “Identification of Belief Structure in Dempster-Shafer Theory”, Foundations of Computing and Decison Sciences, Vol. 21, no. 1, pp. 35-54, 1996.

[32] Kyburg H.E., Jr., “Bayesian and Non Bayesian Evidential Updating”, Artificial Intelligence, Vol. 31, pp. 271-294, 1987.

[33] Lowrance J.D., Garvey T.D.,“Evidential Reasoning: An Implementation for Multisensor Integration”, technical Note 307, Artificial Intelligence Center, Computer Science and Technology Division, SRI International, Menlo Park, CA, 1983.

[34] Lesh S.A.,“An Evidential Theory Approach to Judgment-Based Decision Making”, Ph.D. Thesis, Dept. of Forestry and Environmental Studies, Duke University, Durham, NC, 1986.

[35] Li X.R.,“Probability, Random Signals, and Statistics”,CRC Press, 1999.

[36] Mahler R., “Combining Ambiguous Evidence with Respect to Abiguous a priori Knowledge, I: Boolean Logic”, IEEE Trans. on SMC, Part 1: Systems and Humans, Vol. 26, No. 1, pp. 27-41, 1996.

[37] Mesalkin L.D., “A Generalization of Sperner’s Theorem on the Number of Subsets of a Finite Set”, Theory Prob., 8, pp. 203-204, 1963.

[38] Milner E.C., “A Combinatorial Theorem on Systems of Sets”, J. London Math. Soc., 43, pp. 204-206, 1968.

[39] Neapolitan R.E.,“Probabilistic Reasoning in Expert Systems”, J. Wiley, 1990.

[40] Pearl J.,“Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference”, Computer Science, Artificial Intelligence, Morgan-Kauffman, 1988.

[41] Polya G.,“Les Mathématiques et le Raisonnement Plausible”, Gauthiers-Villars, Paris, 1958.

[42] Provan G.M.,“The validity of Dempster-Shafer Belief Functions”, International Journal of Approximate Reasoning, Vol. 6, pp 389-399, 1992.

[43] Robinson A., “Non-Standard Analysis”, North-Holland Publ. Co., 1966.

[44] Rota G.C.,“Theory of Möbius Functions”, Zeitung für Wahrscheinlichkeitstheorie und Verwandte Gebiete 2, pp. 340-368, 1964.

[45] Ruspini E.H, “The Logical Foundation of Evidential Reasoning”, Tech. Note 408, SRI International, Menlo park, CA, USA, 1986.

[46] Ruspini E.H, Lowrance D.J., Start T.M.,“Understanding Evidential Reasoning”, International Journal of Approximate Reasoning, Vol. 6, pp 401-424, 1992.

[47] Schubert J.,“On Nonspecific Evidence”, International Journal of Intelligent Systems, Vol. 8, pp. 711-725, 1993.

[48] Shafer G.,“A Mathematical Theory of Evidence”, Princeton University Press, Princeton, New Jersey, 1976.

[49] Shafer G.,“Constructive Decision Theory”, Working paper, Dec. 1982 (basis for the following reference).

[50] Shafer G.,“Savage Revisited (with discussion)”, Statistical Science 1, pp. 463-501, Reprinted in Decision Making, edited by David Bell, Howard Raiffa, and Amos Tversky, Cambridge University Press, pp. 193-234, 1988.

[51] Shafer G., Pearl J.,“Readings in Uncertain Reasoning”, Morgan-Kauffman, 1990.

[52] Shafer G.,“Perspectives on the Theory and Practice of Belief Functions”, International Journal of Approximate Reasoning, Vol. 4, pp 323-362, 1990.

[53] Shafer G.,‘The Art of Causal Conjecture”, MIT Press, Cambridge, 1996.

[54] Shapiro, “On the Counting Problem for Monotone Boolean Functions”, Comm. Pure Appl. Math., 23, pp. 299-312, 1970.

[55] Sloane N.J.A., “Sequences A006826/M2469, A007153/M3551, and A014466”, in “An On-Line Version of the Encyclopedia of Interger Sequences” (http://www.research.au.com/˜njas/sequences/eisonline.html), 2001.

[56] Smarandache F.,“An Unifying Field in Logics: Neutrosophic Logic”, (Second Edition), American Research Press, Rehoboth, 2000 (ISBN 1-879585-76-6).

[57] Smarandache F.,“Collected Papers”, Vol. 3, Editura Abaddaba, Oradea, Romania, 159 pages, (ISBN 973-8102-01-4), 2000.

[58] Smarandache F.,“A Unifying Field in Logics : Neutrosophic Logic.”, to appear in Multiple Valued Logic Journal, 2001.

[59] Smets Ph.,“The Combination of Evidence in the Transferable Belief Model”, IEEE Trans. on PAMI, Vol. 12, no. 5, 1990.

[60] Smets Ph.,“Quantifying Beliefs by Belief Functions: An Axiomatic Justification”, Proc. of 13th International Joint Conf. on Artificial Intelligence, Chambéry, France, pp. 598-603, 1993.

[61] Smets Ph.,“Belief Function: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem”, International Journal of Approximate Reasoning, Vol. 9, pp. 1-35, 1993.

[62] Smets Ph.,“Practical Uses of Belief Functions”, Proc. of the 14th Conf. on Uncertainty in Artificial Intelligence, Stokholm, Sweden, 1999.

[63] Smets Ph.,“Data Fusion in the Transferable Belief Model”, Proceedings of 3rd Int. Conf. on Inf. Fusion (Fusion 2000), pp. PS-21–PS33, (http://www.onera.fr/fusion2000), Paris, July 10-13, 2000

[64] Sosnowski Z.A., Walijewski J.,“Generating Fuzzy Decision Rules with the use of Dempster-Shafer Theory”, http://cksr.ac.bialystok.pl/jwal/papers/esm/fds-esm.html.

[65] Sperner E.,“Ein Satz über Untermengen einer endlichen Menge”, Math Z., 27, pp. 544-548, 1928.

[66] Sun H., he K., Zhang B.,“The Performance of Fusion Judgment on Dempster-Shafer Rule”, Chinese Journal of Electronics, Vol. 8, no. 1, Jan. 1999.

[67] Tribus M.,“Rational, Descriptions, Decisions and Designs”, Pergamon Press Inc., 1969 (French Transl.,“Decisions Rationnelles dans l’Incertain”, Masson et Cie, Paris, 1972) .

[68] Voorbraak F.,“On the Justification of Dempster’s rule of Combination”, Artificial Intelligence, Vol. 48, pp. 171-197, 1991.

[69] Wierzchon S.T., Klopotek M.A.,“Evidential Reasoning. An Interpretative Investigation”, Wydawnictwo Akademii Podlaskiej Publisher, PL ISSN 0860-2719, 304 pages, Poland, February 2002 (http://www.ipipan.waw.pl/ klopotek/mak/book2a.htm).

[70] Yager R., “Entropy and Specificity in a Mathematical Theory of Evidence” Int. J. General Systems, Vol. 9, pp. 249-260, 1983.

[71] Yager R., Kacprzyk J., Fedrizzi M. (Editors),“Advances in the Dempster-Shafer Theory of Evidence” ISBN: 0-471-55248-8, Hardcover 608 Pages, February 1994

[72] Zadeh L.A.,“The Concept of a Linguistic variable and its Application to Approximate Reasoning I,II,III” Information Sciences, Vol. 8, Vol. 9, 1975.

[73] Zadeh L.A.,“A Theory of Approximate Reasoning” Machine Intelligence, J. Hayes, D. Michie and L. Mikulich Eds, Vol. 9, pp. 149-194, 1979.