On computing evidential centroid through conjunctive combination: an impossibility theorem

Yiru Zhang, Sébastien Destercke, Zuowei Zhang, Tassadit Bouadi, Arnaud Martin

On computing evidential centroid through conjunctive combination: an impossibility theorem

Yiru Zhang, Sébastien Destercke, Zuowei Zhang, Tassadit Bouadi and Arnaud Martin

***Abstract*—The theory of belief functions (TBF) is now a widespread framework to deal and reason with uncertain and imprecise information, in particular to solve information fusion and clustering problems. Combination functions (rules) and distances are essential tools common to both the clustering and information fusion problems in the context of TBF, which have generated considerable literature. Distances and combination between evidence corpus of TBF are indeed often used within various clustering and classification algorithms, however their interplay and connections have seldom been investigated, which is the topic of this paper. More precisely, we focus on the problem of aggregating evidence corpus to obtain a representative one, and we show through an impossibility theorem that in this case, there is a fundamental contradiction between the use of conjunctive combination rules on the one hand, and the use of distances on the other hand. Rather than adding new methodologies, such results are instrumental in guiding the user among the many methodologies that already exist. To illustrate the interest of our results, we discuss different cases where they are at play.**

***Impact Statement*—Within the theory of belief functions, both distances and conjunctive combination rules can be used to achieve very similar purposes: evaluating the conflict between sources, performing supervised or unsupervised learning in presence of evidential information, or more simply obtaining a synthetic representation of multiple items of information. However, the results obtained by both approaches may show some inconsistency between them. This paper provides some insight as to why this may happen, showing that the two approaches are definitely at odds, and that using distances is, for instance, incompatible with some fundamental notions of the theory of belief functions, such as the least commitment principle. We illustrate the importance of the studied differences on problems such as k-centroid clustering, and discuss the importance of interpretations in such problems, which is rarely done in the literature.**

***Index Terms*—Theory of belief functions, combination rules, metric, uncertainty reasoning, k-centroid clustering.**

I. INTRODUCTION

The theory of belief functions (TBF), also known as Dempster-Shafer theory (DST) or evidence theory, is a mathematical tool to reason under uncertainty. TBF enriches traditional representations of uncertainties by combining probabilities and sets in a unified framework. In the TBF, probability masses are assigned to sets of events rather than to a single event, making it possible to characterise a piece of information with both uncertainty and imprecision simultaneously. Therefore, when dealing with imprecise or missing knowledge, the TBF can encode them without imposing too constraining assumptions. In this paper, we will refer to information encoded by TBF tools as *evidence corpus*.

Combination operation and distance/conflict measures between evidence corpus are essential tools used within TBF to manipulate information. Combination operations, on one hand, are defined by rules that are commonly used to fuse information by combining multiple evidence corpus, resulting in a summary of the different pieces of evidence. There is considerable literature on combination rules (see [1], [2], [3] for some seminal works and [4], [5], [6], [7], [8], [9] for more recent ones), witnessing the importance of this operation within the theory from various aspects. Among those, Dempster's combination rule [1], a product rule acting conjunctively on the sources of information, is the most emblematic of all. It assumes that all information sources are reliable and cognitively independent, and most other rules in the literature try to depart from these two assumptions, such as the disjunctive Dubois-Prade's rule [10] that relaxes the need for reliability, or idempotent conjunctive rules [4], [5], [11] that relax the independence assumption. The combination rules are widely applied in data fusion applications, such as [12], [13], *etc.*

On the other hand, distances between evidence corpus are commonly used to measure their dissimilarities, which may in turn be used to evaluate source reliability to perform a fusion or combination [14], for example. Jousselme and Maupin's critical survey [15] shows a great variety of distances based on the notion of inner products, and other authors such as Perry and Stephanou [16], [17] proposed yet other dissimilarity measures based on extensions of classical probabilistic measures such as entropy or Kullback-Leibler divergence.

Combination and distance measures often play similar roles in different AI fields. For instance, distance-based merging is common in logic [18], [19] and argumentation [20]. Studies on proper representation of distances or manifolds of knowledge are also helpful in the selection combination process [21], [22].

Given all this, it is natural to wonder to which extent, within the TBF, the notions of distance and combination are compatible, particularly when it comes to computing a representative evidence corpus in a group. This is the question we want to address in this paper, or more precisely:

> *Can we use combination rules to compute the centroid of a group of evidence corpus?*

Yiru Zhang is with Laboratoire ETIS, UMR 8051, CY Cergy Paris Universités, ENSEA, CNRS F-95000, Cergy, France. (e-mail: yiru.zhang1@cyu.fr).

Sébastien Destercke is with Sorbonne Universités, Université de technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, 57 Av. de Landshut, 60200 Compiègne, France. (e-mail: sebastien.destercke@hds.utc.fr).

Zuowei Zhang, Tassadit bouadit and Arnaud Martin are with Université Rennes, CNRS, IRISA, DRUID, 22300 Lannion, France. (e-mail: zhangzuowei0720@gmail.com, tassadit.bouadi@irisa.fr, arnaud.martin@univ-rennes1.fr).

The answer to this question can have a significant practical impact. Indeed, both combination and distances are commonly used in clustering issues, sometimes conjointly [23], in the construction of classification models [24], [25] or in the construction of fusion rules [26], hence the importance of better understanding their interplay.

Moreover, the distances for evidential corpus are also used in some learning applications. Hariz *et al.* [27] applies Jousselme distance [28] to define the membership to each cluster of an evidential corpus. Li *et al.* [29] used the same distance to cluster uncertain information represented by evidential corpus. These methods remained on a methodological level with empirical comparisons. In [30], the authors pointed out that learning over evidential corpus representing imprecision may return counter-intuitive results, when they tried to cluster uncertain preferences modeled by TBF. Given the common use of distances and combination rules within evidence theory to perform some tasks (information combination, unsupervised or supervised learning, *etc.*), gaining theoretical insight about the proper use of distances and of combination rules, and their interplay is therefore of essential importance.

Our contribution shows that aggregation operators based on commonly used combination rules and distances necessarily provide different results, indicating that choosing one or the other in specific applications may have a significant impact. The conclusion drawn from a thorough discussion is mathematically proved with very few assumptions. Consequently, the impossibility theorem explains why some tools may fail or differ in performing a particular task, as distance and combination rules cannot be used interchangeably.

The rest of this paper is organized as follows. Section II recalls the basics about the TBF and the combination rules within it, as well as the definition of centroid within this framework. Section III then states our principal result, that shows that the notion of conjunctive combination (as well as other set-based fusion rules) and of centroid cannot be reconciled. Section IV then provides some illustrations showing that the discrepancy between the two approaches can indeed be quite high in practice. Finally, Section V discusses various views about clustering evidential corpus, beyond the centroid notion considered here.

II. Preliminary

In this section, we introduce preliminary concepts used in the paper. Firstly, basic definitions in the theory of belief functions (TBF) are introduced. We then recall the basics of combination rules and centroid computation.

A. Basics on the TBF

In this paper, the transferable belief model (TBM) [31] and belief functions are used to quantify uncertainty. In this model, the set of values that an uncertain quantity can take is defined as a frame of discernment (FoD).

Definition 1. Frame of discernment: *The frame of discernment (FoD) Ω is a finite set of disjoint elements, defining the domain of reference, formally:*

$$\Omega = \{\omega_1, \dots, \omega_p\}, \tag{1}$$

where ω_i are exclusive and exhaustive[1].

The belief functions or their equivalent representations are defined on the power set $2^\Omega = \{X : X \subseteq \Omega\}$. They offer flexible and rich tools to model uncertainty, generalising well-known models such as classical probabilities or possibility distributions. An *evidence corpus* or *evidence object* is represented by a single *mass function*.

Definition 2. Mass function: *A function $m : 2^\Omega \to [0,1]$ is called* mass function *on 2^Ω if it satisfies:*

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{2}$$

Every set A of elements such that $m(A) > 0$ is called a ***focal set*** of m with A named ***focal element***. A mass function is called ***normalized*** if $m(\emptyset) = 0$. A mass function on FoD Ω is usually denoted as m^Ω. We will denote by $\mathcal{E}(\Omega)$ the set of masses, or equivalently of evidence corpus defined on Ω.

Within the TBF, the uncertainty and imprecision of a piece of information are therefore represented by a mass function. We use terms evidence corpus and mass function interchangeably in this paper. Evidence corpus is usually used in the information fusion, aggregation and clustering applications as it bears a semantic, while the mathematical notion of mass function is semantically neutral. An evidence corpus only expressing imprecision and being equivalent to a set is called *categorical mass function*. A categorical mass function m on element $A, A \subseteq \Omega$, i.e., $m(A) = 1$ is denoted as $A^{\mathbf{1}}$, with the value 1 in bold, using the standard notation for simple mass functions[2]. A categorical mass function m on Ω , *i.e.*, $m(\Omega) = 1$, is called *vacuous*, representing ***total ignorance***.

A mass function m is called a *Bayesian mass function* if the focal elements are only singletons:

$$m(A) = 0, \forall A \text{ s.t. } |A| \neq 1. \tag{3}$$

where $|A|$ denotes the carnality of set A. Bayesian mass functions are formally equivalent to probability mass distributions. Within TBF, one also uses various set-functions measuring different aspects of our uncertainty about an event A.

Definition 3. Belief/plausibility/commonality function: *The belief function $Bel(\cdot)$, plausibility function $Pl(\cdot)$, and commonality function $Q(\cdot)$ of a set A are respectively defined as:*

$$Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B); \tag{4}$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B); \tag{5}$$

$$Q(A) = \sum_{B \supseteq A} m(B). \tag{6}$$

Bel represents how much event A is implied by the current evidence, and Pl how much it is consistent with the current evidence. They can also be considered as lower and upper bounds of the probability for event A. However, in this paper,

[1]To facilitate both reading and notations, we do not make an open world assumption, as in Smets TBM [2].

[2]A simple mass function A^α is such that $m(A) = \alpha$, $m(\Omega) = 1 - \alpha$.

we will rather consider Shafer's initial interpretation of masses as pieces of evidence [32], unless specified otherwise.

Commonality function Q is another equivalent representation of a mass function with trickier interpretation but computational interest (notably in the combination calculation).

B. Combination of evidence corpus

Given a set $\mathcal{S}(\Omega) = \{m_1, m_2, \dots, m_\ell\}$ of evidence corpus on FoD Ω, a combination rule f is a mapping function aggregating multiple evidence corpus into one, that aims at building a mass synthesising the information in $\mathcal{S}(\Omega)$. Formally:

$$f : \mathcal{S}(\Omega) \to m, m \in \mathcal{E}(\Omega), \tag{7}$$

where m denotes a mass function defined on the same FoD. While we will discuss some specific combination rules in further sections, such a specification is not needed to prove the main results of the paper. Therefore we will limit our current presentation to the main rule of the TBF, *Dempster's rule*, and delay the description of other rules after proving that most combination rules based on set-operations such as conjunction and disjunction are incompatible with centroid computation. *Dempster's rule* is known to play a pivotal role in the theory of belief functions [11]. Let m_x and m_y be two mass functions, with the inconsistency[3] κ between m_x and m_y defined as:

$$\kappa = \sum_{B,C\subseteq\Omega, B\cap C=\emptyset} m_x(B)m_y(C). \tag{8}$$

If $\kappa < 1$, then the result of Dempster's rule $\bigoplus$ applied to m_x and m_y is:

$$(m_x \bigoplus m_y)(A) = \frac{1}{1-\kappa} \sum_{B\cap C=A} m_x(B)m_y(C), \quad \forall A \subseteq \Omega, A \neq \emptyset. \tag{9}$$

C. Least commitment principle in TBF

Least commitment principle (LCP) is a thumb rule in decision making with multiple evidence sources, defined as:

Definition 4. Least commitment principle (LCP) [33] *When several belief functions are compatible with a set of constraints, the least informative according to some informational ordering (if it exists) should be selected.*

In TBF, the LCP is usually defined based on the plausibility. Given two plausibility functions Pl_1 and Pl_2 on the same FoD Ω, if $Pl_1(A) \leqslant Pl_2(A), \forall A \subseteq \Omega$, we say that Pl_2 is no more committed than Pl_1 (and less committed if there is at least one strict inequality).

D. Centroid of evidence corpus

When aiming at aggregating/clustering evidence corpus by computing a centroid, we must first specify a dissimilarity measure d over such structures such that:

$$d : \mathcal{E}(\Omega) \times \mathcal{E}(\Omega) \to \mathbb{R}_{\geq 0} \tag{10}$$

is a *metric* with the following properties: $\forall m_x,\ m_y,\ m_z \in \mathcal{E}(\Omega)$, d satisfies

- Non-negativity: $d(m_x, m_y) \geq 0$
- Identity of indiscernible: $d(m_x, m_y) = 0 \Leftrightarrow m_x = m_y$
- Symmetry: $d(m_x, m_y) = d(m_y, m_x)$
- Triangle inequality: $d(m_x, m_z) \leq d(m_x, m_y) + d(m_y, m_z)$

Computing the centroid is then the problem of finding the mass $\overline{m}$, defined from an arbitrary set $\{m_1, \dots, m_\ell\}$ of mass functions as:

$$\overline{m} = \arg \min_{m\in\mathcal{E}(\Omega)} \sum_{i=1}^{\ell} d(m_i, m). \tag{11}$$

In the rest of the paper, we assume that Equation (11) has a unique solution to facilitate developments. This is verified, for instance, as soon as d is also continuous and convex. Note that computing such centroid is essential in most quantization procedures, such as the celebrated k-means algorithm and its generalisations [34].

Remark 1 (Centroid and LCP)**.** *Let us note that, just like applying the LCP requires to define an ordering between belief functions (in Definition 4, we took plausibility ordering, while others are possible [35]), with the result of its application depending on this ordering, applying the centroid "principle" and Equation* (11) *requires defining a distance, to which can then be associated desirable properties such as the need to account for focal element interactions. A discussion about the interpretation of distances can be found in [36].*

III. The impossibility theorem

We now deal with the main contribution of this paper, which concerns the clustering and aggregation of evidence corpus, and the fact that computing a centroid is incompatible with a set of properties required by most combination rules in the TBF, and Dempster's rule in particular. More precisely, we will show that obtaining the result of Equation (11), *i.e.*, computing the centroid, cannot be achieved by applying a combination rule $\bigodot$ on the masses as soon as we require a set of commonly requested properties for $\bigodot$.

A. Metric consistency between centroid computation and combination rule

The first property simply formalizes a necessary condition for the centroid to result from a combination.

Property 1. Metric consistency: *The combination rule $\bigodot$ and the dissimilarity measure d must be consistent. Formally, this means that the combined mass function $\overline{m}$ calculated by:*

$$\overline{m} = \bigodot_{i=1}^{\ell} m_i, \tag{12}$$

should be the centroid, i.e.,

$$\sum_{i=1}^{\ell} d(m_i, m') > \sum_{i=1}^{\ell} d(m_i, \overline{m}),\ \forall m' \in \mathcal{E}(\Omega), m' \neq \overline{m}. \tag{13}$$

[3]This inconsistency is considered as a conflict measure only when the FoD is defined in a closed world *i.e.*, all possible results are covered by the FoD. In an open world, the inconsistency is often considered as ignorance.

This property of metric consistency is for example necessary if we want k-centroid clustering algorithm to converge. A k-centroid clustering problem can be regarded as an optimisation problem. In [34], the authors proved that this optimisation is solvable if the inertia of clusters[4] is a decreasing converging function of the number of steps. In the proof, the property given by Equation (13) is applied to guarantee the convergence of the clusters inertia. We should note that Equation (13) would turn into a non-strict inequality if the uniqueness of the centroid is not assumed. However, assuming uniqueness is a classical requirement in combination and in k-centroid problems, and the uniqueness is guaranteed if the metric is a Riemannian distance, as proved in [37]. Given a distance (metric) function, centroid calculation can be considered as a search for *geometric median* or Fermat-Weber point [38].

Note that unconstrained combination rules based on the minimisation of distance can always be made consistent with a corresponding distance applied in Equation (11). However, as we shall see next, adding some common constraints to the combination quickly makes the property of *metric consistency* impossible to satisfy.

B. Ignorance neutrality in combination process

When looking for a synthesis of information provided by sources or evidence corpus, it is common to start with conjunctive rules. Such rules typically assume that the provided information is reliable enough, as in the case of Dempster's rule where sources are assumed to use independent, distinct and reliable information. For example, such combinations are at work in the well-known evidential KNN algorithm proposed by Denœux [25], which has been extended into many versions.

A property satisfied by all conjunctive rules, but also in general by rules trying to combine conjunctive and disjunctive behaviours when dealing with conflict (such as the Dubois/Prade rule [10]), is that vacuous information should not modify the combined, or synthetic mass function. We name this property *ignorance neutrality*, formally defined as follows:

Property 2. Ignorance neutrality: *Vacuous mass functions representing ignorance are neutral elements of the combination rule $\odot$, i.e.,*

$$m \odot \Omega^{\mathbf{1}} = m, \quad \forall m \in \mathcal{E}(\Omega), \tag{14}$$

where $\Omega^{\mathbf{1}}$ denotes a vacuous mass function ($\Omega^{\mathbf{1}} : m(\Omega) = 1$).

We can illustrate the usefulness (if not necessity) of this property by an example of clustering over imperfect preference information, borrowed from [30].

Example III.1. *In evidential preference models [39], [40], the preference relation between two items is modelled by FoD $\Omega_{pref} = \{\omega_{\succ}, \omega_{\prec}, \omega_{\approx}\}$, respectively representing strict preference, inverse strict preference, and indifference. If unobserved sources are represented by vacuous mass function $\Omega^{\mathbf{1}}_{pref}$, it is then natural to require the unobserved data to plays a neutral role in the preference aggregation process (which is also a combination process).*

[4]In clustering, inertia is also known as within-cluster sum-of-squares criterion, describing how internally coherent clusters are. The calculation is given in Algorithm 1.

Ignorance neutrality can be extended to the case of partial ignorance. That is to say, given a mass function m with focal sets $A_1, \ldots, A_n$, and a more imprecise information represented by categorical mass function $B^{\mathbf{1}}$ such that $A_1, \ldots, A_n \subseteq B$, neutrality to partial ignorance can be formalized as follows:

$$m \odot B^{\mathbf{1}} = m, \tag{15}$$

We can extend this property to general cases by considering inclusion notions proper to TBF [36].

Indeed, (partial) ignorance neutrality is also required by the LCP. The proof is straightforward: consider a categorical mass function $m_1 = B^{\mathbf{1}}$ representing partial ignorance, and a mass functions m_2 in FoD Ω, with focal sets $A_1, \ldots, A_n \subseteq B$. According to Equation (5):

$$Pl_1(A) = 1, \forall A \subseteq B. \tag{16}$$

Hence:

$$Pl_2(A) \leqslant Pl_1(A), \forall A \subseteq B, \tag{17}$$

i.e., m_2 should be selected when combining m_1 and m_2.

Finally, we can also note that ignorance neutrality offers protection from artificially adding new vacuous sources of information to the corpus of knowledge, provided for instance by malicious sources.

C. Idempotence in combination process

Idempotence is a desirable property when cautiousness is wanted in an information fusion process. It says that combining two identical evidence corpus should result in the same evidence corpus, therefore not reinforcing the available information. It is formally defined as follows:

Property 3. Idempotence: *The combination rule $\odot$ is idempotent if and only if:*

$$m \odot m = m, \quad \forall m \in \mathcal{E}(\Omega). \tag{18}$$

Idempotence is required by some combination rules, especially *when the dependencies between sources are ill-known, as this property ensures that the same information supplied by two possibly dependent sources will remain unchanged after merging (combination)* [41]. This property also guarantees that identical information items are not counted multiple times in the fusion process, a feature often required in cases where the fusion process is decentralized [42], which may well happen in clustering processes with a high number of evidence corpus.

Obviously, if $m_1 = m_2 = \ldots = m_\ell = m$, then the solution to Equation (11) should be m. As shown in the next proposition, this means that the idempotence property (is necessary) must be satisfied by any combination rule satisfying metric consistency, as formalized in the next proposition.

Theorem 1. *The property of "idempotence" is a necessary condition of "metric consistency".*

Proof: Let us consider a combination rule $\odot$ with the property of consistent metric for a dissimilarity function $d(\cdot)$.

Given a mass function $m, m \in \mathcal{E}(\Omega)$ and $\forall\ m' \neq m, m' \in \mathcal{E}(\Omega)$ according to the property of identity of indiscernible and the non-negativity of the distance d, we have:

$$d(m, m') > 0. \tag{19}$$

Knowing that $d(m, m) = 0$, we have:

$$d(m, m) + d(m, m) < d(m, m') + d(m, m').$$

According to Equation (13), we obtain the following equation.

$$m \odot m = m.$$

The combination rule $\odot$ is idempotent. ■

Note that the property of idempotence can also be related to k-centroid clustering. The centroid of a set of objects with identical value should be assigned with the same value.

D. The impossibility theorem

Theorem 1 concludes that the property of "idempotence" is weaker than the property of "metric consistency", and itself is necessary to obtain a centroid through combination. What we show next is that metric consistency is incompatible with ignorance neutrality.

Theorem 2 (The impossibility theorem)**.** *Given a dissimilarity measure $d : \mathcal{E}(\Omega) \times \mathcal{E}(\Omega) \to \mathbb{R}_{\geq 0}$ on the set of evidence corpus, there is no combination rule $\odot$ that satisfies the properties of* ***metric consistency*** *and* ***ignorance neutrality*** *simultaneously.*

Proof: We already know that for any $m' \neq \overline{m}$,

$$\sum_{i=1}^{\ell} d(m_i, m') > \sum_{i=1}^{\ell} d(m_i, \overline{m}). \tag{20}$$

Let us now assume, *ex absurdo*, that a combination rule $\odot$ satisfying the two properties exists. Let us now consider a mass function m, together with a categorical (possibly vacuous) mass $A^{\mathbf{1}}$ such that A is a super-set of all focal elements of m, and is distinct from m. From our assumptions, it follows that $m \odot A^{\mathbf{1}} = m$ is the centroid of $\{m, A^{\mathbf{1}}\}$. This means in particular that $\forall m' \neq m$,

$$d(m, m) + d(m, A^{\mathbf{1}}) < d(m, m') + d(m', A^{\mathbf{1}}), \tag{21}$$

and since $d(m, m) = 0$ by definition, we have

$$d(m, A^{\mathbf{1}}) < d(m, m') + d(m', A^{\mathbf{1}}).$$

Take $m' = A^{\mathbf{1}}$, and the above inequality becomes

$$d(m, A^{\mathbf{1}}) < d(m, A^{\mathbf{1}}) + d(A^{\mathbf{1}}, A^{\mathbf{1}}),$$

equivalent to:

$$d(A^{\mathbf{1}}, A^{\mathbf{1}}) > 0.$$

This is contradictory with the property of identity discernible. Thus our initial assumption is false. ■

Since all conjunctive rules satisfy the property of ignorance neutrality [43], the following corollary can be easily deduced:

Corollary 1. *No metric is consistent with conjunctive combination rules.*

This corollary includes in particular the Dempster's rule of combination, but also others which we will discuss in the following section. Our impossibility theorem indicates that no conjunctive rules or non-idempotent rules can be made consistent with a centroid computation based on distances. This clearly means that when developing a method within the TBF that requires to build a consensus or to aggregate mass functions, one should carefully think about the view to adopt. Similarly, our discussion at the end of Section III-B indicates that centroid computation cannot be made consistent with a classical LCP approach.

The next section explores some popular rules existing within TBF, summarising their properties and showing on simple examples that the practical differences between a centroid and the combined mass function obtained through such rules can vary from null to quite important.

IV. Case study of combination rules

This section illustrates our results on some well-known combination rules, demonstrating with examples that idempotence is not sufficient to be metric consistent, and that ignorance neutrality is conflicting with a metric consistency requirement. Indeed, all the rules we consider in this section satisfy ignorance neutrality, while some of them satisfy idempotence. The chosen examples also illustrate that the discrepancy between the results of the two approaches (combining *vs.* computing a centroid) may be quite significant. At the end of the section, an illustrative example of k-centroid clustering process using different combination rules is given, showing the importance of the property of metric consistency.

A. Properties of different combination rules

As mentioned before, combination rules are a cornerstone of the TBF, and many have been proposed in the literature. As our proof is quite general and have few assumptions, we will not be exhaustive, and will only focus on the best-known rules that satisfy ignorance neutrality or idempotence. The definitions of these rules are all based on two mass functions m_x and m_y.

Dubois and Prade's (D-P) rule [10]: This rule is a prototypical example of a compromise and non-conjunctive rule satisfying ignorance neutrality. It proposes to solve the conflict by adopting a disjunctive behaviour whenever two pieces of information (two sets) conflict, and a conjunctive one in other situations. This rule is defined as follows:

$$m_x \underset{DP}{\odot} m_y(A) = \sum_{B \cap C = A} m_x(B) m_y(C) + \sum_{\substack{B \cup C = A \\ B \cap C = \emptyset}} m_x(B) m_y(C). \tag{22}$$

Denœux's cautious conjunctive rule [11]: This rule is based on Smets canonical decomposition [44], and on the use of t-norms on these weights. When using the minimum t-norm, we obtain a so-called cautious rule that satisfies idempotence, and can be used, for instance, to tolerate redundancy in the combined information, useful for example, when the information sources are non-independent. The conjunctive version $\owedge$

is defined as:

$$m_x \owedge m_y = \bigoplus_{\emptyset \neq A \subset \Omega} A^{min(w_x(A), w_y(A))}, \tag{23}$$

where $\bigoplus$ refers to Dempster's rule, and $w(A)$ the **weight function** on element A, is formally defined as:

$$w(A) = \prod_{B \supseteq A} Q(B)^{(-1)^{|B|-|A|+1}}, \forall A \subseteq \Omega, \tag{24}$$

where Q refers to the commonality function (Equation (6)).

Klein-Destercke-Colot's (K-D-C) idempotent rule [5]: This rule is based on minimisation of distance, and the conjunctive version is defined as:

$$m_x \bigodot_{f,p} m_y = \underset{\overline{m} \in \mathcal{S}_f(m_x) \cap .. \cap \mathcal{S}_f(m_y)}{\arg\min} d_{f,p}(\overline{m}, m_\Omega), \tag{25}$$

where $\mathcal{S}_f(m_x)$ denotes the set of mass functions more informative than m_x in the sense of f. f is a set function (such as Bel, Pl, or Q), and p the order of the norm of the Minkowski distance, usually taken as $p = 2$ to ensure convexity of the resulting optimisation problem and uniqueness of the solution. This rule is idempotent by definition.

Minimisation of distance sum (Mean): This rule is a simple average of the different mass functions, and provides the combined mass:

$$\overline{m}(A) = \frac{1}{\ell} \sum_{i=1}^{\ell} m_i(A), A \subseteq \Omega, \tag{26}$$

where ℓ denotes the number of evidence corpus. It is obviously idempotent and is not neutral to ignorance. In fact, as an average, it is the centroid when the chosen distance is the L_2 norm taken between mass vectors.

Let us now provide an example showing that these rules, except for the averaging one, do not allow to obtain centroids. Consider three mass functions in a simple binary FoD $\Omega = \{\omega_1, \omega_2\}$. $m_1 = m_2 = \{0, 0.5, 0, 0.5\}$ are Bayesian mass functions with identical values, $m_3 = \{0, 0, 0, 1\}$ is a vacuous mass function. The combination results of the different rules are shown in Table I.

TABLE I
COMBINATION RESULTS OBTAINED BY DIFFERENT RULES:

	$\{\emptyset\}$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
m_1	0	0.5	0	0.5
m_2	0	0.5	0	0.5
m_3	0	0	0	1
Dempster rule [1]				
$m_1 \oplus m_2$	0	0.75	0	0.25
$m_1 \oplus m_3$	0	0.5	0	0.5
D-P rule [10]				
$m_1 \odot_{DP} m_2$	0	0.75	0	0.25
$m_1 \odot_{DP} m_2$	0	0.5	0	0.5
Denœux's cautious rule [11]				
$m_1 \owedge m_2$	0	0.5	0	0.5
$m_1 \owedge m_3$	0	0.5	0	0.5
K-D-C conjunctive rule [5]				
$m_1 \odot_{Pl,2} m_2$	0	0.5	0	0.5
$m_1 \odot_{Pl,2} m_3$	0	0.5	0	0.5
Minimisation of distance				
$m_{1,2}$	0	0.5	0	0.5
$m_{1,3}$	0	0.25	0	0.75

The combination results with blue and red background colours respectively violate the property of *idempotence* and *ignorance neutrality*. It is clear that none of the rules can retrieve simultaneously the two bottom masses of the table, which are the centroid obtained using Jousselme distance [28]. Moreover, they cannot be centroid for another distance, as they are either non-idempotent (D-P and Dempster), or satisfy both idempotence and ignorance neutrality (Denœux and K-D-C), meaning that Theorem 2 applies. The difference is sometimes quite noticeable, for instance if one compares the result of Dempster's or Dubois-Prade's rule applied to m_1 and m_3, with the corresponding centroid: the concentration of weights is completely reversed. Table II summarises the various properties of the rules we consider here. We added some remarks for results associated to the asterisk symbol *.

TABLE II
PROPERTIES OWNED BY COMBINATION RULES:

Combination rules	I.N.	idem	M.C.
Dempster's rule [1]	T	F	F
D-P rule [10]	T	F	F
Denœux's cautious rule [11]	F*	T	F
K-D-C conjunctive rule [5]	T	T	F*
Minimisation of distance	F	T	T

Remark 2. *On Denœux's conjunctive cautious rule: the property of ignorance neutrality holds if and only if the mass to be combined m is separable in the sense of the canonical decomposition. A detailed demonstration is provided in [4].*

Remark 3. *On K-D-C's idempotent rule: Despite the fact that K-D-C's idempotent rule is also distance (metric)-based according to its original definition, it does not respect the property of metric consistency defined in this paper. In this combination rule, distances are not measured between evidence corpus to be combined, but to the vacuous mass function Ω^1, which is the least informative mass function. This strategy guarantees the property of conjunctivity.*

B. An empirical demonstration of metric consistency

As mentioned in Section I, important applications of the notion of centroid are clustering problems. In particular, the property of centroids given by Equation (13) is important to guarantee the convergence of the clustering algorithm to a local minimum, and it is proved in [34] that it is a sufficient condition for converging to a local optimum. Whether this condition is necessary is still an open question. In this part, we empirically demonstrate that departing from this property by using a combination rule may lead to severe degradation of the results.

In order to prove our point, we simply apply an adapted k-centroid algorithm to randomly generated evidential corpus, where the computation of the centroid is replaced by a combination rule. In this experiment, 2000 evidential corpus are randomly generated on a 3-element discernment frame $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The adapted k-centroid algorithm is detailed in Algorithm 1, with the size of cluster set to $|\mathcal{C}| = 8$. To reduce variability, the averages of 20 experiments are reported.

Algorithm 1: Clustering algorithm over evidence corpus

Result: Clustering results of all evidence corpus
Input: Number of evidence corpus N, Cluster set $\mathcal{C}$, combination rule $\odot$, iteration times Max_it
Randomly initialize cluster centers;
iter=0;
while *iter* $< Max\_it$ **do**
  Calculate $|\mathcal{C}|$ cluster center using $\odot$, denoting $\overline{m}_c$ the center of cluster c;
  Assign each evidence corpus m_i to cluster $c_i = \arg\min_{c \in \mathcal{C}} d(\overline{m}_c, m_i)$;
  Calculate the inertia J of clustering result by $J = \sum_{i \in N} d(m_i, \overline{m}_{c_i})$;
  iter=iter+1;
end

Figure 1 displays the results of Algorithm 1 applied to all the combination rules mentioned in Table I and Table II. When the minimisation of distance is applied, the k-centroid algorithm becomes a classical k-means computation. It is clear that while those combination rules seem to converge to some results, those are sub-optimal and not necessarily decreasing. The cases of Dempster's rule, D-P rule and Denœux's cautious are particularly striking, as it brings no change and provide very bad solutions. This shows that the difference proved in Theorem 2 is not merely cosmetic, but may actually lead to very different results.

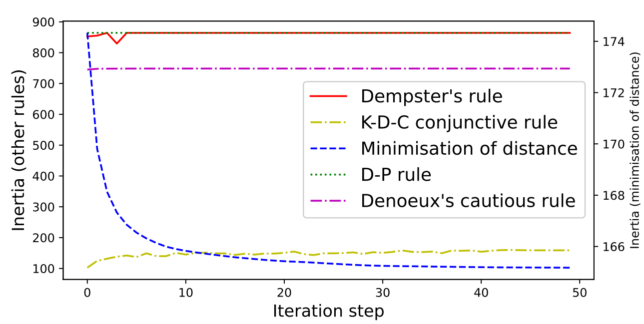

Fig. 1. Inertia of clustering with different combination rules

The experiment results also siren the risk of arbitrarily applying clustering/classification algorithms over evidence corpus, which is routinely done in previous works, such as [7], [30], [27]. The interpretation of evidence corpus should be clarified before executing any application over them. In the following section, we discuss different ways of clustering evidence corpus according to the intended goal and considered interpretation, showing again that those can lead to significantly different results.

V. Various views on clustering evidence corpus

In this section, we want to emphasize that there may be various ways and views to aggregate and cluster evidential objects. In particular, we want to draw attention, by using well-chosen examples, that aggregating knowledge about an object, and that aggregating the objects that are uncertain is not the same thing at all. We will illustrate the different viewpoints considered in this section with brief examples.

Consider a simple binary FoD $\Omega = \{\omega_1, \omega_2\}$, along with three clusters of mass functions respectively close to ω_1, ω_2 and Ω. Since mass functions live on a unit simplex, the values of those masses on ω_1, ω_2 and Ω can be represented in a 3-dimension Cartesian coordinate system as in Figure 2. According to Equation (2), a point representing a mass function will be within the boundaries of the shadowed triangle. This triangle corresponds to the *belief space* $\mathcal{B}$ defined by [45], whereas the edge between (ω_1^0, ω_2^0) represents a probabilistic space $\mathcal{P}$ (still after Cuzzolin [45]). The zone with darker shadow represents less imprecision and vice versa. Nine mass functions are considered with values summarized in Table III.

TABLE III
VALUE SETTING OF EVIDENCE CORPUS:

	Group 1			Group 2			Group 3		
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
$m(\omega_1)$	0	0.1	0	0	0.1	0	0.9	0.9	1
$m(\omega_2)$	0.1	0	0	0.9	0.9	1	0	0.1	0
$m(\Omega)$	0.9	0.9	1	0.1	0	0	0.1	0	0

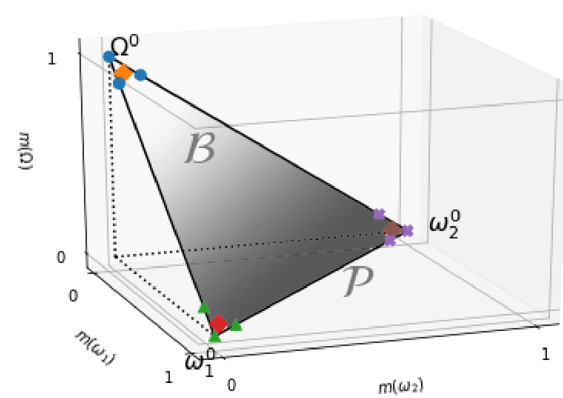

Fig. 2. evidence corpus in binary FoD and k-means clustering result with Minkowski metric

A. A metric view

In the metric view, every corpus expressed by a mass function is directly projected in a space $\mathcal{S}$, in which a dissimilarity function given in Equation (10) with metric properties can be defined. In this projection, each element in the space of 2^Ω is regarded as a dimension orthogonal to others. Table III already clusters or groups the different evidence corpus as they would be if we adopt metric view. In Figure 2, the clusters' centroids calculated with metric (let us say Jousselme distance [28]) are marked by diamond mark ♦. It is clear from Figure 2 that this clustering makes sense as soon as we see mass functions as points in a high-dimensional space, and we want to cluster those points. This also means that in this case, masses are treated as ontic objects [46], and what we are really interested in is summarising their characteristics (*e.g.*, in terms of informative content). Still, as shown in the following possibilities we explore, this is not necessarily coherent with a more epistemic consideration of these mass functions, where, however, exist the interests of reasoning uncertainty by TBF. Indeed, looking at the mass functions and trying to decide whether they could be of type ω_1 or ω_2, one would not necessarily put the objects represented by m_1 and m_2 together, as the first slightly lean towards ω_2, and the second towards ω_3.

B. A conflict-based view

TABLE IV
CONFLICT DEGREE κ BETWEEN EVIDENCE CORPUS:

κ	m_4	m_5	m_6	m_7	m_8	m_9
m_1	0	0.01	0	0.09	0.09	0.1
m_2	0.09	0.09	0.1	0	0.01	0
m_3	0	0	0	0	0	0

Conflict measures have also been used quite widely to decide whether two mass functions are similar/dissimilar [47], usually for combination processes. Given two mass functions m_1 and m_2, a conflict measure $\kappa(m_1, m_2)$ on combination rule is classically defined as:

$$\kappa(m_1, m_2) = (m_1 \bigodot m_2)(\emptyset). \tag{27}$$

In the jungle of combination rules and conflicting measures, the most seminal one is Dempster's rule $\bigoplus$, which we also use as a basis for our discussion. From Equation (8), it can be easily seen that a mass representing ignorance is neutral, *i.e.*, given a mass function m with focal sets $A_1, \ldots, A_n$, and a categorical mass function $B^{\mathbf{1}}$ on element B, such that $A_1, \ldots, A_n \subseteq B$, the following equation is true:

$$\kappa(B^{\mathbf{1}}, m) = 0. \tag{28}$$

Table IV shows the conflict values between m_1, m_2, m_3 and m_4 to m_9 that are given in Table III.

Here, we can easily see that such a conflict view indeed concerns the described object, and not the mass functions themselves. If we look at the conflict between masses measured by Equation (8), it is more reasonable to assign m_1 to the cluster of Group 2 (made up by m_4, m_5, m_6) and m_2 to the cluster of Group 3 (made up by m_7, m_8, m_9), as m_1 is an evidence corpus supporting ω_1 and m_2 supporting ω_2.

Obviously, even though conflict measure satisfies the property of ignorance neutrality, it does not yield to others: for instance it is not a metric as it does not satisfy the *identity of indiscernible*. Therefore, the impossibility theorem still largely applies to conflict-based view, meaning that they are at odd with distance-based approaches. This supports somehow the claim [48] that conflict and distances are two different key notions, yet it should be noticed that other recent works try to reconcile these two notions [14], [49].

C. An (imprecise) probabilistic view

One can also interpret evidence corpus with a probabilistic view in mind, or try to return to conventional probabilities on the singletons of FoD by some transformations. One example of such works is the ones of Cuzzolin [22], which adopt a geometric interpretation of the issue. Here, we are confronted with two different views:

- The first one consists of transforming mass functions into representative probabilities over the initial FoD, for instance, using the well-known pignistic transformation [50], calculated by Equation (29), or plausibility transforms [51] (Equation (5)).

$$BetP(\omega) = \sum_{A \subseteq \Omega, \omega \in A} \frac{1}{|A|} \frac{m(A)}{1 - m(\emptyset)} \tag{29}$$

  In such a view, m_1, m_2 and m_3 are transformed into uniform distributions on the singletons, and one could argue that in the less expressive setting of probabilities, the same clustering as in Table III would make sense.
- The second view consists of seeing mass functions as ill-known probabilities, and consider the bounds given by Equations (4) and (5) as lower and upper probability bounds on singletons, formally $[Bel(\omega), Pl(\omega)], \omega \in \Omega$, as illustrated in Figure 3, in the geometric style of Cuzzolin. In this case, the bounds obtained for m_1 are:

$$[Bel(\omega_1), Pl(\omega_1)] = [0, 0.9],$$
$$[Bel(\omega_2), Pl(\omega_2)] = [0.1, 1],$$

  and the bounds for m_2 are

$$[Bel(\omega_1), Pl(\omega_1)] = [0.1, 1],$$
$$[Bel(\omega_2), Pl(\omega_2)] = [0, 0.9],$$

  which are in line with our more epistemic view, as m_1 would tend to be of kind ω_2, and m_2 of kind ω_1. Still, these would be quite uncertain statements, as the width of the intervals witnesses.

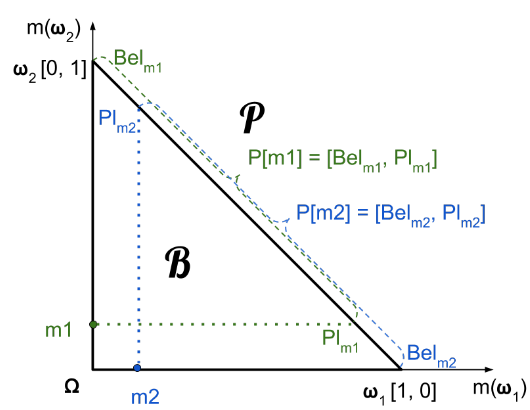

Fig. 3. Projection of m_1 and m_2 from belief space $\mathcal{B}$ to probabilistic space $\mathcal{P}$. Probability intervals are defined by belief function Bel and plausibility function Pl as lower and upper bounds.

The length of the interval represents the uncertainty level of the probability, interpreted from the imprecision in the corresponding evidence corpus. From this representation, it is meaningless to assign m_1, m_2 and m_3 into the same cluster as they are similar at the imprecision level. We can say that the object represented by m_1 (m_2) could be of any class (both ω_1 and ω_2), with a higher possibility to be closer to ω_2 (ω_1). Thus, it is more reasonable to say that m_1 and m_2 could belong to both Group 2 and Group 3, rather than group it with m_3.

The projection from belief space $\mathcal{B}$ to probabilistic space $\mathcal{P}$ gives another illustration of the issue. In this view, the imprecision is kept by probabilistic intervals, fewer assumptions are imposed than in the metric view, and the clustering result is quite consistent with standard interpretation [1], [10], [33] of mass functions.

The main message of this section is that when considering the problem of summarizing or clustering objects, it is crucial to be clear about the interpretation of the structures, and the task we aim to undertake, as what we will consider as acceptable results critically depends on it. Whether the impossibility theorem applies or not highly depend on the chosen interpretation, as illustrated by the comparison between Figure 2 and Figure 3. Therefore, the clarification of the interpretation is essential to reach a practically relevant solution.

VI. Conclusions and perspectives

This paper has studied how compatible combination rules and the computation of centroid were in the problem of building a synthetic mass function from several ones. Our conclusion is that they quickly become incompatible both in theory and practice, and therefore not interchangeable. Our goal was not to propose some new methodologies to combine mass functions or to answer the clustering problem, but rather to clarify some connections between distance and combination. From the aspect of applications, this impossibility theorem can help in explaining the failure of some learning tasks over evidential corpus. It indicates that using a metric-based learning algorithm over evidential corpus primarily concerns our knowledge about an object, and not the object (e.g., to predict, to cluster) itself. After all this, we would like to end this paper by some further discussions.

Centroid vs combination, when use what?: In this paper, we have mainly discussed the fact that using distances or using set-based combination rules to aggregate mass functions are essentially two different things, as well as (in Section V) the fact that one should be careful about what has to be aggregate (knowledge about an object *vs.* the object itself). This means that one should be careful about their purpose when using those tools, as their fitness depends on the context. For instance, while metric-consistency and therefore idempotence is essential in clustering tasks, they are not especially relevant in learning and combining predictive models, where one typically wants reinforcement effects when sources agree.

Beyond impossibility theorem: When summarising the information contained in different mass functions, both centroid computation (through distance minimisation) and combination rules make sense. This paper shows that these two views can quickly become conflicting, as requiring only a few properties leads to our impossibility theorem.

A possibility to obtain more favourable results would be to abandon some of these properties, such as ignorance neutrality. For example, a simple (weighted) average of mass functions is compatible with the minimisation of some specific distances, but clearly violates ignorance neutrality. Disjunctive rules are another kind of rules not satisfying ignorance neutrality, but in their cases, we would be more pessimistic about the fact that they can give the centroid for some distance, given the fact that they are the dual of conjunctive rules.

Some questions as perspectives: While our results show that combination-based and distance-based aggregation are irreconcilable to some extent, the two previous paragraphs indicate that the story has not ended. Indeed, many other questions remain unsolved, to be tackled in further research:

- Simply looking at Equation (11), can we easily characterise those distances (among the many ones existing within TBF) for which the solution is unique and easy to determine? What properties do we need for that? In which situations one is preferable to the other?
- How can we relax the presented properties to turn impossibility into more positive results? We already mentioned the possibility of looking at rules not satisfying ignorance neutrality. However, an alternative could be to relax the requirement made on distances, turning to pseudo distances or non-symmetric divergences (such as the celebrated Kullback-Leibler one).
- Clustering information rather than the objects they refer to may have an interest in itself: to identify different situations regarding the information, *e.g.*, detect those zones with high epistemic uncertainty/imprecision to perform active learning [52]. Also, previous empirical results have shown that in some situation, such as classification, using distances between information about the objects rather than between the objects as a proxy may deliver good results [24]. Yet, the question of how to properly classify/cluster objects (and not the information we have on them) with uncertain distances remains largely open, and is an avenue of research we intend to investigate.

VII. Acknowledgement

We thank GDR IA doctoral visit program, CNRS for their support of the visiting to Heudiasyc, Université de Technologie de Compiègne, where the idea of the paper was initialised.

References

[1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, vol. 38, no. 2, pp. 325–339, 04 1967.

[2] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 447–458, May 1990.

[3] ——, "The normative representation of quantified beliefs by belief functions," *Artificial Intelligence*, vol. 92, no. 1-2, pp. 229–242, 1997.

[4] T. Denœux, "Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence," *Artificial Intelligence*, vol. 172, no. 2-3, pp. 234–264, 2008.

[5] J. Klein, S. Destercke, and O. Colot, "Idempotent conjunctive and disjunctive combination of belief functions by distance minimization," *International Journal of Approximate Reasoning*, vol. 92, pp. 32–48, 2018.

[6] A. Martin, "Conflict management in information fusion with belief functions," in *Information quality in information fusion and decision making*, ser. Information Fusion and Data Science, E. Bossé and G. L. Rogova, Eds., 2019, pp. 79–97.

[7] K. Zhou, A. Martin, and Q. Pan, "Evidence combination for a large number of sources," in *2017 20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–8.

[8] T. Denoeux, N. El Zoghby, V. Cherfaoui, and A. Jouglet, "Optimal object association in the dempster–shafer framework," *IEEE transactions on cybernetics*, vol. 44, no. 12, pp. 2521–2531, 2014.

[9] T. Denoeux, "Distributed combination of belief functions," *Information Fusion*, vol. 65, pp. 179–191, 2021.

[10] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Computational intelligence*, vol. 4, no. 3, pp. 244–264, 1988.

[11] T. Denœux, "The cautious rule of combination for belief functions and some extensions," in *2006 9th International Conference on Information Fusion*. IEEE, 2006, pp. 1–8.

[12] H. Laghmara, T. Laurain, C. Cudel, and J.-P. Lauffenburger, "Heterogeneous sensor data fusion for multiple object association using belief functions," *Information Fusion*, vol. 57, pp. 44–58, 2020.

[13] F. Xiao, "Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy," *Information Fusion*, vol. 46, pp. 23–32, 2019.

[14] A. Martin, "About conflict in the theory of belief functions," in *Belief Functions: Theory and Applications*. Springer, 2012, pp. 161–168.

[15] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, no. 2, pp. 118–145, 2012.

[16] H. E. Stephanou and S.-Y. Lu, "Measuring consensus effectiveness by a generalized entropy criterion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 544–554, 1988.

[17] W. L. Perry and H. E. Stephanou, "Belief function divergence as a classifier," in *Proceedings of the 1991 IEEE International Symposium on Intelligent Control*. IEEE, 1991, pp. 280–285.

[18] S. Konieczny, J. Lang, and P. Marquis, "Distance-based merging: a general framework and some complexity results," *KR*, vol. 2, pp. 97–108, 2002.

[19] S. Benferhat, D. Dubois, S. Kaci, and H. Prade, "Possibilistic merging and distance-based fusion of propositional information," *Annals of Mathematics and Artificial Intelligence*, vol. 34, no. 1, pp. 217–252, 2002.

[20] S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, and P. Marquis, "On the merging of dung's argumentation systems," *Artificial Intelligence*, vol. 171, no. 10-15, pp. 730–753, 2007.

[21] T. Burger, "Geometric views on conflicting mass functions: From distances to angles," *International Journal of Approximate Reasoning*, vol. 70, pp. 36–50, 2016.

[22] F. Cuzzolin, *The Geometry of Uncertainty - The Geometry of Imprecise Probabilities*, ser. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer, 2021.

[23] T. Denœux, O. Kanjanatarakul, and S. Sriboonchitta, "Ek-nnclus: a clustering procedure based on the evidential k-nearest neighbor rule," *Knowledge-Based Systems*, vol. 88, no. C, pp. 57–69, Nov. 2015.

[24] A. Trabelsi, Z. Elouedi, and E. Lefevre, "Decision tree classifiers for evidential attribute values and class labels," *Fuzzy Sets and Systems*, vol. 366, pp. 46–62, 2019.

[25] T. Denœux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE transactions on systems, man, and cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.

[26] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Belief function combination and conflict management," *Information Fusion*, vol. 3, no. 2, pp. 149–162, 2002.

[27] S. B. Hariz, Z. Elouedi, and K. Mellouli, "Clustering approach using belief function theory," in *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*. Springer, 2006, pp. 162–171.

[28] A.-L. Jousselme, D. Grenier, and Éloi Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91 – 101, 2001.

[29] Y. Li, Y. Zhang, D. Wei, and Y. Deng, "Uncertain information clustering based on distance between bpas," in *2012 24th Chinese Control and Decision Conference (CCDC)*, 2012, pp. 3985–3988.

[30] Y. Zhang, T. Bouadi, and A. Martin, "A clustering model for uncertain preferences based on belief functions," in *International Conference on Big Data Analytics and Knowledge Discovery*. Springer, 2018, pp. 111–125.

[31] P. Smets and R. Kennes, "The transferable belief model," *Artificial intelligence*, vol. 66, no. 2, pp. 191–234, 1994.

[32] G. Shafer, *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.

[33] P. Smets, "Belief functions: the disjunctive rule of combination and the generalized bayesian theorem," *International Journal of approximate reasoning*, vol. 9, no. 1, pp. 1–35, 1993.

[34] S. Z. Selim and M. A. Ismail, "K-means-type algorithms: A generalized convergence theorem and characterization of local optimality," *IEEE Transactions on pattern analysis and machine intelligence*, no. 1, pp. 81–87, 1984.

[35] D. Dubois and H. Prade, "A set-theoretic view of belief functions logical operations and approximations by fuzzy sets," *International Journal Of General System*, vol. 12, no. 3, pp. 193–226, 1986.

[36] J. Klein, S. Destercke, and O. Colot, "Interpreting evidential distances by connecting them to partial orders: Application to belief function approximation," *International Journal of Approximate Reasoning*, vol. 71, pp. 15–33, 2016.

[37] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, "The geometric median on riemannian manifolds with application to robust atlas estimation," *NeuroImage*, vol. 45, no. 1, pp. S143–S152, 2009.

[38] P. Bose, A. Maheshwari, and P. Morin, "Fast approximations for sums of distances, clustering and the fermat–weber problem," *Computational Geometry*, vol. 24, no. 3, pp. 135–146, 2003.

[39] M.-H. Masson, S. Destercke, and T. Denœux, "Modelling and predicting partial orders from pairwise belief functions," *Soft Computing*, vol. 20, no. 3, pp. 939–950, 2016.

[40] Y. Zhang, T. Bouadi, and A. Martin, "Preference fusion and condorcet's paradox under uncertainty," in *2017 20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–8.

[41] S. Destercke and D. Dubois, "Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory," *Information Sciences*, vol. 181, no. 18, pp. 3925–3945, 2011.

[42] R. Guyard and V. Cherfaoui, "Study of distributed data fusion using dempster's rule and cautious operator," in *International Conference on Belief Functions*. Springer, 2018, pp. 95–102.

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

[44] P. Smets, "The canonical decomposition of a weighted belief," in *IJCAI*, vol. 95, 1995, pp. 1896–1901.

[45] F. Cuzzolin, "Geometry of dempster's rule of combination," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 2, pp. 961–977, 2004.

[46] I. Couso and D. Dubois, "Statistical reasoning with set-valued information: Ontic vs. epistemic views," *International Journal of Approximate Reasoning*, vol. 55, no. 7, pp. 1502–1518, 2014.

[47] W. Liu, "Analyzing the degree of conflict among belief functions," *Artificial Intelligence*, vol. 170, no. 11, pp. 909–924, 2006.

[48] S. Destercke and T. Burger, "Toward an axiomatic definition of conflict between belief functions," *IEEE transactions on cybernetics*, vol. 43, no. 2, pp. 585–596, 2013.

[49] F. Pichon, A.-L. Jousselme, and N. B. Abdallah, "Several shades of conflict," *Fuzzy Sets and Systems*, vol. 366, pp. 63–84, 2019.

[50] P. Smets, "Decision making in the tbm: the necessity of the pignistic transformation," *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 133–147, 2005.

[51] B. R. Cobb and P. P. Shenoy, "On the plausibility transformation method for translating belief function models to probability models," *International journal of approximate reasoning*, vol. 41, no. 3, pp. 314–330, 2006.

[52] V.-L. Nguyen, S. Destercke, and E. Hüllermeier, "Epistemic uncertainty sampling," in *International Conference on Discovery Science*. Springer, 2019, pp. 72–86.