Information Sciences 756 (2026) 123855

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Using pairwise comparisons to estimate mass assignments in evidence theory

Matteo Brunelli [a,*], Sébastien Destercke [b]

[a] Department of Industrial Engineering, University of Trento, Trento, Italy
[b] Université de Technologie de Compiègne, CNRS UMR 7253, Heudiasyc (Heuristics and Diagnosis of Complex Systems), CS 60319 - 60203, Compiègne Cedex, France

HIGHLIGHTS

- Pairwise comparisons are used to elicit mass assignments.
- Focal elements are compared as evidential hypotheses.
- Incomplete comparisons reduce elicitation burden.
- Expert judgments can be aggregated before deriving masses.
- Simulations support reduced focal-element elicitation.

ARTICLE INFO

Keywords:
Dempster-Shafer theory
Evidence theory
Mass assignment
Pairwise comparisons

ABSTRACT

Evidence Theory (a.k.a. Dempster-Shafer theory) offers a quite general and powerful setting to reason under uncertainty. However, most of the approaches allowing one to instantiate mass functions, the basic building block of the theory, rely on measured, objective data, and there are very few approaches to instantiate them from subjective, expert opinions. Here we suggest a viable solution, based on the concept of pairwise comparisons, for the elicitation of such subjective information. The approach is based on the well-established body of knowledge on the theory of pairwise comparisons and shows its flexibility as it can incorporate multiple representations for the subjective judgments and multiple experts, just to cite two possible extensions.

> Sometimes, on the contrary, I made infinite divisions of each thought and each fact under view, breaking and sectioning them into a vast number of smaller thoughts and facts, easier thus to keep in hand. By this method resolutions difficult to take were broken down into a veritable powder of minute decisions, to be adopted one by one, each leading to the next, and thereby becoming, as it were, easy and inevitable.
>
> Marguerite Yourcenar, *Memoirs of Hadrian*

* Corresponding author.
Email addresses: matteo.brunelli@unitn.it (M. Brunelli), sebastien.destercke@hds.utc.fr (S. Destercke).

https://doi.org/10.1016/j.ins.2026.123855
Received 3 January 2026; Received in revised form 29 June 2026; Accepted 29 June 2026
Available online 30 June 2026

1. Introduction

There are undoubtedly many facets to uncertainty. A common distinction [1] made in the literature is between aleatoric uncertainty, deemed intrinsic to the studied phenomenon, and epistemic uncertainty, originating from a lack of knowledge. The result of a coin flip for a coin that has been tested hundreds of times is an example of a purely aleatoric event and, instead, the uncertainty concerning exact time at which you had dinner yesterday is purely of epistemic nature. While this aleatoric/epistemic distinction may certainly be discussed and/or refined, there is little doubt that the uncertainties in the two mentioned examples have different natures or sources, that may call for formal representations accounting for these differences.

As recalled by Augustin et al. [2], among others, different representations of uncertainty have emerged as generalisations of probability theory able to better account for some aspects of uncertainty such as imprecision or the presence of so-called epistemic uncertainties. Among them, since the seminal contributions by Dempster [3] and Shafer [4], the Dempster-Shafer theory of evidence—hereafter called Evidence Theory—has gained a prominent role. Evidence theory can be seen as a generalization of subjective probability where agents, e.g., sensors or experts, can allocate evidence masses to all subsets of the sample space and not only to its elements representing elementary events. Such allocations, called *mass assignments*, give rise to two dual measures, belief and plausibility. Only when the mass assignment is made on the singletons, do the two measures coincide and become additive. That is, they collapse into a probability measure. On the other hand, when the whole evidence is allocated to a unique set, we recover classical set-valued representations. Evidence theory thus allows one with go from a knowledge state of total imprecision (a set), which one could associate to complete epistemic uncertainty, to a state of precisely knowing the probabilistic process that governs a quantity, which one could associate with complete aleatoric uncertainty.

One of the points that prevent a larger use of evidence theory is that subjective expertise may be required to determine a mass assignment. As we shall see later, there exists a rich quantity of methods to determine mass assignments starting from datasets of objective information, possibly coming from past observations or sensor data. On the contrary, there is a relative lack of proposals to determine mass assignments from purely subjective estimations.

The method of pairwise comparisons, employed in many multi-criteria decision analysis methods [5], has shown for years its effectiveness and versatility in decomposing complex problems into smaller and more tractable subproblems involving pairs of entities, possibly criteria or alternatives. It is worth mentioning that pairwise comparisons have already been employed, for example, to estimate levels of membership of elements in a fuzzy set [6] and subjective probabilities over discrete sample spaces [7]. It is therefore natural to investigate its possible use to extract mass assignments from expert knowledge.

Usually, evidence theory has been used within decision problems to allow decision makers to express epistemic uncertainty in non-probabilistic terms in their judgments on pairs of entities. That is, evidence theory has been used to model pairwise comparisons, as corroborated by the extensions of the Analytic Hierarchy Process [8], the Best-Worst method [9], just to cite a couple of methods. Conversely, in this manuscript we consider the opposite opportunity: using pairwise comparisons to aid the estimation of mass assignments by expert elicitation. More precisely, the paper explores the potential of pairwise comparison techniques and their extension to help experts assess their subjective beliefs, and it can be seen as a *trait d'union* between the two fields of investigation. In essence, we exploit the formal correspondence between weights derived from pairwise comparisons in MCDA and mass assignments over focal elements to propose an elicitation framework in which experts compare potential focal elements pairwise, and the resulting preferences are transformed into a corresponding mass function.

The manuscript is organized as follows. Section 2 discusses related works and the interconnections between evidence theory and pairwise comparisons to further motivate this manuscript. Section 3 includes, by means of two self-contained subsections, more technical preliminary notions on the two key elements: evidence theory and pairwise comparisons. Next, Section 4 shows, also by means of examples, how the pairwise comparison method can be applied to the problem at stake. Issues like consistency, uncertainty, and incompleteness will be considered and some original results will be presented for the case in which multiple experts express their opinions. Section 5 presents a numerical experiment to test the capacity of pairwise comparisons to capture partial knowledge, also considering possible inconsistencies. Finally, Section 6 concludes the manuscript with a discussion and some conclusions.

2. Related works

Our proposal consists of using well-known multi-criteria decision analysis (MCDA) tools to construct mass assignments from expert opinions. As such, the immediately related works are those that propose building mass assignments from available information, and those that connect MCDA approaches with belief functions.

2.1. Mass assignment construction

Most of the approaches presented in the literature that intend to construct mass assignments consider data to do so. This is not surprising, as Dempster's initial proposal [3] was to use belief functions to model imprecisely observed random variables. This is why many proposals consider statistical tools to build mass assignments, such as the extension of the likelihood principle [10].

Beyond statistics, both supervised and unsupervised machine learning are the two fields where most proposals to construct mass assignments have been made. Without going into the details of the dozens if not hundreds of methods proposed to do so, one can mention for example the seminal E-KNN and its extensions, rule-based reasoning approaches, or more recently deep learning-based approaches.

All these approaches have in common that they rely on data to identify mass assignments, and do not apply to the case where a human expert has to be consulted to build a mass assignment through an elicitation process. However, while many uncertainty-based decision problems require considering subjective opinions [11], there is a striking lack of methods designed to support the subjective elicitation of mass assignments from expert judgment. There are only a few works dealing explicitly or implicitly with belief function elicitation. Let us detail them a bit more:

- A first trend [12] consists in considering qualitative assessments of the kind $A \succ B$ or $A \sim B$ and transforming them into constraints over the belief values, for instance mapping $A \succ B$ into $\mathrm{Bel}(A) > \mathrm{Bel}(B)$, and then finding a unique belief function that respects those constraints as much as possible and minimizes a suitable functional (typically an uncertainty quantification tool). Note that this induces many subjective choices, such as the choice of the functional and the translation of $A \succ B$ into corresponding constraints;
- A second trend consists in asking for numerical assessments, often confined to specific rather than general mass functions. In the case of possibility distributions, we can mention the work of Thierry et al. [13], which built expert-based data sets with uncertain labels, where Likert-scales are used to build mass assignments over nested sets, and Sandri et al. [14] introduced general numerical elicitation strategies for possibilities. Likewise, imprecise cumulative distributions, which are special cases of belief functions [15], are often used to model numerical elicitation in risk analysis.

In contrast, our approach relies on the idea of comparing the amount of evidence pointing to subsets, rather than comparing measures. It considers quantitative comparative statements that can be applied to any mass function, therefore mixing the advantages of qualitative and quantitative statements.

This lack of methods to handle expert opinions and human elicitation contrasts strongly with the field of preference modeling and MCDA [16], which benefits from decades of research on the problem of eliciting numerical preference models from experts. It therefore seems natural to investigate how such methods could help in obtaining mass assignments from experts.

2.2. MCDA and belief functions

As just mentioned, belief functions and MCDA are a natural mix to consider, yet when looking at the literature at the cross-road between belief functions and preferences, it quickly becomes obvious that this is a rather one-way street, as most of the works intend to include belief functions within the MCDA process to reflect uncertainties in the decision-maker's opinion.

Although it is difficult to trace the inception of evidential reasoning within the multi-criteria decision analysis community, it is safe to consider the approach of Yang and Singh [17] as one of the precursors, as it generated a number of follow-up studies, where belief measures were used to represent uncertainty of entries of the decision matrix and of the weights of individual criteria. Later on, the use of belief functions has been considered with different MCDA approaches such as BWM [9] or AHP [8]. We can also mention some other related works exploring how belief functions can help in tasks such as preference aggregation [18] or incremental preference construction [19], but those also explore how belief functions can help in preference handling.

As for data-induced mass assignments, making a full review of the dozens of works in this direction would necessitate a review paper of its own, and is not within the scope of the present study, which aims at using MCDA to solve issues within evidence theory, and not the other way around. We will therefore abstain from doing such a review.

3. Preliminaries

Hereafter, we shall use $\mathbb{R}_>$ and $\mathbb{R}_\geq$ to denote the set of all positive real numbers and the set of non-negative real numbers, respectively.

3.1. Evidence theory

In basic evidence theory, the *frame of discernment*, here denoted $\Theta = \{\theta_1 \dots, \theta_p\}$, is a finite non-empty set that plays a similar role as the sample or outcome space in probability theory. If we call 2^Θ the power set of Θ, then a *mass assignment* (MA), also called basic probability assignment, on Θ is a function $m : 2^\Theta \to [0,1]$ such that $m(\emptyset) = 0$ and $\sum_{A\subseteq\Theta} m(A) = 1$. It can be interpreted as an assignment of strengths of evidence to subsets $A \subseteq \Theta$. The set of *focal elements* is the set of all subsets of Θ with a positive mass, i.e., $\mathcal{F} = \{A | A \subseteq \Theta,\ m(A) > 0\}$. In practical applications, focal elements could be subsets of the frame of discernment whose elements share a common feature. For example, if Θ is a set of possible diseases, then $A \subset \Theta$ could be the set of all infectious ones. Two measures are induced by an MA. The first one, called *belief*, is a superadditive measure defined as

$$\mathrm{Bel}(A) = \sum_{B\subseteq A} m(B) \qquad \forall A \subseteq \Theta,$$

and quantifies the amount of evidence directly supporting A. The second measure, subadditive, is called *plausibility* and is defined as

$$\mathrm{Pl}(A) = \sum_{B|B\cap A\neq\emptyset} m(B) \qquad \forall A \subseteq \Theta,$$

and is the sum of the evidence that does not disprove A. If the constraint $m(\emptyset) = 0$ is satisfied, there always exists a probability measure Pr such that

$$\mathrm{Bel}(A) \leq \Pr(A) \leq \mathrm{Pl}(A) \qquad \forall A \subseteq \Theta.$$

For this reason, Bel and Pl have often been interpreted as lower and upper bounds for a "true" probability, respectively. The set of all probability measures compatible with belief and plausibility is called a *credal set*. Together, the two measures are dual, i.e.,

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\overline{A}) \qquad \forall A \subseteq \Theta.$$

where $\overline{A}$ is the complement of A with respect to Θ. Some special cases are often considered. Note that when the set of focal elements corresponds to the elements of the frame of discernment Θ taken as singletons, i.e., $\mathcal{F} = \{\{\theta_1\}, \dots, \{\theta_n\}\}$, then $\mathrm{Bel}(A) = \mathrm{Pl}(A) = \mathrm{Pr}(A)$ for any event A. This case is called *Bayesian* MA and represents the case where evidence theory collapses into subjective probability. If $\mathcal{F}$ is a singleton, we say that the MA is *deterministic*, and among all the deterministic MAs we call *vacuous* the case in which $\mathcal{F} = \{\Theta\}$ and represents total ignorance. On the contrary, $\mathcal{F} = \{\theta_i\}$ represents full certainty on the elementary event θ_i. The set $\mathcal{F}$ is *nested* if there exists a permutation π that orders the focal elements such that $A_{\pi(1)} \subset \cdots \subset A_{\pi(|\mathcal{F}|)}$. If $\mathcal{F}$ is nested, then Pl becomes a *possibility measure*, i.e., $\mathrm{Pl}(A \cup B) = \max\{\mathrm{Pl}(A), \mathrm{Pl}(B)\}$ for all $A, B \subseteq \Theta$.

An uncertainty index[1] is a function $\mathsf{U} : m \mapsto \mathsf{U}(m) \in \mathbb{R}$ with the value $\mathsf{U}(m)$ serving as an estimation of the uncertainty of the MA. The deterministic MA on a singleton has the lowest level of uncertainty and it is common to consider the vacuous MA as the most uncertain. Most proposals in the literature attempt to measure two different aspects of an MA: on the one hand its imprecision (a.k.a. "epistemic uncertainty"), and on the other hand its degree of ambiguity (a.k.a. "aleatoric uncertainty"), notably by proposing extensions of, respectively, the Hartley measure for sets and the Shannon entropy for probabilities. For instance, regarding the Hartley measure that intends to quantify the so-called epistemic uncertainty, one can use the formula

$$\mathsf{U}_D(m) = \sum_{A \in \mathcal{F}} m(A) \log_2 |A|$$

proposed and justified by Dubois and Prade [20]. Similarly, Klir and Ramer [21] proposed and justified the following formula extending Shannon entropy

$$\mathsf{U}_{KR}(m) = -\sum_{A \in \mathcal{F}} m(A) \log_2 \sum_{E \cap A \neq \emptyset} m(E) \frac{|E \cap A|}{|E|}$$

where the term $\sum_{E \cap A \neq \emptyset} m(E) \frac{|E \cap A|}{|E|}$ is nothing else but the so-called pignistic probability.

The value obtained by means of an uncertainty index can be used to decide whether an MA is too uncertain, and the expert can be further inquired to refine it to make it less uncertain or more reliable. Indeed, too vague and inconclusive evidence has often been related to the reliability of evidence. Note that there are two driving forces determining the value of $\mathsf{U}(m)$: the choice of the set of focal elements $\mathcal{F}$ and the assignment of positive masses to its elements $A \in \mathcal{F}$.

3.2. Pairwise comparisons

Given a finite non-empty set of entities $X = \{x_1, \dots, x_n\}$, pairwise comparisons exploit the divide and conquer logic and capture subjective statements over pairs of entities. This is done with the goal of decomposing a complex problem into smaller and more tractable ones. However, the final result that one expects from a set of pairwise comparisons is often a positive normalized vector

$$\mathbf{w} = (w_1, \dots, w_n) \tag{1}$$

that best fits the set of pairwise comparisons. In this sense, components of $\mathbf{w}$ can be interpreted as slices into which a pie of unit area should be divided.

In most MCDA methods, weights constitute a fundamental input used to properly balance the influence of the different attributes describing the alternatives. Consequently, identifying appropriate weights represents a critical step in the decision-making process. Methods for determining weights are generally classified into three main categories: subjective, objective, and hybrid approaches. *Subjective methods* [22] are based on the assumption that weights should reflect the judgments, preferences, and expertise of decision-makers or specialists. Conversely, *objective methods* do not incorporate individual opinions; instead, they derive weights solely from the intrinsic characteristics or performance data of the alternatives. In this work, we focus specifically on subjective methods.

Preference relations are the mathematical structures underlying (subjective) pairwise comparisons. Many definitions of (valued) preference relations have been given in the literature [23]. Here we consider the possibility that two entities may be uncomparable and define a preference relation as a mapping

$$P : X \times X \to D \cup \{*\} \tag{2}$$

where D is the domain of representation of preferences and "$*$" is an additional element denoting incomparability. We say that a preference relation is *complete* if no element of $X \times X$ is mapped to "$*$". Otherwise, we say that the relation is *incomplete*. Note that here, incomplete means incompletely specified, in the sense that some comparisons are not made or measured. This differs from intrinsic, ontic incomparability.

The most widely used representation of preferences is the multiplicative one, where $D = \mathbb{R}_{>}$. It is used among others in the well-known AHP [24] and in Multi-attribute Value Theory. In this case, preferences can be collected into a (positive) *pairwise comparison*

[1] They are usually called uncertainty measures but here we call them indices or indicators to avoid ambiguity with measures (from the measure theoretic point of view) like Pr, Pl, and Bel.

matrix $\mathbf{A} = (a_{ij})_{n\times n}$ with $a_{ii} = 1$ for all i and $a_{ji} = 1/a_{ij}$ for all i, j. The idea is that each comparison is an estimation of the ratio between the weights of the compared entities, i.e., $a_{ij} \approx w_i/w_j$. If equality holds, we say that the pairwise comparison matrix is *consistent*. Equivalently, a pairwise comparison matrix is consistent if and only if

$$a_{ik} = a_{ij}a_{jk} \quad \forall i, j, k. \tag{3}$$

That is, when an expert is fully rational, each direct comparison a_{ik} is supported by indirectly comparing x_i and x_k passing through x_j. An inconsistency index is a function $\mathrm{I} : \mathbf{A} \mapsto \mathrm{I}(\mathbf{A}) \in \mathbb{R}$ where $\mathrm{I}(\mathbf{A})$ is an estimation of the inconsistency of preference contained in $\mathbf{A}$. The most widely used index is probably the CI proposed by Saaty [24] defined as

$$\mathsf{CI}(\mathbf{A}) = \frac{\lambda_{\max} - n}{n-1}$$

where $\lambda_{\max}$ is the Perron-Frobenius eigenvalue of $\mathbf{A}$. Another commonly used inconsistency index is $\mathsf{CR}(\mathbf{A}) = \mathsf{CI}(\mathbf{A})/RI$ where RI is the average inconsistency computed on a very large number of pairwise comparison matrices.

Note that the definition of consistency (3) applies to complete matrices, but can be extended to incomplete ones (based on incomplete preference relations) by considering only the triples (a_{ij}, a_{jk}, a_{ik}) for which the elements are numerical. Inconsistency indices have also been extended to work on incomplete matrices [25].

Each preference relation, and hence also each pairwise comparison matrix $\mathbf{A}$, has an *underlying graph*. In the case of a pairwise comparison matrix we define it as the pair $\mathbf{A} = (X, E)$ where X is the set of vertices, represented by the entities that are compared, and E, the set of edges, is the set of pairs of indices of distinct entities on which the expert expressed her judgments, i.e., $E = \{\{i, j\} \mid x_i, x_j \in X,\ i < j,\ a_{ij} \in \mathbb{R}_>\}$.

The number of pairwise comparisons grows quadratically with the number of focal elements when complete comparison matrices are elicited. This may impose a cognitive burden for large sets of focal elements. A common mitigation strategy is to use incomplete pairwise comparison matrices, where only a subset of comparisons is provided and the remaining values are inferred during the estimation procedure. When the comparison graph is connected, the number of required judgments can be drastically reduced from a quadratic growth to a linear one. In particular, the minimum number of comparisons required is $(n-1)$, while at least n comparisons are typically recommended to enable consistency checks [26]. In practice, the elicitation burden can be further reduced by restricting the number of focal elements, which is a common modeling strategy in belief function applications.

Prioritization is the process that takes a pairwise comparison matrix $\mathbf{A}$ as input and yields a vector $\mathbf{w}$ as output. The best-known method is probably the eigenvector method [24], according to which one should take the Perron-Frobenius eigenvector to obtain a good estimation of the weights. An alternative method is the geometric mean method [27], according to which an estimation of weights can be obtained as

$$w_i = \left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}} \tag{4}$$

If the weight vector needs to be normalized, then it is straightforward to apply the normalization to both methods. Furthermore, the eigenvector and the geometric mean methods were proposed for the basic case in which preferences cannot be incomplete, but they have been extended to incomplete matrices by Harker [28] and Bozóki et al. [29], respectively. For both, the only requirement is the connectedness of the underlying graph.

An extension of pairwise comparison matrices regards the case where *multiple experts*, say q, provide their preferences by means of matrices $\mathbf{A}_1, \dots, \mathbf{A}_q$ and the goal of reaching one final vector $\mathbf{w}$ remains. To solve this problem, Aczél and Saaty [30] stipulated some reasonable properties for the aggregation of ratio (multiplicative) judgments and showed that the weighted geometric mean method is the only aggregation method that satisfies them. Consequently, in the case of complete pairwise comparison matrices associated with q experts with positive normalized importance weights $\omega_1, \dots, \omega_q$, they proposed the following formula:

$$a_{ij}^{G} = \prod_{h=1}^{q} \left(a_{ij}^{(h)}\right)^{\omega_h} \tag{5}$$

where $a_{ij}^{(h)}$ is the (i, j)th entry of the matrix provided by the hth expert. Such aggregation method has been popularized by Forman and Peniwati [31] under the name of aggregation of individual judgments (AIJ). In some contexts, it may be desirable to quantify the degree of compatibility between two complete pairwise comparison matrices, say $\mathbf{A}_1 = (a_{ij}^{(1)})_{n\times n}$ and $\mathbf{A}_2 = (a_{ij}^{(2)})_{n\times n}$. In the following we will use

$$\mathsf{C}(\mathbf{A}_1, \mathbf{A}_2) = \frac{2}{n(n-1)} \sum_{i<j}^{n} \ln^2 \left(a_{ij}^{(1)} a_{ji}^{(2)}\right) \tag{6}$$

A comparative analysis of compatibility indices was proposed by Ágoston et al. [32]. Updated expositions of the pairwise comparison method and its extensions were presented for instance by Mazurek [33].

4. Pairwise comparisons for the estimation of mass functions

We are now ready to consider a basic use of pairwise comparisons on a set of focal elements. In this basic case an expert decides, a priori, on what subsets $A \subseteq \Theta$ she feels confident to express her pairwise opinions. Such sets become focal elements in $\mathcal{F}$. Then, a preference relation is built so that they can be pairwise compared. That is, the reference set X in the preference relation structure (2) is replaced by $\mathcal{F}$ and the preference relation becomes

$$P : \mathcal{F} \times \mathcal{F} \to D \cup \{*\} \tag{7}$$

Note that, in its multiplicative representation by means of pairwise comparisons, the comparison of two subsets A_i, A_j is represented by $a_{ij} > 0$ and could be interpreted as "the evidence pointing to A_i is a_{ij} times as much as the one pointing to A_j". Note that even if $A_j \subseteq A_i$, such a statement could simply mean that the evidence allows narrowing down elements in A_i, but not further due to, e.g., ambiguity or coarseness of the measured quantity. One could be, for instance, convinced that a blurry picture points to a cat or a dog, without necessarily being able to distinguish between the two.

It is worth noting that expressing numerical values directly to compare evidential strength can be challenging for experts. For this reason, several transformations linking verbal assessments to numerical scales have been proposed in the literature. More recent studies have analyzed and compared these transformations, providing empirical support for their validity [34].

Once subjective judgments on some pairs of focal elements are collected into a pairwise comparison matrix we can derive a priority vector, whose components can be interpreted as MAs. Namely, we consider

$$m(A_i) := w_i\,, \tag{8}$$

where w_i is ith component of the normalized eigenvector of $\mathbf{A}$ or the result of the geometric mean method (4) after normalization. Note that, depending on whether we think there exists a true underlying mass or it is a subjective entity, the mass resulting from (8) can either be seen as an approximation, or as a mass representing the subjective opinion of a user. In Section 5 we will see that, if we assume the existence of a true mass, then focusing on a handful of well-chosen focal elements can quickly provide a good approximation. Moreover, a subset A that is not in $\mathcal{F}$ will receive zero mass. In practice, however, not including a subset in the set of focal elements that are compared does not entail that its mass must be null, but more pragmatically that it is sufficiently small to be considered negligible. Methods for generating and structuring alternatives are central in decision analysis; by the same token, the selection of focal elements plays a fundamental role when constructing belief functions. While the ultimate choice should remain with the expert, it is natural to select focal elements that correspond to meaningful and interpretable hypotheses. For example, in medical diagnosis, diseases may be grouped into subsets according to clinically relevant categories such as contagious, sexually transmitted, chronic, cardiovascular, airborne, or genetic conditions. Such domain-driven groupings provide semantically coherent focal elements, especially when some structure (e.g., hierarchical or conceptual) exists over Θ, as in hierarchical classification settings. In addition to expert judgment, the initial selection of focal elements may be supported by optimization-based procedures aimed at satisfying structural desiderata—for instance, balancing the frequency with which atomic events θ_i appear across focal elements [35]. This perspective is consistent with existing approaches that restrict focal sets for modeling or predictive purposes, such as singleton-based belief functions in machine learning and evidential clustering methods, as well as recent findings highlighting the benefits of limiting the number of focal elements. To be general, we do not impose any extra coherence constraint on our set of considered focal elements, but one can easily impose them if needed. Note that such constraints would only limit the possible focal elements to pick from, and not the pairwise assessments between those focal elements.

The next example shows an application of preference relations to determine an MA using a pairwise comparison matrix $\mathbf{A}$ and $D = \mathbb{R}_{>}$.

Example 1. Consider a frame of discernment of 4 possible alternative diseases $\Theta = \{\theta_1, \dots, \theta_4\}$ and a set of 5 focal elements $\mathcal{F} = \{A_1, \dots, A_5\}$ defined as

$$A_1 = \{\theta_1, \theta_2\}$$
$$A_2 = \{\theta_3, \theta_4\}$$
$$A_3 = \{\theta_1\}$$
$$A_4 = \{\theta_2, \theta_3, \theta_4\}$$
$$A_5 = \{\theta_1, \theta_2, \theta_3, \theta_4\} = \Theta$$

A medical doctor pairwise compares, using her expertise, the values of the MAs of such subsets using the following pairwise comparison matrix:

$$\mathbf{A} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{array} \begin{array}{c} \begin{array}{ccccc} A_1 & A_2 & A_3 & A_4 & A_5 \end{array} \\ \begin{pmatrix} 1 & 2 & * & 1/2 & * \\ 1/2 & 1 & * & * & 3 \\ * & * & 1 & 1/5 & * \\ 2 & * & 5 & 1 & 1/3 \\ * & 1/3 & * & 3 & 1 \end{pmatrix} \end{array}$$

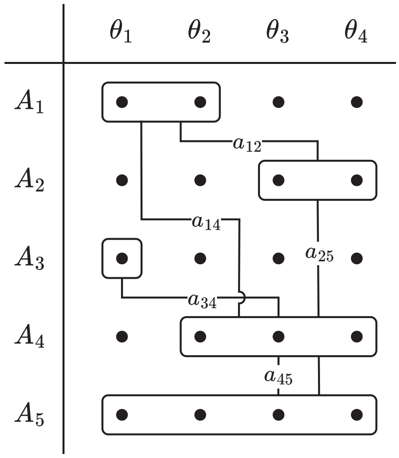

Fig. 1. The structure of the preferences expressed in $\mathbf{A}$ on the set of focal elements $\mathcal{F} = \{A_1, \dots, A_5\}$.

The structure of the preferences in matrix $\mathbf{A}$ is represented in Fig. 1. If we apply Harker's method [28] we obtain[2]

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{pmatrix} = \begin{pmatrix} m(A_1) \\ m(A_2) \\ m(A_3) \\ m(A_4) \\ m(A_5) \end{pmatrix} = \begin{pmatrix} m(\{x_1, x_2\}) \\ m(\{x_3, x_4\}) \\ m(\{x_1\}) \\ m(\{x_2, x_3, x_4\}) \\ m(\Theta) \end{pmatrix} = \begin{pmatrix} 0.2539 \\ 0.3040 \\ 0.0209 \\ 0.1841 \\ 0.2371 \end{pmatrix}.$$

It is worth clarifying the semantic interpretation of these comparisons. Unlike AHP, where pairwise comparisons typically involve mutually exclusive alternatives, in the present framework the objects of comparison are focal elements, which may represent composite hypotheses. Such comparisons are nevertheless meaningful: in decision theory and probabilistic preference models, experts routinely compare evidential support assigned to events rather than to single alternatives. A focal element can thus be regarded as a conceptual hypothesis (e.g., a category such as "infectious diseases" versus "genetic diseases"), and experts are asked to assess the relative strength of evidence supporting these hypotheses as single entities, without explicitly decomposing them into their atomic components.

A clarification is in order regarding overlapping or nested focal elements. Pairwise comparisons in the proposed approach concern the *strength of evidential support* assigned to hypotheses, rather than their set-theoretic relations. For example, let $A = \{\theta_1, \theta_2\}$ denote the hypothesis "the culprit is male" and $B = \{\theta_2, \theta_3\}$ the hypothesis "the culprit has dark hair." Although A and B overlap in θ_2, the evidence supporting them may stem from different sources (e.g., a witness statement versus forensic analysis). We could even argue that such ideas of independent sources of information are at the basis of distinctness Smets and Kennes [36]. An expert may therefore compare the overall support for A and B as conceptual hypotheses, despite their non-disjointness. It should also be noted that, in contrast with elicitation in MCDA approaches that focus on distinct alternatives, elicitation of uncertainty typically involves non-disjoint events and sets, as it is common to compare likelihood of A and B with $A \cap B \neq \emptyset$.

4.1. Multiple-experts

Now we consider the existence of multiple experts who can be interpreted as distinct sources of evidence. This relates to the problem of combination of evidence in Dempster-Shafer theory, on which much research has been carried out with the Dempster rule of combination still playing a dominant role. Given two MAs m_1 and m_2, they can be combined as follows [4, p. 60]:

$$m(A) = \frac{\displaystyle\sum_{\substack{A_i, A_j \\ A_i \cap A_j = A}} m_1(A_i) m_2(A_j)}{1 - \displaystyle\sum_{\substack{A_i, A_j \\ A_i \cap A_j = \emptyset}} m_1(A_i) m_2(A_j)} \tag{9}$$

[2] Harker's method considers an incomplete pairwise comparison matrix $\mathbf{A}$ and transforms it into a complete auxiliary matrix $\mathbf{C}$, using the rule

$$c_{ij} = \begin{cases} a_{ij} & \text{if } i \neq j \text{ and } a_{ij} \in \mathbb{R} \\ 0 & \text{if } i \neq j \text{ and } a_{ij} = \text{"}*\text{"} \\ 1 + m_i & \text{if } i = j \end{cases}$$

where m_i is the number of missing entries on the ith row of $\mathbf{A}$. Then, the eigenvector method can be applied to $\mathbf{C}$.

Let us note that, if $(\cup\mathcal{F}_1)\cap(\cup\mathcal{F}_2)=\emptyset$, then m_1 and m_2 are not *combinable* through the Dempster rule of combination.

We shall show how pairwise comparison matrices can be used to aggregate elicited judgments from multiple experts before deriving a collective mass assignment. First, we call $d_1,\dots,d_q$ the decision makers/experts who are going to express their pairwise judgments. Note that their sets of focal elements $\mathcal{F}_1,\dots,\mathcal{F}_q$ need not be equal as well as their relative weights in the decision process, which we denote as $\omega_1\dots,\omega_q$ with $\omega_h>0$ and $\omega_1+\cdots+\omega_q=1$. We call S_{ij} the set of experts who expressed numerical judgments on the two sets $A_i,A_j\in\mathcal{F}_G$ where $\mathcal{F}_G$ is the union of the sets of focal elements, i.e., $\mathcal{F}_G=\bigcup_{h=1}^q\mathcal{F}_h$. Then, the aggregate preference relation $\mathbf{A}_G:\mathcal{F}_G\times\mathcal{F}_G\to D\cup\{*\}$ can be represented by a pairwise comparison matrix $\mathbf{A}_G=\left(a_{ij}^G\right)_{p\times p}$ with entries defined as

$$a_{ij}^G:=\begin{cases}\prod_{h\in S_{ij}}\left(a_{ij}^{(h)}\right)^{\frac{\omega_h}{\sum_{k\in S_{ij}}\omega_k}}, & \text{if } S_{ij}\neq\emptyset\\ "*", & \text{if } S_{ij}=\emptyset\end{cases}\tag{10}$$

where i,j are the indices of the sets A_i and A_j in $\mathcal{F}_G$.

Interestingly, if we use (10), then under a mild condition the connectedness of $\mathbf{A}_1,\dots,\mathbf{A}_q$ implies the connectedness of $\mathbf{A}_G$.

Proposition 1. *Consider q pairwise comparison matrices $\mathbf{A}_1,\dots,\mathbf{A}_q$, whose underlying graphs are connected on their domains $\mathcal{F}_1,\dots,\mathcal{F}_q$. If there exists a permutation π on $\{1,\dots,q\}$ such that $\mathcal{F}_{\pi(h)}\cap\mathcal{F}_{\pi(h+1)}\neq\emptyset$ for all $h=1,\dots,q-1$, then the graph underlying $\mathbf{A}_G$ is connected too.*

Proof. By definition of a_{ij}^G, i.e., (10), and the connectedness of all the graphs underlying $\mathbf{A}_h$, we know that the graph underlying $\mathbf{A}_G$ is the union of the graphs underlying $\mathbf{A}_1,\dots,\mathbf{A}_q$. Hence, $\{i,j\}$ is an edge of $\mathbf{A}_G$ if and only if it is an edge for at least one $\mathbf{A}_h$. As each underlying graph is associated with a set of vertices $\mathcal{F}_h$ that is connected, we know that we can go from each of its vertices to any other of its vertices. Since $\mathcal{F}_{\pi(h)}\cap\mathcal{F}_{\pi(h+1)}\neq\emptyset$, then we know that, if we consider the union of the two graphs, we can go from any vertex in $\mathcal{F}_{\pi(h)}\cup\mathcal{F}_{\pi(h+1)}$ to any other vertex in the same set. Hence, we know that the union of all the underlying graphs of the pairwise comparison matrices with vertex set $\bigcup_{h=1}^q\mathcal{F}_h$ is also connected, and so must be the graph underlying $\mathbf{A}_G$. □

Remark 1. Note that the implication in Proposition 1 is not true the other way around. Namely, the union of disconnected graphs could be connected. This can be seen as a strong point of this approach as connectedness of individual underlying graphs is not necessary to guarantee the connectedness of the group one.

Given that the sets of focal elements of different experts may be different, we need to generalize the concept of comparability index. If we index the pairwise comparisons with respect to the focal elements that they compare, then we can define the following:

Definition 1. Given two pairwise comparison matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ with underlying graphs $(\mathcal{F}_1,E_1)$ and $(\mathcal{F}_2,E_2)$, we define their compatibility as

$$\tilde{\mathsf{C}}(\mathbf{A}_1,\mathbf{A}_2)=\frac{1}{|E_1\cap E_2|}\sum_{\{A_i,A_j\}\in E_1\cap E_2}\ln^2\left(a_{A_i,A_j}^{(1)}a_{A_j,A_i}^{(2)}\right)$$

Proposition 2. *If $\mathcal{F}_1=\mathcal{F}_2$, and the two pairwise comparison matrices are complete, then $\tilde{\mathsf{C}}(\mathbf{A}_1,\mathbf{A}_2)$ collapses into $\mathsf{C}(\mathbf{A}_1,\mathbf{A}_2)$, as defined in (6).*

Proof. Considering the premise of the proposition we can map the elements of $\mathcal{F}_1$ into the first $|\mathcal{F}_1|$ positive integers and rewrite

$$\sum_{\{A_i,A_j\}\in E_1\cap E_2}\ln^2\left(a_{A_i,A_j}^{(1)}a_{A_j,A_i}^{(2)}\right)=\sum_{1\le i<j\le|\mathcal{F}_1|}\ln^2\left(a_{ij}^{(1)}a_{ji}^{(2)}\right)$$

A complete graph has $\frac{n(n-1)}{2}$ edges, and therefore

$$\frac{1}{|E_1\cap E_2|}=\frac{1}{\frac{n(n-1)}{2}}=\frac{2}{n(n-1)},$$

which shows that C is a special case of $\tilde{\mathsf{C}}$. □

Example 2. Building upon Example 1, we now consider the addition of a medical doctor acting as a supplementary expert and assume that their weights in the decision should be equal. That is $\omega_1=\omega_2=1/2$. Unlike the previous medical doctor (d_1) the supplementary doctor (d_2) considers the set $A_6=\{\theta_3\}$ in his set of focal elements, instead of A_5. In this case their preferences can be collected into pairwise comparison matrices representing preferences on the set $\mathcal{F}_G=\mathcal{F}_1\cup\mathcal{F}_2$. We have,

$$\mathbf{A}_1=\begin{array}{c|c} & \begin{matrix}A_1 & A_2 & A_3 & A_4 & A_5\end{matrix}\\ \begin{matrix}A_1\\A_2\\A_3\\A_4\\A_5\end{matrix} & \begin{pmatrix}1 & \mathbf{2} & * & \mathbf{1/2} & *\\ 1/2 & 1 & * & * & 3\\ * & * & 1 & \mathbf{1/5} & *\\ 2 & * & 5 & 1 & 1/3\\ * & 1/3 & * & 3 & 1\end{pmatrix}\end{array}\qquad \mathbf{A}_2=\begin{array}{c|c} & \begin{matrix}A_1 & A_2 & A_3 & A_4 & A_6\end{matrix}\\ \begin{matrix}A_1\\A_2\\A_3\\A_4\\A_6\end{matrix} & \begin{pmatrix}1 & \mathbf{3} & * & \mathbf{1} & 1/4\\ 1/3 & 1 & * & 1/3 & *\\ * & * & 1 & \mathbf{1/4} & *\\ 1 & 1/3 & 4 & 1 & 1/3\\ 4 & * & * & 3 & 1\end{pmatrix}\end{array}$$

where the boldfaced upper triangular comparisons are those that were elicited by both experts. The compatibility between the two pairwise comparison matrices can be calculated as follows:

$$\tilde{\mathrm{C}}(\mathbf{A}_1, \mathbf{A}_2) = \frac{1}{3}\left[\ln^2\left(2 \cdot \frac{1}{3}\right) + \ln^2\left(\frac{1}{2} \cdot 1\right) + \ln^2\left(\frac{1}{5} \cdot 4\right)\right] = 0.231$$

We use (10) to define the entries of $\mathbf{A}_G$ and we obtain

$$\mathbf{A}_G = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \end{array} \begin{array}{c} \begin{array}{cccccc} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 \end{array} \\ \begin{pmatrix} 1 & 2^{\frac{1}{2}}3^{\frac{1}{2}} & * & \frac{1}{2}^{\frac{1}{2}}1^{\frac{1}{2}} & * & 1/4 \\ \frac{1}{2}^{\frac{1}{2}}\frac{1}{3}^{\frac{1}{2}} & 1 & * & 1/3 & 3 & * \\ * & * & 1 & \frac{1}{5}^{\frac{1}{2}}\frac{1}{4}^{\frac{1}{2}} & * & * \\ 2^{\frac{1}{2}}1^{\frac{1}{2}} & 3 & 5^{\frac{1}{2}}4^{\frac{1}{2}} & 1 & 1/3 & 1/3 \\ * & 1/3 & * & 3 & 1 & * \\ 4 & * & * & 3 & * & 1 \end{pmatrix} \end{array}$$

representing a preference relation $P : \mathcal{F}_G \times \mathcal{F}_G \to \mathbb{R}_> \cup \{*\}$ with a connected underlying graph. The final assessment for the mass function can be obtained using Harker's method and is

$$\mathbf{w} = \begin{pmatrix} m(A_1) \\ m(A_2) \\ m(A_3) \\ m(A_4) \\ m(A_5) \\ m(A_6) \end{pmatrix} = \begin{pmatrix} m(\{\theta_1, \theta_2\}) \\ m(\{\theta_3, \theta_4\}) \\ m(\{\theta_1\}) \\ m(\{\theta_2, \theta_3, \theta_4\}) \\ m(\Theta) \\ m(\{\theta_3\}) \end{pmatrix} = \begin{pmatrix} 0.228 \\ 0.238 \\ 0.043 \\ 0.190 \\ 0.245 \\ 0.056 \end{pmatrix}$$

It appears that the approach outlined in this section first *aggregates* the preferences and only afterwards it *derives* a collective MA. One may wonder under what conditions the two steps of the process are commutative and—if this is the case—what kind of aggregation is applied to the MAs of individual decision makers. Let us consider a set of complete pairwise comparison matrices $\mathbf{A}_1, \dots, \mathbf{A}_q$ defined on the same set of focal elements $\mathcal{F}$, and assume that we use the geometric mean method to extract weights from pairwise comparison matrices. Then, if we discard the cosmetic operation of normalization, the process outlined in this section can be summarized as the left hand side of

$$\underbrace{\prod_{j=1}^{n}\left(\underbrace{\left(\prod_{h=1}^{q}\left(a_{ij}^{(h)}\right)^{\omega_h}\right)}_{\text{eq. (5)}}\right)^{\frac{1}{n}}}_{\text{eq. (4)}} = \underbrace{\prod_{h=1}^{q}\left(\underbrace{\left(\prod_{j=1}^{n} a_{ij}^{(h)}\right)^{\frac{1}{n}}}_{\text{eq. (4)}}\right)^{\omega_h}}_{\text{eq. (5)}} \tag{11}$$

where the aggregation of individual judgments is the internal function and elicitation of priorities is external. However, the two operations commute as shown on the right hand side where a weight vector (in our context corresponding to a MA) is obtained for each expert, and then these are aggregated using the weighted geometric mean. It provides a natural way to avoid the question of whether expert opinions should be first aggregated then transformed, or first transformed then aggregated. Lastly, we considered Zadeh's paradox and tested our approach on two pairwise comparison matrices that would lead to the same MAs considered originally by Zadeh [37] himself. The application of (10) to the two matrices yields $m(A) = 99/199$, $m(B) = 99/199$, $m(C) = 1/199$ which exposes the averaging nature of (10), much more in line with Murphy's rule [38] than with Dempster rule of combination.

4.2. Uncertainty and inconsistency

Inconsistency of elicited preferences, usually captured by an inconsistency index $\mathtt{I}(\mathbf{A})$, has been considered a symptom of irrationality, cognitive errors, cognitive overload, and other factors that could impair the preference elicitation process. Full consistency, and thus full rationality, is hardly ever attainable, and it is common practice to fix a threshold for the inconsistency of preferences beyond which the expert is invited to reconsider his/her statements and reduce inconsistency. In some cases, as recalled by Koksalmis and Kabak [39], the values of inconsistency indices have been used to weigh experts in decision processes to give greater weight to the most rational ones. It should be noted that this notion of conflict is different from the ones usually considered within belief function theory [40], whose intent is to measure a conflict between instantiated MAs, not to measure how inconsistent expert assessments about these potential MAs are. $\mathtt{I}(\mathbf{A})$ could nevertheless be interpreted, to some extent, as the quantity of internal conflict of the information source. A full investigation of its connection with internal conflict within the evidential literature [41], while interesting, is beyond the scope of this paper. We can nevertheless mention that the two notions are different: a consistent matrix could lead to a mass with some internal conflict, while having no inconsistency according to $\mathtt{I}(\mathbf{A})$.

In evidence theory, some decision-making approaches considered $\mathsf{U}(m)$ as an index of the information content of the information contained in the MA m. This indicator has been used in decision support systems. For instance, similarly to what was proposed for inconsistency indices, Liang et al. [9] used the uncertainty of an MA to find suitable weights of experts in a group decision-making problem.

Here we propose a tentative algorithmic approach to the elicitation of MAs by means of pairwise comparisons. Information on inconsistency of preferences and uncertainty of the MA can be incorporated into the MA elicitation process, as shown in Fig. 2, where there are two checks: one on inconsistency and one on uncertainty.

The two thresholds classify pairwise comparison matrices on the set of focal elements and derived MAs into 4 categories, as shown in Fig. 3. The goal is to ask the decision maker to revise his statements to make the information belong to region III. Let us note that this approach is similar to the one proposed by Li et al. [42] for interval-valued pairwise comparison matrices, in whose case uncertainty derives from the interval-valued nature of judgments, while in our case it depends on the obtained MA. Pairwise comparisons in the area III are sufficiently *rational* and provide sufficiently *conclusive* evidence.

Example 3. Consider again the pairwise comparison matrix $\mathbf{A}$ in Example 1 and the two thresholds $\tau_{\mathsf{CR}} = 0.08$ and $\tau_{\mathsf{U}_{KR}} = 0.75$. We can calculate the inconsistency following the approach proposed by Ágoston and Csató [25] in which it is suggested that an incomplete matrix be completed by filling the missing entries in the most consistent way before calculating the inconsistency. Then, the inconsistency can be calculated on the completed matrix, $\tilde{\mathbf{A}}$. In this case, we have $\mathsf{CR}(\tilde{\mathbf{A}}) = 0.1595$, which is above the threshold.

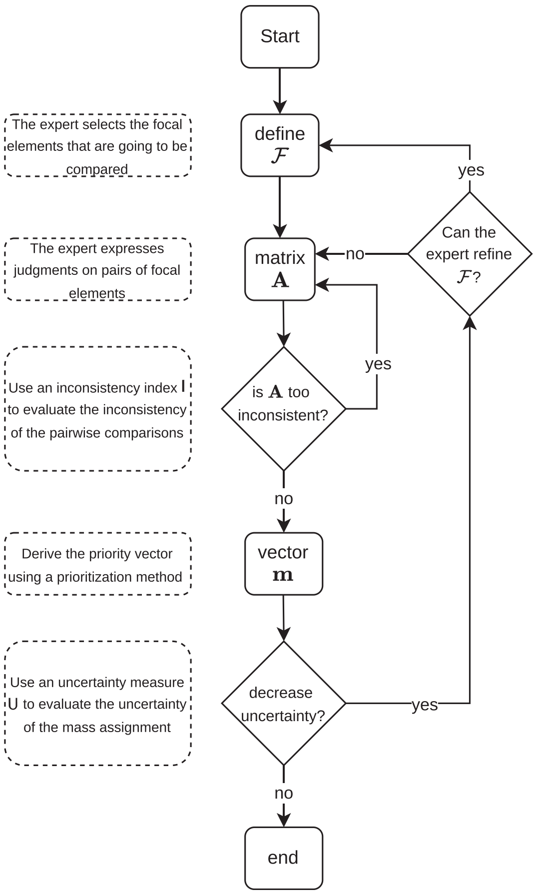

Fig. 2. Workflow of an algorithmic procedure to determine an MA by means of pairwise comparisons with inconsistency and uncertainty thresholds.

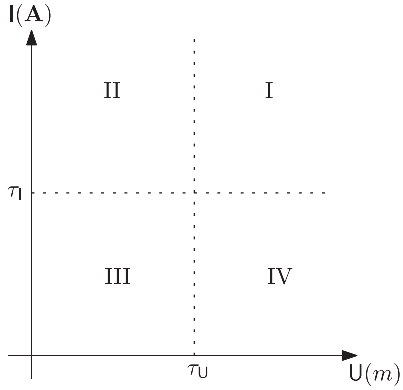

Fig. 3. Representations of pairwise comparisons on a set of focal elements $\mathcal{F}$ can be classified according to inconsistency and uncertainty. Here τ_I and τ_U are the inconsistency and uncertainty thresholds, respectively.

We can then calculate the uncertainty of the mass function, that is $\mathsf{U}_{KR} = 0.6639$ and it respects the threshold $\tau_{\mathsf{U}_{KR}}$. Hence, this case is in the region II in Fig. 3. We ask the expert to reconsider the set of focal elements, to which she adds $A_6 = \{\theta_3\}$. The revised preferences are

$$\mathbf{A} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \end{array} \begin{array}{c} \begin{array}{cccccc} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 \end{array} \\ \begin{pmatrix} 1 & 2 & * & 1/2 & * & * \\ 1/2 & 1 & * & * & 2 & 4 \\ * & * & 1 & 1/3 & * & * \\ 2 & * & 3 & 1 & 1/3 & * \\ * & 1/2 & * & 3 & 1 & 3 \\ * & 1/4 & * & * & 1/3 & 1 \end{pmatrix} \end{array}$$

The inconsistency is now $\mathsf{CR}(\tilde{\mathbf{A}}) = 0.075$, which is now below the threshold; the uncertainty slightly increased to $\mathsf{U}_{KR}(\mathbf{A}) = 0.7062$ but remains below the threshold.

Let us recall that the determination of thresholds for inconsistency or uncertainty measures remains an open issue, as it involves defining cut-off rules for inherently continuous phenomena. Nevertheless, in the literature on pairwise comparisons, several proposals have been developed for inconsistency indices and their acceptable ranges, including extensions to incomplete preference relations [25]. By contrast, relatively little work has addressed threshold selection for uncertainty measures. Progress in this direction will likely depend on the development of measures that are theoretically grounded and easily interpretable by practitioners.

We conclude the subsection by highlighting that inconsistency is *not* a guarantee of reliability as there can be consistent matrices that do not reflect the opinion of an expert. Furthermore, their use should be limited to individual preferences, as their interpretation for group preference relations is questionable; e.g., the aggregation of two inconsistent pairwise comparison matrices could be fully consistent.

4.3. Other representations of preferences

Up to now, we have considered preference relations where the domain of representation of preferences was $\mathbb{R}_>$. There are, however, other domains that can be relevant and among them *probabilistic preference relations* seem particularly appealing and are widely accepted from the perspective of measurement theory. They are also known as reciprocal relations or fuzzy preference relations, and, in their case, the domain of representation of preferences is $D =]0,1[$. Comparisons of this type can be collected into matrices $\mathbf{R} = (r_{ij})_{n\times n} \in]0,1[^{n\times n}$ such that $r_{ii} = 0.5\ \forall i$ and $r_{ij} + r_{ji} = 1\ \forall i,j$. The most common interpretation is that each entry r_{ij} is related to the weight vector through the following relation:

$$r_{ij} \approx \frac{w_i}{w_i + w_j} \tag{12}$$

If (12) holds as equality, then the reciprocal relation is *consistent*. If we consider our problem, then we could interpret weights w_i as weights of evidence allocated to focal elements, i.e., the masses $m(A_i) > 0$. Hence, given the total evidence allocated to two focal elements A_i and A_j, each entry r_{ij} represents the subjective estimation of the percentage of such evidence allocated to A_i, i.e.,

$$r_{ij} \approx \frac{m(A_i)}{m(A_i) + m(A_j)} \tag{13}$$

Similar to pairwise comparison matrices, also for reciprocal relations there are inconsistency indices and prioritization methods, but a more pragmatic approach could be to transform them into a pairwise comparison matrix by means of the following isomorphism

$$a_{ij} = \frac{r_{ij}}{1 - r_{ij}}$$

Given the existence of such an isomorphism, experts can decide to use the representations of preferences that they prefer, and with which they feel more at ease.

4.4. Interval-valued pairwise comparison matrices

In some contexts, it may be unreasonable to assume that an expert can specify exact numerical values, making it appropriate to explore extensions that permit the expression of preferences under uncertainty. As noted in the literature, this need arises both in the case of pairwise comparisons [43] and within evidence theory frameworks [44]. In our setting, a practical way to address this issue is to employ interval-valued pairwise comparisons. In this case, if the domain D is an open subset of the real line, we denote $[D]$ the set of all its sub-intervals, and the interval-valued preference relation, say $\overline{P}$, is defined as

$$\overline{P} : \mathcal{F} \times \mathcal{F} \rightarrow [D] \cup \{*\}$$

The following interval-valued pairwise comparison matrix is an example of numerical representation.

$$\overline{\mathbf{A}} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{pmatrix} [1,1] & [2,3] & * \\ [\frac{1}{3},\frac{1}{2}] & [1,1] & [2,4] \\ * & [\frac{1}{4},\frac{1}{2}] & [1,1] \end{pmatrix} \end{array} \tag{14}$$

Such preference relations are consistent if and only if there exists a vector whose ratios of components are comprised in the specified intervals. That is, $\overline{\mathbf{A}}$ is consistent if and only if, for all determined entries, there exists a *compatible* vector $\mathbf{w}$ such that, $a^-_{ij} \leq \frac{w_i}{w_j} \leq a^+_{ij}$, where a^-_{ij} and a^+_{ij} are the extremes of the interval. In general, as shown in Fig. 4, we can call $W_{\overline{\mathbf{A}}}$ the set of weight vectors compatible with the preferences in $\overline{\mathbf{A}}$.

Depending on whether $W_{\overline{\mathbf{A}}}$ is empty or not, different strategies can be used to estimate a representative weight vector.

- If $W_{\overline{\mathbf{A}}} \neq \emptyset$, one could, for instance, estimate a real-valued weight vector by taking the center of gravity of $W_{\overline{\mathbf{A}}}$ [45], or an interval-valued weight vector $\overline{\mathbf{w}} = ([w^-_i, w^+_i])_{i=1,\dots,n}$ by exploring the maximum and minimum values of weights w_i's that can be found in $W_{\overline{\mathbf{A}}}$ [43]. This second approach could be a starting point to obtain interval-valued MAs, which are equivalent to the interval-valued belief structures defined by Denœux [44].
- If, conversely, $W_{\overline{\mathbf{A}}} = \emptyset$, as surveyed by Wang et al. [46], there is still a plethora of methods to find a representative weight vector. The goal of some of them is to find the vector $\mathbf{w}$ that minimizes the violation of the constraints imposed by the comparisons. Some others, instead, search for the minimum stretch that must be applied to the interval-valued comparisons to make $W_{\overline{\mathbf{A}}} \neq \emptyset$.

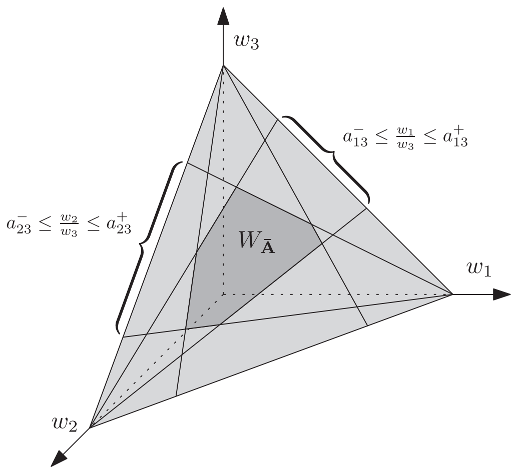

Fig. 4. The effect of the application of interval judgments $[a^-_{13}, a^+_{13}]$ and $[a^-_{23}, a^+_{23}]$ on the set of feasible vectors $W_{\overline{\mathbf{A}}}$ in the case of a pairwise comparison matrix of order 3, e.g, (14).

4.5. Special cases

Some special cases can be identified to help validate the approach based on pairwise comparisons. We shall check that in these special and well-known cases the approach based on pairwise comparisons yields desirable results.

Deterministic MA

An important case refers to the deterministic case, in which there exists an $A \subseteq \Theta$ such that $m(A) = 1$. Within this case, total ignorance is represented by the *vacuous* MA $m(\Theta) = 1$. That is, an expert cannot even compare strict subsets $A \subset \Theta$ as he/she is totally ignorant about their likelihood. Hence, $\mathcal{F} = \{\Theta\}$ and the pairwise comparison matrix collapses into $\mathbf{A} = (1)$, obviously yielding $\mathbf{m} = (m(\Theta)) = (1)$, thus returning the vacuous MA $m(\Theta) = 1$.

Bayesian MA

If $\mathcal{F} = \{\{\theta_1\}, \dots, \{\theta_p\}\}$, then the MA is Bayesian and the weights obtained from $\mathbf{A}$ can be put into a correspondence with the estimated MA using $m(\{\theta_i\}) := w_i$ for all $i = 1, \dots, p$. Furthermore, if the preference relation on $\mathcal{F}$ is complete, then the use of Harker's method makes it collapse to the problem proposed by Yager [7] in which a complete pairwise comparison matrix

$$\mathbf{A} = \begin{array}{c c} & \begin{array}{cccc} \{\theta_1\} & \{\theta_2\} & \cdots & \{\theta_p\} \end{array} \\ \begin{array}{c} \{\theta_1\} \\ \{\theta_2\} \\ \vdots \\ \{\theta_p\} \end{array} & \begin{pmatrix} 1 & a_{12} & \cdots & a_{1p} \\ 1/a_{12} & 1 & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1p} & 1/a_{2p} & \cdots & 1 \end{pmatrix} \end{array}$$

was used, together with the eigenvector method, to estimate subjective probabilities on a sample space Θ.

Consonant MA

If, given a set of focal elements $\mathcal{F} = \{A_1, \dots, A_n\}$, there exists a permutation π such that $A_{\pi(1)} \subset \cdots \subset A_{\pi(n)}$, then the set of focal elements is consonant. In this case, when a prioritization procedure is applied to a pairwise comparison matrix comparing them, the priority vector represents an MA whose associated plausibility measure is, as expected, a possibility measure.

General case

More generally, for each MA m there always exists a complete and consistent preference relation from which it can be elicited. This can be shown using a constructive approach and a pairwise comparison matrix $\mathbf{A}$ of order $|\mathcal{F}|$ with a connected underlying graph and whose known entries are defined as $a_{ij} = m(A_i)/m(A_j)$. At this point, any mathematically sound prioritization should lead to a priority vector whose components correspond to the initial masses of the focal elements. This means that the representative power of the approach based on pairwise comparisons suffices to cover all possible belief functions.

5. A numerical experiment

An important question about our proposed procedure is to determine whether focusing on a few focal elements (defined by the user or analyst, found by an ML algorithm, ...) with possible noisy replies can still provide satisfactory results and approximations even in the case where the full mass is complex. To test the robustness of our approach under different conditions, we consider the possibility that an expert does not compare all the focal elements for which she would have evidence, but only the most relevant ones (in terms of their masses). In practice, we investigate the capacity of a reduced subset of focal elements to be representative of the true preferences of an expert on the full set of focal elements. We set up a Monte Carlo analysis in which we randomly generate MAs with given $|X|$ and $|\mathcal{F}|$. To each MA represented by a vector $\mathbf{m} = \left(m(A_1), \dots, m(A_{|\mathcal{F}|})\right)$ we assign null mass to the $k < |\mathcal{F}|$ focal elements with the smallest MA and renormalize the remaining positive values to form a new vector $\mathbf{m}'$. Then we build a pairwise comparison matrix $\mathbf{A} = \left(m(A_i)/m(A_j)\right)_{(|\mathcal{F}|-k)\times(|\mathcal{F}|-k)}$ to compare the non-zero values and we add a log-normal noise to its entries (while keeping the reciprocity property) to simulate a reasonable level of inconsistency. Then, the geometric mean method is used to find the normalized priority vector from $\mathbf{A}$, which serves to construct a new vector called $\mathbf{m}''$. Finally, we use the distance proposed by Jousselme et al. [47] to quantify the discrepancy between $\mathbf{m}$ and $\mathbf{m}''$,

$$J(\mathbf{m}, \mathbf{m}'') = \sqrt{\frac{1}{2}(\mathbf{m} - \mathbf{m}'')^T \mathbf{D}\,(\mathbf{m} - \mathbf{m}'')}$$

whose value ranges in $[0, 1]$ and where each entry d_{pq} of the matrix $\mathbf{D}$ is defined as

$$d_{pq} = \begin{cases} \dfrac{|A_p \cap A_q|}{|A_p \cup A_q|}, & A_p \cup A_q \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

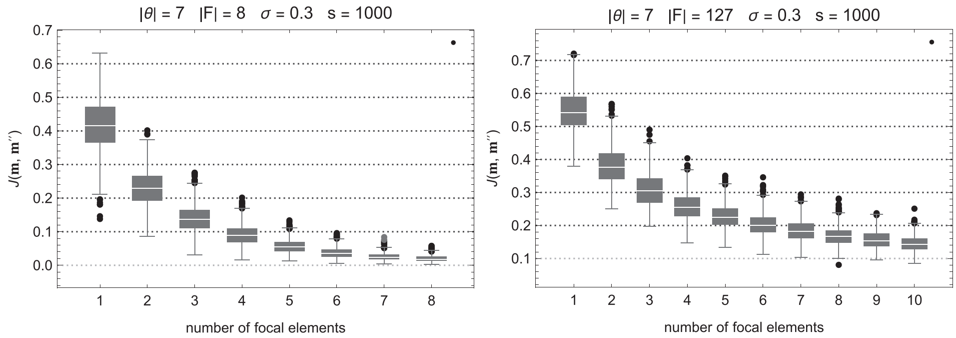

Fig. 5. Jousselme distance between $\mathbf{m}''$ and the original $\mathbf{m}$. The value on the x-axis denotes how many of the original $\mathcal{F}$ focal elements are considered; σ is the standard deviation of the lognormal noise applied to the pairwise comparison matrix; s is the sample size.

The value of $J(\mathbf{m}, \mathbf{m}'')$ can be interpreted as the discrepancy between the "true" MA $\mathbf{m}$ and the MA $\mathbf{m}''$ that is obtained from an expert considering his limited cognitive abilities in terms of number of focal elements that can be identified and his internal consistency. Hence, it attempts to quantify the difference between a purely theoretical normative approach and a pragmatic and prescriptive one.

Fig. 5 reports some of the results for varying cardinalities of frames of discernment and focal elements. We can observe the non-linear increase in the distance when focal elements are discarded. Hence, only a few focal elements are needed to obtain a good approximation of an otherwise complex mass. This is particularly striking for the case of 127 ($= 2^7 - 1$) focal elements, where the distance is already 0.2 with only 5 focal elements being compared. This is possibly also due to the fact that those with the smallest masses were removed, but assuming that an expert can identify the subsets having the most evidence in their favor seems a reasonable assumption. This small experiment suggests that pairwise elicitation done with a reasonable cognitive workload may still result in quite reliable estimates, hence the interest in such approaches. However, these preliminary numerical experiments should be complemented by empirical studies conducted in collaboration with cognitive psychologists to assess their validity. Such studies would also provide an opportunity to compare the proposed approach with alternative elicitation methods, even though only a limited number of them are available for belief functions.

6. Discussion and conclusions

Belief functions, or equivalent uncertainty structures, are complex objects having $2^n - 2$ degrees of freedom, making a full elicitation cognitively unattainable. However, there are still very few works that investigate how to obtain belief functions from the limited information provided by the expert.

In this manuscript, we surveyed the potential of pairwise comparisons as an enabler for the subjective estimation of MAs, and, consequently, of belief measures, thus building a first bridge between the two approaches and methodologies. We saw that most of the existing results obtained for pairwise comparisons (elicitation of priorities, inconsistency indices, group decision-making, interval-valued judgments) can be readily used for the estimation of MAs. Some preliminary experiments also indicate that this approach allows one to obtain faithful mass distributions from a limited budget of elicited values. We believe that the use of pairwise comparisons in evidence theory can be the key to the elicitation of subjective expertise. However, it may not be suitable in application contexts that are data-driven and require real-time responses.

Although we have not yet conducted an empirical validation involving real experts, our proposal builds on the well-established validation of pairwise comparisons reported in the literature. These methods have been successfully applied to the evaluation of various stimuli, such as lengths, areas, weights, and electrical consumption [48], as well as to the elicitation of subjective probabilities [49]. Viewing MAs as expressions of subjective support allocated to hypotheses represented by subsets, it is reasonable, as an initial step, to extend these validation results to the assessment of focal elements.[3]

Future perspectives include testing the approach with human users and checking whether human experiments follow the pattern of our simulation, for instance taking inspiration from Thierry et al. [13].

Another worthy path of investigation is to study how the notions commonly used in pairwise comparisons connect with those used within belief functions: for instance, can we link the notion of conflict of belief functions [40] with the one used within pairwise approaches [39], as already hinted in Section 4.2? Similarly, pairwise comparisons and evidence theory approach the problem of multiple-expert combination in very different ways, and it would be interesting to explore links and differences between the two trends. In short, we believe it would be valuable to strengthen the connections between the two fields, particularly by exploring what pairwise comparison methods can contribute to belief functions, since the reverse direction has already been extensively studied.

Additional future possible extensions include extending the approach to more complex models, both regarding belief functions and elicitation statements. For belief functions, we could for instance look at extensions where masses take complex values, raising the question of interpreting such complex-valued masses and what a comparison between two complex numbers means. Regarding the elicitation statements, we could, for instance, consider listwise approaches rather than pairwise ones Cao et al. [50]. Indeed,

[3] Recall that the term *basic probability assignment* is often used interchangeably with *mass assignment*.

in our view, pairwise comparisons represent a departure from the possibly too naive pointwise evaluations, where each point is scored individually. Pairwise comparisons still retain some simplicity and give the possibility to check the consistency of information, which would be impossible for pointwise evaluations. On the other side of the spectrum, listwise comparisons could be considered, but comparisons requiring the evaluation of multiple elements simultaneously may be perceived as too complex. Also note that numerically assessing intensities for listwise comparisons may be cognitively demanding.

CRediT authorship contribution statement

Matteo Brunelli: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Sébastien Destercke:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the four anonymous reviewers for their constructive comments, which have helped improve the manuscript.

Data availability

No data were used for the research described in the article.

References

[1] B.M. Ayyub, G.J. Klir, Uncertainty Modeling and Analysis in Engineering and the Sciences, Chapman and Hall/CRC, 2006.
[2] T. Augustin, F.P. Coolen, G. De Cooman, M.C. Troffaes, Introduction to Imprecise Probabilities, John Wiley & Sons, 2014.
[3] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[4] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[5] G. Kou, D. Ergu, C. Lin, Y. Chen, Pairwise comparison matrix in multiple criteria decision making, Technol. Econ. Dev. Econ. 22 (2016) 738–765.
[6] T.L. Saaty, Measuring the fuzziness of sets, J. Cybern. 4 (1974) 53–61.
[7] R.R. Yager, An eigenvalue method of obtaining subjective probabilities, Behav. Sci. 24 (1979) 382–387.
[8] M. Beynon, DS/AHP method: a mathematical analysis, including an understanding of uncertainty, Eur. J. Oper. Res. 140 (2002) 148–164.
[9] F. Liang, M. Brunelli, K. Septian, J. Rezaei, Belief-based best worst method, Int. J. Inf. Technol. Decis. Mak. 20 (2021) 287–320.
[10] T. Denoeux, Likelihood-based belief function: justification and some extensions to low-quality data, Int. J. Approx. Reason. 55 (2014) 1535–1547.
[11] T. Denoeux, Decision-making with belief functions: a review, Int. J. Approx. Reason. 109 (2019) 87–110.
[12] A.B. Yaghlane, T. Denœux, K. Mellouli, Elicitation of expert opinions for constructing belief functions, in: Uncertainty and Intelligent Information Systems, World Scientific, 2008, pp. 75–89.
[13] C. Thierry, A. Hoarau, A. Martin, J.-C. Dubois, Y. Le Gall, Real bird dataset with imprecise and uncertain values, in: International Conference on Belief Functions, Springer, 2022, pp. 275–285.
[14] S.A. Sandri, D. Dubois, H.W. Kalfsbeek, Elicitation, assessment, and pooling of expert judgments using possibility theory, IEEE Trans. Fuzzy Syst. 3 (1995) 313–335.
[15] S. Destercke, D. Dubois, E. Chojnacki, Unifying practical uncertainty representations–I: generalized p-boxes, Int. J. Approx. Reason. 49 (2008) 649–663.
[16] D. Bouyssou, T. Marchant, M. Pirlot, A. Tsoukias, P. Vincke, Evaluation and Decision Models with Multiple Criteria: Stepping Stones for the Analyst, Springer, 2006.
[17] J.-B. Yang, M.G. Singh, An evidential reasoning approach for multiple-attribute decision making with uncertainty, IEEE Trans. Syst. Man Cybern. 24 (1994) 1–18.
[18] T. Denœux, M.-H. Masson, Evidential reasoning in large partially ordered sets: application to multi-label classification, ensemble clustering and preference aggregation, Ann. Oper. Res. 195 (2012) 135–161.
[19] P.-L. Guillot, S. Destercke, Preference elicitation with uncertainty: extending regret based methods with belief functions, in: International Conference on Scalable Uncertainty Management, Springer, 2019, pp. 289–309.
[20] D. Dubois, H. Prade, Properties of measures of information in evidence and possibility theories, Fuzzy Sets Syst. 24 (1987) 161–182.
[21] G.J. Klir, A. Ramer, Uncertainty in the Dempster Shafer theory: a critical re-examination, Int. J. Gen. Syst. 18 (1990) 155–166.
[22] B. Kizielewicz, T. Tomczyk, M. Gandor, W. Sałabun, Subjective weight determination methods in multi-criteria decision-making: a systematic review, Procedia Comput. Sci. 246 (2024) 5396–5407.
[23] R. Ureña, F. Chiclana, J.A. Morente-Molinera, E. Herrera-Viedma, Managing incomplete preference relations in decision making: a review and future trends, Inf. Sci. 302 (2015) 14–32.
[24] T.L. Saaty, A scaling method for priorities in hierarchical structures, J. Math. Psychol. 15 (1977) 234–281.
[25] K.C. Ágoston, L. Csató, Inconsistency thresholds for incomplete pairwise comparison matrices, Omega 108 (2022) 102576.
[26] M. Brunelli, Why should not a decision analyst be content with only (n-1) pairwise comparisons? Echoes from the literature, in: The International Workshop on Best-Worst Method, Springer, 2022, pp. 33–40.
[27] G. Crawford, C. Williams, A note on the analysis of subjective judgment matrices, J. Math. Psychol. 29 (1985) 387–405.
[28] P.T. Harker, Alternative modes of questioning in the analytic hierarchy process, Math. Model. 9 (1987) 353–360.
[29] S. Bozóki, J. Fülöp, L. Rónyai, On optimal completion of incomplete pairwise comparison matrices, Math. Comput. Model. 52 (2010) 318–333.
[30] J. Aczél, T.L. Saaty, Procedures for synthesizing ratio judgements, J. Math. Psychol. 27 (1983) 93–102.
[31] E. Forman, K. Peniwati, Aggregating individual judgments and priorities with the analytic hierarchy process, Eur. J. Oper. Res. 108 (1998) 165–169.
[32] K.C. Ágoston, S. Bozóki, L. Csató, A clustering approach for pairwise comparison matrices, J. Oper. Res. Soc. 76 (2025) 971–983.
[33] J. Mazurek, Advances in Pairwise Comparisons: Detection, Evaluation and Reduction of Inconsistency, Multiple Criteria Decision Making, Springer, 2023.
[34] B. Cavallo, A. Ishizaka, Evaluating scales for pairwise comparisons, Ann. Oper. Res. 325 (2023) 951–965.
[35] M. Brunelli, R.P. Kuranage Jayasuriya, V.-N. Huynh, Selection rules for new focal elements in the Dempster Shafer evidence theory, Inf. Sci. 712 (2025) 122160.
[36] P. Smets, R. Kennes, The concept of distinct evidence, in: IPMU 92 Proceedings. PG, 1992, pp. 789–794.
[37] L.A. Zadeh, A simple view of the Dempster Shafer theory of evidence and its implication for the rule of combination, AI Mag. 7 (1986) 85.

[38] C.K. Murphy, Combining belief functions when evidence conflicts, Decis. Support Syst. 29 (2000) 1–9.
[39] E. Koksalmis, Ö. Kabak, Deriving decision makers' weights in group decision making: an overview of objective methods, Inf. Fusion 49 (2019) 146–160.
[40] S. Destercke, T. Burger, Toward an axiomatic definition of conflict between belief functions, IEEE Trans. Cybern. 43 (2013) 585–596.
[41] A. Lepskiy, Decomposition of evidence and internal conflict, Procedia Comput. Sci. 122 (2017) 186–193.
[42] K.W. Li, Z.-J. Wang, X. Tong, Acceptability analysis and priority weight elicitation for interval multiplicative comparison matrices, Eur. J. Oper. Res. 250 (2016) 628–638.
[43] A.A. Salo, R.P. Hämäläinen, Preference programming through approximate ratio comparisons, Eur. J. Oper. Res. 82 (1995) 458–475.
[44] T. Denœux, Reasoning with imprecise belief structures, Int. J. Approx. Reason. 20 (1999) 79–111.
[45] A. Arbel, L. Vargas, Interval judgments and euclidean centers, Math. Comput. Model. 46 (2007) 976–984.
[46] J. Wang, B. Golden, J. Mazurek, Interval pairwise comparisons in the presence of infeasibilities: numerical experiments, Comput. Oper. Res. 173 (2025) 106856.
[47] A.-L. Jousselme, D. Grenier, É. Bossé, A new distance between two bodies of evidence, Inf. Fusion 2 (2001) 91–101.
[48] R. Whitaker, Validation examples of the analytic hierarchy process and analytic network process, Math. Comput. Model. 46 (2007) 840–859.
[49] R.D. Luce, P. Suppes, Preference, utility, and subjective utility, In R.D. Luce, R.R. Bush, E. Galanter (Eds.), Handbook of Mathematical Psychology, vol. 3, Wiley, New York, 1965, pp. 249–409.
[50] Z. Cao, T. Qin, T. Liu, M. Tsai, H. Li, Learning to rank: from pairwise approach to listwise approach, in: ICML, 227 ACM International Conference Proceeding Series, ACM, 2007, pp. 129–136.