Belief functions combination without the assumption of independence of the information sources

Marco E. G. V. Cattaneo

Department of Statistics, LMU Munich, Ludwigstraße 33, 80539 München, Germany

Abstract

This paper considers the problem of combining belief functions obtained from not necessarily independent sources of information. It introduces two combination rules for the situation in which no assumption is made about the dependence of the information sources. These two rules are based on cautious combinations of plausibility and commonality functions, respectively. The paper studies the properties of these rules and their connection with Dempster's rules of conditioning and combination and the minimum rule of possibility theory.

Key words: Dempster-Shafer theory, belief functions, combination rule, dependence, conflict, cautious combination

1. Introduction

A central theme of Dempster-Shafer theory [41, 53] is the combination of the information obtained from different sources. The information obtained from a source is described by a belief function, and the belief functions corresponding to independent sources of information can be combined by means of Dempster's rule. The problem is that it is usually not clear if Dempster's rule is applicable, because the meaning of the independence of the information sources is very abstract (this holds for all interpretations of belief functions).

Dempster-Shafer theory allows the description of partial or complete ignorance, since the belief not accorded to a proposition does not have to be accorded to the negation of that proposition. Hence, it is fully in the spirit of the theory to allow ignorance also about the dependence of the information sources: the present paper studies the combination of belief functions in the extreme case of complete ignorance about the dependence of the information sources. Several combination rules have been recently suggested for that extreme case: the rule proposed by Cattaneo [1] replaces the independence assumption with an assumption of maximal consistency (that is, minimal conflict), but is computationally too demanding for many applications; the rule studied by Denœux [11] (based on the concept of weight of evidence) satisfies important properties, but has some difficulties with non-separable belief functions; finally, the rule suggested by Destercke et al. [13] generalizes the minimum rule of possibility theory, but does not respect the fundamental equivalence between belief functions and their vacuous extensions. The combination rules proposed in the present paper are closely related to these three rules, without being subject to the above weaknesses.

An example of a situation in which the assumption of independence of the information sources leads to absurd results is the generalization of Bayes' theorem studied in [1, Section 5] (see also [46, Subsection 5.2]). In fact, in that situation the conflict of the combination appears to play the role of a measure of disagreement between the involved belief functions, and the conflict of Dempster's rule is not a good measure of disagreement (see for instance [34]). A much better measure of disagreement among belief functions is the minimal conflict defined in Section 2 of the present paper. More generally, Section 2 contains some mathematical definitions and results, in particular on specializations of belief functions and on two classes of Fréchet bounds. The proofs of the theorems are given in Appendix A.

In general, the information sources cannot be independent, because the pieces of information obtained from them are about the same topic, and the independence assumption can thus lead to partial inconsistency (that is, conflict).

Email address: cattaneo@stat.uni-muenchen.de (Marco E. G. V. Cattaneo)

This is the unformatted version of: doi:10.1016/j.ijar.2010.10.006

Hence, the result of Dempster's rule should be interpreted as an approximation by a belief function of the conflictual combination actually resulting from the independence assumption. This is the subject of Section 3, and is important for the justification of the combination rules proposed in Section 5 for the situation in which no assumption is made about the dependence of the information sources, since the results of those rules are approximations by belief functions of most cautious descriptions of the combined information. The rules of Section 5 satisfy many important properties, discussed in Section 4 in relation to the vast literature on combination rules for belief functions.

2. Some mathematical definitions and results

Let S be a finite, nonempty set. The complement of $A \subseteq S$ in S is denoted by $\overline{A}$, while 2^A and $|A|$ denote the power set and the cardinality of A, respectively. An *ordered partition* of $A \subseteq S$ into $n \geq 1$ subsets is an n-tuple $(B_1, \ldots, B_n) \in (2^A)^n$ of pairwise disjoint subsets of A whose union is A (note that the non-emptiness of the B_i's is not assumed). Let $\mathcal{OP}_n(A)$ denote the set of all ordered partitions of A into n subsets.

A function $\mu : 2^S \to [0, 1]$ is said to be *monotonic* if $A \subseteq B \subseteq S$ implies $\mu(A) \leq \mu(B)$, while μ is said to be *anti-monotonic* if $A \subseteq B \subseteq S$ implies $\mu(A) \geq \mu(B)$; moreover, μ is said to be *subadditive* if $\mu(A \cup B) \leq \mu(A) + \mu(B)$ for all disjoint $A, B \subseteq S$, while μ is said to be *quasi-superadditive* if $\mu(A \cup B) \geq \mu(A) + \mu(B) - 1$ for all disjoint $A, B \subseteq S$; finally, μ is said to be 2-*alternating* if $\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B)$ for all $A, B \subseteq S$. Let $\mathcal{MS}_0(S)$ denote the set of all monotonic, subadditive functions $\mu : 2^S \to [0, 1]$ with $\mu(\varnothing) = 0$; and let $\mathcal{AQ}_1(S)$ denote the set of all anti-monotonic, quasi-superadditive functions $\mu : 2^S \to [0, 1]$ with $\mu(\varnothing) = 1$.

A *basic belief assignment (bba)* is a function $m : 2^S \to [0, 1]$ such that $\sum_{A \subseteq S} m(A) = 1$. The value $m(A)$ is the total belief mass assigned to A (without being assigned to any proper subset of A). The *conflict* of a bba m is the total belief mass $m(\varnothing)$ assigned to the empty set. A bba m is said to be *normal* if it has no conflict (that is, $m(\varnothing) = 0$). A bba m with $m(\varnothing) < 1$ can be normalized by setting $m(\varnothing)$ to 0 and rescaling the resulting function on 2^S to a bba.

The *belief function* $Bel : 2^S \to [0, 1]$, *plausibility function* $Pl : 2^S \to [0, 1]$, and *commonality function* $Q : 2^S \to [0, 1]$ associated with a bba m on 2^S satisfy

$$Bel(A) = \sum_{\substack{B \subseteq A: \\ B \neq \varnothing}} m(B), \quad Pl(A) = \sum_{\substack{B \subseteq S: \\ A \cap B \neq \varnothing}} m(B), \quad \text{and} \quad Q(A) = \sum_{\substack{B \subseteq S: \\ A \subseteq B}} m(B),$$

respectively, for all $A \subseteq S$. The value $Bel(A)$ is the total non-conflictual belief mass assigned to A or its subsets, while the value $Pl(A)$ is the total belief mass not assigned to $\overline{A}$ or its subsets. Hence, $Bel \leq Pl$ (in this paper, expressions involving functions without explicit arguments are to be interpreted pointwise), $Bel(\varnothing) = Pl(\varnothing) = 0$, and $Bel(A) + Pl(\overline{A}) = 1 - m(\varnothing)$ for all $A \subseteq S$. Moreover, $Pl \in \mathcal{MS}_0(S)$ and Pl is 2-alternating, while $Q \in \mathcal{AQ}_1(S)$ and Q satisfies $\sum_{A \subseteq S} (-1)^{|A|} Q(A) = m(\varnothing)$ (see for example [41, page 42]). When m_x is a bba, then Bel_x, Pl_x, and Q_x denote the belief, plausibility, and commonality functions associated with m_x, respectively.

A *refinement* of S is a pair (R, r) where R is a finite set and $r : S \to 2^R \setminus \{\varnothing\}$ is a mapping such that the images of the elements of S are pairwise disjoint and their union is R (that is, the images of the elements of S build a nonempty partition of R). When m is a bba on 2^S and (R, r) is a refinement of S, the *vacuous extension* of m to 2^R by means of r is the bba $m^{\uparrow(R,r)}$ on 2^R such that $m^{\uparrow(R,r)}(\varnothing) = m(\varnothing)$ and $m^{\uparrow(R,r)}\left(\bigcup_{x \in A} r(x)\right) = m(A)$ for all nonempty $A \subseteq S$. The plausibility and commonality functions associated with $m^{\uparrow(R,r)}$ are $Pl \circ \tilde{r}$ and $Q \circ \tilde{r}$, respectively, where $\tilde{r} : 2^R \to 2^S$ is defined by $\tilde{r}(A) = \{x \in S : r(x) \cap A \neq \varnothing\}$ for all $A \subseteq R$.

The *focal sets* of a bba m on 2^S are the $A \subseteq S$ such that $m(A) > 0$. The *core* $C(m)$ of a bba m on 2^S is the union of its focal sets; that is, $C(m)$ is the smallest $A \subseteq S$ such that $Pl(\overline{A}) = 0$. A bba m is said to be *consonant* if its focal sets are nested. The *contour function* $\pi : S \to [0, 1]$ associated with a consonant bba m on 2^S satisfies $\pi(x) = Pl(\{x\}) = Q(\{x\})$ for all $x \in S$; therefore, $Pl(A) = \max_{x \in A} \pi(x)$ and $Q(A) = \min_{x \in A} \pi(x)$, for all nonempty $A \subseteq S$. The *simple* bba $m_{A,\alpha}$ on 2^S is the consonant bba defined by $m_{A,\alpha}(A) = \alpha$ and $m_{A,\alpha}(S) = 1 - \alpha$, where $A \subset S$ and $\alpha \in [0, 1]$. In particular, $m_{A,0}$ does not depend on A, and is called the *vacuous* bba on 2^S.

2.1. Joint belief assignments and specializations

Let $m_1, \ldots, m_n$ be $n \geq 1$ bba's on 2^S. A *joint belief assignment (jba)* with marginals $m_1, \ldots, m_n$ is a function $J : (2^S)^n \to [0,1]$ such that

$$\sum_{\substack{B_1,\ldots,B_n \subseteq S:\\ B_i = A}} J(B_1, \ldots, B_n) = m_i(A)$$

for all $A \subseteq S$ and all $i \in \{1, \ldots, n\}$ (hence, in particular, $\sum_{B_1,\ldots,B_n \subseteq S} J(B_1, \ldots, B_n) = 1$). Let $\mathcal{J}(m_1, \ldots, m_n)$ denote the set of all jba's with marginals $m_1, \ldots, m_n$. A jba J on $(2^S)^n$ induces a bba m_J on 2^S defined by

$$m_J(A) = \sum_{\substack{B_1,\ldots,B_n \subseteq S:\\ B_1 \cap \cdots \cap B_n = A}} J(B_1, \ldots, B_n)$$

for all $A \subseteq S$. The induced bba m_J is obtained from J by assigning the belief mass $J(B_1, \ldots, B_n)$ to the set $B_1 \cap \cdots \cap B_n$, for all $B_1, \ldots, B_n \subseteq S$.

A bba m on 2^S can be interpreted as the probability distribution of a random subset X of S with respect to some probability measure P, where $(\Omega, \mathcal{A}, P)$ is some underlying probability space and $X : \Omega \to 2^S$ is a random object. The bba m is normal if and only if X is nonempty a.s. (that is, $P\{X \neq \varnothing\} = 1$), and normalizing m corresponds to conditioning P on $\{X \neq \varnothing\}$. With this interpretation, $Bel(A) = P\{X \subseteq A,\, X \neq \varnothing\}$, $Pl(A) = P\{X \cap A \neq \varnothing\}$, and $Q(A) = P\{A \subseteq X\}$, for all $A \subseteq S$.

If the bba's $m_1, \ldots, m_n$ on 2^S are interpreted as the probability distributions of the random subsets $X_1, \ldots, X_n$ of S, respectively, then the jba's $J \in \mathcal{J}(m_1, \ldots, m_n)$ correspond to the possible joint probability distributions of $X_1, \ldots, X_n$, and m_J to the resulting probability distribution of $X_1 \cap \cdots \cap X_n$. In particular, the jba I corresponding to the independence of $X_1, \ldots, X_n$ is defined by $I(B_1, \ldots, B_n) = m_1(B_1) \cdots m_n(B_n)$ for all $B_1, \ldots, B_n \subseteq S$; in this case, the commonality function Q_I associated with the induced bba m_I satisfies, for all $A \subseteq S$,

$$Q_I(A) = P\{A \subseteq X_1 \cap \cdots \cap X_n\} = P\{A \subseteq X_1\} \cdots P\{A \subseteq X_n\} = Q_1(A) \cdots Q_n(A). \tag{1}$$

A bba m_s on 2^S is a *specialization* of a bba m on 2^S if there is a jba $J \in \mathcal{J}(m_s, m)$ such that $J(B_s, B) > 0$ implies $B_s \subseteq B$, for all $B_s, B \subseteq S$. In this case, $m_J = m_s$, because

$$m_J(A) = \sum_{\substack{B_s,B \subseteq S:\\ B_s \cap B = A}} J(B_s, B) = \sum_{\substack{B_s,B \subseteq S:\\ B_s = A}} J(B_s, B) = m_s(A)$$

for all $A \subseteq S$. Hence, a specialization of a bba on 2^S is obtained by transferring some belief mass from A to some of its subsets, for all $A \subseteq S$. As a consequence, if m_s is a specialization of m, then $Pl_s \leq Pl$ and $Q_s \leq Q$, and all specializations of m_s are also specializations of m.

Let $\mathcal{S}(m)$ denote the set of all specializations of a bba m. It can be easily proved that

$$Pl = \max_{m_s \in \mathcal{S}(m)} Bel_s \quad \text{and} \quad Q = \max_{m_s \in \mathcal{S}(m)} m_s. \tag{2}$$

Hence, for each $A \subseteq S$, the value $Pl(A)$ is the maximum amount of non-conflictual belief mass that can be assigned to A or its subsets by specializing m, while $Q(A)$ is the maximum amount of belief mass that can be assigned to A by specializing m. The following result points out the strong relationship between common specializations and jba's.

Theorem 1. *Let $m_1, \ldots, m_n$ be $n \geq 1$ bba's on 2^S. A bba on 2^S is a common specialization of $m_1, \ldots, m_n$ if and only if it is a specialization of a bba m_J induced by some jba J with marginals $m_1, \ldots, m_n$; that is,*

$$\mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n) = \bigcup_{J \in \mathcal{J}(m_1,\ldots,m_n)} \mathcal{S}(m_J).$$

2.2. Two binary operators

The binary operators $\curlywedge$ and $\wedge$ on the set of all real functions on 2^S are defined by

$$(\mu \curlywedge \nu)(A) = \min_{B \subseteq A} (\mu(B) + \nu(A \setminus B)) \quad \text{and} \quad (\mu \wedge \nu)(A) = \min\{\mu(A), \nu(A)\},$$

respectively, for all $A \subseteq S$ and all real functions μ, ν on 2^S. It can be easily checked that $\curlywedge$ and $\wedge$ are commutative and associative, with

$$(\mu_1 \curlywedge \cdots \curlywedge \mu_n)(A) = \min_{(B_1,\ldots,B_n) \in \mathcal{OP}_n(A)} (\mu_1(B_1) + \cdots + \mu_n(B_n)) \quad \text{and} \quad (\mu_1 \wedge \cdots \wedge \mu_n)(A) = \min\{\mu_1(A), \ldots, \mu_n(A)\},$$

for all $n \geq 1$, all $A \subseteq S$, and all real functions $\mu_1, \ldots, \mu_n$ on 2^S. Hence, in particular,

$$\mu_1 \curlywedge \cdots \curlywedge \mu_n \leq \mu_1 \wedge \cdots \wedge \mu_n$$

for all $n \geq 1$ and all real functions $\mu_1, \ldots, \mu_n$ on 2^S such that $\mu_1(\varnothing) = \cdots = \mu_n(\varnothing) = 0$. That is, $\mu_1 \curlywedge \cdots \curlywedge \mu_n$ is bounded above by the pointwise minimum of $\mu_1, \ldots, \mu_n$, if $\mu_1(\varnothing) = \cdots = \mu_n(\varnothing) = 0$, and this is the case in particular when $\mu_1, \ldots, \mu_n \in \mathcal{MS}_0(S)$.

Theorem 2. *$\curlywedge$ is a commutative and associative binary operator on $\mathcal{MS}_0(S)$. Moreover, for all $\mu, \nu \in \mathcal{MS}_0(S)$, if (R, r) is a refinement of S, then $(\mu \circ \tilde{r}) \curlywedge (\nu \circ \tilde{r}) = (\mu \curlywedge \nu) \circ \tilde{r}$; and if $\mu \leq \nu$, then $\mu \curlywedge \nu = \mu$.*

The following simple result is an analogue of Theorem 2 for the pointwise minimum operator $\wedge$.

Theorem 3. *$\wedge$ is a commutative and associative binary operator on $\mathcal{AQ}_1(S)$. Moreover, for all $\mu, \nu \in \mathcal{AQ}_1(S)$, if (R, r) is a refinement of S, then $(\mu \circ \tilde{r}) \wedge (\nu \circ \tilde{r}) = (\mu \wedge \nu) \circ \tilde{r}$; and if $\mu \leq \nu$, then $\mu \wedge \nu = \mu$.*

2.3. Fréchet bounds

Let $m_1, \ldots, m_n$ be $n \geq 1$ bba's on 2^S and let A be a subset of S. If $m_1, \ldots, m_n$ are interpreted as the probability distributions of the random subsets $X_1, \ldots, X_n$ of S, respectively, then

$$\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J(A) \quad \text{and} \quad \max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J(A) \tag{3}$$

are the maximum values of $P\{X_1 \cap \cdots \cap X_n \cap A \neq \varnothing\}$ and $P\{A \subseteq X_1 \cap \cdots \cap X_n\}$, respectively, over all possible joint probability distributions of $X_1, \ldots, X_n$. Such maxima are called Fréchet bounds in probability theory (see for example [39]). The following theorem states in particular that $(Pl_1 \curlywedge \cdots \curlywedge Pl_n)(A)$ corresponds to the first of the two Fréchet bounds (3) when $n \leq 2$, and bounds it from above when $n \geq 3$.

Theorem 4. *If $m_1, \ldots, m_n$ are $n \geq 1$ bba's on 2^S, then*

$$\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J = \max_{m_s \in \mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)} Pl_s \leq Pl_1 \curlywedge \cdots \curlywedge Pl_n.$$

Moreover, the above inequality is actually an equality when $n \leq 2$; that is, in particular,

$$\max_{J \in \mathcal{J}(m_1,m_2)} Pl_J = \max_{m_s \in \mathcal{S}(m_1) \cap \mathcal{S}(m_2)} Pl_s = Pl_1 \curlywedge Pl_2.$$

The equality between the first of the two Fréchet bounds (3) and $(Pl_1 \curlywedge \cdots \curlywedge Pl_n)(A)$ when $n \leq 2$ can be deduced from a theorem by Strassen [49, Theorem 11], who noted that his theorem cannot be straightforwardly extended to the case with $n \geq 3$ (see also [44]). In fact, the following counterexample implies that for no $n \geq 3$ the inequality in the first part of Theorem 4 is always an equality.

Example 1. *Choose $n \geq 3$ and define $S = \{1, \ldots, n\}$. For each $i \in S$ let m_i be the bba on 2^S assigning the belief mass $\frac{1}{2}$ to both the singleton $\{i\}$ and its complement; that is, $m_i(\{i\}) = m_i(\overline{\{i\}}) = \frac{1}{2}$ for all $i \in S$.*

Then $(Pl_1 \curlywedge \cdots \curlywedge Pl_n)(S) = 1$, because $m_1, \ldots, m_n$ are normal and $Pl_i(A) \geq \frac{1}{2}$ for all nonempty $A \subset S$ and all $i \in S$. But if m_s is a common specialization of $m_1, \ldots, m_n$, then m_s can assign belief mass only to the n singletons and to $\varnothing$, and it can be easily checked that $Pl_s(S)$ is maximal when $m_s(\{i\}) = \frac{1}{2}\frac{1}{n-1}$ for all $i \in S$. Therefore,

$$\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J(S) = \max_{m_s \in \mathcal{S}(m_1)\cap\cdots\cap\mathcal{S}(m_n)} Pl_s(S) = \frac{1}{2}\frac{n}{n-1} \leq \frac{3}{4} < 1 = (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(S).$$

In spite of the simplicity of Example 1, no difference between $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ and $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ has been observed in thousands of randomly generated numerical examples (with various generating probability distributions). This suggests that in general $(Pl_1 \curlywedge \cdots \curlywedge Pl_n)(A)$ is a very good upper approximation of the first of the two Fréchet bounds (3). The second one is simpler: it corresponds to $(Q_1 \wedge \cdots \wedge Q_n)(A)$, as implied by the following result.

Theorem 5. *If $m_1, \ldots, m_n$ are $n \geq 1$ bba's on 2^S, then*

$$\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J = \max_{m_s \in \mathcal{S}(m_1)\cap\cdots\cap\mathcal{S}(m_n)} Q_s = Q_1 \wedge \cdots \wedge Q_n.$$

2.4. Minimal conflict

The *minimal conflict* of $n \geq 1$ bba's $m_1, \ldots, m_n$ on 2^S is the value

$$c_{\min}(m_1, \ldots, m_n) = \min_{J \in \mathcal{J}(m_1,\ldots,m_n)} m_J(\varnothing) = \min_{m_s \in \mathcal{S}(m_1)\cap\cdots\cap\mathcal{S}(m_n)} m_s(\varnothing),$$

where the second equality is implied by Theorem 4, since $m(\varnothing) = 1 - Pl(S)$. Hence, Theorem 4 implies also

$$c_{\min}(m_1, \ldots, m_n) \geq 1 - (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(S) = \max_{(B_1,\ldots,B_n) \in \mathcal{OP}_n(S)} (1 - Pl_1(B_1) - \cdots - Pl_n(B_n)).$$

As noted above, in general $1 - (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(S)$ seems to be a very good lower approximation of $c_{\min}(m_1, \ldots, m_n)$.

The second part of Theorem 4 implies that the minimal conflict of the bba's m_1, m_2 on 2^S is $1 - (Pl_1 \curlywedge Pl_2)(S)$; that is,

$$c_{\min}(m_1, m_2) = m_1(\varnothing) + \max_{A \subseteq S} (Bel_1(A) - Pl_2(A)). \tag{4}$$

In the case with $|S|$ a power of 2 and m_1, m_2 normal, the expression (4) was proved directly in [1, Proposition 2]. The equality between the first of the two Fréchet bounds (3) and $(Pl_1 \curlywedge \cdots \curlywedge Pl_n)(A)$ when $n \leq 2$ can be easily deduced from this result.

The minimal conflict $c_{\min}(m_1, \ldots, m_n) \in [0, 1]$ can be interpreted as a measure of disagreement between the bba's $m_1, \ldots, m_n$. The following theorem collects some important properties of this measure of disagreement.

Theorem 6. *Let $m_1, \ldots, m_n$ be $n \geq 1$ bba's on 2^S. Then $c_{\min}(m_1, \ldots, m_n) = 0$ if and only if $m_1, \ldots, m_n$ have a common normal specialization; while $c_{\min}(m_1, \ldots, m_n) = 1$ if and only if $C(m_1) \cap \cdots \cap C(m_n) = \varnothing$. If (R, r) is a refinement of S, then*

$$c_{\min}(m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)}) = c_{\min}(m_1, \ldots, m_n).$$

Moreover, if $n \geq 2$ and m_1 is a specialization of m_n, then $c_{\min}(m_1, \ldots, m_n) = c_{\min}(m_1, \ldots, m_{n-1})$; and in particular, if m_1 is a common specialization of $m_2, \ldots, m_n$, then $c_{\min}(m_1, \ldots, m_n) = m_1(\varnothing)$.

Theorem 6 implies in particular that if $c_{\min}(m_1, \ldots, m_n) = 0$, then $m_1, \ldots, m_n$ are normal and $Bel_1 \vee \cdots \vee Bel_n \leq Pl_1 \wedge \cdots \wedge Pl_n$ (where $\vee$ denotes the pointwise maximum operator), since $m_1, \ldots, m_n$ have a common normal specialization. The converse holds when $n \leq 2$, as follows from the expression (4); but the following counterexample implies that it does not hold for any $n \geq 3$.

Example 2. *In the situation of Example 1, $m_1, \ldots, m_n$ are normal and $Bel_1 \vee \cdots \vee Bel_n \leq Pl_1 \wedge \cdots \wedge Pl_n$, since*

$$(Bel_1 \vee \cdots \vee Bel_n)(A) = (Pl_1 \wedge \cdots \wedge Pl_n)(A) = \begin{cases} 0 & \text{if } A = \varnothing, \\ 1 & \text{if } A = S, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Therefore, $c_{\min}(m_i, m_j) = c_{\min}(m_i) = 0$ for all $i, j \in S$; but $c_{\min}(m_1, \ldots, m_n) = 1 - \frac{1}{2}\frac{n}{n-1} \geq \frac{1}{4} > 0$.

3. On the assumption of independence of the information sources

In Dempster-Shafer theory, a piece of information about the uncertain value of $x \in S$ is usually described by a normal bba m on 2^S: for each $A \subseteq S$, the value $m(A)$ is the total belief mass assigned to the proposition "$x \in A$" without being assigned to any more specific proposition. A central theme of the theory is the combination of the information obtained from different sources; that is, the combination of several normal bba's on 2^S (describing the information obtained from different sources) into a single normal bba on 2^S (describing the combined information).

If the normal bba's $m_1, \ldots, m_n$ on 2^S are interpreted as the probability distributions of the random subsets $X_1, \ldots, X_n$ of S, respectively (that is, the belief masses are interpreted probabilistically), then the combination of $m_1, \ldots, m_n$ is the probability distribution of $X_1 \cap \cdots \cap X_n$. The possible joint probability distributions of $X_1, \ldots, X_n$ correspond to the jba's $J \in \mathcal{J}(m_1, \ldots, m_n)$, and the resulting combination of $m_1, \ldots, m_n$ is m_J. In particular, when the independence of $X_1, \ldots, X_n$ is assumed, the combination of $m_1, \ldots, m_n$ is m_I (where $I \in \mathcal{J}(m_1, \ldots, m_n)$ is the jba corresponding to the independence of $X_1, \ldots, X_n$), but m_I can be non-normal. Normalizing m_I corresponds to conditioning on $\{X_1 \cap \cdots \cap X_n \neq \emptyset\}$, but in general $X_1, \ldots, X_n$ are not independent anymore when conditioned on $\{X_1 \cap \cdots \cap X_n \neq \emptyset\}$. Hence, the independence assumption and the normality of the combination are not compatible in general.

3.1. On the possibility of independence

Dempster [8] interpreted the belief masses probabilistically, and avoided the problem of the incompatibility between the independence assumption and the normality of the bba's by allowing non-normal bba's and including the normalization step in the definition of *Bel*, *Pl*, and Q. That is, Dempster defined the belief, plausibility, and commonality functions (under other names) associated with a bba m on 2^S as Bel_n, Pl_n, and Q_n, respectively, where m_n is the normalized version of m. This corresponds to using equivalence classes of bba's instead of bba's: an equivalence class contains all bba's leading to the same normal bba when normalized. In particular, the independence assumption poses no problem: the combination of the equivalence classes of bba's represented by $m_1, \ldots, m_n$ is the equivalence class of bba's represented by m_I. If we use the normal bba's as representatives of the equivalence classes, then the combination of the normal bba's $m_1, \ldots, m_n$ is the normalized version of m_I; that is, we obtain Dempster's rule of combination. However, this mathematical expedient does not solve the real problem of the impossibility of independence for the normal bba's (on which *Bel*, *Pl*, and Q are based). Moreover, the approach with equivalence classes of bba's is not applicable when the independence is not assumed, because the normalized possible combinations of $m_1, \ldots, m_n$ (that is, the normalized versions of m_J, for all jba's $J \in \mathcal{J}(m_1, \ldots, m_n)$) depend on the conflicts $m_1(\emptyset), \ldots, m_n(\emptyset)$ of the particular representatives of the respective equivalence classes.

In his monograph [41], Shafer abandoned the probabilistic interpretation of the belief masses, and justified Dempster's rule of combination on intuitive grounds, by interpreting geometrically the independence of the information sources. However, this geometrical interpretation does not avoid the problem of the incompatibility between the independence assumption and the normality of the combination. Later [42], Shafer gave a new interpretation of the belief masses, implicitly relying on fiducial probability, but it is not clear if fiducial inference is applicable for the combination of $m_1, \ldots, m_n$ when independence is assumed and $m_I(\emptyset) > 0$ (see for instance [27, 37]). Hence, the question of the compatibility between the independence assumption and the normality of the combination remains open.

Smets [45] followed [41] in the non-probabilistic interpretation of the belief masses, and replaced the independence assumption by equivalent technical requirements. Like Dempster, Smets avoided the problem of the incompatibility with the normality of the bba's by allowing non-normal bba's, but differently from Dempster, he did not include the normalization step in the definition of *Bel*, *Pl*, and Q. However, all uses of the non-normal bba's (such as the decisions based on the pignistic transformation [48]) seem to involve an explicit or implicit normalization, and under the usual closed-world assumption we recover Dempster's rule of combination. Hence, as in the approach with equivalence classes of bba's, the mathematical expedient of non-normality does not solve the real problem of the impossibility of independence for the normal bba's (on which all uses of the model seem to be actually based).

3.2. Approximating independence

In general, the normalization step in Dempster's rule of combination seems to be justified only as an approximation step: the result of the combination of $m_1, \ldots, m_n$ can be interpreted as the best approximation of m_I by a normal bba on 2^S (this idea is studied also in [31]). That is, in general Dempster's rule of combination corresponds to an assumption of approximate independence; in particular, when the conflict $m_I(\varnothing)$ is large, the reasonableness of that assumption can be questionable (because the approximation would be poor), and the result of Dempster's rule of combination can be rather arbitrary (compare with [41, page 254]).

Anyway, the independence assumption is always problematic, because the meaning of the dependence between the information sources is very abstract (this holds for all interpretations of the belief masses). For the same reason, the assumption of other specific dependence structures for the information sources (as described for example in [35]) is usually even more problematic. In this paper we study combination rules for the situation in which no assumption is made about the dependence of the information sources.

4. On combination rules and their properties

Let $m_1, \ldots, m_n$ be $n \geq 1$ normal bba's on 2^S. We study rules for combining $m_1, \ldots, m_n$ into a normal bba $\langle m_1, \ldots, m_n \rangle$ on 2^S, with $\langle m_1 \rangle = m_1$ when $n = 1$. Many combination rules have been proposed in the literature: see for instance [40, 47] for partial reviews. However, most of these rules require the independence of the information sources: exceptions are for example the combination rules proposed in [33, 29, 20, 1, 31, 35, 11, 13, 21, 28, 12].

The above formulation imposes some strong constraints on the rules considered. First of all, the combination $\langle m_1, \ldots, m_n \rangle$ can only depend on the bba's $m_1, \ldots, m_n$, while for example the rules studied in [33, 35, 21, 28] require additional information about the dependence of the information sources. Secondly, the combination $\langle m_1, \ldots, m_n \rangle$ must be defined for all finite, nonempty sets S, and all normal bba's $m_1, \ldots, m_n$ on 2^S, while for instance the rules proposed in [1, 25] impose constraints on S, and the rules studied in [29, 20, 11] are limited to particular classes of bba's. Finally, the combination $\langle m_1, \ldots, m_n \rangle$ must be a normal bba on 2^S, while for example the rules proposed in [6, 12] do not lead in general to a bba, and the rules studied in [45, 13] can lead to a non-normal bba.

It seems that no combination rule satisfying these three constraints without assuming the independence of the information sources has ever been published. However, we obtain such a rule if we generalize to all finite, nonempty sets S the combination rule proposed by Cattaneo [1], if we extend also to dogmatic bba's the normalized cautious rule studied by Denœux [11], or if we completely specify (by introducing a second criterion after maximal expected cardinality) and normalize the combination suggested by Destercke et al. [13].

4.1. Basic requirements

We are considering the combination of n bba's, without assuming that these are ordered in any particular way. Hence, a basic requirement for combination rules is *commutativity*: a rule is commutative if

$$\langle m_{\pi(1)}, \ldots, m_{\pi(n)} \rangle = \langle m_1, \ldots, m_n \rangle$$

for all permutations π of $\{1, \ldots, n\}$. Most combination rules in the literature are binary: the combination of n bba's is obtained through $n - 1$ applications of the binary rule. The commutativity of the resulting n-ary combination rule is implied by the commutativity and associativity of the binary rule; hence, for binary rules associativity is also fundamental. Many binary rules proposed in the literature are commutative, but not associative: however, most of them could be extended to commutative n-ary combination rules. In general, associativity is not so fundamental for n-ary combination rules as it is for binary ones, but it can have important advantages from the computational point of view.

The simultaneous consideration of different frames of discernment on which the same beliefs are described is a central feature of Dempster-Shafer theory (see for example [41, 43]). In particular, the ability of describing exactly the same information also on more refined frames of discernment is a fundamental property of the theory. In fact, if m is a bba on 2^S, and (R, r) is a refinement of S, then m and its vacuous extension $m^{\uparrow(R,r)}$ describe the same information, while for instance two probability distributions on R and S, respectively, cannot describe exactly the same information when $|R| > |S|$. Since a bba and its vacuous extensions describe the same information, they can be considered as equivalent. The original theory by Dempster and Shafer respects this equivalence, while surprisingly many methods proposed by other authors in the literature on Dempster-Shafer theory do not respect it. For example, the equivalence between a bba and its vacuous extensions is not respected by the pignistic transformation [48] or any other transformation of belief functions in probability distributions (such as those studied in [7, 3, 4, 5]), by the second component of the measure of conflict proposed by Liu [34], or by the combination rule suggested by Destercke et al. [13]. In fact, that rule does not satisfy the basic requirement of *equivariance with respect to vacuous extensions*: a combination rule is equivariant with respect to vacuous extensions if

$$\langle m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)} \rangle = \langle m_1, \ldots, m_n \rangle^{\uparrow(R,r)}$$

for all refinements (R, r) of S. This basic requirement implies in particular that all focal sets of $\langle m_1, \ldots, m_n \rangle$ are elements of the algebra of subsets of S generated by all focal sets of $m_1, \ldots, m_n$ (since $m_1, \ldots, m_n$ can be seen as vacuous extensions to 2^S of n bba's on that algebra of subsets). Hence, besides the above theoretical justification, the equivariance with respect to vacuous extensions can also have important advantages from the computational point of view. The combination rule proposed by Cattaneo [1] is equivariant with respect to vacuous extensions in the restricted framework of [1], but in order to maintain this property when generalizing the rule to all finite, nonempty sets S, the measure of nonspecificity should be replaced with a more suitable measure of noncommitment.

Another central feature of Dempster-Shafer theory is that it generalizes propositional logic (see for example [9, 36, 30, 1]). When a normal bba m on 2^S describes a piece of information about the uncertain value of $x \in S$, the proposition "$x \in A$" is certain according to that piece of information if and only if $C(m) \subseteq A$ (that is, if and only if $Bel(A) = 1$). In propositional logic, the certainty of a proposition is preserved when new information is acquired. Hence, a combination rule generalizes propositional logic only if it satisfies the basic requirement of *certainty preservation*: a rule preserves certainty if

$$C(\langle m_1, \ldots, m_n \rangle) \subseteq C(m_1) \cap \cdots \cap C(m_n)$$

when the right-hand side is not empty (that is, when $c_{\min}(m_1, \ldots, m_n) < 1$). We could have restricted the definition of the combination rules to the case with $c_{\min}(m_1, \ldots, m_n) < 1$ (that is, the case in which all certainties described by $m_1, \ldots, m_n$ are compatible), but it is simpler to require $\langle m_1, \ldots, m_n \rangle$ to be defined even when $c_{\min}(m_1, \ldots, m_n) = 1$: for example as the vacuous bba on 2^S (we tacitly assume that this is the case for those rules that are usually not defined when $c_{\min}(m_1, \ldots, m_n) = 1$, such as Dempster's rule of combination). Certainty preservation is useful because it allows Dempster-Shafer theory to handle certain as well as uncertain knowledge, but it is important to underline that the degree of belief 1 should be assigned only to absolutely certain propositions. In particular, any combination rule that preserves certainty gives the same result as Dempster's rule of combination in the medical diagnosis example of Zadeh [55]. Even though it is the only reasonable result in that example (see for instance [43, 22, 23]), several authors have suggested alternative combination rules in order to "correct" it (for example in [51, 18, 32, 25, 6, 26]): of course, none of these rules preserves certainty.

Dempster's rule of combination satisfies the above three basic requirements (commutativity, equivariance with respect to vacuous extensions, and certainty preservation), and so do the combination rule proposed by Cattaneo [1] and the normalized cautious rule studied by Denœux [11], when suitably extended to the present framework. It is interesting to note that these basic requirements for n-ary combination rules are very similar to the criteria for binary rules considered by Smets [47, Section A.4]: that is, commutativity and associativity (implying the commutativity of the resulting n-ary combination rules), equivariance with respect to vacuous extensions (called "resistance to refinement"), a special case of certainty preservation (called "plausibility of false"), and two criteria strictly related to it ("duplicate conditioning" and "iterated conditioning").

4.2. *Absorption and idempotency*

All combination rules that have been proposed for the situation in which no assumption is made about the dependence of the information sources (such as the rules studied in [29, 20, 1, 11, 13, 12]) seem to satisfy the property of *idempotency*: a rule is idempotent if

$$m_1 = \cdots = m_n \quad \Rightarrow \quad \langle m_1, \ldots, m_n \rangle = m_1.$$

When the bba's $m_1, \ldots, m_n$ on 2^S are equal, they could describe exactly the same piece of information (in which case the sources would be completely dependent), or the information sources could be independent or have any other dependence structure. A combination rule that does not use additional information about the dependence of the sources cannot distinguish among these cases: the idempotency corresponds to the cautious choice of assuming that the total amount of information is the minimum possible. In fact, the total amount of information is minimal when

$m_1, \ldots, m_n$ describe exactly the same piece of information, and in this case $\langle m_1, \ldots, m_n \rangle = m_1$ describes the combined information.

The cautious choice of assuming that the total amount of information is the minimum possible leads also to the more general property of *absorption*: a combination rule is absorbing if

$$m_1 \in \mathcal{S}(m_n) \quad \Rightarrow \quad \langle m_1, \ldots, m_n \rangle = \langle m_1, \ldots, m_{n-1} \rangle$$

for all $n \geq 2$. In fact, if m_1 is a specialization of m_n, then m_1 could describe the same information as m_n plus some additional information. Hence, when the total amount of information is the minimum possible, the combined information of $m_1, \ldots, m_{n-1}$ corresponds to the combined information of $m_1, \ldots, m_n$. The property of absorption implies by induction that if m_1 is a common specialization of $m_2, \ldots, m_n$, then $\langle m_1, \ldots, m_n \rangle = m_1$; therefore, in particular, absorption implies idempotency. The property of absorption is satisfied by the combination rules studied in [20, 1, 13, 12], but not by the idempotent rules proposed in [29, 11]. In particular, the following example shows that in the fundamental problem of the combination of two simple bba's, the cautious rule suggested by Denœux [11] gives very often the same result as Dempster's rule of combination (that is, the result obtained when the two sources of information are assumed to be independent).

Example 3. *Let A, B be nonempty, proper subsets of S, and let $\alpha, \beta \in (0, 1)$ satisfy $\alpha \geq \beta$. If $A \subsetneq B$, then the simple bba $m_{A,\alpha}$ on 2^S is a specialization of the simple bba $m_{B,\beta}$ on 2^S, and therefore the combination of $m_{A,\alpha}, m_{B,\beta}$ is $m_{A,\alpha}$ for any rule satisfying the property of absorption. By contrast, the cautious rule studied by Denœux [11] (which in this situation corresponds to the conjunctive rule proposed by Kennes [29]) gives the same result as Dempster's rule of combination when $A \neq B$, and gives the result $m_{A,\alpha}$ only when $A = B$.*

4.3. Dempster's rule of conditioning and the minimum rule of possibility theory

Dempster's rule of conditioning is the special case of Dempster's rule of combination for two normal bba's m_1, m_2 on 2^S when m_2 assigns the total belief mass 1 to a nonempty subset of S (that is, m_2 is either the vacuous bba on 2^S or a simple bba $m_{A,1}$ on 2^S such that $A \subset S$ is not empty). In this case, $\mathcal{J}(m_1, m_2)$ is a singleton, because if m_1, m_2 are interpreted as the probability distributions of the random subsets X_1, X_2 of S, respectively, then X_1, X_2 are certainly independent (since X_2 is constant a.s.), and therefore there is exactly one jba I with marginals m_1, m_2. Hence, when the belief masses are interpreted probabilistically (as in [8, 42]), the result of Dempster's rule of conditioning can be justified as the best approximation of m_I by a normal bba on 2^S (see Subsection 3.2), without need of any assumptions about the dependence of the information sources. When the belief masses are interpreted non-probabilistically (as in [41, 45]), Dempster's rule of conditioning can be justified by considering specializations, which are a fundamental concept in Dempster-Shafer theory independently of the interpretation of the belief masses: see for example the expressions (2). As noted in Subsection 4.2, a specialization of a bba m can be interpreted as describing the same information as m plus some additional information; hence, it is natural to assume that the combination of $m_1, \ldots, m_n$ is a common specialization of $m_1, \ldots, m_n$, or an approximation thereof. In the case of conditioning, Theorem 1 implies $\mathcal{S}(m_1) \cap \mathcal{S}(m_2) = \mathcal{S}(m_I)$, and therefore m_I is the most cautious choice of a common specialization of m_1, m_2, in the sense of describing the least possible amount of information. In fact, all other common specializations of m_1, m_2 can be interpreted as describing more information than m_I, since they are specializations of m_I as well. Hence, the result of Dempster's rule of conditioning can be justified also as the best normal approximation of the least specialized common specialization of m_1, m_2 (see also [47, Theorem 3.2]).

A combination rule *generalizes Dempster's rule of conditioning* if it gives the same result as Dempster's rule of combination in the case of conditioning (that is, in the case of combining two normal bba's m_1, m_2 on 2^S such that m_2 assigns the total belief mass 1 to a nonempty subset of S). In particular, Dempster's rule of conditioning is generalized (at least up to normalization) by the combination rules proposed in [29, 20, 1, 13, 12] for the situation in which no assumption is made about the dependence of the information sources. By contrast, the cautious rule suggested by Denœux [11] does not generalize Dempster's rule of conditioning: the result of the combination of a non-separable bba m on 2^S with the vacuous bba on 2^S is not m (see [11, Propositions 8 and 9]), in contrast with the usual interpretation of the vacuous bba as the bba describing no information. Non-separable bba's pose difficulties also with the interpretation of Denœux's cautious rule (since the interpretation of negative weights of evidence is rather difficult, see also [21, Subsection 2.3]), and when restricted to separable bba's, Denœux's cautious rule reduces to the conjunctive rule already studied by Kennes [29] (who notes that it was suggested by Smets).

The plausibility functions associated with the consonant bba's on 2^S correspond to the possibility measures on 2^S, considered in possibility theory [54, 17]. The usual (and most cautious) conjunctive combination rule in possibility theory is the minimum rule, which corresponds to combining the consonant bba's $m_1, \ldots, m_n$ on 2^S into the consonant bba on 2^S associated with the contour function $\pi_1 \wedge \cdots \wedge \pi_n$ (or into its normalized version), where $\pi_1, \ldots, \pi_n$ are the contour functions associated with $m_1, \ldots, m_n$, respectively, and $\wedge$ is the pointwise minimum operator. It can be easily proved (by explicit construction, see for example [15, 19]) that there is a jba $J \in \mathcal{J}(m_1, \ldots, m_n)$ such that m_J is the result of the unnormalized version of the minimum rule of possibility theory (that is, m_J is the consonant bba on 2^S associated with the contour function $\pi_1 \wedge \cdots \wedge \pi_n$). Hence, Theorem 1 implies that m_J is a common specialization of $m_1, \ldots, m_n$, and Theorem 5 implies that Q_J is the pointwise maximum of the commonality functions associated with the common specializations of $m_1, \ldots, m_n$, since

$$Q_J(A) = \min_{x \in A} (\pi_1 \wedge \cdots \wedge \pi_n)(x) = \min_{i \in \{1,\ldots,n\}} Q_i(A) = (Q_1 \wedge \cdots \wedge Q_n)(A) \tag{5}$$

for all nonempty $A \subseteq S$. Therefore, the result of the unnormalized version of the minimum rule of possibility theory is a least specialized common specialization of $m_1, \ldots, m_n$ (in the sense that it is not a specialization of another common specialization of $m_1, \ldots, m_n$), but the following simple example shows that in general the least specialized common specialization of $m_1, \ldots, m_n$ is not unique (hence, Theorem 4 of [20] is wrong).

Example 4. *Let A, B be nonempty, disjoint subsets of S, and for each $\alpha \in [0, \frac{1}{2}]$ let m_α be the bba on 2^S defined by $m_\alpha(A) = m_\alpha(B) = \alpha$ and $m_\alpha(S) = m_\alpha(\varnothing) = \frac{1}{2} - \alpha$. It can be easily checked that the bba's m_α are the ones induced by the jba's with as marginals the simple bba's $m_{A,1/2}, m_{B,1/2}$ on 2^S, and that*

$$(Pl_{A,1/2} \wedge Pl_{B,1/2})(C) = \begin{cases} 0 & \text{if } C = \varnothing, \\ 1 & \text{if } C \cap A \neq \varnothing \text{ and } C \cap B \neq \varnothing, \\ \frac{1}{2} & \text{otherwise,} \end{cases} \quad \text{and} \quad (Q_{A,1/2} \wedge Q_{B,1/2})(C) = \begin{cases} 1 & \text{if } C = \varnothing, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Hence, $Q_{A,1/2} \wedge Q_{B,1/2} = Q_0$; moreover, $Pl_{A,1/2} \curlywedge Pl_{B,1/2} = Pl_{A,1/2} \wedge Pl_{B,1/2}$, since $Pl_{A,1/2}(C) \geq \frac{1}{2}$ and $Pl_{B,1/2}(C) \geq \frac{1}{2}$ for all nonempty $C \subset S$. Therefore, $Pl_{A,1/2} \curlywedge Pl_{B,1/2} = Pl_{1/2}$ when $A \cup B = S$, while $Pl_{A,1/2} \curlywedge Pl_{B,1/2}$ is not a plausibility function on 2^S when $A \cup B \subset S$, because for instance

$$(Pl_{A,1/2} \curlywedge Pl_{B,1/2})(\overline{A}) + (Pl_{A,1/2} \curlywedge Pl_{B,1/2})(\overline{B}) = \tfrac{1}{2} + \tfrac{1}{2} < 1 + \tfrac{1}{2} = (Pl_{A,1/2} \curlywedge Pl_{B,1/2})(\overline{A} \cup \overline{B}) + (Pl_{A,1/2} \curlywedge Pl_{B,1/2})(\overline{A} \cap \overline{B}),$$

whereas plausibility functions are 2-alternating.

Hence, the minimum rule of possibility theory combines $m_{A,1/2}, m_{B,1/2}$ into m_0 (or into its normalized version: the vacuous bba on 2^S), but Theorem 1 implies that each bba m_α is a least specialized common specialization of $m_{A,1/2}, m_{B,1/2}$, since m_α is not a specialization of $m_{\alpha'}$ when $\alpha' \neq \alpha$. That is, m_0 is not the unique least specialized common specialization of $m_{A,1/2}, m_{B,1/2}$, but it is the only one pointwise maximizing the associated commonality function. For instance, $m_{1/2}$ is another least specialized common specialization of $m_{A,1/2}, m_{B,1/2}$, and it is the only one pointwise maximizing the associated plausibility function when $A \cup B = S$, as follows from Theorem 4. By contrast, when $A \cup B \subset S$, Theorem 1 implies that each bba m_α is a common specialization of $m_{A,1/2}, m_{B,1/2}$ pointwise maximizing the associated plausibility function (in the sense that Pl_α is not pointwise dominated by any plausibility function associated with another common specialization of $m_{A,1/2}, m_{B,1/2}$), since Pl_α is not dominated by $Pl_{\alpha'}$ when $\alpha' \neq \alpha$.

As noted above, it is natural to assume that the combination of $m_1, \ldots, m_n$ is a common specialization of $m_1, \ldots, m_n$ (or an approximation thereof), and when the least specialized common specialization of $m_1, \ldots, m_n$ is not unique, the usual ways of choosing a most cautious common specialization of $m_1, \ldots, m_n$ are by pointwise maximizing the associated plausibility or commonality functions (see for example [14, 50, 16, 24, 46, 20]). When $m_1, \ldots, m_n$ are consonant, the result of the unnormalized version of the minimum rule of possibility theory is the unique common specialization of $m_1, \ldots, m_n$ pointwise maximizing the associated commonality function, while Example 4 shows that in general the common specialization of $m_1, \ldots, m_n$ pointwise maximizing the associated plausibility function is not unique. A combination rule *generalizes the minimum rule of possibility theory* if it gives the same result as the normalized version of that rule when $m_1, \ldots, m_n$ are consonant. In particular, the minimum rule of possibility theory is generalized (up to normalization) by the combination rules studied in [13, 12] (if we completely specify them in a suitable way). By contrast, the combination rules proposed in [29, 1, 11] do not generalize the minimum rule of possibility theory: when applied to the situation of Example 4, the rules studied by Kennes [29] and Denœux [11] give the same result (at least up to normalization) as Dempster's rule of combination, while the rule proposed by Cattaneo [1] gives the result $m_{1/2}$.

4.4. *Quasi-associativity*

A property that can have important advantages from the computational point of view is *quasi-associativity*: a combination rule is quasi-associative if there are an associative binary operator $\star$ on a set $\mathcal{F}$, a function f assigning to each normal bba on 2^S an element of $\mathcal{F}$, and a function g on $\mathcal{F}$ such that

$$\langle m_1, \ldots, m_n \rangle = g(f(m_1) \star \cdots \star f(m_n)).$$

This definition basically corresponds to the idea of Yager [52]; it implies in particular $(g \circ f)(m) = \langle m \rangle = m$ for all normal bba's m on 2^S (hence, f is an injection, and g is a quasi-inverse of f). The computational advantages of quasi-associativity are related to the possibility of performing the actual combinations in the set $\mathcal{F}$ by means of the associative binary operator $\star$: the application of the function g can be interpreted as an approximation step, in the sense that $f(\langle m_1, \ldots, m_n \rangle)$ can be interpreted as the best approximation of $f(m_1) \star \cdots \star f(m_n)$ by an element of the image of f. An n-ary combination rule obtained through $n-1$ applications of an associative binary rule is trivially quasi-associative (with $f = g$ the identity function and $\star$ the binary rule); another example of quasi-associative combination rule is the one proposed by Daniel [6], where the elements of $\mathcal{F}$ are generalized belief functions. Since the normalized version of the minimum rule of possibility theory is not associative, a combination rule generalizing the minimum rule of possibility theory cannot be associative, but it can be quasi-associative, as will be shown in the following section.

5. Two combination rules not requiring assumptions about the dependence of the information sources

Let $m_1, \ldots, m_n$ be $n \geq 1$ normal bba's on 2^S, describing the information obtained from different sources, and consider the situation in which no assumption is made about the dependence of the information sources. When the belief masses are interpreted probabilistically (as in [8, 42]), the combined information is described by a bba induced by some jba with marginals $m_1, \ldots, m_n$, while as noted in Subsection 4.3, when the belief masses are interpreted non-probabilistically (as in [41, 45]), it is natural to assume that the combined information is described by a common specialization of $m_1, \ldots, m_n$. Hence, we have a set $\{m_J : J \in \mathcal{J}(m_1, \ldots, m_n)\}$ or $\mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$ of bba's on 2^S possibly describing the combined information, and the usual way to proceed corresponds to the cautious choice of excluding the elements describing more information than is strictly necessary (see for example [16, 24, 46, 20]). The resulting subsets depend on the exact definition of information content of a bba, but the usual definitions are compatible with the concept of specialization, in the sense that a more specialized bba is also more informative (see for instance [14, 50, 16, 24, 46, 20]). Therefore, we can restrict attention to the two subsets of all least specialized elements of $\{m_J : J \in \mathcal{J}(m_1, \ldots, m_n)\}$ and $\mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$, respectively: Theorem 1 implies that these two subsets are equal; that is, we have a subset of least specialized bba's on 2^S possibly describing the combined information, independently of the interpretation of the belief masses.

If there is a unique least specialized common specialization m_s of $m_1, \ldots, m_n$, then $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ and $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ are the plausibility and commonality functions associated with m_s, respectively, and we can consider the normalized version m of m_s as a cautious combination of $m_1, \ldots, m_n$: it can be interpreted as the best approximation of m_s by a normal bba on 2^S (see Subsection 3.2), or equivalently Pl and Q can be interpreted as the best approximations of $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ and $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ by a plausibility and a commonality functions associated with some normal bba's on 2^S, respectively.

As noted in Subsection 4.3, when the least specialized common specialization of $m_1, \ldots, m_n$ is not unique, the usual ways of choosing a most cautious common specialization of $m_1, \ldots, m_n$ are by pointwise maximizing the associated plausibility or commonality functions. If there is a unique common specialization m_s of $m_1, \ldots, m_n$ pointwise maximizing the associated plausibility function, then $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ is the plausibility function associated with m_s, and we can consider the normalized version m of m_s as a cautious combination of $m_1, \ldots, m_n$: it can be interpreted as the best approximation of m_s by a normal bba on 2^S, or equivalently Pl can be interpreted as the best approximation of $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ by a plausibility function associated with some normal bba on 2^S. Analogously, if there is a unique common specialization m_s of $m_1, \ldots, m_n$ pointwise maximizing the associated commonality function, then $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ is the commonality function associated with m_s, and we can consider the normalized version m of m_s as a cautious combination of $m_1, \ldots, m_n$: it can be interpreted as the best approximation of m_s by a normal bba on 2^S, or equivalently Q can be interpreted as the best approximation of $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ by a commonality function associated with some normal bba on 2^S. However, Example 4 and Theorem 4 imply that in general $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ is not a plausibility function on 2^S, while the following example and Theorem 5 imply that in general $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ is not a commonality function on 2^S.

Example 5. *In the situation of Example 1,*

$$(Q_1 \wedge \cdots \wedge Q_n)(A) = \begin{cases} 1 & \text{if } A = \varnothing, \\ \frac{1}{2} & \text{if } |A| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $Q_1 \wedge \cdots \wedge Q_n$ is not a commonality function on 2^S, because for instance

$$\sum_{A \subseteq S} (-1)^{|A|} (Q_1 \wedge \cdots \wedge Q_n)(A) = 1 - \tfrac{n}{2} \leq -\tfrac{1}{2} < 0,$$

whereas all commonality functions Q on 2^S satisfy $\sum_{A \subseteq S} (-1)^{|A|} Q(A) \geq 0$.

Hence, pointwise maximizing the associated plausibility or commonality functions does not always lead to a unique common specialization of $m_1, \ldots, m_n$. To avoid this problem, we could consider a measure of information content that always allows to choose a unique least informative bba from a set of bba's on 2^S: if this measure is compatible with the concept of specialization (in the sense that a more specialized bba is also more informative), then Theorem 1 implies that the least informative bba in $\{m_J : J \in \mathcal{J}(m_1, \ldots, m_n)\}$ corresponds to the least informative bba in $\mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$, and it can be easily proved that the combination rule obtained by considering the normalized version of that least informative bba as the combination of $m_1, \ldots, m_n$ would satisfy the properties of commutativity, quasi-associativity, idempotency, absorption, and certainty preservation, and would generalize Dempster's rule of conditioning. This is the approach followed for instance by Destercke et al. [13], but their measure of information (based on expected cardinality) does not always allow to choose a unique least informative common specialization of $m_1, \ldots, m_n$, and thus their combination rule is not completely specified (in another paper [12], Destercke and Dubois consider sets of least informative bba's as possible results of combinations). In general, this approach has difficulties with the basic requirement of equivariance with respect to vacuous extensions (for example, the rule studied by Destercke et al. [13] does not satisfy that requirement), and the minimization of the measure of information can be computationally too demanding for many applications of the combination rule.

The idea of the present section is to consider as cautious combination of $m_1, \ldots, m_n$ a normal bba m on 2^S such that Pl can be interpreted as the best approximation of $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ by a plausibility function associated with some normal bba on 2^S, even when $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ is not a plausibility function on 2^S; or alternatively, to consider as cautious combination of $m_1, \ldots, m_n$ a normal bba m on 2^S such that Q can be interpreted as the best approximation of $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ by a commonality function associated with some normal bba on 2^S, even when $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ is not a commonality function on 2^S. Hence, we need an approximation method more general than normalization: ideally, we could choose Pl and Q by minimizing some suitable measures of distance from $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ and $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$, respectively, but the study of such distance measures and of algorithms for minimizing them goes beyond the scope of the present paper. In the following two subsections, we shall present results that hold for whole classes of approximation methods, and we shall only give two simple examples of approximation methods belonging to these classes.

5.1. The $\curlywedge$Pl-rule

We first consider the combination rule obtained by defining $\langle m_1, \ldots, m_n \rangle$ as the normal bba on 2^S associated with the plausibility function best approximating $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$. The exact calculation of $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Pl_J$ would be computationally too demanding for many applications of the rule, but the simple upper approximation $Pl_1 \curlywedge \cdots \curlywedge Pl_n$

suggested by Theorem 4 is exact when $n \leq 2$, and often exact when $n \geq 3$. Hence, we consider the combination rule obtained by defining $\langle m_1, \ldots, m_n \rangle$ as the normal bba on 2^S associated with the plausibility function best approximating $Pl_1 \curlywedge \cdots \curlywedge Pl_n$. Actually, the approximation by a plausibility function can be unnecessary in many applications, since the monotonic, subadditive function $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ on 2^S can be used directly to make inferences and decisions: for instance by simply comparing the values assigned by $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ to competing subsets of S, or more generally by comparing the nonadditive integrals (with respect to $Pl_1 \curlywedge \cdots \curlywedge Pl_n$) of the utility or loss functions on S associated with competing decisions (see for example [10, 2]).

Theorem 7. *Let ξ be a function assigning to each $\mu \in \mathcal{MS}_0(S)$ a normal bba on 2^S, and satisfying the following three conditions:*

- *if m is a bba on 2^S such that $m(\varnothing) < 1$, then $\xi(Pl)$ is the normalized version of m;*
- *$\xi(\mu \circ \tilde{r}) = (\xi(\mu))^{\uparrow(R,r)}$ for all $\mu \in \mathcal{MS}_0(S)$ and all refinements (R, r) of S;*
- *$C(\xi(\mu)) \subseteq \{x \in S : \mu(\{x\}) > 0\}$ when the right-hand side is not empty, for all $\mu \in \mathcal{MS}_0(S)$.*

Then the combination rule defined by

$$\langle m_1, \ldots, m_n \rangle = \xi(Pl_1 \curlywedge \cdots \curlywedge Pl_n)$$

satisfies the properties of commutativity, quasi-associativity, idempotency, absorption, equivariance with respect to vacuous extensions, and certainty preservation, and generalizes Dempster's rule of conditioning.

We call $\curlywedge$*Pl-rule* any combination rule of the form considered in Theorem 7, independently of the particular approximation method ξ. A simple example of approximation method satisfying the conditions of Theorem 7 is the following: for each $A \subseteq S$, in order of increasing cardinality, define

$$m(A) = \max\left\{0,\ 1 - \mu(\overline{A}) - \textstyle\sum_{B \subset A} m(B)\right\},$$

and if $\sum_{B \subseteq S\,:\,B \neq \varnothing} m(B) > 0$, then define $(\xi(\mu))\,(\varnothing) = 0$ and

$$(\xi(\mu))\,(A) = \frac{m(A)}{\sum_{B \subseteq S\,:\,B \neq \varnothing} m(B)} \tag{6}$$

for all nonempty $A \subseteq S$, while if $\sum_{B \subseteq S\,:\,B \neq \varnothing} m(B) = 0$, then define $\xi(\mu)$ as the vacuous bba on 2^S.

When μ is a plausibility function on 2^S, the bba m associated with μ can be obtained by calculating $m(A) = 1 - \mu(\overline{A}) - \sum_{B \subset A} m(B)$ for each $A \subseteq S$, in order of increasing cardinality. Hence, the approximation method (6) satisfies the first condition of Theorem 7, and can be interpreted as simply enforcing the nonnegativity and normality of m also when μ is not a plausibility function. It can be easily proved that the approximation method (6) satisfies also the second and third conditions of Theorem 7, and that $\sum_{B \subseteq S\,:\,B \neq \varnothing} m(B) = 0$ in the algorithm (6) if and only if $c_{\min}(m_1, \ldots, m_n) = 1$.

Example 6. *In the situation of Example 1, $Pl_1 \curlywedge \cdots \curlywedge Pl_n = Pl_1 \wedge \cdots \wedge Pl_n$, since $Pl_i(A) \geq \frac{1}{2}$ for all nonempty $A \subset S$ and all $i \in S$. Hence, $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ is not 2-alternating, because for instance (see Example 2)*

$$(Pl_1 \curlywedge \cdots \curlywedge Pl_n)(\{1,2\}) + (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(\{2,\ldots,n\}) = \tfrac{1}{2} + \tfrac{1}{2} < 1 + \tfrac{1}{2} = (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(S) + (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(\{2\}).$$

Therefore, $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ is not a plausibility function on 2^S, and the result of the $\curlywedge$Pl-rule depends on the approximation method. In the algorithm (6) we obtain $m(\varnothing) = 0$, $m(\{i\}) = \frac{1}{2}$ for all $i \in S$, and $m(A) = 0$ for all $A \subseteq S$ such that $|A| \geq 2$. Hence, the result of the $\curlywedge$Pl-rule with the approximation method (6) is the bba m' on 2^S such that $m'(\{i\}) = \frac{1}{n}$ for all $i \in S$.

In the situation of Example 3, the result of the $\curlywedge$Pl-rule is $m_{A,\alpha}$, independently of the approximation method, since the $\curlywedge$Pl-rule satisfies the property of absorption.

In the situation of Example 4, if $A \cup B = S$, then $Pl_{A,1/2} \curlywedge Pl_{B,1/2} = Pl_{1/2}$, and the result of the $\curlywedge$Pl-rule is $m_{1/2}$, independently of the approximation method; while if $A \cup B \subset S$, then $Pl_{A,1/2} \curlywedge Pl_{B,1/2}$ is not a plausibility function on 2^S, and the result of the $\curlywedge$Pl-rule depends on the approximation method: with the method (6) the result is $m_{1/2}$ in this case too, since in the algorithm (6) we obtain $m = m_{1/2}$.

5.2. The $\wedge Q$-rule

We now consider the combination rule obtained by defining $\langle m_1, \ldots, m_n \rangle$ as the normal bba on 2^S associated with the commonality function best approximating $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$. An advantage over the $\curlywedge Pl$-rule is that $\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J$ can be easily calculated as $Q_1 \wedge \cdots \wedge Q_n$ thanks to Theorem 5, while a drawback is that the interpretation of commonality functions is less straightforward than the interpretation of plausibility functions, and in particular the approximation of $Q_1 \wedge \cdots \wedge Q_n$ by a commonality function seems to be necessary in most applications (since usually the anti-monotonic, quasi-superadditive function $Q_1 \wedge \cdots \wedge Q_n$ on 2^S cannot be used directly to make inferences and decisions). The expression (1) implies that this rule is strictly connected with Dempster's rule of combination, whose result can be interpreted as the normal bba on 2^S associated with the commonality function best approximating $Q_1 \cdots Q_n$; and the expression (5) shows that the present rule can also be interpreted as a straightforward generalization of the minimum rule of possibility theory. Moreover, there is an interesting similarity with the normalized cautious rule studied by Denœux [11], where weight functions are used instead of commonality functions: an advantage of Denœux's cautious rule is that no approximation method more general than normalization is necessary, while drawbacks are the problems with non-separable bba's (see Subsection 4.3) and the strange behavior outlined in Example 3.

Theorem 8. *Let χ be a function assigning to each $v \in \mathcal{AQ}_1(S)$ a normal bba on 2^S, and satisfying the following three conditions:*

- *if m is a bba on 2^S such that $m(\varnothing) < 1$, then $\chi(Q)$ is the normalized version of m;*
- *$\chi(v \circ \tilde{r}) = (\chi(v))^{\uparrow(R,r)}$ for all $v \in \mathcal{AQ}_1(S)$ and all refinements (R, r) of S;*
- *$C(\chi(v)) \subseteq \{x \in S : v(\{x\}) > 0\}$ when the right-hand side is not empty, for all $v \in \mathcal{AQ}_1(S)$.*

Then the combination rule defined by

$$\langle m_1, \ldots, m_n \rangle = \chi(Q_1 \wedge \cdots \wedge Q_n)$$

satisfies the properties of commutativity, quasi-associativity, idempotency, absorption, equivariance with respect to vacuous extensions, and certainty preservation, and generalizes Dempster's rule of conditioning and the minimum rule of possibility theory.

We call $\wedge Q$*-rule* any combination rule of the form considered in Theorem 8, independently of the particular approximation method χ. A simple example of approximation method satisfying the conditions of Theorem 8 is the following: for each $A \subseteq S$, in order of decreasing cardinality, define

$$m(A) = \max\left\{0,\ v(A) - \textstyle\sum_{B \subseteq S : A \subset B} m(B)\right\},$$

and if $\sum_{B \subseteq S : B \neq \varnothing} m(B) > 0$, then define $(\chi(v))(\varnothing) = 0$ and

$$(\chi(v))(A) = \frac{m(A)}{\sum_{B \subseteq S : B \neq \varnothing} m(B)} \tag{7}$$

for all nonempty $A \subseteq S$, while if $\sum_{B \subseteq S : B \neq \varnothing} m(B) = 0$, then define $\chi(v)$ as the vacuous bba on 2^S.

When v is a commonality function on 2^S, the bba m associated with v can be obtained by calculating $m(A) = v(A) - \sum_{B \subseteq S : A \subset B} m(B)$ for each $A \subseteq S$, in order of decreasing cardinality. Hence, the approximation method (7) satisfies the first condition of Theorem 8, and can be interpreted as simply enforcing the nonnegativity and normality of m also when v is not a commonality function. It can be easily proved that the approximation method (7) satisfies also the second and third conditions of Theorem 8, and that $\sum_{B \subseteq S : B \neq \varnothing} m(B) = 0$ in the algorithm (7) if and only if $c_{\min}(m_1, \ldots, m_n) = 1$.

Example 7. *In the situation of Example 1, $Q_1 \wedge \cdots \wedge Q_n$ is not a commonality function on 2^S (see Example 5), and the result of the $\wedge Q$-rule depends on the approximation method. In the algorithm (7) we obtain $m(A) = 0$ for all $A \subseteq S$*

such that $|A| \geq 2$, $m(\{i\}) = \frac{1}{2}$ for all $i \in S$, and $m(\varnothing) = 0$. Hence, the result of the $\wedge Q$-rule with the approximation method (7) is the bba m' on 2^S such that $m'(\{i\}) = \frac{1}{n}$ for all $i \in S$.

In the situation of Example 3, the result of the $\wedge Q$-rule is $m_{A,\alpha}$, independently of the approximation method, since the $\wedge Q$-rule satisfies the property of absorption.

In the situation of Example 4, the result of the $\wedge Q$-rule is the vacuous bba on 2^S, independently of the approximation method, since the $\wedge Q$-rule generalizes the minimum rule of possibility theory.

6. Conclusion

In the present paper the problem of combining belief functions obtained from not necessarily independent sources of information has been studied. The minimal conflict of belief functions has been defined: it is a much better measure of disagreement among belief functions than the conflict of Dempster's rule of combination, which assumes the independence of the information sources.

Two combination rules for the situation in which no assumption is made about the dependence of the information sources have been proposed: the $\curlywedge Pl$-rule and the $\wedge Q$-rule. They consist in approximating by a plausibility and a commonality functions the combinations $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ and $Q_1 \wedge \cdots \wedge Q_n$ of the plausibility and commonality functions obtained from the information sources, respectively, where $\curlywedge$ and $\wedge$ are two simple commutative and associative binary operators. The proposed combination rules satisfy the properties of commutativity, quasi-associativity, idempotency, absorption, equivariance with respect to vacuous extensions, and certainty preservation, and generalize Dempster's rule of conditioning. Moreover, the $\wedge Q$-rule generalizes the minimum rule of possibility theory, and is closely connected with both Dempster's rule of combination and the cautious rule studied by Denœux [11], but the $\curlywedge Pl$-rule has the advantage that the interpretation of plausibility functions is more straightforward than the interpretation of commonality functions.

As regards the approximations of $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ and $Q_1 \wedge \cdots \wedge Q_n$ by a plausibility and a commonality functions, respectively, in the present paper only two simple algorithms have been proposed. Better algorithms would improve the $\curlywedge Pl$-rule and the $\wedge Q$-rule: in particular, it would be interesting to have upper approximations of $Pl_1 \curlywedge \cdots \curlywedge Pl_n$ and $Q_1 \wedge \cdots \wedge Q_n$ by a plausibility and a commonality functions, respectively, such that the conditions of Theorems 7 and 8 are satisfied.

A. Proofs of theorems

Proof of Theorem 1. If $m \in \mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$, then there are $J_i \in \mathcal{J}(m, m_i)$ such that $J_i(B, B_i) > 0$ implies $B \subseteq B_i$, for all $i \in \{1, \ldots, n\}$ and all $B, B_i \subseteq S$. Let J be the function on $(2^S)^n$ defined by

$$J(B_1, \ldots, B_n) = \sum_{\substack{B \subseteq S:\\ m(B)>0}} \frac{J_1(B, B_1) \cdots J_n(B, B_n)}{m(B)^{n-1}}$$

for all $B_1, \ldots, B_n \subseteq S$. Then $J \in \mathcal{J}(m_1, \ldots, m_n)$, since $J : (2^S)^n \to [0, \infty)$ and, for all $A \subseteq S$ and all $i \in \{1, \ldots, n\}$,

$$\sum_{\substack{B_1,\ldots,B_n \subseteq S:\\ B_i = A}} J(B_1, \ldots, B_n) = \sum_{\substack{B \subseteq S:\\ m(B)>0}} \sum_{\substack{B_1,\ldots,B_n \subseteq S:\\ B_i = A}} \frac{J_1(B, B_1) \cdots J_n(B, B_n)}{m(B)^{n-1}} = \sum_{\substack{B \subseteq S:\\ m(B)>0}} \frac{J_i(B, A)\, m(B)^{n-1}}{m(B)^{n-1}} = m_i(A).$$

Let J' be the function on $(2^S)^2$ defined by

$$J'(B, A) = \begin{cases} 0 & \text{if } m(B) = 0, \\ \sum_{B_1,\ldots,B_n \subseteq S:\, B_1 \cap \cdots \cap B_n = A} \frac{J_1(B,B_1)\cdots J_n(B,B_n)}{m(B)^{n-1}} & \text{if } m(B) > 0, \end{cases}$$

for all $B, A \subseteq S$. If $m(B) = 0$, then $\sum_{A \subseteq S} J'(B, A) = 0 = m(B)$, and if $m(B) > 0$ then

$$\sum_{A \subseteq S} J'(B, A) = \sum_{B_1,\ldots,B_n \subseteq S} \frac{J_1(B, B_1) \cdots J_n(B, B_n)}{m(B)^{n-1}} = \frac{m(B)^n}{m(B)^{n-1}} = m(B),$$

for all $B \subseteq S$. Hence, $J' \in \mathcal{J}(m, m_J)$, because $J' : (2^S)^2 \to [0, \infty)$ and, for all $A \subseteq S$,

$$\sum_{B \subseteq S} J'(B, A) = \sum_{\substack{B \subseteq S:\\ m(B)>0}} \sum_{\substack{B_1,\dots,B_n \subseteq S:\\ B_1 \cap \cdots \cap B_n = A}} \frac{J_1(B, B_1) \cdots J_n(B, B_n)}{m(B)^{n-1}} = \sum_{\substack{B_1,\dots,B_n \subseteq S:\\ B_1 \cap \cdots \cap B_n = A}} J(B_1, \dots, B_n) = m_J(A).$$

To prove that $m \in \mathcal{S}(m_J)$, it suffices to show that $J'(B, A) > 0$ implies $B \subseteq A$, for all $B, A \subseteq S$. If $J'(B, A) > 0$, then there are $B_1, \dots, B_n \subseteq S$ such that $B_1 \cap \cdots \cap B_n = A$ and $J_i(B, B_i) > 0$ for all $i \in \{1, \dots, n\}$. Therefore, $B \subseteq B_i$ for all $i \in \{1, \dots, n\}$, and so $B \subseteq B_1 \cap \cdots \cap B_n = A$.

We have proved that $\mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n) \subseteq \bigcup_{J \in \mathcal{J}(m_1,\dots,m_n)} \mathcal{S}(m_J)$, and we now turn to the proof of the converse inclusion. If $J \in \mathcal{J}(m_1, \dots, m_n)$ and $m \in \mathcal{S}(m_J)$, then there is $J' \in \mathcal{J}(m, m_J)$ such that $J'(B, B_J) > 0$ implies $B \subseteq B_J$, for all $B, B_J \subseteq S$. Let $J_1, \dots, J_n$ be the functions on $(2^S)^2$ defined by

$$J_i(B, A) = \begin{cases} 0 & \text{if } m(B) = 0,\\ \sum_{B_1,\dots,B_n \subseteq S:\, B_i = A,\, m_J(B_1 \cap \cdots \cap B_n) > 0} \frac{J'(B, B_1 \cap \cdots \cap B_n)\, J(B_1,\dots,B_n)}{m_J(B_1 \cap \cdots \cap B_n)} & \text{if } m(B) > 0, \end{cases}$$

for all $B, A \subseteq S$ and all $i \in \{1, \dots, n\}$. If $m(B) = 0$, then $\sum_{A \subseteq S} J_i(B, A) = 0 = m(B)$, and if $m(B) > 0$ then

$$\sum_{A \subseteq S} J_i(B, A) = \sum_{\substack{C \subseteq S:\\ m_J(C)>0}} \sum_{\substack{B_1,\dots,B_n \subseteq S:\\ B_1 \cap \cdots \cap B_n = C}} \frac{J'(B, C)\, J(B_1, \dots, B_n)}{m_J(C)} = \sum_{\substack{C \subseteq S:\\ m_J(C)>0}} J'(B, C) = m(B),$$

for all $B \subseteq S$ and all $i \in \{1, \dots, n\}$. Hence, $J_i \in \mathcal{J}(m, m_i)$ for all $i \in \{1, \dots, n\}$, because $J_i : (2^S)^2 \to [0, \infty)$ and, for all $A \subseteq S$ and all $i \in \{1, \dots, n\}$,

$$\sum_{B \subseteq S} J_i(B, A) = \sum_{\substack{B_1,\dots,B_n \subseteq S:\\ B_i = A,\, m_J(B_1 \cap \cdots \cap B_n) > 0}} \sum_{B \subseteq S} \frac{J'(B, B_1 \cap \cdots \cap B_n)\, J(B_1, \dots, B_n)}{m_J(B_1 \cap \cdots \cap B_n)} = \sum_{\substack{B_1,\dots,B_n \subseteq S:\\ B_i = A,\, m_J(B_1 \cap \cdots \cap B_n) > 0}} J(B_1, \dots, B_n) = m_i(A).$$

To prove that $m \in \mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$, it suffices to show that $J_i(B, A) > 0$ implies $B \subseteq A$, for all $B, A \subseteq S$ and all $i \in \{1, \dots, n\}$. If $J_i(B, A) > 0$, then there are $B_1, \dots, B_n \subseteq S$ such that $B_i = A$ and $J'(B, B_1 \cap \cdots \cap B_n) > 0$, and thus $B \subseteq B_1 \cap \cdots \cap B_n \subseteq A$. □

Proof of Theorem 2. To prove that $\curlywedge$ is a commutative and associative binary operator on $\mathcal{MS}_0(S)$, it suffices to show that $\mu, \nu \in \mathcal{MS}_0(S)$ implies $\mu \curlywedge \nu \in \mathcal{MS}_0(S)$. Since $\mu(\varnothing) = \nu(\varnothing) = 0$, we have $(\mu \curlywedge \nu)(\varnothing) = 0$ and $0 \leq \mu \curlywedge \nu \leq \mu \wedge \nu \leq 1$. If $A \subseteq B \subseteq S$, then

$$(\mu \curlywedge \nu)(B) = \min_{C \subseteq B} (\mu(C) + \nu(B \setminus C)) \geq \min_{C \subseteq B} (\mu(A \cap C) + \nu(A \setminus C)) = (\mu \curlywedge \nu)(A),$$

because μ and ν are monotonic; hence, $\mu \curlywedge \nu$ is monotonic as well. If $A, B \subseteq S$ are disjoint, then

$$\begin{aligned}(\mu \curlywedge \nu)(A) + (\mu \curlywedge \nu)(B) &= \min_{C \subseteq A,\, D \subseteq B} (\mu(C) + \nu(A \setminus C) + \mu(D) + \nu(B \setminus D)) \geq \\ &\geq \min_{C \subseteq A,\, D \subseteq B} (\mu(C \cup D) + \nu((A \cup B) \setminus (C \cup D))) = (\mu \curlywedge \nu)(A \cup B),\end{aligned}$$

because μ and ν are subadditive; hence, $\mu \curlywedge \nu$ is subadditive as well. That is, $\mu, \nu \in \mathcal{MS}_0(S)$ implies $\mu \curlywedge \nu \in \mathcal{MS}_0(S)$.

Let (R, r) be a refinement of S, and for each $A \subseteq R$ let f_A be the function on 2^A defined by $f_A(\varnothing) = \varnothing$ and $f_A(B) = \bigcup_{x \in \tilde{r}(B)} (r(x) \cap A)$ for all nonempty $B \subseteq A$. Hence, $\tilde{r}(f_A(B)) = \tilde{r}(B)$ and $\tilde{r}(A \setminus f_A(B)) = \tilde{r}(A) \setminus \tilde{r}(f_A(B)) \subseteq \tilde{r}(A \setminus B)$, for all $B \subseteq A \subseteq R$. Therefore, the image of $\tilde{r} \circ f_A$ is $2^{\tilde{r}(A)}$, and since ν is monotonic,

$$((\mu \circ \tilde{r}) \curlywedge (\nu \circ \tilde{r}))(A) = \min_{B \subseteq A} (\mu(\tilde{r}(B)) + \nu(\tilde{r}(A \setminus B))) = \min_{B \subseteq A} (\mu(\tilde{r}(f_A(B))) + \nu(\tilde{r}(A) \setminus \tilde{r}(f_A(B)))) = (\mu \curlywedge \nu)(\tilde{r}(A)),$$

for all $A \subseteq R$. That is, $(\mu \circ \tilde{r}) \curlywedge (\nu \circ \tilde{r}) = (\mu \curlywedge \nu) \circ \tilde{r}$.

If $\mu \leq \nu$ and $A \subseteq S$, then

$$(\mu \curlywedge \nu)(A) \leq (\mu \wedge \nu)(A) = \mu(A) = \min_{B \subseteq A} (\mu(B) + \mu(A \setminus B)) \leq \min_{B \subseteq A} (\mu(B) + \nu(A \setminus B)) = (\mu \curlywedge \nu)(A),$$

because μ is subadditive and $\mu(\varnothing) = \nu(\varnothing) = 0$. That is, $\mu \leq \nu$ implies $\mu \curlywedge \nu = \mu$. □

Proof of Theorem 3. To prove that $\wedge$ is a commutative and associative binary operator on $\mathcal{AQ}_1(S)$, it suffices to show that $\mu, \nu \in \mathcal{AQ}_1(S)$ implies $\mu \wedge \nu \in \mathcal{AQ}_1(S)$. Clearly, $0 \leq \mu \wedge \nu \leq 1$, and since $\mu(\varnothing) = \nu(\varnothing) = 1$, we have $(\mu \wedge \nu)(\varnothing) = 1$. If $A \subseteq B \subseteq S$, then $(\mu \wedge \nu)(A) = \min\{\mu(A), \nu(A)\} \geq \min\{\mu(B), \nu(B)\} = (\mu \wedge \nu)(B)$, because μ and ν are anti-monotonic; hence, $\mu \wedge \nu$ is anti-monotonic as well. If $A, B \subseteq S$ are disjoint, then

$$(\mu \wedge \nu)(A \cup B) = \min\{\mu(A \cup B), \nu(A \cup B)\} \geq \min\{\mu(A) + \mu(B) - 1, \nu(A) + \nu(B) - 1\} \geq$$
$$\geq \min\{\mu(A), \nu(A)\} + \min\{\mu(B), \nu(B)\} - 1 = (\mu \wedge \nu)(A) + (\mu \wedge \nu)(B) - 1,$$

because μ and ν are quasi-superadditive; hence, $\mu \wedge \nu$ is quasi-superadditive as well. That is, $\mu, \nu \in \mathcal{AQ}_1(S)$ implies $\mu \wedge \nu \in \mathcal{AQ}_1(S)$.

If (R, r) is a refinement of S, and $A \subseteq R$, then $((\mu \circ \tilde{r}) \wedge (\nu \circ \tilde{r}))(A) = \min\{\mu(\tilde{r}(A)), \nu(\tilde{r}(A))\} = (\mu \wedge \nu)(\tilde{r}(A))$; hence, $(\mu \circ \tilde{r}) \wedge (\nu \circ \tilde{r}) = (\mu \wedge \nu) \circ \tilde{r}$. Moreover, if $\mu \leq \nu$, then $\mu \wedge \nu = \mu$, because $\wedge$ is the pointwise minimum operator. □

Proof of Theorem 4. The use of max instead of sup in the first of the two Fréchet bounds (3) is correct, because a J' maximizing $J \mapsto Pl_J(A)$ over $\mathcal{J}(m_1, \ldots, m_n)$ certainly exists, since $Pl_J(A) = \sum_{B_1,\ldots,B_n \subseteq S : A \cap B_1 \cap \cdots \cap B_n \neq \varnothing} J(B_1, \ldots, B_n)$, and thus the problem of maximizing $J \mapsto Pl_J(A)$ over $\mathcal{J}(m_1, \ldots, m_n)$ can be interpreted as a problem of linear optimization over a nonempty, convex polytope in the $2^{n|S|}$-dimensional Euclidean space. Hence, $m_{J'}$ maximizes $m \mapsto Pl(A)$ over $\bigcup_{J \in \mathcal{J}(m_1,\ldots,m_n)} \mathcal{S}(m_J)$, because $Pl(A) \leq Pl_J(A)$ when $m \in \mathcal{S}(m_J)$. The equality of the two maxima in the first part of the theorem thus follows from Theorem 1.

The inequality in the first part of the theorem can be easily proved, since for all $m_s \in \mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$ and all $A \subseteq S$,

$$Pl_s(A) \leq \min_{(B_1,\ldots,B_n) \in \mathcal{OP}_n(A)} (Pl_s(B_1) + \cdots + Pl_s(B_n)) \leq \min_{(B_1,\ldots,B_n) \in \mathcal{OP}_n(A)} (Pl_1(B_1) + \cdots + Pl_n(B_n)) = (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(A),$$

where the first inequality follows by induction from the subadditivity of Pl_s, while the second one is a consequence of m_s being a specialization of m_i, for all $i \in \{1, \ldots, n\}$.

We now turn to the proof of the second part of the theorem. The case with $n = 1$ is obvious. In the case with $n = 2$, let A be a subset of S, define $\mathcal{A} = \{(A_1, A_2) \in (2^S)^2 : A_1 \cap A_2 \cap A \neq \varnothing\}$, and for each $C \subseteq S$ define $\mathcal{B}_C = \{B \subseteq S : B \cap A \cap C \neq \varnothing\}$. Theorem 4 of [39] implies

$$\max_{J \in \mathcal{J}(m_1,m_2)} Pl_J(A) = \min_{\substack{\mathcal{A}_1, \mathcal{A}_2 \subseteq 2^S : \\ \mathcal{A} \subseteq (\mathcal{A}_1 \times 2^S) \cup (2^S \times \mathcal{A}_2)}} \left(\sum_{A_1 \in \mathcal{A}_1} m_1(A_1) + \sum_{A_2 \in \mathcal{A}_2} m_2(A_2) \right). \tag{8}$$

Let $\mathcal{A}_1, \mathcal{A}_2$ be subsets of 2^S such that $\mathcal{A} \subseteq (\mathcal{A}_1 \times 2^S) \cup (2^S \times \mathcal{A}_2)$, and define $C = \cup_{A_2 \in 2^S \setminus \mathcal{A}_2} A_2$. Then $\mathcal{B}_C \subseteq \mathcal{A}_1$, because $B \in \mathcal{B}_C$ implies that there is an $A_2 \in 2^S \setminus \mathcal{A}_2$ such that $B \cap A \cap A_2 \neq \varnothing$, and thus $B \in \mathcal{A}_1$ (since $(B, A_2) \in \mathcal{A}$ and $A_2 \notin \mathcal{A}_2$). If $A_2 \in 2^S \setminus \mathcal{A}_2$, then $A_2 \subseteq C$ and so $A_2 \notin \mathcal{B}_{\overline{C}}$; that is, $\mathcal{B}_{\overline{C}} \subseteq \mathcal{A}_2$. Moreover, $\mathcal{A} \subseteq (\mathcal{B}_C \times 2^S) \cup (2^S \times \mathcal{B}_{\overline{C}})$, because $A_1 \cap A_2 \cap A \neq \varnothing$ implies $A_1 \cap A_2 \cap A \cap C \neq \varnothing$ or $A_1 \cap A_2 \cap A \cap \overline{C} \neq \varnothing$. Hence, in the right-hand side of the expression (8) it suffices to minimize over pairs $(\mathcal{A}_1, \mathcal{A}_2)$ of the form $(\mathcal{B}_C, \mathcal{B}_{\overline{C}})$; that is,

$$\max_{J \in \mathcal{J}(m_1,m_2)} Pl_J(A) \geq \min_{C \subseteq S} \left(\sum_{A_1 \in \mathcal{B}_C} m_1(A_1) + \sum_{A_2 \in \mathcal{B}_{\overline{C}}} m_2(A_2) \right) = \min_{C \subseteq S} \left(Pl_1(A \cap C) + Pl_2(A \cap \overline{C}) \right) = (Pl_1 \curlywedge Pl_2)(A).$$

This result and the first part of the theorem imply the second part. □

Proof of Theorem 5. Theorem 6 of [38] implies that the use of max instead of sup in the second of the two Fréchet bounds (3) is correct, and

$$\max_{J \in \mathcal{J}(m_1,\ldots,m_n)} Q_J(A) = \min\{Q_1(A), \ldots, Q_n(A)\} = (Q_1 \wedge \cdots \wedge Q_n)(A).$$

Hence, there is a J' maximizing $J \mapsto Q_J(A)$ over $\mathcal{J}(m_1, \ldots, m_n)$, and therefore $m_{J'}$ maximizes $m \mapsto Q(A)$ over $\bigcup_{J \in \mathcal{J}(m_1,\ldots,m_n)} \mathcal{S}(m_J)$, because $Q(A) \leq Q_J(A)$ when $m \in \mathcal{S}(m_J)$. The equality of the two maxima in the statement of the theorem thus follows from Theorem 1. □

Proof of Theorem 6. $c_{\min}(m_1, \ldots, m_n) = 0$ if and only if there is an $m_s \in \mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$ with $m_s(\emptyset) = 0$; that is, if and only if $m_1, \ldots, m_n$ have a common normal specialization. If $C(m_1) \cap \cdots \cap C(m_n) = \emptyset$, then $c_{\min}(m_1, \ldots, m_n) = 1$, because $m_s \in \mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n)$ implies $C(m_s) \subseteq C(m_1) \cap \cdots \cap C(m_n)$. On the other hand, if $C(m_1) \cap \cdots \cap C(m_n) \neq \emptyset$, then there are $A_1, \ldots, A_n \subseteq S$ such that $A_1 \cap \cdots \cap A_n \neq \emptyset$ and $m_i(A_i) > 0$ for all $i \in \{1, \ldots, n\}$; hence, $c_{\min}(m_1, \ldots, m_n) < 1$, because the bba m_s on 2^S defined by $m_s(A_1 \cap \cdots \cap A_n) = 1 - m_s(\emptyset) = \min\{m_1(A_1), \ldots, m_n(A_n)\}$ is a common specialization of $m_1, \ldots, m_n$.

Let (R, r) be a refinement of S, and let $f : \mathcal{J}(m_1, \ldots, m_n) \to \mathcal{J}(m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)})$ be defined by

$$(f(J))(B_1, \ldots, B_n) = \begin{cases} J(\tilde{r}(B_1), \ldots, \tilde{r}(B_n)) & \text{if } B_1, \ldots, B_n \text{ are focal sets of } m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)}, \text{ respectively,} \\ 0 & \text{otherwise.} \end{cases}$$

Then f is a bijection, because $\tilde{r}$ describes the one-to-one correspondences between the focal sets of $m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)}$ and the focal sets of $m_1, \ldots, m_n$, respectively, and a jba can take positive values only on n-tuples of focal sets of the respective marginals. Moreover, $m_{f(J)}(\emptyset) = m_J(\emptyset)$ for all $J \in \mathcal{J}(m_1, \ldots, m_n)$, because if $B_1, \ldots, B_n$ are focal sets of $m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)}$, respectively, then $B_1 \cap \cdots \cap B_n = \emptyset$ if and only if $\tilde{r}(B_1) \cap \cdots \cap \tilde{r}(B_n) = \emptyset$. Hence,

$$c_{\min}(m_1, \ldots, m_n) = \min_{J \in \mathcal{J}(m_1, \ldots, m_n)} m_{f(J)}(\emptyset) = c_{\min}(m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)}).$$

If $n \geq 2$ and m_1 is a specialization of m_n, then $\mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_n) = \mathcal{S}(m_1) \cap \cdots \cap \mathcal{S}(m_{n-1})$, and therefore $c_{\min}(m_1, \ldots, m_n) = c_{\min}(m_1, \ldots, m_{n-1})$. By induction it follows that if m_1 is a common specialization of $m_2, \ldots, m_n$, then $c_{\min}(m_1, \ldots, m_n) = c_{\min}(m_1) = m_1(\emptyset)$. □

Proof of Theorem 7. The first condition for ξ implies in particular $\langle m_1 \rangle = \xi(Pl_1) = m_1$, and thus $\langle m_1, \ldots, m_n \rangle$ is a combination rule. The commutativity and quasi-associativity of the rule follow from the commutativity and associativity of the binary operator $\curlywedge$ on $\mathcal{MS}_0(S)$, proved in Theorem 2. The combination rule satisfies the property of absorption (and thus idempotency), since Theorem 2 implies $Pl_1 \curlywedge Pl_n = Pl_1$ when m_1 is a specialization of m_n, and therefore

$$\langle m_1, \ldots, m_n \rangle = \xi((Pl_1 \curlywedge Pl_n) \curlywedge Pl_2 \curlywedge \cdots \curlywedge Pl_{n-1}) = \xi(Pl_1 \curlywedge Pl_2 \curlywedge \cdots \curlywedge Pl_{n-1}) = \langle m_1, \ldots, m_{n-1} \rangle$$

for all $n \geq 2$. The rule is also equivariant with respect to vacuous extensions, because

$$\langle m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)} \rangle = \xi((Pl_1 \circ \tilde{r}) \curlywedge \cdots \curlywedge (Pl_n \circ \tilde{r})) = \xi((Pl_1 \curlywedge \cdots \curlywedge Pl_n) \circ \tilde{r}) = \langle m_1, \ldots, m_n \rangle^{\uparrow(R,r)}$$

for all refinements (R, r) of S, where the second equality follows by induction from Theorem 2, and the third one is implied by the second condition for ξ. Since

$$\{x \in S : (Pl_1 \curlywedge \cdots \curlywedge Pl_n)(\{x\}) > 0\} = \{x \in S : Pl_1(\{x\}) > 0\} \cap \cdots \cap \{x \in S : Pl_n(\{x\}) > 0\} = C(m_1) \cap \cdots \cap C(m_n),$$

the third condition for ξ implies that the combination rule preserves certainty. Finally, the rule generalizes Dempster's rule of conditioning, because in the case of conditioning, $Pl_1 \curlywedge Pl_2 = Pl_I$ follows from Theorem 4 (where I is the unique jba with marginals m_1, m_2), and thus the first condition for ξ implies that $\langle m_1, m_2 \rangle = \xi(Pl_1)$ is the normalized version of m_I (when $m_I(\emptyset) < 1$). □

Proof of Theorem 8. The first condition for χ implies in particular $\langle m_1 \rangle = \chi(Q_1) = m_1$, and thus $\langle m_1, \ldots, m_n \rangle$ is a combination rule. The commutativity and quasi-associativity of the rule follow from the commutativity and associativity of the binary operator $\wedge$ on $\mathcal{AQ}_1(S)$, proved in Theorem 3. Since $Q_1 \wedge \cdots \wedge Q_n = Q_1 \wedge \cdots \wedge Q_{n-1}$ when $Q_1 \leq Q_n$ and $n \geq 2$, the combination rule satisfies the property of absorption (and thus idempotency). The rule is also equivariant with respect to vacuous extensions, because

$$\langle m_1^{\uparrow(R,r)}, \ldots, m_n^{\uparrow(R,r)} \rangle = \chi((Q_1 \circ \tilde{r}) \wedge \cdots \wedge (Q_n \circ \tilde{r})) = \chi((Q_1 \wedge \cdots \wedge Q_n) \circ \tilde{r}) = \langle m_1, \ldots, m_n \rangle^{\uparrow(R,r)}$$

for all refinements (R, r) of S, where the last equality is implied by the second condition for χ. Since

$$\{x \in S : (Q_1 \wedge \cdots \wedge Q_n)(\{x\}) > 0\} = \{x \in S : Q_1(\{x\}) > 0\} \cap \cdots \cap \{x \in S : Q_n(\{x\}) > 0\} = C(m_1) \cap \cdots \cap C(m_n),$$

the third condition for χ implies that the combination rule preserves certainty. Moreover, the rule generalizes Dempster's rule of conditioning, because in the case of conditioning, $Q_1 \wedge Q_2 = Q_I$ follows from Theorem 5 (where I is the unique jba with marginals m_1, m_2), and thus the first condition for χ implies that $\langle m_1, m_2 \rangle = \chi(Q_1)$ is the normalized version of m_I (when $m_I(\varnothing) < 1$). Finally, the first condition for χ implies also that the combination rule generalizes the minimum rule of possibility theory, since when $m_1, \ldots, m_n$ are consonant, $Q_1 \wedge \cdots \wedge Q_n$ is the commonality function associated with the result of the unnormalized version of that rule, as follows from the expression (5). □

Acknowledgments

The author gratefully acknowledges financial support from the Stefano Franscini Fund of ETH Zurich, and thanks the anonymous referees for their valuable comments.

References

[1] M. Cattaneo, Combining belief functions issued from dependent sources, in: J.-M. Bernard, T. Seidenfeld, M. Zaffalon (eds.), ISIPTA '03, Proceedings in Informatics, Vol. 18, Carleton Scientific, Waterloo, 2003, pp. 133–147.
[2] M. Cattaneo, Statistical Decisions Based Directly on the Likelihood Function, PhD thesis, No. 17122, ETH Zurich, 2007.
[3] B.R. Cobb, P.P. Shenoy, On the plausibility transformation method for translating belief function models to probability models, Int. J. Approx. Reasoning 41 (3) (2006) 314–330.
[4] F. Cuzzolin, Two new Bayesian approximations of belief functions based on convex geometry, IEEE Trans. Syst. Man Cybern., Part B: Cybern. 37 (4) (2007) 993–1008.
[5] F. Cuzzolin, Credal semantics of Bayesian transformations, in: T. Augustin, F.P.A. Coolen, S. Moral, M.C.M. Troffaes (eds.), ISIPTA '09, SIPTA, Manno, 2009, pp. 129–138.
[6] M. Daniel, Associativity in combination of belief functions; a derivation of minC combination, Soft Comput. 7 (5) (2003) 288–296.
[7] M. Daniel, On transformations of belief functions to probabilities, Int. J. Intell. Syst. 21 (3) (2006) 261–282.
[8] A.P. Dempster, Upper and lower probability induced by a multivalued mapping, Ann. Math. Stat. 38 (2) (1967) 325–339.
[9] A.P. Dempster, Logicist statistics II: Inference, in: R.R. Yager, L. Liu (eds.), Classic Works of the Dempster-Shafer Theory of Belief Functions, Studies in Fuzziness and Soft Computing, Vol. 219, Springer, Berlin, 2008, pp. 761–785.
[10] D. Denneberg, Non-Additive Measure and Integral, Theory and Decision Library, Series B: Mathematical and Statistical Methods, Vol. 27, Kluwer, Dordrecht, 1994.
[11] T. Denœux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artif. Intell. 172 (2–3) (2008) 234–264.
[12] S. Destercke, D. Dubois, Can the minimum rule of possibility theory be extended to belief functions?, in: C. Sossai, G. Chemello (eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Artificial Intelligence, Vol. 5590, Springer, Berlin, 2009, pp. 299–310.
[13] S. Destercke, D. Dubois, E. Chojnacki, Cautious conjunctive merging of belief functions, in: K. Mellouli (ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Artificial Intelligence, Vol. 4724, Springer, Berlin, 2007, pp. 332–343.
[14] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, Int. J. Gen. Syst. 12 (3) (1986) 193–226.
[15] D. Dubois, H. Prade, On the unicity of Dempster rule of combination, Int. J. Intell. Syst. 1 (2) (1986) 133–142.
[16] D. Dubois, H. Prade, The principle of minimum specificity as a basis for evidential reasoning, in: B. Bouchon, R.R. Yager (eds.), Uncertainty in Knowledge-Based Systems, Lecture Notes in Computer Science, Vol. 286, Springer, Berlin, 1987, pp. 75–84.
[17] D. Dubois, H. Prade, Possibility Theory, Plenum Press, New York, 1988.
[18] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Comput. Intell. 4 (3) (1988) 244–264.
[19] D. Dubois, R.R. Yager, Fuzzy set connectives as combinations of belief structures, Inf. Sci. 66 (3) (1992) 245–276.
[20] D. Dubois, H. Prade, P. Smets, New semantics for quantitative possibility theory, in: S. Benferhat, P. Besnard (eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Artificial Intelligence, Vol. 2143, Springer, Berlin, 2001, pp. 410–421.
[21] M. Ha-Duong, Hierarchical fusion of expert opinions in the Transferable Belief Model, application to climate sensitivity, Int. J. Approx. Reasoning 49 (3) (2008) 555–574.
[22] R. Haenni, Are alternatives to Dempster's rule of combination real alternatives? Comments on "About the belief function combination and the conflict management problem"—Lefevre et al, Inf. Fusion 3 (3) (2002) 237–239.
[23] R. Haenni, Shedding new light on Zadeh's criticism of Dempster's rule of combination, in: 2005 8th International Conference on Information Fusion (FUSION), Vol. 2, IEEE, Piscataway, 2005, pp. 879–884.
[24] Y.-T. Hsia, Characterizing belief with minimum commitment, in: J. Mylopoulos, R. Reiter (eds.), IJCAI-91, Vol. 2, Morgan Kaufmann, San Mateo, 1991, pp. 1184–1189.
[25] A. Jøsang, The consensus operator for combining beliefs, Artif. Intell. 141 (1–2) (2002) 157–170.
[26] A. Jøsang, M. Daniel, P. Vannoorenberghe, Strategies for combining conflicting dogmatic beliefs, in: Proceedings of the Sixth International Conference on Information Fusion, Vol. 2, ISIF, Fairborn, 2003, pp. 1133–1140.
[27] J.G. Kalbfleisch, D.A. Sprott, Fiducial probability, Stat. Pap. 8 (2) (1967) 99–109.

[28] A. Kallel, S. Le Hégarat-Mascle, Combination of partially non-distinct beliefs: The cautious-adaptive rule, Int. J. Approx. Reasoning 50 (7) (2009) 1000–1021.
[29] R. Kennes, Evidential reasoning in a categorial perspective: Conjunction and disjunction of belief functions, in: B. D'Ambrosio, P. Smets, P. Bonissone (eds.), Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Mateo, 1991, pp. 174–181.
[30] J. Kohlas, P.-A. Monney, A Mathematical Theory of Hints, Lecture Notes in Economics and Mathematical Systems, Vol. 425, Springer, Berlin, 1995.
[31] J. Lawry, J. Recasens, I. González Rodríguez, A similarity approach to evidence combination in Dempster-Shafer theory, unpublished manuscript, 2005.
[32] E. Lefevre, O. Colot, P. Vannoorenberghe, Belief function combination and conflict management, Inf. Fusion 3 (2) (2002) 149–162.
[33] X. Ling, W.G. Rudd, Combining opinions from several experts, Appl. Artif. Intell. 3 (4) (1989) 439–452.
[34] W. Liu, Analyzing the degree of conflict among belief functions, Artif. Intell. 170 (11) (2006) 909–924.
[35] P.-A. Monney, M. Chan, Modelling dependence in Dempster-Shafer theory, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 15 (1) (2007) 93–114.
[36] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, 1988.
[37] J.G. Pedersen, Fiducial inference, Int. Stat. Rev. 46 (2) (1978) 147–170.
[38] L. Rüschendorf, Sharpness of Fréchet-bounds, Z. Wahrscheinlichkeitstheorie verw. Gebiete 57 (2) (1981) 293–302.
[39] L. Rüschendorf, Fréchet-bounds and their applications, in: G. Dall'Aglio, S. Kotz, S. Salinetti (eds.), Advances in Probability Distributions with Given Marginals, Mathematics and its Applications, Vol. 67, Kluwer, Dordrecht, 1991, pp. 151–187.
[40] K. Sentz, S. Ferson, Combination of evidence in Dempster-Shafer theory, Tech. Report SAND 2002-0835, Sandia National Laboratories, Albuquerque, 2002.
[41] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[42] G. Shafer, Constructive probability, Synthese 48 (1) (1981) 1–60.
[43] G. Shafer, Belief functions and possibility measures, in: J.C. Bezdek (ed.), Analysis of Fuzzy Information, Vol. 1: Mathematics and Logic, CRC Press, Boca Raton, 1987, pp. 51–84.
[44] R.M. Shortt, Strassen's marginal problem in two or more dimensions, Z. Wahrscheinlichkeitstheorie verw. Gebiete 64 (3) (1983) 313–325.
[45] P. Smets, The combination of evidence in the transferable belief model, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990) 447–458.
[46] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reasoning 9 (1) (1993) 1–35.
[47] P. Smets, Analyzing the combination of conflicting belief functions, Inf. Fusion 8 (4) (2007) 387–412.
[48] P. Smets, R. Kennes, The transferable belief model, Artif. Intell. 66 (2) (1994) 191–234.
[49] V. Strassen, The existence of probability measures with given marginals, Ann. Math. Stat. 36 (2) (1965) 423–439.
[50] R.R. Yager, The entailment principle for Dempster-Shafer granules, Int. J. Intell. Syst. 1 (4) (1986) 247–262.
[51] R.R. Yager, On the Dempster-Shafer framework and new combination rules, Inf. Sci. 41 (2) (1987) 93–137.
[52] R.R. Yager, Quasi-associative operations in the combination of evidence, Kybernetes 16 (1) (1987) 37–41.
[53] R.R. Yager, L. Liu (eds.), Classic Works of the Dempster-Shafer Theory of Belief Functions, Studies in Fuzziness and Soft Computing, Vol. 219, Springer, Berlin, 2008.
[54] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets Syst. 1 (1) (1978) 3–28.
[55] L.A. Zadeh, Review of Shafer's "A Mathematical Theory of Evidence", AI Mag. 5 (3) (1984) 81–83.