Canonical Decomposition of Belief Functions Based on Teugels' Representation of the Multivariate Bernoulli Distribution

Frédéric Pichon

Canonical Decomposition of Belief Functions Based on Teugels' Representation of the Multivariate Bernoulli Distribution

Frédéric Pichon

Univ. Artois, EA 3926,
Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A),
Béthune, F-62400, France
frederic.pichon@univ-artois.fr

October 14, 2017

Abstract

A canonical decomposition of belief functions is a unique decomposition of belief functions into elementary pieces of evidence. Smets found an equivalent representation of belief functions, which he interpreted as a canonical decomposition. However, his proposal is not entirely satisfactory as it involves elementary pieces of evidence, corresponding to a generalisation of belief function axioms, whose semantics lacks formal justifications. In this paper, a new canonical decomposition relying only on well-defined concepts is proposed. In particular, it is based on a means to induce belief functions from the multivariate Bernoulli distribution and on Teugels' representation of this distribution, which consists of the means and the central moments of the underlying Bernoulli random variables. According to our decomposition, a belief function results from as many crisp pieces of information as there are elements in its domain, and from simple probabilistic knowledge concerning their marginal reliability and the dependencies between their reliability. In addition, we show that instead of interpreting with some difficulty Smets' representation of belief functions as a canonical decomposition, it is possible to give it a different and well-defined semantics in terms of measures of information associated with the reliability of the pieces of information in our decomposition.

Keywords: Dempster-Shafer Theory; Belief Function; Canonical Decomposition; Multivariate Bernoulli Distribution; Moment; Mutual Information; Random Set.

1 Introduction

A belief function [34, 40] is a rich mathematical object for representing uncertain information on the actual value taken by a variable. Belief functions have been successfully applied to numerous problems such as, recently, clustering of large dissimilarity data [6], multi-label classification [7], object association [25, 8], and rule-based classification [17]; we refer the reader to [5] and the references therein for a much larger account of belief function applications.

Belief functions were originally introduced by Shafer [34], as a formal object for the representation of evidence. An important result shown in [34] is that the so-called separable belief functions can be canonically decomposed, that is, decomposed uniquely into elementary pieces of evidence. Smets [37] extended this result and proposed a solution to canonically decompose any belief function. The concept of canonical decomposition is important at a fundamental level in that it lays bare the underlying elementary components of a complex belief state. It also raises the interesting question as to how can one accumulate evidence progressively to construct probability judgements on the value taken by a variable of interest. Moreover, it can be useful to tackle several problems, as exemplified in recent years where proposals based on Smets' decomposition have appeared to address the issues of belief function combination [4, 19, 29], correction [24, 26, 31] and clustering [33].

In this paper, this concept is deeply revisited. First, a critical review of Smets' canonical decomposition is conducted (Section 2): we challenge his solution and argue that it is not entirely satisfactory. Then, a new canonical decomposition of belief functions is proposed (Section 3). This new decomposition is in the same spirit as Smets' – a belief function is viewed as resulting from partially reliable sources providing crisp pieces of information –, but it follows a completely different approach based on some results concerning the representation of the multivariate Bernoulli distribution [42]. Next, some comments on our solution are provided and, in particular, our solution is compared to Smets' and is considered in the context of random sets [21, 23] (Section 4). Finally, a new perspective on Smets' decomposition is brought to light using measures of information [14] associated with the multivariate Bernoulli distribution (Section 5). Section 6 concludes the paper.

2 Review of Smets' Canonical Decomposition

In this section, necessary background on belief functions and on their canonical decomposition as proposed by Smets is provided. Then, a general information fusion approach based on explicit assumptions about the reliability of information sources is recalled and used for critical examination of Smets' canonical decomposition.

2.1 Basics of Belief Function Theory

Belief function theory [34, 40] is a framework for uncertainty modelling and reasoning. In this theory, uncertainty regarding the actual value taken by a variable $\boldsymbol{y}$ defined on a finite domain $\mathcal{Y} = \{y_1, ..., y_n\}$, is represented by a so-called *mass* function (MF) defined as a mapping $m : 2^{\mathcal{Y}} \rightarrow [0, 1]$ satisfying $\sum_{A \subseteq \mathcal{Y}} m(A) = 1$. The quantity $m(A)$ may be interpreted as the probability of knowing only that $\boldsymbol{y} \in A$, $A \subseteq \mathcal{Y}$ [2]. Subsets A of $\mathcal{Y}$ such that $m(A) > 0$ are called *focal sets* of m. A mass function is called: *dogmatic* if $\mathcal{Y}$ is not a focal set; *normal* if $\emptyset$ is not a focal set; *vacuous* if $\mathcal{Y}$ is its only focal set; *simple* if it has at most two focal sets, and if it has two, $\mathcal{Y}$ is one of those; *Bayesian* if its focal sets are singletons. A non normal MF m can be normalised, *i.e.*, transformed into a normal MF m^*, by the following operation:

$$m^*(A) = \begin{cases} \kappa \cdot m(A) & \text{for all } A \subseteq \mathcal{Y}, A \neq \emptyset, \\ 0 & \text{if } A = \emptyset, \end{cases} \tag{1}$$

with $\kappa = (1 - m(\emptyset))^{-1}$.

The *belief* function *bel* is an equivalent representation of a mass function m. It is defined as

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad \forall A \subseteq \mathcal{Y}.$$

The degree of belief $bel(A)$ evaluates to what extent event A is logically implied by the available evidence [13]. Other equivalent representations of m that are of interest for this paper are the *plausibility* pl and *commonality* q functions:

$$\begin{aligned} pl(A) &= \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \mathcal{Y}, \\ q(A) &= \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \mathcal{Y}. \end{aligned} \tag{2}$$

The degree of plausibility $pl(A)$ evaluates to what extent event A is consistent with the available evidence [13]. The commonality function has a technical role, which will

be described later. Functions m, bel, pl and q are in one-to-one correspondence, in particular mass function m can be recovered from any of these functions. For instance, we have:

$$m(A) = \sum_{B \supseteq A} (-1)^{|B|-|A|} q(B), \quad \forall A \subseteq \mathcal{Y}, \tag{3}$$

with $|A|$ denoting the cardinality of $A \subseteq \mathcal{Y}$. Let us note that the plausibility function restricted to the singletons of $\mathcal{Y}$ is called the *contour* function. Besides, the plausibility of any singleton of $\mathcal{Y}$ is equal to its commonality, *i.e.*, $pl(\{y\}) = q(\{y\})$ for all $y \in \mathcal{Y}$.

Matrix calculus is useful to simplify the mathematics of belief function theory [38]. A mass function m and its associated functions, such as q, can be seen as column vectors of size $2^{|\mathcal{Y}|} = 2^n$, whose elements are ordered according to the so-called binary order detailed hereafter. Let k be an integer such that $1 \leq k \leq 2^n$. k can be written in a binary expansion, *i.e.*,

$$k = 1 + \sum_{i=1}^{n} k_i 2^{i-1}, \tag{4}$$

where $k_i \in \{0, 1\}$. Expansion (4) induces a one-to-one correspondence $k \leftrightarrow (k_1, \ldots, k_n)$, that is Eq. (4) associates to each integer k, $1 \leq k \leq 2^n$, a binary vector $(k_1, \ldots, k_n) \in \{0, 1\}^n$. Let A_k, $1 \leq k \leq 2^n$, be the subset of $\mathcal{Y}$, such that $y_i \in A_k$ if $k_i = 1$ and $y_i \notin A_k$ if $k_i = 0$, with k_i, $i = 1, \ldots, n$, the terms in the binary expansion (4) of k. In the binary order, the k-th element of the vector $\mathbf{m}$ corresponds to the set A_k. Thus $\emptyset$ is the first element, $\{y_1\}$ the second element, $\{y_2\}$ the third element, $\{y_1, y_2\}$ the fourth element, etc. For instance, for $n = 4$, $A_{14} = \{y_1, y_3, y_4\}$ since $k = 14$ is in one-to-one correspondence with $(k_1 = 1, k_2 = 0, k_3 = 1, k_4 = 1)$. Eqs. (2) and (3) become in matrix form [38]:

$$\begin{aligned} \mathbf{q} &= \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) \mathbf{m}, \\ \mathbf{m} &= \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right) \mathbf{q}, \end{aligned} \tag{5}$$

where $\otimes$ denotes Kronecker product.

Several so-called belief function combination rules have been proposed to combine multiple pieces of information about a variable [39]. The most classical ones are Dempster's rule of combination [3, 34] and its unnormalised version called conjunctive rule [39]. We denote the former by $\oplus$ and the latter by $\textcircled{\cap}$. They are defined as follows. Let m_1 and m_2 be two mass functions representing uncertainty about a variable $\boldsymbol{y}$. Let

$m_{1\oplus 2} = m_1 \oplus m_2$ and $m_{1\textcircled{\cap}2} = m_1 \textcircled{\cap} m_2$ denote the mass functions resulting from the combination of m_1 and m_2 by $\oplus$ and by $\textcircled{\cap}$ respectively. We have

$$m_{1\textcircled{\cap}2}(A) \quad = \quad \sum_{B\cap C=A} m_1(B)m_2(C), \quad \forall A \subseteq \mathcal{Y}, \tag{6}$$

and, assuming that $m_{1\textcircled{\cap}2}(\emptyset) \neq 1$, $m_{1\oplus 2} = m^*_{1\textcircled{\cap}2}$, *i.e.*, combination by Dempster's rule amounts to combination by the conjunctive rule (6) followed by normalisation (1). Both rules are commutative, associative and admit the vacuous mass function as only neutral element. These rules are appropriate when m_1 and m_2 represent independent bodies of evidence, and the conjunctive rule is sensible under the so-called open world assumption, whereas Dempter's rule corresponds to the closed world assumption [39]. The conjunctive rule has a simple expression using the commonality function:

$$q_{1\textcircled{\cap}2}(A) \quad = \quad q_1(A) \cdot q_2(A), \quad \forall A \subseteq \mathcal{Y}, \tag{7}$$

with q_1, q_2 and $q_{1\textcircled{\cap}2}$ the commonality functions associated to m_1, m_2 and $m_{1\textcircled{\cap}2}$, respectively.

The inverse of the conjunctive rule, denoted by $\textcircled{\not\cap}$, can be defined [37, 4]. It is the rule which restores m_1 from $m_{1\textcircled{\cap}2}$ and m_2, *i.e.*, $m_{1\textcircled{\cap}2}\textcircled{\not\cap}m_2 = m_1$. It is sensible if we learn that m_2 is actually not supported by evidence and should thus be removed from $m_{1\textcircled{\cap}2}$. Let q_1 and q_2 be the commonality functions associated respectively to any two mass functions m_1 and m_2, the inverse of the conjunctive rule is defined as:

$$q_{1\textcircled{\not\cap}2}(A) \quad = \quad \frac{q_1(A)}{q_2(A)}, \quad \forall A \subseteq \mathcal{Y}. \tag{8}$$

This operation is well-defined as long as (i) $m_{1\textcircled{\not\cap}2}$ is a mass function, which is not necessarily the case since the quotient of two commonality functions is not always a commonality function, and (ii) m_2 is non dogmatic, which ensures $q_2(A) > 0$ for all $A \subseteq \mathcal{Y}$.

2.2 Smets' Canonical Decomposition of Belief Functions

Some mass functions can be obtained as the result of the combination by Dempster's rule of simple mass functions; they are called *separable* by Shafer [34]. A simple mass function having two focal sets $A \subset \mathcal{Y}$ and $\mathcal{Y}$, with respective masses $1 - w$ and w, $w \in [0, 1]$, may be simply denoted A^w. For every separable mass function m, one has

$$m \quad = \quad \bigoplus_{\emptyset \neq A \subset \mathcal{Y}} A^{w(A)}, \tag{9}$$

with $w(A) \in [0,1]$ for all $A \subset \mathcal{Y}, A \neq \emptyset$. This *canonical decomposition* of m is unique if m is non dogmatic (in which case $w(A) > 0$ for all $A \subset \mathcal{Y}, A \neq \emptyset$).

Smets [37] extended the concept of separability to the more general case of non normal mass functions. This extended concept is called u-separability in [4] (u stands for unnormalised) and relies on the unnormalised version of Dempster's rule, *i.e.*, the conjunctive rule. A mass function m is thus u-separable if it can be expressed as:

$$m = \bigcirc\!\!\!\!\!\cap_{A \subset \mathcal{Y}} A^{w(A)}, \tag{10}$$

with $w(A) \in [0,1]$ for all $A \subset \mathcal{Y}$. This decomposition is also unique if m is non dogmatic.

Smets [37] extended this latter decomposition to all non dogmatic mass functions. The decomposition of non dogmatic mass functions relies on the concept of *generalised* simple mass function defined as a function $\mu : 2^{\mathcal{Y}} \to \mathbb{R}$ verifying:

$$\begin{aligned} \mu(A) &= 1 - w, \\ \mu(\mathcal{Y}) &= w, \\ \mu(B) &= 0, \quad \forall B \in 2^{\mathcal{Y}} \backslash \{A, \mathcal{Y}\}, \end{aligned}$$

for some $A \neq \mathcal{Y}$ and some $w \in [0, +\infty)$. Any generalised simple mass function μ can be noted A^w for some $A \neq \mathcal{Y}$ and $w \in [0, +\infty)$. When $w \leq 1$, A^w is a simple mass function. When $w > 1$, A^w is called *inverse* simple mass function; Smets [37] proposed to interpret such function as reasons *not* to believe in A (he also speaks of "debt of belief" in A) since combining $A^{1/w}$ with A^w by $\textcircled{\cap}$ (using a trivial extension of the conjunctive rule to combine generalised simple mass functions) yields the vacuous mass function, that is, the reasons to believe in A represented by $A^{1/w}$ are counter-balanced by A^w.

Using the concept of generalised simple mass function, Smets showed that for any non dogmatic mass function m, we have:

$$m = \bigcirc\!\!\!\!\!\cap_{A \subset \mathcal{Y}} A^{w(A)}, \tag{11}$$

with $w(A) \in (0, +\infty)$ for all $A \subset \mathcal{Y}$. The *weight* function $w : 2^{\mathcal{Y}} \backslash \{\mathcal{Y}\} \to (0, +\infty)$ that appears in (11), is an equivalent representation of a non dogmatic mass function. It can be obtained from the commonality function as follows:

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}, \quad \forall A \subset \mathcal{Y}. \tag{12}$$

Note that the conjunctive rule has a simple expression using the weight function:

$$w_{1\textcircled{\cap}2}(A) \quad = \quad w_1(A) \cdot w_2(A), \quad \forall A \subset \mathcal{Y}. \tag{13}$$

Using function w, Smets [37] further showed that any non dogmatic mass function m can be written as

$$m \quad = \quad m^c \textcircled{\oslash} m^d, \tag{14}$$

with m^c and m^d two u-separable mass functions such that

$$m^c \quad = \quad \bigcirc\!\!\!\!\!\!\cap_{A \subset \mathcal{Y}} A^{1 \wedge w(A)},$$

and

$$m^d \quad = \quad \bigcirc\!\!\!\!\!\!\cap_{A \subset \mathcal{Y}} A^{1 \wedge \frac{1}{w(A)}},$$

where $\wedge$ denotes the minimum operator. In other words, Smets [37] showed with (14) that any non dogmatic mass function can be uniquely obtained from simple mass functions. The m^c and m^d mass functions in (14) are called the *confidence and diffidence components*, respectively, of m by Smets [37], who proposed to view m^c as representing "good reasons to believe" in some propositions $A \subset \mathcal{Y}$, and m^d as representing "good reasons not to believe" in some other propositions.

Smets' canonical decomposition of non dogmatic mass functions is illustrated by Example 1.

Example 1 ("Example 2, continuation" of [37])**.** *Let $\mathcal{Y} = \{y_1, y_2, y_3\}$ be the domain of a variable $\boldsymbol{y}$ and let m be a mass function on $\mathcal{Y}$ such that*

$$m(\{y_1, y_2\}) = m(\{y_1, y_3\}) = m(\{y_1, y_2, y_3\}) = 1/3. \tag{15}$$

We have

$$\begin{aligned} w(\{y_1, y_2\}) &= 1/2, \\ w(\{y_1, y_3\}) &= 1/2, \\ w(\{y_1\}) &= 4/3, \end{aligned}$$

and $w(A) = 1$ for all $A \in 2^{\mathcal{Y}} \backslash \{\mathcal{Y}, \{y_1, y_2\}, \{y_1, y_3\}, \{y_1\}\}$. Hence, using (11) *and the fact that the vacuous mass function is the neutral element of $\textcircled{\cap}$ and can be noted A^1 for any $A \subset \mathcal{Y}$, we have*

$$m \quad = \quad \{y_1, y_2\}^{1/2} \textcircled{\cap} \{y_1, y_3\}^{1/2} \textcircled{\cap} \{y_1\}^{4/3},$$

and using (14) *we have*

$$m \;=\; \{y_1, y_2\}^{1/2} \textcircled{\cap} \{y_1, y_3\}^{1/2} \textcircled{\varnothing} \{y_1\}^{3/4}.$$

According to Smets [37], m is thus the result of the following independent pieces of evidence:

- *a first source provides the evidence* believe $\{y_1, y_2\}$ *and this source is given reliability* $1/2$*;*
- *a second source provides the evidence* believe $\{y_1, y_3\}$ *and this source is given reliability* $1/2$*;*
- *a third source provides the evidence* do not believe $\{y_1\}$ *and this source is given reliability* $3/4$*.*

Smets' decomposition is quite elegant and has sparked several contributions such as fusion [4, 19, 29], correction [24, 26, 31] and clustering schemes [33]. Nonetheless, despite its appeal and the success that it has enjoyed, we will try and argue in Section 2.4 that Smets' decomposition is not entirely satisfactory and we will provide an alternative decomposition in Section 3. Both our critique of Smets' decomposition and our alternative to it rely in part on a general approach to information fusion, which is recalled in the next section.

2.3 Behaviour-Based Fusion

In [30], Pichon *et al.* introduced a general approach to information fusion. In this approach, an agent builds his belief about the actual value of a variable $\boldsymbol{y}$ defined on a domain $\mathcal{Y}$, given pieces of information about $\boldsymbol{y}$ provided by some sources and given his knowledge about the behaviour of these sources (called *meta-knowledge* in [30] as it is higher order knowledge that differs from the knowledge supplied by the sources). This approach is recalled hereafter in the particular case where the sources provide crisp pieces of information about $\boldsymbol{y}$ and meta-knowledge on the sources concern their reliability, as this particular case is instrumental to the present paper.

Let us consider the simple situation where there is a single source $\mathfrak{s}_1$ providing to an agent a piece of information about $\boldsymbol{y}$ and this piece of information is crisp, *i.e.*, it is of the form $\boldsymbol{y} \in A$ for some $A \subseteq \mathcal{Y}$. Besides, the agent assumes that the source can be in only one of two states: reliable or not reliable. If the source is reliable, then the agent

should deduce that $\boldsymbol{y} \in A$ from the piece of information provided by $\mathfrak{s}_1$. If the source is not reliable, then the piece of information provided by $\mathfrak{s}_1$ is useless and the agent knows only that $\boldsymbol{y} \in \mathcal{Y}$, *i.e.*, he knows nothing.

Let $\mathcal{X}_1 = \{0, 1\}$ be the space denoting the reliability of $\mathfrak{s}_1$, where 0 means that $\mathfrak{s}_1$ is reliable and 1 means that $\mathfrak{s}_1$ is not reliable. The above reasoning can be encoded by multi-valued mappings Γ_A, $A \subseteq \mathcal{Y}$, from $\mathcal{X}_1$ to $\mathcal{Y}$ such that

$$\begin{aligned} \Gamma_A(0) &= A, \\ \Gamma_A(1) &= \mathcal{Y}. \end{aligned}$$

$\Gamma_A(k_1)$ interprets the testimony $\boldsymbol{y} \in A$ in each configuration $k_1 \in \mathcal{X}_1$ of the source $\mathfrak{s}_1$.

More generally, meta-knowledge on the source may be uncertain and specifically $\mathfrak{s}_1$ may be assumed to be reliable with probability $1-\pi_1$ and not reliable with probability π_1, $\pi_1 \in [0, 1]$. In such case, if $\mathfrak{s}_1$ provides the piece of information $\boldsymbol{y} \in A$, then probability $1 - \pi_1$ will be transferred to $\Gamma_A(0)$ and probability π_1 to $\Gamma_A(1)$, *i.e.*, the knowledge of the agent about $\boldsymbol{y}$ is represented by a mass function defined as [30][1]:

$$\begin{aligned} m(A) &= 1 - \pi_1, \\ m(\mathcal{Y}) &= \pi_1. \end{aligned} \tag{16}$$

We may remark that mass function defined by (16) is a simple mass function, which may be denoted by A^{π_1}. Besides, the transformation of testimony $\boldsymbol{y} \in A$ according to (16) is nothing but the discounting operation proposed by Shafer [34] to integrate the reliability of information sources.

Let us now consider the case where there are N sources providing crisp pieces of information about $\boldsymbol{y}$ and each source can either be reliable or not reliable. Let $\mathcal{X}_i = \{0, 1\}$ be the space denoting the reliability of source $\mathfrak{s}_i$, $i = 1, \ldots, N$, where 0 means that $\mathfrak{s}_i$ is reliable and 1 means that $\mathfrak{s}_i$ is not reliable. The set of elementary joint states on the sources is therefore the Cartesian product $\mathcal{X}^N := \times_{i=1}^{N} \mathcal{X}_i$. To any state $(k_1, \ldots, k_N) \in \mathcal{X}^N$, we associate the number k, $1 \leq k \leq 2^N$, such that

$$k = 1 + \sum_{i=1}^{N} k_i 2^{i-1}, \tag{17}$$

[1]Mass function m may be induced formally as follows. $(\mathcal{X}_1, 2^{\mathcal{X}_1}, P)$ is a probability space, with P a probability measure on $\mathcal{X}_1$ such that $P(\{0\}) = 1 - \pi_1$ and $P(\{1\}) = \pi_1$, and $(\mathcal{Y}, 2^{\mathcal{Y}})$ is a measurable space. Since Γ_A is *strongly measurable* [27] with respect to $2^{\mathcal{Y}}$ and $2^{\mathcal{X}_1}$, the four-tuple $(\mathcal{X}_1, 2^{\mathcal{X}_1}, P, \Gamma_A)$ induces a belief function on $\mathcal{Y}$ (obtained by composition of P and the *lower inverse* [27] of Γ_A) with associated mass function m defined by (16) [27, 20].

i.e., we have $k \leftrightarrow (k_1, \ldots, k_N) \in \mathcal{X}^N$. Furthermore for any $k \leftrightarrow (k_1, \ldots, k_N) \in \mathcal{X}^N$, we define a mapping $\Gamma_{\mathbf{A}}$ for any $\mathbf{A} = (A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N}) \subseteq \times_{i=1}^N \mathcal{Y}$ as

$$\Gamma_{\mathbf{A}}(k) \quad = \quad \bigcap_{i=1}^{N} \Gamma_{A_{\mathfrak{s}_i}}(k_i).$$

$\Gamma_{\mathbf{A}}(k)$ represents the information deduced on $\mathcal{Y}$ from crisp testimonies $(A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N})$ provided by sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$, when they are in states $(k_1, \ldots, k_N)$ [30]. We have $\Gamma_{\mathbf{A}}(k) \subseteq \mathcal{Y}$ since $\Gamma_{\mathbf{A}}(k)$ is defined as the intersection of subsets of $\mathcal{Y}$.

In the more general case where meta-knowledge on the sources is uncertain and each joint state $k \in \mathcal{X}^N$ has a probability p_k, the knowledge of the agent about $\mathcal{Y}$ is represented by a mass function defined as [30]:

$$m(B) \quad = \quad \sum_{k:\Gamma_{\mathbf{A}}(k)=B} p_k, \quad \forall B \subseteq \mathcal{Y}. \tag{18}$$

An important special case of uncertain meta-knowledge is when for each $k \in \mathcal{X}^N$, the probability p_k that the sources are in joint state $(k_1, \ldots, k_N)$ is equal to the product of the marginal probabilities of the individual states k_i, $i = 1, \ldots, N$, *i.e.*, the probabilities p_k satisfy the following property

$$p_k \quad = \quad \prod_{i=1}^{N} (1 - \pi_i)^{1-k_i} \pi_i^{k_i}, \quad \forall k \in \mathcal{X}^N, \tag{19}$$

with k_i, $i = 1, \ldots, N$, the terms in the binary expansion (17) of k, and π_i the marginal probability that $\mathfrak{s}_i$ is not reliable, *i.e.*,

$$\pi_i \quad = \quad \sum_{k:k_i=1} p_k. \tag{20}$$

This property is a case of so-called *meta-independence* between the sources [30], which corresponds to the situation where pieces of meta-knowledge regarding the states of each source are independent. From [30, Theorem 1], if the probabilities p_k satisfy (19), then the mass function m given by (18) can be equivalently written as:

$$m \quad = \quad \bigcirc\!\!\!\!\!\!\bigcap_{i=1}^{N} A_{\mathfrak{s}_i}^{\pi_i}, \tag{21}$$

with $A_{\mathfrak{s}_i}^{\pi_i}$ the simple mass function such that $m(A_{\mathfrak{s}_i}) = 1 - \pi_i$ and $m(\mathcal{Y}) = \pi_i$.

2.4 Discussion on Smets' Decomposition

It is clear that the canonical decomposition (10) is an instance of (21) and thus can be given an interpretation using Pichon *et al.*'s scheme recalled in the preceding section. Precisely, any u-separable mass function m defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ and with associated weight function w, can be seen as originating from the following pieces of evidence:

- there are $2^n - 1$ sources, with source $\mathfrak{s}_i$, $i = 1, \ldots, 2^n - 1$, providing the crisp piece of information $\boldsymbol{y} \in A_i$, with A_i the i-th subset of $\mathcal{Y}$ according to the binary order;
- each source $\mathfrak{s}_i$ is non reliable with marginal probability $w(A_i)$;
- the sources are meta-independent.

This interpretation of the canonical decomposition (10), obtained using Pichon *et al.*'s scheme, is totally in line with Smets' interpretation of this decomposition, as can be recognised when comparing the above pieces of evidence to the ones at play in Example 1. The above interpretation is a bit more cumbersome due to its heavier formalisation but it has the advantage to rely on well-known concepts since it follows Dempster's original approach [3], where belief functions are induced from a space equipped with a probability measure and a multi-valued mapping from this space to another one. Overall, one may argue that both interpretation are essentially one and the same, and that Pichon *et al.*'s scheme merely provides a formal ground to Smets' interpretation using well-known concepts.

Due to the similarity between (10) and (11), one may wonder whether the straightforward extension of the above interpretation to non-dogmatic mass functions decomposed using (11) is possible. The answer is negative since it would require allowing marginal probabilities $w(A_i)$ greater than 1. We feel that it is a drawback of Smets' decomposition that it is incompatible with the partially reliable source model relying only on well-known concepts and leading to (21), while this formal model seems to be totally in line with his view on his decomposition.

This leads us to discuss further Smets' decomposition of non-dogmatic mass functions, irrespective of our previous comment. His decomposition involves so-called inverse simple mass functions A^w, $w > 1$, which we recall are *not* mass functions since they do not satisfy $m(A) \in [0, 1]$ for all $A \subseteq \mathcal{Y}$. Smets [37] proposed to interpret these functions as pieces of evidence of the form *a source provides the evidence* do not believe A *and this*

source is given reliability $1/w$. While Smets [37] provides some intuition on why such semantics might suit such functions, it lacks an operational definition[2], as is necessary for any uncertainty representation (see, *e.g.*, [13]). Moreover, if one is to accept the existence of this notion of "good reasons *not to believe*" and its associated mathematical representation, and one wants to model every possible state of belief involving "good reasons to believe" and "good reasons not to believe" in some propositions, then one actually needs to consider what Smets calls latent belief structures [37], which are couples of belief functions (representing respectively confidence and diffidence) that cannot be reduced in general to a single belief function. For instance, replacing the third piece of evidence in Example 1 *a third source provides the evidence* do not believe $\{y_1\}$ *and this source is given reliability* $3/4$ by *a third source provides the evidence* do not believe $\{y_2\}$ *and this source is given reliability* $3/4$ does not yield a mass function (one can easily check that $\{y_1, y_2\}^{1/2} \textcircled{\cap} \{y_1, y_3\}^{1/2} \textcircled{\oslash} \{y_2\}^{3/4}$ is not a mass function); even modifying only slightly the third piece of evidence to *a third source provides the evidence* do not believe $\{y_1\}$ *and this source is given reliability* 0.74 does not yield a mass function. A theory based on latent belief structures and involving inverse simple mass functions clearly goes beyond a theory based only on belief functions. A fully operationalised theory based on this richer mathematical object remains to be proposed.

This section has reviewed Smets' canonical decomposition and has also recalled Pichon *et al.*'s fusion scheme, which allows one to account explicitly for uncertain knowledge about source reliability. In the next section, we use this latter scheme to unveil a new canonical decomposition of belief functions. As will be seen, this new canonical decomposition relies only on well-defined concepts, contrary to Smets' decomposition.

3 A New Canonical Decomposition of Belief Functions

In this section, a specific connection between the multivariate Bernoulli distribution and belief functions is established and used in conjunction with Pichon *et al.*'s fusion scheme to lay the foundations for a new canonical decomposition of belief functions. Then, Teugels' representations of the multivariate Bernoulli distribution [42] are recalled and used together with the aforementioned connection to propose a new canonical decomposition of belief functions.

[2]Their relevance has also been questioned in [28].

3.1 Multivariate Bernoulli Distribution Induced Belief Function

In this section, one way is provided to induce from the multivariate Bernoulli distribution any belief function on domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$.

Let $\{X_i : i = 1, \ldots, n\}$ be a sequence of Bernoulli random variables with ranges $\mathcal{X}_i = \{0, 1\}$, $i = 1, \ldots, n$, *i.e.*, for $i = 1, \ldots, n$,

$$P(X_i = 1) = \pi_i, \ P(X_i = 0) = \xi_i,$$

where $0 \leq \pi_i = 1 - \xi_i \leq 1$. Recall that $\mathbb{E}[X_i] = \pi_i$.

Consider the multivariate Bernoulli distribution (MBD)

$$p_{k_1,\ldots,k_n} := P(X_1 = k_1, \ldots, X_n = k_n) \tag{22}$$

where $k_i \in \{0, 1\}$, $i = 1, \ldots, n$. Let k be an integer such that $1 \leq k \leq 2^n$. Using the binary expansion of k based on the k_i, $i = 1, \ldots, n$, that is the correspondence $k \leftrightarrow (k_1, \ldots, k_n)$, we can write

$$p_k = p_{k_1,\ldots,k_n}, \quad 1 \leq k \leq 2^n,$$

and represent a MBD p by the vector $\mathbf{p}$ containing 2^n components, with p_k its k-th component, $1 \leq k \leq 2^n$. This latter convenient notation for the MBD is borrowed from [42].

Definition 1 (MBD-equivalent)**.** *Let m be any mass function on $\mathcal{Y}$ and let p be the MBD* (22) *such that $p_k = m(A_k)$, $1 \leq k \leq 2^n$, with A_k the k-th subset of $\mathcal{Y}$ according to the binary order. p is called the* MBD-equivalent *of m.*

Accordingly, the vector $\mathbf{p}$ associated to the MBD-equivalent of a mass function m, is the same vector as $\mathbf{m}$.

In addition, for $i = 1, \ldots, n$, let Γ_i be a multi-valued mapping from $\mathcal{X}_i$ to $\mathcal{Y}$ such that

$$\begin{aligned} \Gamma_i(0) &= \mathcal{Y} \backslash \{y_i\} = \overline{\{y_i\}}, \\ \Gamma_i(1) &= \mathcal{Y}. \end{aligned} \tag{23}$$

Let $\mathcal{X}^n := \times_{i=1}^n \mathcal{X}_i$ and T be a multi-valued mapping from $\mathcal{X}^n$ to $\mathcal{Y}$ s.t.

$$\mathrm{T}(k) = \bigcap_{i=1}^{n} \Gamma_i(k_i), \tag{24}$$

for all $k \leftrightarrow (k_1, \ldots, k_n) \in \mathcal{X}^n$. Remark 1 provides a setting in which mapping T naturally occurs.

Remark 1. *Let $\mathcal{Y} = \{y_1, \ldots, y_n\}$ be the domain of a variable $\boldsymbol{y}$. Assume there are n sources of information $\mathfrak{s}_i$, $i = 1, \ldots, n$, with $\mathfrak{s}_i$ providing the following piece of information about $\boldsymbol{y}$: $\boldsymbol{y} \in \overline{\{y_i\}}$. Furthermore, an agent who receives these pieces of information, considers that each source can either be reliable or not reliable. Let $\mathcal{X}_i = \{0, 1\}$ be the space denoting the reliability of $\mathfrak{s}_i$, where 0 means that $\mathfrak{s}_i$ is reliable and 1 means that $\mathfrak{s}_i$ is not reliable. Following Pichon* et al.*'s fusion scheme recalled in Section 2.3, a multi-valued mapping Γ_i from $\mathcal{X}_i$ to $\mathcal{Y}$ can be defined as* (23)*; $\Gamma_i(k_i)$ interprets the testimony $\boldsymbol{y} \in \overline{\{y_i\}}$ in each configuration $k_i \in \mathcal{X}_i$ of the source $\mathfrak{s}_i$. More generally, a multi-valued mapping* T *from $\mathcal{X}^n$ to $\mathcal{Y}$ can be defined as* (24)*; $\mathrm{T}(k)$ interprets the testimonies $\boldsymbol{y} \in \overline{\{y_i\}}$, $i = 1, \ldots, n$, in each joint configuration $k \in \mathcal{X}^n$ of the sources $\mathfrak{s}_i$, $i = 1, \ldots, n$.*

Proposition 1. *Let m be any mass function on $\mathcal{Y}$ and let p be its MBD-equivalent. Then transferring p via the multi-valued mapping* T *defined by* (24) *induces m.*

Proof. Following [3], transferring distribution p via mapping T yields a mass function m' on $\mathcal{Y}$ defined, for $1 \leq k \leq 2^n$, by

$$m'(A_k) = \sum_{k' \in \mathcal{X}^n : \mathrm{T}(k') = A_k} p_{k'}.$$

For any $1 \leq k \leq 2^n$, we have

$$\begin{aligned}
\mathrm{T}(k) &= \bigcap_{i=1}^{n} \Gamma_i(k_i) \\
&= \left(\bigcap_{i:k_i=0} \Gamma_i(0) \right) \bigcap \left(\bigcap_{i:k_i=1} \Gamma_i(1) \right) \\
&= \bigcap_{i:k_i=0} \Gamma_i(0) \\
&= \bigcap_{i:k_i=0} \mathcal{Y} \backslash \{y_i\} \\
&= \mathcal{Y} \backslash \{y_i : i, k_i = 0\} \\
&= \{y_i : i, k_i = 1\} = A_k.
\end{aligned}$$

Hence $\sum_{k' \in \mathcal{X}^n : \mathrm{T}(k') = A_k} p_{k'} = p_k$ and thus m' is m. □

Example 2 illustrates Proposition 1.

Example 2 (Example 1 continued)**.** *Consider the same mass function m as in Example 1. This mass function can be equivalently written as*

$$m(A_4) = m(A_6) = m(A_8) = 1/3,$$

since $A_4 = \{y_1, y_2\}$, $A_6 = \{y_1, y_3\}$ and $A_8 = \{y_1, y_2, y_3\}$.

In addition, let p be the MBD-equivalent of m, i.e.*, the MBD defined by*

$$\begin{aligned} p_4 &= p_{110} = m(\{y_1, y_2\}), \\ p_6 &= p_{101} = m(\{y_1, y_3\}), \\ p_8 &= p_{111} = m(\{y_1, y_2, y_3\}), \end{aligned}$$

and $p_k = 0$ for $k = 1, 2, 3, 5, 7$.

We have

$$\begin{aligned} \mathrm{T}(4) &= \mathrm{T}[(1,1,0)] \\ &= \Gamma_1(1) \bigcap \Gamma_2(1) \bigcap \Gamma_3(0) \\ &= \mathcal{Y} \cap \mathcal{Y} \cap \overline{\{y_3\}} \\ &= \{y_1, y_2\} \end{aligned}$$

and, similarly, $\mathrm{T}(6) = \{y_1, y_3\}$ *and* $\mathrm{T}(8) = \{y_1, y_2, y_3\}$*. Hence, probability p_4 is transferred to* $\mathrm{T}(4) = \{y_1, y_2\} = A_4$*, probability p_6 to A_6 and probability p_8 to A_8,* i.e.*, the mass function m defined by* (15) *is recovered.*

Considering the setting of Remark 1, this means that mass function m can be viewed as resulting from having three sources $\mathfrak{s}_i$, each telling $\boldsymbol{y} \in \overline{\{y_i\}}$, $i = 1, 2, 3$, and such that the agent has the following knowledge on the reliability of the sources: with probability p_4, $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are not reliable and $\mathfrak{s}_3$ is reliable; with probability p_6, $\mathfrak{s}_1$ and $\mathfrak{s}_3$ are not reliable and $\mathfrak{s}_2$ is reliable; with probability p_8, the three sources are not reliable.

Let us finally state a result, which links the contour function of a mass function and the marginal expectations (called means in [42]) of its MBD-equivalent.

Lemma 1. *Let m be any mass function on $\mathcal{Y}$ and let p be its MBD-equivalent. The contour function of m is equal to the expectations of the r.v. $\{X_i : i = 1, \ldots, n\}$:*

$$pl(\{y_i\}) = \pi_i, \quad 1 \leq i \leq n.$$

Proof. For any $1 \leq i \leq n$, we have

$$\pi_i \quad = \quad \sum_{k:k_i=1} p_k$$

and

$$pl(\{y_i\}) \quad = \quad \sum_{k:k_i=1} m(A_k).$$

The lemma holds since $p_k = m(A_k)$ for all $1 \leq k \leq 2^n$, by definition of p. □

This section has provided a way to meaningfully induce any belief function from the MBD. This relation between the MBD and belief functions will be exploited in conjunction with Teugels' representations of the MBD [42] recalled in the next section, to introduce a new canonical decomposition of belief functions in Section 3.3.

3.2 Teugels' Representations of the MBD

In [42], Teugels considers two vectors associated to the MBD (22): the vector $\boldsymbol{\mu}$ and the vector $\boldsymbol{\sigma}$ defined respectively as

$$\begin{aligned} \boldsymbol{\mu} &= (\mu_1, \ldots, \mu_{2^n})', \\ \boldsymbol{\sigma} &= (\sigma_1, \ldots, \sigma_{2^n})', \end{aligned}$$

where, for $1 \leq k \leq 2^n$,

$$\begin{aligned} \mu_k &= \mathbb{E}\left[\prod_{i=1}^{n} X_i^{k_i}\right], \\ \sigma_k &= \mathbb{E}\left[\prod_{i=1}^{n} (X_i - \pi_i)^{k_i}\right], \end{aligned}$$

with k_i, $i = 1, \ldots, n$, the terms in the binary expansion of k. Let us remark that μ_k comes down to the (marginal) probability that each of the variables in $\{X_i : i, k_i = 1\}$ equals 1. This is not to be confused with p_k, which is the probability that each of the variables in $\{X_i : i, k_i = 1\}$ equals 1 *and* each of the variables in $\{X_i : i, k_i = 0\}$ equals 0. In addition, we note that vector $\boldsymbol{\mu}$ contains the moments of order 1 of the r.v. in $\{X_i : i = 1, \ldots, n\}$, and it contains also so-called product moments (or, joint moments) [41, 32] of all subsets of at least two r.v. in $\{X_i : i = 1, \ldots, n\}$. Following [42], $\boldsymbol{\mu}$ is simply referred to in this paper as the vector of *moments* of the MBD (22). Vector $\boldsymbol{\sigma}$ contains the central moments of order 1 of the r.v. in $\{X_i : i = 1, \ldots, n\}$, which

equal 0, as well as product (or, joint) central moments [41, 32] of all subsets of at least two r.v. in $\{X_i : i = 1, \ldots, n\}$, and in particular the covariances of all pairs of r.v. in $\{X_i : i = 1, \ldots, n\}$ (this will be illustrated later by Example 3). Following again [42], $\boldsymbol{\sigma}$ is simply referred to as the vector of *central moments* of the MBD (22).

Teugels [42] shows that interesting relations hold between vectors $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and the MBD vector $\mathbf{p}$. We reproduce his results for convenience:

Theorem 1 (Theorem 1 and p. 261 of [42])**.** *We have*

(i)

$$\mathbf{p} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right) \boldsymbol{\mu} \tag{25}$$

and

$$\boldsymbol{\mu} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) \mathbf{p}. \tag{26}$$

(ii)

$$\mathbf{p} = \begin{bmatrix} \xi_n & -1 \\ \pi_n & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \boldsymbol{\sigma} \tag{27}$$

and

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 & 1 \\ -\pi_n & \xi_n \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 1 \\ -\pi_1 & \xi_1 \end{bmatrix} \mathbf{p}. \tag{28}$$

(iii)

$$\boldsymbol{\mu} = \begin{bmatrix} 1 & 0 \\ \pi_n & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ \pi_1 & 1 \end{bmatrix} \boldsymbol{\sigma} \tag{29}$$

and

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 & 0 \\ -\pi_n & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ -\pi_1 & 1 \end{bmatrix} \boldsymbol{\mu}. \tag{30}$$

Of particular interest with respect to the new canonical decomposition of a belief function to be introduced in the next section, are relations (27) and (28); the former is illustrated by Example 3.

Example 3 (Example 1.2 of [42]). *Let $n = 3$. In this case, we have*

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \\ \sigma_7 \\ \sigma_8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \mathbb{E}\left[(X_1 - \pi_1)(X_2 - \pi_2)\right] \\ 0 \\ \mathbb{E}\left[(X_1 - \pi_1)(X_3 - \pi_3)\right] \\ \mathbb{E}\left[(X_2 - \pi_2)(X_3 - \pi_3)\right] \\ \mathbb{E}\left[(X_1 - \pi_1)(X_2 - \pi_2)(X_3 - \pi_3)\right] \end{bmatrix}.$$

σ_4 is nothing but the covariance between r.v. X_1 and X_2; accordingly we may denote σ_4 by $\sigma_{1,2}$. Similarly, we may denote σ_6 by $\sigma_{1,3}$, σ_7 by $\sigma_{2,3}$, and σ_8 by $\sigma_{1,2,3}$. Thus, we can write:

$$\boldsymbol{\sigma} = (1, 0, 0, \sigma_{1,2}, 0, \sigma_{1,3}, \sigma_{2,3}, \sigma_{1,2,3})'.$$

From (27)*, one obtains*

$$\mathbf{p} = \begin{bmatrix} p_{000} \\ p_{100} \\ p_{010} \\ p_{110} \\ p_{001} \\ p_{101} \\ p_{011} \\ p_{111} \end{bmatrix} = \left(\begin{bmatrix} \xi_3 & -1 \\ \pi_3 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_2 & -1 \\ \pi_2 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \right) \boldsymbol{\sigma}$$

$$= \begin{bmatrix} \xi_1\xi_2\xi_3 + \xi_3\sigma_{1,2} + \xi_2\sigma_{1,3} + \xi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \pi_1\xi_2\xi_3 - \xi_3\sigma_{1,2} - \xi_2\sigma_{1,3} + \pi_1\sigma_{2,3} + \sigma_{1,2,3} \\ \xi_1\pi_2\xi_3 - \xi_3\sigma_{1,2} + \pi_2\sigma_{1,3} - \xi_1\sigma_{2,3} + \sigma_{1,2,3} \\ \pi_1\pi_2\xi_3 + \xi_3\sigma_{1,2} - \pi_2\sigma_{1,3} - \pi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \xi_1\xi_2\pi_3 + \pi_3\sigma_{1,2} - \xi_2\sigma_{1,3} - \xi_1\sigma_{2,3} + \sigma_{1,2,3} \\ \pi_1\xi_2\pi_3 - \pi_3\sigma_{1,2} + \xi_2\sigma_{1,3} - \pi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \xi_1\pi_2\pi_3 - \pi_3\sigma_{1,2} - \pi_2\sigma_{1,3} + \xi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \pi_1\pi_2\pi_3 + \pi_3\sigma_{1,2} + \pi_2\sigma_{1,3} + \pi_1\sigma_{2,3} + \sigma_{1,2,3} \end{bmatrix}.$$

As detailed in [42, Section 2.3 (i)], both representations (25) and (27) contain as many parameters as $\mathbf{p}$, that is $2^n - 1$, since $\mu_1 = \sigma_1 = 1$. Specifically, in (27), $2^n - n - 1$

parameters are given by $\boldsymbol{\sigma}$, as $\sigma_k = 0$ when $k_1 + \cdots + k_n = 1$, and the n remaining parameters are π_i, $i = 1, \ldots, n$. The non-null components of $\boldsymbol{\sigma}$ represent the $2^n - n - 1$ possible dependencies between any subset (of at least two) of the r.v. $\{X_i : i = 1, \ldots, n\}$; under independence of all these r.v., we have $\boldsymbol{\sigma} = \mathbf{e}_1$. As a matter of fact, Teugels [42] refers to $\boldsymbol{\sigma}$ as the *dependency vector*. The central moments stored in $\boldsymbol{\sigma}$ may be particularly useful as illustrated by Teugels and Van Horebeek [43], who use them to formulate hypothesis tests about the interactions between treatments by several kinds of drugs.

Representation (27) can be insightfully depicted graphically using a Venn diagram[3], as illustrated in Figure 1 for the case $n = 3$:

- each Bernoulli random variable X_i underlying the MBD is represented by a circle (and more generally a closed curve for the cases $n > 3$);
- the dependency (central moment) σ_k between the variables $\{X_i : i, k_i = 1\}$ is in one-to-one correspondence with an overlap region, *i.e.*, the non-null components of $\boldsymbol{\sigma}$ are mapped to overlap regions;
- the mean π_i of each variable X_i is in one-to-one correspondence with a non overlap region, *i.e.*, the means π_i, $i = 1, \ldots, n$, are mapped to non overlap regions.

3.3 A New Canonical Decomposition of Belief Functions Based on Teugels' Representation of the MBD

It will be convenient to introduce the following definition associated with representation (27).

Definition 2 (Teugels' representation of the MDB)**.** *The Teugels' representation of the MBD* (22) *is the vector* $\boldsymbol{\tau}$ *of size* 2^n *such that, for* $1 \leq k \leq 2^n$*,*

$$\tau_k = \begin{cases} \pi_{\arg\max_{1 \leq i \leq n} k_i} & \text{if } k_1 + \cdots + k_n = 1, \\ \sigma_k & \text{otherwise.} \end{cases} \tag{31}$$

Definition 2 basically introduces a vector which stores the $2^n - 1$ parameters of representation (27), *i.e*, the (marginal) mean values π_i, $i = 1, \ldots, n$, and the non-null

[3] Although one must be aware that some individual (specifically overlap) regions can have positive or negative value.

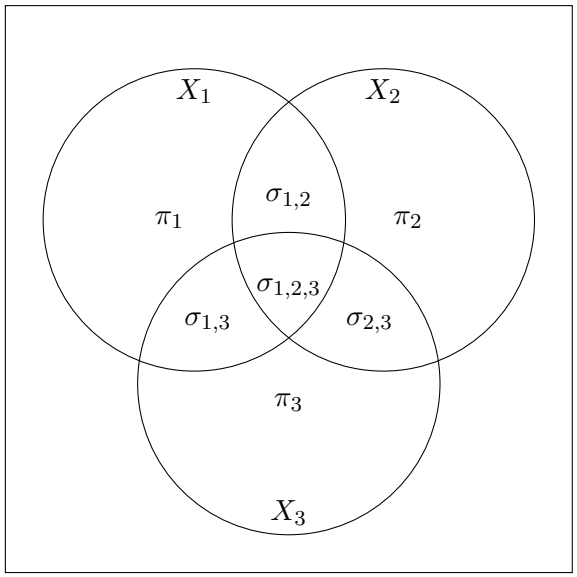

Figure 1: Venn diagram for representation (27) of the multivariate Bernoulli distribution, using the notation for $\boldsymbol{\sigma}$ introduced in Example 3.

components of $\boldsymbol{\sigma}$. Note that $\tau_1 = \sigma_1 = 1$. For instance, for $n = 2$, Equation (27) is

$$\begin{aligned}
\mathbf{p} &= \begin{bmatrix} \xi_2 & -1 \\ \pi_2 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \end{bmatrix} \\
&= \begin{bmatrix} \xi_2 & -1 \\ \pi_2 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \sigma_4 \end{bmatrix}
\end{aligned}$$

and it can be rewritten using $\boldsymbol{\tau}$ as

$$\mathbf{p} = \begin{bmatrix} 1-\tau_3 & -1 \\ \tau_3 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1-\tau_2 & -1 \\ \tau_2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \tau_4 \end{bmatrix}.$$

Proposition 1 and Definition 2 allow us to propose the following decomposition of a mass function into elementary items, which have well-defined semantics.

Definition 3 (t-canonical decomposition)**.** *Let m be a mass function defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ and let τ be the Teugels' representation of its MBD-equivalent.*

Its t-canonical decomposition (or t-decomposition for short) is the mapping t on $2^{\mathcal{Y}}\backslash\{\emptyset\}$, called the Teugels function and defined by:

$$t(A_k) \quad = \quad \tau_k, \quad 1 < k \leq 2^n, \tag{32}$$

with A_k the k-th subset of $\mathcal{Y}$ according to the binary order.

The t-canonical decomposition of a mass function m is thus basically Teugels' representation of its MBD-equivalent. It is unique since the Teugels function t is in one-to-one correspondence with function m (function τ is in such correspondence with the MBD-equivalent p of m and the vector $\mathbf{p}$ is equal to the vector $\mathbf{m}$).

The Teugels function t can be computed from Eqs. (32), (31), (28) and Lemma 1, which comes down to the following definition for t, for all $A \in 2^{\mathcal{Y}}\backslash\{\emptyset\}$:

$$t(A) \quad = \quad \begin{cases} pl(A), & \text{if } |A| = 1, \\ \left(\begin{bmatrix} 1 & 1 \\ -pl(\{y_n\}) & 1 - pl(\{y_n\}) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 1 \\ -pl(\{y_1\}) & 1 - pl(\{y_1\}) \end{bmatrix} \mathbf{m} \right)(A), & \text{otherwise.} \end{cases} \tag{33}$$

Adopting the setting of Remark 1, the t-decomposition of a mass function can be interpreted in the same vein as Smets' decomposition, *i.e.*, a mass function can be viewed as resulting from partially reliable sources providing crisp pieces of information. More precisely, any mass function m can be recovered from the following basic components:

- crisp testimonies $\boldsymbol{y} \in \overline{\{y_i\}}$ provided by sources $\mathfrak{s}_i$, $i = 1, \ldots, n$;
- knowledge on the individual (marginal) reliability of each source $\mathfrak{s}_i$, represented by the mean $t(\{y_i\})$ (which happens to be given by the contour function, *i.e.*, $t(\{y_i\}) = pl(\{y_i\})$);
- knowledge on the dependency between the source reliabilities for each subset of sources $\{\mathfrak{s}_i : i, k_i = 1\}$, represented by the central moment $t(A_k)$, $|A_k| > 1$, with k_i, $i = 1, \ldots, n$, the terms in the binary expansion of k.

This is illustrated by Examples 4 and 5.

Example 4. *Let m be a mass function about a variable $\boldsymbol{y}$ defined on $\mathcal{Y} = \{y_1, y_2\}$, with associated Teugels function t. From* (33) *we obtain*

$$\begin{aligned}
t(\{y_1\}) &= pl(\{y_1\}) \\
&= m(\{y_1\}) + m(\{y_1, y_2\}), \qquad (34) \\
t(\{y_2\}) &= pl(\{y_2\}) \\
&= m(\{y_2\}) + m(\{y_1, y_2\}), \qquad (35) \\
t(\{y_1, y_2\}) &= \begin{bmatrix} pl(\{y_2\})pl(\{y_1\}) \\ -pl(\{y_2\})(1 - pl(\{y_1\})) \\ -pl(\{y_1\})(1 - pl(\{y_2\})) \\ (1 - pl(\{y_1\}))(1 - pl(\{y_2\})) \end{bmatrix}' \mathbf{m} \\
&= m(\{y_1, y_2\})m(\emptyset) - m(\{y_1\})m(\{y_2\}). \qquad (36)
\end{aligned}$$

Let us now consider the following pieces of evidence:

- *There are two sources $\mathfrak{s}_1$ and $\mathfrak{s}_2$, with $\mathfrak{s}_i$ providing the crisp piece of information $\boldsymbol{y} \in \overline{\{y_i\}}$,* i.e., *$\mathfrak{s}_1$ tells $\boldsymbol{y} \in \{y_2\}$ and $\mathfrak{s}_2$ tells $\boldsymbol{y} \in \{y_1\}$;*
- *$\mathfrak{s}_1$ and $\mathfrak{s}_2$ are believed to be non reliable with probabilities $t(\{y_1\})$ and $t(\{y_2\})$ respectively;*
- *The dependence between their reliabilities is given by covariance $t(\{y_1, y_2\})$.*

These latter two pieces of evidence yield, using (27)*, the following knowledge about the reliability of the sources*

$$\begin{aligned}
P(X_1 = 0, X_2 = 0) &= (1 - t(\{y_1\}))(1 - t(\{y_2\})) + t(\{y_1, y_2\}), \\
P(X_1 = 1, X_2 = 0) &= t(\{y_1\})(1 - t(\{y_2\})) - t(\{y_1, y_2\}), \\
P(X_1 = 0, X_2 = 1) &= (1 - t(\{y_1\}))t(\{y_2\}) - t(\{y_1, y_2\}), \\
P(X_1 = 1, X_2 = 1) &= t(\{y_1\})t(\{y_2\}) + t(\{y_1, y_2\}).
\end{aligned}$$

Besides, the mapping T *representing the interpretations of the source testimonies in*

each of their joint configurations in terms of reliability is:

$$\begin{aligned}
\mathrm{T}[(0,0)] &= \Gamma_1(0) \cap \Gamma_2(0) \\
&= \{y_2\} \cap \{y_1\} \\
&= \emptyset, \\
\mathrm{T}[(1,0)] &= \Gamma_1(1) \cap \Gamma_2(0) \\
&= \mathcal{Y} \cap \{y_1\} \\
&= \{y_1\}, \\
\mathrm{T}[(0,1)] &= \{y_2\}, \\
\mathrm{T}[(1,1)] &= \mathcal{Y}.
\end{aligned}$$

Using Eqs. (34), (35) *and* (36), *it can easily be checked that*

$$\begin{aligned}
(1 - t(\{y_1\}))(1 - t(\{y_2\})) + t(\{y_1, y_2\}) &= m(\emptyset), \\
t(\{y_1\})(1 - t(\{y_2\})) - t(\{y_1, y_2\}) &= m(\{y_1\}, \\
(1 - t(\{y_1\}))t(\{y_2\}) - t(\{y_1, y_2\}) &= m(\{y_2\}, \\
t(\{y_1\})t(\{y_2\}) + t(\{y_1, y_2\}) &= m(\{y_1, y_2\}.
\end{aligned}$$

Mass function m is thus clearly recovered when transferring P to $\mathcal{Y}$ via T. *Remark also that the greater the covariance $t(\{y_1, y_2\})$ is, the greater are the probabilities $P(X_1 = 1, X_2 = 1)$ and $P(X_1 = 0, X_2 = 0)$ that the sources are jointly non reliable and that they are jointly reliable, respectively (and the smaller are the probabilities $P(X_1 = 0, X_2 = 1)$ and $P(X_1 = 1, X_2 = 0)$ that the first source is reliable whereas the second is not and that the first source is not reliable whereas the second is, respectively). Hence, the greater this covariance is, the greater are the masses on $\mathcal{Y}$ and on $\emptyset$ (and the smaller are the masses on $\{y_1\}$ and on $\{y_2\}$).*

Example 5 (Example 2 continued)**.** *Let us compute the Teugels function t associated to the mass function m of Example 2. We find using* (33)

$$\begin{aligned}
t(\{y_1\}) &= pl(\{y_1\}) = 1, \\
t(\{y_2\}) &= pl(\{y_2\}) = 2/3, \\
t(\{y_3\}) &= pl(\{y_3\}) = 2/3,
\end{aligned}$$

and since (using $\pi_i = pl(\{y_i\})$ and $\xi_i = 1 - \pi_i$, $1 \leq i \leq 3$, to shorten expressions)

$$\begin{bmatrix} 1 & 1 \\ -pl(\{y_3\}) & 1 - pl(\{y_3\}) \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ -pl(\{y_2\}) & 1 - pl(\{y_2\}) \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ -pl(\{y_1\}) & 1 - pl(\{y_1\}) \end{bmatrix} \mathbf{m}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\pi_1 & \xi_1 & -\pi_1 & \xi_1 & -\pi_1 & \xi_1 & -\pi_1 & \xi_1 \\ -\pi_2 & -\pi_2 & \xi_2 & \xi_2 & -\pi_2 & -\pi_2 & \xi_2 & \xi_2 \\ \pi_2\pi_1 & -\pi_2\xi_1 & -\pi_1\xi_2 & \xi_2\xi_1 & \pi_2\pi_1 & -\pi_2\xi_1 & -\pi_1\xi_2 & \xi_2\xi_1 \\ -\pi_3 & -\pi_3 & -\pi_3 & -\pi_3 & \xi_3 & \xi_3 & \xi_3 & \xi_3 \\ \pi_3\pi_1 & -\pi_3\xi_1 & \pi_3\pi_1 & -\pi_3\xi_1 & -\pi_1\xi_3 & \xi_3\xi_1 & -\pi_1\xi_3 & \xi_3\xi_1 \\ \pi_3\pi_2 & \pi_3\pi_2 & -\pi_3\xi_2 & -\pi_3\xi_2 & -\pi_2\xi_3 & -\pi_2\xi_3 & \xi_3\xi_2 & \xi_3\xi_2 \\ -\pi_1\pi_3\pi_2 & \pi_3\pi_2\xi_1 & \pi_1\pi_3\xi_2 & -\pi_3\xi_2\xi_1 & \pi_1\pi_2\xi_3 & -\pi_2\xi_3\xi_1 & -\pi_1\xi_3\xi_2 & \xi_3\xi_2\xi_1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ m(\{y_1, y_2\}) \\ 0 \\ m(\{y_1, y_3\}) \\ 0 \\ m(\{y_1, y_2, y_3\}) \end{bmatrix},$$

we find

$$\begin{aligned}
t(\{y_1, y_2\}) &= m(\{y_1, y_2\})\xi_2\xi_1 - m(\{y_1, y_3\})\pi_2\xi_1 + m(\{y_1, y_2, y_3\}\xi_2\xi_1 = 0 \text{ (since } \xi_1 = 0\text{)}, \\
t(\{y_1, y_3\}) &= -m(\{y_1, y_2\})\pi_3\xi_1 - m(\{y_1, y_3\})\xi_3\xi_1 + m(\{y_1, y_2, y_3\}\xi_3\xi_1 = 0 \text{ (since } \xi_1 = 0\text{)}, \\
t(\{y_2, y_3\}) &= -m(\{y_1, y_2\})\pi_3\xi_2 - m(\{y_1, y_3\})\pi_2\xi_3 + m(\{y_1, y_2, y_3\}\xi_3\xi_2 \\
&= -(1/3)(2/3)(1/3) - (1/3)(2/3)(1/3) + (1/3)(1/3)(1/3) = -1/9, \\
t(\{y_1, y_2, y_3\}) &= -m(\{y_1, y_2\})\pi_3\xi_2\xi_1 - m(\{y_1, y_3\})\pi_2\xi_3\xi_1 + m(\{y_1, y_2, y_3\}\xi_3\xi_2\xi_1 = 0 \text{ (since } \xi_1 = 0\text{)}.
\end{aligned}$$

In other words, considering as in Example 2 the setting of Remark 1, mass function m can be seen as originating from the following pieces of evidence:

- *There are three sources sources $\mathfrak{s}_i$, each telling $\boldsymbol{y} \in \overline{\{y_i\}}$, $i = 1, 2, 3$;*
- *$\mathfrak{s}_1$, $\mathfrak{s}_2$ and $\mathfrak{s}_3$ are believed to be non reliable with probabilities 1, 2/3 and 2/3 respectively;*
- *The covariance between the reliabilities of $\mathfrak{s}_2$ and $\mathfrak{s}_3$ is $-1/9$, and all other subsets of variables representing the reliabilities of the sources have null central moments.*

Note that this is clearly a different decomposition to that of Smets provided for the same mass function in Example 1.

Since it relies on representation (27) of the MBP, the t-decomposition can also be represented by a Venn diagram. For instance, the pieces of evidence underlying mass function m in Example 5 are shown in Figure 2.

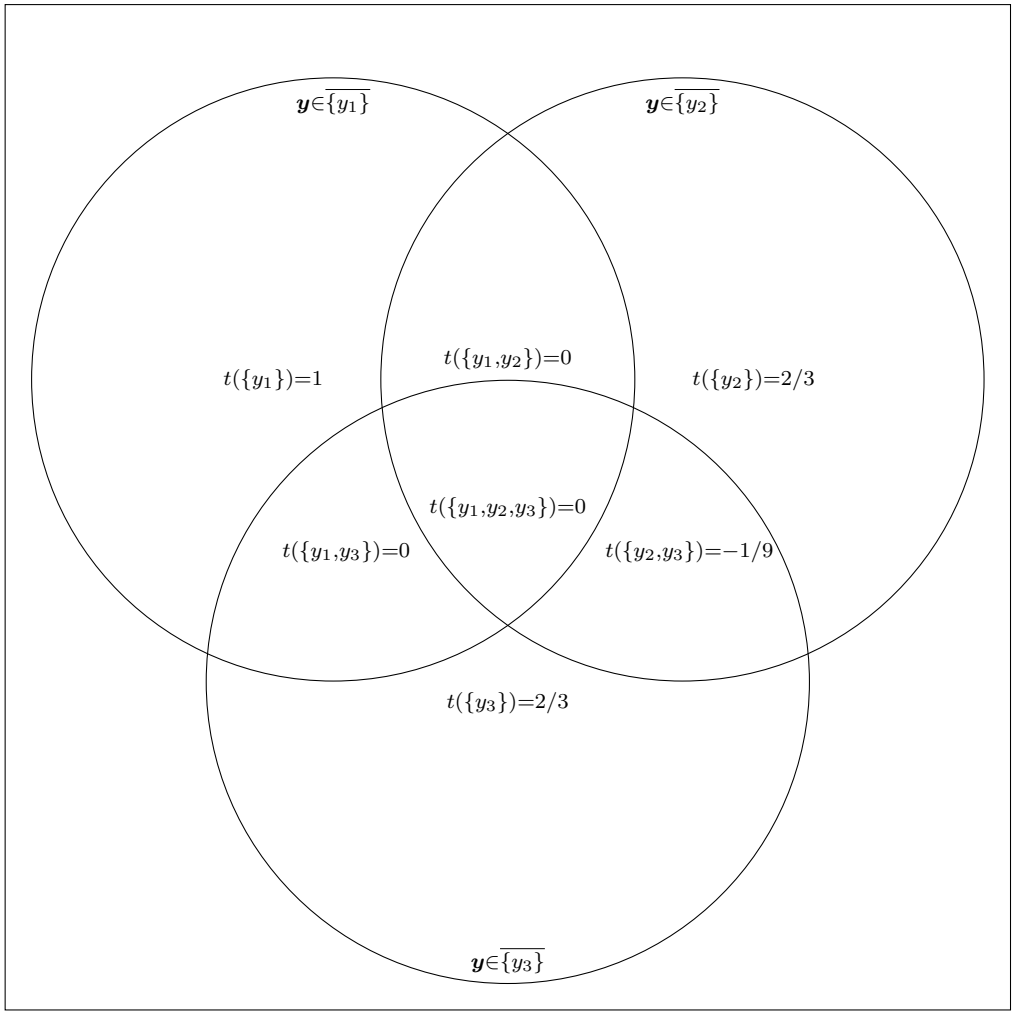

Figure 2: t-decomposition of mass function m in Example 5, with $t(\{y_i\})$ the marginal probability that source $\mathfrak{s}_i$ is not reliable, $t(A_k)$, $|A_k| > 1$, the central moment between the reliabilities of the subset of sources $\{\mathfrak{s}_i : i, k_i = 1\}$, and $\boldsymbol{y} \in \overline{\{y_i\}}$ the testimony provided by source $\mathfrak{s}_i$.

Up until now, the Teugels function t has been computed using (33), where t is obtained from the contour function and the mass function. However, as will be explained in the remainder of this section, t can also conveniently be obtained using only the commonality function.

Theorem 1 leads to the following straightforward remark.

Remark 2. *Let m be a mass function on $\mathcal{Y}$ and p its MBD-equivalent. By comparing Eqs.* (26) *and* (5)*, it is clear that the vector $\boldsymbol{\mu}$ of moments of p is equal to the commonality vector $\mathbf{q}$ associated to m,* i.e.*, $\mu_k = q(A_k)$ for all $1 \leq k \leq 2^n$.*

Adopting the setting of Remark 1 and using Remark 2, a new interpretation is obtained for the commonality function: $q(A_k)$ is the moment between the random variables $\{X_i : i, k_i = 1\}$, representing the reliabilities of sources $\{\mathfrak{s}_i : i, k_i = 1\}$. More simply, $q(A_k)$ is the marginal probability that each of the sources in $\{\mathfrak{s}_i : i, k_i = 1\}$ is not reliable.

In particular, Remark 2 yields an alternative and direct proof for Lemma 1. Indeed, for k such that $k_1 + \cdots + k_n = 1$, we have

$$\begin{aligned} q(A_k) &= \mathbb{E}\left[X_{\arg\max_{1\leq i\leq n} k_i}\right] \\ &= \pi_{\arg\max_{1\leq i\leq n} k_i}, \end{aligned}$$

and thus $q(\{y_i\}) = \pi_i$, for all $1 \leq i \leq n$.

In addition, from Lemma 1 and Eqs. (32), (31) and (30), the following equivalent definition based only on the commonality function is obtained for t:

$$t(A) = \begin{cases} q(A), & \text{if } |A| = 1, \\ \left(\begin{bmatrix} 1 & 0 \\ -q(\{y_n\}) & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ -q(\{y_1\}) & 1 \end{bmatrix} \mathbf{q}\right)(A), & \text{otherwise.} \end{cases} \tag{37}$$

For instance, for $n = 2$, the Teugels function t associated to a mass function with commonality function q is

$$\begin{aligned} t(\{y_1\}) &= q(\{y_1\}), \\ t(\{y_2\}) &= q(\{y_2\}), \\ t(\{y_1, y_2\}) &= \begin{bmatrix} q(\{y_1\})q(\{y_2\}) & -q(\{y_2\}) & -q(\{y_1\}) & 1 \end{bmatrix} \mathbf{q} \\ &= q(\{y_1, y_2\} - q(\{y_1\})q(\{y_2\}). \end{aligned}$$

This section has proposed a new solution to decompose a belief function (Definition 3). This solution relies on the following building blocks: (i) the notion of MBD-equivalent of a belief function (Definition 1) and a particular multi-valued mapping allowing one to recover a belief function from its MBD-equivalent (Proposition 1), (ii) a setting that provides a meaning to this connection between the MBD and belief functions (Remark 1), and (iii) a particular representation of the MBD (Definition 2). Next section provides some comments on this solution.

4 Some Comments on the t-Canonical Decomposition

In this section, some comments are provided on the t-canonical decomposition. First, we study a particular case of the general approach proposed in [11] for the conjunctive combination of bodies of evidence with known dependence structure. Then, this study is used to express the t-canonical decomposition in terms of simple mass functions, similarly as Smets' decomposition. Next, the t-canonical decomposition is compared further with Smets' decomposition. Finally, the t-canonical decomposition is considered in the context of random sets.

4.1 Conjunctive combination with known dependence structure

The general approach proposed in [11] for the conjunctive combination of bodies of evidence represented by mass functions, allows for other dependence structures among them besides independence. This approach is the following. Let $m_1, \ldots, m_N$ be N mass functions on $\mathcal{Y}$. The approach [11] defines a mass function $m_\cap$ on $\mathcal{Y}$ resulting from a conjunctive combination of $m_1, \ldots, m_N$ as the result of the following procedure:

1. A so-called joint mass function $jm : \times_{i=1}^{N} 2^{\mathcal{Y}} \rightarrow [0,1]$ is built, preserving $m_1, \ldots, m_N$ as marginals, which means that $\forall A_i \in \mathcal{F}_i$, with $\mathcal{F}_i$ the set of focal sets of m_i,

$$m_i(A_i) \quad = \sum_{A_1 \in \mathcal{F}_1, \ldots, A_{i-1} \in \mathcal{F}_{i-1}, A_{i+1} \in \mathcal{F}_{i+1}, \ldots, A_N \in \mathcal{F}_N} jm(A_1, \ldots, A_{i-1}, A_i, A_{i+1}, \ldots, A_N). \tag{38}$$

2. Each joint mass $jm(A_1, \ldots, A_N)$ is allocated to the subset $\bigcap_{i=1}^{N} A_i$ in the final mass function $m_\cap$, *i.e.*, for all $A \subseteq \mathcal{Y}$

$$m_\cap(A) \quad = \sum_{\bigcap_{i=1}^{N} A_i = A} jm(A_1, \ldots, A_N).$$

As noted in [11], Eq. (38) indicates that the information m_i representing each body of evidence can be retrieved from the richer information jm that includes a representation of their mutual dependence. In particular, the combination of $m_1, \ldots, m_N$ by the conjunctive rule $\textcircled{\cap}$ is a particular case of this approach, retrieved for $jm(A_1, \ldots, A_N) = m_1(A_1)m_2(A_2)\ldots m_N(A_N)$ in Step 1; this latter equality corresponds to the assumption that the bodies of evidence represented by mass functions $m_1, \ldots, m_N$ are independent.

Let us now consider a particular case of this general approach where each mass function m_i has only two focal sets, denoted A_{i_0} and A_{i_1}, for some $A_{i_0}, A_{i_1} \subseteq \mathcal{Y}$, such that $m(A_{i_1}) = \pi_i$ and $m(A_{i_0}) = 1 - \pi_i$ for some $\pi_i \in [0, 1]$. Such kind of mass function will be called *binary* hereafter and may be simply denoted $(A_{i_0}, A_{i_1})^{\pi_i}$.

When each mass function m_i is binary and such that $m_i = (A_{i_0}, A_{i_1})^{\pi_i}$, it is clear that only subsets $\mathbf{A}_k := (A_{1_{k_1}}, A_{2_{k_2}}, \ldots, A_{N_{k_N}}) \subseteq \times_{i=1}^{N} \mathcal{Y}$, $1 \leq k \leq 2^N$ with $k \leftrightarrow (k_1, \ldots, k_N) \in \{0, 1\}^N$, can receive a non null joint mass in jm. Moreover, by associating a Bernoulli r.v. X_i to each m_i, such that $P(X_i = 1) = m_i(A_{i_1}) = \pi_i$ and $P(X_i = 0) = m_i(A_{i_0}) = 1 - \pi_i$, we can establish a one-to-one correspondence between the MBD with underlying r.v. X_i, $i = 1, \ldots, N$, and the joint mass function jm, by setting

$$P(X_1 = k_1, \ldots, X_N = k_N) = jm(A_{1_{k_1}}, A_{2_{k_2}}, \ldots, A_{N_{k_N}}),$$

or for short, using $k \leftrightarrow (k_1, \ldots, k_N)$, $p_k = jm(\mathbf{A}_k)$ with $p_k := P(X_1 = k_1, \ldots, X_N = k_N)$. From Theorem 1, we have that the joint mass function jm can be fully specified by parameters π_i, $i = 1, \ldots, N$, and vector $\boldsymbol{\sigma}$ associated with its corresponding MBD. Most interestingly, the dependence structure encoded in jm between the binary mass functions m_i is actually entirely captured by vector $\boldsymbol{\sigma}$, since any MBD (and thus any corresponding joint mass function) with given marginals is obtained for a unique vector $\boldsymbol{\sigma}$. More specifically, since σ_k, $k_1 + \cdots + k_N > 1$, represents the dependency among variables $\{X_i : i, k_i = 1\}$, and since variable X_i is associated to mass function m_i, $i = 1, \ldots, N$, we may regard σ_k as capturing the dependency among mass functions $\{m_i : i, k_i = 1\}$.

Since the conjunctive combination of N binary mass functions with dependence structure represented by a joint mass function jm, is completely determined by vector $\boldsymbol{\sigma}$ associated to jm, then this combination can be expressed as a parameterised combination rule for binary mass functions, with parameter $\boldsymbol{\sigma}$ representing the dependence structure. Formally, let $\mathcal{B}$ denote the set of all binary mass functions on $\mathcal{Y}$ and $\mathcal{M}$ the set of all mass functions on $\mathcal{Y}$. In addition, let $m_i = (A_{i_0}, A_{i_1})^{\pi_i}$, $i = 1, \ldots, N$, be N

binary mass functions. Then, we refer to the operator $\textcircled{\cap}_{\boldsymbol{\sigma}} : \mathcal{B}^N \to \mathcal{M}$ such that

$$\textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, \ldots, m_N) = \textcircled{\cap}_{\boldsymbol{\sigma}}\left((A_{1_0}, A_{1_1})^{\pi_1}, \ldots, (A_{N_0}, A_{N_1})^{\pi_N}\right) := m_{\cap},$$

with $m_{\cap}$ the result of the conjunctive combination of mass functions m_i with dependence represented by the joint mass function jm determined by vector $\boldsymbol{\sigma}$ and having marginals m_i, $i = 1, \ldots, N$, as the *conjunctive combination with dependence* $\boldsymbol{\sigma}$ (or, for short, $\boldsymbol{\sigma}$-*conjunctive combination*) of these mass functions. This is illustrated by Example 6.

Example 6. *Let* $m_1 = (A_{1_0}, A_{1_1})^{\pi_1}$ *and* $m_2 = (A_{2_0}, A_{2_1})^{\pi_2}$ *be two binary mass functions on* $\mathcal{Y} = \{y_1, y_2, y_3\}$ *such that* $A_{1_0} = \{y_2\}, A_{1_1} = \mathcal{Y}, \pi_1 = 0.4$, *and* $A_{2_0} = \{y_2, y_3\}, A_{2_1} = \{y_1, y_2\}, \pi_2 = 0.5$. *In other words,* $m_1 = (\{y_2\}, \mathcal{Y})^{0.4}$ *and* $m_2 = (\{y_2, y_3\}, \{y_1, y_2\})^{0.5}$. *Let their dependence structure be represented by the joint mass function* jm *such that*

$$\begin{aligned} jm(\mathbf{A}_1) &= jm(\{y_2\}, \{y_2, y_3\}) = 0.4, \\ jm(\mathbf{A}_2) &= jm(\mathcal{Y}, \{y_2, y_3\}) = 0.1, \\ jm(\mathbf{A}_3) &= jm(\{y_2\}, \{y_1, y_2\}) = 0.2, \\ jm(\mathbf{A}_4) &= jm(\mathcal{Y}, \{y_1, y_2\}) = 0.3. \end{aligned}$$

One can easily check that jm *satisfies* (38) *for* $i = 1, 2$.

Performing Step 2 of the conjunctive combination of m_1 *and* m_2, *we obtain:*

$$\begin{aligned} m_{\cap}(\{y_2\}) &= jm(\{y_2\}, \{y_2, y_3\}) + jm(\{y_2\}, \{y_1, y_2\}) = 0.6, \\ m_{\cap}(\{y_2, y_3\}) &= jm(\mathcal{Y}, \{y_2, y_3\}) = 0.1, \\ m_{\cap}(\{y_1, y_2\}) &= jm(\mathcal{Y}, \{y_1, y_2\}) = 0.3. \end{aligned}$$

Now, the MBD in one-to-one correspondence with jm *is*

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= p_1 = jm(\{y_2\}, \{y_2, y_3\}), \\ P(X_1 = 1, X_2 = 0) &= p_2 = jm(\mathcal{Y}, \{y_2, y_3\}), \\ P(X_1 = 0, X_2 = 1) &= p_3 = jm(\{y_2\}, \{y_1, y_2\}), \\ P(X_1 = 1, X_2 = 1) &= p_4 = jm(\mathcal{Y}, \{y_1, y_2\}). \end{aligned}$$

The vector $\boldsymbol{\sigma}$ *associated to this MBD is, using* (28),

$$\begin{aligned} \boldsymbol{\sigma} &= \begin{bmatrix} 1 & 1 \\ -\pi_2 & 1 - \pi_2 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ -\pi_1 & 1 - \pi_1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} \\ &= (1, 0, 0, 0.1)' \end{aligned}$$

From the definition of $\textcircled{\cap}_{\boldsymbol{\sigma}}$*, we have then*

$$m_{\cap} = \textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, m_2) = \textcircled{\cap}_{(1,0,0,0.1)'} \left((\{y_2\}, \mathcal{Y})^{0.4}, (\{y_2, y_3\}, \{y_1, y_2\})^{0.5} \right).$$

The operation $\textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, \ldots, m_N)$ is well-defined as long as there exists a joint mass function jm compatible with vector $\boldsymbol{\sigma}$ and binary mass functions m_i.

In addition, it is clear that the case where the bodies of evidence, represented by binary mass functions $m_1, \ldots, m_N$, are independent, comes down to $\boldsymbol{\sigma} = \mathbf{e}_1$, which means that the following equality holds:

$$\textcircled{\cap}_{\mathbf{e}_1}(m_1, \ldots, m_N) = m_1 \textcircled{\cap} \ldots \textcircled{\cap} m_N,$$

that is, the conjunctive rule is a particular case of the $\boldsymbol{\sigma}$-conjunctive rule, recovered for $\boldsymbol{\sigma} = \mathbf{e}_1$.

Finally, note that $\boldsymbol{\sigma}$-conjunctive combination $\textcircled{\cap}_{\boldsymbol{\sigma}} \left((A_{1_0}, A_{1_1})^{\pi_1}, \ldots, (A_{N_0}, A_{N_1})^{\pi_N} \right)$ of N binary mass functions $(A_{i_0}, A_{i_1})^{\pi_i}$ may be denoted for short $\textcircled{\cap}_{\boldsymbol{\sigma}} \left(A_{1_0}^{\pi_1}, \ldots, A_{N_0}^{\pi_N} \right)$ when $A_{i_1} = \mathcal{Y}$, $i = 1, \ldots, N$, since a binary mass function $(A_{i_0}, A_{i_1})^{\pi_i}$ such that $A_{i_1} = \mathcal{Y}$ is nothing but the simple mass function $A_{i_0}^{\pi_i}$.

4.2 t-Canonical Decomposition in Terms of Simple Mass Functions

In Section 2.3, it was recalled that when some sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$ provide crisp testimonies $\mathbf{A} = (A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N})$, and when knowledge about the source reliability is uncertain such that each joint state $k \in \mathcal{X}^N$ has a probability p_k, then the induced knowledge of the agent about $\mathcal{Y}$ is represented by a mass function m defined by

$$m(B) \quad = \quad \sum_{k : \Gamma_{\mathbf{A}}(k) = B} p_k, \quad \forall B \subseteq \mathcal{Y}. \tag{39}$$

This induced knowledge may be seen as the 'trace' on $\mathcal{Y}$ of the available pieces of information, that is, of the source testimonies and of the knowledge about their reliability. Besides, let us note that the vector $\boldsymbol{\sigma}$, obtained using (28) from the MBD such that $P(X_1 = k_1, \ldots, X_N = k_N) = p_k$ with X_i representing the reliability of $\mathfrak{s}_i$, represents the meta-dependences among the sources, that is, the dependences (in terms of central moments) between the reliabilities of all subsets of (at least two) sources.

Considering solely the testimony of source $\mathfrak{s}_i$ and what is known of its reliability, *i.e.*, it is not reliable with marginal probability π_i given by (20), then the induced knowledge on $\mathcal{Y}$ given this testimony is represented according to (16) by the simple mass

function $A^{\pi_i}_{\mathfrak{S}_i}$. This latter mass function is the trace on $\mathcal{Y}$ of the pieces of information pertaining only to $\mathfrak{s}_i$, that is, of the testimony of $\mathfrak{s}_i$ and of what is known of its reliability. Accordingly, it will be referred to as the individual (or, marginal) trace associated with $\mathfrak{s}_i$.

Theorem 2. *The trace m defined by* (39) *verifies*

$$m = \circledcirc_{\boldsymbol{\sigma}}(A^{\pi_1}_{\mathfrak{S}_1}, \ldots, A^{\pi_N}_{\mathfrak{S}_N}), \tag{40}$$

with $A^{\pi_i}_{\mathfrak{S}_i}$ the marginal trace associated with $\mathfrak{s}_i$ and $\boldsymbol{\sigma}$ the vector obtained using (28) *from the MBD such that $P(X_1 = k_1, \ldots, X_N = k_N) = p_k$ with X_i representing the reliability of $\mathfrak{s}_i$.*

Proof. Since $\Gamma_{\mathbf{A}}(k) = \bigcap_{i=1}^{N} \Gamma_{A_{\mathfrak{S}_i}}(k_i)$, the mass function m defined by (39) is recovered by considering a joint mass function jm defined by

$$jm(\Gamma_{A_{\mathfrak{S}_1}}(k_1), \ldots, \Gamma_{A_{\mathfrak{S}_N}}(k_N)) = p_k,$$

and by allocating the joint mass $jm(\Gamma_{A_{\mathfrak{S}_1}}(k_1), \ldots, \Gamma_{A_{\mathfrak{S}_N}}(k_N))$ to $\bigcap_{i=1}^{N} \Gamma_{A_{\mathfrak{S}_i}}(k_i)$. Besides, it is direct to see that the marginals of jm are the marginal traces $A^{\pi_i}_{\mathfrak{S}_i}$, $i = 1, \ldots, N$ associated with sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$.

Hence, the mass function m defined by (39) may be obtained as the conjunctive combination of elementary pieces of evidence represented by the simple mass functions $A^{\pi_i}_{\mathfrak{S}_i}$ with dependence structure jm, or, equivalently, we have

$$m = \circledcirc_{\boldsymbol{\sigma}}(A^{\pi_1}_{\mathfrak{S}_1}, \ldots, A^{\pi_N}_{\mathfrak{S}_N}),$$

since the MBD in one-to-one correspondence with jm is formally equivalent to the one representing the meta-knowledge about the sources and thus the dependence structure in jm is represented by $\boldsymbol{\sigma}$. □

The fact that the mass function given by (39) can be equivalently obtained with (40) corresponds to the intuition. It means that a situation where some sources provide crisp testimonies $A_{\mathfrak{S}_i}$, are assumed to be marginally non reliable with probabilities π_i and to have a meta-dependence structure $\boldsymbol{\sigma}$, can be equivalently regarded with respect to the knowledge it induces on $\mathcal{Y}$, as a situation where one has accumulated elementary pieces of evidence, represented by the marginal traces $A^{\pi_i}_{\mathfrak{S}_i}$, and one assumes that these pieces of evidence have a dependence structure $\boldsymbol{\sigma}$ among them. In particular, we remark that the

dependence structure at the meta level 'flows down' to the trace level, and specifically $\sigma_k,\ k_1 + \cdots + k_N > 1$, which represents the dependency among the marginal traces $\{A^{\pi_i}_{\mathfrak{s}_i} : i, k_i = 1\}$ as explained in Section 4.1, is thus nothing but the inheritance of the dependency σ_k between the reliability of the sources $\{\mathfrak{s}_i : i, k_i = 1\}$ that are at the origin of these marginal traces.

An important particular case of this is when the dependence structure at the meta level is that of independence, which corresponds to $\boldsymbol{\sigma} = \mathbf{e}_1$ (in which case Eq. (40) reduces to (21)). Indeed, in such a case where the sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$ are meta-independent, then the induced knowledge on $\mathcal{Y}$ is given by (21), as explained in Section 2.3. Yet, according to the conjunctive combination approach recalled in Section 4.1, m defined by (21) is also obtained if one receives N elementary pieces of evidence represented by simple mass functions $A^{\pi_i}_{\mathfrak{s}_i}$, $i = 1, \ldots, N$, *i.e.*, one receives the marginal traces associated with the sources, and assumes that these pieces of evidence are independent. That is, one obtains m by assuming the same dependence structure at the trace level than that at the meta level.

Theorem 2 is further illustrated by Example 7.

Example 7. *Let $\mathfrak{s}_1$ and $\mathfrak{s}_2$ be two sources providing the testimonies $A_{\mathfrak{s}_1} = A$ and $A_{\mathfrak{s}_2} = B$, for some $A, B \subseteq \mathcal{Y}$, $A \neq B$. Assume that $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are both reliable with probability p_1, $\mathfrak{s}_1$ is not reliable and $\mathfrak{s}_2$ is reliable with probability p_2, $\mathfrak{s}_1$ is reliable and $\mathfrak{s}_2$ is not reliable with probability p_3, and they are both not reliable with probability p_4,* i.e.*, we have*

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= p_1, \\ P(X_1 = 1, X_2 = 0) &= p_2, \\ P(X_1 = 0, X_2 = 1) &= p_3, \\ P(X_1 = 1, X_2 = 1) &= p_4. \end{aligned}$$

Then, from (39)*, the induced knowledge on $\mathcal{Y}$ is represented by the mass function m defined as*

$$m(A \cap B) = p_1, m(B) = p_2, m(A) = p_3, m(\mathcal{Y}) = p_4.$$

Now, the marginal probabilities that sources $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are non reliable are $\pi_1 = p_2 + p_4$ and $\pi_2 = p_3 + p_4$, respectively. In addition, their meta-dependence structure is $\boldsymbol{\sigma} = (1, 0, 0, \sigma_4)'$, with $\sigma_4 = p_4 \cdot p_1 - p_2 \cdot p_3$. Hence, from Theorem 2, m can be equivalently

obtained as the conjunctive combination of the simple mass functions A^{π_1} and B^{π_2} with dependence structure $\boldsymbol{\sigma}$, that is, as

$$m = \textcircled{\cap}_{(1,0,0,\sigma_4)'}(A^{\pi_1}, B^{\pi_2}).$$

Theorem 2 is particularly useful to obtain an expression of the t-decomposition in terms of a conjunctive combination of some simple mass functions:

Theorem 3. *Any mass function m defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ with Teugels function t satisfies*

$$m = \textcircled{\cap}_{\boldsymbol{\sigma}} \left(\overline{\{y_1\}}^{t(\{y_1\})}, \ldots, \overline{\{y_n\}}^{t(\{y_n\})} \right), \tag{41}$$

with $\boldsymbol{\sigma}$ the vector such that $\sigma_1 = 1$ and, for $1 < k \leq 2^n$, $\sigma_k = t(A_k)$ if $|A_k| > 1$, and $\sigma_k = 0$ otherwise.

Proof. As shown in Section 3.3, the t-canonical decomposition allows one to view any mass function on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ as resulting from crisp testimonies $\overline{\{y_i\}}$ provided by sources $\mathfrak{s}_i$, $i = 1, \ldots, n$, assumed to be non reliable with marginal probabilities $t(\{y_i\})$ and having some meta-dependence represented by central moments $t(A_k), |A_k| > 1$. The theorem follows from this latter fact and Theorem 2. □

Theorem 3 shows that according to the t-decomposition, any mass function results from the conjunctive combination of $|\mathcal{Y}|$ elementary pieces of evidence represented by simple mass functions having some dependence structure. This is illustrated by Examples 8 and 9.

Example 8 (Example 4 continued)**.** *Let m be a mass function defined on $\mathcal{Y} = \{y_1, y_2\}$, with associated Teugels function t. From Theorem 3, m results from the conjunctive combination of simple mass functions $\overline{\{y_1\}}^{t(\{y_1\})}$ and $\overline{\{y_2\}}^{t(\{y_2\})}$ with dependence structure $\boldsymbol{\sigma} = [1, 0, 0, t(\{y_1, y_2\})]'$,* i.e.*, we have*

$$m = \textcircled{\cap}_{[1,0,0,t(\{y_1,y_2\})]'} \left(\overline{\{y_1\}}^{t(\{y_1\})}, \overline{\{y_2\}}^{t(\{y_2\})} \right).$$

Example 9 (Example 5 continued)**.** *From Theorem 3, the mass function m of Example 2, with function t provided in Example 5, results from the conjunctive combination of simple mass functions $\overline{\{y_1\}}^{1}$, $\overline{\{y_2\}}^{2/3}$ and $\overline{\{y_3\}}^{2/3}$ with dependence structure $\boldsymbol{\sigma} = (1, 0, 0, 0, 0, 0, -\frac{1}{9}, 0)'$,* i.e.*, we have*

$$m = \textcircled{\cap}_{(1,0,0,0,0,0,-\frac{1}{9},0)'} \left(\overline{\{y_1\}}^{1}, \overline{\{y_2\}}^{2/3}, \overline{\{y_3\}}^{2/3} \right).$$

Remark 3 (Bayesian mass function on a binary frame)**.** *Let $\boldsymbol{y}$ be a variable defined on $\mathcal{Y} = \{y_1, y_2\}$. Let $\mathfrak{s}_1$ and $\mathfrak{s}_2$ be two sources providing the testimonies $\boldsymbol{y} \in \overline{\{y_1\}} = \{y_2\}$ and $\boldsymbol{y} \in \overline{\{y_2\}} = \{y_1\}$, respectively. Assume that $\mathfrak{s}_1$ is reliable and $\mathfrak{s}_2$ is not reliable with probability α, and that $\mathfrak{s}_1$ is not reliable and $\mathfrak{s}_2$ is reliable with probability $1-\alpha$, for some $\alpha \in [0,1]$,* i.e.*, we have*

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= 0, \\ P(X_1 = 1, X_2 = 0) &= 1 - \alpha, \\ P(X_1 = 0, X_2 = 1) &= \alpha, \\ P(X_1 = 1, X_2 = 1) &= 0. \end{aligned}$$

Then, from (39)*, the induced knowledge on $\mathcal{Y}$ is represented by the Bayesian mass function m such that*

$$m(\{y_1\}) = 1 - \alpha, \quad m(\{y_2\}) = \alpha. \tag{42}$$

The marginal probabilities that sources $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are non reliable are $\pi_1 = 1-\alpha$ and $\pi_2 = \alpha$, respectively. In addition, their meta-dependence structure is $\boldsymbol{\sigma} = (1, 0, 0, \alpha^2 - \alpha)'$. Hence, from Theorem 2, m can be equivalently obtained as

$$m = \textcircled{\cap}_{(1,0,0,\alpha^2-\alpha)'}(\overline{\{y_1\}}^{1-\alpha}, \overline{\{y_2\}}^{\alpha}),$$

which is nothing but the t-decomposition of the Bayesian mass function (42)*.*

Now, suppose we are given only the marginal probabilities $\pi_1 = 1-\alpha$ and $\pi_2 = \alpha$. Then, assume that there is a perfect dependence *[15] between $\mathfrak{s}_1$ being reliable and $\mathfrak{s}_2$ being not reliable, which means that the probability $P(X_1 = 0, X_2 = 1)$ that $\mathfrak{s}_1$ is reliable and $\mathfrak{s}_2$ is not reliable is such that*

$$P(X_1 = 0, X_2 = 1) = P(X_1 = 0) \wedge P(X_2 = 1) = (1 - \pi_1) \wedge \pi_2,$$

where $\wedge$ denotes the minimum operator. This latter assumption is equivalent[4] *to assuming $\boldsymbol{\sigma} = (1, 0, 0, \alpha^2 - \alpha)'$, since by definition $\sigma_1 = 1$, $\sigma_2 = 0$, $\sigma_3 = 0$ and from* (27)

$$\begin{aligned} P(X_1 = 0, X_2 = 1) &= (1 - \pi_1)\pi_2 - \sigma_4 \\ \Leftrightarrow \sigma_4 &= (1 - \pi_1)\pi_2 - (1 - \pi_1) \wedge \pi_2 \\ &= \alpha^2 - \alpha. \end{aligned}$$

[4]Vector $\boldsymbol{\sigma}$ is equivalently obtained by three other assumptions: i) perfect dependence between $\mathfrak{s}_1$ being not reliable and $\mathfrak{s}_2$ being reliable, ii) *opposite dependence* [15] between $\mathfrak{s}_1$ being reliable and $\mathfrak{s}_2$ being reliable, iii) opposite dependence between $\mathfrak{s}_1$ being not reliable and $\mathfrak{s}_2$ being not reliable.

Hence, any Bayesian mass function on a binary domain is obtained, according to the t-decomposition, from two sources such that the marginal probability that the first source is reliable is equal to the marginal probability that the second source is not reliable, and such that there is a perfect dependence between the first source being reliable and the second source being not reliable.

Remark 4. *The cautious rule* $\circledwedge$*, which is based on Smets' decomposition, was proposed in [4] for the conjunctive combination of bodies of evidence, which cannot be assumed to be independent. The definition of this rule for the combination of two simple mass functions is the following. For all* $A, B \subset \mathcal{Y}$ *such that* $A \neq B$*, and* $\pi_1, \pi_2 \in (0, 1]$*, we have*

$$A^{\pi_1} \circledwedge A^{\pi_2} = A^{\pi_1 \wedge \pi_2}, \tag{43}$$

$$A^{\pi_1} \circledwedge B^{\pi_2} = A^{\pi_1} \circledcap B^{\pi_2}. \tag{44}$$

The behaviour of the cautious rule can be analysed in the light of our framework of meta-dependent and partially reliable sources. Specifically, viewing A^{π_i} *as the trace of a source* $\mathfrak{s}_i$ *providing testimony* $\boldsymbol{y} \in A$ *and assumed to be not reliable with marginal probability* π_i*,* $i = 1, 2$*, then Eq.* (43) *is recovered by assuming that there is a* perfect dependence *between* $\mathfrak{s}_1$ *not being reliable and* $\mathfrak{s}_2$ *not being reliable, which is equivalent to assuming that their meta-dependence is* $\boldsymbol{\sigma} = (1, 0, 0, \pi_1 \wedge \pi_2 - \pi_1\pi_2)'$*. From Theorem 2, we obtain then:*

$$A^{\pi_1} \circledwedge A^{\pi_2} = \circledcap_{(1,0,0,\pi_1 \wedge \pi_2 - \pi_1\pi_2)'} \left(A^{\pi_1}, A^{\pi_2}\right).$$

Furthermore, viewing B^{π_2} *as the trace of a source* $\mathfrak{s}_3$ *providing testimony* $\boldsymbol{y} \in B$ *and assumed to be not reliable with marginal probability* π_2*, then Eq.* (44) *is recovered by assuming* independence *between* $\mathfrak{s}_1$ *not being reliable and* $\mathfrak{s}_3$ *not being reliable, which is equivalent to assuming that their meta-dependence is* $\boldsymbol{\sigma} = \mathbf{e}_1$*. In other words, we have:*

$$A^{\pi_1} \circledwedge B^{\pi_2} = \circledcap_{\mathbf{e}_1} \left(A^{\pi_1}, B^{\pi_2}\right).$$

Hence, the combination by the cautious rule of two elementary pieces of evidence is recovered as different particular cases of the rule $\circledcap_\sigma$*, that is, by assuming different dependence structures among these pieces of evidence – the dependence structure to be used in a given situation being determined by whether the pieces of evidence support the same subset or not.*

4.3 Comparison with Smets' Decomposition

(i) The t-canonical decomposition of belief functions bears some resemblance with that of Smets, in that both of these decompositions view a belief function defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ as resulting from partially reliable sources providing crisp pieces of information about a variable of interest $\boldsymbol{y}$. As a matter of fact, these decompositions coincide in the following special case.

Let m be a mass function with Teugels function t such that for all $A \subseteq \mathcal{Y}, |A| > 1$, $t(A) = 0$. Then, we have

$$\begin{aligned} m &= \textcircled{\cap}_{\mathbf{e}_1} \left(\overline{\{y_1\}}^{t(\{y_1\})}, \ldots, \overline{\{y_n\}}^{t(\{y_n\})} \right), \\ &= \bigcirc\!\!\!\!\!\!\cap_{i=1}^{n} \overline{\{y_i\}}^{t(\{y_i\})}. \end{aligned} \tag{45}$$

Mass functions satisfying (45) will be called $\mathbf{e}_1$-separable hereafter. Moreover, for mass function m, Smets' decomposition yields:

$$m = \bigcirc\!\!\!\!\!\!\cap_{i=1}^{n} \overline{\{y_i\}}^{w(\overline{\{y_i\}})},$$

with $w(\overline{\{y_i\}}) = t(\{y_i\})$, $i = 1, \ldots, n$. Hence, according to both decompositions, any $\mathbf{e}_1$-separable mass function m is obtained from the conjunctive combination of $|\mathcal{Y}|$ independent elementary pieces of evidence represented by simple mass functions $\overline{\{y_i\}}^{t(\{y_i\})}$. In sum, the decompositions coincide for $\mathbf{e}_1$-separable mass functions.

In contrast, it may be interesting to note that the decompositions do not coincide in general for u-separable mass functions as shown by Example 10.

Example 10. *Let m be a u-separable mass function on $\mathcal{Y} = \{y_1, y_2\}$ such that $m = \emptyset^{w(\emptyset)} \textcircled{\cap} \{y_1\}^{w(\{y_1\})}$, with $w(\emptyset), w(\{y_1\}) \in (0, 1)$. We may remark that*

$$m = \textcircled{\cap}_{\mathbf{e}_1} \left(\emptyset^{w(\emptyset)}, \{y_1\}^{w(\{y_1\})} \right),$$

but that does not mean that m is $\mathbf{e}_1$-separable. Indeed, for m to be $\mathbf{e}_1$-separable, we must have $m = \textcircled{\cap}_{\mathbf{e}_1} \left(\overline{\{y_1\}}^{t(\{y_1\})}, \overline{\{y_2\}}^{t(\{y_2\})} \right)$.

Function t associated with m is

$$t(\{y_1\}) = w(\emptyset), \quad t(\{y_2\}) = w(\emptyset)w(\{y_1\}), \quad t(\{y_1, y_2\}) = w(\emptyset)w(\{y_1\})(1 - w(\emptyset)).$$

Hence, $m = \textcircled{\cap}_{[1,0,0,w(\emptyset)w(\{y_1\})(1-w(\emptyset))]'} \left(\overline{\{y_1\}}^{w(\emptyset)}, \overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})} \right)$, that is, m results according to the t-decomposition from the conjunctive combination of simple mass functions $\overline{\{y_1\}}^{w(\emptyset)}$ and $\overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})}$ with dependence structure $\boldsymbol{\sigma} = [1, 0, 0, w(\emptyset)w(\{y_1\})(1-$

$w(\emptyset))]'$, *which is different from Smets' decomposition for which* m *results from the conjunctive combination of the independent simple mass functions* $\emptyset^{w(\emptyset)}$ *and* $\{y_1\}^{w(\{y_1\})}$.

The fact that the decompositions coincide for $\mathbf{e}_1$-separable mass functions and do not coincide for u-separable mass functions actually follows from Lemma 2.

Lemma 2. *Let* m *be a mass function on* $\mathcal{Y}$. *We have*

- m *is* $\mathbf{e}_1$*-separable* $\Rightarrow$ m *is u-separable.*
- m *is* $\mathbf{e}_1$*-separable* $\not\Leftarrow$ m *is u-separable.*

Proof. $\Rightarrow$ follows from the definitions of $\mathbf{e}_1$-separability and u-separability.

$\not\Leftarrow$: Example 10 shows that a u-separable mass function m admits the t-decomposition $m = \textcircled{\cap}_{[1,0,0,w(\emptyset)w(\{y_1\})(1-w(\emptyset))]'}\left(\overline{\{y_1\}}^{w(\emptyset)}, \overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})}\right) \neq \textcircled{\cap}_{\mathbf{e}_1}\left(\overline{\{y_1\}}^{w(\emptyset)}, \overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})}\right)$. □

In general, Smets' decomposition and the t-decomposition significantly differ on various aspects. First, Smets' decomposition involves 2^n-1 sources while the t-decomposition involves n sources, and the sources are in general not meta-independent in the t-decomposition whereas Smets' decomposition involves meta-independence. In terms of elementary pieces of evidence, this means that the t-decomposition breaks down a mass function as the result of the conjunctive combination of n simple mass functions that have a dependency structure, whereas with Smets' decomposition it is the result of the conjunctive combination of $2^n - 1$ (generalised) simple mass functions that are independent. Second, the t-decomposition can be obtained for any mass function, whereas Smets' decomposition is restricted to non dogmatic mass function[5]. Third and most importantly, the t-decomposition involves only well-defined concepts with clear semantics, in particular means and central moments of Bernoulli variables, which is not the case of Smets' decomposition.

(ii) Smets' decomposition has a nice behaviour (13) with respect to the unnormalised Dempster's rule. It seems thus interesting to comment on the behaviour of our solution with respect to this rule. It is actually easy to uncover the assumptions associated with the use of this rule in our solution since Dempster's rule was originally introduced

[5]Smets [37] proposed a technical means to decompose a dogmatic mass function m, which consists essentially in assigning an ϵ to $\mathcal{Y}$. However, this comes down to approximating m by a non dogmatic mass function.

in a similar setting as ours [3]. Let m_1 and m_2 be two mass functions defined on $\mathcal{Y} = \{y_1, \ldots, y_n\}$. In our approach, mass function m_j, $j = 1, 2$, can be viewed as resulting from having n sources S_i^j, $i = 1, \ldots, n$, each telling $\boldsymbol{y} \in \overline{\{y_i\}}$, and such that uncertainty with respect to the reliability of these sources is represented by a probability distribution, with p_{k^j} the probability that sources S_i^j, $i = 1, \ldots, n$, are in the joint state k^j. Assuming that the sources S_i^1, $i = 1, \ldots, n$, underlying m_1 are meta-independent from the sources S_i^2, $i = 1, \ldots, n$, underlying m_2, we obtain that the joint probability $p_{(k^1,k^2)}$ that sources S_i^1 and S_i^2, $i = 1, \ldots, n$, are in the joint state (k^1, k^2) is $p_{(k^1,k^2)} = p_{k^1} \cdot p_{k^2}$. Besides, if the sources S_i^1 and S_i^2, $i = 1, \ldots, n$, are in the joint state (k^1, k^2), then we should deduce that $\boldsymbol{y} \in \mathrm{T}(k^1) \cap \mathrm{T}(k^2)$. Hence assuming meta-independence between the sources underlying m_1 and m_2 yield the mass function m_{12} such that

$$m_{12}(A) \quad = \sum_{k^1,k^2:\mathrm{T}(k^1)\cap\mathrm{T}(k^2)=A} p_{(k^1,k^2)}. \tag{46}$$

It can easily be checked that $m_{12} = m_{1 \textcircled{\cap} 2}$ and thus the unnormalised Dempster's rule has a nice interpretation in our approach: it corresponds simply to assuming meta-independence between the sources underlying m_1 and the sources underlying m_2.

In addition, let us remark that the mass function $m_{1 \textcircled{\cap} 2}$ can be viewed using our solution as originating from n sources, which we denote by S_i^{12}, $i = 1, \ldots, n$, and an associated probability distribution p^{12} representing uncertainty with respect to the reliability of these latter sources. We may remark that due to (7) the vector $\boldsymbol{\mu}^{12}$ of moments associated with p^{12} is such that $\mu^{12} = \mu^1 \cdot \mu^2$, with $\boldsymbol{\mu}^i$ the vector of moments associated with the MBD-equivalent of m_i, $i = 1, 2$. In other words, any moment between the reliabilities of the sources underlying $m_{1 \textcircled{\cap} 2}$ is equal to the pointwise product of the same moments between the reliabilities of the sources underlying m_1 and m_2. This means that in our approach the unnormalised Dempster's rule has a similar behaviour to the one it has in Smets' solution: it comes down to a simple pointwise product of some functions (μ^1 and μ^2 in our case, w_1 and w_2 in Smets' case).

4.4 t-Canonical Decomposition of Random Sets

A random set is a random element taking values as subsets of some space [21, 23]. A random set is thus defined in the finite case by a probability distribution m on the power set $2^{\mathcal{X}}$ of some finite set $\mathcal{X} = \{x_1, \ldots, x_n\}$ such that $\sum_{A \subseteq \mathcal{X}} m(A) = 1$. Distribution m is formally equivalent to a mass function on $\mathcal{X}$ [27], but it has different semantics [35].

For instance, borrowing from [2], let $\mathcal{X} = \{\text{English, French, Spanish}\}$ denote the set of languages that a person can speak, then $m(A)$ is the probability that someone speaks *exactly* all the languages in A (and not other ones), $A \subseteq \mathcal{X}$. Furthermore, the commonality degree $q(A) = \sum_{B \supseteq A} m(B)$ is the probability that someone speaks *at least* all the languages in A; for instance, the probability $q(\{\text{English}\}$ that someone speaks at least English is obtained by summing the probabilities $m(\{\text{English}\}$ of speaking only English, $m(\{\text{English, French}\}$ of speaking only English and French, $m(\{\text{English, Spanish}\}$ of speaking only English and Spanish, and $m(\{\text{English, French, Spanish}\}$ of speaking the three languages.

Definition 4 (MBD-RS-equivalent)**.** *Let m be the probability distribution of some finite random set, defined on the power set of some set $\mathcal{X} = \{x_1, \ldots, x_n\}$ and let p be the MBD* (22) *such that $p_k = m(A_k)$, $1 \leq k \leq 2^n$, with A_k the k-th subset of $\mathcal{X}$ according to the binary order. p is called the* MBD-Random Set-equivalent *(MBD-RS-equivalent for short) of m.*

Although they seem fairly similar, the bond between the probability distribution of a random set and its MBD-RS-equivalent is much stronger than that of a mass function and its MBD-equivalent. Indeed, the notion of MBD-equivalent becomes interesting mostly in conjunction with Proposition 1. In contrast, the semantics of a random set and its MBD-RS-equivalent are in immediate correspondence. For instance, let m be the preceding probability distribution representing the languages that someone speaks among the $n = 3$ languages English, French and Spanish. Furthermore, let p be the MBD-RS-equivalent of m, where Bernoulli random variables X_1, X_2 and X_3 underlying p represent respectively whether someone speaks English, French and Spanish, with $X_i = 1$ meaning that someone speaks the i-th language and $X_i = 0$ meaning that someone does not speak the i-th language. Then p_k has the same interpretation as $m(A_k)$, $1 \leq k \leq 2^n$: it is the probability that someone speaks exactly all the languages in A_k (and not other ones). We have for instance $A_4 = \{\text{English, French}\}$ and the probability $m(A_4)$ that someone speaks only English and French is equal to $p_4 = p_{110} = P(X_1 = 1, X_2 = 1, X_3 = 0)$.

It is clear that the vector $\boldsymbol{\mu}$ computed through (26) from the MBD-RS-equivalent p of the probability distribution m of a random set, is the vector $\mathbf{q}$ of commonality degrees associated to m. Most importantly, thanks to the Teugels' representation τ of p, we can propose a canonical decomposition of a random set, which has a well-defined semantics.

This canonical decomposition is given by function t computed using (37). For instance, continuing the language example, let m be such that

$$m(A_4) = m(A_6) = m(A_8) = 1/3, \tag{47}$$

with $A_4 = \{\text{English, French}\}$, $A_6 = \{\text{English, Spanish}\}$ and $A_8 = \{\text{English, French, Spanish}\}$. We have

$$\begin{aligned} t(\{\text{English}\} &= 1, \\ t(\{\text{French}\} &= 2/3, \\ t(\{\text{English,French}\} &= 0, \\ t(\{\text{Spanish}\} &= 2/3, \\ t(\{\text{English,Spanish}\} &= 0, \\ t(\{\text{French,Spanish}\} &= -1/9, \\ t(\{\text{English,French,Spanish}\} &= 0. \end{aligned}$$

Hence, the random set defined by m can be seen as originating from the following pieces of information

- Someone speaks English, French and Spanish with probabilities 1, 2/3 and 2/3 respectively;
- The covariance between someone speaking French and speaking Spanish is $-1/9$, and all other subsets of variables representing the languages spoken by someone have null central moments.

Let us conclude this section by remarking that the correspondence established between finite random sets and the MBD can be used together with the multi-valued mapping T (24) to relate random sets and belief functions in a different way than in [27], and specifically as a means to obtain a belief function from a random set. Indeed, let p be the MBD-RS-equivalent of the probability distribution m of some random set, defined on $2^{\mathcal{X}}$ with $\mathcal{X} = \{x_1, \ldots, x_n\}$. Then, transferring MBD p via multi-valued mapping T (24) yield a mass function m' on $\mathcal{Y} = \{y_1, \ldots, y_n\}$ defined, for $1 \leq k \leq 2^n$, by

$$m'(A_k) = \sum_{k' \in \mathcal{X}^n : \mathrm{T}(k') = A_k} p_{k'},$$

with A_k the k-th subset of $\mathcal{Y}$. For instance, consider again probability distribution m defined by (47) and let p be its MBD-RS-equivalent, where Bernoulli random variables

X_i, $i = 1, 2, 3$, underlying p represent respectively whether someone speaks English, French and Spanish. Let $\boldsymbol{y}$ be a variable defined on

$$\begin{aligned} \mathcal{Y} &= \{\text{British citizen, French citizen, Spanish citizen}\} \\ &= \{y_1, y_2, y_3\} \end{aligned}$$

and denoting the citizenship of someone (we assume someone has only one citizenship). Furthermore, assume the following pieces of information:

- Someone who does not speak the language of a given country, is surely not a citizen of that country;
- A citizen of a given country may speak the language of another country (*e.g.*, a Spanish man may speak French).

For each language i, $i = 1, 2, 3$, these pieces of information can be represented by a multi-valued mapping Γ_i defined by (23). For instance, we have $\Gamma_2(0) = \{y_1, y_3\}$ and $\Gamma_2(1) = \mathcal{Y}$ since someone who does not speak French ($X_2 = 0$) is surely not a French citizen ($\boldsymbol{y} \in \{y_1, y_3\}$), and since someone who speaks French ($X_2 = 1$) may be a citizen of any country ($\boldsymbol{y} \in \mathcal{Y}$). More generally, what can be deduced about the citizenship of someone given the languages that she speaks can be represented by multi-valued mapping T (24). For instance, someone who speaks English and French but not Spanish is a British citizen or a French citizen, as

$$\begin{aligned} \mathrm{T}[(1,1,0)] &= \Gamma_1(1) \bigcap \Gamma_2(1) \bigcap \Gamma_3(0) \\ &= \mathcal{Y} \cap \mathcal{Y} \cap \overline{\{y_3\}} \\ &= \{y_1, y_2\}. \end{aligned}$$

Hence, if knowledge about the languages spoken by someone is represented by probability distribution m defined by (47), we obtain that the citizenship of someone is represented by mass function m' defined on $\mathcal{Y}$ by $m'(\{y_1, y_2\}) = m'(\{y_1, y_3\}) = m'(\mathcal{Y}) = 1/3$.

5 A MBD-Based Perspective on the Weight Function

We have argued that the semantics provided by Smets for the function w, as weights associated with statements of the form *believe* and *do not believe* in some propositions, is not totally satisfactory. However, this does not mean that this function does not

represent some facet of the information contained in a mass function, similarly as, *e.g.*, *pl* or *bel* – it just means that the interpretation given by Smets for this function as representing a canonical decomposition, that is, a decomposition into elementary pieces of evidence, is not totally acceptable, at least for the time being, and deserves further justifications.

As a matter of fact, a completely new perspective on the weight function w associated to a mass function m is brought to light in this section, using measures of information associated with the MBD-equivalent of m. As will be seen, the proposed interpretation is not as appealing as Smets', and especially is unrelated to the concept of canonical decomposition, but it relies only on well-defined concepts. Then, a similar result is provided for the disjunctive counterpart of w known as the *disjunctive* weight function [4] and a disjunctive counterpart of the expression (41) is also unveiled for the t-decomposition.

5.1 Weights as Measures of Information

Instead of using function w, Smets' decomposition (11) can be equivalently presented using a function $s : 2^{\mathcal{Y}} \setminus \{\mathcal{Y}\} \to (-\infty, +\infty)$ such that $s(A) = -\ln w(A)$ for all $A \subset \mathcal{Y}$ (see, *e.g.*, [16]). We note that simple mass functions correspond to the case where $s(A) \geq 0$ and that inverse simple mass functions correspond to the case where $s(A) < 0$. Function s is actually the function Shafer originally used in his monograph [34] to present the canonical decomposition of separable mass functions (9). Accordingly and following [16], $s(A)$, $A \subset \mathcal{Y}$, may be referred to as *Shafer's weights*.

Let p be the MBD-equivalent of some mass function m and let μ be the moments of p. By replacing q in (12) by μ, we obtain the following expression for function s, for all $A_k \subset \mathcal{Y}$:

$$\begin{aligned} s(A_k) &= -\ln\Big(\prod_{A_{k'} \supseteq A_k} q(A_{k'})^{(-1)^{|A_{k'}|-|A_k|+1}}\Big), \\ &= -\ln\Big(\prod_{A_{k'} \supseteq A_k} \mu_{k'}^{(-1)^{|A_{k'}|-|A_k|+1}}\Big). \end{aligned} \tag{48}$$

In particular, let m be a mass function on $\mathcal{Y} = \{y_1, y_2\}$. We have:

$$\begin{aligned} s(\emptyset) &= -\ln \frac{\mu_2\mu_3}{\mu_1\mu_4} \\ &= -\ln \frac{\mu_2\mu_3}{\mu_4} \\ &= \ln \frac{\mu_4}{\mu_2\mu_3}, \\ s(\{y_1\}) &= -\ln \frac{\mu_4}{\mu_2}, \\ s(\{y_2\}) &= -\ln \frac{\mu_4}{\mu_3}. \end{aligned}$$

Interestingly, these latter combinations of the marginal probabilities $\mu_{k'}$, and thus Shafer's weights, correspond to known quantities associated with the MBD. This is detailed hereafter.

Let X_1 and X_2 be two Bernoulli random variables and let p be the MBD such that $p_k = p_{k_1,k_2}$, $1 \leq k \leq 4$, with $p_{k_1,k_2} := P(X_1 = k_1, X_2 = k_2)$ where $k_i \in \{0, 1\}$, $i = 1, 2$, are the terms in the binary expansion of k. Furthermore, let $\boldsymbol{\mu}$ be the vector of moments of p. Recall (cf [14, p. 28]) that the *mutual information* $I(X_1 = k_1; X_2 = k_2)$ between two events $X_1 = k_1$ and $X_2 = k_2$ is[6]

$$\begin{aligned} I(X_1 = k_1; X_2 = k_2) &= \ln \frac{P(X_1 = k_1, X_2 = k_2)}{P(X_1 = k_1)P(X_2 = k_2)} \\ &= \ln \frac{p_k}{(\sum_{k':k'_1=k_1} p_{k'})(\sum_{k':k'_2=k_2} p_{k'})}. \end{aligned}$$

Besides, the *conditional self information* $I(X_1 = k_1|X_2 = k_2)$ of event $X_1 = k_1$ given event $X_2 = k_2$ is [14, p. 36]

$$\begin{aligned} I(X_1 = k_1|X_2 = k_2) &= -\ln \frac{P(X_1 = k_1, X_2 = k_2)}{P(X_2 = k_2)} \\ &= -\ln \frac{p_k}{\sum_{k':k'_2=k_2} p_{k'}}, \end{aligned}$$

and the conditional self information $I(X_2 = k_2|X_1 = k_1)$ of event $X_2 = k_2$ given event $X_1 = k_1$ is

$$\begin{aligned} I(X_2 = k_2|X_1 = k_1) &= -\ln \frac{P(X_1 = k_1, X_2 = k_2)}{P(X_1 = k_1)} \\ &= -\ln \frac{p_k}{\sum_{k':k'_1=k_1} p_{k'}}. \end{aligned}$$

[6]Usually, log rather than ln is used, but this is just a change of unit of information: the unit is bits when using log and nats when using ln [14].

In particular, we have for $k = 4 \leftrightarrow (k_1 = 1, k_2 = 1)$

$$\begin{aligned}
I(X_1 = 1; X_2 = 1) &= \ln \frac{p_4}{(\sum_{k':k'_1=1} p_{k'})(\sum_{k':k'_2=1} p_{k'})} \\
&= \ln \frac{\mu_4}{\mu_2 \mu_3}, \\
I(X_2 = 1 | X_1 = 1) &= -\ln \frac{p_4}{\sum_{k':k'_1=1} p_{k'}} \\
&= -\ln \frac{\mu_4}{\mu_2}, \\
I(X_1 = 1 | X_2 = 1) &= -\ln \frac{p_4}{\sum_{k':k'_2=1} p_{k'}} \\
&= -\ln \frac{\mu_4}{\mu_3}.
\end{aligned}$$

Hence, when m is a mass function defined on the binary domain $\mathcal{Y} = \{y_1, y_2\}$, with MBD-equivalent p, we have

$$\begin{aligned}
s(\emptyset) &= I(X_1 = 1; X_2 = 1), \\
s(\{y_1\}) &= I(X_2 = 1 | X_1 = 1), \\
s(\{y_2\}) &= I(X_1 = 1 | X_2 = 1).
\end{aligned}$$

Adopting the setting of Remark 1, $s(\emptyset)$ is then equal to the mutual information between both sources underlying m not being reliable, $s(\{y_1\})$ is equal to the conditional self information of the second source not being reliable given that the first source is not reliable, and $s(\{y_2\})$ is equal to the conditional self information of the first source not being reliable given that the second source is not reliable.

Remark that for any mass function m on $\mathcal{Y} = \{y_1, y_2\}$, we have $s(\{y_i\}) = -\ln \frac{q(\{y_1, y_2\})}{q(\{y_i\})} \geq 0$, $i = 1, 2$, and $s(\emptyset) = \ln \frac{q(\{y_1, y_2\})}{q(\{y_1\}) q(\{y_2\})} \in (-\infty, +\infty)$. In other words, when $\mathcal{Y}$ is binary, only for $A = \emptyset$ one can have $s(A) < 0$, which is a case of debt of belief for A according to Smets' interpretation of s. Our new perspective on s provides a completely different meaning to $s(\emptyset) < 0$ than that of debt of belief, and especially a well-established meaning which is that of mutual information. More generally, our alternative interpretation for the whole function s is admittedly less appealing than the one proposed by Smets, but it is rigorous.

When m is a mass function on $\mathcal{Y} = \{y_1, y_2, y_3\}$, its associated weight function s can also be interpreted using measures of information, as explained hereafter. The MBD-equivalent p of m relies in this case on three Bernoulli random variables $\{X_1, X_2, X_3\}$. Let us recall that the mutual information $I(X_1 = 1; X_2 = 1; X_3 = 1)$ between the three

events $X_1 = 1$, $X_2 = 1$ and $X_3 = 1$ is [14, p. 57-58]

$$\begin{aligned} I(X_1 = 1; X_2 = 1; X_3 = 1) &= \ln \frac{P(X_1 = 1, X_2 = 1)P(X_1 = 1, X_3 = 1)P(X_2 = 1, X_3 = 1)}{P(X_1 = 1)P(X_2 = 1)P(X_3 = 1)P(X_1 = 1, X_2 = 1, X_3 = 1)} \\ &= \ln \frac{\mu_4\mu_6\mu_7}{\mu_2\mu_3\mu_5\mu_8}. \end{aligned}$$

with $\boldsymbol{\mu}$ the vector of moments of p. The *conditional mutual information* $I(X_2 = 1; X_3 = 1|X_1 = 1)$ between events $X_2 = 1$ and $X_3 = 1$ given event $X_1 = 1$ is [14, p. 28]

$$\begin{aligned} I(X_2 = 1; X_3 = 1|X_1 = 1) &= \ln \frac{P(X_1 = 1)P(X_1 = 1, X_2 = 1, X_3 = 1)}{P(X_1 = 1, X_2 = 1)P(X_1 = 1, X_3 = 1)} \\ &= \ln \frac{\mu_2\mu_8}{\mu_4\mu_6}. \end{aligned}$$

The conditional self information $I(X_3 = 1|X_1 = 1, X_2 = 1)$ of event $X_3 = 1$ given events $X_1 = 1$ and $X_2 = 1$ is

$$\begin{aligned} I(X_3 = 1|X_1 = 1, X_2 = 1) &= -\ln \frac{P(X_1 = 1, X_2 = 1, X_3 = 1)}{P(X_1 = 1, X_2 = 1)} \\ &= -\ln \frac{\mu_8}{\mu_4}. \end{aligned}$$

For m a mass function on $\mathcal{Y} = \{y_1, y_2, y_3\}$, we have thus

$$\begin{aligned} s(\emptyset) &= -\ln \frac{\mu_2\mu_3\mu_5\mu_8}{\mu_1\mu_4\mu_6\mu_7} \\ &= I(X_1 = 1; X_2 = 1; X_3 = 1), \quad (\text{since } \mu_1 = 1) \\ s(\{y_1\}) &= -\ln \frac{\mu_4\mu_6}{\mu_2\mu_8} \\ &= I(X_2 = 1; X_3 = 1|X_1 = 1), \\ s(\{y_1, y_2\}) &= -\ln \frac{\mu_8}{\mu_4} \\ &= I(X_3 = 1|X_1 = 1, X_2 = 1), \end{aligned}$$

and similarly we obtain

$$\begin{aligned} s(\{y_2\}) &= I(X_1 = 1; X_3 = 1|X_2 = 1), \\ s(\{y_3\}) &= I(X_1 = 1; X_2 = 1|X_3 = 1), \\ s(\{y_1, y_3\}) &= I(X_2 = 1|X_1 = 1, X_3 = 1), \\ s(\{y_2, y_3\}) &= I(X_1 = 1|X_2 = 1, X_3 = 1). \end{aligned}$$

Hence, for instance, $s(\{y_2\})$ is equal to the conditional mutual information between the first and third sources underlying m not being reliable given that the second source is not reliable.

The above interpretation for function s in terms of measures of information can be extended to belief functions defined on domains $\mathcal{Y}$ of any cardinality. This extension is based on the notion of conditional mutual information between an arbitrary number of events [14] recalled hereafter.

Let $\{V_i : i = 1, \ldots, \ell\}$ be a sequence of discrete random variables with ranges $\mathcal{V}_i$, $i = 1, \ldots, \ell$. Consider the multivariate distribution $P(V_1 = v_1, \ldots, V_\ell = v_\ell)$ with $v_i \in \mathcal{V}_i, i = 1, \ldots, \ell$. Then, the *conditional mutual information* between events $V_i = v_i, i = 1, \ldots, \ell - 1$, given event $V_\ell = v_\ell$, for some $v_i \in \mathcal{V}_i, i = 1, \ldots, \ell$, is [14, p. 58]

$$I(V_1 = v_1; \ldots; V_{\ell-1} = v_{\ell-1} | V_\ell = v_\ell) \quad = \quad \ln \frac{\prod_{E \subseteq \Xi, (V_\ell = v_\ell) \in E, |E| \notin 2\mathbb{N}} P(E)}{\prod_{E \subseteq \Xi, (V_\ell = v_\ell) \in E, |E| \in 2\mathbb{N}} P(E)}, \tag{49}$$

with $\Xi := \{V_i = v_i : i = 1, \ldots, \ell\}$. It will be convenient to denote the conditional mutual information between all the events belonging to some set $\mathcal{S}$, given another event E, as $I(Mut(\mathcal{S}))|E)$. For instance, the quantity (49) may be equivalently denoted as $I(Mut(\mathcal{S})|V_\ell = v_\ell)$ with $\mathcal{S} := \{V_i = v_i : i = 1, \ldots, \ell - 1\}$.

Theorem 4. *Let m be a mass function on $\mathcal{Y} = \{y_1, \ldots, y_n\}$ with MBD-equivalent p and with $\{X_i : i = 1, \ldots, n\}$ the Bernoulli random variables underlying p. We have*

$$s(A_k) \quad = \quad I(Mut(\mathcal{S}_k)|C_k = 1), \quad 1 \leq k < 2^n,$$

with, using $k \leftrightarrow (k_1, \ldots, k_n)$, $C_k := \prod_{i=1}^n X_i^{k_i}$ and $\mathcal{S}_k := \{X_i = 1 : i, k_i = 0\}$.

Proof. Eq. (48) can be rewritten as

$$s(A_k) \quad = \quad \sum_{A_{k'} \supseteq A_k} (-1)^{|A_{k'}| - |A_k|} \ln \mu_{k'}. \tag{50}$$

For all $A_{k'}, A_k \subseteq \mathcal{Y}$, we have $A_{k'} \supseteq A_k \Leftrightarrow \forall 1 \leq i \leq n$ s.t. $k_i = 1$, we have $k'_i = 1$, using $k \leftrightarrow (k_1, \ldots, k_n)$ and $k' \leftrightarrow (k'_1, \ldots, k'_n)$.

Let $1 \leq k \leq 2^n$ and $1 \leq k' \leq 2^n$. If $\forall 1 \leq i \leq n$ s.t. $k_i = 1$, we have $k'_i = 1$, we will write $k' \sqsupseteq k$. Hence, for any $A_{k'}, A_k \subseteq \mathcal{Y}$, $A_{k'} \supseteq A_k \Leftrightarrow k' \sqsupseteq k$.

Let $|k| := |\{i : i, k_i = 1\}|$. Hence, for all $A_k \subseteq \mathcal{Y}$, we have $|A_k| = |k|$.

Using these notations, Eq. (50) can be rewritten as

$$s(A_k) \quad = \quad \sum_{k' \sqsupseteq k} (-1)^{|k'| - |k|} \ln \mu_{k'}. \tag{51}$$

For $1 \leq k < 2^n$, let $\Xi_k := \mathcal{S}_k \cup \{C_k = 1\}$. Since for $1 \leq k < 2^n$, $C_k = 1$ is equivalent to the event $\{X_i = 1 : i, k_i = 1\}$ with $k \leftrightarrow (k_1, \ldots, k_n)$, any $E \subseteq \Xi_k$ s.t. $(C_k = 1) \in E$

is equivalent to the event $E\backslash\{(C_k = 1)\} \cup \{X_i = 1 : i, k_i = 1\}$ and has thus probability $P(E) = \mu_{k'}$ with $\boldsymbol{\mu}$ the vector of moments of the MBD p and $k' \leftrightarrow (k'_1, \ldots, k'_n)$ s.t., for all $1 \leq i \leq n$, $k'_i = 1$ if $k_i = 1$ or $(X_i = 1) \in E\backslash\{(C_k = 1)\}$, and $k'_i = 0$ otherwise. Hence, for the conditional mutual information between events in $\mathcal{S}_k$ given event $C_k = 1$, we have

$$\begin{aligned} I(Mut(\mathcal{S}_k)|C_k = 1) &= \ln \frac{\prod_{E\subseteq\Xi_k,(C_k=1)\in E,|E|\notin 2\mathbf{N}} P(E)}{\prod_{E\subseteq\Xi_k,(C_k=1)\in E,|E|\in 2\mathbf{N}} P(E)} \\ &= \ln \frac{\prod_{k'\sqsupseteq k,|k'|-|k|\in 2\mathbf{N}} \mu_{k'}}{\prod_{k'\sqsupseteq k,|k'|-|k|\notin 2\mathbf{N}} \mu_{k'}} \\ &= \sum_{k'\sqsupseteq k,|k'|-|k|\in 2\mathbf{N}} \ln \mu_{k'} - \sum_{k'\sqsupseteq k,|k'|-|k|\notin 2\mathbf{N}} \ln \mu_{k'} \\ &= \sum_{k'\sqsupseteq k} (-1)^{|k'|-|k|} \ln \mu_{k'}. \end{aligned}$$

□

Since for $1 \leq k < 2^n$, $C_k = 1$ is equivalent to the event $\{X_i = 1 : i, k_i = 1\}$, Theorem 4 shows, using the setting of Remark 1, that $s(A_k)$ is equal to the conditional mutual information between sources $\mathfrak{s}_i$, i such that $k_i = 0$, not being reliable, given that the sources $\mathfrak{s}_i$, i such that $k_i = 1$, are not reliable.

5.2 Disjunctive Counterparts

The *implicability* function b is another equivalent representation of a mass function m, defined as

$$b(A) = \sum_{B\subseteq A} m(B), \quad \forall A \subseteq \mathcal{Y}.$$

It allows an expression similar to Eq. (7) for the combination by the disjunctive rule $\textcircled{\cup}$ [12, 36], which is a combination rule that has the same definition as $\textcircled{\cap}$ except that $\cap$ is replaced by $\cup$ in (6): we have $b_{1\textcircled{\cup}2}(A) = b_1(A) \cdot b_2(A)$, for all $A \subseteq \mathcal{Y}$. In matrix form, we have [38]

$$\mathbf{m} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \right) \mathbf{b}. \tag{52}$$

Let us consider the MBD p (22) and its associated vector $\boldsymbol{\lambda}$ defined as

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{2^n}),$$

where, for $1 \leq k \leq 2^n$,

$$\lambda_k \;=\; \mathbb{E}\left[\prod_{i=1}^{n}(1-X_i)^{1-k_i}\right],$$

with k_i, $i = 1, \ldots, n$, the terms in the binary expansion of k. λ_k comes down to the (marginal) probability that each of the variables in $\{X_i : i, k_i = 0\}$ equals 0.

Using a similar[7] proof to that of (25) provided in [42], it is straightforward to show that the following relation holds

$$\mathbf{p} \;=\; \left(\bigotimes_{i=1}^{n}\begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\right)\boldsymbol{\lambda}. \tag{53}$$

Let m be a mass function on $\mathcal{Y}$ with associated implicability function b, and let p be its MBD-equivalent, with associated vector $\boldsymbol{\lambda}$. By comparing Eqs. (52) and (53), it is clear that $\lambda_k = b(A_k)$ for all $1 \leq k \leq 2^n$. Hence, adopting the setting of Remark 1, $b(A_k)$ is equal to the (marginal) probability that each of the sources in $\{\mathfrak{s}_i : i, k_i = 0\}$ is reliable.

As shown in [4], there exists a decomposition of non normal belief functions based on the disjunctive rule. This decomposition relies on the *disjunctive* weight function $v : 2^{\mathcal{Y}}\backslash\{\emptyset\} \rightarrow (0, +\infty)$, which is an equivalent representation of a non normal mass function m, defined as

$$v(A) \;=\; \prod_{B \subseteq A} b(B)^{(-1)^{|A|-|B|+1}}, \quad \forall A \neq \emptyset,$$

with b the implicability function associated to m. We have, for any non normal mass function m,

$$m \;=\; \bigcirc\!\!\!\!\!\!\bigcup_{A \neq \emptyset} A_{v(A)}, \tag{54}$$

with v the disjunctive weight function associated to m and $A_{v(A)} : 2^{\mathcal{Y}} \rightarrow \mathbb{R}$ a mapping called a *negative* generalised simple mass function allocating $v(A)$ to $\emptyset$, $1 - v(A)$ to $A \subseteq \mathcal{Y}, A \neq \emptyset$, and 0 to all $B \in 2^{\mathcal{Y}}\backslash\{A, \emptyset\}$ (if $v(A) \leq 1$, then this mapping is a proper mass function called a negative simple mass function). Decomposition (54) is unique and

[7] The main difference with the proof of (25) is that one needs to use

$$\begin{bmatrix}1 - X_i\\ X_i\end{bmatrix} \;=\; \begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\begin{bmatrix}1 - X_i\\ 1\end{bmatrix}, \quad i = 1, \ldots, n.$$

is the disjunctive counterpart of (11). The interpretation of disjunctive weights $v(A)$, $A \neq \emptyset$, is not discussed in [4]. However, using a similar reasoning to the one followed in Section 5.1 for the weight function w, disjunctive weights can be related to measures of information, as detailed below.

Much as it is possible to present decomposition (11) using function s rather than w, the decomposition (54) can be presented using a function $r : 2^{\mathcal{Y}} \setminus \{\emptyset\} \rightarrow (-\infty, +\infty)$ such that $r(A) = -\ln v(A)$. Let m be a mass function defined on $\mathcal{Y} = \{y_1, y_2\}$, with MBD-equivalent p. We obtain

$$\begin{aligned}
r(\{y_1\}) &= -\ln \frac{b(\emptyset)}{b(\{y_1\})} \\
&= -\ln \frac{\lambda_1}{\lambda_2} \\
&= I(X_1 = 0 | X_2 = 0), \\
r(\{y_2\}) &= -\ln \frac{b(\emptyset)}{b(\{y_2\})} \\
&= -\ln \frac{\lambda_1}{\lambda_3} \\
&= I(X_2 = 0 | X_1 = 0), \\
r(\mathcal{Y}) &= -\ln \frac{b(\{y_1\})b(\{y_2\})}{b(\emptyset)b(\mathcal{Y})} \\
&= \ln \frac{\lambda_1}{\lambda_2 \lambda_3} \\
&= I(X_1 = 0; X_2 = 0),
\end{aligned}$$

since we have, for $k = 1 \leftrightarrow (k_1 = 0, k_2 = 0)$,

$$\begin{aligned}
I(X_1 = 0; X_2 = 0) &= \ln \frac{P(X_1 = 0, X_2 = 0)}{P(X_1 = 0)P(X_2 = 0)} \\
&= \ln \frac{p_1}{(\sum_{k' : k'_1 = 0} p_{k'})(\sum_{k' : k'_2 = 0} p_{k'})} \\
&= \ln \frac{\lambda_1}{\lambda_3 \lambda_2}, \\
I(X_2 = 0 | X_1 = 0) &= -\ln \frac{P(X_1 = 0, X_2 = 0)}{P(X_1 = 0)} \\
&= -\ln \frac{\lambda_1}{\lambda_3}, \\
I(X_1 = 0 | X_2 = 0) &= -\ln \frac{P(X_1 = 0, X_2 = 0)}{P(X_2 = 0)} \\
&= -\ln \frac{\lambda_1}{\lambda_2}.
\end{aligned}$$

Adopting the setting of Remark 1, we have for instance that $r(\mathcal{Y})$ is thus equal to the mutual information between both sources underlying m being reliable.

More generally, the above interpretation for function r in terms of measures of information can be extended to belief functions defined on domains $\mathcal{Y}$ of any cardinality, thanks to Theorem 5, which is the counterpart to Theorem 4 for function r.

Theorem 5. *Let m be a mass function on $\mathcal{Y} = \{y_1, \ldots, y_n\}$ with MBD-equivalent p and with $\{X_i : i = 1, \ldots, n\}$ the Bernoulli random variables underlying p. We have*

$$r(A_k) \quad = \quad I(Mut(\mathcal{R}_k)|D_k = 1), \quad 1 < k \leq 2^n,$$

with, using $k \leftrightarrow (k_1, \ldots, k_N)$, $D_k := \prod_{i=1}^{n}(1 - X_i)^{1-k_i}$ and $\mathcal{R}_k := \{X_i = 0 : i, k_i = 1\}$.

Proof. The proof is similar to that of Theorem 4. □

Since for $1 < k \leq 2^n$, $D_k = 1$ is equivalent to the event $\{X_i = 0 : i, k_i = 0\}$, Theorem 5 shows, using the setting of Remark 1, that $r(A_k)$ is equal to the conditional mutual information between sources $\mathfrak{s}_i$, i such that $k_i = 1$, being reliable, given that the sources $\mathfrak{s}_i$, i such that $k_i = 0$, are reliable.

Let us conclude by remarking that much as Eq. (54) is the disjunctive counterpart of Eq. (11), it is possible to obtain a disjunctive counterpart of Theorem 3 as follows. In [11], a general approach for the *disjunctive* combination of bodies of evidence is also mentioned: it consists in a simple modification of Step 2 of the general approach to the conjunctive combination of bodies of evidence recalled in Section 4.1, where each joint mass $jm(A_1, \ldots, A_N)$ is now allocated to the subset $\bigcup_{i=1}^{N} A_i$. As a result, the disjunctive combination of N mass functions $m_1, \ldots, m_N$ on $\mathcal{Y}$ is the mass function $m_\cup$ defined as

$$m_\cup(A) \quad = \quad \sum_{\bigcup_{i=1}^{N} A_i = A} jm(A_1, \ldots, A_N). \quad \forall A \subseteq \mathcal{Y}.$$

Let us consider the case where the mass functions $m_1, \ldots, m_N$ are binary and such that $m_i = (A_{i_0}, A_{i_1})^{\pi_i}$. In such case, the dependence structure represented by a joint mass function jm, is completely determined as explained in Section 4.1 by vector $\boldsymbol{\sigma}$ associated with the MBD such that $p_k = jm(\mathbf{A}_k)$ with $\mathbf{A}_k := (A_{1_{k_1}}, \ldots, A_{N_{k_N}})$ using $k \leftrightarrow (k_1, \ldots, k_N)$, $1 \leq k \leq 2^N$. As a result, the disjunctive combination of N binary mass functions can be expressed as a parameterised combination rule $\textcircled{\cup}_{\boldsymbol{\sigma}} : \mathcal{B}^N \to \mathcal{M}$, with parameter $\boldsymbol{\sigma}$ representing the dependence structure, defined as

$$\textcircled{\cup}_{\boldsymbol{\sigma}}((A_{1_0}, A_{1_1})^{\pi_1}, \ldots, (A_{N_0}, A_{N_1})^{\pi_N}) := m_\cup,$$

with $m_\cup$ the result of the disjunctive combination of mass functions m_i with dependence represented by the joint mass function jm determined by vector $\boldsymbol{\sigma}$ and having marginals m_i, $i = 1, \ldots, N$.

When $A_{i_0} = \emptyset$, $i = 1, \ldots, N$, we may denote $\circledcup_{\boldsymbol{\sigma}} \left((A_{1_0}, A_{1_1})^{\pi_1}, \ldots, (A_{N_0}, A_{N_1})^{\pi_N}\right)$ for short as $\circledcup_{\boldsymbol{\sigma}} \left(A_{1_1 1-\pi_1}, \ldots, A_{N_1 1-\pi_N}\right)$, since a binary mass function $(A_{i_0}, A_{i_1})^{\pi_i}$ such that $A_{i_0} = \emptyset$ is nothing but the negative simple mass function $A_{i_1 1-\pi_i}$.

Theorem 6. *Any mass function m defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ with Teugels function t satisfies*

$$m = \circledcup_{\boldsymbol{\sigma}} \left(\{y_1\}_{1-t(\{y_1\})}, \ldots, \{y_n\}_{1-t(\{y_n\})}\right), \tag{55}$$

with $\boldsymbol{\sigma}$ the vector such that $\sigma_1 = 1$ and, for $1 < k \leq 2^n$, $\sigma_k = t(A_k)$ if $|A_k| > 1$, and $\sigma_k = 0$ otherwise.

Proof. The MBD in one-to-one correspondence with the joint mass function jm underlying the disjunctive combination (55), is such that its marginals satisfy $P(X_i = 1) = \pi_i = t(\{y_i\})$, $i = 1, \ldots, n$, and its vector of central moments is $\boldsymbol{\sigma}$. Hence, its Teugels' representation τ is such that $\tau_k = t(A_k)$, $1 < k \leq 2^n$. In other words, the MBD associated with jm is the MBD-equivalent of m, and thus $p_k = m(A_k)$, $1 \leq k \leq 2^n$, from which we obtain $jm(\mathbf{A}_k) = m(A_k)$ since $p_k = jm(\mathbf{A}_k)$, with $\mathbf{A}_k := (A_{1_{k_1}}, \ldots, A_{n_{k_n}})$.

Moreover, for $1 \leq k \leq 2^n$, we have

$$\begin{aligned}
\bigcup_{i=1}^{n} A_{i_{k_i}} &= \Big(\bigcup_{i=1, k_i=0}^{n} A_{i_{k_i}}\Big) \bigcup \Big(\bigcup_{i=1, k_i=1}^{n} A_{i_{k_i}}\Big) \\
&= \bigcup_{i=1, k_i=1}^{n} A_{i_{k_i}} \\
&= \bigcup_{i=1, k_i=1}^{n} \{y_i\} \\
&= A_k.
\end{aligned}$$

Hence, $jm(\mathbf{A}_k) = m(A_k)$ is allocated to A_k. □

Theorem 6 is the counterpart of Theorem 3. It shows that any mass function results from the disjunctive combination of $|\mathcal{Y}|$ negative simple mass functions having some dependence structure.

6 Conclusions

The problem of decomposing uniquely any belief function into elementary items received a solution from Smets [37], extending previous ideas from Shafer [34]. As argued in this paper, Smets' solution has a major weakness, which is that it involves elementary items whose proposed semantics lacks formal justifications.

In Dempster's seminal work [3], a belief function is induced from a space equipped with a probability measure and a multi-valued mapping from this space to another one. In this paper we have considered a particular case of this framework where the probability measure is multivariate Bernoulli and where the multi-valued mapping is the conjunction of multi-valued mappings associated to the Bernoulli random variables underlying the multivariate Bernoulli distribution. Using Pichon *et al.*'s general approach to information fusion [30], we have provided a setting in which this particular case of Dempster's framework receives a concrete meaning: it may be associated to a simple situation where partially reliable sources provide crisp pieces of information. Furthermore, using this latter setting and a representation of the multivariate Bernoulli distribution due to Teugels [42], we have been able to propose a new canonical decomposition of belief functions and an associated new equivalent representation of a belief function which we called the Teugels function. This decomposition resembles Smets' decomposition in that a belief function is also seen as the result of partially reliable pieces of evidence. However, all the elementary items that it involves have well-defined semantics. In a nutshell, a belief function stems according to this decomposition, from as many crisp pieces of information as there are elements in its domain of definition $\mathcal{Y}$, and from simple probabilistic knowledge concerning both the marginal reliability of each of these pieces of information and the dependencies between their reliability. In addition, we showed that computing the Teugels function of a finite random set provides a meaningful canonical decomposition of the random set.

We were also led to consider the weight function associated with Smets' decomposition in light of our framework. We showed that the weight function corresponds to measures of information of the case where the sources involved in our framework are not reliable. This constitutes a completely different perspective on this function than the one proposed by Smets. This new semantics for the weight function is unfortunately less appealing than Smets', but it is well-defined, contrary to Smets'. In addition, we provided a similar semantics for the disjunctive weight function, whose interpretation

had never been discussed.

The t-canonical decomposition and its building blocks open a potentially fruitful path. It seems indeed interesting to revisit in light of our framework, the main problems concerning belief functions and their associated solutions, such as combination [39] and specifically combination under ill-known dependency [11, 4], distance evaluation [18], building methods [1], measurement of uncertainty [22] and of consistency [10].

On the one hand, revisiting an existing solution or notion, as we have done in this paper for instance for the unnormalised Dempster's rule and the equivalent representations of a belief function that are the commonality, implicability, weight and disjunctive weight functions, may provide a new and useful perspective on the solution or notion, as was the case in particular for the weight functions that we were able to relate to information measures.

On the other hand, tackling these main problems using our framework might lead to new solutions for them. For instance, the problem of combining belief functions under ill-known dependency could be looked at in light of the elementary pieces of evidence underlying a belief function according to the t-decomposition. New distances between belief functions, based on Teugels function t, could be investigated. Uncertainty measures and building methods (such as the one used in [43, Section 2.2.1]) for the MBD could yield new uncertainty measures and building methods for belief functions. The conjunctive combination with dependence $\boldsymbol{\sigma}$ of simple mass functions could present some interest in preference modelling as a means to resolve inconsistency between elementary information items, represented by simple mass functions and assumed so far to be independent [9], since the dependence assumption made between pieces of information is known to have potentially a significant impact on consistency [10].

Acknowledgments

The author thanks the anonymous referees for their valuable comments that helped to improve this paper.

References

[1] N. Ben Abdallah, N. Mouhous-Voyneau, and T. Denœux. Combining statistical and expert evidence using belief functions: Application to centennial sea level esti-

mation taking into account climate change. *International Journal of Approximate Reasoning*, 55(1, Part 3):341 – 354, 2014.

[2] I. Couso and D. Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7):1502–1518, 2014.

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

[4] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2008.

[5] T. Denœux. 40 years of Dempster–Shafer theory. *International Journal of Approximate Reasoning*, 79:1 – 6, 2016.

[6] T. Denœux, S. Sriboonchitta, and O. Kanjanatarakul. Evidential clustering of large dissimilarity data. *Knowledge-Based Systems*, 106:179 – 195, 2016.

[7] T. Denœux, Z. Younes, and F. Abdallah. Representing uncertainty on set-valued variables using belief functions. *Artificial Intelligence*, 174(7):479 – 499, 2010.

[8] T. Denœux, N. El Zoghby, V. Cherfaoui, and A. Jouglet. Optimal object association in the Dempster-Shafer framework. *IEEE Trans. Cybernetics*, 44(12):2521 – 2531, 2014.

[9] S. Destercke. A generic framework to include belief functions in preference handling for multi-criteria decision. In A. Antonucci, L. Cholvy, and O. Papini, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 14th European Conference, ECSQARU 2017*, volume 10369 of *Lecture Notes in Artificial Intelligence*, pages 179–189. Springer, 2017.

[10] S. Destercke and T. Burger. Toward an axiomatic definition of conflict between belief functions. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 43(2):585–596, 2013.

[11] S. Destercke and D. Dubois. Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. *Information Sciences*, 181(18):3925 – 3945, 2011.

[12] D. Dubois and H. Prade. A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12(3):193–226, 1986.

[13] D. Dubois, H. Prade, and P. Smets. A definition of subjective possibility. *International Journal of Approximate Reasoning*, 48(2):352 – 364, 2008.

[14] R. Fano. *Transmission of Information: A Statistical Theory of Communications*. MIT Press, Cambridge, MA, 1961.

[15] S. Ferson, R. Nelsen, J. Hajagos, D. Berleant, J. Zhang, W. T. Tucker, L. Ginzburg, and W. L. Oberkampf. Dependence in Probabilistic Modeling, Dempster-Shafer Theory, and Probability Bounds Analysis. Technical Report SAND2004-3072, Sandia National Laboratories, Albuquerque, New Mexico, 2004.

[16] M. Ha-Duong. Hierarchical fusion of expert opinions in the transferable belief model, application to climate sensitivity. *International Journal of Approximate Reasoning*, 49(3):555–574, 2008.

[17] L. Jiao, Q. Pan, T. Denœux, Y. Liang, and X. Feng. Belief rule–based classification system: Extension of FRBCS in belief functions framework. *Information Sciences*, 309:26 – 49, 2015.

[18] A.-L. Jousselme and P. Maupin. Distances in evidence theory: Comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 53(2):118 – 145, 2012.

[19] A. Kallel and S. Le Hégarat-Mascle. Combination of partially non-distinct beliefs: The cautious-adaptive rule. *International Journal of Approximate Reasoning*, 50(7):1000 – 1021, 2009.

[20] O. Kanjanatarakul, T. Denœux, and S. Sriboonchitta. Prediction of future observations using belief functions: A likelihood-based approach. *International Journal of Approximate Reasoning*, 72:71–94, 2016.

[21] D. G. Kendall. Foundations of a theory of random sets. In E. F. Harding and D. G. Kendall, editors, *Stochastic Geometry*, pages 322–376. Wiley, New York, 1974.

[22] G. J. Klir. *Uncertainty and Information: Foundations of Generalized Information Theory*. Wiley-IEEE Press, 2005.

[23] G. Matheron. *Random Sets and Integral Geometry*. Wiley, New York, 1975.

[24] D. Mercier, E. Lefèvre, and F. Delmotte. Belief functions contextual discounting and canonical decompositions. *International Journal of Approximate Reasoning*, 53(2):146–158, 2012.

[25] D. Mercier, E. Lefèvre, and D. Jolly. Object association with belief functions, an application with vehicles. *Information Sciences*, 181(24):5485 – 5500, 2011.

[26] D. Mercier, F. Pichon, and E. Lefèvre. Corrigendum to "Belief functions contextual discounting and canonical decompositions" [International Journal of Approximate Reasoning 53 (2012) 146-158]. *International Journal of Approximate Reasoning*, 70:137–139, 2016.

[27] H. T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978.

[28] F. Pichon and T. Denœux. On latent belief structures. In K. Mellouli, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 4724 of *Lecture Notes in Computer Science*, pages 368–380. Springer, 2007.

[29] F. Pichon and T. Denœux. The unnormalized Dempster's rule of combination: A new justification from the least commitment principle and some extensions. *Journal of Automated Reasoning*, 45(1):61–87, 2010.

[30] F. Pichon, D. Dubois, and T. Denœux. Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53(2):159–175, 2012.

[31] F. Pichon, D. Mercier, E. Lefèvre, and F. Delmotte. Proposition and learning of some belief function contextual correction mechanisms. *International Journal of Approximate Reasoning*, 72:4–42, 2016.

[32] C. Rose and M. D. Smith. *Mathematical Statistics with Mathematica*. Springer-Verlag, New York, 2002.

[33] J. Schubert. Clustering decomposed belief functions using generalized weights of conflict. *International Journal of Approximate Reasoning*, 48(2):466 – 480, 2008.

[34] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

[35] P. Smets. The Transferable Belief Model and random sets. *International Journal of Intelligent Systems*, 7:37–46, 1992.

[36] P. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1–35, 1993.

[37] P. Smets. The canonical decomposition of a weighted belief. In *Proceedings of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI'95), San Mateo, California, USA, 1995*, pages 1896–1901. Morgan Kaufmann, 1995.

[38] P. Smets. The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31(1–2):1–30, 2002.

[39] P. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412, 2007.

[40] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[41] I. Song and S. Lee. Explicit formulae for product moments of multivariate gaussian random variables. *Statistics and Probability Letters*, 100:27–34, 2015.

[42] J. L. Teugels. Some representations of the multivariate Bernoulli and binomial distributions. *Journal of Multivariate Analysis*, 32(2):256 – 268, 1990.

[43] J. L. Teugels and J. Van Horebeek. Algebraic descriptions of nominal multivariate discrete data. *Journal of Multivariate Analysis*, 67(2):203 – 226, 1998.