Prejudice in uncertain information merging: Pushing the fusion paradigm of evidence theory further

Didier Dubois, Francis Faux, Henri Prade

Prejudice in uncertain information merging: pushing the fusion paradigm of evidence theory further

Didier Dubois[1], Francis Faux[1], Henri Prade[1]

IRIT, CNRS & Université de Toulouse,
118 Route de Narbonne, F-31062 Toulouse Cedex 9, France

Abstract

In his 1976 book, G. Shafer reinterprets Dempster lower probabilities as degrees of belief. He studies the fusion of independent elementary partially reliable pieces of evidence coming from different sources, showing that not all belief functions can be seen as the combination of simple support functions, representing such pieces of evidence, using Dempster rule. It only yields a special kind of belief functions called separable. In 1995, Ph. Smets has indicated that any non-dogmatic belief function can be seen as the combination of so-called *generalized* simple support functions, whose masses may lie outside the unit interval. It comes down to viewing a belief function as the result of combining two separable belief functions, one of which models reports from sources, and the other one expresses doubt, via a retraction operation. We propose a new interpretation of the latter belief function in terms of prejudice of the receiver, and consider retraction as a special kind of belief change. Its role is to weaken the support of some focal sets of a belief function, possibly stemming from the fusion of the incoming information. It provides an alternative extensive account of non-dogmatic belief functions as a theory of merging pieces of evidence *and* prejudices, which partially differs from Shafer approach's based on support functions and coarsenings. Retraction differs from discounting, revision, and from the symmetric combination of conflicting evidence. The approach relies on a so-called diffidence function on the positive reals ranging from full confidence to full diffidence. We also discuss information orderings and combination rules that rely on diffidence functions. Finally, we study the diffidence-based ordering and combination in the consonant case, and show that the diffidence view suggests a new branch of possibility theory, in agreement with likelihood functions.

Keywords: Evidence theory, information fusion, separable belief functions, likelihood functions, belief change, possibility theory, information ordering

Email addresses: dubois@irit.fr (Didier Dubois), francis.faux@univ-jfc.fr (Francis Faux), prade@irit.fr (Henri Prade)

This paper is a revised and extended version of a conference paper [8]. Section 4 has been significantly expanded; Sections 5 and 6 are new.

1. Introduction

The theory of belief functions originates in the work of Dempster [4] on upper and lower probabilities induced by imprecise observations from a sample space. Mathematically, Dempster lower probabilities are induced by a random set and coincide with Shafer belief functions. In his seminal book [34], Shafer revisits Dempster's view and presents his theory of evidence essentially as an approach to the fusion, by means of an aggregation function (called *orthogonal sum* or yet Dempster's rule of combination), of independent, more or less reliable pieces of evidence. He tries to reconstruct belief functions by combining elementary pieces of uncertain evidence in the form of so-called *simple support functions*, each representing a Boolean proposition coming from a source along with a reliability coefficient. As recalled in [37] (and in detail in [36]), such simple-support functions already appear in the works of scholars like Hooper and Lambert from the turn of the XVIIIth century on, with the purpose of merging uncertain testimonies at courts of law.

However, in evidence theory, only some belief functions, called *separable*, can be decomposed in this way. Many others prove to be not separable, i.e., are not the orthogonal sum of simple support functions. In his book [34], Shafer considers a more general class of belief functions called support functions that can be induced from separable belief functions via a coarsening of the frame of discernment. Finally, he showed that the other belief functions can be obtained as a limit of sequences of support functions.

Smets [40] proposed another approach to decompose so-called non-dogmatic belief functions into elementary components. He first generalized simple support functions, allowing some masses to escape the unit interval. Then he showed that any non-dogmatic belief function is the conjunctive combination (i.e., the orthogonal sum without the normalisation factor) of such generalized elementary belief functions. Actually, any non-dogmatic belief function can be decomposed into two standard separable belief functions combined by a so-called retraction operation. The former belief function (expressing confidence) results from the fusion of independent testimonies (understood here in a general sense, from expert opinions to sensor measurements), while the other (expressing doubt) plays the role of an inhibitor that can attenuate, via retraction, the strength of information supplied by the former. This pair of belief functions is called "Latent Belief Structure" by Smets [40]. In this bipolar approach, retraction behaves in such a way that if the strength of the reasons to believe a proposition and the strength of those not to believe it counterbalance each other, a vacuous belief function is obtained. This setting was further studied by Denœux [5] who introduces a new information ordering and an idempotent combination rule. Our aim is to revisit Smets latent belief structures, so as to use them as a basis for a more intuitive and systematic approach to an important part of evidence theory in the spirit of Shafer's book.

Our paper first revisits the concept of degree of support, valued on the unit interval, introduced by Shafer [34], then generalized to the positive real line by Smets. We interpret the mass allocated to the tautology in a simple support function in terms of diffidence with respect to a proposition taking the form of a subset. The degree of diffidence can express a range of attitudes from belief reinforcement (if less than 1), to belief attenuation (if more than 1). Such degrees can be synthetized by a full-fledged set-function that is able to represent any non-dogmatic belief function. We show how to express Moebius mass functions in terms of the diffidence functions and conversely.

In order to avoid the use of deviant simple support functions (with diffidence values greater than 1), we propose to use the retraction operation. We consider it as a belief change operation, that is capable of erasing focal elements from a belief functions, due to prior meta-information or prejudices against such focal elements. The paper proposes a systematic study of this form of belief change, which corresponds to an asymmetric merging of reasons to believe and, to quote Smets [40],"reasons not to believe". This approach also provides a systematic account of non-dogmatic belief functions as resulting from the attenuation of the support granted to the result of merging unreliable evidence, due to meta-information possessed by the receiver (that can take the form of prejudices), which provides reasons not to believe the result of the merging.

For instance, consider a variant of an example in [44]: suppose we are in the presence of two uncertain reports. The first one claims that a certain person, say Linda, is a banker. Another source asserts that Linda is a philanthropist. The conjunctive rule (typically the Dempter's rule of combination) allocates a positive belief degree to the conjunction, i.e., the fact that she is a philanthropist banker. If we suppose now that, prior to receiving these pieces of information, the prejudiced receiver does not believe in philanthropist bankers, this degree should be less than it would have been if this prejudice was not present in the mind of the receiver. Such a situation can be captured by a retraction operation applied to latent belief structures. It leads us to consider retraction as a specific belief change operation.

The organization of the paper is as follows. In Section 2, some necessary background on belief functions is introduced. We review the general bipolar decomposition of belief functions into positive and negative elementary information items proposed by Smets [40]. Section 3 presents a version of evidence theory based on diffidence functions, introduced by Smets, extended to the whole power set. Section 4 presents a generalized setting for the merging of elementary testimonies in the presence of meta-information such as prejudices, focusing on the process of belief attenuation by means of the retraction operation. This framework is illustrated on the Linda example, highlighting the difference between belief retraction and source discounting, belief revision and the fusion with conflicting inputs. We also reinterpret the normalization factor in Dempster's rule of combination as a retraction due to prejudice against contradiction. Section 5 tries to highlight the meaning of the diffidence-based information ordering of belief functions, as opposed to other better known proposals. Also the assumption underlying the idempotent diffidence combination rule from [5] is laid bare. Finally, in Section 6, we reconsider possibility theory in terms of the diffidence function, and indicate that the diffidence-based ordering is related to the comparison of likelihood ratios in statistics.

2. Background

Consider a finite set Ω, called the frame of discernment, whose elements represent descriptions of possible situations, states of the world, one of which, denoted by ω^*, corresponds to the truth. In Shafer evidence theory [34], the uncertainty concerning an agent's state of belief on the real situation $\omega^* \in \Omega$ is represented by a *basic probability assignment* (BPA) [2] defined as a mapping $m : 2^\Omega \longrightarrow [0,1]$ such that $m(\emptyset) = 0$ and verifying:

$$\sum_{E \subseteq \Omega} m(E) = 1.$$

The mass $m(E)$ represents the probability attached to the claim of knowing only that $\omega^* \in E$ and nothing more. Each subset $E \subseteq \Omega$ such as $m(E) > 0$ is called a *focal set* of m. The constraint $m(\emptyset) = 0$ has been relaxed in Smets' Transferable Belief Model (TBM) [42]. A BPA m is called *normal* if $\emptyset$ is not a focal set (subnormal otherwise), *vacuous* if Ω is the only focal element, *non-dogmatic* if Ω is a focal set, *categorical* if m has only one focal set, and it differs from Ω.

A *simple* BPA (SBPA) is a BPA whose focal sets consist of two sets $E \subset \Omega$ and Ω, of the form:

$$\begin{array}{rcl} m(E) & = & s \in [0,1], \\ m(\Omega) & = & d \in [0,1] \\ & \text{with} & s + d = 1. \end{array}$$

and is denoted by $m = E^d$, where $E \neq \Omega$, in [40, 5]. When $E \neq \emptyset$, an SBPA models an elementary testimony, where the value $d \in [0,1]$ represents the probability that the piece of information $\omega^* \in E$ is useless (can be deleted) either because a sensor is flawed or a human witness is not competent. We call the value d *diffidence value*, as it evaluates the lack of reliability of the source of information. In contrast, the piece of information $\omega^* \in E$ can be trusted with strength $s = 1 - d$, called *degree of support*. At the limit, a vacuous BPA can thus be denoted by E^1 (useless testimony) for any $E \subset \Omega$, and a categorical BPA $E \neq \Omega$ can be denoted by E^0 (safe testimony).[3]

A belief function $Bel(A)$ is a non-additive set function which represents the total quantity of evidence supporting the proposition $\omega^* \in A \subseteq \Omega$; it is defined

[2] Sometimes called mass function.

[3] Smets [40] uses the notation E^w for SBPA's, with diffidence value w. However, the symbol w denotes the weight of evidence in Shafer's book, defined by $w = -\log(1-s)$ where s is the degree of support. The reason why Shafer calls w weight of evidence is because such values add when combining two SBPAs focused on the same proposition E ($E^{d_1} \oplus E^{d_2} = E^{d_1 d_2}$, corresponding to a weight of evidence $w_1 + w_2$, where $\oplus$ stands for Dempter's rule of combination). In this paper, we prefer to use other symbols than w for diffidence numbers, so as to avoid confusion between Shafer weights of evidence and Smets diffidence values.

by:

$$Bel(A) = \sum_{\emptyset \neq E \subseteq A} m(E).$$

A belief function based on an SBPA is called a *simple support function* (SSF). A BPA m can be equivalently represented by its associated plausibility or commonality functions respectively defined for all $A \subseteq \Omega$ by:

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) = 1 - Bel(\overline{A}),$$
$$Q(A) = \sum_{B \supseteq A} m(B).$$

The plausibility $Pl(A)$ reflects the total quantity of evidence that does not support the proposition $\omega^* \notin A$. The commonality function is instrumental for combining belief functions, but its meaning is less obvious. However, we can see $Q(A)$ as the total quantity of incomplete evidence that makes *all* elements of A possible.

A BPA m can be recovered from Bel using the Möbius transform [34] or from the commonality function Q in a similar way, reversing the direction of set inclusion:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} Bel(B), \ \forall A \subseteq \Omega,$$
$$m(A) = \sum_{B \supseteq A} (-1)^{|A|-|B|} Q(B), \ \forall A \subseteq \Omega,$$

where $|X|$ represents the cardinality of X.

In [9] information fusion is described as a specific aggregation process which aims at extracting truthful knowledge from incomplete or uncertain information coming from various sources. In [1] an informal definition of fusion is proposed: "fusion consists in conjoining or merging information that stems from several sources and exploiting that conjoined or merged information in various tasks such as answering questions, making decisions, numerical estimation, etc." We now present the main conjunctive fusion rules in evidence theory.

The conjunctive combination [11] of BPA's m_j derived from k distinct sources, denoted by $m_{\bigcirc\!\!\!\!\cap} = m_1 \bigcirc\!\!\!\!\cap \ldots \bigcirc\!\!\!\!\cap m_k$ is expressed by:

$$\forall A \subseteq \Omega, m_{\bigcirc\!\!\!\!\cap}(A) = \sum_{A_1,\ldots,A_k \subseteq \Omega : A_1 \cap \ldots \cap A_k = A} \left(\prod_{j=1}^{k} m_j(A_j) \right). \tag{1}$$

Note that $m_{\bigcirc\!\!\!\!\cap}$ is not always normal, since there may exist focal sets $A_1, \ldots, A_k$ whose intersection is empty. This combination has been extensively used by Smets in the TBM [42].

The orthogonal sum of belief functions, denoted by $\oplus$, is a normalized version of the conjunctive combination rule that was first proposed in [4]. It is extensively used by Shafer [34] and is defined by:

$$m_{\oplus}(A) = K \cdot m_{\textcircled{\cap}}(A), A \neq \emptyset, \qquad (2)$$
$$m_{\oplus}(\emptyset) = 0.$$

It is also known as the orthogonal combination rule. The normalization factor K is of the form $(1 - c(m_1, \ldots, m_k))^{-1}$ where

$$c(m_1, \ldots m_k) = \sum_{A_1 \cap \ldots \cap A_k = \emptyset} \left(\prod_{j=1}^{k} m_j(A_j) \right) = m_{\textcircled{\cap}}(\emptyset)$$

represents the amount of conflict between the k sources. Note that this combination rule is defined only if $c(m_1, \ldots, m_k) < 1$.

These two combination rules are commutative, associative, and the vacuous BPA A^1 is an identity for both combination rules: $m \textcircled{\cap} A^1 = m \oplus A^1 = m$. These rules are generally used to combine BPAs from independent (hence distinct) sources.

The conjunctive rule can be expressed by multiplying commonality functions as we have $Q_{m_1 \textcircled{\cap} \ldots \textcircled{\cap} m_k} = Q_1 \cdot Q_2 \cdots Q_k$. Dempster rule is simply expressed by normalizing the latter and it holds $Q_{m_1 \oplus \ldots \oplus m_k} = K.Q_1 \cdot Q_2 \cdots Q_k$.

Shafer [34] calls *separable support function* a belief function that is the result of Dempster's rule of combination applied to SSFs with SBPAs $A_i^{d_i}$, where $A_i \neq \Omega, 0 < d_i < 1, i = 1, \ldots, k$, namely $m = \oplus_{i=1}^{k} A_i^{d_i}$. It is obtained by merging elementary independent testimonies. If all the focal sets A_i are distinct, this representation is unique. Since A^1 is an identity for $\oplus$, a separable belief function can be equivalently represented by means of a function $\delta : 2^{\Omega} \setminus \{\Omega\} \to (0, 1]$ called *diffidence function* in this paper:

$$m = \bigoplus_{\emptyset \neq A \subset \Omega} A^{\delta(A)}, \delta(A) \in (0, 1],\ \forall A \subset \Omega, A \neq \emptyset, \qquad (3)$$

where $\delta(A_i) = d_i, i = 1, \ldots, k$, and $\delta(A) = 1$ otherwise. This is called the normalized canonical decomposition of m, when it is separable (hence non-dogmatic since $m(\Omega) = \prod_{i=1}^{k} d_k > 0$).

Denœux [5] has modified this concept using the conjunctive combination of subnormal BPA's. Any BPA m that can be written as:

$$m = \textcircled{\cap}_{A \subset \Omega} A^{\delta(A)},\ \forall A \subset \Omega. \qquad (4)$$

where $0 < \delta(A) \leq 1$, is then said to be unnormalized separable (*u-separable*).

We can easily express commonality functions in terms of diffidence functions. Note that if $m_i = A_i^{d_i}, i = 1, \ldots, n$, then its commonality function is $Q_i(A) = 1$ if $A \subseteq A_i$ and d_i otherwise. So, if $m = \textcircled{\cap}_{i=1}^{n} m_i$, then it is clear that if $A \neq \emptyset$,

$$Q(A) = \prod_{i: A \not\subseteq A_i} d_i = \prod_{B: A \not\subseteq B} \delta(B).$$

3. Evidence theory based on diffidence functions

This view of belief functions as the result of merging unreliable testimonies goes back to early works in the history of probability as recalled by Shafer [35, 36], and this problem was still viewed as central for the notion of probability in the XVIIIth century as shown by the corresponding item in the Encyclopedia of Diderot and d'Alembert, as recalled by Pichon et al. [33]. It contrasts with the view of Dempster [4] in terms of upper and lower probabilities induced by imprecise statistical information. Shafer [34] [Th. 7.2 p.143] shows that if Bel is a separable belief function, and A and B are two of its focal sets such as $A\cap B \neq \emptyset$, then $A\cap B$ is a focal set of Bel.[4] But the converse is not true. This necessary condition clearly indicates that not all belief functions are separable.

To overcome this difficulty, Smets [40] generalized the concept of simple support function, allowing the diffidence values to range on the positive reals. A generalized simple support function (GSSF) focused on E is still denoted by E^d, but now $d \in (0, +\infty)$. According to Smets [40], a GSSF focusing on a subset E with a small diffidence value $d < 1$ represents the idea that "one has some reason to believe that the actual world is in E (and nothing more)", whereas, when $d > 1$, it expresses the idea that "one has some reason not to believe that the actual world is in E".

Smets has shown that any *non dogmatic* belief function can be decomposed into the conjunctive combination of GSSF's, i.e. (4) holds if we extend the range of diffidence functions δ from $(0, 1]$ to $(0, +\infty)$. He showed that for every $A \subset \Omega$, the values $\delta(A)$ can be obtained from the commonality function of m as:

$$\delta(A) = \prod_{B\supseteq A} Q(B)^{(-1)^{|B|-|A|+1}} = \frac{\prod_{C\cap A=\emptyset,|C|_{\text{odd}}} Q(A\cup C)}{\prod_{C\cap A=\emptyset,|C|_{\text{even}}} Q(A\cup C)}. \tag{5}$$

Note that in the above expression (5), there are as many factors in the numerator as in the denominator (see Shafer [34] Lemmas 2.1 and 2.2 p. 47-48). In the following we are interested in retrieving the mass function m from its diffidence function δ via the commonality function rather than through the conjunctive combination.

3.1. Diffidence functions, revisited

The range of the diffidence function δ can be extended to the whole of 2^Ω, even if only sets $A \subset \Omega$ appear in the decomposition formula (4). This idea was first suggested in [22]. Indeed, since $\delta(A) = \prod_{B\supseteq A} Q(B)^{(-1)^{|B|-|A|+1}}$, we can define $\delta(\Omega) = Q(\Omega)^{(-1)^{|\Omega|-|\Omega|+1}}$ and we get:

$$\delta(\Omega) = 1/Q(\Omega). \tag{6}$$

[4]The condition $A\cap B \neq \emptyset$ can be dropped if we allow for u-separable sub-normalized belief functions.

This expression is well-defined for non-dogmatic belief functions only since for them, $Q(\Omega) > 0$ by definition. Of course, $\delta(\Omega) > 1$,[5] but this will be also the case for the diffidence values of other subsets for all non-dogmatic belief functions.

Moreover, note that $Q(\Omega) = \prod_{B:\Omega\not\subseteq B} \delta(B) = \prod_{B\neq\Omega} \delta(B)$. So, we can define an (extended) diffidence function as a mapping $\delta : 2^\Omega \to (0, +\infty)$, such that:

$$\prod_{A\subseteq\Omega} \delta(A) = 1, \text{and} \quad \delta(\Omega) \geq 1. \tag{7}$$

Based on this extended definition, we can express the commonality function of a BPA m in terms of its diffidence function in a way more faithful to the usual expression of commonality:

Proposition 1. *The commonality function Q can be obtained from the diffidence function δ of a BPA m as : for any $A \subseteq \Omega$,*

$$Q(A) = \frac{1}{\prod_{A\subseteq B} \delta(B)}. \tag{8}$$

Proof: $Q(A) = \prod_{B:A\not\subseteq B} \delta(B) = \frac{1}{\prod_{A\subseteq B} \delta(B)}$ since $\prod_{A\subseteq\Omega} \delta(A) = 1$. $\square$.

Note that in (8), the value $\delta(\Omega)$ appears explicitly in all the expressions of $Q(A)$ for all subsets A.

Finally, we can retrieve the BPA m, from the diffidence function δ computed from it.

Theorem 1. *For any non dogmatic belief function, the BPA $m : 2^\Omega \longrightarrow [0, 1]$ is obtained from the diffidence function δ directly as:*

$$m(A) = \sum_{A\subseteq B} (-1)^{|B|-|A|} \left(\frac{1}{\prod_{B\subseteq C} \delta(C)}\right). \tag{9}$$

Proof: The BPA m is obtained from the commonality function as: $m(A) = \sum_{A\subseteq B} (-1)^{|B|-|A|} Q(B)$. Due to (8), $m(A) = \sum_{B:A\subseteq B} (-1)^{|B|-|A|} \left(\frac{1}{\prod_{B\subseteq C} \delta(C)}\right)$. $\square$

Note that the mass function m_δ derived, from any function δ obeying (7), via (9), is not always positive. Indeed, suppose that $\delta(A) = \lambda < 1, \delta(B) = \mu > 1, \delta(C) = 1, C \neq \Omega$ otherwise, and $\lambda\mu < 1$ (so $\delta(\Omega) = 1/\lambda\mu > 1$). By means of the conjunctive rule, one gets the BPA: $m_\delta(A \cap B) = (1 - \lambda)(1 - \mu), m_\delta(A) = (1-\lambda)\mu, m_\delta(B) = (1-\mu)\lambda, m_\delta(\Omega) = \lambda\mu$. It is clear that $m_\delta(A\cap B)$, and $m_\delta(B)$ are negative.

So the function $m \mapsto \delta$ is injective, but it is not surjective. Namely, given a non-negative function δ such that $\prod_{A\subseteq\Omega} \delta(A) = 1$ and $\delta(\Omega) \geq 1$, Q_δ obtained

[5] This value cannot be used to define a GSSF focused on Ω, since Ω^d is not defined.

by (8) is a set-function that may range partially outside $[0, 1]$, i.e., Q_δ is not always a commonality function.

Example 1. *Consider* $\Omega = \{a, b\}$ *and* $\delta(\emptyset) = 1, \delta(\{a\}) = 4, \delta(\{b\}) = 0.1$. *Then* $\delta(\Omega) = 1/0.4$ *and* $Q(\{b\}) = 4$.

However diffidence functions such that $\delta(A) \leq 1, \forall A \subset \Omega$ are in one to one correspondence with BPAs of separable belief functions. In fact as the sets A with $\delta(A) < 1$ are characteristic of the separable BPA with diffidence function δ, we can claim that

Proposition 2. *Given a non-dogmatic BPA* m*, and* δ *its diffidence function, if* $\delta(A) < 1$ *then* A *is a focal set of* m.

Proof: Suppose $\delta(A) < 1$. When reconstructing m using the conjunctive rule of combination, we combine GSSFs m_B such that $m_B(B) = 1 - \delta(B); m_B(\Omega) = \delta(B), \forall B \in \Omega$. Of course, in the combination, we can delete m_B when $\delta(B) = 1$. It is clear that via such combination, we get that the resulting mass of set A is $m(A) \geq m_A(A) \prod_{B \neq A} m_B(\Omega) = (1 - \delta(A)) \prod_{B \neq A} \delta(B) > 0$. Hence if $\delta(A) < 1$, then A is a focal set of m. □

The converse is false, namely there may be focal sets C for which $\delta(C) \geq 1$.

Example 2. *Let* m *be the non-dogmatic BPA such that* $m(A) = 0.088, m(B) = 0.528, m(A \cap B) = 0.032, m(\Omega) = 0.352$. *It canonically decomposes as* $m = A^{0.8} \circledcirc B^{0.4} \circledcirc (A \cap B)^{1.1}$. *Clearly,* $A \cap B$ *is focal (*$m(A \cap B) > 0$*), but* $\delta(A \cap B) > 1$. *And* $A \cap B$ *is also a focal set of the decomposable BPA* $m' = A^{0.8} \circledcirc B^{0.4}$ *since* $m'(A \cap B) = (1 - 0.8)(1 - 0.4) = 0.12$ *and* $\delta'(A \cap B) = 1$.

The interpretation of the δ function in terms of diffidence contrasts with the proposal in [31] to interpret its logarithm in terms of mutual measures of information related to sources.

3.2. The bipolar representation of belief functions

We start from the expression of a BPA m in terms of the conjunctive combination of GSSF's given by (4) using the diffidence function δ. Let $\mathcal{T}, \mathcal{P} \subseteq 2^\Omega$ be families of subsets such that $A \in \mathcal{T} \subseteq 2^\Omega$ if and only if $\delta(A) < 1$ and $B \neq \Omega \in \mathcal{P}$ if and only if $\delta(B) > 1$. We call sets in $\mathcal{T}$ *testimonies*. Factors of the form $A^{\delta(A)}$, where $A \in \mathcal{T}$, represent evidence in favor of A. The sets $B \in \mathcal{P}$ are propositions against which there are prejudices (we call such propositions *prejudices* for short). More generally, according to Smets [40], they are propositions you have "some reason not to believe" to the extent suggested by the fusion of pieces of evidence in $\mathcal{T}$.

By construction, $\mathcal{T} \cap \mathcal{P} = \emptyset$. As we shall see later, the role of such factors as $B^{\delta(B)}, \delta(B) > 1$ in the decomposition of m will be to weaken the supports (possibly make them vanish) of intersections of sets representing propositions coming from several independent sources. For instance, for the BPA m in the above Example 2, $\mathcal{T} = \{A, B\}$ and $\mathcal{P} = \{A \cap B\}$, and the mass assigned to $A \cap B$ is decreased from 0.12 (when $\mathcal{P} = \emptyset$) to 0.032.

Define δ^+ and δ^- to be standard diffidence functions valued in $(0, 1)$ defined from the original one δ associated to m, such that: $\delta^+(A) = \min(1, \delta(A))$, and $\delta^-(A) = \min(1, 1/\delta(A)), \forall A \subset \Omega$. The decomposition (4) of a separable belief function can be extended to non-dogmatic belief functions as:

$$m = \textcircled{\cap}_{A \subset \Omega} A^{\delta(A)} = (\textcircled{\cap}_{A \subset \Omega} A^{\delta^+(A)}) \textcircled{\cap} (\textcircled{\cap}_{B \subset \Omega} B^{\frac{1}{\delta^-(B)}}).$$

Let $m^+ = \textcircled{\cap}_{A \subset \Omega} A^{\delta^+(A)}$ be the separable BPA with diffidence function δ^+ and $m^- = \textcircled{\cap}_{B \subset \Omega} B^{\delta^-(B)}$ be the separable BPA with diffidence function δ^-.

The commonality function of m can be expressed using (8) in terms of the commonality functions of m^+ and m^-, say Q^+ and Q^- respectively as:

$$Q(C) = \frac{1}{\prod_{C \subseteq A} \delta(A)} = \frac{\delta^-(\Omega) \prod_{C \subseteq B, B \in \mathcal{P}} \delta^-(B)}{\delta^+(\Omega) \prod_{C \subseteq A, A \in \mathcal{T}} \delta^+(A)} = \frac{Q^+(C)}{Q^-(C)}$$

noticing that $\delta(\Omega) = \frac{\delta^+(\Omega)}{\delta^-(\Omega)}$. An operation $\oslash$, called *retraction*, can thus be defined by the division of commonality functions [40]:

$$Q_1 \oslash Q_2 = \frac{Q_1}{Q_2}.$$

It is called *decombination* by Smets [40] or sometimes removal [38]. We can write the decomposition (4) of a non-dogmatic belief function using retraction as

$$m = (\textcircled{\cap}_{A \in \mathcal{T}} A^{\delta^+(A)}) \oslash (\textcircled{\cap}_{B \in \mathcal{P}} B^{\delta^-(B)}).$$

A non-dogmatic BPA m can thus be decomposed in a unique irredundant way as a pair (m^+, m^-), of separable belief functions induced by separable BPAs m^+ and m^-, such that $m = m^+ \oslash m^-$. The existence of positive and negative information is generally coined under the term bipolarity [15], an idea already exploited to discuss latent belief structures more than 15 years ago in [16].

The confidence component denoted by m^+ is a BPA obtained from the merging of SBPAs, with focal sets in $\mathcal{T}$, and the diffidence component denoted by m^- is a BPA obtained likewise, with focal sets in $\mathcal{P}$. The pair (m^+, m^-) of separable BPAs is called a *latent belief structure* by Smets [40].

Testimonies in $\mathcal{T}$ are focal sets, but not all focal sets are in $\mathcal{T}$, as focal sets may include intersections of source sets. A belief function with diffidence function δ is separable if and only if $\delta = \delta^+$. So a u-separable belief function will be of the unique form: $m = \textcircled{\cap}_{A \in \mathcal{T}} A^{\delta^+(A)}$, in which case, as shown by Shafer [34] any intersection of a set of sources $\cap_{i \in \mathcal{T}} A_i$ of m will be a focal set as well.

It is easy to check that a belief function is separable by inspecting its diffidence function but it is much less obvious by considering its BPA. Nevertheless, replacing commonality by its expression in terms of m, equation (5) leads to a characteristic condition for separability on the BPA: the BPA m is separable if and only if

$$\forall A \subset \Omega, \frac{\prod_{C \cap A = \emptyset, |C|_{\text{odd}}} \sum_{D \supseteq C} m(A \cup D)}{\prod_{C \cap A = \emptyset, |C|_{\text{even}}} \sum_{D \supseteq C} m(A \cup D)} \leq 1.$$

On a two-element frame, all belief functions are separable; this is not the case for more than 2 elements. In the general case, checking this condition is quite unwieldy. The problem of proving separability directly from the BPA proves to be not trivial and has exponential complexity in the number of elements of the frame due to the use of focal sets. In [8] these separability conditions are completely solved in the simple case of two overlapping focal sets on a 4-element frame.

3.3. Related work

Shafer's theory of evidence [34] includes as a pivotal notion the rule of combination of pieces of evidence known as orthogonal sum. It seems clear that the thesis in this book is that a belief function should be the result of merging elementary pieces of evidence. Doing so only yields separable belief functions. However there is a need to go beyond separable belief functions. In Shafer's book this is done by considering support functions, that are defined as belief functions that result from coarsening a separable belief function. Coarsening consists in building an approximation space on Ω, in the form of a partition $\{A_1, A_2, \ldots, A_q\}$. Each subset $A \subseteq \Omega$ is replaced by its upper approximation $A^* = \cup\{A_i : A_i \cap A \neq \emptyset\}$ (in the sense of rough set theory [29]).[6] Conversely, Ω is a refinement of $\Theta = \{\theta_1, \theta_2, \ldots, \theta_q\}$ if all elements θ_i are expanded into disjoint sets $A_i, \forall i = 1, \ldots, q$. Given a BPA m on Ω, the new BPA m^* on the coarsened space $\Theta = \{\theta_1, \theta_2, \ldots, \theta_q\}$ is such that

$$\forall B \subseteq \Theta, m^*(B) = \sum_{E \subset \Omega : E^* = \cup_{\theta_i \in B} A_i} m(E)$$

(the sum of masses of focal sets of m with the same upper approximation). A BPA μ on Θ defines a support function if and only if there exists a refinement Ω of Θ and a separable BPA m on Ω such that $\mu = m^*$. As noticed by Smets [40], coarsening is not in agreement with Dempster rule of combination, i.e., $(m_1 \oplus m_2)^* \neq m_1^* \oplus m_2^*$. He advocates his decomposition method as avoiding this problem.

Interestingly, Shafer proves that a characteristic property of support functions is that their core defined by $\cup\{E \in \mathcal{F}(m)\}$ is a focal set of m. Support functions of Shafer are closely related to non-dogmatic belief functions. It is obvious that the latter are support functions, since $m(\Omega) > 0$. Conversely, support functions are non dogmatic on their cores (eliminating impossible elements from Ω). Smets' decomposition in terms of latent belief structures can thus be viewed as an alternative interpretation of Shafer support functions.

[6] In fact, Chapter 6 of Shafer's book contains the basic notions of rough set theory, written before Pawlak proposed them.

Ten years before Smets' 1995 paper [40], Ginsberg [20] had proposed a special case of the retraction operator applied to belief functions in a frame of discernment with only two elements $\Omega = \{a, b\}$. Two coefficients $(\alpha, \beta), \alpha + \beta \leq 1$, are given, where $\alpha = m(a)$ and $\beta = m(b)$. An operation of inversion $\oslash$ is defined such that the pair $(x, y) = (\alpha, \beta) \oslash (\gamma, \delta)$ is the solution of $(\gamma, \delta) \oplus (x, y) = (\alpha, \beta)$ using the orthogonal rule. The solution is of the form $x = \dfrac{(\alpha\bar{\delta} - \bar{\beta}\gamma)(\bar{\gamma})}{\bar{\gamma}\bar{\delta} - \bar{\beta}\gamma\bar{\gamma} - \bar{\alpha}\delta\bar{\delta}}$ and $y = \dfrac{(\beta\bar{\gamma} - \bar{\alpha}\delta)(\bar{\delta})}{\bar{\gamma}\bar{\delta} - \bar{\beta}\gamma\bar{\gamma} - \bar{\alpha}\delta\bar{\delta}}$ with $\bar{x} = 1 - x$ for $x \in \{\alpha, \beta, \gamma\}$. It is clearly a retraction of the BPA defined by (γ, δ) from the BPA defined by (α, β) such that:

$$(\alpha, \beta) \oslash (\gamma, \delta) = \{a\}^{\delta_a} \oplus \{b\}^{\delta_b} \oplus \{a\}^{1/\delta_c} \oplus \{b\}^{1/\delta_d} = \{a\}^{\delta_a/\delta_c} \oplus \{b\}^{\delta_b/\delta_d}$$

with $\delta_a = \frac{1-\alpha-\beta}{1-\beta}$, $\delta_b = \frac{1-\alpha-\beta}{1-\alpha}$, $\delta_c = \frac{1-\gamma-\delta}{1-\delta}$ and $\delta_d = \frac{1-\gamma-\delta}{1-\gamma}$. Ginsberg's notion can therefore be viewed as the forerunner of the retraction operation of Smets.

In [26], Kramosil proposed an alternative solution to solve the inversion problem (find m such that $m \oplus m_1 = m_2$) based on the apparatus of measure theory. His approach was to replace probability measures, used for defining classical belief functions, by the so-called signed measures which can take values outside the unit interval of real numbers including the negative and even infinite ones. A basic signed measure assignment (BSMA) defined on Ω is a mapping $m : 2^\Omega \to \mathbb{R}^\star = \mathbb{R} \cup \{+\infty, -\infty\}$ such that $m(\emptyset) = 0$ and m takes at most one of the infinite values $-\infty, +\infty$. A commonality function induced by a BSMA m is the mapping $q : 2^\Omega \to \mathbb{R}^\star$ defined by $q(A) = \sum_{B \supseteq A} m(B)$ for each $A \subseteq \Omega$. A finite BSMA m over Ω is called invertible, if $q(A) \neq 0$ holds for each $A \subseteq \Omega$. Kramosil suggested that the notion of q-invertibility may be seen as generalizing non-dogmaticism. It is interesting to point out that the idea of prejudice was evoked in [25]. In the latter paper, Kramosil suggested that a negative degree of belief could quantify "some a priori prejudice of the subject in question which needs to be neutralized by some positive degree of belief to arrive at the absolutely neutral outcome position defined by the vacuous belief function". In [30], Pichon has pursued Kramosil's seminal work by defining the so-called *conjunctive signed weight function* as a mapping $s\delta : 2^\Omega \backslash \{\Omega\} \to \mathbb{R} \backslash \{0\}$ such that $s\delta(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}$ for all $A \subseteq \Omega$. But the absence of a semantic and intuitive interpretation of such generalized belief functions and the fact that only the conjunctive rule is used to combine BSMAs (normalisation cannot be applied) are impediments to the potential use of this approach.

Some sufficient conditions for separability and non-separability of belief functions are given by Ke et al. [23]. In [31], Pichon makes a critical review of Smets' canonical decomposition. He argues that Smets' solution has a major weakness, namely, it involves elementary items whose proposed semantics lacks formal justifications. He proposed a new canonical decomposition based on a means to induce belief functions from the multivariate Bernoulli distribution and on Teugels representation of this distribution [43]. According to this decomposition, a belief function results from as many crisp pieces of information as there are elements in its domain, from simple probabilistic knowledge concerning their marginal reliability, *and dependencies* between their reliabilities. All belief functions are then viewed as the result of dependent merging of SSFs focused on complements of singletons, which is enough to generate all subsets of the frame by conjunctive combination. The use of a dependency model is a crucial additional degree of freedom that extends the range of belief functions generated from elementary SSFs beyond separable ones, namely it is powerful enough to cover all belief functions. In this approach, the use of the Bernoulli distribution facilitates the mastering of dependences between sources. The purely technical restriction to such SSFs is counterbalanced by the freedom in choosing a dependence structure.

In contrast, our paper sticks to the spirit of Smets' decomposition, trying to interpret the non-standard SSF's involved as prejudices[7] or meta-information and to envisage the retraction operation as a natural form of non-symmetric combination between prior information and new input ones coming from sources.

4. Information fusion using retraction

In this section, we focus on the notion of retraction as a tool for belief change. First we study the retraction of an SSF from another SSF, and then the retraction of an SSF from a separable belief function. We interpret the effect of retraction in terms of prejudices affecting the result of an information fusion process. It leads us to view any non-dogmatic belief function as the result of a prejudiced fusion of elementary pieces of evidence.

4.1. Properties of retraction

Consider the retraction $A^x \oslash B^y = A^x \bigcirc\!\!\!\!\!\cap\, B^{1/y}$ of a simple BPA $B^y, y < 1$ supporting B from a simple support function $A^x, x < 1$. It corresponds to diffidence functions $\delta_A(E) = x$ if $E = A, 1/x$ for $E = \Omega$ and 1 otherwise; $\delta_B(E) = y$ if $E = B, 1/y$ for $E = \Omega$ and 1 otherwise. So the retraction of B^y from A^x yields the diffidence function $\delta = \delta_A/\delta_B$.

However the result $A^x \oslash B^y$ is NOT a belief function in general as the BPA induced by δ may fail to be positive. Actually, retracting an SSF B^y from A^x when $B \neq A$, never yields a belief function since we get $\delta(B) = 1/y > 0$. This is summarized by the following observation:

Proposition 3. *If $x, y \in (0, 1)$, $A^x \oslash B^y$ is a belief function if only if $A = B$ and $x/y < 1$.*

Proof: The mass $m(B)$ resulting from $A^x \oslash B^y$ is $m(B) = (1 - 1/y)x < 0$ if $B \neq A$. So it is not a belief function. If $A = B, m(A) = 1 - x/y > 0$ if and only

[7]Taken literally, i.e., propositions judged true or false by the agent prior to receiving the new evidence

if $x < y$. $\square$

In this simple case, we could write the result of the retraction operation as $A^x \oslash A^y = A^{\min(1,x/y)}$. At the limit, $A^x \oslash A^x$ yields the vacuous belief function ($m(\Omega) = 1$), resulting in total ignorance.

Besides, retraction does not work if we try to retract from a belief function a set that is not focal.

Proposition 4. *Consider a non-dogmatic BPA m with focal sets forming a family $\mathcal{F}$, and let $B \notin \mathcal{F}$. Then, if $y < 1$, $m' = m \oslash B^y$ is not a belief function.*

Proof: Compute $m'(B)$ using the conjunctive rule. Since $B \notin \mathcal{F}$ and since m is non-dogmatic, B will be a focal set of m' with support $m'(B) = (1 - 1/y) \sum_{A:B \subset A} m(A) < 0$. This mass is negative, hence the result. $\square$

The same problem will occur if we we try to retract from a belief function a focal set that only overlaps some other focal sets, without containment.

Proposition 5. *Consider a non-dogmatic normal BPA m with focal sets forming a family $\mathcal{F}$, and let $B \neq \Omega \in \mathcal{F}$ such that B neither contains nor is contained in another focal set different from Ω. Then, if $y < 1$, $m' = m \oslash B^y$ is not a belief function.*

Proof: It is clear that the set $\mathcal{F}'$ of focal sets of m' contains $\mathcal{F}$ since Ω is focal for B^y. Moreover the other focal sets of m' are of the form $A \cap B, A \in \mathcal{F}$. However $A \cap B \notin \mathcal{F}$ by assumption. So $m'(A \cap B) = m(A)(1 - 1/y) < 0$. $\square$

Clearly the belief function in the previous proposition is not separable. It is easy to see from the above example that in order to avoid the conditions in Proposition 5, it is necessary and sufficient to require that the set of focal sets of m be closed under intersection, which is a necessary but not sufficient condition of separability.

Example 3. *Consider the belief function with focal sets A, B, Ω, where $A \cap B \neq \emptyset$ and $A \not\subset B, B \not\subset A$, and let us try to retract B by computing $m' = m \oslash B^y$ for $y < 1$. It is easy to see that $m'(A \cap B) = (1 - 1/y)m(A)$ and also $m'(A) = m(A)/y; m'(B) = m(B) + (1 - 1/y)m(\Omega); m'(\Omega) = m(\Omega)/y$. Clearly, $m'(A \cap B) < 0$ so that we cannot retract B without going out of the range of belief functions. However if $A \cap B$ is also focal for m, then $m'(A \cap B) = (1 - 1/y)m(A) + m(A \cap B)$. Now, if $1 > y > \max(\frac{m(\Omega)}{m(B)+m(\Omega)}, \frac{m(A)}{m(A \cap B)+m(A)})$, the result is a belief function where the masses of B and $A \cap B$ are decreased.*

As we have seen, retracting a focal set B from m consists in the conjunctive combination $m' = m \circledcap B^z$ with $z = 1/y > 1$. It is clear that the resulting BPA is of the form:

$$\forall E \subseteq \Omega, m'(E) = zm(E) + (1 - z) \sum_{A: A \cap B = E} m(A).$$

Note that the term $\sum_{A:A\cap B=E} m(A)$ is the result of (un-normalized) conditioning of m by B, and could be written $[m\circledcirc B^0](E)$, using the categorical belief function B^0.

Proposition 6. *Suppose the set of focal sets of m is closed under conjunction. Then $m' = m\circledcirc B^z$ with $z > 1$ will be a regular BPA provided that*

$$z < \min(\frac{Q(B)}{Q(B) - m(B)}, \min_{E\subset B} \frac{\sum_{D\subseteq \overline{B}} m(E\cup D)}{\sum_{\emptyset\neq D\subseteq \overline{B}} m(E\cup D)}, \min_{E\not\subseteq B} 1/m(E)).$$

Proof: Clearly, focal sets of m' are the same as those of m. Then we have, using the previous identity:

$$\forall E \in \mathcal{F}, m'(E) = \begin{cases} zm(B) + (1-z)Q(B) & \text{if } E = B, \\ m(E) + (1-z)\sum_{\emptyset\neq D\subseteq \overline{B}} m(E\cup D) & \text{if } E \subset B, \\ zm(E) & \text{otherwise.} \end{cases} \tag{10}$$

Enforcing $m'(E) > 0$ directly yield the upper bounds on z. □

We can see that retraction will decrease the masses of B and its subsets that are focal. But the condition for ensuring the result is a regular mass function is non-trivial.

4.2. Retraction of focal set from a separable belief function

Retraction will behave better on a separable belief function. Indeed, not only Proposition 5 does not apply since the set of focal sets is closed under conjunction, but the masses will have specific regularities. Let $m = \circledcirc_{i=1}^k E_i^{\delta_i}$, where the E_i's are distinct, $\delta_i < 1, \forall i$, and, for the moment being, there is no containment relation between them. Denote by I, J subsets of indices in $\{1,\ldots,n\}$. Focal sets of m are of the form $E_I = \cap_{i\in I} A_i; I \subseteq \{1,\ldots,k\}$ with masses:

$$m(E_I) = \prod_{i\in I}(1-\delta_i)\prod_{i\notin I}\delta_i \tag{11}$$

(where we allow that some E_I's may be identical).

$\circledcirc$	E_J	$E_I, I\subset J$	$E_I : J \subset I$	E_I, no inclusion
$E_J\,(1-z)$	E_J	E_J	E_I	$E_{J\cup I}$
$\Omega\ (z>1)$	E_J	E_I	E_I	E_I

Table 1: Combination with a focal set

Suppose we combine m with a generalized SSF $B^z, z > 1$, with $B = E_J$. Combining this BPA with $(E_J)^z$ yields a BPA $m' = m\circledcirc(E_J)^z$ whose focal sets are obtained as follows (see Table 1):

- E_J: $m'(E_J)$ is obtained by combining $(E_J, m(E_J))$ with (Ω, z), or by combining $(E_I, m(E_I))$, for $I \subseteq J$ with $(E_J, (1-z))$. Hence $m'(E_J) = zm(E_J) + (1-z)(\sum_{I \subseteq J} m(E_I)) = m(E_J) + (1-z)(\sum_{I \subset J} m(E_I))$ (where $E_\emptyset = \Omega$ by convention). It is also $m'(E_J) = zm(E_J) + (1-z)Q(E_J)$.

- $\{E_I : J \subset I\}$: $m'(E_I)$ is obtained by combining $(E_I, m(E_I))$ with (Ω, z), or with $(E_K, m(E_K))$ for $K \cup J = I$ (equivalently, $I \backslash J \subseteq K \subseteq I$) with $(E_J, (1-z))$. Hence $m'(E_I) = zm(E_I) + (1-z)(\sum_{I \backslash J \subseteq K \subseteq I} m(E_K))$

- $\{E_I : J \not\subset I\}$: $m'(E_I)$ is only obtained by combining $(E_I, m(E_I))$ with (Ω, z), hence $m'(E_I) = zm(E_I)$.

Example 4. *Suppose a separable belief function with BPA m is the result of combining three SSFs $E_i^{d_i}, i = 1, 2, 3$ where no inclusion between the sets E_i holds. Suppose we try to retract E_1 with strength $z > 1$. Table 2 provides the resulting focal sets of the combination with $E_i \cap \ldots \cap E_k$ denoted by $E_{i \ldots k}$*

$\bigcirc\!\!\!\!\cap$	E_1	E_2	E_3	E_{12}	E_{13}	E_{23}	E_{123}	Ω
E_1	E_1	E_{12}	E_{13}	E_{12}	E_{13}	E_{123}	E_{123}	E_1
Ω	E_1	E_2	E_3	E_{12}	E_{13}	E_{23}	E_{123}	Ω
E_{12}	E_{12}	E_{12}	E_{123}	E_{12}	E_{123}	E_{123}	E_{123}	E_{12}

Table 2: Combination with focal sets E_1 or $E_1 \cap E_2$

We get the following masses after retraction

- $m'(E_1) = m(E_1) + (1-z)m(\Omega)$
- $m'(\Omega) = zm(\Omega)$
- $m'(E_1 \cap E_2) = m(E_1 \cap E_2) + (1-z)m(E_2)$
- $m'(E_1 \cap E_3) = m(E_1 \cap E_3) + (1-z)m(E_3)$
- $m'(E_1 \cap E_2 \cap E_3) = m(E_1 \cap E_2 \cap E_3) + (1-z)m(E_2 \cap E_3)$
- $m'(E_2) = zm(E_2); m'(E_3) = zm(E_3); m'(E_2 \cap E_3) = zm(E_2 \cap E_3)$

In terms of diffidence values for the SSFs it reads

- $m'(E_1) = (1 - zd_1)d_2d_3$
- $m'(\Omega) = zd_1d_2d_3$
- $m'(E_1 \cap E_2) = (1-d_1)(1-d_2)d_3 + (1-z)d_1(1-d_2)d_3 = (1-d_2)d_3(1-zd_1)$
- $m'(E_1 \cap E_3) = (1 - zd_1)d_2(1 - d_3)$
- $m'(E_1 \cap E_2 \cap E_3) = (1-d_1)(1-d_2)(1-d_3) + (1-z)d_1(1-d_2)(1-d_3) = (1-d_2)(1-d_3)(1-zd_1)$

- $m'(E_2) = zd_1(1-d_2)d_3; m'(E_3) = zd_1d_2(1-d_3)$; $m'(E_2 \cap E_3) = zd_1(1-d_2)(1-d_3)$

When z increases from 0 to $1/d_1$ the masses of E_1 and its subsets decrease in the same proportion, while those of E_2, E_3 and its intersection increase. It is clear that when z attains its maximal value $1/d_1$, focal set E_1 is deleted along with its subsets $E_1 \cap E_2, E_1 \cap E_3, E_1 \cap E_2 \cap E_3$ and other focal sets are improved accordingly, in fact we observe the unsurprizing result that $E_1^{d_1} \bigcirc\!\!\!\!\!\cap\, E_2^{d_2} \bigcirc\!\!\!\!\!\cap\, E_3^{d_3} \oslash E_1^{1/d_1} = E_2^{d_2} \bigcirc\!\!\!\!\!\cap\, E_3^{d_3}$ (since $zd_1 = 1$).

When retracting $E_1 \cap E_2$ (line 3 of Table 2) one gets the following results

- $m'(E_1 \cap E_2) = m(E_1 \cap E_2) + (1-z)(m(E_1) + m(E_2) + m(\Omega)) = d_3(1 - z(d_1 + d_2 - d_1d_2))$
- $m'(E_1) = z(m(E_1)) = z(1-d_1)d_2d_3$, $m'(E_2) = z(m(E_2)) = zd_1(1-d_2)d_3; m'(\Omega) = zm(\Omega) = zd_1d_2d_3$
- $m'(E_1 \cap E_2 \cap E_3) = m(E_1 \cap E_2 \cap E_3) + (1-z)(m(E_3) + m(E_1 \cap E_3) + m(E_2 \cap E_3)) = (1-d_3)(1 - z(d_1 + d_2 - d_1d_2))$
- $m'(E_3) = z(m(E_3)) = zd_1d_2(1-d_3); m'(E_2 \cap E_3) = z(m(E_2 \cap E_3)) = zd_1(1-d_2)(1-d_3); m'(E_1 \cap E_3) = z(m(E_1 \cap E_3)) = z(1-d_1)d_2(1-d_3)$

When z increases from 0 to $1/(d_1 + d_2 - d_1d_2)$ the masses of $E_1 \cap E_2$ and its subset $E_1 \cap E_2 \cap E_3$ decrease in the same proportion, while those of other subsets increase. It is clear that when z attains its maximal value, focal sets $E_1 \cap E_2$ and $E_1 \cap E_2 \cap E_3$ are deleted and other focal sets are improved accordingly.

Proposition 7. *Retracting a focal set E_J from a separable BPA m will decrease its mass and may delete it.*

Proof: Note that $m'(E_J)$ can be expressed as:

$$\begin{aligned}
m'(E_J) &= \sum_{I \subseteq J} m(E_I) - z \sum_{I \subset J} m(E_I) \\
&= \sum_{I \subseteq J} \prod_{i \in I} (1-\delta_i) \prod_{i \notin I} \delta_i - z \sum_{I \subset J} \prod_{i \in I} (1-\delta_i) \prod_{i \notin I} \delta_i \\
&= (\prod_{i \notin J} \delta_i)(\sum_{I \subseteq J} \prod_{i \in I} (1-\delta_i) \prod_{i \in J \setminus I} \delta_i - z \sum_{I \subset J} \prod_{i \in I} (1-\delta_i) \prod_{i \in J \setminus I} \delta_i
\end{aligned}$$

But notice that $\sum_{I \subseteq J} \prod_{i \in I}(1-\delta_i) \prod_{i \in J \setminus I} \delta_i = ((1-\delta_i) + \delta_i)^{|J|} = 1$.

Hence

$$m'(E_J) = (\prod_{i \notin J} \delta_i)(1 - z(1 - \prod_{i \in J}(1-\delta_i)))$$

Clearly the focal set E_J is deleted in m' if and only if

$$z = \frac{\sum_{I \subseteq J} m(E_I)}{\sum_{I \subset J} m(E_I)} = \frac{Q(E_J)}{Q(E_J) - m(E_J)} = \frac{1}{1 - \prod_{i \in J}(1 - \delta_i)}.$$

□

Proposition 8. *Retracting a focal set E_J from a separable BPA m affects and may delete all focal sets $E_I \subset E_J$ as well, namely all combinations between the merging of information E_J from sources indexed in J, with information from other sources.*

Proof: Suppose $J \subset I$, so that $E_I \subset E_J$ then we have seen that

$$\begin{aligned} m'(E_I) &= zm(E_I) + (1-z)(\sum_{I \setminus J \subseteq K \subseteq I} m(E_K)) \\ &= z \prod_{i \in I}(1-\delta_i) \prod_{i \notin I} \delta_i + (1-z) \sum_{I \setminus J \subseteq K \subseteq I} \prod_{i \in K}(1-\delta_i) \prod_{i \notin K} \delta_i \\ &= \prod_{i \notin I} \delta_i \prod_{i \in I \setminus J}(1-\delta_i)(z \prod_{i \in J}(1-\delta_i) + (1-z)Z) \end{aligned}$$

where $Z = \sum_{I \setminus J \subseteq K \subseteq I} \prod_{i \in K \setminus J}(1-\delta_i) \prod_{i \in J \setminus K} \delta_i$. Letting $K = (I \setminus J) \cup L$ with $\emptyset \subseteq L \subseteq J$, it reads $Z = \sum_{L \subseteq J} \prod_{i \in L}(1-\delta_i) \prod_{i \in J \setminus L} \delta_i = 1$ again. So, $m'(E_I) = (\prod_{i \notin I} \delta_i) \prod_{i \in I \setminus J}(1-\delta_i)(1 - z + z \prod_{i \in J}(1-\delta_i)) = 0$ if and only if $z = 1/(1 - \prod_{i \in J}(1-\delta_i))$ again. In other words, when $E_I \subset E_J$, $m'(E_I)$ vanishes along with $m'(E_J)$. □

This result highlights the resemblance between the retraction operation and the contraction operation in the logical theory of belief change after Alchourron, Gärdenfors and Makinson (see the book written by the second author [19]). Given a set of closed set K of propositional formulas interpreted as a belief set, the contraction of K by a formula ϕ (whose set of models is a proposition) consists in forgetting $\phi \in K$, constructing a closed set $K \ominus \phi$ that no longer contains ϕ. It is not enough to delete ϕ because due to the closed nature of $K \ominus \phi$, if K contains ψ that implies ϕ then the closure of $K \setminus \{\phi\}$ still contains ϕ. So, the contraction by ϕ requires the deletion from K of all formulas that imply ϕ as well, just like the retraction of E_J from the focal sets also deletes focal subsets $E_I \subset E_J$ of it.

Proposition 9. *When retracting a focal set E_J from a separable BPA m, the focal sets $E_I \not\subset E_J$ retain a positive mass $m'(E_I) \leq 1$.*

Proof: When $J \not\subset I$, $m'(E_I) = zm(E_I)$, where $1 < z \leq 1/(1 - \prod_{i \in J}(1-\delta_i))$. Since $\sum_{E \subseteq \Omega} m'(E) = 1$, we can write it as $\sum_{E \subseteq \Omega} m'(E) = \sum_{E_I \subseteq E_J} m'(E_I) + \sum_{E_I \not\subseteq E_J} zm(E_I) = 1$. Clearly, $0 \leq m'(E_I) \leq 1$ in the first term due to the preceding propositions, and since $z > 1$, we must have $zm(E_I) < 1$ for $E_I \not\subseteq E_J$ to respect the normalisation condition. Hence $0 \leq m'(E_I) \leq 1$ when $E_I \not\subseteq E_J$. $\square$

Note that while fully retracting A_i leads to retracting its subsets made of conjunctions as well, it does not lead to retracting any set $A_k \in \mathcal{T}$ even when $A_k \subset A_i$. The idea is that questioning a piece of information A_i received from a source i does not affect information coming from other sources, even if they are fully coherent with A_i.

Example 5. *Consider again the case in Example 4 with three testimonies $E_i, i = 1, 2, 3$. Suppose $E_2 \subset E_1$, and we again retract E_1. The results in terms of diffidence values can be grouped noticing that $E_1 \cap E_2 = E_2$ and $E_1 \cap E_2 \cap E_3 = E_2 \cap E_3$. It yields*

- $m'(E_1) = (1 - zd_1)d_2d_3; m'(E_3) = zd_1d_2(1 - d_3)$
- $m'(E_2) = zd_1(1 - d_2)d_3 + (1 - d_2)d_3(1 - zd_1) = (1 - d_2)d_3$
- $m'(\Omega) = zd_1d_2d_3$
- $m'(E_1 \cap E_3) = (1 - zd_1)d_2(1 - d_3)$
- $m'(E_2 \cap E_3) = (1-d_2)(1-d_3)(1-zd_1)+zd_1(1-d_2)(1-d_3) = (1-d_2)(1-d_3)$

It is clear that when $z = 1/d_1$, the focal sets E_1 and $E_1 \cap E_3$ are deleted and $m' = E_2^{d_2} \circledcirc E_3^{d_3}$ is obtained, namely E_2 remains as a full-fledged source even if it is a more precise information than E_1.

4.3. Toward the prejudiced merging of non-dogmatic belief functions

The intuition behind retraction is that the agent possessing a reason of strength $y < 1$ not to believe E (modeled by $E^{1/y}$) is ready to doubt about the truth of E whenever receiving a testimony $E^d, d < 1$ claiming that E is true. More precisely, the degree of belief $1 - d$ in E based on evaluating the reliability of the source will be attenuated by this prior information, understood as a prejudice, to an extent that depends on the diffidence value $y < 1$. The lesser y, the stronger the reason not to believe in the truth of E. The condition $y = d$ is enough to fully erase E.

Example 6. *Consider a formal model of an informal example due to Smets [40, 16]. A newspaper reports that the economic situation in a region called Ukalvia is pretty good. You had never heard of Ukalvia and have no idea which newspaper is involved. So you have some reason (support $1 - d$), to think that the information is correct. It is an SSF denoted by G^d (G for good). Later on a friend you trust lets you know that Ukalvia lies in a totalitarian country and the newspaper is handled by the government of this country. You now have a prejudice against the truth of news published in this journal, and then you start to doubt about the good economic situation in Ukalvia. So you are led to* *downgrade the strength in your previous belief in G from d to d/y by retracting G^y, if $d \leq y < 1$. If $d > y$, the result $G^{d/y}$ is not a belief function, and can be interpreted, as Smets [40, 16] suggests, as a debt of belief that will attenuate the strength of further testimonies of the form G^d.*

In this paper, when $y > 1$, we call E^y a *prejudice against* E. It means a resistance to believing input information asserting $\omega^* \in E$. A prejudice may be due to the fact that we have a reason to believe the complementary statement $\omega^* \in \overline{E}$ (which would mean believing the complement $\overline{E}$ to extent $1/d < 1$). Indeed, if an agent possesses the prior entrenched belief (i.e. not based on recent testimonies) that E is false, it may be used as a prejudice against an input information claiming that E is true. Alternatively, a prejudice against E can be due to the fact that we have a reason to distrust the source of information claiming that E is true (which, in that case, does not presuppose any reason to believe its negation).

In view of the above results, Smets' decomposition of a BPA m as a conjunction of GSSF's comes down to considering that any non-dogmatic belief function comes from merging unreliable elementary pieces of information $\mathcal{T} = \{A_1, \ldots, A_n\}$ (forming the separable BPA m^+), followed by a retraction of statements in $\mathcal{P}$ (collected in the separable BPA m^-). Note that $B \in \mathcal{P}$ is of the form $\cap_{j \in J} A_j$ where $|J| > 1$. It results in weakening the support pertaining to the *conjunction* of information items coming from sources, thus eroding the strength of these fusion results. Following our intuition, the decomposition suggests that there are prejudices against conjunctions of statements $\omega^* \in A_j, j \in J$. The set of prejudices in such a decomposition is exactly

$$\mathcal{P} \subseteq \{\cap_{j \in J} A_j : J \subseteq [n], |J| > 1\} \setminus \mathcal{T}$$

since $\mathcal{P} \cap \mathcal{T} = \emptyset$ and some $\cap_{j \in J} A_j$ may be equal to some A_k. Indeed, $\mathcal{T}$ is made of focal sets of m that are not retracted in the decomposition; only other focal sets of m^+, i.e. the non-trivial intersections of the $A_i's$ are more or less retracted to form m. The idea is that when receiving information from outside sources, the merging operation performed by the receiver may also involve prior information that questions the validity of received uncertain testimonies. Such prior information is viewed as prejudice and does not play the same role as the input information.

It is indeed natural to consider that information we receive from the outside may be challenged by our prior beliefs. These prior beliefs may take the form of stereotypes, or prejudices that one is often unaware of. The receiver is for some reason reluctant to consider as credible the result of the conjunction of some reported propositions. For instance, consider a variant of the Linda problem [44].[8]

[8] In the original example, the bank teller Linda, depicted as a philanthropist, is found by participants to a psychological experiment, more likely to be a philanthropist bank teller than a bank teller, because the former looks more "representative" or typical of persons who might

Example 7. *We receive two testimonies, namely one (B with diffidence value $y < 1$) claiming that Linda is a banker and another one (A with diffidence value $x < 1$) that she is a philanthropist. The fusion process leads us to allocate a belief degree $(1-x)(1-y)$ to the fact that she is a philanthropist bank teller. However, a prejudiced individual would hardly believe that a bank teller can be philanthropist. In other words, we have prior information of the form $(\overline{A} \cup \overline{B})^u$. On this basis, we would like to erode, possibly delete, the focal set $A \cap B$ from the merged inputs by retracting the SSF $(A \cap B)^u$ from the result of the fusion of A^x with B^y. Thus we compute $m = (A^x \bigcirc\!\!\!\!\cap\, B^y) \oslash (A \cap B)^u$ which leads to a mass on $A \cap B$ equal to $1 - \frac{(x+y-xy)}{u}$, that is all the lesser as the prejudice is strong. When $u > x + y - xy$, the result is still a belief function. The focal set $A \cap B$ is deleted when $u = x + y - xy$. Note that if $u < x + y - xy$, the result is no longer a belief function. It is a "debt of belief", as Smets says, that will oppose any future new input of the form $A \cap B$ with an attenuated strength.*

		Comb. with negation	Revision [27]	Retraction
	$A^x \bigcirc\!\!\!\!\cap\, B^y$	$(A^x \bigcirc\!\!\!\!\cap\, B^y) \bigcirc\!\!\!\!\cap\, \overline{(A \cap B)}^u$	$(A^x \bigcirc\!\!\!\!\cap\, B^y) \star \overline{(A \cap B)}^u$	$(A^x \bigcirc\!\!\!\!\cap\, B^y) \oslash (A \cap B)^u$
$\emptyset$	0	$(1-x)(1-y)(1-u)$	0	0
A	$(1-x)y$	$(1-x)yu$	$(1-x)yu$	$(1-x)y/u$
B	$x(1-y)$	$x(1-y)u$	$x(1-y)u$	$x(1-y)/u$
$\overline{AB}$	0	$xy(1-u)$	$(1-x-y+2xy)(1-u)$	0
$A\overline{B}$	0	$(1-x)y(1-u)$	$(1-x)y(1-u)$	0
$\overline{A}B$	0	$x(1-y)(1-u)$	$x(1-y)(1-u)$	0
AB	$(1-x)(1-y)$	$(1-x)(1-y)(u)$	$(1-x)(1-y)u$	$1 - \frac{(x+y-xy)}{u}$
Ω	xy	xyu	xyu	xy/u

Table 3: Comparison of change rules in the Linda case of Example 7.

One might be tempted to consider that the retraction operation is redundant and that retracting a focal set could be simulated by combining the belief function with the negation of this focal set. This conjecture is not valid, as shown in Table 3 on the Linda case of Example 7. Namely, instead of retracting $(A \cap B)^u$ from $(A^x \bigcirc\!\!\!\!\cap\, B^y)$ (4th column in Table 3) suppose we symmetrically combine the latter with the prior information $\overline{(A \cap B)}^u$ (2d column in Table 3). We can see the result of combining with the negation of the retracted proposition is rather confusing since $A \cap B$ remains a focal set along with its complement, and the contradiction receives a positive weight. In contrast, retraction just deletes (or decreases the weight of) $A \cap B$ without adding any new focal set. Instead of this combination with the opposite proposition, we may revise $(A^x \bigcirc\!\!\!\!\cap\, B^y)$ by means of $\overline{(A \cap B)}^u$ using an asymmetric revision rule (denoted by $\star$) proposed in [27]. Revising m by m' consists, in the conjunctive rule, of replacing the intersection of any focal set E of m and any focal set F of m', by the revision rule $E \star F = F$ if $E \cap F = \emptyset$, and $E \cap F$ otherwise. In the Linda case, it comes down to replacing the positive mass of $\emptyset$ in column 2 by 0 and transferring it to $\overline{(A \cap B)}$ in line 4 (see the result of this change in column 3). Up to this change, the result of

fit the description of Linda. It illustrates the so-called representativeness heuristic [44] which consists in basing one's judgement on personalized rather than statistical information.

revision is similar to the one of combination; in particular, some mass remains attached to the same non-empty focal sets in both columns.[9]

More generally we can consider the retraction of a non-dogmatic belief function from another one. Consider two non-dogmatic belief functions m_1 and m_2. The general retraction operation is of the form $m_1 \oslash m_2$. Using the decomposition of m_1 and m_2 into latent belief structures (m_i^+, m_i^-), it is easy to express this retraction operation in terms of prejudiced fusion of SSF's, namely $m_1 \oslash m_2$ takes the form:

$$
\begin{aligned}
&((\circledcap_{A\in\mathcal{T}_1} A^{\delta_1^+(A)}) \circledcap (\circledcap_{B\in\mathcal{P}_1} B^{\frac{1}{\delta_1^-(B)}})) \oslash ((\circledcap_{A\in\mathcal{T}_2} A^{\delta_2^+(A)}) \circledcap (\circledcap_{B\in\mathcal{P}_2} B^{\frac{1}{\delta_2^-(B)}}))\\
=&((\circledcap_{A\in\mathcal{T}_1} A^{\delta_1^+(A)}) \circledcap (\circledcap_{B\in\mathcal{P}_2} B^{\delta_2^-(B)})) \oslash ((\circledcap_{B\in\mathcal{P}_1} B^{\delta_1^-(B)}) \circledcap (\circledcap_{A\in\mathcal{T}_2} A^{\delta_2^+(A)}))\\
=&m_{12}^+ \oslash m_{12}^-
\end{aligned}
$$

where m_{12}^+ and m_{12}^- are *separable* belief functions with respective testimony sets $\mathcal{T}_1 \cup \mathcal{P}_2$ and $\mathcal{T}_2 \cup \mathcal{P}_1$. Due to commutativity and associativity of $\circledcap$, the retraction of m_{12}^- from m_{12}^+ can proceed step by step by retracting SSFs focused on each element of $\mathcal{T}_2 \cup \mathcal{P}_1$ one by one in any order. The result will be a non-dogmatic belief function only if the conditions pointed out in Propositions 4, 5 and 6 are satisfied.

4.4. Additional comments

We conclude this section by some remarks concerning the Dempster rule of combination ~~and~~ the discounting method, in the light of the retraction operation. Moreover we briefly discuss the possibility of setting the mass of one focal set to a prescribed value using conjunctive combination with a generalised SSF.

The orthogonal combination rule as a retraction process. In [5], Denœux noticed that normalizing a subnormal BPA m using (2) amounts to combining it with a GSSF of the form $\emptyset^d$.

Proposition 10. *Consider a belief function $m_{\circledcap}$ obtained as the conjunctive combination of two non-dogmatic belief functions m_1 and m_2 such that $m_{\circledcap}(\emptyset) > 0$. Then their orthogonal sum $m_\oplus = m_1 \oplus m_2$ is of the form $m_{\circledcap} \oslash \emptyset^d$ where $d = 1 - m_{\circledcap}(\emptyset)$.*

Proof: Let us compute $m' = m_{\circledcap} \circledcap \emptyset^x$. It is clear that for any focal set A different from $\emptyset$, $m'(A) = x m_{\circledcap}(A)$, and $m'(\emptyset) = m_{\circledcap}(\emptyset)x + 1 - x$ since $\emptyset$ has mass $1 - x$ in $\emptyset^x$. The full retraction of $\emptyset$ from $m_{\circledcap}$ comes down to enforcing $m_{\circledcap}(\emptyset)x + 1 - x = 0$ that is, $x = 1/(1 - m_{\circledcap}(\emptyset))$. So $m' = \frac{m_{\circledcap}}{1 - m_{\circledcap}(\emptyset)} = m_\oplus$. $\square$

[9] An alternative combination would be to consider the disjunctive rule between $A^x \circledcup B^y$ and $\overline{(A \cap B)}^u$ [11]. However, the only focal set obtained is Ω, whatever the values of x, y, u, so that the result is always vacuous.

As the orthogonal rule of combination is a building block of evidence theory, this result shows that the retraction operation is already present from the start in the approach. It is a conjunctive fusion, with a prejudice against contradiction, leading to retract the latter and restore internal consistency of the resulting BPA. The renormalisation step in Dempster's rule of combination is a retraction operation.

Retraction versus Discounting. In the belief function framework, additional doubt about the reliability of a source of information is taken into account through the discounting operation [34], which transforms, in its simpler form, each belief function provided by a source into a weaker, less informative one. Namely, discounting a belief function m by a factor $\alpha \in [0, 1]$ reduces the mass $m(A)$ bearing on A and reassigns the remaining mass to Ω. It yields a BPA m_α such that:

$$m_\alpha(A) = \begin{cases} \alpha \cdot m(A), & \text{if } A \neq \Omega, \\ (1-\alpha) + \alpha \cdot m(\Omega), & \text{otherwise.} \end{cases}$$

In particular, an SSF A^d can be viewed as discounting a sure piece of information ($m(A) = 1$) by a factor $1 - d$. Discounting A^d by a factor α yields $m_\alpha(A) = A^{(1-\alpha)+\alpha.d}$. It is similar to retracting A from A^d as $A^d \bigcirc\!\!\!\!\cap\, A^x$ with $x > 1$, which yields A^{dx}. It is equal to $A^{(1-\alpha)+\alpha.d}$ provided that $dx = (1-\alpha) + \alpha.d$. Namely retracting A with strength $1 < x \leq 1/d$ comes down to discounting A^d with a factor $\alpha = \frac{1-dx}{1-d}$. However, this result does not extend to more general belief functions: while discounting affects all focal sets to the same extent, we can retract a focal set B from a separable BPA m, such that $m(B) > 0$, and delete it from the focal sets, while maintaining other focal sets and increasing their masses, as seen above. A comparison of retraction with contextual discounting [21, 28] could also be carried out.

Mass change via retraction. Consider a BPA m and suppose for some reason one must change $m(A)$ into $m'(A) = \alpha m(A)$ for some focal set A and some fixed factor $0 < \alpha < 1$. It is not clear how this change can be propagated to other masses $m(B), B \neq A$ in order to keep the normalization condition for m'. One way of proceeding is to assume the result m' of the change to be of the form $m \bigcirc\!\!\!\!\cap\, A^z$ with $z > 1$. As we have seen earlier, there is a risk for $m \bigcirc\!\!\!\!\cap\, A^z$ to have negative masses and this risk is mitigated if we assume that the focal sets of m are closed under intersection. In that case, $\forall E \in \mathcal{F}(m), A \cap E \in \mathcal{F}(m)$. We can compute the value of z such that $m \bigcirc\!\!\!\!\cap\, A^z = \alpha m(A)$. Using results in subsection 4.1, we must ensure the equality $zm(A) + (1 - z)Q(A) = \alpha m(A)$, which yields $z = \frac{Q(A)-\alpha m(A)}{Q(A)-m(A)}$, where we can see that $z > 1$ if and only if $\alpha < 1$ (which corresponds to entering a prejudice against A). We can update the other masses for $E \neq A$ as follows, using (10):

$$\forall E \in \mathcal{F}, m'(E) = \begin{cases} m(E) - \frac{(1-\alpha)m(A)}{Q(A)-m(A)} \sum_{\emptyset \neq D \subseteq \overline{A}} m(E \cup D) & \text{if } E \subset A, \\ \frac{Q(A)-\alpha m(A)}{Q(A)-m(A)} m(E) & \text{otherwise.} \end{cases}$$

But these formulas can be used only if the updated masses remain positive and less than 1. In fact, it is not true that $\forall \alpha \in [0, 1/m(A)], \exists z > 0$ such that $m' = m \textcircled{\cap} A^z$. In other words, modifying the value of a mass cannot always be expressed as the retraction of a focal set nor by the combination with an SSF.

The iterability of retraction. If we notice that retraction can be performed by a simple division of commonalities, it is obvious to conclude that retraction can be iterated. The only delicate point is to ensure that the result of retracting a belief function from another one remains a belief function. In subsection 4.2, we have devised conditions for ensuring this behavior in the case of retracting an SSB from a separable belief function. The generalization of these results to the retraction of any non-dogmatic belief function from another one is a matter of further research.

5. Informational orderings in agreement with diffidence functions

An important issue in the various uncertainty theories is how to compare uncertain pieces of information from the point of view of their informative content. In each framework a "more-informative-than" relation is at work, as shown for instance in [9]. This relation makes it possible to adopt a cautious attitude when representing information, namely, apply a least commitment principle which states that one should never presuppose more beliefs than justified.

In the case of belief functions there have been several proposals for such a notion of relative informativeness [45, 11, 5]. This fact is partly due to the existence of several interpretive settings for belief functions. In this paper, we shall focus on the relative informativeness that is in agreement with the interpretation of belief functions as the result of merging testimonies and prejudices. We consider relations of the form $m_1 \sqsubseteq_x m_2$ that intend to mean "m_1 is at least as informed as m_2". We can also read it as a form of entailment of m_2 from m_1 as, in most cases, it reduces to set-inclusion $A_1 \subseteq A_2$ when $m_i(A_i) = 1$ for $i = 1, 2$.

5.1. Various informational orderings

Formally, no less than seven definitions can be found in the literature:

1. *cf-ordering*[17]: $m_1 \sqsubseteq_{cf} m_2$ iff $cf_1(\{\omega\}) \leq cf_2(\{\omega\}), \forall \omega \in \Omega$ where $cf_i(\{\omega\}) = Pl_i(\{\omega\})$ are contour functions of m_i.
2. *pl-ordering* [11]: $m_1 \sqsubseteq_{pl} m_2$ iff $Pl_1(A) \leq Pl_2(A), \forall A \subseteq \Omega$,
3. *q-ordering* [11]: $m_1 \sqsubseteq_q m_2$ iff $Q_1(A) \leq Q_2(A), \forall A \subseteq \Omega$,
4. *s-ordering* [45]: $m_1 \sqsubseteq_s m_2$ iff there exists a stochastic matrix $S(A, B)$ where A is focal for m_1 and B is focal for m_2 such that $\sum_{A:A \subseteq B} S(A, B) = 1$ (so $S(A, B) = 0$ if $A \not\subseteq B$), and $m_1 = S \cdot m_2$ (short for $m_1(A) = \Sigma_{B:A \subseteq B} S(A, B) m_2(B)$). Then, m_1 is called a *specialization* of m_2.
5. *d-ordering* [24]: $m_1 \sqsubseteq_d m_2$ iff there exists a BPA m such that $m_1 = m \textcircled{\cap} m_2$. Then, m_1 is said to be a *Dempsterian specialization* of m_2

6. *dif-ordering* [5]: $m_1 \sqsubseteq_w m_2$ iff $\delta_1(A) \leq \delta_2(A), \forall A \subseteq \Omega$. We call *diffidence ordering.*[10]
7. *l-ordering* [32]: $m_2 \sqsubseteq_l m_1$ (l stands for latent) iff $m_1^+ \sqsubseteq_w m_2^+$ and $m_1^- \sqsubseteq_w m_2^-$, considering latent belief structures obtained by Smets canonical decomposition of m_1 as $L_1 = (m_1^+, m_1^-)$ and m_2 as $L_2 = (m_2^+, m_2^-)$.

These information orderings are more or less stringent. It is known that [11, 7, 5]:

$$m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \sqsubseteq_d m_2 \Rightarrow m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \Rightarrow m_1 \sqsubseteq_{cf} m_2, \\ m_1 \sqsubseteq_q m_2 \Rightarrow m_1 \sqsubseteq_{cf} m_2 \end{cases}$$

where all implications are strict. They make sense in some frameworks, not in other ones. For instance, the cf-ordering is a minimal generalisation of relative specificity in possibility theory [12], since the contour functions reduce to possibility distributions if the belief functions are consonant.

In the pl-ordering we may use belief functions by duality: $m_1 \sqsubseteq_{pl} m_2$ if and only if $Bel_1(A) \geq Bel_2(A), \forall A \subseteq \Omega$. The pl-ordering actually compares the credal sets induced by the BPAs. Namely, a BPA m_i can be viewed as encoding a convex set of probability functions $\mathcal{P}_i = \{P : P(A) \leq Pl_i(A), \forall A \subseteq \Omega\}$ provided that $m_i(\emptyset) = 0$. Then we also have the equivalence $m_1 \sqsubseteq_{pl} m_2$ if and only if $\mathcal{P}_1 \subseteq \mathcal{P}_2$. This view requires to see belief functions as encoding for instance, frequencies of ill-observed data, that would have given standard probability distributions had the observations been precise. It corresponds to Dempster upper and lower probabilities induced by a set-valued random variable [4], where the sets observed have an epistemic flavor (incomplete pieces of information).

These five information orderings reduce to standard set-inclusion when the belief functions have each a single focal set with mass 1. The cf-ordering corresponds to fuzzy set inclusion. The specialization ordering is a direct extension of set inclusion to random sets [11], where the stochastic matrix induces a joint BPA $m(A, B) = S(A, B)m_1(A)$ bearing on Cartesian products of focal sets of m_1 and m_2, and whose marginals are m_1 and m_2. This information ordering, like the pl-ordering, is completely independent of the orthogonal or conjunctive rules of combination. In contrast, the d-ordering add constraints on the joint mass function, forcing m_1 to be the result of merging m_2 with another belief function. In the case of comparing categorical belief functions focused on A and B, the d-ordering becomes $\exists C : A \cap C = B$ to express the inclusion $B \subseteq A$.

5.2. Informational orderings and the conjunctive rule of combination

There are four informational orderings that have connections with the orthogonal and conjunctive rules of combination: the q-ordering, the d-ordering, the dif-ordering and the l-ordering. In the case of the q-ordering, the relation

[10]We keep the original notation $\sqsubseteq_w$ of Denœux who denotes diffidence functions by w. It should be called diffidence ordering, though.

with the conjunctive rule is because $Q_1(A) \leq Q_2(A), \forall A \subseteq \Omega$ if and only if there exists a function $Q : 2^\Omega \to [0,1]$ such that $Q_1(A) = Q(A) \cdot Q_2(A) \forall A \subseteq \Omega$. However, Q is generally not a commonality function. When it is, then it corresponds to a BPA m such that $m_1 = m \textcircled{\cap} m_2$, that is, we recover the Dempsterian specialisation $m_1 \sqsubseteq_d m_2$ and the fact that the latter is stronger than the q-ordering [24].

Likewise in the case of the dif-ordering, the relation with the conjunctive rule is because $\delta_1(A) \leq \delta_2(A), \forall A \subseteq \Omega$ if and only if there exists a function $w : 2^\Omega \to [0,1]$ such that $\delta_1(A) = \delta(A) \cdot \delta_2(A) \forall A \subseteq \Omega$. However, as by construction, $\delta(A) \leq 1, \forall A \subseteq \Omega$, δ is the diffidence function of a separable BPA m and $\delta_1 = \delta \cdot \delta_2$ is the same as $m_1 = m \textcircled{\cap} m_2$. Then the dif-ordering corresponds to a special case of d-ordering where the BPA m is separable, that is, we recover the implication $m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \sqsubseteq_d m_2$ [5]. These two orderings collapse on separable belief functions. Indeed, if $m_2 \textcircled{\cap} m = m_1$ and m_1, m_2 are separable, then m should be separable as well; for if not, we would have $m_2 \textcircled{\cap} m$ not separable.

In order to be meaningful in the setting of belief functions resulting from the merging of testimonies and prejudices, the information ordering must be compatible with the conjunctive rule of combination. We can define this compatibility as follows.

Definition 1. *An information ordering $\sqsubset_x$ is compatible with a combination rule $\odot$ if and only if $m_1 \sqsubseteq_x m_2$ and $m_3 \sqsubseteq_x m_4$ imply $m_1 \odot m_3 \sqsubseteq_x m_2 \odot m_4$.*

Proposition 11. *The information orderings $\sqsubset_x, x = cf, q, s, d, w$ are compatible with the conjunctive rule $\textcircled{\cap}$.*

Proof: Suppose $m_1 \sqsubseteq_q m_2$ and $m_3 \sqsubseteq_q m_4$. It means $Q_1 \leq Q_2$ and $Q_3 \leq Q_4$, and by the conjunctive rule $Q_{m_1 \textcircled{\cap} m_3} = Q_1 Q_3 \leq Q_{m_2 \textcircled{\cap} m_4} = Q_2 Q_4$. The same holds for the cf-ordering as the contour function of m is $cf(\omega) = Q(\{\omega\})$.

Suppose $m_1 \sqsubseteq_d m_2$ and $m_3 \sqsubseteq_d m_4$. It means $m_1 = m_2 \textcircled{\cap} m_2'$ and $m_3 = m_4 \textcircled{\cap} m_4'$, so (by associativity) $m_1 \textcircled{\cap} m_3 = m_2 \textcircled{\cap} m_4 \textcircled{\cap} (m_2' \textcircled{\cap} m_4')$. Hence $m_1 \textcircled{\cap} m_3 \sqsubseteq_d m_2 \textcircled{\cap} m_4$. For the dif-ordering the proof is similar.

For specialization entailment, suppose $m_1 = S \cdot m_2$ and $m_3 = S' \cdot m_4$. Let $A_i, B_i, \ldots$ denote the focal sets of m_i. Consider the coefficient $S(A_1, A_2)$ of S. It is the portion of mass $m(A_2)$ allocated to A_1 where $A_1 \subseteq A_2$. According to the conjunctive rule, the mass $m_1(A_1)m_3(A_3)$ is allocated to set $A_1 \cap A_3$, and the mass $m_2(A_2)m_4(A_4)$ is allocated to set $A_2 \cap A_4$. Since $A_1 \cap A_3 \subseteq A_2 \cap A_4$ we can define another stochastic matrix $S_{\textcircled{\cap}}$ such that $S_{\textcircled{\cap}}(A_1 \cap A_3, A_2 \cap A_4) = S(A_1, A_2)S'(A_3, A_4)$. Since by assumption $m_1(A_1) = \sum_{A_2 : A_1 \subseteq A_2} S(A_1, A_2)m_2(A_2)$ and $m_3(A_3) = \sum_{A_4 : A_3 \subseteq A_4} S'(A_3, A_4)m_4(A_4)$, it follows that

$$m_1(A_1)m_3(A_3) = \sum_{A_1 \subseteq A_2, A_3 \subseteq A_4} S(A_1, A_2)S'(A_3, A_4)m_2(A_2)m_4(A_4).$$

As $\sum_{A_1 : A_1 \subseteq A_2} S(A_1, A_2) = 1$ and $\sum_{A_3 : A_3 \subseteq A_4} S(A_3, A_4) = 1$, it implies that $\sum_{A_1, A_3 : A_1 \subseteq A_2; A_3 \subseteq A_4} S(A_1, A_2)S'(A_3, A_4) = 1$ as well. Note that the fact that there may be sets A_1, A_2, B_1, B_2 such that $A_1 \cap A_3 = B_1 \cap B_3$ is innocuous as we can add the coefficients $S_{\textcircled{\cap}}(A_1 \cap A_3, A_2 \cap A_4)$ and $S_{\textcircled{\cap}}(B_1 \cap B_3, A_2 \cap A_4)$, if we consider the focal set $C_{13} = A_1 \cap A_3 = B_1 \cap B_3$ of $m_1 \textcircled{\cap} m_3$. In other words, by adding lines of the matrix $S_{\textcircled{\cap}}$ pertaining to the same focal set of $m_1 \textcircled{\cap} m_3$, and then adding columns of terms pertaining to the same focal set of $m_2 \textcircled{\cap} m_4$, we get a reduced stochastic matrix $S^{''}$ such that $m_1 \textcircled{\cap} m_3 = S^{''} \cdot m_2 \textcircled{\cap} m_4$. $\square$

However, it is not clear whether the plausibility ordering $\sqsubseteq_{pl}$ is compatible with the conjunctive rule or not.

Another form of compatibility between the conjunctive combination rule and an information ordering can be defined by requiring that the result of the conjunctive combination $m_1 \textcircled{\cap} m_2$ is more informed than any of m_1 and m_2. It is easy to see that $m_1 \textcircled{\cap} m_2 \sqsubseteq_d m_i, i = 1, 2$, by construction. As a consequence, $m_1 \textcircled{\cap} m_2 \sqsubseteq_x m_i, i = 1, 2$ for $x = cf, pl, q, s$, as well.

5.3. Notes on the diffidence ordering

Note that only the d-ordering and the dif-ordering explicitly involve the conjunctive combination rule: $m_1 \sqsubseteq_x m_2, x = d, w$ means that m_1 results from combining m_1 with another source. Recall that these two orderings collapse on separable belief functions.

Consider two SSF's $A_1^{\delta_1}$ and $A_2^{\delta_2}$. It is natural to consider that $A_1^{\delta_1}$ is at least as informed as $A_2^{\delta_2}$ whenever $A_1 \subset A_2$ and $\delta_1 < \delta_2$: the statement A_1 is more precise and has less diffidence. And indeed we have that $A_1^{\delta_1} \sqsubset_{cf} A_2^{\delta_2}$, and also for $x = pl, q, s$, since SSFs induce possibility measures and these orderings collapse in the consonant case [11].

Yet, we *do not* have that $A_1^{\delta_1} \sqsubseteq_x A_2^{\delta_2}$, for $x = d, w$, even if $A_1 \subset A_2$ and $\delta_1 < \delta_2$. This is because the result of combining two SSF's with different focal sets yields a belief function that has at least two focal sets different from Ω (i.e., finding B^δ such that $A_1^{\delta_1} = B^\delta \textcircled{\cap} A_2^{\delta_2}$ is impossible as the result of the merging will have more than one focal set different from Ω). However $A^{\delta_1} \sqsubset_w A^{\delta_2}$ is true if $A_1 = A_2 = A$. However, it implies that SSF's can seldom be compared by these ordering relations. These remarks lead to highlight the meaning of relation $\sqsubset_w$.

Consider the case of separable belief functions $m_i = \textcircled{\cap}_{A_i \in \mathcal{T}_i} A_i^{\delta_i(A)}$ where $\delta_i(A) < 1$ if $A \in \mathcal{T}_i$, i.e., the prejudice sets $\mathcal{P}_i$ are empty. It is then clear that $m_1 \sqsubseteq_w m_2$ is equivalent to $\mathcal{T}_2 \subseteq \mathcal{T}_1$ and $\delta_1(A) \leq \delta_2(A), \forall A \in \mathcal{T}_1$. In words, the BPA m_2 results from merging a subset of the pieces of information whose merging yields m_1, and the former grants less confidence in these pieces of information than m_1. For instance, if $A_1 \neq A_2$, we never have that $A_1^{\delta_1} \sqsubset_w A_2^{\delta_2}$ even if $A_1 \subset A_2$ and $\delta_1 < \delta_2$, because $\mathcal{T}_1 = \{A_1\}$ and $\mathcal{T}_2 = \{A_2\}$ so, $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$.

One can figure out how stringent the relation $m_1 \sqsubseteq_w m_2$ is: it means that m_1 is based on the same pieces of information as m_2 (the former with more confidence), plus other ones. In particular, if m_2 is obtained from m_1 by enlarging the focal sets of the latter, we shall not have $m_1 \sqsubseteq_w m_2$ because then $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$ again, while obviously $m_1 \sqsubseteq_s m_2$. Since $A^\delta \not\sqsubseteq_w B^\delta$ when $A \subset B$, letting δ vanish yields categorical belief functions focusing on A and B but at the limit we do not have that $A \sqsubseteq_w B$: since their (extended) diffidence functions are $\delta_A(A) = 0, \delta_A(E) = 1$ otherwise for $E \neq A \subset \Omega$ and $\delta_B(B) = 0, \delta_B(E) = 1$ otherwise for $E \neq B \subset \Omega$, we do not have that $\delta_A \leq \delta_B$ at the limit. It may be seen as anomalous that the diffidence ordering does not generalize set inclusion. But this is because this ordering insists on having the inclusion of *sets* of propositions ($\mathcal{T}_1 = \{A\} \not\subset \mathcal{T}_2 = \{B\}$), even in the categorical case.

However, there is a tiny difference between $\sqsubseteq_w$ and $\sqsubseteq_d$ on separable belief functions: the former is only defined in the non-dogmatic case (so the above anomaly of the diffidence ordering is only obtained as a limit process, assuming continuity at the limit), while the latter still makes full sense for dogmatic (in particular categorical) belief functions. Since $A \subseteq B$ is equivalent to $A = B \cap C$ for some set C, and the conjunctive rule extends set-intersection and applies to dogmatic belief functions, $A \sqsubseteq_d B$ holds.

Consider now the canonical decomposition of a non dogmatic BPA m_i into its latent belief structure $L_i = (m_i^+, m_i^-)$. The diffidence function of m_i is of the form:

$$\delta_i(A) = \begin{cases} \delta_i^+(A) < 1 \text{ if } A \in \mathcal{T}_i \\ 1/\delta_i^-(A) > 1 \text{ if } A \in \mathcal{P}_i \\ 1 \text{ otherwise.} \end{cases}$$

In this general case, $m_1 \sqsubseteq_w m_2$ expresses that not only $\mathcal{T}_2 \subseteq \mathcal{T}_1$ and $\delta_1^+(A) \leq \delta_2^+(A), \forall A \in \mathcal{T}_2$, but also $\mathcal{P}_1 \subseteq \mathcal{P}_2$ and $\delta_1^-(A) \geq \delta_2^-(A), \forall A \in \mathcal{P}_1$. Namely, m_1 results from more testimonies and less prejudices than m_2, and prejudices common to m_1 and m_2 are less strong for m_1. In m_2 some sources of information may be missing and some focal sets resulting from combining the remaining ones may be deleted by prejudices.

In the same vein, it is easy to see that generally, $m_1 \circledcirc m_2 \not\sqsubseteq_w m_i, i = 1, 2$ except if m_1 and m_2 are separable. This is because in the general case of non-dogmatic belief functions, the diffidence function $\delta_{12} = \delta_1 \cdot \delta_2$ of $m_1 \circledcirc m_2$ does not satisfy the inequality $\forall A \subset \Omega, \delta_{12}(A) \leq \delta_1(A)$ if m_2 is not separable. This inequality holds only for sets A such that $\delta_2(A) \leq 1$. These results highlight the limited compatibility between the diffidence ordering and the conjunctive rule of combination for non-separable belief functions.

5.4. Notes on the latent ordering

Finally, the l-ordering [32] explicitly relies on latent belief structures of m_1 and m_2 namely, $L_1 = (m_1^+, m_1^-)$ and $L_2 = (m_2^+, m_2^-)$. $m_1 \sqsubseteq_l m_2$ means that the positive part of m_1 is more informed than the positive part of m_2 and likewise for the prejudice parts.

Notice that we do not have that $m_1 \sqsubseteq_l m_2$ implies $m_1 \sqsubseteq_d m_2$. Indeed, suppose that $m_1^+ \sqsubseteq_w m_2^+$ and $m_1^- \sqsubseteq_w m_2^-$. As $m_1^+, m_1^-, m_2^+, m_2^-$ are separable, this is equivalent to $m_1^+ \sqsubseteq_d m_2^+$ and $m_1^- \sqsubseteq_d m_2^-$. More explicitly, $\exists m^+, m^-$ separable, such that $m_1^+ = m^+ \circledcirc m_2^+$ and $m_1^- = m^- \circledcirc m_2^-$. Hence

$$m_1 = m_1^+ \oslash m_1^- = (m^+ \cap m_2^+) \oslash (m^- \cap m_2^-) = (m^+ \oslash m^-) \cap (m_2^+ \oslash m_2^-).$$

So $m_1 = m \circledcirc m_2$, but there is no reason why m should be a regular BPA, since m^+ and m^- are unrelated.

Of course, for separable belief functions the latent, diffidence and Dempsterian specialisation orderings coincide trivially. But generally, the two information orderings $\sqsubseteq_l$ and $\sqsubseteq_w$ are at odds for non-separable functions. Indeed, when comparing m_1 and m_2, the inequalities for the negative parts of the latent belief structures have opposite directions for $\sqsubseteq_l$ and $\sqsubseteq_w$, respectively $\delta_1^- \leq \delta_2^-$ and $1/\delta_1^- \leq 1/\delta_2^-$.

At the interpretation level, $m_1 \sqsubset_l m_2$ means that m_1 results from more testimonies (like the diffidence ordering) and *more* prejudices (unlike the diffidence ordering) than m_2, and prejudices common to m_1 and m_2 are stronger for m_1. In particular, when $m_1^+ = m_2^+$ we can have that $m_1 \sqsubset_l m_2$ while $m_2 \sqsubset_w m_1$.

Example 8. *Consider the belief function of the form $A^{\delta_A} \circledcirc B^{\delta_B} \circledcirc (A \cap B)^x$.*

- *When $x = 1$, it is a decomposable belief function m_1 such that $m_1(A) = (1-\delta_A)\delta_B, m_1(B) = (1-\delta_B)\delta_A, m_1(A \cap B) = (1-\delta_A)(1-\delta_B), m_1(\Omega) = \delta_A\delta_B$.*
- *When $x = 1/(\delta_A + \delta_B - \delta_A\delta_B) > 1$, it is a non-dogmatic belief function m_2 such that $m_2(A) = \frac{(1-\delta_A)\delta_B}{\delta_A+\delta_B-\delta_A\delta_B}, m_2(B) = \frac{(1-\delta_B)\delta_A}{\delta_A+\delta_B-\delta_A\delta_B}, m_2(A \cap B) = 0, m_2(\Omega) = \frac{\delta_A\delta_B}{\delta_A+\delta_B-\delta_A\delta_B}$*

It is easy to see that $\delta_1 < \delta_2$, since $\delta_1(E) = \delta_2(E)$, except that $\delta_1(A \cap B) = 1 > \delta_2(A \cap B) = 1/(\delta_A + \delta_B - \delta_A\delta_B)$, so $m_1 \sqsubset_w m_2$. However $m_2 \sqsubset_l m_1$ since $m_1^+ = m_2^+$ but $m_2^- = m_1^-$, i.e., $1/x = \delta_A + \delta_B - \delta_A\delta_B < 1$. It is clear that m_1 is more informative than m_2 (in the sense of the six information orderings other than $\sqsubseteq_l$).

In fact, the stronger (i.e. informative) are the prejudices the less informative the resulting belief function after retraction of these prejudices, which is not the intuition that rules the relation $\sqsubseteq_l$.

Example 9. [11] *Consider $L_1 = (m_1^+, m_1^-)$ and $L_2 = (m_2^+, m_2^-)$ such that $m_1^+ = m_1^- = m_\Omega$ and $m_2^+ = m_2^- = A^w \circledcirc B^v$. Hence, $L_2 \sqsubseteq_l L_1$ but the BPA's $m_i = m_i^+ \oslash m_i^-, i = 1,2$ are indistinguishable with respect to any of the x-orderings, since we have $m_1 = m_2 = m_\Omega$.*

Furthermore, consider $L_3 = (A^{0.4}, A^{0.5})$ and $L_4 = (A^{0.2}, A^{0.2})$. Here, $L_4 \sqsubset_l L_3$, however $m_3 \sqsubset_x m_4$ as $A^{0.8} \sqsubset_x m_\Omega$.

As a conclusion, the l-ordering compares the informative contents of latent belief structures, but not the informative contents of resulting belief functions. Consequently, the l-ordering $\sqsubseteq_l$ conveys a specific meaning distinct from the one captured by the other six x-orderings $\sqsubseteq_x$.

[11] Suggested by a reviewer.

5.5. Diffidence-based combinations in evidence theory

In the framework of fusion, the conjunctive rule is justified when the sources of information are independent. However this hypothesis is often unrealistic and combination rules dealing with non-independent information sources are necessary. In general, idempotent rules can be used as a cautious approach when dependencies between sources are ill-known. But it is not easy to find simple idempotent combination rules in evidence theory (see the discussion in [7]). In the context of belief functions described by diffidence functions, Denœux [5] suggests an idempotent combination rule relying on the diffidence ordering. The result of combining m_1 and m_2 is the least informative BPA m such that $m \sqsubseteq_w m_1$ and $m \sqsubseteq_w m_2$. The result, we denote by $m_1 \owedge m_2$, is obviously the BPA m with diffidence function $\delta = \min(\delta_1, \delta_2)$. In contrast, the conjunctive rule of combination performs the product instead of the minimum. Note that since δ_1 and δ_2 are diffidence functions issued from standard BPA's, so is $\delta = \min(\delta_1, \delta_2)$, since there exist separable belief functions m'_1 and m'_2 such that $m_1 \owedge m_2 = m'_1 \ocap m_1$ and $m_1 \owedge m_2 = m'_2 \ocap m_2$.

Note that it is difficult to define a combination rule using the same approach based on other information comparisons, i.e., there is not often one least informative belief function such that $m \sqsubseteq_x m_1$ and $m \sqsubseteq_x m_2$, for x = s, pl, q, cf. In [7], the case of x = cf is studied in detail, as it comes down to finding an idempotent combination rule in agreement with the minimum rule of possibility theory. It is shown that in the non-consonant case, the question has sometimes several answers, sometimes no answer. In the following we compare the idempotent combination of diffidence functions and the conjunctive rule rule as well as the minimum rule of possibility theory.

First we can show that there are cases when the two combinations $\owedge$ and $\ocap$ coincide:

Proposition 12. *Consider two separable BPA's m_1 and m_2 with respective testimony sets $\mathcal{T}_i, i = 1, 2$. Then $m_1 \owedge m_2 = m_1 \ocap m_2$ if and only if $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$*

Proof: First, $\forall A \subset \Omega, \delta_i(A) \leq 1, i = 1, 2$ due to separability. If $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$, it means that $\forall A \subset \Omega, \delta_1(A) = 1$ or $\delta_2(A) = 1$; it implies that $\min(\delta_1, \delta_2) = \delta_1 \cdot \delta_2$. The converse is obvious since $\min(x, y) = xy$ on the positive real line only if $\max(x, y) = 1$. □

This proposition sheds light on the meaning of the idempotent diffidence merging rule on separable belief functions. It assumes that sources delivering the same information are dependent, while sources delivering distinct pieces of information are independent.[12] It explains that $m_1 \owedge m_1 = m_1$, while the conjunctive rule is recovered with disjoint testimony sets.

On the contrary, if $\mathcal{T}_1 = \mathcal{T}_2$ while $\delta_1 \neq \delta_2$, the two BPA's stem from the same testimonies with varying strengths, and then $m_1 \owedge m_2 \neq m_1 \ocap m_2$. In the

[12] Note that, in the particular case of simple support functions this result was highlighted by Pichon in [31, Remark 4]).

general case when $\mathcal{T}_1 \cap \mathcal{T}_2 \neq \emptyset$, under the separability assumption, let $m_{1\setminus 2} = \circledcirc_{A \in \mathcal{T}_1 \setminus \mathcal{T}_2} A^{\delta_1(A)}$, $m_{2\setminus 1} = \circledcirc_{A \in \mathcal{T}_2 \setminus \mathcal{T}_1} A^{\delta_2(A)}$ and $m_{12} = \circledcirc_{A \in \mathcal{T}_1 \cap \mathcal{T}_2} A^{\min(\delta_1(A), \delta_2(A))}$ (the idempotent part of the combination), we have that

$$m_1 \owedge m_2 = m_{1\setminus 2} \circledcirc m_{12} \circledcirc m_{2\setminus 1}. \tag{12}$$

These results does not extend to non-dogmatic belief functions that have non-empty sets of prejudices $\mathcal{P}_i, i = 1, 2$ for which $\delta_i(A) > 1$. However we have the following result:

Proposition 13. *Consider two non-dogmatic BPA's m_1 and m_2 with respective prejudice sets $\mathcal{P}_i, i = 1, 2$. Then if $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$, the resulting BPA $m_1 \owedge m_2$ is separable.*

Proof: As $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$, it means that $\forall A \in \mathcal{P}_1 \cup \mathcal{P}_2, \delta_1(A) \leq 1$ or $\delta_2(A) \leq 1$. As $\delta_i(A) > 1, \forall A \in \mathcal{P}_i, i = 1, 2$ it means that $\min(\delta_1(A), \delta_2(A)) \leq 1, \forall A \in \mathcal{P}_1 \cup \mathcal{P}_2$. And obviously, $\forall A \notin \mathcal{P}_1 \cup \mathcal{P}_2, \delta_i(A) \leq 1$, so $\min(\delta_1(A) \leq 1, \delta_2(A)) \leq 1$ as well. Hence $m_1 \owedge m_2$ is separable and its set of testimonies is $\mathcal{T}_1 \cup \mathcal{T}_2$. $\square$

Combining the two propositions is then clear that if both $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$ and $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$, then $m_1 \owedge m_2 = m_1^+ \circledcirc m_2^+$. And in the non-separable case we still have that

$$m_1 \owedge m_2 = m_{1\setminus 2} \circledcirc m_{12} \circledcirc m_{2\setminus 1}.$$

where $m_{1\setminus 2}$ and $m_{1\setminus 2}$ are still defined as above and are separable, and now $m_{12} = \circledcirc_{A \in (\mathcal{T}_1 \cap \mathcal{T}_2) \cup (\mathcal{P}_1 \cap \mathcal{P}_2)} A^{\min(\delta_1(A), \delta_2(A))}$ (the idempotent part) is not separable.

6. Possibility theory as the merging of consonant information

In the consonant (non-dogmatic) case, the set of focal sets are nested, and the plausibility measure is a possibility measure. Then, it is possible to directly express the plausibility of singletons in terms of the diffidence function. We try to interpret diffidence ordering in the case of possibility theory, and compare the idempotent rule of combination of possibility theory with the diffidence-based idempotent combination rule.

6.1. Possibilistic BPA's

A possibilistic BPA has nested focal sets of the form $\{E_1 \subset E_2 \subset \cdots \subset E_k\}$, plus $E_{k+1} = \Omega$. Its contour function, called possibility distribution, denoted by π, is of the form

$$\pi(\omega) = \sum_{\omega \in E_i} m(E_i) = \sum_{i=j}^{k+1} m(E_i) \text{ if } \omega \in E_j \setminus E_{j-1}, i = 2, \ldots k+1$$

and $\pi(\omega) = 1$ if $\omega \in E_1$. Moreover $Pl(A) = \max_{\omega \in A} \pi(\omega)$ (called a possibility measure) so that it is maxitive [34, 12], i.e.,

$$Pl(A \cup B) = \max(Pl(A), Pl(B)) \text{ and } Bel(A \cap B) = \min(Bel(A), Bel(B)).$$

In the consonant case, the BPA can be recovered from the contour function. Let $\pi^i = \pi(\omega)$ if $\omega \in E_j \setminus E_{j-1}$ where $\pi^1 \geq \pi^2 \geq \dots \pi^k > \pi^{k+1} > 0$. Then we have that [12]:

$$m(E_i) = \pi^i - \pi^{i+1}, i = 1, \dots, k$$

and $m(\Omega) = \pi^{k+1}$. In other words all the information contained in the BPA is contained in the contour function. In particular the $\alpha-$*cuts* of π, say $E_\alpha = \{\omega \in \Omega : \pi(\omega) \geq \alpha\}$ for $0 < \alpha \leq 1$ coincide with the focal sets of the corresponding BPA.

In terms of information ordering, it holds that for consonant belief functions, we have that $m_1 \sqsubseteq_s m_2$ if and only if $m_1 \sqsubseteq_{cf} m_2$, i.e. $\pi_1 \leq \pi_2$ (specificity ordering): the four loosest information orderings coincide. As we show in the sequel this is not the case for the diffidence ordering.

6.2. From possibility distributions to diffidence functions and back

A consonant belief function is separable, and comes down to the merging of consonant pieces of information of the form $E_i^{\delta_i}$ with $E_1 \subset E_2 \subset \cdots \subset E_k$ and $\delta_i \leq 1$. Due to consonance, it is easy to see [5] that:

Proposition 14. *The BPA m_k induced by merging consonant sets of the form $E_i^{\delta_i}, i = 1 \dots k$ is such that $m_k(E_1) = 1 - \delta_1, m_k(E_i) = (1 - \delta_i)\prod_{j=1}^{i-1}\delta_j$ for $i = 2, \dots k$, and $m_k(\Omega) = \prod_{i=1}^{k}\delta_i$. Moreover the contour function of m_k is the possibility distribution such that if $\omega \in E_i \setminus E_{i-1}, i > 1, \pi_k(\omega) = \prod_{j=1}^{i-1}\delta_j$.*

Proof: Suppose $k = 2$, it is easy to see that $E_1^{\delta_1} \circledcirc E_2^{\delta_2}$ has a BPA such that $m_2(E_1) = 1 - \delta_1, m_2(E_2) = \delta_1(1 - \delta_2), m(\Omega) = \delta_1\delta_2$. Now suppose the result holds till $k - 1$, and compute $(\circledcirc_{i=1}^{k-1} E_i^{\delta_i}) \circledcirc E_k^{\delta_k}$ where E_k is a superset of all E_i's, $i < k$. As $E_i \subset E_k, \forall i < k$ the resulting mass of E_i is $m_k(E_i) = m_{k-1}(E_i)(1-\delta_k) + m_{k-1}(E_i)\delta_k$ (corresponding to $E_i \cap E_k$ and $E_i \cap \Omega$ respectively), that is $m_k(E_i) = (1 - \delta_i)\prod_{j=1}^{i-1}\delta_j = m_{k-1}(E_i)$ (it remains the same for $i < k$). Besides $m_k(E_k) = m_{k-1}(\Omega)(1 - \delta_k) = (1 - \delta_k)\prod_{j=1}^{k-1}\delta_j$, and $m_k(\Omega) = m_{k-1}(\Omega)\delta_k = \prod_{i=1}^{k}\delta_i$. Besides it is clear that if $\omega \in \Omega \setminus E_k, \pi(\omega) = m_k(\Omega) = \prod_{i=1}^{k}\delta_i$. If $\omega \in E_k \setminus E_{k-1}, \pi_k(\omega) = m_k(\Omega) + m_k(E_k) = \prod_{i=1}^{k-1}\delta_i$, and more generally it is obvious that $\pi_k^i = \delta_i \pi_k^{i-1}$. □

Conversely given the possibility degrees $\pi^i, i = 1, \dots k+1$ where $\pi^1 = 1 > \pi^2 > \dots \pi^{k+1} > 0$ defined as above, the diffidence function of the corresponding necessity measure can be immediately recomputed as

$$\delta(A) = \begin{cases} \dfrac{\pi^{i+1}}{\pi^i} \text{ if } A = E_i, i = 1, \dots, k \\ 0 \text{ otherwise.} \end{cases}$$

This result is already hinted in Shafer's book [34] using weights of evidence.

6.3. Diffidence ordering between possibility distributions

At this point it is interesting to express the diffidence ordering between two possibility distributions π_1 and π_2. Suppose π_1 and π_2 have nested focal sets $\mathcal{F}_1$, and $\mathcal{F}_2$. It is clear that

$$\pi_1 \sqsubseteq_w \pi_2 \text{ if and only if } \mathcal{F}_2 \subseteq \mathcal{F}_1 \text{ and } \frac{\pi_1^{i+1}}{\pi_1^i} \leq \frac{\pi_2^{i+1}}{\pi_2^i}, \forall E_i \in \mathcal{F}_2.$$

In the above definition, note that if E_i is focal for π_1 but not for π_2, it is clear that if $\omega \in E_i \setminus E_{i-1}$ and $\omega' \in E_{i+1} \setminus E_i$, we have that $\pi_2(\omega) = \pi_2(\omega')$ so that $\pi_2^{i+1} = \pi_2^i$ in the above equivalence. So not only the cuts of π_2 must be among the cuts of π_1, but there should be some inequalities between possibility ratios. These inequalities are equivalent to $\frac{\pi_1^i}{\pi_1^j} \leq \frac{\pi_2^i}{\pi_2^j}$ for all $j \leq i$.

This information comparison may seem counter-intuitive in the usual contexts of possibility theory. For instance, if we consider possibility measures as upper probability bounds [14, 3] it is clear that the usual specificity relation ($\pi_1 \leq \pi_2$ pointwise), i.e. the cf-ordering, is the most natural one, as it corresponds to an inclusion of the corresponding convex probability sets. However the diffidence ordering seems to make sense for likelihood functions.

It has been known for a long time that possibility distributions can be viewed as likelihood functions. This is first explained in the book [34], chapter 11, where Shafer assumes that likelihood functions can be viewed as proportional to contour functions of consonant belief functions, obtained by renormalizing the likelihood function so that its maximal value is 1. The same message was emphasized by Smets [39]. Later on, it was explained in [10] that the function $\lambda(\omega) = P(\mathcal{D}|\omega)$, where $\mathcal{D}$ is a data set, and P is a probabilistic model based on parameter ω for the data set, behaves like a possibility distribution because (i) for any subset $A \subset \Omega$, $\lambda(A) = P(\mathcal{D}|A)$ is a weighted average of $P(\mathcal{D}|\omega), \omega \in A$ if prior probabilities are known, hence upper bounded by $\max\{\lambda(\omega) : \omega \in A\}$ (ii) $\lambda(A)$ should be monotonic with inclusion (it represents the likelihood of $\omega \in A$). Hence $\lambda(A) = \max\{\lambda(\omega) : \omega \in A\}$. Other authors proposed justifications of this proposal such as Coletti and Scozzafava [2], and more recently Denœux [6].

It is tempting (and it has been done indeed) to compare likelihood functions pointwisely using the cf-ordering. However, likelihood theory (Edwards [18]) specifies that we can only compare *likelihood ratios* relative to results of distinct data sets $\mathcal{D}_1$ and $\mathcal{D}_2$, namely compare $\frac{P(\mathcal{D}_1|\omega)}{P(\mathcal{D}_1|\omega')}$ and $\frac{P(\mathcal{D}_2|\omega)}{P(\mathcal{D}_2|\omega')}$. Letting $\pi_i(\omega) = c_i P(\mathcal{D}_1|\omega)$ where c_i is a value such that $\max_{\omega \in \Omega} \pi_i(\omega) = 1$, it is obvious that

$$\frac{P(\mathcal{D}_1|\omega)}{P(\mathcal{D}_1|\omega')} \leq \frac{P(\mathcal{D}_2|\omega)}{P(\mathcal{D}_2|\omega')} \iff \pi_1 \sqsubseteq_w \pi_2$$

provided that $P(\mathcal{D}_1|\omega)$ and $P(\mathcal{D}_2|\omega)$ induce the same orderings. These considerations seem to indicate that the approach of possibility distributions as likelihood functions follows different principles from those of the usual approach, and the specificity ordering between them is questionable.

6.4. *The diffidence-based idempotent combination rule for possibility distributions*

Lastly let us focus on combination rules in possibility theory. Given two possibility distributions the usual idempotent conjunctive rule is the minimum: $\pi_{1\wedge 2} = \min(\pi_1, \pi_2)$ (see for instance [13]). This combination rule is in agreement with the q-ordering of belief functions as the consonant belief function based on $\pi_{1\wedge 2}$ is the least q-informative belief function that is more informative than both π_1 and π_2 [41]. There is also a pointwise conjunctive combination rule using product : $\pi_{1\circledcirc 2} = \pi_1 \cdot \pi_2$. It is the contour function of the (generally not consonant) result of combining the two consonant belief functions using the conjunctive rule. The diffidence approach brings us another idempotent conjunction in possibility theory, using $\owedge$, namely $\pi_{1\owedge 2}$.

In the consonant case, the set of focal sets is exactly the set of testimonies used for building the BPA (plus Ω). Suppose π_1, π_2 have nested focal sets $\mathcal{F}_1$, and $\mathcal{F}_2$. The diffidence-based idempotent combination rule $\owedge$ has interesting special cases:

- If $\mathcal{F}_1 \cap \mathcal{F}_2 = \{\Omega\}$, then $\pi_{1\owedge 2} = \pi_{1\circledcirc 2}$, i.e., we get the product rule. However the result is generally not consonant since $\owedge$ coincides with the conjunctive rule. It is consonant only if $\mathcal{F}_1 \cup \mathcal{F}_2$ is a nested sequence.
- If $\mathcal{F}_1 = \mathcal{F}_2$, then the result is consonant, but $\pi_{1\owedge 2} \neq \pi_{1\wedge 2}$, generally, i.e., we do not recover the minimum rule of possibility theory.

In the latter case the resulting possibility distribution can be derived from the consonant belief function $\circledcirc_{i=1}^{k} E_k^{\delta_k}$ where $\delta_i = \min(\frac{\pi_1^{i+1}}{\pi_1^i}, \frac{\pi_2^{i+1}}{\pi_2^i})$. We can see that

- if $\omega \in E_1$ then $\pi_{1\owedge 2}(\omega) = 1$
- if $\omega \in E_2 \setminus E_1$ then $\pi_{1\owedge 2}(\omega) = \min(\pi_1^1, \pi_2^1)$
- if $\omega \in E_{i+1} \setminus E_i$ then $\pi_{1\owedge 2}(\omega) = \min(\pi_1^1, \pi_2^1) \cdot \prod_{j=1}^{i-1} \min(\dfrac{\pi_1^{j+1}}{\pi_1^j}, \dfrac{\pi_2^{j+1}}{\pi_2^j})$.

It is clear that $\pi_{1\owedge 2} = \pi_{1\wedge 2}$ is the minimum rule when combining SSF focused on the same subset E. However it is no longer the case for combining possibility distributions having, say, two cuts $E_1 \subset E_2 \neq \Omega$. For $\omega \notin E_2$ we get $\pi_{1\owedge 2}(\omega) = \min(\pi_1^2, \pi_2^2) \cdot \min(\frac{\pi_1^3}{\pi_1^2}, \frac{\pi_2^3}{\pi_2^2})$.

In the general case, where $\mathcal{F}_1$ and $\mathcal{F}_2$ have sets other than Ω in common, the diffidence-based combination can take the following form induced by (12): $\pi_{1\owedge 2} = \pi_{1\setminus 2} \circledcirc \pi_{12} \circledcirc \pi_{2\setminus 1}$, where $\pi_{1\setminus 2}$ is the contour function of the consonant belief function based on $\mathcal{F}_1 \setminus \mathcal{F}_2$ using the diffidence weights from π_1, $\pi_{2\setminus 1}$ is the same, exchanging 1 and 2, and π_{12} is the result of combining by $\owedge$ the consonant belief functions obtained by restricting the focal sets of π_1 and π_2 to $\mathcal{F}_1 \cap \mathcal{F}_2$. All partial results are consonant belief functions but the final result may not be so if $\mathcal{F}_1 \cup \mathcal{F}_2$ is not nested. As a consequence we of course do not have that if $\pi_1 \leq \pi_2$ then $\pi_{1\owedge 2} = \min(\pi_1, \pi_2)$. Likewise, we do have that obviously $\pi_1 \circledcirc \pi_2 \sqsubset_w \pi_{1\owedge 2}$ by construction. But we do not have that $\pi_1 \circledcirc \pi_2 < \pi_{1\owedge 2}$.

Example 10. *We illustrate on some instances the above mentioned difference between the minimum rule, the conjunctive rule and the idempotent diffidence-based rule of combination. Consider a five-element set $\Omega = a, b, c, d, e$. Possibility distributions are denoted by 5-component vectors $(\pi(a), \pi(b), \pi(c), \pi(d), \pi(e))$. Subsets are denoted ab, abc etc.*

1. *Consider $\pi_1 = (1, 0.9, 0.7, 0.5, 0.2)$ and $\pi_2 = (1, 0.7, 0.6, 0.4, 0.1)$. Here, $\pi_1 > \pi_2$ and they have the same focal sets. In terms of diffidence, they are: $(a)^{0.9}(ab)^{7/9}(abc)^{5/7}(abcd)^{2/5}$ and $(a)^{0.7}(ab)^{6/7}(abc)^{4/6}(abcd)^{1/4}$ respectively. The minimum of diffidence weights is $(a)^{0.7}(ab)^{7/9}(abc)^{2/3}(abcd)^{1/4}$ hence $\pi_{1\owedge 2} = (1, 0.7, 49/90, 98/270, 98/1080) = (1, 0.7, 0.544, 0.363, 0.09)$, which is more specific than $\pi_{1\wedge 2} = \pi_2$.*
2. *Consider $\pi_1 = (1, 0.9, 0.9, 0.5, 0.5)$ and $\pi_2 = (1, 1, 0.8, 0.8, 0.2)$. They have different focal sets, but $\mathcal{F}_1 \cup \mathcal{F}_2$ is a nested sequence. In terms of diffidence, they can be written as $(a)^{0.9}(abc)^{5/9}$ and $(ab)^{0.8}(abcd)^{1/4}$ respectively. Then $\pi_{1\owedge 2}$ has decomposition $(a)^{0.9}(ab)^{0.8}(abc)^{5/9}(abcd)^{1/4}$, using conjunctive fusion of the elementary testimonies. It is clearly consonant. So, $\pi_{1\owedge 2} = (1, 0.9, 0.72, (72 \times 5)/9, (72 \times 5)/36) = (1, 0.9, 0.72, 0.4, 0.1) = \pi_1 \cdot \pi_2$ We recover the conjunctive rule on possibility distributions.*
3. *Consider $\pi_1 = (1, 0.9, 0.7, 0.5, 0.5)$ and $\pi_2 = (1, 0.8, 0.6, 0.6, 0.4)$. They have some focal sets in common. Decomposition: $(a)^{0.9}(ab)^{7/9}(abc)^{5/7}$ and $(a)^{0.8}(ab)^{3/4}(abcd)^{2/3}$.*
   $\pi_{1\owedge 2}$ has decomposition $(a)^{\min(0.9,0.8)}(ab)^{\min(7/9,3/4)}(abc)^{5/7}(abcd)^{2/3}$ where $\pi_{12} = (a)^{0.8}(ab)^{3/4}$, $\pi_{1\setminus 2} = (abc)^{5/7}$ and $\pi_{2\setminus 1} = (abcd)^{2/3}$. We obtain $\pi_{1\owedge 2} = (1, 0.8, 0.6, 30/70, 20/70)$. And clearly $\pi_1 \circledcirc \pi_2 \not< \pi_{1\owedge 2} < \min(\pi_1, \pi_2)$ here.
4. *Consider $\pi_1 = (1, 0.9, 0.7, 0.5, 0.2)$ and $\pi_2 = (1, 0.5, 0.7, 0.9, 0.2)$. They have some focal sets in common, but $\mathcal{F}_1 \cup \mathcal{F}_2$ is not nested. The decompositions are $(a)^{0.9}(ab)^{7/9}(abc)^{5/7}(abcd)^{2/5}$ and $(a)^{0.9}(ad)^{7/9}(acd)^{5/7}(abcd)^{2/5}$. Then $\pi_{1\owedge 2}$ corresponds to $(a)^{0.9}(ab)^{7/9}(abc)^{5/7}(ad)^{7/9}(acd)^{5/7}(abcd)^{2/5}$, and the result is not a consonant belief function.*

The above examples confirm that the diffidence-based idempotent rule of combination differs from the minimum rule of possibility theory, and yields more specific results than the latter. Moreover, just like for the conjunctive rule, the result of combining two consonant belief functions by the diffidence-based idempotent rule may fail to be consonant.

7. Conclusion

This paper revisits the decomposition of a non-dogmatic belief function into a combination of generalized simple support functions proposed by Smets [40] based on the diffidence function, showing that it can be viewed as the merging of uncertain testimonies and of reasons not to believe the result of their partial conjunctions, which we have called prejudices. In other words, starting from the intuition of Shafer, who considered belief functions resulting from the merging of simple uncertain pieces of information, our paper tries to provide an extensive presentation of non-dogmatic belief functions that articulate more recent findings of Smets and Denœux into an intuitively coherent picture of the theory. Our paper is an attempt to explain generalized SSFs with diffidence weights larger than 1, showing they accurately model the idea of prejudice whose role is to resist to the input of some new pieces of information. We also show that the operation of retraction avoids the explicit use of deviant belief functions such as GSSFs. We suggest that prejudices are due to some prior knowledge that is more entrenched than incoming new pieces of uncertain evidence. In this sense, the question of elicitation of the strength of prejudices comes down to the elicitation of prior knowledge of an agent in the form of belief functions.

A contribution of this paper is a detailed investigation of how such prior information can affect a belief function, considering retraction as a special asymmetric belief change operation that avoids the explicit use of negative mass functions. Besides, we have studied properties of information orderings and the idempotent combination rule based on the diffidence function. Our results strengthen the approach to belief functions based on the merging of pieces of evidence, as opposed to the approach based on upper and lower probability and imprecise statistics.

It is important to notice that prejudices, as we conceive them, differ from prior probabilities. In the Bayesian approach prior probabilities are merged with likelihood functions pertaining to sure pieces of evidence, and Shafer has shown that this process is a special case of the orthogonal rule of combination, where prior information and incoming information play the same role. This is problematic if the prior information severely conflicts with the new evidence, as the results of the combination may become questionable due to contradiction. On the contrary, while in the Bayesian setting, acquiring evidence can never increase ignorance, retraction, akin to contraction in belief revision theory, may lead from knowledge to ignorance: a prejudice conflicting with a piece of evidence or with the result of merging such independent pieces tends to weaken or even erase it in such a way that a state of ignorance results. It seems that this mathematical model of resistance in front of new information could be useful in the area of information fusion.

In the future it would be of interest to check the cognitive plausibility of latent belief structures and the role of prejudices in the way humans process new information, in line with previous studies highlighting the bipolar nature of human knowledge (see Dubois and Prade [15] for a bibliography).

References

[1] Bloch, I., Hunter, A. (Eds.), 2001. Special issue on Data and Knowledge Fusion. volume 16 (10-11) of *International Journal of Intelligent Systems*. Wiley.

[2] Coletti, G., Scozzafava, R., 2003. Coherent conditional probability as a measure of uncertainty of the relevant conditioning events, in: Proc. of ECSQARU03, Springer Verlag, Aalborg. pp. 407–418.

[3] De Cooman, G., Aeyels, D., 1999. Supremum-preserving upper probabilities. Information Sciences 118, 173–212.

[4] Dempster, A., 1967. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339.

[5] Denœux, T., 2008. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial Intelligence 172, 234–264.

[6] Denœux, T., 2014. Likelihood-based belief function: Justification and some extensions to low-quality data. Int. J. Approx. Reasoning 55, 1535–1547.

[7] Destercke, S., Dubois, D., 2011. Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. Inf. Sci. 181, 3925–3945.

[8] Dubois, D., Faux, F., Prade, H., 2018. Prejudiced information fusion using belief functions, in: Destercke, S., Denoeux, T., Cuzzolin, F., Martin, A. (Eds.), Belief Functions: Theory and Applications - 5th International Conference, BELIEF 2018, Compiègne, France, September 17-21, 2018, Proceedings, Springer. pp. 77–85.

[9] Dubois, D., Liu, W., Ma, J., Prade, H., 2016. The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. Information Fusion 32, 12 – 39.

[10] Dubois, D., Moral, S., Prade, H., 1997. A semantics for possibility theory based on likelihoods. J. of Mathematical Analysis and Applications 205, 359 – 380.

[11] Dubois, D., Prade, H., 1986. A set-theoretic view of belief functions logical operations and approximations by fuzzy sets. International Journal of General Systems 12, 193–226.

[12] Dubois, D., Prade, H., 1988a. Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press.

[13] Dubois, D., Prade, H., 1988b. Representation and combination of uncertainty with belief functions and possibility measures. Computational Intelligence 4, 244–264.

[14] Dubois, D., Prade, H., 1992. When upper probabilities are possibility measures. Fuzzy Sets and Systems 49, 65–74.

[15] Dubois, D., Prade, H., 2008. An introduction to bipolar representations of information and preference. International Journal of Intelligent Systems 23, 866–877.

[16] Dubois, D., Prade, H., Smets, P., 2001. "Not impossible" vs. "guaranteed possible" in fusion and revision, in: Benferhat, S., Besnard, P. (Eds.), Proc. 6th ECSQARU 2001, Springer. pp. 522–531.

[17] Dubois, D., Prade, H., Smets, P., 2008. A definition of subjective possibility. Int. J. Approx. Reasoning 48, 352–364.

[18] Edwards, A., 1972. Likelihood. Cambridge University Press. Expanded edition, 1992, Johns Hopkins University Press, Baltimore.

[19] Gärdenfors, P., 1988. Knowledge in Flux: Modeling the Dynamics of Epistemic States. MIT Press.

[20] Ginsberg, M.L., 1984. Non-monotonic reasoning using Dempster's rule, in: Proc. Nat. Conf. on Artificial Intelligence. Austin, TX, August 6-10, 1984., pp. 126–129.

[21] Guyard, R., Cherfaoui, V., 2018. Study of discounting methods applied to canonical decomposition of belief functions, in: 21st International Conference on Information Fusion, Cambridge, United Kingdom. pp. 2505–2512.

[22] Kallel, A., Le Hégarat-Mascle, S., Hubert-Moy, L., Ottlé, C., 2008. Fusion of vegetation indices using continuous belief functions and cautious-adaptive combination rule. IEEE Transactions on Geoscience and Remote Sensing 46, 1499–1513.

[23] Ke, X., Ma, L., Wang, Y., 2014. Some notes on canonical decomposition and separability of a belief function, in: Cuzzolin, F. (Ed.), Belief Functions: Theory and Applications (Proc. 3d Conf. BELIEF 2014). Springer. volume 8764 of *LNCS*, pp. 153–160.

[24] Klawonn, F., Smets, P., 1992. The dynamic of belief in the transferable belief model and specialization-generalization matrices, in: UAI '92: Proceedings of the Eighth Annual Conference on Uncertainty in Artificial Intelligence, Stanford University, Stanford, CA, USA, July 17-19, 1992, pp. 130–137.

[25] Kramosil, I., 1999. Measure-theoretic approach to the inversion problem for belief functions. Fuzzy Sets and Systems 102, 363 – 369.

[26] Kramosil, I., 2001. Probabilistic Analysis of Belief Functions. Kluwer, New York.

[27] Ma, J., Liu, W., Dubois, D., Prade, H., 2011. Bridging Jeffrey's rule, AGM revision and Dempster conditioning in the theory of evidence. International Journal on Artificial Intelligence Tools 20, 691–720.

[28] Mercier, D., Quost, B., Denoeux, T., 2008. Refined modeling of sensor reliability in the belief function framework using contextual discounting. Information Fusion 9, 246–258.

[29] Pawlak, Z., 1991. Rough Sets - Theoretical Aspects of Reasoning about Data. Kluwer Academic Publ., Dordrecht.

[30] Pichon, F., 2009. Belief functions: canonical decompositions and combination rules. Ph.D. thesis. Université de Technologie de Compiègne.

[31] Pichon, F., 2018. Canonical decomposition of belief functions based on Teugels representation of the multivariate bernoulli distribution. Inf. Sci. 428, 76–104.

[32] Pichon, F., Denoeux, T., 2007. On latent belief structures, in: Mellouli, K. (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2007, Springer. pp. 368–380.

[33] Pichon, F., Denœux, T., Dubois, D., 2012. Relevance and truthfulness in information correction and fusion. Int. J. of Approximate Reasoning 53, 159–175.

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

[35] Shafer, G., 1978. Non-additive probabilities in the work of Bernoulli and Lambert. Archive for History of Exact Sciences 19, 309–370.

[36] Shafer, G., 1986. The combination of evidence. International Journal of Intelligent Systems 1, 155–179.

[37] Shafer, G., 2016. A mathematical theory of evidence turns 40. International Journal of Approximate Reasoning 79, 7 – 25.

[38] Shenoy, P.P., 1994. Conditional independence in valuation-based systems. Int. J. Approx. Reasoning 10, 203–234.

[39] Smets, P., 1982. Possibilistic inference from statistical data, in: Proc. of the 2nd World Conf. on Mathematics at the Service of Man, Las Palmas (Canary Island), Spain. pp. 611–613.

[40] Smets, P., 1995. The canonical decomposition of a weighted belief, in: Proc. 14th Int. Joint Conf. on Artificial Intelligence (IJCAI), Montreal, Aug. 20-25, pp. 1896–1901.

[41] Smets, P., 2000. Quantified possibility theory seen as an hypercautious transferable belief model, in: Rencontres Francophones sur les Logiques Floues et ses Applications (LFA 2000), Cepadues-Editions, La Rochelle, France. pp. 343–353.

[42] Smets, P., Kennes, R., 1994. The transferable belief model. Artif. Intell. 66, 191–234.

[43] Teugels, J.L., 1990. Some representations of the multivariate Bernoulli and binomial distributions. Journal of Multivariate Analysis 32, 256–268.

[44] Tversky, A., Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review 90, 293–315.

[45] Yager, R., 1986. The entailment principle for Dempster-Shafer granules. I. J. of Intelligent Systems 1, 247–262.