



HAL
open science

A generic framework to include Belief functions in preference handling and multi-criteria decision

S. Destercke

► **To cite this version:**

S. Destercke. A generic framework to include Belief functions in preference handling and multi-criteria decision. *International Journal of Approximate Reasoning*, 2018, 98, pp.62-77. 10.1016/j.ijar.2018.04.005 . hal-01781327

HAL Id: hal-01781327

<https://hal.science/hal-01781327>

Submitted on 4 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A GENERIC FRAMEWORK TO INCLUDE BELIEF FUNCTIONS IN PREFERENCE HANDLING AND MULTI-CRITERIA DECISION

SÉBASTIEN DESTERCKE

ABSTRACT. Modelling the preferences of a decision maker about alternatives having multiple criteria usually starts by collecting preference information (comparisons of alternatives, importance of criteria, ...), which are then used to fit a preference model issued from some set of hypothesis (weighted average, CP-net, lexicographic orderings, ...). In practice, this process may often lead to inconsistencies that may be due to inaccurate information provided by the decision maker, who can be unsure about the provided information, or to a poor choice of hypothesis set, which can be too restrictive or not well adapted to the decision process. In this paper, we propose to use belief functions as a way to quantify and resolve such inconsistencies, notably by allowing the decision maker to express her/his certainty about the provided preferential information. Our framework is generic, in the sense that it does not assume a given set of hypothesis a priori, and is consistent with precise methods, in the sense that in the absence of uncertainty and inconsistencies in the information, precise models are ultimately retrieved.

Keywords: preferences, inconsistency handling, uncertainty, multi-criteria, evidence theory, AHP, lexicographic orders

1. INTRODUCTION

Preference modelling and multi-criteria decision analysis (MCDA for short) are increasingly used in our everyday lives. Generally speaking, their goal is to help decision makers to model their preferences about multi-variate alternatives, to then formulate recommendations on unseen or unanalysed alternatives. Such recommendations can take various shapes, but three common problems can be differentiated [2]:

- **the choice problem**, in which a (set of) best alternative(s) has to be recommended to the decision maker (DM);
- **the ranking problem**, in which the alternatives have to be ranked before the ranking is presented to the DM;
- **the sorting problem**, in which each alternative has to be attributed to a class among a set of ordinal classes, the attribution being then presented to the decision maker.

In this paper, we will be interested in the first two problems, who are usually related, since the choice problem roughly consists in presenting only those elements that would be ranked highest in the ranking problem. The sorting problem can be considered to some extent as different, as it requires to first define the discrete classes, and only then can it be seen as the task of summarizing the (complete)

ordering of alternatives into these discrete classes. This latter problem is typically considered by outranking models [32, 21], and is close to the learning problem known as ordinal regression or classification [12].

One common task, in preference modelling as well as in MCDA, is to collect or elicit preferences of decision makers (DM). This elicitation process can take various forms: from asking directly the values of the parameters to comparing known alternatives; from being static, where all preference information are collected at once, to being dynamic, where collected preference information depends on the process history; from aiming at identifying a preference model with enough precision, to aiming at making a robust recommendation of known alternatives (possibly with an imprecisely identified model). In all these cases, each piece of collected information then helps to better identify the preference model of the DM. Also, the elicitation processes may differ greatly between the different MCDA models, be it Choquet Integrals [28], GAI-networks [27], CP-net [6], lexicographic orderings [26, 22] and their conditional extensions [4], . . .

While we will not discuss in details how to perform the elicitation step in this paper, a common problem in this task is to ensure that the information provided by the DM is consistent with the chosen model, as the DM may be uncertain about the given preference information, or as the chosen model may not have a sufficient expressive power to describe accurately these preferences. It is therefore desirable to encode the uncertainty associated to the various pieces of information, as well as to combine them. Common ways to handle such a task is to identify the parameters of the preference models minimising some error term, for instance the quadratic error [28], to adopt a probabilistic version of preference models and to develop a corresponding probabilistic (e.g., Bayesian) learning method [48], or to relax the model constraints by following some minimal change principle [38]. While such methods try to solve inconsistency between the assessments and the chosen model in principled way, most do not consider the initial information to be uncertain.

Another problem when modelling preferences is to choose an adequate family of models, expressive enough to capture the DM preferences, but also sufficiently simple so that it can be identified with a reasonable amount of information. While there are some works around comparing the expressiveness of different model families [37, 35], few of them have actually investigated how to choose a family among a set of possible ones.

In this paper, we propose to model uncertainty in preference information through belief functions, arguing that they can bring interesting answers to both issues (i.e., inconsistency handling and model choice). Indeed, belief functions are adequate models to model subjective uncertainty about non-statistical quantities (in our case the preferences of an individual decision maker), and many works have been devoted to the question of how to combine such information and assess the resulting inconsistency [18].

It is not the first work that tries to combine belief functions with MCDA and preference modelling, however the past works that dealt with such issues can be split into two main lines of works:

- those that start from a specific MCDA model and propose some adaptation to embed belief functions within it. This has mainly been done for the AHP model [3, 24], but also for the TOPSIS model [20].
- those that start from belief functions defined on the various criteria, and then propose to perform inferences about preferences using such belief functions as well as tools issued from evidence theory, possibly but not necessarily inspired from existing MCDA techniques. This is for instance the case of the evidential reasoning approach [49] or the case of the out-ranking approach developed by Boujelben *et. al.* [5].

The approach investigated and proposed in this paper differs from those works in at least two ways:

- we do not make any a priori assumption about the kind of model used, that is we do not start from an existing method and propose a corresponding extension. This means that the proposal can be directly applied to various methods (including those already well-studied);
- when selecting a particular model, we can retrieve the precise, certain version of the method as a particular instance of our approach, meaning that we are fully consistent with the standard models.

Section 2 describes the framework we propose, with an illustrative running example using weighted averages. Needed notions issued from evidence theory are introduced gradually when they are needed. To show that our approach can be adapted to various models, Section 3 provides three other illustrations to show the generality of our approach. the first one uses a simple dominance rules, the second a lexicographic qualitative approach to describe the provided information, and the last one is a simplified application of the popular AHP method. Section 4 then discusses how the framework of belief functions can be instrumental to deal with the problems we mentioned in this introduction: handling inconsistent assessments of the DM, and choosing a rich enough family of models. This paper is an extended version of [17], that beyond additional explanations, includes examples dealing with other models and an extended discussion of the solutions proposed for the ranking and choice problems.

2. THE BASIC SCHEME

We assume that we want to describe preferences over alternatives X issued from a multivariate space $\mathcal{X} = \times_{i=1}^C \mathcal{X}^i$ constituted of C different criteria X^i . For instance, \mathcal{X} may be the space of hotels, cars, applicants, meals . . . and a given criteria \mathcal{X}^i may be the price, age, skill in a given language, main courses, . . .

Unless stated otherwise and to facilitate the reading, we will also assume that X^i is either real-valued within the unit interval $[0, 1]$, or finite in which case the elements x_1^i, x_2^i, \dots will be assumed to be completely ordered according to a relation \succ_i , and indexed such that $x_k^i \succ_i x_\ell^i$ whenever $k > \ell$. These are common assumptions in many preference or MCDA models, yet not in all of them (for instance, CP-nets do not make such an assumption). In theory, the proposed framework can also be

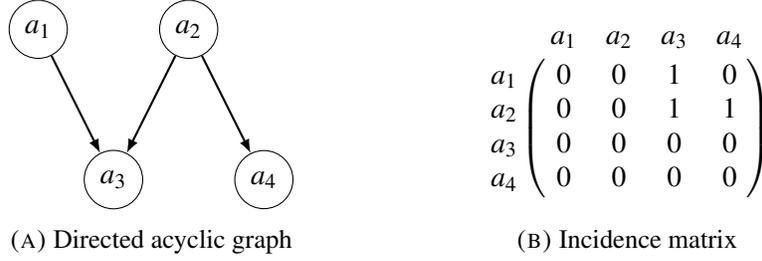


FIGURE 1. Directed acyclic graph and incidence matrix of Example

applied to such models, yet such an application will require dealing with specific computational issues, that we will not discuss here.

We will denote by $\mathbb{P}_{\mathcal{X}}$ the set of partial orders defined over \mathcal{X} . Recall that a strict partial order P is a binary relation over \mathcal{X}^2 that satisfies

- Irreflexivity: not $P(x, x)$ for any $x \in \mathcal{X}$
- Transitivity: $P(x, y)$ and $P(y, z)$ implies $P(x, z)$ for any $(x, y, z) \in \mathcal{X}^3$
- Asymmetry: either $P(x, y)$ or $P(y, x)$, but not both

and where $P(x, y)$ can be read “ x is preferred to y ” that we will also denote $x \succ_P y$. When P concerns only a finite set of $\mathcal{A} = \{a_1, \dots, a_n\} \subseteq \mathcal{X}$ of alternatives, convenient ways to represent it are by its associated directed acyclic graph $\mathcal{G}_P = (V, E)$ with $V = \mathcal{A}$ and $(a_i, a_j) \in E$ iff $P(a_i, a_j)$, and by its incidence matrix whose elements denoted P_{ij} will be such that $P_{ij} = 1$ iff $P(a_i, a_j)$. Given a partial order P and a subset \mathcal{A} , we will denote by Max_P the set of its maximal elements, i.e.,

$$(1) \quad Max_P = \{a \in \mathcal{A} : \nexists a' \in \mathcal{A} \text{ s.t. } a' \succ_P a\}$$

Example 1. Consider four alternatives $\{a_1, \dots, a_4\}$ and the partial order

$$P = \{(a_1, a_3), (a_2, a_3), (a_2, a_4)\}.$$

Its associated directed graphs and incidence matrix are shown in Figure 1, while its set of maximal elements is $Max_P = \{a_1, a_2\}$

2.1. Elementary information item. Our approach is based on the use of elementary information items, about which we assume the following:

- the decision-maker (DM) provides items of preferential information \mathcal{I}_i together with some certainty degree $\alpha_i \in [0, 1]$ ($\alpha_i = 1$ meaning the DM is certain about her/his preference). We remain quite agnostic about the nature of \mathcal{I}_i : it can be a comparison between alternatives of \mathcal{X} (“I prefer menu A to menu B”) or between criteria (“the price is the most important criteria”, “I prefer being close to the city than having a big room”), direct information about the model, etc.
- given a selected space \mathcal{H} of possible models or hypothesis (weighted mean, Choquet integral, ...), each item \mathcal{I}_i can be translated into constraints inducing a subset H_i of possible models consistent with this information.

- Each model $h \in \mathcal{H}$ maps subsets of \mathcal{X} to a partial order relation $P_h \in \mathbb{P}_{\mathcal{X}}$. A subset $H \subseteq \mathcal{H}$ maps subsets of \mathcal{X} to the partial order relation $P_H = \bigcap_{h \in H} P_h$.

We then model this information as a simple support mass function m_i defined as follows over \mathcal{H} :

$$(2) \quad m_i(H_i) = \alpha_i, \quad m_i(\mathcal{H}) = 1 - \alpha_i.$$

Mass functions are the basic building block of evidence theory [43]. A mass function over some space \mathcal{H} is defined as a non-negative mapping from subsets of \mathcal{H} (possibly including the empty-set) to the unit interval, so that it sums up to one. That is, $m : \wp(\mathcal{H}) \rightarrow [0, 1]$ with $\sum m(E) = 1$ and $\wp(\mathcal{H})$ the power set of \mathcal{H} . The mass $m(\emptyset)$ can (and will be in our case) be interpreted as the amount of conflict in the information we have [18]. A subset $H \subseteq \mathcal{H}$ such that $m(H) > 0$ is often called a *focal set*, and we will denote by $\mathcal{F} = \{H \subseteq \mathcal{H} : m(H) > 0\}$ the collection of focal sets of m .

Example 2. Consider the problem with three criteria X^1, X^2, X^3 that are the average notes over 10 of students in Physics, Math, French (for easiness, we will use P, M, F). This means that $X^i \in [0, 10]$, and that \mathcal{X} is the set of possible students. Assume furthermore that the chosen hypothesis space \mathcal{H} is the space of weighted averages, and a model $h \in \mathcal{H}$ is then summarised by a positive vector (w_1, w_2, w_3) where $\sum w_i = 1$. A student a_i is evaluated by $a_i = w_1P + w_2M + w_3F$, and an alternative a_i is better than a_j if $a_i > a_j$.

This means that any subset of models can be summarized by a subset of the space $\mathcal{H} = \{(w_1, w_2) : w_1 + w_2 \leq 1\}$, since the last weight can be inferred from the two firsts. For instance, let us assume that the information item \mathcal{I} is $(0, 8, 5) \succ (8, 4, 5)$, meaning that

$$0w_1 + 8w_2 + 5w_3 > 8w_1 + 4w_2 + 5w_3 \rightarrow w_2 > 2w_1$$

The resulting subspace H of models is then pictured in Figure 2. The decision maker can then provide some assessment of how certain she/he is about this information by providing a value α . For instance, if the DM is certain to choose a student with grades $(0, 8, 5)$ over one with grades $(8, 4, 5)$, then α should be close to 1. Yet if the DM is quite uncertain about this choice, then α should be closer to 0. For example, one could use a linguistic scale to assess α .

2.2. Combining elements of information. In practice, the DM will deliver multiple items of information, that should be combined in a single representation. If m_1 and m_2 are two mass functions over the space \mathcal{H} , then their conjunctive combination in evidence theory [44, 47] is defined as the mass

$$(3) \quad m_{1 \cap 2}(H) = \sum_{\substack{H_1 \in \mathcal{F}_1 \\ H_1 \cap H_2 = H}} m_1(H_1)m_2(H_2),$$

which is applicable if we consider that the provided information items are distinct, i.e., if the DM answers one query without considering strongly the others, which is a reasonable assumption in a preference learning settings where the DM seldom

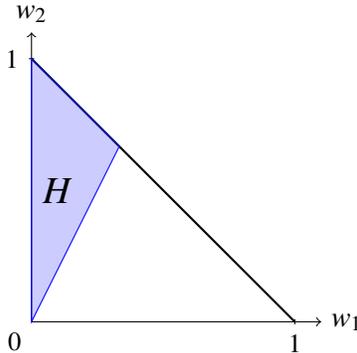


FIGURE 2. Information item subset

answers a question by thinking about the ones already answered. Such a combination rule is also consistent with Bayesian approaches [48], where evidence are accumulated and revised according to a product rule, only using a more specific model of uncertainty. Note that this assumption of information items independence can be dropped by adopting so-called cautious or idempotent combination rules [13, 19, 33].

If we have n masses m_1, \dots, m_n to combine, for instance corresponding to n information items $\mathcal{I}_1, \dots, \mathcal{I}_n$, we can iteratively apply Equation (3), as it is commutative and associative. From Equation (3), we get that if each m_i has two focal elements (say, H_i and \mathcal{H}), then the number of focal elements of the combined mass double after each application of 3. This of course limits the number n we can consider before proceeding to approximations, yet in frameworks where individual decision makers are asked about their preferences, this number is often small.

Of course, it may happen that the provided preferential information items are conflicting, in which case a non-null mass $m(\emptyset) > 0$ may be given to the empty set, corresponding to the fact that no models in \mathcal{H} satisfies all preferential information items at once. In evidence theory, two main ways to deal with this situation exist:

- W1 Ignore the fact that some conflicting information exists, redistributing $m(\emptyset)$ to other focal elements of \mathcal{F} to normalize m into m' . This is the principle we will follow for now, and explore some alternatives in Section 4. Multiple different proposals exist in the literature [46], but the most commonly used consists in considering m' such that for any $H \in \mathcal{F} \setminus \emptyset$ we have

$$(4) \quad m'(H) = \frac{m(H)}{1 - m(\emptyset)}.$$

This rule, originally proposed by Dempster, is similar to other basic normalization rules restoring consistency in possibilistic and probabilistic contexts.

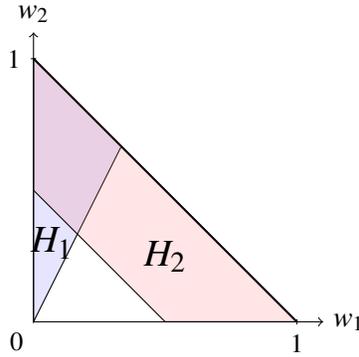


FIGURE 3. Information item subsets

W2 Use the value of $m(\emptyset)$ as a trigger to resolve the conflicting situation rather than just relocating $m(\emptyset)$. A typical solution is then to use alternative combination rules [40], such as the disjunctive rule that replaces the conjunction \cap in Equation (3) by the disjunction \cup . Such rules can be considered in our case, but often involve taking some disjunctions and typically do not satisfy the associativity property, thus potentially generating many additional computational problems.

W3 Pick a subset of information sources (or items in our case) such that the resulting conflict is limited. A similar strategy will be proposed in Section 4.

In Section 4, we describe how the amount of inconsistency can be made instrumental in our context, either to select the relevant information and/or to select an appropriate hypothesis spaces.

Example 3. Consider again the setting of Example 2, The first information delivered, $H_1 = \{(w_1, w_2) \in \mathcal{H} : w_2 \geq 2w_1\}$ is that $(0, 8, 5) \succ (8, 4, 5)$ with a mild certainty, say $\alpha_1 = 0.6$. The second item of information provided by the DM is that ‘sciences are more important than language’, which we interpret as the inequality

$$w_1 + w_2 \geq w_3 \rightarrow w_1 + w_2 \geq 0.5$$

obtained from the fact that $\sum w_i = 1$. The DM is pretty sure about it, meaning $\alpha_2 = 0.9$ and that $H_2 = \{(w_1, w_2) \in \mathcal{H} : w_2 + w_1 \geq 0.5\}$. Again, Figure 3 provides an illustration of the obtained focal elements. The non-conflicting mass resulting from the application of (3) to m_1, m_2 is then

$$m(H_1) = 0.06, m(H_2) = 0.36, m(H_1 \cap H_2) = 0.54, m(\mathcal{H}) = 0.04.$$

2.3. Inferences: choice and ranking. Let us now consider that we are provided with a finite set $\mathcal{A} = \{a_1, \dots, a_n\}$ of alternatives, and that the mass resulting from previous combinations has m focal elements H_1, \dots, H_m that can be the combination of simple ones, and can include the empty set. Two common tasks in MCDM or in preference handling are: to provide a recommendation to the DM, either in the form of one alternative a^* or of a set A^* of them, and to provide a ranking of the

alternatives in \mathcal{A} , possibly partial. We suggest some means to achieve both tasks in the next sections.

2.3.1. *Choice.* When a partial order P is defined over the set \mathcal{A} , a natural recommendation is to provide the set $A^* = \text{Max}_P$ of maximal items derived from P using Equation (1). Providing a choice in an evidential framework, based on the mass m , then requires to extend this notion. Assuming that the best representation of the DM preferences we could have is a partial order P^* , a simple way to do so is to measure the so-called belief and plausibility measures that a given subset $A \subseteq \mathcal{A}$ is a subset of the set of maximal elements A^* , considering that the subset Max_{P_i} derived from the focal element H_i represents a superset of A^* . These two values are easy to compute, as under these assumptions we have that the plausibility is equal to

$$(5) \quad \text{Pl}(A \subseteq A^*) = \sum_{A \subseteq \text{Max}_{P_i}} m(H_i) = \sum_{A \in 2^{\text{Max}_{P_i}}} m(H_i),$$

since if there is an element $a_i \in A$ such that $a_i \notin \text{Max}_{P_i}$, then A cannot be a subset of the maximal elements of P^* , as it cannot be a subset of the imprecise knowledge P_i we have about it. On the other hand, if A is a subset of Max_{P_i} , then it is consistent with the assumption that A is a subset of the maximal elements of P^* . On the other hand, the corresponding belief degree can be computed as

$$(6) \quad \text{Bel}(A \subseteq A^*) = \sum_{A=2^{\text{Max}_{P_i} \setminus \emptyset}} m(H_i) = \begin{cases} 0 & \text{if } |A| > 1, \\ \sum_{\text{Max}_{P_i}=\{a\}} m(H_i) & \text{if } A = \{a\}. \end{cases}$$

This particular form of Bel is due to the fact that information H_i implies that A is a subset of maximal elements of the sought partial orders only when all possible sets of maximal elements induced by H_i are supersets of A . Now, to exclude single alternatives from the possible sets, we need H_i to be the union of partial orders that all have more than one element being maximal, hence to have positive, explicit information about items being incomparable. This would require at least two things: consider incomparability not as a lack of knowledge about preferences, but as a knowledge of lack of preferences, and to build up elicitation protocol allowing for incomparability. Although quite interesting, we will not consider this aspects here, given that it would be a research topic in itself (some elements can be found in [15, 8]). Some noteworthy properties of Equations (5)- (6) are the following:

- for an alternative $a \in \mathcal{A}$, $\text{Pl}(\{a\}) = 1$ if and only if $\{a\}$ is a maximal element of all possible partial orders (in particular, $m(\emptyset) = 0$). In fact $\text{Pl}(\{a\})$ can be considered as the number of focal elements “voting” for a in a setting of approval voting [7] where voters (the sets H_i) can separate desirable candidates (the sets Max_{P_i}) from the undesirable ones;
- for an alternative $a \in \mathcal{A}$, the interval $[\text{Bel}(\{a\}), \text{Pl}(\{a\})]$ represents our belief about a being the only best alternative. In particular, $\text{Bel}(\{a\})$ will be close to one only if most possible partial orders P_i have a as unique maximal element, and $[\text{Bel}(\{a\}), \text{Pl}(\{a\})] = [0, 1]$ if a is always a possible maximal element, but never the only one;

	P	M	F
a_1	4	3	9
a_2	5	9	6
a_3	8	7	3
a_4	7	1	7

TABLE 1. A set of alternatives

- given two subsets $A \subseteq B \subseteq \mathcal{A}$, we can have $Pl(A \subseteq A^*) \geq Pl(B \subseteq A^*)$, meaning in particular that it makes sense to look for the most plausible set of maximal elements, as it may not be \mathcal{A} . The most left-hand term of Equation (5) gives an idea of why this happens, as we are somehow considering a plausibility function not over $2^{\mathcal{A}}$, but over the power set of it. Indeed, for example if $\{a_1, a_3\}$ is the set of maximal elements (hence a subset of \mathcal{A}), then the associated set of possible true values for A^* is $\{\{a_1\}, \{a_3\}, \{a_1, a_3\}\}$, i.e., a subset of subsets. Within such a view, A and B are not subsets but singletons, hence the possible non-monotonicity.

Still, there could be many subsets whose plausibility is equal, in which case the least commitment principle dictates that we should consider the largest possible set, in absence of additional information. If \overline{Pl} denotes this maximal plausible value, this leads to take the set

$$(7) \quad \hat{A} = \arg \max_{|A|} \{A \in 2^{\mathcal{A}} : Pl(A \subseteq A^*) = \overline{Pl}\}$$

Example 4. Consider the example where $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$ consists in the four alternatives presented in Table 1. We then consider the mass of four focal elements given in Example 3 with the following renaming:

$$H_1 = H_1, H_2 = H_2, H_3 = H_1 \cap H_2, H_4 = \mathcal{H}$$

From these, we can for instance deduce that $P_1 = \{(a_1, a_4), (a_2, a_3)\}$ using simple linear programming. For example, that $(a_1, a_4) \in P_1$ comes from the fact that the difference between a_1 and a_4 evaluation is always positive in H_1 , that is

$$\min_{(w_1, w_2, w_3) \in H_1} (4w_1 + 3w_2 + 9w_3) - (7w_1 + w_2 + 7w_3) > 0.$$

Similarly, we can obtain $P_3 = \{(a_1, a_4), (a_2, a_1), (a_2, a_3), (a_3, a_1), (a_3, a_4), (a_2, a_4)\}$ and $P_2 = P_4 = \{\}$, from which we have $Max_{P_1} = \{a_1, a_2\}$, $Max_{P_3} = \{a_2\}$, $Max_{P_2} = Max_{P_4} = \mathcal{A}$. Interestingly, this shows us that while information \mathcal{I}_2 leading to H_2 does not provide sufficient information to recommend any student in \mathcal{A} , when combined with \mathcal{I}_1 , it does improve our recommendation, as $|Max_{P_3}| = 1$.

Table 2 shows the plausibilities and belief resulting from Equations (5)- (6), when applied to subsets of maximal elements counting one or two elements. Clearly, $\{a_2\}$ is the most plausible answer, as well as the most credible, and hence should be chosen as the predicted set of maximal elements.

	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$	$\{a_4\}$	$\{a_1, a_2\}$	$\{a_1, a_3\}$	$\{a_1, a_4\}$	$\{a_2, a_3\}$	$\{a_2, a_4\}$	$\{a_3, a_4\}$
<i>Pl</i>	0.46	1	0.4	0.4	0.46	0.4	0.4	0.4	0.4	0.4
<i>Bel</i>	0	0.54	0	0	0	0	0	0	0	0

TABLE 2. Plausibilities and belief on sets of one and two alternatives

2.3.2. *Ranking.* A second common task is to provide a (possibly partial) ranking of the alternatives. Since each (non-empty) focal element can be associated to a partial order over \mathcal{A} , this problem is close to the one of aggregating partial orders. It is known that solutions to such problems typically do not meet all desirable properties [41]. However, we can take inspirations from proposals existing in the field of preference modelling and aggregation to derive possible solutions. For instance, we can use previous results about belief functions [36, 15], or classical aggregation rules of pairwise scores to predict rankings [16]. Following Denoeux and Masson [15], we can try to find the hypothesis \hat{H} maximizing the commonality function, where the commonality function of H is

$$Q(H) = \sum_{H \subseteq H_i} m(H_i)$$

Hence the commonality of H will increase if it is a refinement of many (ideally, all) hypothesis H_i , or said otherwise if it does not conflict with H_i . More precisely, if P denotes the partial order corresponding to H , then $H \subseteq H_i$ if and only if $P_i(k, l) \Rightarrow P(k, l)$ for all pairs k, l of alternatives. Again, if we denote by $\bar{Q} = \max_{H \subseteq \mathcal{H}} Q(H)$, there may be multiple H whose commonality reaches \bar{Q} . Again, the principle of least commitment suggest to choose the coarsest partial order we can deduce from our information, that is to pick

$$(8) \quad \hat{P} = \arg \min_{|P|} \{H \in \mathcal{H} : Q(H) = \bar{Q}\}$$

with P the partial order corresponding to H , and $|P|$ the number of relations within the partial order. However, computing (8) can be quite difficult, as \mathcal{H} is usually a quite complex space.

An easier way to build a ranking is to decompose preferences in pairwise information, and then to combine it into a final ranking¹. Focusing on pairwise information, we can compute, for two alternatives a_i, a_j , the plausibilities and belief that one is preferred to the other, that can be computed as follows:

$$(9) \quad Pl(a_i \succ a_j) = \sum_{P_k, P_{k,ij} \neq 0} m(H_k),$$

$$(10) \quad Bel(a_i \succ a_j) = \sum_{P_k, P_{k,ij} = 1} m(H_k),$$

¹Note that here we go from a mass function defined on the complete relation to pairwise information, which is different from common procedures [36, 14] that go from pairwise information to the complete relation

where $P_{k,ij}$ is the (i, j) value of the incidence matrix of P_k . In practice, Pl comes down to sum all partial orders that have a linear extension with $a_i \succ a_j$, and Bel the partial orders whose all linear extensions have $a_i \succ a_j$. The result of this procedure can be seen as an interval-valued matrix R with $R_{i,j} = [Bel(a_i \succ a_j), Pl(a_i \succ a_j)]$. It can also be noted that, if $m(\emptyset) = 0$, we do have $Pl(a_i \succ a_j) = 1 - Bel(a_j \succ a_i)$. Each pairwise information must then be combined to obtain a full ranking. Two classical ways to obtain such a ranking are the following:

- Threshold the matrix, and declare that $a_i \succ a_j$ if $Bel(a_i \succ a_j) > 0$, i.e., as soon as some evidence implies $a_i \succ a_j$. Such a method is quite easy to apply, yet may lead to cycles in the produced ranking [9], in which case the threshold can be increased to eliminate such cycles, at the expense of producing more incomparabilities. We will call it the *thresholding approach*. The reason why we use a 0 threshold rather than the minimal 0.5 value used in probabilistic approaches [10] will become clear in Section 2.3.4.
- Compute for each alternative a_i , the interval-valued score

$$[\underline{s}_i, \bar{s}_i] = \sum_{a_j \neq a_i} [Bel(a_i \succ a_j), Pl(a_i \succ a_j)]$$

and then consider the resulting partial order. In addition to being straightforward to apply, has the advantage of always producing a partial order (in fact, an interval order). One drawback, though, is that imprecision of intervals will add up, potentially producing a very partial order. We will call it the *scoring approach*. Note that in label ranking settings, such a combination leads to minimize the loss function associated to the Spearman footrule [30], and it would be interesting to study whether a similar result can be obtained in our case.

Example 5. Assume we have three focal elements H_1, H_2, H_3 , each with $m(H_i) = 1/3$, that over a set $\mathcal{A} = \{a_1, a_2, a_3\}$ induce the orders

$$P_1 = a_3 \succ a_1 \succ a_2; P_2 = a_1 \succ a_2 \succ a_3; P_3 = a_2 \succ a_3 \succ a_1.$$

From this, we have that $Bel(a_1 \succ a_2) = Bel(a_2 \succ a_3) = Bel(a_3 \succ a_1) = 2/3$, from which we get that the thresholding approach would lead to a cycle, even with a value above 0.5.

This example illustrates the fact that the thresholding approach can lead to inconsistent results, and uses arguments similar to the one of the celebrated Condorcet paradox. However, we shall see in Section 2.3.4 that this only happens when some inconsistencies are observed within the obtained information, hence when $m(\emptyset) > 0$ at some point (in particular, the mass function of Example 5 could only be obtained after a renormalisation in our context).

Example 6. The matrix R and the scores $[\underline{s}_i, \bar{s}_i]$ resulting from Example 4 are

$$\begin{array}{c} a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ 0 & [0, 0.46] & [0, 1] & [0.6, 1] \\ [0.54, 1] & 0 & [0.6, 1] & [0.54, 1] \\ [0, 1] & [0, 0.4] & 0 & [0.54, 1] \\ [0, 0.4] & [0, 0.46] & [0, 0.46] & 0 \end{pmatrix} \Sigma = \begin{pmatrix} [\underline{s}_i, \bar{s}_i] \\ [0.6, 2.46] \\ [1.68, 3] \\ [0.54, 2.4] \\ [0, 1.32] \end{pmatrix}$$

from which we get the final partial order $P^* = \{(a_2, a_4)\}$. In contrast, using the pairwise threshold approach gives $P^* = \{(a_2, a_4), (a_2, a_1), (a_2, a_3), (a_1, a_4), (a_3, a_4)\}$, identifying a_4 as the least preferred element, and a_2 as the most preferred one.

In general, we cannot expect these three methods to provide identical result, especially if preferential information presented some conflict which was then cancelled through a normalisation process. Yet, in the next section, we show that the results they provide are consistent with each other when the provided preferential information are not conflicting with respect to the chosen hypothesis space \mathcal{H} .

2.3.3. Combining hypothesis vs combining choices or ranks. So far, we have proposed (see Equation (3)) to combine the focal elements corresponding to sets H_i of models induced by the DM information. Another possibility, if we consider a set \mathcal{A} of alternatives, would be to first map the different sets H_i to their corresponding partial orders P_i or set of maximal elements Max_{P_i} , and to combine these together. That is, replacing $H_1 \cap H_2$ by $P_1 \cup P_2$ or by $Max_{P_1} \cap Max_{P_2}$ in Equation (3).

The upside of such combinations is that, while sets H_1, H_2 may be disjoint sets of possible models, they may still induce partial orders and sets of maximal elements having a non-empty intersection. The downside is that we would have information that is specific to \mathcal{A} , hence could not be applied or re-used for new sets of alternatives.

Example 7. Consider the simple example where H_1 corresponds to the statement ‘‘Physics is strictly more important than Math, whose in turn is at least as important than French’’ and H_2 to ‘‘Math is strictly more important than Physics, whose in turn is at least as important than French’’. H_1 corresponds to $\{(w_1, w_2, w_3) : w_1 > w_2 \geq w_3\}$ and H_2 to $\{(w_1, w_2, w_3) : w_2 > w_1 \geq w_3\}$, meaning that $H_1 \cap H_2 = \emptyset$.

Yet, if we consider the set \mathcal{A} with two alternatives $a_1 = (4, 4, 6)$ and $a_2 = (6, 6, 4)$, we would have $P_1 = P_2 = \{(a_2, a_1)\}$, $Max_{P_1} = Max_{P_2} = \{a_2\}$

Similarly, two partial orders P_i, P_j may well be incompatible, but have sets of maximal elements that have a non-empty intersection. Take for instance a set of four alternatives $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$ with $P_1 = \{(a_1, a_3), (a_2, a_3), (a_3, a_4)\}$ and $P_2 = \{(a_1, a_4), (a_2, a_4), (a_4, a_3)\}$. $P_1 \cup P_2$ is not a partial order (as it contains (a_3, a_4) and (a_4, a_3)), but both have $\{a_1, a_2\}$ as sets of maximal elements.

2.3.4. The case of consistent information. We will now discuss our previous proposals in the specific case where preferential information items are consistent with one another, that is when $\bigcap_{i=1}^m H_m \neq \emptyset$. In practice, this case does not necessitate any uncertainty model on top of the information items, but a minimal requirement is that our method behave as one would expect in such situations.

In the choice problem, any $A \subseteq \cap_{i=1}^m \text{Max}_{P_i}$ would have a plausibility of 1 in Equation (5), while all other sets would have a value strictly lower than one. Among these sets, the one having the maximum cardinality, hence the recommended one would simply be $\hat{A} = \cap_{i=1}^m \text{Max}_{P_i}$, as should be expected.

Using similar arguments, in the ranking problem it is clear that all $H \subseteq \cap_{i=1}^m H_i$, and only those, would receive a commonality equal to 1. The resulting least committed hypothesis maximizing this commonality would then be $\hat{H} = \cap_{i=1}^m H_i$, which corresponds to the predicted partial order $\hat{P} = \cup_{i=1}^m P_i$, again what would be expected.

Let us now show that the different proposed ranking methods, while they can provide different answers in general (as shows for instance Example 6), will not produce contradictory conclusions in the case of consistent information items.

Proposition 1. *When $\cap_{i=1}^m H_m \neq \emptyset$, we have*

$$\text{Bel}(a_k \succ a_l) > 0 \Leftrightarrow \exists P_i \text{ s.t. } P_i(a_k, a_l)$$

Proof. Simply note that $\text{Bel}(a_k \succ a_l) > 0$ if and only if $P_i(a_k, a_l)$ for at least one focal element H_i . \square

This means that in case of consistency, the partial order produced by commonality maximization will be exactly the one produced by pairwise thresholding (see Example 6 to have an illustration). This also implies that in this case the thresholding approach will not contain cycles, as may happen in case of inconsistency. Proposition 1 also explains why, in the thresholding approach, we should start with a threshold equal to 0.

The next propositions show that, in case of consistency, the thresholding and scoring approaches provide coherent answers, in the sense that they cannot contradict each other.

Proposition 2. *When $\cap_{i=1}^m H_m \neq \emptyset$, we have*

$$\text{Bel}(a_k \succ a_l) > 0 \Rightarrow \bar{s}_k > \underline{s}_l$$

Proof. First, let us note that in case of consistency, having $\text{Bel}(a_k \succ a_l) > 0$ implies that there is no P_i with $P_i(l, k)$, and therefore that $\text{Bel}(a_l \succ a_k) = 0$ and $\text{Pl}(a_k \succ a_l) = 1$. We now need to prove that from $\text{Bel}(a_k \succ a_l) > 0$ follows that

$$\begin{aligned} \bar{s}_k &> \underline{s}_l \\ \sum_{j \neq k} \text{Pl}(a_k \succ a_j) &> \sum_{j \neq l} \text{Bel}(a_l \succ a_j) \\ \text{Pl}(a_k \succ a_l) + \sum_{j \neq k, l} \text{Pl}(a_k \succ a_j) &> \text{Bel}(a_l \succ a_k) + \sum_{j \neq l, k} \text{Bel}(a_l \succ a_j), \end{aligned}$$

where it follows from $\text{Bel}(a_k \succ a_l) > 0$ that $\text{Pl}(a_k \succ a_l) = 1$ and $\text{Bel}(a_l \succ a_k) = 0$, hence we can write

$$1 + \sum_{j \neq k, l} \text{Pl}(a_k \succ a_j) > \sum_{j \neq l, k} \text{Bel}(a_l \succ a_j)$$

Now, for a particular value of j in the summations we have

$$\begin{aligned} Pl(a_k \succ a_j) &> Bel(a_l \succ a_j) \\ \sum_{H_i, \neg P_i(j,k)} m(H_i) &> \sum_{H_i, P_i(l,j) \wedge \neg P_i(j,k)} m(H_i). \end{aligned}$$

Clearly the last inequality is true, as hypothesis H_i with $P_i(l, j) \wedge \neg P_i(j, k)$ are included in those where only $\neg P_i(j, k)$. That $Bel(a_l \succ a_j)$ is obtained by summing only the hypothesis where $P_i(l, j) \wedge \neg P_i(j, k)$ comes from the fact that no hypothesis can contain $P_i(l, j) \wedge P_i(j, k)$, as those would also contains $P_i(l, k)$ by transitivity (contradicting the hypothesis $Bel(a_k \succ a_l) > 0$). \square

Proposition 3. *When $\bigcap_{i=1}^m H_m \neq \emptyset$, we have*

$$\underline{s}_k > \bar{s}_l \Rightarrow Bel(a_l \succ a_k) = 0$$

Proof. We will proceed by contradiction. First, note that if $Bel(a_l \succ a_k) > 0$ (in the other case, the result immediately follows), we have

$$\begin{aligned} \underline{s}_k &> \bar{s}_l \\ \sum_{j \neq k} Bel(a_k \succ a_j) &> \sum_{j \neq l} Pl(a_l \succ a_j) \\ Bel(a_k \succ a_l) + \sum_{j \neq k, l} Bel(a_k \succ a_j) &> Pl(a_l \succ a_k) + \sum_{j \neq l, k} Pl(a_l \succ a_j) \\ \sum_{j \neq k, l} Bel(a_k \succ a_j) &> 1 + \sum_{j \neq l, k} Pl(a_l \succ a_j) \\ \sum_{j \neq k, l} Bel(a_k \succ a_j) &> n - 1 - \sum_{j \neq l, k} Bel(a_j \succ a_l) \end{aligned}$$

with the last inequality following from $Pl(a_l \succ a_j) = 1 - Bel(a_j \succ a_l)$. We then have

$$\sum_{j \neq l, k} (Bel(a_k \succ a_j) + Bel(a_j \succ a_l)) > n - 1.$$

Now, we do have $Bel(a_k \succ a_j) + Bel(a_j \succ a_l) \leq 1$. Indeed, $\sum_{H_i, P_i(k, j)} m(H_i) + \sum_{H_i, P_i(j, l)} m(H_i) \leq 1$, since the sets $\{H_i, P_i(k, j)\}$ and $\{H_i, P_i(j, l)\}$ are disjoint, else there would be a focal set H_i with $P_i(k, l)$ (by transitivity), contradicting our assumption. \square

These proposals clearly show that the three ranking approaches will not contradict one another in case of consistency, which is a reasonable minimal requirement.

3. FURTHER ILLUSTRATIVE EXAMPLES

In this section, we provide other examples showing how our approach can be applied to different multi-criteria models, starting from an absence of knowledge. In these two examples, we will keep the assumption that each criteria X^i is completely ordered, and for easiness will consider that when the spaces $\mathcal{X}^i = \{1, \dots, |\mathcal{X}^i|\}$

are finite sets, they will be encoded as natural numbers, with $\ell \prec_i k$ for any $\ell < k$. An alternative in a finite setting will then be a vector of C natural numbers.

To simplify our exposure, we will not deal with fully-fledged complex models such as GAI networks, AHP with hierarchy or CP-nets. Note that starting from a total absence of knowledge for such models means that the starting space H would be quite complex, and probably far too combinatorial (for instance, in GAI we would have to consider the set of all possible decompositions, in AHP all the possible hierarchies, and in CP-net all the possible graph structures and conditional orderings). In these cases, a more sensible scenario would be to start with a pre-specified model (for instance, assuming that the CP-net graph is known, or that the AHP hierarchy is specified).

3.1. Simple dominance rule. When each criteria is already ordered by preferences and can be so independently of each others, it is hard to argue against the Pareto ordering, that simply states that $a_k \succ_P a_l$ if $a_k^i \geq a_l^i$ for all $i \in \{1, \dots, C\}$ and $a_k^i > a_l^i$ for at least one criteria. If following this rule is the only modelling assumption we make, then a typical way to gather more information is to propose pairs of alternatives to the DM that are not already Pareto ordered, and then use this information to deduce further preferences.

In this case, the most natural information item \mathcal{I}_i would be to consider pairs (a_l, a_k) of preferred alternatives, with $a_l \succ a_k$. Let us now denote, given an alternative a , the subsets $a^+ = \{b \in \mathcal{X} : b \succ_P a\}$ and $a^- = \{b \in \mathcal{X} : b \prec_P a\}$ (Note that checking whether an element is in one of those subsets is straightforward). The corresponding subspace H_i would then be, in addition to the Pareto ordering, the fact that

$$H_i = \{a < b \ \forall a \in a_k^-, b \in a_l^+\}.$$

Combining these items of information would then just amounts to list all the pairs of compared alternatives so far, and check whether a preference on new instances can be deduced from the collected pairs. Note that other MCDA approaches adopts such a minimalist view, such aggregation-free proposals [31] based on logical rules, or Rough-set dominance models [29] that concentrate on the sorting problem. Studying such models is out of the scope of the current paper, but they are promising candidates to apply our approach.

Such a minimalist model (in terms of assumptions) is close to other models such as Rough-set dominance models [29] or aggregation-free proposals [31].

Example 8. Consider a 2-dimensional space with $\mathcal{X}^1 = \mathcal{X}^2 = \{0, \dots, 8\}$. The DM delivers the following information items

$$H_1 = \{(5, 2), (2, 6)\}, \quad H_2 = \{(3, 3), (6, 1)\}, \quad H_3 = \{(1, 5), (4, 4)\}$$

with the reliabilities $\alpha_1 = 0.4$, $\alpha_2 = \alpha_3 = 0.8$, as the DM may have some difficulties in comparing quite different alternatives, while being more at ease with comparing unbalanced alternatives with balanced ones. Figure 4 displays the corresponding regions in the bi-dimensional space.

Later on, we are provided with the new set of alternatives $\mathcal{A} = \{a_1 = (5, 1), a_2 = (3, 4), a_3 = (2, 5)\}$ (also noted on Figure 4), none of which is Pareto dominated by

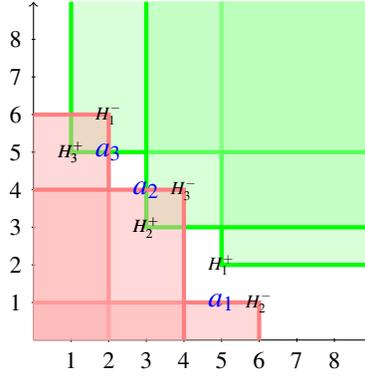


FIGURE 4. Dominance rule example.

another. Given \mathcal{A} and our information, we cannot deduce anything from H_1 , from H_2 we can deduce $a_1 \prec a_2$, as $a_1 \in H_2^-$, $a_2 \in H_2^+$, and from H_3 we can deduce $a_2 \prec a_3$. From the conjunction $H_2 \cap H_3$, we can deduce the ordering $a_1 \prec a_2 \prec a_3$. After combination of the information items, and focusing on the pair a_1, a_3 , we would have for example

$$Bel(a_1 \prec a_3) = 0.64 = m(H_1 \cap H_2 \cap H_3) + m(H_2 \cap H_3).$$

Again, the set of maximal elements would here be reduced to the single element a_3 , yet if we added the element $a_4 = (0, 7)$ to the list, both would be present in \hat{A} , according to Equation (5).

3.2. Lexicographic ordering. Lexicographic orderings [26] assume that all criteria X^i are totally ordered by an ordering \succ_{lex} , and that an alternative a_l is preferred to another a_k if a_l has a higher value than a_k on the first criteria on which they differ, according to \succ_{lex} . Provided the rankings of values for each criteria is known, learning a lexicographic order then amounts to determine the preferences between criteria. Such orderings are arguably among the simplest models, yet experiments suggest that they are frequently use to choose among equally costly alternatives [11]. Also note that lexicographic orderings (as simple dominance rule) are purely qualitative models, in the sense that they do not rely on any numerical scale, nor do use such scales when they are available.

In the case of lexicographic ordering, our space \mathcal{H} is simply the space of possible orderings between the criteria, and information items should therefore provide information about those. In contrast with the dominance rule of Section 3.1, there may be different ways to provide information about the underlying ordering, as one could directly ask whether a criteria is more important than one another, confront the DM with alternatives that differ only on pairs of criteria, or require the DM to specify what could be his top criteria (without necessarily requiring it to be a precise information).

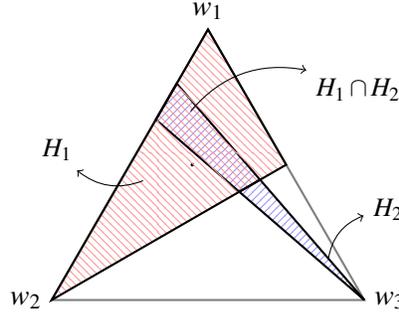


FIGURE 5. AHP example focal elements in barycentric coordinates.

Example 9. Let us assume that we have three criteria $\mathcal{X}^1 = \mathcal{X}^2 = \mathcal{X}^3 = \{1, 2, 3\}$. Then the DM may give us the two following information items:

$$\mathcal{I}_1 : X^1 \succ_{lex} X^3 \quad \mathcal{I}_2 : (1, 2, 2) \succ (1, 3, 1)$$

Assume we get $\alpha_1 = 0.8$ and $\alpha_2 = 0.6$ for the first and second information items, as comparing alternatives can be more difficult than providing definite statements about criteria. The first information item can directly be translated into the corresponding subspace $H_1 = \{X^1 \succ_{lex} X^3\}$, while the second $H_2 = \{X^2 \succ_{lex} X^3\}$ can be deduced from the fact that only criteria 2 and 3 differ (in opposite ways) in the two alternatives. We therefore have that $H_1 \cap H_2 = \{(X^2 \succ_{lex} X^3), (X^1 \succ_{lex} X^3)\}$, still not a fully specified model, as for instance under $H_1 \cap H_2$ the alternatives $(1, 3, 2)$ and $(3, 1, 2)$ would remain incomparable.

Note that many extensions of lexicographic orders have been proposed, among which are conditional lexicographic order [4] and preference trees [34]. In particular, such extensions could be used in the strategy and algorithm described in Section 4.1 further on to select adequate models.

3.3. An AHP minimal example. AHP is a popular MCDM method that, roughly speaking, elicit utilities and weights of a weighted averages by performing ratio comparisons of the weights and utilities, eventually doing so using a hierarchy between the criteria. As the original AHP methods uses precise assessments and precise models, it is not very well fitted to our current method. Yet, latter interval versions of the method [25] can be easily adapted to our framework.

In this example and to simplify our exposure, we will assume that utilities of the alternative criteria are directly given (as in the student example), and that we are trying to elicit the weights of each criterion. In this simple instance, AHP consists in asking ratio comparisons to the DM, in the form $a_{ij} \simeq w_i/w_j$. When a_{ij} is higher than one, this can be read “criterion X^i is a_{ij} times more important than X^j ”, which is indeed an intuitive way to express preferences for a DM. However, giving a set precise, meaningful and consistent a_{ij} values is almost impossible, and indeed one of the main issue of AHP is to make them consistent.

This changes when we allow for a_{ij} to become interval-values, i.e., when we allow the DM to state $w_i/w_j \in [\underline{a}_{ij}, \bar{a}_{ij}]$, in which case assessments such as “ X^i is at least twice as important as X^j ” translate into $[\underline{a}_{ij}, \bar{a}_{ij}] = [2, \infty]$. The next example illustrates the kinds of focal elements we would get in such a situation.

Example 10. As in Example 2, we will assume that we have three criteria X^1, X^2, X^3 , and that we are searching to obtain the values w_1, w_2, w_3 of a weighted average. The DM could for example provide us with the following items of information:

- \mathcal{I}_1 : “ X^1 is at least twice as important as X^2 , but less than four times as important”, which becomes $a_{12} \in [2, 4]$. As this comparison is quite precise, the DM may be quite uncertain about it, and give $\alpha_1 = 0.4$
- \mathcal{I}_2 : “ X^1 is more important than X^3 ”, which translates as $a_{13} \in [1, \infty]$, and as the DM is quite certain about this rather imprecise assessment, we could have $\alpha_2 = 0.9$

as we have that $\sum_i w_i = 1$ and $w_i \geq 0$, we can represent the associated focal elements in barycentric coordinates (as this kind of representation is more adapted to represent ratio comparisons), as shown in Figure 5. Computations are then similar to those performed in the weighted average example.

Note that in this example, there should be a direct connection between the certainty expressed by the DM and the width of intervals $[\underline{a}_{ij}, \bar{a}_{ij}]$. This would put a lot of strain on the DM, who would have to decide upon the interval length and the associated uncertainty degree. In practice, one could well fix the width of the intervals, and proceed with a fixed scale as the classical *AHP* does, or the other way around (fix a level of uncertainty and ask about the interval length).

4. INCONSISTENCY AS A USEFUL INFORMATION

So far, we have largely ignored the problem of dealing with inconsistent information, avoiding the issue of having a strictly positive $m(\emptyset)$. As mentioned in Section 2.2, this issue can be solved through the use of alternative combination rules, yet in the setting of preference learning, other treatments that we discuss in this section appear at least as interesting. These are, respectively, treatments selecting models of adequate complexity and selecting the “best” subset of consistent information. To illustrate our purpose, consider the following addition to the previous examples.

Example 11. Consider that in addition to previously provided information in Example 3, the DM now affirms us (with great certainty, $\alpha_3 = 0.9$) that the overall contribution of mathematics (X^1) should not be more than three tenths. In practice, if \mathcal{H} is the set of weighted means, this can be translated into $H_3 = \{(w_1, w_2) : w_1 \leq 3/10\}$. Figure 6 shows the situation, from which we get that H_1, H_2 and H_3 do not intersect, with $m(\emptyset) = 0.6 \cdot 0.9 \cdot 0.9 = 0.486$, a number high enough to trigger some warning.

4.1. Model selection. A possibility for $m(\emptyset)$ to be high is that the hypothesis space \mathcal{H} is not rich or complex enough to properly model a user preference. By considering more complex space \mathcal{H}' , we may decrease the value $m(\emptyset)$, as if $\mathcal{H} \subseteq \mathcal{H}'$,

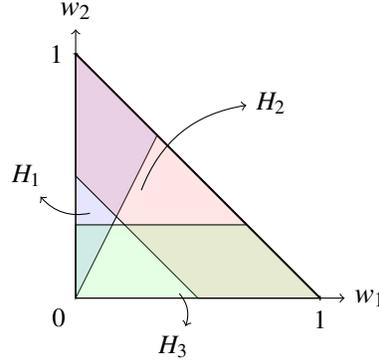


FIGURE 6. Inconsistent information items

we will have that for any information \mathcal{I}_i , the corresponding sets of models will be such that $H_i \subseteq H_i'$ (as all models from \mathcal{H} satisfying the constraints of \mathcal{I}_i will also be in \mathcal{H}'), hence we may have $H_i \cap H_j = \emptyset$ but $H_i' \cap H_j' \neq \emptyset$.

Example 12. Consider again Example 11, where \mathcal{H}' is the set of all 2-additive Choquet integrals. A 2-additive Choquet integral can be defined by a set of weights w_i and w_{ij} , $i \neq j$ where w_i and w_{ij} are the weights of groups of criteria $\{\mathcal{X}^i\}$ and $\{\mathcal{X}^i, \mathcal{X}^j\}$. The evaluation of alternatives for a 2-additive Choquet integral then simply reads

$$a_i = \sum_j w_j x_j + \sum_{j < k} w_{kj} \min(x_j, x_k).$$

For the evaluation function to respect the Pareto ordering, these weights must satisfy the following constraints

$$(11) \quad \begin{aligned} w_i &\geq 0 \text{ for all } i, \\ w_{ij} + w_i + w_j &\geq \max(w_i, w_j) \text{ for all } i, j, \\ \sum_i w_i + \sum_{ij} w_{ij} &= 1. \end{aligned}$$

Also, the contribution ϕ_i of a criterion i can be computed through the Shapley value

$$\phi_i = w_i + 1/2 \sum_{j \neq i} w_{ij}.$$

In the case of Example 11, this means that \mathcal{H} corresponds to the set of vectors (w_i, w_{ij}) that satisfy the constraints given by Equation (11). In this case, the information items H_1, H_2 provided in Example 2 and H_3 in Example 11 induce the

following constraints:

$$H_1 = \{\mathbf{w} \in \mathcal{H}' : 4w_2 + w_{23} \geq 8w_1 + 4w_{12} + 5w_{13}\}$$

$$H_2 = \{\mathbf{w} \in \mathcal{H}' : \phi_1 + \phi_2 \geq \phi_3\}$$

$$= \{\mathbf{w} \in \mathcal{H}' : w_1 + w_2 + w_{12} \geq w_3\}$$

$$H_3 = \{\mathbf{w} \in \mathcal{H}' : \phi_1 \leq 3/10\}$$

$$= \{\mathbf{w} \in \mathcal{H}' : w_1 + 1/2w_{12} + 1/2w_{13} \leq 3/10\}$$

These constraints are not inconsistent, as for example the solution where $w_1 = 0.15, w_2 = 0.3, w_3 = 0.45, w_{23} = 0.1$ are the only non-null values is within H_1, H_2 and H_3 . Among other things, this means that combining m_1, m_2, m_3 within the hypothesis space \mathcal{H}' leads to $m(\emptyset) = 0$

When considering a discrete nested sequence $\mathcal{H}^1 \subseteq \dots \subseteq \mathcal{H}^K$ of hypothesis spaces, then a simple procedure to select a model is to iteratively increase its complexity as summarised in Algorithm 1, where H_j^i is the set of possible hypothesis induced by information \mathcal{I}_j in space \mathcal{H}^j . It should be noted that the mass given to the empty set is guaranteed to decrease as the hypothesis spaces are nested. One could apply the same procedures to non-nested hypothesis spaces $\mathcal{H}^1, \dots, \mathcal{H}^K$ (e.g., considering lexicographic orderings and weighted averages), yet in this case there would be no guaranteed relations between the conflicting mass induced by each hypothesis spaces.

Algorithm 1: Algorithm to select preference model

Input: Spaces $\mathcal{H}^1 \subseteq \dots \subseteq \mathcal{H}^K$, Information $\mathcal{I}_1, \dots, \mathcal{I}_F$, threshold τ

Output: Selected hypothesis space \mathcal{H}^*

$i = 1$;

repeat

foreach $j \in \{0, \dots, m\}$ **do** Evaluate H_j^i ;

 Combine m_1^i, \dots, m_F^i into m^i ;

$i \leftarrow i + 1$

until $m^i(\emptyset) \leq \tau$ or $i = K + 1$;

4.2. Information selection. If we assume that the set \mathcal{H} is sufficiently rich to describe accurately the DM preferences, then it means that the mass $m(\emptyset)$ results from the fact that the DM has provided, at some point, erroneous information. In this case, it makes sense to discard those information items that are the most uncertain and introduce inconsistency in the results. In a short word, given a subset $S \subseteq \{1, \dots, n\}$, if we denote by m_S the mass obtained by combining the masses $\{m_i : i \in S\}$, then we can try to find the subset S such that

$$(12) \quad m_S(\emptyset) = 0$$

and whose amount of uncertainty $Cer(S) = \sum_{i \notin S} \alpha_i$ is minimal among the subsets inducing no conflict.

Remark 1. If the uncertainty degree is the same for all preferential information items (i.e., $\alpha_i = \alpha$ for all i), then minimising $Cer(S)$ while having $m_S(\emptyset) = 0$ comes down to pick the subset S counting the highest number of items. In other words, to pick the maximal subset of consistent sources, a typical way to resolve conflict in logic [42] that is also used in MCDA [38].

An easy, but potentially sub-optimal way to implement this strategy is to consider first the set $S^0 = \{1, \dots, n\}$, and then to iteratively consider subsets of immediate smaller size. In Example 11, this would have amounted to consider first $S^1 = \{2, 3\}$ (with $Cer(S^1) = 0.6$), then either $S^2 = \{1, 3\}$ or $S^2 = \{1, 2\}$ (with $Cer(S^1) = 0.9$). From Figure 6, we can however see that for $S = \{2, 3\}$, we already have $m_S(\emptyset) = 0$, and we do not need to go any further. When n is small enough (which is often the case if MCDA), then such an iterative, naive search may remain affordable. Improving upon it then depends on the nature of the space \mathcal{H} . It seems also fair to assume that the DM makes his/her best to be consistent, and therefore the number of information items to remove from $S^0 = \{1, \dots, n\}$ should be small in general.

A less complex alternative would be to suppress information items in a greedy way, i.e., removing iteratively the information item that would decrease the amount of conflict by the greatest value. This could be done by not only storing the result of focal element conjunctions, but also the sets whose intersection is taken. An efficient way, when using the conjunctive rule of belief functions, would then be to adapt the decombination operator proposed by Smets [45] using the canonical decomposition, that up to now requires a full enumeration of the subsets of space \mathcal{H} (which is impossible in our case, where \mathcal{H} is either continuous or of a combinatorial nature).

One can combine the two previously described approach, i.e., to first increase the model complexity if the conflict is too important at first, and then to eliminate those pieces of information that are the most unreliable and bring inconsistency in the result. There is clearly a balance between the two: increasing complexity allows us to keep all the gathered information but may lead us to over-fit the model or to computational problems, while letting go of some information allows us to reduce the computational burden, but will also deliver more conservative conclusions.

5. CONCLUSION

In this paper, we have described a generic way to handle uncertain preference information within the belief function framework. In contrast with previous works, our proposal is not tailored to a specific method but can handle a great variety of preference models. It is also consistent with the considered preference model, in the sense that if enough fully reliable information is provided, we retrieve a precise preference model.

Our proposal is quite general, and maybe more or less difficult to apply depending on the choice of \mathcal{H} . In the future, it would be interesting to study specific preference models and to propose efficient algorithmic procedures to perform the different calculi proposed in this paper. For instance, how do the computations

look like when we consider numerical models? Indeed, all procedures described in this paper can be applied to numerical as well as to non-numerical models, but numerical models may offer specific computational advantages (e.g., use of linear programming techniques).

On the other hand, one aspect of our proposal that may be criticized is the use of the conjunctive rule of combination, which assumes information items to be independent, meaning that two information items leading to the same subspace $H \subseteq \mathcal{H}$ will reinforce our belief in H being the right hypothesis, and that the number of focal elements will grow exponentially with the number of information items. One possible alternative, especially since we are working with simple support functions, would be to work in a possibilistic setting with a max/min-based aggregation, and to interpret degrees α_i as necessity degrees (see, for example, Dubois and Prade [23] for an introduction). This would also have the advantage that the provided degrees could be ordinal. The price to pay would then be to accept potentially much more conservative conclusions, as the fusion process would then be much more cautious. We would also be less consistent with Bayesian approaches, that also use product-based combination rules.

Finally, it seems quite interesting to explore how the current framework can be articulated with the recent trend in preference learning consisting of providing optimal queries for numerical models with a minimax regret strategy [2, 1, 39]. Indeed, our proposal could help to solve a current problem of such approaches, namely the fact that the DM can never contradict itself in such strategies. It is our belief that by adding uncertainty degrees, the DM could be led to revisit some previously refined spaces. However, articulating the two approaches would also require to solve more complex optimisation problems.

ACKNOWLEDGEMENTS

I wish to thank Olivier Cailloux, khaled Belahcene and Mylène Masson for earlier discussions about this work. This work was carried out in the framework of the Labex MS2T (Reference ANR-11-IDEX-0004-02) and of the ANR Project UMLNET (Reference ANR-14-CE24-0026), which were funded by the French Government.

REFERENCES

- [1] N. Benabbou, S. D. S. D. Diodoro, P. Perny, and P. Viappiani. Incremental preference elicitation in multi-attribute domains for choice and ranking with the borda count. In *Scalable Uncertainty Management - 10th International Conference, SUM 2016, Nice, France, September 21-23, 2016, Proceedings*, pages 81–95, 2016.
- [2] N. Benabbou, P. Perny, and P. Viappiani. Incremental elicitation of choquet capacities for multicriteria decision making. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, pages 87–92. IOS Press, 2014.

- [3] M. J. Beynon. Understanding local ignorance and non-specificity within the ds/ahp method of multi-criteria decision making. *European Journal of Operational Research*, 163(2):403–417, 2005.
- [4] R. Booth, Y. Chevaleyre, J. Lang, J. Mengin, and C. Sombattheera. Learning conditionally lexicographic preference relations. In *ECAI*, pages 269–274, 2010.
- [5] M. A. Boujelben, Y. De Smet, A. Frikha, and H. Chabchoub. A ranking model in uncertain, imprecise and multi-experts contexts: The application of evidence theory. *International journal of approximate reasoning*, 52(8):1171–1194, 2011.
- [6] C. Boutilier, R. I. Brafman, C. Domshlak, H. H. Hoos, and D. Poole. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *J. Artif. Intell. Res.(JAIR)*, 21:135–191, 2004.
- [7] S. J. Brams and P. C. Fishburn. Approval voting. *American Political Science Review*, 72(03):831–847, 1978.
- [8] O. Cailloux and S. Destercke. Reasons and means to model preferences as incomplete. In *International Conference on Scalable Uncertainty Management*, pages 17–30. Springer, 2017.
- [9] W. Cheng and K. Dembczynski. Label ranking methods based on the Plackett-Luce model. *Proceedings of the 27th Annual International Conference on Machine Learning - ICML '10*, 2010.
- [10] W. Cheng, E. Hüllermeier, W. Waegeman, and V. Welker. Label ranking with partial abstention based on thresholded probabilistic models. In *Advances in neural information processing systems*, pages 2501–2509, 2012.
- [11] A. M. Colman and J. A. Stirk. Singleton bias and lexicographic preferences among equally valued alternatives. *Journal of Economic Behavior & Organization*, 40(4):337–351, 1999.
- [12] S. Corrente, S. Greco, M. Kadziński, and R. Słowiński. Robust ordinal regression in preference learning and ranking. *Machine Learning*, 93(2-3):381–422, 2013.
- [13] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2008.
- [14] T. Denoeux, N. El Zoghby, V. Cherfaoui, and A. Jouglet. Optimal object association in the dempster-shafer framework. *IEEE transactions on cybernetics*, 44(12):2521–2531, 2014.
- [15] T. Denœux and M.-H. Masson. Evidential reasoning in large partially ordered sets. *Annals of Operations Research*, 195(1):135–161, 2012.
- [16] S. Destercke. A pairwise label ranking method with imprecise scores and partial predictions. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part II*, pages 112–127, 2013.
- [17] S. Destercke. A generic framework to include belief functions in preference handling for multi-criteria decision. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 179–189.

- Springer, 2017.
- [18] S. Destercke and T. Burger. Toward an axiomatic definition of conflict between belief functions. *IEEE transactions on cybernetics*, 43(2):585–596, 2013.
 - [19] S. Destercke and D. Dubois. Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. *Information Sciences*, 181(18):3925–3945, 2011.
 - [20] J. Dezert, D. Han, and H. Yin. A new belief function based approach for multi-criteria decision-making support. In *Information Fusion (FUSION), 2016 19th International Conference on*, pages 782–789. IEEE, 2016.
 - [21] L. C. Dias and V. Mousseau. Eliciting multi-criteria preferences: Electre models. In *Elicitation*, pages 349–375. Springer, 2018.
 - [22] J. Dombi, C. Imreh, and N. Vincze. Learning lexicographic orders. *European Journal of Operational Research*, 183(2):748–756, 2007.
 - [23] D. Dubois and H. Prade. Possibilistic logic: a retrospective and prospective view. *Fuzzy sets and Systems*, 144(1):3–23, 2004.
 - [24] A. Ennaceur, Z. Elouedi, and E. Lefevre. Reasoning under uncertainty in the ahp method using the belief function theory. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 373–382. Springer, 2012.
 - [25] T. Entani and M. Inuiguchi. Pairwise comparison based interval analysis for group decision aiding with multiple criteria. *Fuzzy Sets and Systems*, 274:79–96, 2015.
 - [26] P. C. Fishburn. Exceptional paperlexicographic orders, utilities and decision rules: A survey. *Management science*, 20(11):1442–1471, 1974.
 - [27] C. Gonzales and P. Perny. Gai networks for utility elicitation. *KR*, 4:224–234, 2004.
 - [28] M. Grabisch, I. Kojadinovic, and P. Meyer. A review of methods for capacity identification in choquet integral based multi-attribute utility theory: Applications of the kappalab r package. *European journal of operational research*, 186(2):766–785, 2008.
 - [29] S. Greco, B. Matarazzo, and R. Slowinski. Rough sets theory for multicriteria decision analysis. *European journal of operational research*, 129(1):1–47, 2001.
 - [30] E. Hüllermeier, J. Furnkranz, W. Cheng, and K. Brinker. Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172(16-17):1897–1916, 2008.
 - [31] S. Kaci and H. Prade. Constraints associated with choquet integrals and other aggregation-free ranking devices. In *Proceedings of IPMU*, volume 8, page 1345, 2008.
 - [32] M. Kadziński, T. Tervonen, and J. R. Figueira. Robust multi-criteria sorting with the outranking preference model and characteristic profiles. *Omega*, 55:126–140, 2015.
 - [33] J. Klein, S. Destercke, and O. Colot. Idempotent conjunctive and disjunctive combination of belief functions by distance minimization. *International*

- Journal of Approximate Reasoning*, 92:32–48, 2018.
- [34] X. Liu and M. Truszczynski. Learning partial lexicographic preference trees over combinatorial domains. In *AAAI*, volume 15, pages 1539–1545, 2015.
 - [35] T. Lust. Choquet integral versus weighted sum in multicriteria decision contexts. In *International Conference on Algorithmic Decision Theory*, pages 288–304. Springer, 2015.
 - [36] M. Masson, S. Destercke, and T. Denoeux. Modelling and predicting partial orders from pairwise belief functions. *Soft Comput.*, 20(3):939–950, 2016.
 - [37] P. Meyer and M. Pirlot. On the expressiveness of the additive value function and the choquet integral models. In *DA2PL 2012: from Multiple Criteria Decision Aid to Preference Learning*, pages 48–56, 2012.
 - [38] V. Mousseau, L. C. Dias, and J. Figueira. Dealing with inconsistent judgments in multiple criteria sorting models. *4OR*, 4(2):145–158, 2006.
 - [39] P. Perny, P. Viappiani, and A. Boukhatem. Incremental preference elicitation for decision making under risk with the rank-dependent utility model. In *Uncertainty in Artificial Intelligence*, 2016.
 - [40] F. Pichon, S. Destercke, and T. Burger. A consistency-specificity trade-off to select source behavior in information fusion. *IEEE transactions on cybernetics*, 45(4):598–609, 2015.
 - [41] M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh. Aggregating partially ordered preferences: impossibility and possibility results. In *Proceedings of the 10th conference on Theoretical aspects of rationality and knowledge*, pages 193–206. National University of Singapore, 2005.
 - [42] N. Rescher and R. Manor. On inference from inconsistent premisses. *Theory and decision*, 1(2):179–217, 1970.
 - [43] G. Shafer. *A mathematical Theory of Evidence*. Princeton University Press, New Jersey, 1976.
 - [44] P. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on pattern analysis and machine intelligence*, 12(5):447–458, 1990.
 - [45] P. Smets. The canonical decomposition of a weighted belief. In *Proc. Int. Joint. Conf. on Artificial Intelligence*, pages 1896–1901, Montreal, 1995.
 - [46] P. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8:387–412, 2007.
 - [47] P. Smets and R. Kennes. The transferable belief model. *Artificial intelligence*, 66(2):191–234, 1994.
 - [48] P. Viappiani and C. Boutilier. Optimal bayesian recommendation sets and myopically optimal choice query sets. In *Advances in neural information processing systems*, pages 2352–2360, 2010.
 - [49] J.-B. Yang and D.-L. Xu. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32(3):289–304, 2002.