Artificial Intelligence 175 (2011) 1410–1448

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

A general framework for explaining the results of a multi-attribute preference model

Christophe Labreuche

Thales Research & Technology, RD128, 91767 Palaiseau cedex, France

ARTICLE INFO

Article history:
Received 27 February 2009
Received in revised form 13 August 2010
Accepted 13 August 2010
Available online 1 December 2010

Keywords:
Preferences
Decision theory
Argumentation
Weight

ABSTRACT

The automatic generation of an explanation of the prescription made by a multi-attribute decision model is crucial in many applications, such as recommender systems. This task is complex since the quantitative models are not designed to be easily explainable. The major limitation of the previous research is that there is no formal justification of the arguments that are selected in the explanation. The goal of this paper is to define a general framework to justify which arguments shall be selected, in the case where the decision model is based on weights assigned to the attributes. Due to the complexity of explaining a preference model based on utility theory, several explanation reasonings are necessary to cover all cases – ranging from situations where the prescription is trivial to situations where the prescription is much more tight. The set of selected arguments is, in this framework, a non-dominated element of a combinatorial structure in the sense of an order relation. Our general approach is instantiated precisely on three models: the probabilistic expected utility model, the qualitative pessimistic minmax model and the concordance rule, which are all constructed from a weight vector.

1. Introduction

In decision making under uncertainty, social choice and multi-criteria decision making, which are the three main domains of decision theory [40,48], explicit analytical models are constructed to represent the preferences of a decision maker regarding how to combine various dimensions, which are the states of nature, the voters and the criteria respectively. Decision theory mainly focuses on specifying how a rational agent should behave, which results in the justification, through axiomatic characterizations, of the decision models that should be used [54,51,37,41]. Another well-developed research area in decision theory concerns the elicitation of the decision maker preferences, and has led to the design of elaborate elicitation methods.

Decision models are traditionally mainly quantitative, which is the case for instance of the expected utility model [54, 51]. More recently, qualitative models have been developed in AI in order to overcome the difficulty of the elicitation of the information necessary for these models [24,19,12,20]. A wide class of quantitative and qualitative models are parameterized by a weight vector where a weight is assigned to each dimension [26,49]. We are interested in these models in this paper.

The final part of the decision process, after the model has been elicited, is usually not studied in decision theory. It is generally reduced to the application of the decision model on the options of interest. If an individual constructs his decision model and is convinced about its relevance, then it is not necessary to explain to him the result of the application of the decision model. However there are many practical situations in which the decision needs to be justified to some actors who did not participate to its construction. These actors are not interested in the technicality of the decision model. On the

E-mail address: christophe.labreuche@thalesgroup.com.

0004-3702/$ – see front matter
doi:10.1016/j.artint.2010.11.008

contrary, they wish to have a synthetic explanation of the decision. It shall be automatically generated. According to Klein [35], such an explanation should be *intuitive*, *comprehensive* and *persuasive*.

The automatic generation of an explanation of the outcome of a decision model is not an easy task since the models from decision theory are not designed to provide the reasons that support the recommendation [35]. By contrast, decision frameworks from AI have, by construction, the ability to naturally provide such an explanation. This is for instance the case of the belief–desire–intention representation architecture [11], argumentation [25] and conditional preference networks [9]. In [18,3,2], an argumentation-based framework for making and explaining a decision is proposed. A preference relation over the candidate options is derived from the positive and negative acceptable arguments that support each option and from an ordering on the arguments. Another extension of argumentation to incorporate preferences can be found in [42].

In the context of multi-criteria decision making and more specifically multi-attribute value theory (MAVT) [26,34,30], there are a few works which aim at generating an explanation of the outcome of the decision model [35,14,43]. The generation process is split into two parts: the selection of the arguments (i.e. the criteria) to be presented, and the structuration and expression of the selected arguments in natural language. The second part is well-developed in Ref. [14]. One can note that the expression of a selected content in natural language has been extensively studied in the literature – see for instance [13,27] to cite a few. Concerning the selection part, the three previously mentioned works [35,14,43] use the same idea. It consists in selecting the k (where k is a parameter) criteria that have the largest contribution to the overall utility. This approach is not satisfactory for the following reasons. First of all, the textual explanation does not mention the weights of the criteria [14]. This is a major drawback since the weights are essential in the MAVT model. Secondly, there is no formal justification of the arguments that are selected, and in particular of the choice of the k parameter.

The aim of this paper is to develop a formal framework that justifies the selection of the arguments. This work is especially dedicated to situations in which the recipient of the explanation is not the individual who has designed the decision model. This general framework is designed for any decision model based on weights. It adapts itself automatically to the complexity of the decision. The easier the decision, the simpler the explanation. Premises of this work can be found in two conference papers [38,39].

Section 2 describes weighted decision models. We introduce three particular models that will be used to validate our framework: the expected utility model [54,51], the weighted minmax model [22] and the weighted majority model [44]. The general explanation framework is presented in Section 3. In order to adapt to the complexity of the situation, several argumentation reasonings are introduced. They are called *anchors* by analogy to the concept of anchor defined by Grize to refer to some implicit information used to convince an audience [32]. A subset of criteria can be selected for the explanation if these criteria are decisive in some sense depending on the anchor. Such a subset is called an *explanation set*. The set of these explanation sets forms a combinatorial structure. One then aims at finding the non-dominated explanation sets in this structure, in the sense of an order relation expressing the simplicity of the explanation set. In the following Sections 4 through 7, this general framework is thoroughly developed on each anchor and each of the three weighted decision models. From the properties satisfied by the non-dominated explanation sets, we show that the explanation to be generated can be derived. A method or algorithm to compute a non-dominated explanation set is given in each case. We will not deal with the expression of the selected arguments in natural language. However, examples of texts that can be generated will be presented. Some experimental results are presented in Section 8. The proofs of the results are given in Section 10.

2. Decision models based on a weight vector

Decision theory [40,48] is interested in preference representation and gathers different domains such as Multi-Criteria Decision Making (MCDM) [49,28,10], Decision Making under Uncertainty (DMU) [36,51] and Social Choice (SC) [6,29]. The typical decision problem studied in decision theory consists in selecting one alternative among a set X of candidate options, where the alternatives are described by several dimensions. This selection is obtained by the construction of a preference relation over X. The set of finite dimensions is denoted by $N = \{1, \ldots, n\}$, and the alternatives are characterized on each dimension $i \in N$ by a value in a set X_i. In MCDM, N is the set of decision criteria, X_i is the attribute representing criterion i, and each alternative is characterized by a value on each attribute, that is $X = X_1 \times \cdots \times X_n$. The criteria are often conflicting [49]. As an example, one may have cost criteria and performance criteria, which cannot be met at the same time. The main difficulty is then to find a good compromise between the criteria. In DMU, the elements of N are the states representing the possible situations, $X_1 = \cdots = X_n =: C$ is the set of possible consequences, and an alternative (also called act) is a mapping from N to C, that is $X = C^N$. The consequence of selecting a particular alternative depends on which state of nature will occur. Moreover the attitude of the decision maker (DM) towards uncertainty influences his choice strategy [51]. In SC, N is the set of voters, $X_1 = \cdots = X_n =: C$ is the set of candidates, and the alternatives are also the candidates, that is $X = C$. The difficulty is to find a fair consensus among the opinions of the voters [5].

The preference relation over the alternatives in X is usually constructed from a preference relation $\succeq_i$ over each set X_i. We denote by $\succ_i$ and $\sim_i$ the asymmetric and symmetric parts of $\succeq_i$. In MCDM, $\succeq_i$ represents the preferences of the DM over the elements of attribute X_i; in DMU, $\succeq_1 = \cdots = \succeq_n$ depict the preferences of the DM over the set C of consequences; in SC, $\succeq_i$ models the preferences of voter i over the set C of candidates. There exist many preference models in which the overall preference of an alternative $y \in X$ over another alternative $x \in X$ is obtained by weighing up the pros and the cons regarding this preference. The sets $A^+(y,x) = \{i \in N,\ y_i \succ_i x_i\}$, $A^-(y,x) = \{i \in N,\ x_i \succ_i y_i\}$ and $A^=(y,x) = \{i \in N,\ y_i \sim_i x_i\}$ are the positive, negative and null arguments respectively concerning the preference of y over x on all dimensions N. The

arbitrage between $A^+(y,x)$ and $A^-(y,x)$ is often based on the priorities allotted to the dimensions. This is quantified by a weight w_i assigned to each $i \in N$ together with an aggregation function depending on $w = (w_1, \ldots, w_n)$ [26,49,51]. The semantics of the weights depends on the method used. Basically, w_i is interpreted as the importance of criterion i in MCDM, as the probability or possibility of the state of nature i in DMU, and as the power of voter i in SC. The weight vector w is normalized in some sense depending on each method used. We denote hereafter by $\wedge$ and $\vee$ the min and max operators respectively.

We are interested in several families of preference relations characterized by a weight vector $w = (w_1, \ldots, w_n)$. For a given family $\mathcal{F}$ of preference models, the set of weights is denoted by $\mathcal{W}(\mathcal{F})$, and the corresponding preference relation is denoted by $\succeq_w^{\mathcal{F}}$, for $w \in \mathcal{W}(\mathcal{F})$. We denote by $\succ_w^{\mathcal{F}}$ and $\sim_w^{\mathcal{F}}$ the asymmetric and symmetric parts of $\succeq_w^{\mathcal{F}}$. We are interested in the models derived from an aggregated value:

$$\forall x, y \in X \quad y \succeq_w^{\mathcal{F}} x \iff H_w^{\mathcal{F}}(y,x) \geqslant 0.$$

A classical representation is to summarize each option $x \in X$ by an overall utility $h_w^{\mathcal{F}}(x)$. This gives $H_w^{\mathcal{F}}(y,x) = h_w^{\mathcal{F}}(y) - h_w^{\mathcal{F}}(x)$.

The most well-known family corresponds to the expected-utility model (labelled "EU" hereafter) of von Neumann and Morgenstern [54] and Savage [51]: $h_w^{\mathrm{EU}}(x) = \sum_{i \in N} w_i u_i(x_i)$, where $w_i \in [0,1]$ and $u_i : X_i \to [0,1]$ is a value function measuring the attractiveness of the elements of X_i. The value function u_i quantifies the preference relation $\succeq_i$: for $a, b \in X_i$, $a \succeq_i b \iff u_i(a) \geqslant u_i(b)$. In MCDM, this model is the Multi-Attribute Value Theory Model [26,34]. One has $\mathcal{W}(\mathrm{EU}) = [0,1]^n$.

There exist accurate elicitation methods to construct the quantitative model h_w^{EU} [7,50]. The main drawback of these methods is their complexity, which does not fit with all applications. Qualitative decision theory has been defined in AI to overcome this difficulty [24,19,12,20]. A pessimistic weighted extension (labelled "Pess" hereafter) of the Wald minmax function [56] has been defined in [22]: $h_w^{\mathrm{Pess}}(x) = \bigwedge_{i \in N}(u_i(x_i) \vee (1 - w_i))$, where w is interpreted as a possibility distribution in DMU [23]. This model has been also used in MCDM. One has $\mathcal{W}(\mathrm{Pess}) = [0,1]^n$. For simplicity, we consider the interval $[0,1]$ also for the Pess model but any linearly ordered scale with top and bottom elements can be used as well.

The EU and Pess models both require the existence of a value representation for each $\succeq_i$ and commensurateness between the preference scale and the weight/uncertainty scale, which is usually not easy to satisfy in practice. The majority rule (labelled "Maj" hereafter) defined in SC [44] gets rid of these two assumptions. It has also been studied in AI in a more general framework [20]. It reads $H_w^{\mathrm{Maj}}(y,x) = \sum_{i \in A^+(y,x)} w_i - \sum_{i \in A^-(y,x)} w_i$. This model has some limitations due to the Arrow's theorem [5]. The majority rule is known in MCDM under the name *concordance rule* [49]. One has $\mathcal{W}(\mathrm{Maj}) = \mathbb{R}_+^n$.

Despite some limitations, the three models EU, Pess and Maj that we have just described are used in many applications covering very different domains. For this reason, we will focus on these models.

The set of *normalized* weights w.r.t. a model $\mathcal{F}$ is denoted by $\overline{\mathcal{W}}(\mathcal{F})$. For the EU and Maj models, we have $\overline{\mathcal{W}}(\mathrm{EU}) = \overline{\mathcal{W}}(\mathrm{Maj}) = \{w \in [0,1]^n : \sum_{i \in N} w_i = 1\}$. Let us mention that, for the EU model, the normalization condition $\sum_{i \in N} w_i = 1$ is equivalent to the idempotency property $h_w^{\mathrm{EU}}(\alpha, \ldots, \alpha) = \alpha$ for all $\alpha \in [0,1]$ [31]. The weights are normalized for the model Pess if $\bigvee_{i \in N} w_i = 1$ and thus $\overline{\mathcal{W}}(\mathrm{Pess}) = \{w \in [0,1]^n : \bigvee_{i \in N} w_i = 1\}$ [22]. For each of the three above models, there exists a particular normalized weight vector $w^{\mathcal{F}}$ that is characterized by the property that the dimensions are symmetric in the aggregation process. Therefore $w_1^{\mathcal{F}} = \cdots = w_n^{\mathcal{F}}$. These are the weights that one will apply in the absence of information. They results from the application of the principles of insufficient reason, maximum entropy and minimum specificity. In the Bayesian approach, the lack of information is represented by the uniform probability $w^{\mathcal{F}}$, and in MCDM and SC, all criteria and voters are assigned to the same weight when there is no reason to proceed differently. The vector $w^{\mathcal{F}}$ will thus be called *reference weight vector*. For models Maj and EU, we obtain for all $i \in N$

$$w_i^{\mathrm{Maj}} = w_i^{\mathrm{EU}} = \frac{1}{n}. \tag{1}$$

$\succeq_{w^{\mathrm{Maj}}}^{\mathrm{Maj}}$ corresponds to a Condorcet majority rule, and $h_{w^{\mathrm{EU}}}^{\mathrm{EU}}$ is the arithmetic mean. Let $w \in \overline{\mathcal{W}}(\mathrm{EU})$ or $w \in \overline{\mathcal{W}}(\mathrm{Maj})$. Since $\sum_{i \in N} w_i = 1$, the value $w_i^{\mathrm{Maj}} = w_i^{\mathrm{EU}} = \frac{1}{n}$ corresponds to the mean importance of a dimension in w. Hence relation $w_i > \frac{1}{n}$ (resp. $w_i < \frac{1}{n}$) means that dimension i is more important (resp. less important) than the average value $\frac{1}{n}$ and thus can be said to be an *important* (resp. *unimportant*) dimension.

For model Pess, one has for all $i \in N$

$$w_i^{\mathrm{Pess}} = 1 \tag{2}$$

and $h_{w^{\mathrm{Pess}}}^{\mathrm{Pess}}$ is the min operator.

3. General explanation framework

3.1. Why generating automatically an explanation?

From now on, we assume that $\mathcal{F} \in \{\mathrm{EU}, \mathrm{Pess}, \mathrm{Maj}\}$ is a fixed family of models, and the weight vector $v \in \overline{\mathcal{W}}(\mathcal{F})$ has already been specified. Let

$$\mathcal{D}(\mathcal{F}) = \left\{(x, y, v) \in X \times X \times \overline{\mathcal{W}}(\mathcal{F}):\ y \succ_v^{\mathcal{F}} x\right\}.$$

We consider two options $x, y \in X$ and we assume that $(x, y, v) \in \mathcal{D}(\mathcal{F})$. We aim at explaining why y is strictly preferred to x according to the model.

If an individual constructs his decision model $\succ_v^{\mathcal{F}}$ and is convinced about its relevance, then it is not necessary to explain to him the comparison $y \succ_v^{\mathcal{F}} x$. This is no more true when several actors are involved in the decision process, as shown in the following examples.

(i) The individual that is referred to as *decision maker* (DM) is usually the person who is responsible for the decision. He often has to explain his decision to other actors – for instance, his chief, his executive board or his shareholders. These actors are not interested in the technicality of the decision model. In order to convince them on the merits of the decision, a synthetic explanation needs to be given.

(ii) A decision support system usually generates an ordered list of recommended options from which the user of the system decides which one to choose. Since the user is not the expert from whom the decision model has been elicited, the recommendation shall be explained. Moreover, the user is not necessarily highly qualified, and there may be time constraints under which the operator needs to make the decisions. Hence the explanation shall be non-technical, simple and fast to understand for the user. An example of this is the multi-criteria decision function that recommends the assignment of priorities to the radar tasks in radar management [8].

(iii) There are many situations in AI in which several (artificial) agents have their own preferences and thus their own decision models. This is for instance the case in negotiation. In negotiation protocols, an offer is made by an actor at each iteration of the protocol, and the other actors give their feedback [45,1]. They do not reveal their preference models to the other actors since they want to keep their models private. Each actor only provides to the other ones some clues on why the offer is, for instance, not satisfactory to him.

In the previous examples, an explanation of the recommendations produced by the decision model must be generated. According to Klein [35], the explanation cannot be simply that, for the EU model, "*y is preferred to x since the overall score* $\sum_{i\in N} v_i y_i$ *of y is larger than the overall score* $\sum_{i\in N} v_i x_i$ *of x*" since it does not *appeal to intuition* and it does not compare the alternatives explicitly.

Yet, in the examples (ii) and (iii) above, a synthetic explanation needs to be automatically generated and it cannot be produced by a human. When the number of criteria is relatively large, the cognitive load to interpret the figures x, y and v is relatively high. This load is too high for the situations in which the decision activity is very repetitive, as it is the case in the radar management example (ii) in which the operator has to perform the same activity for several hours in a raw. In example (iii), the values of v cannot be sent to the other agents for privacy reasons, and a synthetic explanation must be generated by the artificial agent.

The arguments of this explanation are the elements of N. The explanation will be based on the satisfaction degrees $u_i(x)$ and $u_i(y)$ of x and y on each dimension (for the models EU and Pess for which value functions u_i exist). Hence the precise form and expression of the value functions u_i will not matter. As a result, for the sake of simplicity in the notation, we will assume throughout the paper that the value functions u_i are the identity function. Hence one can set $X = [0, 1]^n$ for the models EU and Pess. Our framework will apply on the three domains MCDM, DMU and SC. However, for the sake of conciseness, we now restrict to MCDM for the interpretations of the results. In particular, from now on, N will correspond to decision criteria. For $x \in X$, x_i is the mark or score of x on criterion i. For the models EU and Pess, $x_i \in [0, 1]$ is interpreted as a satisfaction degree, where value 1 is perfectly satisfactory on the criterion and value 0 is unacceptable. Our explanation framework can be transposed to DMU and SC.

3.2. Notation and definitions

For $\mathcal{F} \in \{\text{EU}, \text{Pess}\}$, we have $X = [0, 1]^n$ and we define $\Delta \in \mathbb{R}^n$ by $\Delta_i = y_i - x_i$ for all $i \in N$. For $\mathcal{F} = \text{Maj}$, we define $\text{sgn} \in \{-1, 0, 1\}^N$ by $\text{sgn}_i := +1$ if $i \in A^+(y, x)$, $\text{sgn}_i := 0$ if $i \in A^=(y, x)$ and $\text{sgn}_i := -1$ if $i \in A^-(y, x)$. For $S \subseteq N$ and $Z \in \mathbb{R}^N$, we define $\pi_S^Z : \{1, \ldots, s\} \to S$, with $s = |S|$, by $Z_{\pi_S^Z(1)} \leqslant \cdots \leqslant Z_{\pi_S^Z(s)}$. Throughout this paper, we will apply this definition to the vectors Δ, x, y, v leading to π_S^Δ, π_S^x, π_S^y,π_S^v respectively. Let $\Pi(S)$ be the set of all permutations on a coalition $S \subseteq N$. For $\pi \in \Pi(N)$ and $w \in \mathcal{W}(\mathcal{F})$, $\pi \circ w$ is the weight vector defined by $(\pi \circ w)_i = w_{\pi(i)}$ for all $i \in N$. For $w, w' \in \mathcal{W}(\mathcal{F})$ and $S \subseteq N$, we define the compound weight vector $(w_S, w'_{N\setminus S}) \in \mathcal{W}(\mathcal{F})$ by $(w_S, w'_{N\setminus S})_i = w_i$ if $i \in S$, and $(w_S, w'_{N\setminus S})_i = w'_i$ otherwise.

Let

$$Ex = \left\{\{A_1, \ldots, A_p\}:\ p \in \mathbb{N},\ \emptyset \neq A_1, \ldots, A_p \subseteq N \text{ and } A_i \cap A_j = \emptyset \text{ for all } i \neq j\right\}.$$

For $\mathcal{A}, \mathcal{A}' \in Ex$, we write $\mathcal{A} \sqsubseteq \mathcal{A}'$ if for every $A \in \mathcal{A}$, there exists $A' \in \mathcal{A}'$ such that $A \subseteq A'$. We write $\mathcal{A} \subseteq \mathcal{A}'$ if for every $A \in \mathcal{A}$, we have $A \in \mathcal{A}'$.

For any permutation $\pi \in \Pi(N)$, one can determine the finest partition (in the sense of $\sqsubseteq$) $\mathcal{A}(\pi) := \{A_1, \ldots, A_p\}$ of N such that all A_i are invariant under π (i.e. $\pi(A_i) = A_i$). For all $i \in \{1, \ldots, p\}$, one can write

$$k_2 = \pi(k_1), \quad k_3 = \pi(k_2), \quad \ldots, \quad k_{q_i} = \pi(k_{q_i-1}) \quad \text{and} \quad k_1 = \pi(k_{q_i})$$

where $A_i = \{k_1, \ldots, k_{q_i}\}$ and $q_i = |A_i|$. By abuse of language, A_i is thus called a *cycle* of π.

One can give a taxonomy of the arguments from their sign and strength. The sign of an argument follows from that of Δ_i (see Section 2). The strength of an argument is related to the sign of $v_i - \frac{1}{n}$ (for the EU and Maj models): an argument is *strong*, *medium* and *weak* if $v_i \gg \frac{1}{n}$, $v_i \approx \frac{1}{n}$ and $v_i \ll \frac{1}{n}$ respectively.

3.3. Motivation of the approach on the EU model

To give the general idea of our approach, let us first focus on the EU model.

Ideally, one would like to produce an argumentation of the relation $y \succ_v^{\text{EU}} x$. There are many difficulties to extend argumentation to a quantitative setting. In logic-based argumentation, an argument is a pair $\langle H, h\rangle$ where h is the conclusion and H contains the minimal elements of the knowledge base that logically entails h (i.e. $H \vdash h$) [52]. Here one would like to construct such an argument where h is the statement $y \succ_v^{\text{EU}} x$ and $H \subseteq N$ is a subset of criteria. Intuitively, *the criteria H explain the decision if the decision remains unchanged whatever the values of x and y on the criteria in $N \setminus H$*. It writes

$$H \vdash \left[y \succ_v^{\text{EU}} x\right] \iff \forall x'_{N\setminus H}, y'_{N\setminus H} \in [0,1]^{N\setminus H} \quad \left(y_H, y'_{N\setminus H}\right) \succ_v^{\text{EU}} \left(x_H, x'_{N\setminus H}\right).$$

This is equivalent to $(y_H, 0_{N\setminus H}) \succ_v^{\text{EU}} (x_H, 1_{N\setminus H})$. This condition is too strong as it is often satisfied only when H is almost equal to N. The reason is that all criteria compensate each other in the EU model (see Section 2) and it is rare that a criterion is completely useless in the decision. The difficulty mentioned earlier is also expressed by the *fallacy of composition/division* which applies on divisible objects defined as the concatenation of smaller parts [33]. Quantitative models are indeed characterized by the presence of several concatenation operators. From a measurement standpoint, the value functions u_i are obtained from a quantitative scale, and this latter can be constructed from a concatenation operator over X_i and a preference relation [37]. The aggregation of the marks by h_v^{EU} also results from a concatenation materialized by an arithmetic operator.

Let $\pi \in \Pi(N)$ such that

$$v_{\pi(1)}|\Delta_{\pi(1)}| \leqslant \cdots \leqslant v_{\pi(n)}|\Delta_{\pi(n)}|.$$

The explanations of the preference $y \succ_v^{\text{EU}} x$ proposed in [35,14,43] are similar and consist in displaying to the user the criteria $\{\pi(q), \pi(q+1), \ldots, \pi(n)\}$ (where $q \in \{1, \ldots, n\}$ is a parameter). This idea of simplifying the explanation reasoning by focusing only on a subset of the arguments is originated from the theory of argumentation. The quantity $\text{compel}_i := v_i|\Delta_i|$ measuring the contribution of criterion i in the overall evaluation $H_v^{\text{EU}}(y, x)$ is called *compellingness* by Klein [35]. There are two main limitations of this approach. Firstly, there is no formal justification why selecting these particular criteria in the explanation. Secondly, the textual explanation that is generated does not refer to the weight vector v. Yet the behaviour of the decision model is highly influenced by its weight vector v, and thus the explanation of the preference $y \succ_v^{\mathcal{F}} x$ should mention the weights.

The weights play a central role in our approach. However, as shown in the following example, there are circumstances in which it is not necessary to mention the specificity of the weights in the explanation.

Example 1. The most simple situation arises when y is strictly better than x on all criteria. There is no negative and null argument. Clearly $y \succ_v^{\mathcal{F}} x$ for all $\mathcal{F} \in \{\text{EU}, \text{Pess}, \text{Maj}\}$ and all $v \in \overline{\mathcal{W}}(\mathcal{F})$. There is no need in this trivial situation to mention the weights v in the explanation.

In Example 1, any $w \in \overline{\mathcal{W}}(\mathcal{F})$ yields the same comparison of y and x. This suggests to look, in the general case, at the set $\mathcal{V}^{\mathcal{F}}(y, x)$ of weights $w \in \overline{\mathcal{W}}(\mathcal{F})$ for which y is strictly preferred to x. In the following lemma, we restrict ourselves to the EU model.

Lemma 1. *Let $\mathcal{V}^{\text{EU}}(y, x) := \{w \in \overline{\mathcal{W}}(\text{EU}):\ y \succ_w^{\text{EU}} x\}$. Let $\Delta_i^+ = \Delta_i \vee 0$ and $\Delta_i^- = (-\Delta_i) \vee 0$ for all $i \in N$. If $A^+(y, x) \neq \emptyset$ and $A^-(y, x) \neq \emptyset$, then*:

$$\text{If } \left|A^+(y,x)\right| > 1 \text{ then } \forall i \in A^+(y,x) \quad \left\{w_i,\ w \in \mathcal{V}^{\text{EU}}(y,x)\right\} = [0,1],$$

$$\text{If } \left|A^+(y,x)\right| = 1 \text{ then } \forall i \in A^+(y,x) \quad \left\{w_i,\ w \in \mathcal{V}^{\text{EU}}(y,x)\right\} = \left(\min_{j \in A^-(y,x)} \frac{\Delta_j^-}{\Delta_i^+ + \Delta_j^-}, 1\right],$$

$$\forall j \in A^-(y,x) \quad \left\{w_j,\ w \in \mathcal{V}^{\text{EU}}(y,x)\right\} = \left[0, \max_{i \in A^+(y,x)} \frac{\Delta_i^+}{\Delta_i^+ + \Delta_j^-}\right).$$

It is not easy to construct from the intervals obtained in Lemma 1 a general explanation, except for some very particular cases.

As said in Section 3.1, the recipient of the explanation we wish to generate is supposed to have no prior on the specificities of the decision model and thus on the weight vector v. Anyway, if the recipient wishes to have a prior idea of the decision only from x and y, he will consciously or unconsciously use the less specific weights – i.e. the reference weight w^{EU} – on the options x and y. Hence if x has much better marks than y on average, i.e. x is much preferred to y according to $\succeq^{\mathrm{EU}}_{w^{\mathrm{EU}}}$, then the recipient would a priori expect that x is preferred to y. He may thus be surprised by the relation $y \succ^{\mathrm{EU}}_{v} x$. Hence the explanation shall focus on explaining why the weights v and w^{EU} yield opposite decisions.

Example 2. Assume that $x = (0.7, 0.7, 0.5)$, $y = (1, 0, 0.5)$ and $v = (0.7, 0.2, 0.1)$, which gives $y \succ^{\mathrm{EU}}_{v} x$ and $x \succeq^{\mathrm{EU}}_{w^{\mathrm{EU}}} y$. Hence, the weights v play here an important role in the outcome $y \succ^{\mathrm{EU}}_{v} x$. One feels that since x is better than y on average but y is preferred to x with the weight vector v, then the criteria for which y is better than x are important and the criteria for which x is better than y are not important. Recall that a criterion i is important (resp. unimportant) if $v_i > \frac{1}{n}$ (resp. $v_i < \frac{1}{n}$). One has $A^+(y, x) = \{1\}$, $A^-(y, x) = \{2\}$ and $A^=(y, x) = \{3\}$. The weight vector v is important on the positive argument (the first criterion) and is not important on the negative argument (the second criterion). The conjunction of these two phenomena actually explains the decision. Finally, the third criterion has no impact on the decision and can be removed from the explanation.

In Lemma 1, we have studied how much one can change the weight vector v while maintaining the same preference between x and y. We keep the same idea but instantiate it in a different manner. We look at some changes in the weights that yield a switch of preference between x and y.

Following Example 2, a possible reasoning to select the arguments consists in understanding why the use of the reference weight vector $w^{\mathcal{F}}$ leads to the opposite comparison of x and y, compare to the weight vector v. This switch of preference necessarily comes from the criteria for which v_i is (significantly) different from $w^{\mathcal{F}}_i$. If it is possible to change v_i by $w^{\mathcal{F}}_i$ for some $i \in N$, while y remains preferred to x, then the specificity of criterion i may not be necessary in the explanation so that i may discarded from the explanation. If the wording "*x is on average at least as good as y*" is contained in the generated explanation, then there is no mistake in the reasoning consisting in not mentioning explicitly these criteria in the explanation.

We will compare v to other weights in the following example.

Example 3. Assume that $x = (0, 1, 0.6)$, $y = (1, 0, 1)$ and $v = (0.4, 0.2, 0.4)$, which yields $y \succ^{\mathrm{EU}}_{v} x$. We notice that y is on average strictly better than x. One has $A^+(y, x) = \{1, 3\}$ and $A^-(y, x) = \{2\}$. One feels that the explanation can be the same as in the previous example: the weight on the first criterion which is a positive argument is important, and the weight on the second criterion which is a negative argument is not important. Even though the last criterion is positive, it can be discarded from the explanation. The intuition is that the first criterion is a sufficiently positive argument since $\Delta_1 = 1 > \Delta_3 = 0.4$ and $v_1 = 0.4 = v_3$. How can we justify more formally that the last criterion can be removed from the explanation? Here is a possible justification. The weight vector v could have been assigned to the three criteria in a different manner. If v_1 were assigned to the second criterion, v_2 were assigned to the first criterion – leading to the weight $v' = (0.2, 0.4, 0.4)$ – then the decision would have been the opposite since $x \succeq^{\mathrm{EU}}_{v'} y$. The inversion of the preferences between x and y comes from a switch of the weight of criteria 1 and 2, and criterion 3 has not played any role in this inversion of preference.

The explanation will be based on the identification of the decisive criteria. As suggested by Example 3, one simple way to determine whether a set of criteria are decisive is to look at the influence of permuting their weights. If the positioning of these weights is very crucial, then a permutation will invert the decision. Our last example shows that several explanations can be generated for a given choice $y \succ^{\mathrm{EU}}_{v} x$.

Example 4. Assume that $x = (0.5, 0.5, 1)$, $y = (1, 1, 0)$ and $v = (0.4, 0.4, 0.2)$, which gives $y \succ^{\mathrm{EU}}_{v} x$. One has $A^+(y, x) = \{1, 2\}$ and $A^-(y, x) = \{3\}$. One can proceed as in Example 3. If one switches the weights of criteria 1 and 3 – leading to $v' = (0.2, 0.4, 0.4)$ – we obtain $x \succeq^{\mathrm{EU}}_{v'} y$. Hence the explanation of $y \succ^{\mathrm{EU}}_{v} x$ could be that y is better than x on criterion 1 which is important and x is better than y on criterion 3 which is not important. Another explanation can be generated. One could indeed switch criteria 2 and 3, and thus the explanation could focus as well on criteria 2 and 3.

3.4. Description of the general framework

Based on the previous Examples 1 through 4, the following points can be raised.

- First of all, three different types of explanation reasonings have been made in the previous examples: one for Example 1, one for Example 2 and one for the two Examples 3 and 4. The presentation of each type of explanation to the user requires a different reasoning. This latter is based on some implicit information and is thus related to the concept of an

anchor. In argumentation, an anchor refers to admitted facts, knowledge of the world or common sense rules [55,32,46]. The anchor corresponds to some implicit reasoning rules that are not explicitly quoted in an argumentation [55]. The missing causal relations in the argumentation are then drawn automatically by the audience. In our case, we will use the term *anchor* to denote a generic way of reasoning in the explanation. The set of all anchors is denoted by Ψ.
An anchor cannot be used in all situations, that is for any value of x, y and v. We denote by $\mathcal{D}(\mathcal{F},\psi)$ the set of values of $(x,y,v)\in\mathcal{D}(\mathcal{F})$ for which anchor ψ can be applied.

- Secondly, in Examples 3 and 4, the explanation concerns a permutation of the weights of two criteria. In more complex situations, several cycles may be necessary. In this case, the explanation is composed of a set of disjoint coalitions of criteria, and is thus an element of *Ex*. Each coalition of criteria in this set is called an *elementary argument*. Let $Ex(x,y,v,\mathcal{F},\psi)\subseteq Ex$ be the set of explanation sets allowed with anchor $\psi\in\Psi$ to explain the preference $y\succ_v^{\mathcal{F}} x$. We have $Ex(x,y,v,\mathcal{F},\psi)=\emptyset$ when $(x,y,v)\notin\mathcal{D}(\mathcal{F},\psi)$.

Definition 1. For $(x,y,v)\in\mathcal{D}(\mathcal{F})$, an *explanation set* of the prescription $y\succ_v^{\mathcal{F}} x$ is an element of the following set

$$Ex(x,y,v,\mathcal{F})=\left\{(\psi,\mathcal{A})\colon\ \psi\in\Psi \text{ and } \mathcal{A}\in Ex(x,y,v,\mathcal{F},\psi)\right\}.$$

For $(\psi,\mathcal{A})\in Ex(x,y,v,\mathcal{F})$, $\overline{\mathcal{A}}\cap A^{+}(y,x)$, $\overline{\mathcal{A}}\cap A^{=}(y,x)$ and $\overline{\mathcal{A}}\cap A^{-}(y,x)$ are the positive, null and negative arguments respectively used in the explanation, where $\overline{\mathcal{A}}:=\bigcup_{S\in\mathcal{A}} S$. In Examples 2, 3 and 4, we have seen that some arguments may be absent from the explanation. This means that $\overline{\mathcal{A}}$ is not necessarily equal to N.

We describe in this paragraph the formalism we will use to justify the selection of some arguments in the explanation and the removal of the other ones.

The set $Ex(x,y,v,\mathcal{F})$ is a combinatorial structure composed of many elements. In order to select the simplest explanation set, we define an order relation $\trianglelefteq$ over *Ex*. Relation $ex\trianglelefteq ex'$ means that the explanation set ex is not more complex to understand than ex'. The anchors are more or less simple to understand. There is thus an order relation $\trianglelefteq$ over Ψ. We denote by $\triangleleft$ the asymmetric part of $\trianglelefteq$ and by $\equiv$ the symmetric part of $\trianglelefteq$. One prefers unconditionally an explanation set using a simple anchor than any explanation set using a more complex anchor. The order relation $\trianglelefteq$ over $Ex(x,y,v,\mathcal{F})$ is thus defined as follows: $(\psi,\mathcal{A})\trianglelefteq(\psi',\mathcal{A}')$ if one of the following two conditions is met

- $\psi\triangleleft\psi'$,
- $\psi\equiv\psi'$, $\mathcal{A}\sqsubseteq\mathcal{A}'$ (i.e. $\mathcal{A}$ is simpler than $\mathcal{A}'$).

We are interested in the simplest explanation sets of $y\succ_v^{\mathcal{F}} x$, that is the minimal elements of $Ex(x,y,v,\mathcal{F})$ in the sense of $\trianglelefteq$.

The set *Ex* forms a combinatorial structure. The number of partitions of k blocks in a set of p elements is the Stirling number S_p^k of the second kind defined by [4]

$$S_p^k:=\frac{1}{k!}\sum_{i=0}^{k}(-1)^{k-i}\binom{k}{i}i^p.$$

Hence the cardinality of *Ex* is

$$\sum_{p=1}^{n}\sum_{k=1}^{p}S_p^k.$$

The exploration of *Ex* to find the simplest explanation set is expected to be complex.

3.5. Descriptions of the anchors

There remains to describe the anchors. Generalizing Examples 1, 2, 3 and 4, the set Ψ is composed of four anchors $\Psi=\{\psi_{\text{ALL}},\psi_{\text{NOA}},\psi_{\text{IVT}},\psi_{\text{RMG}}\}$.

- **Anchor "*all*"** ψ_{ALL} **(generalization of Example 1)**. When y is preferred to x on all criteria (i.e. $A^{-}(y,x)=A^{=}(y,x)=\emptyset$), y is preferred to x according to any preference model (see Example 1).

  $$\mathcal{D}(\mathcal{F},\psi_{\text{ALL}})=\left\{(x,y,v)\in\mathcal{D}(\mathcal{F})\colon\ A^{+}(y,x)=N\right\}.$$

  No specificity of $\succeq_v^{\mathcal{F}}$ needs to be quoted in the explanation. There is thus only one elementary argument which is the grand coalition $\{N\}$: $Ex(x,y,v,\mathcal{F},\psi_{\text{ALL}})=\{\{N\}\}$.

- **Anchor "*not on average*"** ψ_{NOA} **(generalization of Example 2)**. We are interested here in the case where the reference weight $w^{\mathcal{F}}$ leads to the opposite decision (see Example 2), that is $y \succ_v^{\mathcal{F}} x$ whereas $x \succeq_{w^{\mathcal{F}}}^{\mathcal{F}} y$. The reference weights $w^{\mathcal{F}}$ act as an anchor since the recipient of the explanation may think of $w^{\mathcal{F}}$. Hence

$$\mathcal{D}(\mathcal{F}, \psi_{\text{NOA}}) = \left\{(x, y, v) \in \mathcal{D}(\mathcal{F}) \colon x \succeq_{w^{\mathcal{F}}}^{\mathcal{F}} y\right\}.$$

  Following Example 2, the criteria i for which v_i can be replaced by $w_i^{\mathcal{F}}$, while y remains preferred to x, are discarded from the explanation. Therefore, for an explanation set $\mathcal{A}$, y is strictly preferred to x with the weights $(v_{\overline{\mathcal{A}}}, w^{\mathcal{F}}_{N\setminus\overline{\mathcal{A}}})$. Note that weights $(v_{\overline{\mathcal{A}}}, w^{\mathcal{F}}_{N\setminus\overline{\mathcal{A}}})$ may not be normalized, but this does not matter. The criteria in $N \setminus \overline{\mathcal{A}}$ are not decisive in the sense that their weights are either not influencing the decision or close to the reference weight. If the criteria $\overline{\mathcal{A}}$ are described in the explanation, then $N \setminus \overline{\mathcal{A}}$ may not be mentioned in the explanation.
  The set of elementary arguments is the coalition structure $\langle N\rangle$ composed of the singletons of N. Hence if $x \succeq_{w^{\mathcal{F}}}^{\mathcal{F}} y$, we define

$$Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}}) := \left\{\mathcal{A} \subseteq \langle N\rangle \colon y \succ^{\mathcal{F}}_{(v_{\overline{\mathcal{A}}}, w^{\mathcal{F}}_{N\setminus\overline{\mathcal{A}}})} x\right\}$$

  and otherwise we set $Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}}) := \emptyset$.
- **Anchor "*invert*"** ψ_{IVT} **(generalization of Examples 3 and 4)**. The idea of this anchor is to say that the decision would not have been the same if some weights of v were switched. Hence we set

$$\mathcal{D}(\mathcal{F}, \psi_{\text{IVT}}) = \left\{(x, y, v) \in \mathcal{D}(\mathcal{F}) \colon \exists \pi \in \Pi(N)\ x \succeq^{\mathcal{F}}_{\pi\circ v}\ y\right\}.$$

  The elementary arguments are the cycles of π. In the explanation that is generated, an implicit justification of why some arguments are discarded shall be given. In Examples 3 and 4, the criteria that are not used in the explanation keep their original weight v. We denote by $\mathcal{A}$ the explanation set. The preference between x and y is switched when the weights assigned to the criteria $\overline{\mathcal{A}}$ are $\pi \circ v$ and the remaining criteria $N \setminus \overline{\mathcal{A}}$ keep their original weight v. The set $\mathcal{A}$ must be a union of cycles of π, and thus $\mathcal{A} \subseteq \mathcal{A}(\pi)$. Therefore, we define

$$Ex(x, y, v, \mathcal{F}, \psi_{\text{IVT}}) := \left\{\mathcal{A} \colon \exists \pi \in \Pi(N) \text{ with } x \succeq^{\mathcal{F}}_{(\pi\circ v_{\overline{\mathcal{A}}}, v_{N\setminus\overline{\mathcal{A}}})} y \text{ and } \mathcal{A} \subseteq \mathcal{A}(\pi)\right\}.$$

  The analysis of the situations in which a different assignment of the weights to the criteria yields an inversion of preference between x and y helps to understand what are the decisive criteria.
- **Anchor "*remaining*"** ψ_{RMG}. This case occurs when none of the previous anchors applies. Hence

$$\mathcal{D}(\mathcal{F}, \psi_{\text{RMG}}) = \left\{(x, y, v) \in \mathcal{D}(\mathcal{F}) \colon A^+(y, x) \neq N,\ y \succ^{\mathcal{F}}_{w^{\mathcal{F}}} x \text{ and } \forall \pi \in \Pi(N),\ y \succ^{\mathcal{F}}_{\pi\circ v} x\right\}.$$

  As we will see in Section 7, there is only one elementary argument which is the grand coalition $\{N\}$: $Ex(x, y, v, \mathcal{F}, \psi_{\text{RMG}}) = \{\{N\}\}$.

All the previous anchors are ordered according to their intrinsic complexity:

$$\psi_{\text{ALL}} \lhd \psi_{\text{NOA}} \lhd \psi_{\text{IVT}} \lhd \psi_{\text{RMG}}.$$

The four anchors are derived from the previous examples. In the rest of the paper, we prove that the intuition given in the examples of Section 3.3 on the EU model holds in all situations of x, y and v, and that this intuition also generalizes to the other models Pess and Maj. More precisely, we first derive the important properties of the minimal elements of $Ex(x, y, v, \mathcal{F})$. From these properties we show that convincing explanations can be generated in all cases. Sections 4, 5, 6 and 7 study the anchors ψ_{ALL}, ψ_{NOA}, ψ_{IVT} and ψ_{RMG} respectively. In each section, the three decision models EU, Maj and Pess are analysed.

4. Anchor "*all*"

When $A^+(y, x) = N$, the fact that $y \succ_v^{\mathcal{F}} x$ is trivial since y is preferred to x according to any decision model.
The following sentence can be generated to explain $y \succ_v^{\mathcal{F}} x$ for this anchor.

`y is preferred to x since y is better than x on ALL criteria.`

5. Anchor "*not on average*"

We assume throughout this section that the anchor "*not on average*" holds and thus $y \preceq_{w^{\mathcal{F}}}^{\mathcal{F}} x$. The explanation synthesis is obtained from the understanding of why the weight vectors v and $w^{\mathcal{F}}$ yield different decisions regarding x and y, where

$w^{\mathcal{F}}$ is a representation of the priors of the recipient of the explanation on the weights. Our approach identifies the minimal set of criteria that need to keep their original weight, while the other criteria can take the weight value of $w^{\mathcal{F}}$, so that y remains preferred to x.

For the sake of simplicity in the notation, $\langle N \rangle$ is assimilated to N in this section. Hence the elements of $Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}})$ are assimilated to coalitions of N. The minimal explanation sets with anchor ψ_{NOA} according to $\trianglelefteq$ are exactly the minimal coalitions of $Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}})$ according to $\subseteq$.

Lemma 2. *A coalition $S \subseteq N$ is minimal in $Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}})$ if and only if*

$$S \in Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}}) \quad \text{and} \quad \forall k \in S, \quad S \setminus \{k\} \notin Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}}). \tag{3}$$

Note that $N \in Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}})$ and $\emptyset \notin Ex(x, y, v, \mathcal{F}, \psi_{\text{NOA}})$.

5.1. Expected utility model

Proposition 1. *Let $S \in Ex(x, y, v, \text{EU}, \psi_{\text{NOA}})$ be minimal in the sense of $\subseteq$. Then necessarily one has $v_k \neq \frac{1}{n}$ for all $k \in S$ and $S \subseteq A^+(y, x) \cup A^-(y, x)$. Moreover, for all $k \in S$,*

$$k \in A^+(y, x) \quad \Leftrightarrow \quad v_k > \frac{1}{n},$$
$$k \in A^-(y, x) \quad \Leftrightarrow \quad v_k < \frac{1}{n}.$$

Interpretation 1. From Proposition 1, the selected criteria S are clearly decisive since S contains no null argument, the positive arguments $S \cap A^+(y, x)$ are strong (i.e. their importance is larger than the mean weight $\frac{1}{n}$) and the negative arguments $S \cap A^-(y, x)$ are weak (i.e. their importance is lower than $\frac{1}{n}$). We have thus shown that the intuition of Example 2 is true in the general case. Hence the arguments that are selected can be presented in a very natural way to the recipient. If the set $S \subseteq N$ is selected, the reason why the other criteria are not mentioned is that the outcome would have been the same if these criteria had a medium importance. Hence a general sentence about the situation of the criteria in $N \setminus S$ is enough and one needs only to show the specificity of the criteria in S.

We are now interested in the generation of an explanation to the user. The explanation that is automatically generated can take the following structure: a concession to the reverse preference (namely $y \preceq^{\text{EU}}_{w^{\text{EU}}} x$) that is disproved, followed by a statement of the preference (namely $y \succ^{\text{EU}}_{v} x$) and the list $S \in Ex(x, y, v, \text{EU}, \psi_{\text{NOA}})$ (with S minimal) of arguments. This structure is classical in argumentation [17]. Relation $y \preceq^{\text{EU}}_{w^{\text{EU}}} x$ means that x is better than y on average. The following sentence can thus be generated to explain $y \succ^{\text{EU}}_{v} x$:

> Even though x is better than y on average, y is preferred to x since y is better than x on the criteria $S \cap A^+(y, x)$ that are important whereas y is worse than x on the criteria $S \cap A^-(y, x)$ that are not important.

The following additional sentence can be added if $\sum_{i \in N \setminus S} v_i \Delta_i > 0$

> Moreover, y is on average better than x on the other criteria.

This explanation aims at convincing an audience that would a priori think that x is preferred to y (because x is better on average), that the opposite preference holds. This is illustrated in Section 8.2.1 on several examples. Section 8.2.1 also presents a comparison of our approach with the Klein method.

Let us now give a simple method to compute a minimal element of $Ex(x, y, v, \text{EU}, \psi_{\text{NOA}})$. Let $\pi \in \Pi(N)$ such that

$$\left(v_{\pi(1)} - \frac{1}{n}\right)\Delta_{\pi(1)} \leqslant \cdots \leqslant \left(v_{\pi(n)} - \frac{1}{n}\right)\Delta_{\pi(n)}.$$

Proposition 2. *Let p be the largest integer in $\{1, \ldots, n\}$ such that $\{\pi(p), \ldots, \pi(n)\} \in Ex(x, y, v, \text{EU}, \psi_{\text{NOA}})$. Then $\{\pi(p), \ldots, \pi(n)\}$ is minimal in $Ex(x, y, v, \text{EU}, \psi_{\text{NOA}})$. Moreover, there is no minimal coalition of $Ex(x, y, v, \text{EU}, \psi_{\text{NOA}})$ with a strictly smaller cardinality than $\{\pi(p), \ldots, \pi(n)\}$.*

5.2. Weighted majority model

Proposition 3. *Let $S \in Ex(x, y, v, \text{Maj}, \psi_{\text{NOA}})$ be minimal. Then necessarily one has $v_k \neq \frac{1}{n}$ for all $k \in S$, and $S \subseteq A^+(y,x) \cup A^-(y,x)$. Moreover, for all $k \in S$,*

$$k \in A^+(y,x) \quad \Leftrightarrow \quad v_k > \frac{1}{n},$$
$$k \in A^-(y,x) \quad \Leftrightarrow \quad v_k < \frac{1}{n}.$$

From Proposition 3, for any minimal explanation set, the positive arguments are strong and the negative arguments are weak. This is similar to Interpretation 1.

The explanation of the preference of y over x follows the same structure as in Section 5.1. Relation $y \preceq^{\text{Maj}}_{w^{\text{Maj}}} x$ means that there are more criteria for which x is preferred to y than criteria for which y is preferred to x. The following sentence can be generated to explain $y \succ^{\text{Maj}}_{v} x$:

> Even though there are more criteria for which x is preferred to y than criteria for which y is preferred to x, y is nevertheless preferred to x since the criteria $S \cap A^+(y,x)$ for which y is better than x are important whereas the criteria $S \cap A^-(y,x)$ for which x is better than y are not important.

Let us now give a simple method to compute a minimal element of $Ex(x, y, v, \text{Maj}, \psi_{\text{NOA}})$. Let $\pi \in \Pi(N)$ such that

$$\left(v_{\pi(1)} - \frac{1}{n}\right) \text{sgn}_{\pi(1)} \leqslant \cdots \leqslant \left(v_{\pi(n)} - \frac{1}{n}\right) \text{sgn}_{\pi(n)}.$$

Proposition 4. *Let p be the largest element of $\{1, \ldots, n\}$ such that $\{\pi(p), \ldots, \pi(n)\} \in Ex(x, y, v, \text{Maj}, \psi_{\text{NOA}})$. Then $\{\pi(p), \ldots, \pi(n)\}$ is minimal in $Ex(x, y, v, \text{Maj}, \psi_{\text{NOA}})$. Moreover, there is no minimal coalition of $Ex(x, y, v, \text{Maj}, \psi_{\text{NOA}})$ with a strictly smaller cardinality than $\{\pi(p), \ldots, \pi(n)\}$.*

5.3. Pessimistic qualitative model

Let $h^{\text{Pess},S}_w(z) = \bigwedge_{i \in S}(z_i \vee (1 - w_i))$, with $S \subseteq N$. In the weighted minimum aggregation function $h^{\text{Pess}}_v(x)$, the contribution of a criterion $i \in N$ is $x_i \vee (1 - v_i)$. The value of $1 - v_i$ is small when criterion i is important. Hence a bad score on an important criterion cannot be saved by the other criteria, and the overall score is necessarily bad. On the contrary, a bad score on an unimportant criterion is saved by its small weight.

Proposition 5. *Let $S \in Ex(x, y, v, \text{Pess}, \psi_{\text{NOA}})$ be minimal. Then $S \subseteq A^-(y,x) \cup A^=(y,x)$ and for all $k \in S$, $y_k < 1 - v_k$. Moreover,*

$$h^{\text{Pess},S}_v(x) \geqslant h^{\text{Pess},S}_v(y) > \bigwedge_{i \in N \setminus S} x_i \quad \text{and} \quad \bigwedge_{i \in N \setminus S} y_i > \bigwedge_{i \in N \setminus S} x_i.$$

Finally, we have $h^{\text{Pess},S}_v(x) \geqslant h^{\text{Pess},S}_v(y) > h^{\text{Pess},N\setminus S}_v(x)$ and $h^{\text{Pess},N\setminus S}_v(y) > h^{\text{Pess},N\setminus S}_v(x)$.

Relation $y_k < 1 - v_k$ for $k \in S$ means that the weight of criterion k hides the bad score y_k. Hence S consists only in weak negative or null arguments. This proposition also shows that the worse mark of x on the remaining criteria (i.e. $N \setminus S$) is lower than any score of y in $N \setminus S$ and is also lower than the aggregated score of y in S. Moreover, $h^{\text{Pess}}_v(x)$ is attained in $N \setminus S$.

A positive argument cannot be selected in $Ex(x, y, v, \text{Pess}, \psi_{\text{NOA}})$ since this set only consists of negative or null arguments. The following proposition shows that some criteria not in S (for $S \in Ex(x, y, v, \text{Pess}, \psi_{\text{NOA}})$) might be advantageously be added in the explanation.

Proposition 6. *Assume that $h^{\text{Pess}}_v(x)$ is attained at $k^x \in N$: $h^{\text{Pess}}_v(x) = x_{k^x} \vee (1 - v_{k^x})$. Then $k^x \in A^+(y,x)$ and $y_{k^x} > 1 - v_{k^x}$.*

Let $M := \{k \in N,\ y_k \leqslant h^{\text{Pess}}_v(x)\}$. For every $S \in Ex(x, y, v, \text{Pess}, \psi_{\text{NOA}})$ minimal, we have $S \subseteq M$. Moreover, $1 - v_k > x_{k^x} \vee (1 - v_{k^x})$ for every $k \in M$.

Interpretation 2. Thanks to Propositions 5 and 6, this anchor is relatively simple. The set of arguments can indeed be divided into parts. On the one hand, the selected arguments S are negative and weak arguments. On these criteria, y has a worse score than x but this is saved by the relatively small value of the weight (since $y_k < 1 - v_k$). The value of the weight is thus crucial for these criteria. On the other hand, the non-selected arguments $N \setminus S$ are criteria for which the value of

the weight is either not decisive or close to the reference weight 1. The only really decisive criteria in $N \setminus S$ is k^x for which the overall score $h_v^{\text{Pess}}(x)$ of x is attained. This criterion is sufficiently important so that it does not hide the score of x (see conditions on $1 - v_k$ in Proposition 6). The remaining criteria $N \setminus (S \cup \{k^x\})$ may not be mentioned. If one wishes to provide a comprehensive explanation, one may add that the bad evaluation of y on the criteria in $M^* := \{k \in N \setminus S,\ y_k \leqslant h_v^{\text{Pess}}(x)\} = M \setminus S$ is saved by their relatively small importance. Indeed, from Proposition 6, we have $1 - v_k > x_{k^x} \vee (1 - v_{k^x})$ for all $k \in M^*$, and thus, the criteria in M^* are relatively unimportant. Note that M^* may contain negative or null arguments even though this set will be most of the time composed of only positive arguments. Finally, y is relatively good on the remaining criteria $N \setminus (S \cup M^* \cup \{k^x\})$.

The following sentences can be generated to explain $y \succ_v^{\text{Pess}} x$:

```
Even though the worst score of y is worse than that of x, y is nevertheless preferred
to x since the relatively bad scores of y compared to that of x on criteria S are
saved by their relatively small importance.
  The overall score of x is attained at criterion k^x for which x is worse than y and
the relatively large importance does not save x.
  On the criteria M*, the bad evaluation of y is saved by the relatively small
importance of these criteria. Finally, y is sufficiently good on the remaining
criteria N \ (S ∪ M* ∪ {k^x}).
```

The last sentence means that the score of y on the criteria $N \setminus (S \cup M^* \cup \{k^x\})$ needs just to be larger than the overall score of x. An illustration is proposed in Example 7 in Section 8.2.2.

The following proposition provides a simple way to compute a minimal explanation set.

Proposition 7. *Let p be the smallest integer of $\{1, \ldots, n\}$ such that $\{\pi_N^y(1), \ldots, \pi_N^y(p)\} \in Ex(x, y, v, \text{Pess}, \psi_{\text{NOA}})$, where $\pi_N^y \in \Pi(N)$ is defined in the beginning of Section 3. Then $\{\pi_N^y(1), \ldots, \pi_N^y(p)\}$ is minimal in $Ex(x, y, v, \text{Pess}, \psi_{\text{NOA}})$.*

It might seem surprising that x is not taken into account in the permutation used to compute a minimal explanation set. In fact, since $\bigwedge_{i \notin S} x_i < \bigwedge_{i \notin S} y_i$, it is clear that $\pi_N^y(1) \in A^=(y, x) \cup A^-(y, x)$.

6. Anchor "*invert*"

We assume in this section that there exists $\pi \in \Pi(N)$ such that $y \preceq_{\pi \circ v}^{\mathcal{F}} x$. We aim at determining the arguments explaining why $y \succ_v^{\mathcal{F}} x$ even though $y \preceq_{\pi \circ v}^{\mathcal{F}} x$. The minimal arguments with anchor ψ_{IVT} are exactly the minimal elements of $Ex(x, y, v, \mathcal{F}, \psi_{\text{IVT}})$ in the sense of $\sqsubseteq$.

Proposition 8. *Let $\mathcal{A} \in Ex(x, y, v, \mathcal{F}, \psi_{\text{IVT}})$ be minimal in the sense of $\sqsubseteq$. Then for every $S \in \mathcal{A}$, $|S| \geqslant 2$.*

6.1. Expected utility model

We need the following result.

Lemma 3. *Let $S \subseteq N$ and $s = |S|$. Then $\max_{\pi \in \Pi(S)} \sum_{i \in S} v_{\pi(i)} \Delta_i$ is attained at $\overline{\pi}_S \in \Pi(S)$ defined by $\overline{\pi}_S(j) = \pi_S^y((\pi_S^\Delta)^{-1}(j))$ for all $j \in S$. Moreover, $\min_{\pi \in \Pi(S)} \sum_{i \in S} v_{\pi(i)} \Delta_i$ is attained at $\underline{\pi}_S \in \Pi(S)$ defined by $\underline{\pi}_S(j) = \pi_S^y(s - (\pi_S^\Delta)^{-1}(j) + 1)$ for all $j \in S$.*

The minimal explanation sets are characterized in the next proposition.

Proposition 9. *Let $\mathcal{A} \in Ex(x, y, v, \text{EU}, \psi_{\text{IVT}})$ be minimal in the sense of $\sqsubseteq$. Let $\pi \in \Pi(N)$ such that $x \succeq^{\text{EU}}_{(\pi \circ v_{\overline{\mathcal{A}}}, v_{N \setminus \overline{\mathcal{A}}})} y$ and $\mathcal{A} \subseteq \mathcal{A}(\pi)$. We can choose π such that $\pi(i) = i$ for all $i \in N \setminus \overline{\mathcal{A}}$.*

Let $S \in \mathcal{A}$. Then for all $k, j \in S$ with $j \neq k$ we have $\Delta_k \neq \Delta_j$, $v_{\pi(j)} \neq v_{\pi(k)}$ and

$$(v_{\pi(j)} - v_{\pi(k)}) \times (\Delta_k - \Delta_j) > 0. \tag{4}$$

Moreover, $\Delta_{\pi_S^\Delta(1)} < \Delta_{\pi_S^\Delta(2)} < \cdots < \Delta_{\pi_S^\Delta(|S|)}$, and

$$v_{\pi \circ \pi_S^\Delta(1)} > v_{\pi \circ \pi_S^\Delta(2)} > \cdots > v_{\pi \circ \pi_S^\Delta(|S|)}. \tag{5}$$

Hence for all $j \in S$, $\pi(j) = \underline{\pi}_S(j)$. Finally,

$$\sum_{j \in S} v_j \Delta_j > \sum_{j \in S} v_{\underline{\pi}_S(j)} \Delta_j.$$

From the relation $\Delta_{\pi_S^{\Delta}(1)} < \Delta_{\pi_S^{\Delta}(2)} < \cdots < \Delta_{\pi_S^{\Delta}(|S|)}$, there is at most one null argument in every $S \in \mathcal{A}$. Moreover, since $\pi = \underline{\pi}_S$ over each $S \in \mathcal{A}$, the allocation of the weights to the criteria according to π is the most favourable one relatively to an inversion of the comparison between x and y.

We are now interested in the generation of the explanation to the user. Let $\mathcal{A} \in Ex(x, y, v, \mathcal{F}, \psi_{\text{IVT}})$ be minimal in the sense of $\sqsubseteq$. Relation (5) means that, among each cycle S of π, the ordering π assigns the largest weights to the most negative criteria (i.e. to the criteria that have the smallest value of Δ) and the smallest weights to the most positive criteria. Proposition 9 shows that if the weights are assigned differently in a cycle S, then the preference between x and y is not reversed. According to Proposition 9, we have

$$\forall i, j \in S \quad \text{if } \Delta_i < \Delta_j, \quad \text{then } v_{\pi(i)} > v_{\pi(j)}.$$

We have $y \succ_v^{\text{EU}} x$ since there exist $i, j \in S$ with $\Delta_i < \Delta_j$ such that the weight v_i assigned to criterion i is not larger than the weight v_j assigned to criterion j. The explanation can thus focus on the following pairs:

$$R_S := \left\{(i, j) \in S^2 \colon \Delta_i < \Delta_j \text{ and } v_i < v_j\right\}.$$

For $R \subset S^2$, if R is seen as a binary relation, we define the transitive closure $\overline{R}$ of R as the smallest subset of S^2 such that $R \subseteq \overline{R}$, and

$$\left[(i, j) \in \overline{R} \text{ and } (j, k) \in \overline{R}\right] \quad \Rightarrow \quad (i, k) \in \overline{R}.$$

The set R_S is stable under the transitive closure. Let us denote by $R_S^* \subseteq R_S$ the smallest subset R of S^2 such that $\overline{R} = R_S$. For the explanation, we can restrict ourselves to the pairs in R_S^* since the other pairs $R_S \setminus R_S^*$ can be deduced from R_S^* by transitivity. Let $R^* = \bigcup_{S \in \mathcal{A}} R_S^*$.

The explanation does not consist of a concatenation of an elementary text for each pair in R^*. We organize the arguments in R^* into four sets of arguments C_{PS} (positive and strong), C_{PRS} (positive and relatively strong), C_{NW} (negative and weak), C_{NRW} (negative and relatively weak), and a set of pairs of arguments C_{PN} (a positive and a negative argument). Let $(i, j) \in R^*$. We have three cases.

(i) $j \in A^+(y, x)$, $i \in A^-(y, x)$ and $v_i < v_j$. This situation is illustrated by Example 10 in Section 8.3.1. The strength of the positive argument j is larger than that of the negative argument i. If $v_j \geqslant \frac{1}{n}$ and $v_i \leqslant \frac{1}{n}$, then j is added in C_{PS} and i is added in C_{NW}. If the previous condition does not hold but $v_i \ll v_j$ then the pair (i, j) is added in C_{PN}. Finally, if the previous conditions do not hold, then j is added in C_{PS} when $v_j \geqslant \frac{1}{n}$ and i is added in C_{NW} when $v_i \leqslant \frac{1}{n}$. Note that one of the previous two cases necessarily holds.

(ii) $i, j \in A^+(y, x)$, $\Delta_j > \Delta_i$ and $v_i < v_j$. This situation is illustrated by Example 10 in Section 8.3.1. The fact that the pair (i, j) is selected emphasises the fact that criterion j is stronger and more positive than i. If criterion j had the smaller weight v_i of criterion i and if criterion i had the larger weight v_j, then the decision would have been the opposite. Criterion i is present in the selected pair (i, j) only as a comparison to the situation of criterion j. Hence, criterion i is not mentioned in the explanation. Therefore, if $v_j \geqslant \frac{1}{n}$, then j is added in C_{PS}, otherwise j is added in C_{PRS}.

(iii) $i, j \in A^-(y, x)$, $\Delta_j > \Delta_i$ and $v_i < v_j$. The situation is dual to (ii). Likewise, the most striking criterion is i since it is more negative and weaker than j. Criterion j is present only to emphasis this difference and is thus not mentioned. Therefore, if $v_j \leqslant \frac{1}{n}$, then i is added in C_{NW}, otherwise i is added in C_{NRW}.

Interpretation 3. In anchor ψ_{IVT}, there are more positive reasons regarding option y compared to the situation of the anchor ψ_{NOA} since y is here better on average than x. We are looking for the permutations in the assignment of the weights to the criteria, that invert the decision. These permutations are used to identify the decisive criteria. More precisely, from Proposition 9, the permutation π that is selected assigns in each cycle of π, the largest weights to the most negative criteria and the smallest weights to the most positive criteria. Hence a criterion for which this property is satisfied without the permutation, cannot be selected. The criteria that are likely to be selected are thus the positive arguments that are strong and the negative ones that are weak. Yet the selection process in ψ_{IVT} can be seen as a relaxation of that in ψ_{NOA} as one may now consider a positive criterion which weight is lower than $\frac{1}{n}$, and a negative criterion which weight is greater than $\frac{1}{n}$.

The arguments that are displayed among the elements of R^* are contained in $C = \langle C_{\text{PS}}, C_{\text{PRS}}, C_{\text{NW}}, C_{\text{NRW}}, C_{\text{PN}}\rangle$. Several examples of the computation of C are given in Section 8.3.1. The explanation can be the following one:

> y is preferred to x since y is better than x on the criteria C_{PS} that are important and on the criteria C_{PRS} that are relatively important, x is better than y on the criteria C_{NW} that are not important and on the criteria C_{NRW} that are not really important, and [criterion j for which y is better than x is more important than criterion i for which y is worse than x]$_{\text{for all } (i,j) \in C_{\text{PN}}}$.

The argument in the brackets is repeated for all $(i, j) \in C_{\text{PN}}$.

In the weight $(\pi \circ v_{\overline{\mathcal{A}}}, v_{N\setminus\overline{\mathcal{A}}})$, the weights remain the same for the non-selected criteria $N \setminus \overline{\mathcal{A}}$. In this case, unlike anchor ψ_{NOA}, there is no property of the non-selected criteria. It is then possible to consider all minimal explanation sets $\mathcal{A}$ with the smallest number of elements and to concatenate the explanation for these sets. This is illustrated in Example 15 in Section 8.4.

We now wish to compute a minimal element $\mathcal{A}$ of $Ex(x, y, v, \text{EU}, \psi_{\text{IVT}})$. First of all, one can search for $\mathcal{A}$ without searching for a permutation π. Proposition 9 shows indeed that, for $\mathcal{A}$ minimal, the restriction of π on S is equal to $\underline{\pi}_S$ for all $S \in \mathcal{A}$, and $\pi(i) = i$ for all $i \in N \setminus \overline{\mathcal{A}}$. Let us define for $S \subseteq N$

$$D_S^{\text{EU}} := \sum_{j\in S} v_j \Delta_j - \min_{\pi\in\Pi(S)} \sum_{j\in S} v_{\pi(j)} \Delta_j.$$

By Lemma 3, $D_S^{\text{EU}} = \sum_{j\in S}(v_j - v_{\underline{\pi}_S(j)})\Delta_j$. By Proposition 9, $D_S^{\text{EU}} > 0$ for $S \in \mathcal{A}$. The subsets S for which $D_S^{\text{EU}} = 0$ are not candidate coalitions for $\mathcal{A}$. We set

$$\mathcal{S} = \left\{S \subseteq N\colon\ D_S^{\text{EU}} > 0 \text{ and } \mathcal{A}(\underline{\pi}_S) = \{S\}\right\}.$$

If $\mathcal{A}(\underline{\pi}_S) \sqsubset \{S\}$, then the coalition structure $\mathcal{A}(\underline{\pi}_S) =: \{S_1, \ldots, S_r\}$ can be decomposed in sub-coalitions $S_1, \ldots, S_r$ of S. Since $D_S^{\text{EU}} = D_{S_1}^{\text{EU}} + \cdots + D_{S_r}^{\text{EU}}$, coalition S can be replaced by the simpler coalitions $\{S_1, \ldots, S_r\}$, and thus S is discarded. This explains the condition $\mathcal{A}(\underline{\pi}_S) = \{S\}$ in the definition of $\mathcal{S}$.

The elements of $\mathcal{S}$ are labelled in the following order

$$\mathcal{S} = \{T_1, T_2, \ldots, T_p\}$$

with $p = |\mathcal{S}|$, $T_1 \trianglelefteq_{\text{lexi}} T_2 \trianglelefteq_{\text{lexi}} \cdots \trianglelefteq_{\text{lexi}} T_p$, and $\trianglelefteq_{\text{lexi}}$ is defined by

$$S \trianglelefteq_{\text{lexi}} T \iff \text{either} \quad |S| < |T| \quad \text{or} \quad \left[|S| = |T| \text{ and } D_S^{\text{EU}} \geqslant D_T^{\text{EU}}\right].$$

This is a lexicographic ordering where one looks first at the cardinality and then at the value of D^{EU}. It can be interpreted in terms of the simplicity of presenting a coalition in the explanation. This order relation can be extended to explanation sets. We define $\sqsubset_{\text{discri}}$ over Ex as follows: For $\{A_1, \ldots, A_q\}, \{B_1, \ldots, B_r\} \in Ex$ with $A_1 \trianglelefteq_{\text{lexi}} \cdots \trianglelefteq_{\text{lexi}} A_q$ and $B_1 \trianglelefteq_{\text{lexi}} \cdots \trianglelefteq_{\text{lexi}} B_r$, we have $\{A_1, \ldots, A_q\} \sqsubset^*_{\text{discri}} \{B_1, \ldots, B_r\}$ if

$$\begin{aligned}&\left[\exists k \in \{1, \ldots, t\}\ |A_q| = |B_r|, \ldots, |A_{q-k+1}| = |B_{r-k+1}| \text{ and } |A_{q-k}| < |B_{r-k}|\right]\\ &\quad \text{or} \quad \left[|A_q| = |B_r|, \ldots, q < r \text{ and } |A_1| = |B_{r-q+1}|\right]\end{aligned}$$

where $t = q \wedge r$, and we have $\{A_1, \ldots, A_q\} \sqsubset_{\text{discri}} \{B_1, \ldots, B_r\}$ if $\{A_1, \ldots, A_q\} \sqsubset^*_{\text{discri}} \{B_1, \ldots, B_r\}$ or

$$\begin{aligned}&q = r \quad \text{and} \quad |A_1| = |B_1|, \ldots, |A_q| = |B_q|\\ &\quad \text{and} \quad \exists k \in \{1, \ldots, t\} \quad D_{A_1} = D_{B_1}, \ldots, D_{A_{k-1}} = D_{B_{k-1}} \quad \text{and} \quad D_{A_k} > D_{B_k}.\end{aligned}$$

In the explanation set $\{A_1, \ldots, A_q\}$, all coalitions will be explained to the user and A_q is the most complex coalition of this set. Hence when comparing two explanation sets, we compare the most complex coalition of the first explanation set with that of the second one, then, in case of equality, do the same with the second most complex coalition, and so on. This is described by $\sqsubset^*_{\text{discri}}$. When the two explanations sets have coalitions of the same cardinality, one then looks at the value of D. This is depicted in $\sqsubset_{\text{discri}}$. We define $\equiv_{\text{discri}}$ over Ex as follows: with the same notation as before, we set $\{A_1, \ldots, A_q\} \equiv_{\text{discri}} \{B_1, \ldots, B_r\}$ if

$$q = r \quad \text{and} \quad |A_1| = |B_1|, \ldots, |A_q| = |B_q| \quad \text{and} \quad D_{A_1} = D_{B_1}, \ldots, D_{A_q} = D_{B_q}.$$

Finally, we define $\sqsubseteq_{\text{discri}}$ by $\mathcal{A} \sqsubseteq_{\text{discri}} \mathcal{B}$ if either $\mathcal{A} \sqsubset_{\text{discri}} \mathcal{B}$ or $\mathcal{A} \equiv_{\text{discri}} \mathcal{B}$. The order $\sqsubseteq_{\text{discri}}$ is a complete order that refines $\sqsubseteq$ in the sense that [21]:

(i) $\mathcal{A} \sqsubset \mathcal{B} \Rightarrow \mathcal{A} \sqsubset_{\text{discri}} \mathcal{B}$,
(ii) $\exists \mathcal{A}, \mathcal{B} \in Ex,\ \mathcal{A} \not\sqsubset \mathcal{B},\ \mathcal{B} \not\sqsubset \mathcal{A}$ and $\mathcal{A} \sqsubset_{\text{discri}} \mathcal{B}$.

We define Algorithm **Algo-**EU (see Table 1) to determine a minimal explanation set $\mathcal{A}$. In this algorithm, $\mathcal{B}$ contains the best explanation set (in the sense of the complete order $\sqsubseteq_{\text{discri}}$) found so far.

Proposition 10. *Algorithm* **Algo-**EU *always returns a non-empty explanation set. Moreover, the outcome of Algorithm* **Algo-**EU *is an element of* $Ex(x, y, v, \text{EU}, \psi_{\text{IVT}})$ *that is minimal in the sense of* $\sqsubseteq$. *Finally, the outcome of Algorithm* **Algo-**EU *is a minimal argumentation set in* $Ex(x, y, v, \text{EU}, \psi_{\text{IVT}})$ *in the sense of* $\sqsubseteq_{\text{discri}}$.

Table 1
Algorithm for the determination of $\mathcal{A}$.

Algorithm **Algo-**EU:

- The algorithm returns the output of $Algo(\emptyset, \emptyset, 0)$.
- $Algo(\mathcal{A}, \mathcal{B}, k)$:

$L1$: For $i = k+1, \ldots, p$:
$L2$: If $T_i \cap \overline{\mathcal{A}} = \emptyset$:
$L3$: If $\sum_{S \in \mathcal{A}} D_S^{\mathrm{EU}} + D_{T_i}^{\mathrm{EU}} \geqslant H_v^{\mathrm{EU}}(y, x)$
$L4$: then $\mathcal{C} \leftarrow \mathcal{A} \cup \{\{T_i\}\}$,
$L5$: else // *Branching:*
$L6$: $\mathcal{C} \leftarrow Algo(\mathcal{A} \cup \{\{T_i\}\}, \mathcal{B}, i)$.
$L7$: // *Updating the best explanation set:*
$L8$: If $\mathcal{C} \neq \emptyset$ and $[\mathcal{B} = \emptyset$ or $\mathcal{C} \sqsubset_{\mathrm{discri}} \mathcal{B}]$ then $\mathcal{B} \leftarrow \mathcal{C}$.
$L9$: // *Bounding:*
$L10$: If $\mathcal{B} \neq \emptyset$ and $\mathcal{A} \cup \{\{T_i\}\} \not\sqsubset_{\mathrm{discri}} \mathcal{B}$ then return $\mathcal{B}$.
$L11$: return $\emptyset$.

To end this section, we show that Algorithm **Algo-**EU can be easily modified to generate all minimal explanation sets in the sense of the partial order $\sqsubset^*_{\mathrm{discri}}$. As said earlier, when one wish to enrich the generated text by considering several explanation sets, all these sets shall be minimal elements of $Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{IVT}})$ in the sense of $\sqsubset^*_{\mathrm{discri}}$. To this end, $\sqsubset_{\mathrm{discri}}$ is replaced by $\sqsubset^*_{\mathrm{discri}}$ in the algorithm. Moreover, the output $\mathcal{B}$ of the algorithm is no more a unique explanation set but a collection of explanation sets. Accordingly, $\mathcal{B}$ is now a collection of explanation sets. Then the line $L8$ is replaced by the following two lines:

If $\mathcal{C} \neq \emptyset$ and $\mathcal{B} = \emptyset$ then $\mathcal{B} \leftarrow \{\mathcal{C}\}$.

If $\mathcal{C} \neq \emptyset$ and $\mathcal{B} \neq \emptyset$ and $\nexists \mathcal{D} \in \mathcal{B}$: $\mathcal{D} \sqsubset^*_{\mathrm{discri}} \mathcal{C}$ then $\mathcal{B} \leftarrow (\mathcal{B} \setminus \{\mathcal{D} \in \mathcal{B}\colon\ \mathcal{C} \sqsubset^*_{\mathrm{discri}} \mathcal{D}\}) \cup \{\mathcal{C}\}$.

Moreover, line $L10$ is replaced by:

If $\mathcal{B} \neq \emptyset$ and $[\forall \mathcal{D} \in \mathcal{B}\colon\ \mathcal{D} \sqsubset^*_{\mathrm{discri}} \mathcal{A} \cup \{\{T_i\}\}]$ then return $\mathcal{B}$.

Extending Proposition 10, we obtain that the modified Algorithm **Algo-**EU returns the minimal elements of $Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{IVT}})$ in the sense of $\sqsubset^*_{\mathrm{discri}}$.

6.2. Weighted majority model

Proposition 11. *Let $\mathcal{A} \in Ex(x, y, v, \mathrm{Maj}, \psi_{\mathrm{IVT}})$ be minimal in the sense of $\sqsubseteq$. Let $\pi \in \Pi(N)$ such that $x \succeq^{\mathrm{EU}}_{(\pi \circ v_{\mathcal{A}}, v_{N \setminus \mathcal{A}})} y$ and $\mathcal{A} \subseteq \mathcal{A}(\pi)$. Let $S \in \mathcal{A}$. Then $2 \leqslant |S| \leqslant 3$. For all $k, j \in S$, with $k \neq j$, k and j cannot belong to the same set $A^+(y, x)$, $A^=(y, x)$ or $A^-(y, x)$. Moreover, we have*

- *If $[j \in A^+(y, x)$ and $k \in A^-(y, x)]$ or $[j \in A^+(y, x)$ and $k \in A^=(y, x)]$ or $[j \in A^=(y, x)$ and $k \in A^-(y, x)]$, then $v_{\pi(k)} > v_{\pi(j)}$.*
- *If $[k \in A^+(y, x)$ and $j \in A^-(y, x)]$ or $[k \in A^+(y, x)$ and $j \in A^=(y, x)]$ or $[k \in A^=(y, x)$ and $j \in A^-(y, x)]$, then $v_{\pi(k)} < v_{\pi(j)}$.*

From Proposition 11, following the permutation π, the largest weights among the coalition S are assigned to the negative arguments, and the smallest weights are assigned to the positive arguments. When $|S| = 3$, with $S = \{i_+, i_=, i_-\}$, $i_+ \in A^+(y, x)$, $i_= \in A^=(y, x)$ and $i_- \in A^-(y, x)$. The idea is that, among criteria S, the largest weight is not assigned to the negative argument i_- and the smallest weight is not assigned to the positive argument i_+. The explanation can be the same as in the previous section. First, the set R^* is computed. Then the five sets of arguments C_{PS}, C_{PRS}, C_{NW}, C_{NRW} and C_{PN} are constructed. The generated text is similar to that in the previous section.

Algorithm **Algo-**EU described in Section 6.1 can be adapted with minor changes to determine a minimal explanation set with the Maj model. One only needs to change D_S^{EU} by

$$D_S^{\mathrm{Maj}} := \sum_{j \in S} v_j \operatorname{sgn}_j - \min_{\pi \in \Pi(S)} \sum_{j \in S} v_{\pi(j)} \operatorname{sgn}_j = \sum_{j \in S} (v_j - v_{\underline{\pi}_S(j)}) \operatorname{sgn}_j$$

and condition $\sum_{S \in \mathcal{B}} D_S^{\mathrm{EU}} + D_{S_i}^{\mathrm{EU}} \geqslant H_v^{\mathrm{EU}}(y, x)$ by $\sum_{S \in \mathcal{B}} D_S^{\mathrm{Maj}} + D_{S_i}^{\mathrm{Maj}} \geqslant H_v^{\mathrm{Maj}}(y, x)$.

6.3. Pessimistic qualitative model

Proposition 12. *Let $\mathcal{A} \in Ex(x, y, v, \mathrm{Pess}, \psi_{\mathrm{IVT}})$ be minimal in the sense of $\sqsubseteq$. Let $\pi \in \Pi(N)$ such that $x \succeq^{\mathrm{Pess}}_{(\pi \circ v_{\mathcal{A}}, v_{N \setminus \mathcal{A}})} y$ and $\mathcal{A} \subseteq \mathcal{A}(\pi)$.*

For every $S \in \mathcal{A}$, there exist $K_S \subseteq (A^+(y, x) \cup A^=(y, x)) \cap S$ and $J_S \subseteq A^-(y, x) \cap S$ such that

$$\{v_{\pi(i)}\colon i \in (A^-(y,x) \cap S) \setminus J_S\} > \{v_{\pi(k)}\colon k \in K_S\} \geqslant \{v_{\pi(j)}\colon j \in J_S\} > \{v_{\pi(i)}\colon i \in ((A^+(y,x) \cup A^=(y,x)) \cap S) \setminus K_S\}.$$

The sets K_S and J_S, for $S \in \mathcal{A}$, satisfy the following properties:

- *There exists at most a subset $S \in \mathcal{A}$ such that $K_S \neq \emptyset$. For this coalition S, one has $|K_S| = 1$.*
- *For all $S \in \mathcal{A}$, $J_S = \emptyset$ whenever $K_S = \emptyset$.*
- *If $S = N$, then $K_S \cup J_S \neq N$.*

Let $S \in \mathcal{A}$. If $K_S \neq \emptyset$ and $K_S = \{k\}$, then

$$h^{\text{Pess},N\setminus\{k\}}_{\pi\circ v}(x) = h^{\text{Pess},N\setminus(J_S\cup\{k\})}_{\pi\circ v}(x).$$

The last relation in Proposition 12 means that criteria J_S do not count in the overall evaluations of x and y.

Interpretation 4. Without the permutation π, we are not in the situation described in Proposition 12. Hence the largest weights among a coalition $S \in \mathcal{A}$ are not assigned to the negative arguments (except for J_S), and the smallest weights are not assigned to the positive arguments (except for K_S). The idea is that, among the selected criteria, the positive arguments have a larger weight than the negative ones, which is quite easy to understand.

As in Section 8.3.1, one may compute R^* and $C = \langle C_{\text{PS}}, C_{\text{PRS}}, C_{\text{NW}}, C_{\text{NRW}}, C_{\text{PN}}\rangle$. More precisely, for all $S \in \mathcal{A}$, we define the pair (A_S, B_S) (with $A_S = (S \cap A^-(y,x)) \setminus J_S$ and $B_S = [S \cap (A^+(y,x) \cup A^-(y,x))] \setminus K_S$) such that $\forall i \in A_S$, $\forall j \in B_S$, i is negative, j is positive or null, and $v_i < v_j$. The explanation can thus be the following one:

> y is preferred to x since [criteria A_S for which y is better than x are more important than criteria B_S for which y is worse than x]$_{\text{for all } S\in\mathcal{A}}$.

When $K_S \cup J_S = S$, another explanation must be given. From the inequalities contained in (17) (see the proof of Proposition 12), the marks of x and y on the positive argument k are bad but are saved by the small weight of criterion k, since $1 - v_{\pi(j)}$ is relatively large for all $j \in S \setminus \{k\}$. Moreover, the marks of x and y on the negative arguments J_S are good. From the last relation in Proposition 12, the relatively bad overall utility of x is not attained in S but in the other criteria. The explanation can thus be the following one:

> ... [the bad scores of x and y on the positive argument k are saved by its small weight, the marks of x and y on the other criteria of S are good, and the relatively bad overall utility of x compared to y is due to the criteria not in S]$_S$.

Let us turn now to the computation of a minimal element of $Ex(x, y, v, \text{Pess}, \psi_{\text{IVT}})$. It is not necessary to explore this combinatorial structure. We use the characterization of the situation of the anchor ψ_{RMG} described in Proposition 19 given in Section 7.3. According to this proposition, if there exists a permutation that inverts the preference between x and y, then one of the three conditions (i), (ii) or (iii) in Proposition 19 is violated.

Proposition 20 gives an equivalent condition for the satisfaction of (i). This condition can be easily checked. In the proof of this proposition, a permutation $\pi_1^\star$ is constructed when (i) is violated. In this case, we have $y \preceq^{\text{Pess}}_{\pi_1^\star} x$ by construction.

Condition (ii) in Proposition 19 can be easily checked in practice. When it is violated, a permutation $\pi_2^\star$ is constructed in the proof of this proposition. It satisfies $y \preceq^{\text{Pess}}_{\pi_2^\star} x$ by construction.

Condition (iii) in Proposition 19 is equivalent to relation (12) according to Proposition 21. Condition (12) can be easily checked in practice. This condition is the conjunction of two conditions. When the first part of (12) is violated, one has $y \preceq^{\text{Pess}}_{\pi_3^\star} x$, where $\pi_3^\star$ is defined in the proof of Proposition 21. When the second part of (12) is violated, one has $y \preceq^{\text{Pess}}_{\pi_4^\star} x$, where $\pi_4^\star$ is defined in the proof of Proposition 21.

A permutation that ensures the switch of preferences is thus easily constructed in all cases. One can refine these permutations to obtain cycles with the smallest cardinality. This provides a minimal element of $Ex(x, y, v, \text{Pess}, \psi_{\text{IVT}})$.

7. Anchor "*remaining case*"

We assume in this section that the anchor ψ_{RMG} applies, and thus

$$A^+(y,x) \neq N, \quad y \succ^{\mathcal{F}}_{w^{\mathcal{F}}} x \quad \text{and} \quad \forall \pi \in \Pi(N), \quad y \succ^{\mathcal{F}}_{\pi\circ v} x. \tag{6}$$

The following result shows that the last relation in (6) implies that the middle relation in (6) also holds.

Lemma 4. *For all three models (namely $\succeq_w^{\text{Maj}}$, $\succeq_w^{\text{Pess}}$ and $\succeq_w^{\text{EU}}$), one has*

$$\forall \pi \in \Pi(N), \quad y \succ^{\mathcal{F}}_{\pi \circ v} x \quad \Longrightarrow \quad y \succ^{\mathcal{F}}_{w^{\mathcal{F}}} x.$$

It proves that when anchor "*invert*" does not apply, the decision would be the same with the reference weights. Furthermore, the preference of y over x is probably strong with the reference weights.

7.1. Expected utility model

By Lemma 3, if there does not exist $\pi \in \Pi(N)$ such that $H^{\text{EU}}_{\pi \circ v}(y,x) \leqslant 0$, then

$$0 < \min_{\pi \in \Pi(N)} H^{\text{EU}}_{\pi \circ v}(y,x) = H^{\text{EU}}_{\underline{\pi}_N \circ v}(y,x).$$

A necessary and sufficient condition for the anchor ψ_{RMG} to be applied is that

$$A^+(y,x) \neq N \quad \text{and} \quad H^{\text{EU}}_{\underline{\pi}_N \circ v}(y,x) > 0.$$

Lemma 5. *One has for any $v \in \overline{\mathcal{W}}(\text{EU})$ and any $z \in [0,1]^n$*

$$h^{\text{EU}}_{w^{\text{EU}}}(z) = \frac{1}{n!} \sum_{\pi \in \Pi(N)} h^{\text{EU}}_{\pi \circ v}(z).$$

There is a geometrical interpretation of this result. The set of weights $\pi \circ v$ where π can be any permutation forms the vertices of a polyhedron (interpreted as a set of normalized weights). The arithmetic mean is the centre of gravity of these vertices.

Proposition 13. *We have*

$$H^{\text{EU}}_{\underline{\pi}_N \circ v}(y,x) \leqslant H^{\text{EU}}_{w^{\text{EU}}}(y,x).$$

Moreover, assume that

$$\exists i, j \in N, \quad \Delta_i \neq \Delta_j. \tag{7}$$

Then $H^{\text{EU}}_{\underline{\pi}_N \circ v}(y,x) = H^{\text{EU}}_{w^{\text{EU}}}(y,x)$ if and only if for all $i \in N$, $v_i = \frac{1}{n}$.

Lemma 4 when $\mathcal{F} = \text{EU}$ follows from the first part of Proposition 13.

When anchor ψ_{RMG} applies, there are both positive and negative arguments. Hence (7) holds. Proposition 13 suggests that there are basically two cases to explain why y remains preferred to x even when the weights are switched. The first case occurs when the weights are more or less the same (hence $v \approx w^{\text{EU}}$). A permutation of the weights has indeed no impact on the comparison of x and y. In the other case, the weights are significantly different. Hence, since y remains preferred to x for every permutation of the weights, y is expected to be much better than x on average. We need the following two propositions to understand the separation between the previous two cases. More precisely, they establish a relationship between the closeness of v to the reference weight w^{EU}, and the closeness of $H^{\text{EU}}_{w^{\text{EU}}}(y,x)$ to $H^{\text{EU}}_{\underline{\pi}_N \circ v}(y,x)$.

Proposition 14. *If (7) is satisfied and $v \in \overline{\mathcal{W}}(\text{EU})$ is different from w^{EU}, then*

$$\frac{H^{\text{EU}}_{w^{\text{EU}}}(y,x) - \min_{\pi \in \Pi(N)} H^{\text{EU}}_{\pi \circ v}(y,x)}{\max_{\pi \in \Pi(N)} H^{\text{EU}}_{\pi \circ v}(y,x) - \min_{\pi \in \Pi(N)} H^{\text{EU}}_{\pi \circ v}(y,x)} \geqslant \frac{1}{n}.$$

Moreover this inequality is sharp in the sense that one can find x, y, v such that the previous relation is satisfied with an equality.

Proposition 15. *We set*

$$\nu := \max_{i \in N} \left| v_i - \frac{1}{n} \right|, \qquad \delta := \max_{i,j} |\Delta_i - \Delta_j| \quad \textit{and} \quad \chi^{\text{EU}} := \left| H^{\text{EU}}_{w^{\text{EU}}}(y,x) - H^{\text{EU}}_{\underline{\pi}_N \circ v}(y,x) \right|.$$

Then

$$\nu \leqslant \frac{(n-1)}{n} \frac{H^{\text{EU}}_{\overline{\pi}_N \circ v}(y,x) - H^{\text{EU}}_{\underline{\pi}_N \circ v}(y,x)}{\delta} \quad \textit{and} \quad \nu \leqslant \frac{(n-1)\chi^{\text{EU}}}{\delta}.$$

Moreover, these inequalities are sharp.

Following Propositions 14 and 15, we have two cases.

- $\nu \approx 0$ *(say* $\nu \leqslant \underline{\varepsilon}_\nu$ *in practice, where* $\underline{\varepsilon}_\nu$ *is a parameter).*

  Interpretation 5. The weights ν of the criteria are neutral in the sense that there is no specificity that could favour a criterion or disfavour another criterion. Hence H^{EU}_ν is almost equal to the arithmetic mean $H^{\mathrm{EU}}_{w^{\mathrm{EU}}}$. This means that $y \succ^{\mathrm{EU}}_\nu x$ follows from the relation $y \succ^{\mathrm{EU}}_{w^{\mathrm{EU}}} x$. This latter relation is easy to understand and to be checked by the user.

  The explanation can thus be the following one:

  > y is preferred to x since y is on average better than x and all the criteria have almost the same weights.

- $\nu \not\approx 0$ *(say* $\nu > \underline{\varepsilon}_\nu$ *in practice).*

  Interpretation 6. When the weights of the criteria are very different, the large weights can be assigned indifferently on the positive and negative arguments, without inverting the preference of y over x. One feels intuitively that this comes from the fact that the value of Δ on $A^+(y,x)$ is on average much larger than the value of Δ on $A^-(y,x)$.

  From Proposition 15, when $\nu > \underline{\varepsilon}_\nu$, quantity $\frac{\chi^{\mathrm{EU}}}{\delta}$ is not small. In this ratio, the denominator δ acts as a normalization. We do not focus on the absolute values of the difference between $H^{\mathrm{EU}}_{w^{\mathrm{EU}}}(y,x)$ and $H^{\mathrm{EU}}_{\nu_{\underline{\pi}_N}}(y,x)$, since a small value of χ^{EU} does not necessarily mean that the comparison of x and y is obvious. It may indeed result from the fact that the two options x and y have relatively close marks.
  Here $H^{\mathrm{EU}}_{w^{\mathrm{EU}}}(y,x)$ is significantly larger than $H^{\mathrm{EU}}_{\nu_{\underline{\pi}_N}}(y,x) > 0$. The fact that $H^{\mathrm{EU}}_{w^{\mathrm{EU}}}(y,x)$ is significantly larger than zero means that the positive arguments in favour of y are on average significantly larger than the negative arguments. We consider two subcases.
  Firstly, we consider the case when ν is relatively small (say $\underline{\varepsilon}_\nu < \nu < \overline{\varepsilon}_\nu$ in practice, where $\overline{\varepsilon}_\nu$ is a parameter). The explanation can thus be the following one:

  > y is preferred to x since the intensity of preference y over x on $A^+(y,x)$ is significantly larger than the intensity of preference of x over y on $A^-(y,x)$, and all the criteria have more or less the same weights.

  There remains to consider the case when ν is large (say $\nu \geqslant \overline{\varepsilon}_\nu$ in practice). Then from Proposition 15, quantity $\frac{\chi^{\mathrm{EU}}}{\delta}$ is large. The explanation can thus be the following one:

  > y is preferred to x since the intensity of preference y over x on $A^+(y,x)$ is much larger than the intensity of preference of x over y on $A^-(y,x)$.

7.2. Weighted majority model

Lemma 6. *One has for any* $\nu \in \overline{\mathcal{W}}(\mathrm{Maj})$

$$H^{\mathrm{Maj}}_{w^{\mathrm{Maj}}}(y,x) = \frac{1}{n!} \sum_{\pi \in \Pi(N)} H^{\mathrm{Maj}}_{\pi \circ \nu}(y,x).$$

Proposition 16. *Consider two alternatives* $x, y \in X$*. We have*

$$H^{\mathrm{Maj}}_{w^{\mathrm{Maj}}}(y,x) \geqslant H^{\mathrm{Maj}}_{\underline{\pi}_N \circ \nu}(y,x).$$

Assume that

$$A^+(y,x) \neq N, \qquad A^-(y,x) \neq N \quad \text{and} \quad A^=(y,x) \neq N. \tag{8}$$

Then, $H^{\mathrm{Maj}}_{\underline{\pi}_N \circ \nu}(y,x) = H^{\mathrm{Maj}}_{w^{\mathrm{Maj}}}(y,x)$ *if and only if* $\nu = w^{\mathrm{Maj}}$.

Condition (8) means that x does not dominates strictly y on all criteria, y does not dominates strictly x on all criteria, and x is not identical to y on all criteria.

A necessary and sufficient condition for the anchor ψ_{RMG} to be applied is that

$$A^+(y,x) \neq N \quad \text{and} \quad H^{\text{Maj}}_{\underline{\pi}_N \circ \nu}(y,x) > 0.$$

Lemma 4 when $\mathcal{F} = \text{Maj}$ follows from the first part of Proposition 16.

Proposition 17. *Let* $\chi^{\text{Maj}} := |H^{\text{Maj}}_{w^{\text{Maj}}}(x,y) - H^{\text{Maj}}_{\nu_{\underline{\pi}_N}}(x,y)|$. *Then for all* $i \in N$

$$\left| \nu_i - \frac{1}{n} \right| \leqslant 4n(n-1)\chi^{\text{Maj}}.$$

The explanation that can be generated is more or less similar to Section 7.1. We have the following cases.

- $\nu \approx 0$ *(say* $\nu \leqslant \underline{\varepsilon}_\nu$ *in practice)*. Hence H^{EU}_ν is almost equal to the arithmetic mean $H^{\text{EU}}_{w^{\text{EU}}}$. This means that $y \succ^{\text{Maj}}_\nu x$ follows from the relation $y \succ^{\text{Maj}}_{w^{\text{EU}}} x$. The explanation can thus be the following one:

  `y is preferred to x since there are more criteria for which y is better than x than criteria for which x is better than y, and the aggregation model is almost the majority rule.`

- $\nu \not\approx 0$ *(say* $\nu > \underline{\varepsilon}_\nu$ *in practice)*. Proposition 17, quantity χ^{Maj} is not small. Hence $H^{\text{EU}}_{w^{\text{Maj}}}(y,x)$ is significantly larger than $H^{\text{Maj}}_{\underline{\pi}_N \circ \nu}(y,x)$. Hence the decision made by the majority rule is very strong. The explanation tells just a little bit of the specificities of x and y, namely

  `y is preferred to x since the criteria for which y is better than x are ON AVERAGE MUCH stronger than the criteria for which x is better than y.`

7.3. Pessimistic qualitative model

Lemma 4 when $\mathcal{F} = \text{Pess}$ follows from the following proposition.

Proposition 18. *Let* $\nu \in \overline{\mathcal{W}}(\text{Pess})$. *If for every permutation* $\pi \in \Pi(N)$, $h^{\text{Pess}}_{\pi \circ \nu}(y) > h^{\text{Pess}}_{\pi \circ \nu}(x)$ *then*

$$\bigwedge_{i=1}^n y_i > \bigwedge_{i=1}^n x_i.$$

For $w \in \mathcal{W}(\text{Pess})$, we have $h^{\text{Pess}}_w(z) = h^{\text{Pess},A^+(y,x)}_w(z) \wedge h^{\text{Pess},A^=(y,x)}_w(z) \wedge h^{\text{Pess},A^-(y,x)}_w(z)$. We now give a characterization of anchor ψ_{RMG} for the Pess model through the following three propositions.

Proposition 19. *Let*

$$E := \{1\} \cup \left\{x_i \colon i \in A^+(y,x)\right\} \cup \left\{1 - \nu_{\pi^\nu_N(j)} \colon j \in \left\{1, \ldots, \left|A^+(y,x)\right|\right\}\right\}.$$

Let e *be the median value of the discrete set* E.

We have

$$\forall \pi \in \Pi(N) \quad h^{\text{Pess},A^+(y,x)}_{\pi \circ \nu}(x) \leqslant e. \tag{9}$$

Moreover, $H^{\text{Pess}}_{\pi \circ \nu}(y,x) > 0$ *for every* $\pi \in \Pi(N)$ *if and only if the following three propositions hold*:

(i) $\forall \pi \in \Pi(N)$, $h^{\text{Pess},A^+(y,x)}_{\pi \circ \nu}(y) > h^{\text{Pess},A^+(y,x)}_{\pi \circ \nu}(x)$,
(ii) $\forall i \in A^-(y,x) \cup A^=(y,x)$, $y_i > e$,
(iii) $\forall \pi \in \Pi(N)$, $h^{\text{Pess}}_{\pi \circ \nu}(x) < h^{\text{Pess},A^-(y,x)}_{\pi \circ \nu}(x) \wedge h^{\text{Pess},A^=(y,x)}_{\pi \circ \nu}(x)$.

Proposition 20. *Let* $i^+ \in A^+(y,x)$ *such that* $\bigwedge_{i \in A^+(y,x)} y_i = y_{i^+}$. *Let* $V_{i^+} = \{k \in N \colon 1 - \nu_k \geqslant y_{i^+}\}$, *and* $k^+ \in V_{i^+}$ *(defined when* $V_{i^+} \neq \emptyset$*) such that* $1 - \nu_{k^+} = \bigwedge_{k \in V_{i^+}} (1 - \nu_k)$. *The following two propositions are equivalent*:

(i) $\forall \pi \in \Pi(N)$, $h^{\text{Pess},A^+(y,x)}_{\pi \circ \nu}(y) > h^{\text{Pess},A^+(y,x)}_{\pi \circ \nu}(x)$,
(ii) *either* $V_{i^+} = \emptyset$ *or*

$$|V_{i^+}| < \left|\left\{j \in A^+(y,x) \colon x_j < 1 - \nu_{k^+}\right\}\right|. \tag{10}$$

Proposition 21. *Let* $j^+ \in A^-(y,x) \cup A^=(y,x)$ *such that* $x_{j^+} = \bigwedge_{j \in A^-(y,x) \cup A^=(y,x)} x_j$. *Let*

$$I = \left\{ i \in A^+(y,x) \colon x_i < x_{j^+} \right\}.$$

Then the following relations are equivalent:

$$\forall \pi \in \Pi(N) \colon \quad h^{\mathrm{Pess}}_{\pi \circ v}(x) < h^{\mathrm{Pess}, A^-(y,x)}_{\pi \circ v}(x) \wedge h^{\mathrm{Pess}, A^=(y,x)}_{\pi \circ v}(x) \tag{11}$$

and

$$I \neq \emptyset \quad \text{and} \quad \left| \{ k \in N \colon 1 - v_k \geqslant x_{j^+} \} \right| < |I|. \tag{12}$$

A necessary and sufficient condition for having $y \succ^{\mathrm{Pess}}_{\pi \circ v} x$ for all $\pi \in \Pi(N)$ is the satisfaction of both condition (ii) in Proposition 19, condition (ii) in Proposition 20 and relation (12) in Proposition 21. From these properties, one can easily check whether anchor ψ_{RMG} can be applied.

We are now interested in the generation of the explanation to the user. From Proposition 19, the three conditions (i), (ii) and (iii) hold. According to Proposition 20, one has either $V_{i^+} = \emptyset$ or (10).

Assume first that $V_{i^+} = \emptyset$. The worse scores of x on $A^+(y,x)$ are clearly smaller than y_{i^+}. When $V_{i^+} = \emptyset$, there is no weight that can hide the difference between the worst score y_{i^+} of y on $A^+(y,x)$ and the even worse score of x. This means that all the weights are large enough, and thus that the aggregation function is close to the minimum. The first part of the explanation can thus be as follows:

> y is preferred to x since there is no weight that can hide the worse score of x compared to y on $A^+(y,x)$. [The aggregation function is almost the Minimum.] Recall that y is strictly better than x on $A^+(y,x)$. ...

The other case is when (10) holds. The set $L := \{ j \in A^+(y,x) \colon x_j < 1 - v_{k^+} \}$ contains the criteria in $A^+(y,x)$ for which x has a very bad score. The set V_{i^+} gathers all the weights that can hide the worst score y_{i^+} of y on $A^+(y,x)$. If $|V_{i^+}| < |L|$, then there is at least one weight for which the worse score of x compared to y cannot be saved. The first part of the explanation can thus be as follows:

> y is preferred to x since there is at least a weight that cannot hide the worse scores of x compared to y on $A^+(y,x)$. [The aggregation function is close to the Minimum.] Recall that y is strictly better than x on $A^+(y,x)$. ...

The mid-part of the explanation concerns the point (ii) in Proposition 19. The overall score of x is lower than e. Thus y is better than the overall utility of x, over $A^-(y,x) \cup A^=(y,x)$. The next part of the explanation can thus be the following one:

> ... y is better than the overall score of x, on the criteria $A^-(y,x) \cup A^=(y,x)$. Recall that y is not better than x on the criteria $A^-(y,x) \cup A^=(y,x)$. ...

The last part of the explanation concerns the point (iii) in Proposition 19, and Proposition 21. From relation $I \neq \emptyset$, the worse scores of x are attained in $A^+(y,x)$. From relation $|\{k \in N \colon 1 - v_k \geqslant x_{j^+}\}| < |I|$, there is at least one weight for which the worse scores of x in $A^+(y,x)$ cannot be saved up to the scores of x on $A^-(y,x) \cup A^=(y,x)$. The end of the explanation can thus be as follows:

> ... Finally, the worse scores of x are attained in $A^+(y,x)$. There is at least one weight for which the worse scores of x in $A^+(y,x)$ cannot be saved up to the scores of x on $A^-(y,x) \cup A^=(y,x)$. [The aggregation function is close to the Minimum.]

8. Illustration and experimental results

The aim of this section is threefold. Firstly, the process for generating the explanation is summarized in Section 8.1.

Secondly, we wish to show that our approach helps the user to better understand the decision than when he tries to do it on his own. The anchors can be seen as meta-explanations or explanation schemas [57]. It aims at selecting the decisive criteria in the comparison $y \succ^{\mathcal{F}}_{v} x$. The decisive criteria are the most influencing criteria in the decision, and their determination helps the user to better understand the decision and also to take actions on the decision such as improvement actions. We illustrate how the anchors behave on several examples.

The illustrations focus on the two models EU and Pess. The reason why the model Maj is not considered is that the Maj model can be seen as a particular case of the EU model applied to binary alternatives. Indeed, we have

$$y \succ_v^{\text{Maj}} x \iff y^* \succ_v^{\text{EU}} x^*$$

where x^* and y^* are defined from x and y by the relations $x_i^* = 1$ if $x_i \geqslant y_i$ and $x_i^* = 0$ otherwise, and $y_i^* = 1$ if $y_i \geqslant x_i$ and $y_i^* = 0$ otherwise.

Sections 8.2 and 8.3 discuss the two anchors ψ_{NOA} and ψ_{IVT} respectively, and show several illustrative examples. We do not consider specifically the two other anchors ψ_{ALL} and ψ_{RMG} since their interpretation is more straightforward and in particular they do not require the selection of a subset of criteria. Moreover, Section 8.4 presents a comparative study of several explanations on the three anchors ψ_{NOA}, ψ_{IVT} and ψ_{RMG} for the EU model.

The third aim of this section is to present some experimental results concerning the practical determination of the explanation (see Section 8.5).

8.1. Generation process of the explanations

The generation of the explanation for a given instance $(x, y, v) \in \mathcal{D}(\mathcal{F})$ is obtained as follows. One aims at determining a non-dominated element of $Ex(x, y, v, \mathcal{F})$ w.r.t. $\trianglelefteq$.

Due to the lexicographic structure of $\trianglelefteq$ over $Ex(x, y, v, \mathcal{F})$, one first identifies the anchor to be applied. The conditions under which each anchor can be applied have been presented throughout this paper. More precisely, these are, for every $\psi \in \Psi$, necessity and sufficient conditions on (x, y, v) to have $(x, y, v) \in \mathcal{D}(\mathcal{F}, \psi)$. We identify the most preferred anchor $\hat{\psi}$ in the set $\{\psi \in \Psi\colon\ (x, y, v) \in \mathcal{D}(\mathcal{F}, \psi)\}$ according to $\trianglelefteq$.

Then we compute a non-dominated explanation set $\hat{\mathcal{A}}$ in $Ex(x, y, v, \mathcal{F}, \hat{\psi})$ for this anchor $\hat{\psi}$. This was presented in the previous sections. Considering the EU model, it is trivial for the ψ_{ALL} and ψ_{RMG} anchors, it requires to sort a vector for the ψ_{NOA} anchor, and it is solved by Algorithm **Algo-**EU for the anchor ψ_{IVT}.

Finally, from $(\hat{\psi}, \hat{\mathcal{A}})$, the corresponding textual explanation is generated. Examples of text have been presented earlier in the paper.

8.2. Illustration of the anchor "not on average"

8.2.1. Model EU

Let us first compare the Klein [35] approach with our. In our approach, a selected criterion i for anchor ψ_{NOA} necessarily fulfils $(v_i - \frac{1}{n})\Delta_i > 0$ (see Proposition 1). A criterion is likely to be selected if it is a negative and weak argument, or if it is a positive and strong argument. On the other hand, in the Klein approach, the criteria for which $|\Delta_i| \approx 0$ and $v_i \approx 0$ are unlikely to be selected.

There are three main situations in which the two approaches give different results. The first two ones are not specific to the anchor ψ_{NOA} while the third one is specific to this anchor. In the first one, criterion i is a weak and negative argument. It is very intuitive to show such a criterion. Our approach selects this criterion while Klein's approach usually does not. In the second situation, criterion i is a strong and negative argument. Klein's approach selects this criterion while ours does not. This non-decisive criterion is compensated by positive arguments since y is after all preferred to x. Hence, it is not necessary at all to show this criterion to the user in a synthesis. In the last situation, criterion i is a medium argument ($v_i \approx \frac{1}{n}$) with $|\Delta_i| \gg 0$. Klein's approach often selects this criterion while ours usually does not. The strength of the argument corresponds to the reference weight $\frac{1}{n}$ that one might have in mind. Since it is not different from the prior, it is not useful to show it explicitly to the user. It is better to focus on the criteria that have less standard weights (very small or large).

We now give several examples to illustrate the main properties of our approach. As shown in Example 5, our approach selects very few arguments (only one in Example 5) when the decision is clear cut. On the opposite side, when the decision is very tight, more criteria are selected with our approach (see Example 6). It is indeed intuitive that the tighter the decision, the more arguments are selected. This does not occur with the Klein approach (see Examples 5 and 6). Other examples with more criteria can be found in Section 8.4

Example 5 *(The decision is relatively clear-cut, and the two approaches select opposite arguments).* Consider the situation where $x = (0.42, 0.66, 0.66, 0.57)$, $y = (0.54, 0.04, 0.89, 0.76)$ and $v = (0.41, 0.06, 0.24, 0.29)$. Then $\Delta = (0.12, -0.62, 0.23, 0.19)$. y is significantly preferred to x since $h_v^{\text{EU}}(y) = 0.66$ and $h_v^{\text{EU}}(x) = 0.54$.

Criteria	1	2	3	4
$(v_i - \frac{1}{n}) \times \Delta_i$	0.019	**0.118**	−0.002	0.008
$\text{compel}_i = v_i\lvert\Delta_i\rvert$	**0.049**	0.037	**0.055**	**0.055**

The Klein approach selects most of the arguments in this case – namely criteria 1, 3 and 4 – even though the decision is relatively clear-cut. On the other hand, our approach selects only one argument – namely criterion 2 – which is enough in this example. Looking at the example, criterion 2 is indeed the main argument why x is better on average than y but y is preferred to x with the decision model.

Example 6 *(The decision is tight and our approach returns almost all arguments).* Let us consider $x = (0.95, 0.67, 0.64, 0.27, 0.39)$, $y = (0.3, 0.37, 0.41, 0.94, 0.49)$ and $v = (0.18, 0.11, 0.12, 0.24, 0.35)$. Then $\Delta = (-0.65, -0.3, -0.23, 0.67, 0.1)$. y is just slightly preferred to x since $h_v^{\mathrm{EU}}(y) = 0.54$ and $h_v^{\mathrm{EU}}(x) = 0.52$.

Criteria	1	2	3	4	5
$(v_i - \frac{1}{n}) \times \Delta_i$	0.013	**0.027**	**0.018**	**0.027**	**0.015**
compel_i	**0.117**	0.033	0.028	**0.161**	0.035

The Klein approach selects criteria 1 and 4. On the other hand, in our approach, criteria 2, 3, 4, 5 are selected.

8.2.2. Model Pess

Several examples are given. Example 7 illustrates the case when $M^* = \emptyset$. This example shows that when the decision is clear-cut, the number of selected criteria is small. The more general case $M^* \neq \emptyset$ is represented by Example 8.

Example 7 *(Clear-cut decision with* 10 *criteria).* Consider the following values of x, y and v.

$$y = (0.1, 0., 0.66, 0.97, 0.45, 0.71, 0.57, 0.77, 0.09, 0.69),$$

$$x = (0.14, 0.46, 0.26, 0.8, 0.69, 0.64, 0.07, 0.17, 0.94, 0.45),$$

$$v = (0.13, 0.53, 0.45, 0.12, 0.83, 0.57, 1.0, 0.19, 0.58, 0.23).$$

We have $h_v^{\mathrm{Pess}}(x) = 0.07$, $h_v^{\mathrm{Pess}}(y) = 0.42$ and $y \succ_v^{\mathrm{Pess}} x$. The selected criterion is 2. Moreover, criterion $k^x = 7$ is decisive since the worse score of x is reached on this criterion, it is a positive and strong argument. We have $M = \{2\}$ and $M^* = \emptyset$. The generic text can be generated from S, k^x and M^*. The last sentence in this generic explanation text is: "*y is sufficiently good on the remaining criteria* 1, 3, 4, 5, 6, 8, 9 *and* 10". We remark that the scores of y on the two criteria 1 and 9 are not good (0.1 and 0.09) but they are larger than the over score $h_v^{\mathrm{Pess}}(x) = 0.07$ of x. The decision would not have changed if the weight $w_i^{\mathrm{Pess}} = 1$ were assigned to the criteria 1 and 9. It turns out that these two criteria have a medium importance, which implies that the overall score of y is much larger than that of x. But it is not necessary to show this in the explanation.

Example 8 *(Tight decision with* 10 *criteria).* Consider the following values of x, y and v.

$$y = (0.36, 0.01, 0.22, 0.11, 0.61, 0.4, 0.06, 0.07, 0.43, 0.61),$$

$$x = (0.25, 0.21, 0.14, 0.08, 0.53, 0.15, 0.2, 0.53, 0.87, 0.75),$$

$$v = (1.0, 0.02, 0.86, 0.23, 0.91, 0.53, 0.36, 0.17, 0.26, 0.53).$$

We have $h_v^{\mathrm{Pess}}(x) = 0.14$, $h_v^{\mathrm{Pess}}(y) = 0.22$ and $y \succ_v^{\mathrm{Pess}} x$. The selected criteria are 2, 7, 8. Moreover, $k^x = 3$, $M = \{2, 4, 7, 8\}$ and $M^* = \{4\}$. Indeed, the worse score of x is reached on this criterion 4, which is a positive and strong argument.

8.3. Illustration of the anchor "invert"

8.3.1. Model EU

In most of the cases that we have encountered, the cycles of a minimal permutation are of cardinality 2. A cycle of size 3 is presented in Example 9. The second example contains 10 criteria. It is also possible to consider all the explanation sets of minimal cardinality in order to generate the textual explanation (see Example 15 later).

Example 9 *(Permutation between three criteria).* Let $x = (0.89, 0.03, 0.07, 0.32, 0.38)$, $y = (0.36, 0.76, 0.6, 0.25, 0.75)$ and $v = (0.06, 0.11, 0.21, 0.29, 0.33)$. Then we have $\Delta = (-0.53, 0.73, 0.53, -0.07, 0.37)$. We have $\mathrm{compel} = (0.031, 0.08, \mathbf{0.111}, 0.02, \mathbf{0.122})$. The Klein approach selects criteria 5 and 3. On the other hand, the minimal element in $Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{IVT}})$ is $\{\{1, 3, 5\}\}$. For the associated permutation π, the score of criteria 1, 5 and 3 are assigned to the weight of criteria 5, 3 and 1 respectively (i.e. $\pi(5) = 1$, $\pi(3) = 5$ and $\pi(1) = 3$). We note that π is the most disfavourable permutation for y on $\{1, 3, 5\}$. It is easy to see that $R^* = \{(1, 3), (1, 5)\}$. Finally, $C_{\mathrm{PS}} = \{3, 5\}$, $C_{\mathrm{NW}} = \{1\}$, and the other sets of the tuple C are empty (see Section 6.1).

Example 10 *(Example with* 10 *criteria).* Let

$$y = (0.45, 0.64, 0.86, 0.76, 0.87, 0.54, 0.17, 0.04, 0.55, 0.05),$$

$$x = (0.61, 0.28, 0.08, 0.02, 0.81, 0.15, 0.16, 0.38, 0.24, 0.75),$$

$$\Delta = (-0.16, 0.36, 0.78, 0.74, 0.06, 0.39, 0.01, -0.34, 0.31, -0.7),$$

$$v = (0.13, 0.04, 0.12, 0.1, 0.07, 0.19, 0.15, 0.03, 0.01, 0.16).$$

We have compel = (0.021, 0.014, **0.094**, **0.074**, 0.004, **0.074**, 0.001, 0.01, 0.003, **0.112**). The Klein approach selects criteria 10, 3, 6 and 4. On the other hand, in our approach, the explanation set is $\{\{6, 8\}, \{3, 9\}\}$. Regarding the pair $\{6, 8\}$, criterion 6 is positive and important and criterion 8 is negative and not important. Regarding the pair $\{3, 9\}$, both criteria are positive arguments, and criterion 3 is more positive and more important than criterion 9. Criterion 9 is not mentioned. Finally, $C_{\text{PS}} = \{3, 6\}$, $C_{\text{NW}} = \{8\}$, and the other sets of the tuple C are empty.

8.3.2. Model Pess

The following examples illustrate different situations regarding the size of the minimal explanation sets: the minimal explanation set is composed of one subset of cardinality 3 in Example 11, and of two subsets of cardinality 2 in Example 12.

Example 11 *(A selected set of cardinality* 3*).* Let $y = (0.83, 0.66, 0.19, 0.55, 0.94, 0.75)$, $x = (0.63, 0.0, 0.15, 0.98, 0.97, 0.81)$ and $v = (0.82, 0.8, 1.0, 0.1, 0.28, 0.26)$. We have $y \succ_v^{\text{Pess}} x$. The minimum selected coalitions in $Ex(x, y, v, \text{Pess}, \psi_{\text{IVT}})$ are of cardinality 3: $\{\{2, 3, 4\}\}$, $\{\{2, 3, 5\}\}$ and $\{\{2, 3, 6\}\}$. We note that criteria 2 and 3 are present in all three selected sets. These two criteria are positive arguments with a large weight. The overall score of x and y is attained at criterion 3 that has a very large importance. The other criteria 4, 5 and 6 that appear in the selected criteria are negative arguments with a small weight.

Example 12 *(Example with* 9 *criteria).* Consider the following values of x, y and v.

$$y = (0.43, 0.57, 0.28, 0.5, 0.46, 0.3, 0.6, 0.48, 0.83),$$

$$x = (0.05, 0.89, 0.22, 0.0, 0.82, 0.74, 0.86, 0.17, 0.76),$$

$$v = (1.0, 0.7, 0.67, 0.29, 0.27, 0.29, 0.78, 0.87, 0.67).$$

We have $y \succ_v^{\text{Pess}} x$. The selected set is composed of two cycles: $\{\{1, 2\}, \{6, 8\}\}$. In the cycle $\{1, 2\}$, criterion 1 is a positive argument with a very large importance, and criterion 2 is a negative argument with an importance lower than that of criterion 1. Moreover, the overall score of x is attained at the very important criterion 1 for which x has a very small score. Concerning the cycle $\{6, 8\}$, criterion 6 is a negative argument with a small weight and criterion 8 is a positive argument with a large weight. The explanation says that criterion 1 that is a positive argument is more important than criterion 2 that is a negative argument, and criterion 8 that is a positive argument is more important than criterion 6 that is a negative argument. Finally, on the remaining criteria, y has rather good evaluations.

8.4. Comparison of several anchors on the EU *model*

In order to compare the anchors, we fix the two vectors x and y, and we will consider different values of the weight v. Clearly, if anchor ψ_{ALL} holds for one weight vector v, then this anchor also holds for any other value of v. Hence, we focus on the three other anchors ψ_{NOA}, ψ_{IVT} and ψ_{RMG}. We consider the following values of x and y

$$y = (0.99, 0.35, 0.31, 0.51, 0.62, 0.57, 0.52),$$

$$x = (0.5, 0.06, 0.03, 0.95, 0.87, 0.2, 0.95).$$

Option y is on average better than x. The following examples of v are given.

Example 13 *(Anchor* ψ_{NOA}*).* Let $v = (0.06, 0.11, 0.19, 0.11, 0.31, 0.08, 0.14)$. Option x is preferred to y. The Klein approach selects the criteria 5 and 7. The selected criteria for this anchor are the criteria 1 and 5. Criterion 1 is a weak and negative argument, and criterion 5 is a positive and strong argument. The other criteria are not explicitly mentioned since their importance are relatively close to the mean importance $\frac{1}{n}$.

Example 14 *(Anchor* ψ_{NOA}*).* Let $v = (0.14, 0.05, 0.17, 0.23, 0.17, 0.11, 0.13)$. Option x is preferred to y. The Klein approach highlights the criteria 4 and 1. Our approach selects the criteria 2 and 4. Criterion 2 is a weak and negative argument, and criterion 4 is a positive and strong argument. The other criteria have, as in the previous case, an importance that is close to the mean importance $\frac{1}{n}$, and are thus not explicitly mentioned.

Example 15 *(Anchor* ψ_{IVT}*).* Let $v = (0.11, 0.14, 0.13, 0.02, 0.27, 0.25, 0.08)$. Option y is preferred to x. The Klein approach highlights the criteria 6 and 5. There are two elements of $Ex(x, y, v, \text{EU}, \psi_{\text{IVT}})$ of cardinality 2: $\{\{4, 6\}\}$ and $\{\{6, 7\}\}$. In the selection $\{4, 6\}$, criterion 6 is a positive and strong argument, and criterion 4 is a very weak and negative argument. We note that criterion 6 also belongs to the second selection and that criterion 7 in the second selection is a weak and negative argument. At the end, the three criteria 4, 6 and 7 are displayed in the explanation.

n	ψ_{ALL}	ψ_{NOA}	ψ_{IVT}				ψ_{RMG}
	occur.	occur.	occur.	mean execut. time	perc. tree explor.	$1st$ coalition expl. set	occur.
4	12.5%	14.9%	34.8%	$0.017ms$	10.4%	84.8%	37.8%
6	3.1%	15.7%	49.8%	$0.03ms$	0.4%	73.2%	31.4%
8	0.8%	16.0%	59.5%	$0.07ms$	0.01%	65.8%	23.8%
10	0.2%	16.2%	66.3%	$0.45ms$	$3.\ 10^{-4}\%$	60.6%	17.3%
12	$5.\ 10^{-2}\%$	16.5%	70.7%	$8.8ms$	$9.\ 10^{-6}\%$	57.2%	12.8%
14	$1.\ 10^{-2}\%$	16.9%	73.3%	$53.8ms$	$2.\ 10^{-124}\%$	56.1%	9.8%
16	$3.\ 10^{-3}\%$	17.8%	74.8%	$3874.0ms$	$1.\ 10^{-253}\%$	54.5%	7.4%
18	$8.\ 10^{-4}\%$	18.3%	77.1%	$9.2\ 10^{4}ms$	–	49.9%	4.6%
20	$2.\ 10^{-4}\%$	18.5%	78.7%	$8.6\ 10^{5}ms$	–	47.6%	2.8%

Fig. 1. Results of the experimental study. The columns "occur." show the percentage of occurrence of each anchor. The mean computation time is presented in the column "mean execut. time". The mean percentage of the tree that is explored during the search appears in the column "perc. tree explor.". Finally, the column "1*st* coalition expl. set" corresponds to the percentage of times the algorithm terminates at the very first step.

Example 16 *(Anchor ψ_{IVT}).* Let $v = (0.24, 0.2, 0.25, 0.06, 0.02, 0.19, 0.04)$. Option y is preferred to x. The Klein approach highlights the criteria 1 and 6. The preferred explanation set in $Ex(x, y, v, \text{EU}, \psi_{\text{IVT}})$ is $\{\{1, 7\}, \{3, 4\}\}$. In the pair $\{1, 7\}$, criterion 1 is a positive and strong argument, and criterion 7 is a weak and negative argument. In the pair $\{3, 4\}$, criterion 3 is a positive and strong argument, and criterion 4 is a weak and negative argument. The four criteria 1, 3, 4 and 7 are displayed in the explanation.

Example 17 *(Anchor ψ_{RMG}).* Let $v = (0.16, 0.14, 0.15, 0.1, 0.16, 0.15, 0.14)$. Option y is preferred to x. The Klein approach selects the criteria 1 and 7. We are in the situation of Interpretation 6: the intensity of preference of y over x on the positive criteria 1, 2, 3 and 6 is significantly larger than the intensity of preference of x over y on the negative criteria 4, 5 and 7. Moreover, all the criteria have more or less the same weights.

Example 18 *(Anchor ψ_{RMG}).* Let $v = (0.12, 0.16, 0.15, 0.16, 0.15, 0.14, 0.12)$. Option y is preferred to x. The Klein approach selects the criteria 4 and 1. We are in the situation of Interpretation 5. Compared to the previous situation, the weights are closer to the arithmetic mean. Even though y is on average significantly better than x, we do not need to use this argument here.

We have seen that the three anchors are triggered for different values of v. The seven criteria have been selected in the previous examples for the anchors ψ_{NOA} and ψ_{IVT}. This shows the influence of the weights on the selection process.

8.5. Experimental results

We present in this section the results of experimentations conducted on our approach. The tests are mainly concerned with the computational performance of the determination of a non-dominated explanation set. Among the three models EU, Maj and Pess, we restrict ourselves to the case of the EU model. We have seen that the methods and algorithms that select an explanation set are very similar for the two models EU and Maj. Hence it is not necessary to perform an experiment on both models. Moreover, the determination of a non-dominated explanation set can be almost done *by hand* for the model Pess (see the end of Section 6.3), and requires thus less computation time.

The computation that is necessary to determine a non-dominated explanation set is not time consuming for the anchors ψ_{ALL}, ψ_{NOA} and ψ_{RMG} since one needs at most to sort a vector of n components. By contrast, Algorithm **Algo-**EU used in anchor ψ_{IVT} requires the search of the explanation among a large tree. The computation time is thus shown only for the anchor ψ_{IVT}.

Our approach has been implemented in Java and tested on an Intel Pentium Core 2 computer with 2.66 GHz. The experimentations are performed on randomly generated instances from the set $\mathcal{D}(\text{EU})$. Fig. 1 shows the percentage of occurrence of each anchor in the experiment. Concerning Algorithm **Algo-**EU, the mean execution time is given. To analyse the efficiency of the branching strategy of this algorithm, Fig. 1 also indicates the mean percentage of the tree that is explored during the search. We have noticed that the algorithm often terminates at the first iteration. This situation arises when $\{T_1\}$ is an explanation set, where T_1 is the smallest element of $\mathcal{S}$ according to $\trianglelefteq_{\text{lexi}}$. Fig. 1 presents the results for values of n between 4 and 20. These are the most commonly encountered values for the number of attributes, in practice. According to Fig. 1, the algorithm terminates at the very first coalition in $\mathcal{S}$, in at least one case over two. This shows that the strategy that was chosen for the ranking of the coalitions of $\mathcal{S}$ according to $\trianglelefteq_{\text{lexi}}$ is efficient. Moreover, the percentage of the search tree that is explored by the algorithm decreases very rapidly with n. The worse scenario in Algorithm **Algo-**EU occurs when the only explanation set is the grand coalition N, that is for a permutation π such that $\mathcal{A}(\pi) = \{N\}$. In this case, the $2^{|\mathcal{S}|}$ subsets of $\mathcal{S}$, where $|\mathcal{S}| \leqslant 2^n - 1$, are explored before finding the explanation set.

For small values of n, the two anchors that occur most likely are ψ_{IVT} and ψ_{RMG}. The probability of occurrence for anchor ψ_{ALL} is $\frac{1}{2^{n-1}}$ and thus decreases very rapidly with n. For larger values of n, the preponderant case becomes anchor ψ_{IVT}.

If $\nu > \frac{1}{n}$, the weight vector ν is clearly not close to the reference weight w^{EU}. Concerning the anchor ψ_{RMG}, we have obtained a good separation of the three sub-cases on this anchor with the values $\underline{\varepsilon}_\nu = \frac{0.15}{n}$ and $\overline{\varepsilon}_\nu = \frac{0.3}{n}$.

9. Conclusion

We propose an approach to select the arguments to be used in the explanation of the prescription made by a multi-attribute decision model parameterized by weights assigned to the criteria. It is based on the analysis of the values of the weights together with the relative scores of the options to be compared. The general approach is applied on three decision models: the expected utility model (EU), the weighted majority model (Maj) and the weighted minmax model (Pess). These three models are the most well-known weight-based models in decision theory. Moreover, they represent very different visions of decision theory.

The idea of our approach is to look for some changes in the weight vector ν that yield an inversion of the prescription made by the decision model. The explanation focuses then on the criteria for which the weight vector has changed. The remaining criteria do not play any role in the inversion of the prescription and are thus not mentioned in the explanation. Not any change of the weights can be used since this change shall be explainable to the user. In this paper, two strategies for the modification of the weights are considered: the replacement of ν by some reference weights $w^{\mathcal{F}}$, and a permutation of the weights ν among the criteria. These two strategies lead to two different explanation strategies. They are called anchors ψ_{NOA} and ψ_{IVT} respectively.

There are several possible changes of the weights compatible to each anchor. All these admissible changes can be represented in a single combinatorial structure $Ex(x, y, \nu, \mathcal{F})$ containing all *explanation sets*. Since one is interested in the simplest explanation, the simplest changes are sought. In the structure $Ex(x, y, \nu, \mathcal{F})$, this is expressed by an order relation $\sqsubseteq$. One then looks at the non-dominated elements of $Ex(x, y, \nu, \mathcal{F})$, in the sense of $\sqsubseteq$.

The properties of the non-dominated explanation sets have been studied for the two anchors ψ_{NOA} and ψ_{IVT}. Concerning anchor ψ_{NOA}, we have shown, for the three models EU, Maj and Pess, that the positive selected arguments turn out to be strong and the negative selected arguments are weak. Moreover, the selected arguments for the qualitative Pess model are necessarily negative. This comes from the fact that the min operator expresses the principle of elimination among the criteria. An explanation can then easily be generated from these properties. The computation of a particular non-dominated explanation set is easy since one needs only to rank a vector. Concerning anchor ψ_{IVT}, we obtain roughly the same idea but in a weaker form. More precisely, the allocation of the weights to the criteria is not the most unfavourable one relatively to an inversion of the prescription, for the models EU and Maj. This implies that the positive arguments have more important weights than the negative arguments. A branch and bound algorithm has been proposed to compute a particular non-dominated explanation set. The underlying combinatorial structure is large since it is isomorphic to the set of coalition structures. In the algorithm, the coalition structures with coalitions of small cardinality are explored first. Experimental tests have shown that this strategy enables to find very quickly the explanation set in most of the cases.

The two anchors ψ_{NOA} and ψ_{IVT} do not cover all cases of $(x, y, \nu) \in \mathcal{D}(\mathcal{F})$. Another case, which is the simplest one (ψ_{ALL}), occurs when there is only positive arguments. The explanation does not need to mention the specificities of the decision model in this trivial situation. The last case (ψ_{RMG}) gathers all the situations not covered by the other anchors. Interestingly enough, we were able to divide $\mathcal{D}(\mathcal{F}, \psi_{\text{IVT}})$ in very few typical sub-cases. For instance, for the EU model, either the weights are close to the reference weights, or there are clearly more positive arguments than negative ones.

We have shown through numerous illustrative examples, that the explanation that is generated really helps to understand the decision. This comes from the fact that our approach selects the decisive criteria, that is the criteria for which a given change in the weight switches the decision. The type of relevant change is dependant on the anchor.

This work could benefit from several extensions. First of all, our framework could be applied to other weight-based decision models. An example of such model is the weighted maxmin function [22]. This model is the optimistic counterpart of the weighted minmax function h^{Pess}. Some preliminary results were given in [39] and suggest that the extension of our approach is possible for this model. Secondly, the approach could also be extended to models using more complex parameters than a simple weight assigned to each criterion. One may think of the Choquet and Sugeno integrals which extend the EU and Pess models respectively [15,53]. The parameters of these discrete integrals correspond to the concept of a capacity, which contains 2^n coefficients [15]. The generation of an explanation for the Choquet integral w.r.t. a particular case of a capacity were already addressed but it needs to be further studied [38,43].

In this paper, we have not put our attention in this paper to the structuration and expression of the selected arguments. The textual explanations that we proposed throughout this paper were given only to show the type of explanation that each anchor yields. This can be improved by incorporating dedicated techniques for the generation of natural language.

The explanations resulting from our approach aim at finding the very arguments that are at the root of the prescription. The explanation that we produce is more complex than that generated in [14,35,43], but our explanation reasoning is more sound. We analyse more deeply the specificity of the model and deduce from that the set of selected arguments. For this reason, our approach is more suited to applications where a high added-value of the decision model is expected. This means that the recipients of this approach are more domain experts than simple users of the Internet. To cite an example of application, one can mention the decision aid for the selection of candidate architectural options in the design of complex

systems, which requires elaborate decision models [47]. However, we have shown that the explanation that is generated is far from being complex for the EU and Maj models. An interesting property of our framework is indeed that the explanation adapts itself automatically to the complexity of the prescription. On the other hand, the explanation is more complex for the Pess model. The min and max binary operators are extensions of the Boolean AND and OR operator on non-Boolean spaces. The weighted minmax can thus be seen as a complex condition combining many AND and OR operators. This indicates why the explanation is complex and needs many arguments. An extension of this work could be to generate a simplified explanation for this model.

10. Proofs

10.1. Proof of Section 3

Proof of Lemma 1. In order to compute the maximum and minimum values of each component of the elements of the polytope $\mathcal{V}^{\mathrm{EU}}(y,x)$, it is enough to consider its vertices since $\mathcal{V}^{\mathrm{EU}}(y,x)$ is a polytope [16]. First of all, the following set

$$\left\{w \in \overline{\mathcal{W}}(\mathrm{EU})\colon\ \forall i \in A^-(y,x) \cup A^=(y,x)\ w_i = 0\right\}$$

is included in $\mathcal{V}^{\mathrm{EU}}(y,x)$. Hence, when $|A^+(y,x)| > 1$, the set admissible values of the weight w_i, with $i \in A^+(y,x)$, is the $[0,1]$ interval.

For $\varepsilon > 0$, let U_ε be the set of $w \in \mathbb{R}^n$ defined by $n+1$ inequalities $w_1 \geqslant 0, \ldots, w_n \geqslant 0$ and $\sum_{k\in N} \Delta_k w_k \geqslant \varepsilon$, and one equality $\sum_{k\in N} w_k = 1$. Then $\mathcal{V}^{\mathrm{EU}}(y,x) = \bigcup_{\varepsilon>0} U_\varepsilon$. A point $w \in U_\varepsilon$ is a vertex of U_ε if and only if $n-1$ inequalities are replaced by the corresponding equalities [16, Theorem 18.1]. Let $\{i,j\} \subseteq \{1,\ldots,n+1\}$ be the index of the two inequalities that are not transformed into equalities.

The first case is when $\{i,j\} \subseteq \{1,\ldots,n\}$. For ε small enough, one necessarily has $i \in A^+(y,x)$ and $j \in A^-(y,x)$. From the relations $\Delta_i^+ w_i - \Delta_j^- w_j = \varepsilon$ and $w_i + w_j = 1$, we obtain

$$w_i = \frac{\Delta_j^- + \varepsilon}{\Delta_i^+ + \Delta_j^-}, \qquad w_j = \frac{\Delta_i^+ - \varepsilon}{\Delta_i^+ + \Delta_j^-} \quad \text{and} \quad \forall k \in N \setminus \{i,j\} \quad w_k = 0.$$

The second case arises when $j = n+1$. We obtain

$$i \in A^+(y,x), \qquad w_i = 1 \quad \text{and} \quad \forall k \in N \setminus \{i\} \quad w_k = 0.$$

Assume that $A^+(y,x) \neq \emptyset$ and $A^-(y,x) \neq \emptyset$. Hence $w_i \in [\min_{j\in A^-(y,x)} \frac{\Delta_j^- + \varepsilon}{\Delta_i^+ + \Delta_j^-}, 1]$ for all $i \in A^+(y,x)$, and $w_j \in [0, \max_{i\in A^+(y,x)} \frac{\Delta_i^+ - \varepsilon}{\Delta_i^+ + \Delta_j^-}]$ for all $j \in A^-(y,x)$. The lemma is shown since this relations hold for all $\varepsilon > 0$. By construction, the boundaries of the intervals are reached so that the intervals are sharp. □

10.2. Proofs of Section 5

Proof of Lemma 2. The *only if* part of the lemma is clear. Let us show the *if* part. Consider thus $S \in Ex(x,y,v,\mathcal{F},\psi_{\mathrm{NOA}})$, and assume thus that (3) holds.

When $\mathcal{F} = \mathrm{EU}$, we show in the proof of Proposition 1 that if (3) holds, then we have $(v_k - \frac{1}{n})\Delta_k > 0$ for all $k \in S$. Let $T \subseteq S$ with $T \neq S$. For $k \in S \setminus T$, we have from (13)

$$H^{\mathrm{EU}}_{(v_T, w^{\mathrm{EU}}_{N\setminus T})}(y,x) = H^{\mathrm{EU}}_{(v_{S\setminus\{k\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{k\}})}(y,x) - \sum_{i\in S\setminus(T\cup\{k\})} \left(v_i - \frac{1}{n}\right)\Delta_i.$$

By (3), $H^{\mathrm{EU}}_{(v_{S\setminus\{k\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{k\}})}(y,x) \leqslant 0$. We conclude that $H^{\mathrm{EU}}_{(v_T, w^{\mathrm{EU}}_{N\setminus T})}(y,x) < 0$ and thus $T \notin Ex(x,y,v,\mathrm{EU},\psi_{\mathrm{NOA}})$. This proves that the coalitions satisfying (3) are necessary minimal in $Ex(x,y,v,\mathrm{EU},\psi_{\mathrm{NOA}})$.

The proof is similar when $\mathcal{F} = \mathrm{Maj}$, thanks to the relation (14) given below.

Lastly, consider the case when $\mathcal{F} = \mathrm{Pess}$. Let $T \subseteq S$ with $T \neq S$. From (3), we have $h^{\mathrm{Pess}}_{(v_{S\setminus\{k\}}, w^{\mathrm{Maj}}_{(N\setminus S)\cup\{k\}})}(x) \geqslant h^{\mathrm{Pess}}_{(v_{S\setminus\{k\}}, w^{\mathrm{Maj}}_{(N\setminus S)\cup\{k\}})}(y)$ for all $k \in S \setminus T$. Lemma 7 (shown below) holds with the strict inequalities are replaced by non-strict inequalities. We obtain thus

$$\bigwedge_{k\in S\setminus T} h^{\mathrm{Pess}}_{(v_{S\setminus\{k\}}, w^{\mathrm{Maj}}_{(N\setminus S)\cup\{k\}})}(x) \geqslant \bigwedge_{k\in S\setminus T} h^{\mathrm{Pess}}_{(v_{S\setminus\{k\}}, w^{\mathrm{Maj}}_{(N\setminus S)\cup\{k\}})}(y).$$

For every $z \in X$,

$$\begin{aligned}\bigwedge_{k\in S\setminus T} h^{\mathrm{Pess}}_{(v_{S\setminus\{k\}}, w^{\mathrm{Maj}}_{(N\setminus S)\cup\{k\}})}(z) &= \bigwedge_{i\in T}\big(z_i\vee(1-v_i)\big)\wedge\bigwedge_{i\notin S} z_i\wedge\bigwedge_{k\in S\setminus T}\big[\big(z_k\vee(1-v_k)\big)\wedge z_k\big]\\ &= \bigwedge_{i\in T}\big(z_i\vee(1-v_i)\big)\wedge\bigwedge_{i\notin S} z_i\wedge\bigwedge_{k\in S\setminus T} z_k\\ &= h^{\mathrm{Pess}}_{(v_T, w^{\mathrm{Maj}}_{N\setminus T})}(z).\end{aligned}$$

Hence we conclude that

$$h^{\mathrm{Pess}}_{(v_T, w^{\mathrm{Maj}}_{N\setminus T})}(x)\geqslant h^{\mathrm{Pess}}_{(v_T, w^{\mathrm{Maj}}_{N\setminus T})}(y).$$

This proves that $T\notin Ex(x, y, v, \mathrm{Pess}, \psi_{\mathrm{NOA}})$. □

Proof of Proposition 1. Let $S\in Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{NOA}})$ be minimal, and $k\in S$. Clearly, (3) is satisfied. Hence $H^{\mathrm{EU}}_{(v_S, w^{\mathrm{EU}}_{N\setminus S})}(y, x)>0$ and $H^{\mathrm{EU}}_{(v_{S\setminus\{k\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{k\}})}(y, x)\leqslant 0$, which gives $H^{\mathrm{EU}}_{(v_S, w^{\mathrm{EU}}_{N\setminus S})}(y, x)-H^{\mathrm{EU}}_{(v_{S\setminus\{k\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{k\}})}(y, x)>0$. From the relation

$$H^{\mathrm{EU}}_{(v_S, w^{\mathrm{EU}}_{N\setminus S})}(y, x)-H^{\mathrm{EU}}_{(v_{S\setminus\{k\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{k\}})}(y, x)=\left(v_k-\frac{1}{n}\right)\Delta_k \tag{13}$$

we can infer that $(v_k-\frac{1}{n})\Delta_k>0$. This concludes the proof. □

Proof of Proposition 2. The integer p as defined in Proposition 2 exists since $N\in Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{NOA}})$ and $\emptyset\notin Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{NOA}})$. We set $S=\{\pi(p), \ldots, \pi(n)\}$. Let $k\in S$. We write

$$\begin{aligned}H^{\mathrm{EU}}_{(v_{S\setminus\{k\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{k\}})}(y, x) &= H^{\mathrm{EU}}_{(v_S, w^{\mathrm{EU}}_{N\setminus S})}(y, x)-\left(v_k-\frac{1}{n}\right)\Delta_k \quad \text{by (13)}\\ &\leqslant H^{\mathrm{EU}}_{(v_S, w^{\mathrm{EU}}_{N\setminus S})}(y, x)-\left(v_{\pi(p)}-\frac{1}{n}\right)\Delta_{\pi(p)} \quad \text{since } \pi^{-1}(k)\geqslant p\\ &= H^{\mathrm{EU}}_{(v_{S\setminus\{\pi(p)\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{\pi(p)\}})}(y, x)\\ &\quad+\left(H^{\mathrm{EU}}_{(v_S, w^{\mathrm{EU}}_{N\setminus S})}(y, x)-H^{\mathrm{EU}}_{(v_{S\setminus\{\pi(p)\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{\pi(p)\}})}(y, x)-\left(v_{\pi(p)}-\frac{1}{n}\right)\Delta_{\pi(p)}\right)\\ &= H^{\mathrm{EU}}_{(v_{S\setminus\{\pi(p)\}}, w^{\mathrm{EU}}_{(N\setminus S)\cup\{\pi(p)\}})}(y, x) \quad \text{by (13)}\\ &\leqslant 0\end{aligned}$$

since $S\setminus\{\pi(p)\}\notin Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{NOA}})$. Hence S is minimal in $Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{NOA}})$.

By (13), we have for every $S\subseteq N$

$$H^{\mathrm{EU}}_{(v_S, w^{\mathrm{EU}}_{N\setminus S})}(y, x)=H^{\mathrm{EU}}_{w^{\mathrm{EU}}}(y, x)+\sum_{k\in S}\left(v_k-\frac{1}{n}\right)\Delta_k.$$

By definition of p, it is clear that S is the set with the smallest cardinality for which $\sum_{k\in S}(v_k-\frac{1}{n})\Delta_k>-H^{\mathrm{EU}}_{w^{\mathrm{EU}}}(y, x)$. Hence there is no minimal element of $Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{NOA}})$ with a strictly lower cardinality than S. □

Proof of Proposition 3. Let $S\in Ex(x, y, v, \mathrm{Maj}, \psi_{\mathrm{NOA}})$ be minimal and $k\in S$. We have

$$H^{\mathrm{Maj}}_{(v_S, w^{\mathrm{Maj}}_{N\setminus S})}(y, x)-H^{\mathrm{Maj}}_{(v_{S\setminus\{k\}}, w^{\mathrm{Maj}}_{(N\setminus S)\cup\{k\}})}(y, x)=\left(v_k-\frac{1}{n}\right)\mathrm{sgn}_k\,. \tag{14}$$

Hence the result is shown thanks to (3). □

Proof of Proposition 4. Thanks to (14), the proof is similar to that of Proposition 2. □

Proof of Proposition 5. Let $S\in Ex(x, y, v, \mathrm{Pess}, \psi_{\mathrm{NOA}})$ be minimal. Let $k\in S$. We set $a=\bigwedge_{i\in S\setminus\{k\}}(y_i\vee(1-v_i))\wedge\bigwedge_{i\notin S} y_i$ and $b=\bigwedge_{i\in S\setminus\{k\}}(x_i\vee(1-v_i))\wedge\bigwedge_{i\notin S} x_i$. We obtain by (3)

$$\begin{aligned}&a\wedge\big(y_k\vee(1-v_k)\big)>b\wedge\big(x_k\vee(1-v_k)\big),\\ &a\wedge y_k\leqslant b\wedge x_k.\end{aligned}$$

We wish to infer inequality relations among a, b, x_k, y_k, $1-v_k$ for the previous two relations. It is possible to do it by a standard reasoning on the inequalities. However, it is tedious, lengthy and not always easy. We propose rather to make a systematic check of all possible cases of comparison between the variables a, b, x_k, y_k, $1-v_k$. Since the previous two relations contain only the min and max operators, their analysis does not depend on the numerical values of a, b, x_k, y_k, $1-v_k$, but only on their relative ordering. Hence, this can be performed by a computer. As a result, we generate by computer all possible comparisons (lower, equal or greater) among the variables a, b, x_k, y_k, $1-v_k$, and see whether the previous two inequalities are satisfied. At the end, for every pair of variables, we analyse which comparisons between these two variables are compatible with the two inequalities. Considering for instance the two variables a and b, the outcome is either no relation between a and b, or one of the following comparisons: $a<b$, $a\leqslant b$, $a=b$, $a\geqslant b$ or $a>b$. This approach has been also used in the proof of Proposition 12. We obtain the following invariant comparisons

$$y_k \leqslant b < \{a, 1-v_k\} \quad \text{and} \quad y_k \leqslant x_k$$

where the notation $b<\{a,1-v_k\}$ means that $b<a$ and $b<1-v_k$.

Hence $k\in A^-(y,x)\cup A^=(y,x)$. Applying this relation for all $k\in S$, we obtain $S\subseteq A^-(y,x)\cup A^=(y,x)$. Since $y_k<1-v_k$, k is a weak negative argument.

We have two cases:

(i) S is of cardinality 1: $S=\{k\}$. Then

$$y_k\vee(1-v_k)\geqslant 1-v_k>b=\bigwedge_{i\notin S}x_i,$$
$$\bigwedge_{i\notin S}y_i=a>b=\bigwedge_{i\notin S}x_i.$$

(ii) S is of cardinality at least 2. Since $a>b$, one has

$$\bigwedge_{i\in S\setminus\{k\}}\big(y_i\vee(1-v_i)\big)\wedge\bigwedge_{i\notin S}y_i>\bigwedge_{i\in S\setminus\{k\}}\big(x_i\vee(1-v_i)\big)\wedge\bigwedge_{i\notin S}x_i.$$

Since $y_i\leqslant x_i$ for all $i\in S$, one has (see Lemma 7)

$$\bigwedge_{i\in S\setminus\{k\}}\big(y_i\vee(1-v_i)\big)\leqslant\bigwedge_{i\in S\setminus\{k\}}\big(x_i\vee(1-v_i)\big).$$

Hence for all $k\in S$

$$\bigwedge_{i\notin S}x_i<\bigwedge_{i\notin S}y_i \quad\text{and}\quad \bigwedge_{i\notin S}x_i<\bigwedge_{i\in S\setminus\{k\}}\big(y_i\vee(1-v_i)\big).$$

This gives

$$\bigwedge_{i\notin S}x_i<\bigwedge_{k\in S}\bigwedge_{i\in S\setminus\{k\}}\big(y_i\vee(1-v_i)\big)=\bigwedge_{i\in S}\big(y_i\vee(1-v_i)\big).$$

In both cases, we have

$$\bigwedge_{i\notin S}x_i<\bigwedge_{i\notin S}y_i \quad\text{and}\quad \bigwedge_{i\notin S}x_i<\bigwedge_{i\in S}\big(y_i\vee(1-v_i)\big). \qquad \square$$

Proof of Proposition 6. Assume that $h_v^{\text{Pess}}(x)$ is attained at $k^x\in N$. Since $h_v^{\text{Pess}}(y)>h_v^{\text{Pess}}(x)$, we have in particular $y_{k^x}\vee(1-v_{k^x})>x_{k^x}\vee(1-v_{k^x})$. We obtain a contradiction if $y_{k^x}\leqslant 1-v_{k^x}$. Hence $y_{k^x}>1-v_{k^x}$. Furthermore, $y_{k^x}>x_{k^x}$.

Moreover, for $k\in M$, we have $y_k\vee(1-v_k)>x_{k^x}\vee(1-v_{k^x})$ and thus $1-v_k>x_{k^x}\vee(1-v_{k^x})$.

Let $S\in Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$ be minimal, and assume that $S\setminus M\neq\emptyset$. Let $k\in S\setminus M$. Then $y_k>h_v^{\text{Pess}}(x)$. We have $h^{\text{Pess}}_{(v_{S\setminus\{k\}},w^{\text{Pess}}_{(N\setminus S)\cup\{k\}})}(y)=h^{\text{Pess}}_{(v_S,w^{\text{Pess}}_{N\setminus S})}(y)\wedge y_k$. Since $S\in Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$, we have $h^{\text{Pess}}_{(v_S,w^{\text{Pess}}_{N\setminus S})}(y)>h^{\text{Pess}}_{(v_S,w^{\text{Pess}}_{N\setminus S})}(x)$. Since S is minimal, by Proposition 5, $h^{\text{Pess}}_{(v_S,w^{\text{Pess}}_{N\setminus S})}(x)=\bigwedge_{i\in N\setminus S}x_i$, and $y_k>h_v^{\text{Pess}}(x)=h_v^{\text{Pess},N\setminus S}(x)\geqslant\bigwedge_{i\in N\setminus S}x_i$. Hence

$$h^{\text{Pess}}_{(v_{S\setminus\{k\}},w^{\text{Pess}}_{(N\setminus S)\cup\{k\}})}(y)>\bigwedge_{i\in N\setminus S}x_i.$$

Since $k\in A^-(y,x)\cup A^=(y,x)$, $x_k\geqslant y_k>\bigwedge_{i\in N\setminus S}x_i$, we have $\bigwedge_{i\in N\setminus S}x_i=\bigwedge_{i\in(N\setminus S)\cup\{k\}}x_i$. Therefore

$$h^{\text{Pess}}_{(v_{S\setminus\{k\}}, w^{\text{Pess}}_{(N\setminus S)\cup\{k\}})}(y) > \bigwedge_{i\in(N\setminus S)\cup\{k\}} x_i \geqslant h^{\text{Pess}}_{(v_{S\setminus\{k\}}, w^{\text{Pess}}_{(N\setminus S)\cup\{k\}})}(x).$$

Hence $S\setminus\{k\}\in Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$, which contradicts the fact that S is minimal. We conclude that $S\subseteq M$. □

Proof of Proposition 7. By definition $S := \{\pi^y_N(1),\dots,\pi^y_N(p)\}\in Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$ and $\{\pi^y_N(1),\dots,\pi^y_N(p-1)\}\notin Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$. To show that S is minimal, we need to show that for every $i\in\{1,\dots,p-2\}$, $D := S\setminus\{\pi^y_N(i)\}\notin Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$.

Let $i\in\{1,\dots,p-2\}$. Set $C=\{\pi^y_N(1),\dots,\pi^y_N(i-1)\}$, $C' = D\setminus C$. One has $C\notin Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$. Moreover, $\bigwedge_{j\in N\setminus C} y_j = y_{\pi^y_N(i)}$ and thus

$$h^{\text{Pess}}_{(v_C, w^{\text{Pess}}_{N\setminus C})}(y) = \bigwedge_{j\in C}\big(y_j\vee(1-v_j)\big)\wedge y_{\pi^y_N(i)}$$

and

$$h^{\text{Pess}}_{(v_D, w^{\text{Pess}}_{N\setminus D})}(y) = \bigwedge_{j\in C}\big(y_j\vee(1-v_j)\big)\wedge y_{\pi^y_N(i)}\wedge\bigwedge_{j\in C'}\big(y_j\vee(1-v_j)\big).$$

Hence

$$h^{\text{Pess}}_{(v_C, w^{\text{Pess}}_{N\setminus C})}(y)\geqslant h^{\text{Pess}}_{(v_D, w^{\text{Pess}}_{N\setminus D})}(y).$$

Furthermore for any option z, relation $C\subseteq D$ implies the following inequality

$$h^{\text{Pess}}_{(v_C, w^{\text{Pess}}_{N\setminus C})}(z)\leqslant h^{\text{Pess}}_{(v_D, w^{\text{Pess}}_{N\setminus D})}(z). \tag{15}$$

Applying that to $z=y$, we conclude that

$$h^{\text{Pess}}_{(v_C, w^{\text{Pess}}_{N\setminus C})}(y) = h^{\text{Pess}}_{(v_D, w^{\text{Pess}}_{N\setminus D})}(y).$$

Since $C\notin Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$, and by (15) applied to x, we get

$$\begin{aligned} h^{\text{Pess}}_{(v_D, w^{\text{Pess}}_{N\setminus D})}(x) &\geqslant h^{\text{Pess}}_{(v_C, w^{\text{Pess}}_{N\setminus C})}(x)\\ &\geqslant h^{\text{Pess}}_{(v_C, w^{\text{Pess}}_{N\setminus C})}(y)\\ &= h^{\text{Pess}}_{(v_D, w^{\text{Pess}}_{N\setminus D})}(y).\end{aligned}$$

Hence $D\notin Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$. This proves that $\{\pi^y_N(1),\dots,\pi^y_N(p)\}$ is minimal in $Ex(x,y,v,\text{Pess},\psi_{\text{NOA}})$. □

10.3. Proofs of Section 6

Proof of Proposition 8. Let $\mathcal{A}\in Ex(x,y,v,\mathcal{F},\psi_{\text{IVT}})$ be minimal in the sense of $\sqsubseteq$. Let $\pi\in\Pi(N)$ such that $\mathcal{A}\subseteq\mathcal{A}(\pi)$. Then $x\succeq^{\mathcal{F}}_{(\pi\circ v_{\mathcal{A}}, v_{N\setminus\mathcal{A}})} y$. Assume by contradiction that there exists $S\in\mathcal{A}$ with $|S|=1$. The condition $|S|=1$ implies that $\pi(i)=i$ for $i\in S$. Hence $x\succeq^{\mathcal{F}}_{(\pi\circ v_{\overline{\mathcal{A}\setminus\{S\}}}, v_{(N\setminus\overline{\mathcal{A}\setminus\{S\}})}} y$. This contradicts the minimality of $\mathcal{A}$ since $\mathcal{A}\setminus\{S\}\sqsubset\mathcal{A}$ and $\mathcal{A}\setminus\{S\}\subseteq\mathcal{A}(\pi)$. □

Proof of Lemma 3. The proof is done by induction on $s=|S|$. Let $H^S_w := \sum_{i\in S} w_i\Delta_i$ for $w\in\mathcal{W}(\text{EU})$.

When $s=2$, we have

$$\begin{aligned} H^S_{\overline{\pi}_S\circ v} - H^S_{\underline{\pi}_S\circ v} &= (v_{\pi^v_S(1)}\Delta_{\pi^\Delta_S(1)} + v_{\pi^v_S(2)}\Delta_{\pi^\Delta_S(2)}) - (v_{\pi^v_S(2)}\Delta_{\pi^\Delta_S(1)} + v_{\pi^v_S(1)}\Delta_{\pi^\Delta_S(2)})\\ &= (v_{\pi^v_S(2)} - v_{\pi^v_S(1)})\times(\Delta_{\pi^\Delta_S(2)} - \Delta_{\pi^\Delta_S(1)})\geqslant 0.\end{aligned}$$

Hence the result is proved when $s=2$.

Assume that the result is shown for all subsets of cardinality strictly lower than s. Let $S\subseteq N$ with $|S|=s$.

Let $\pi\in\Pi(S)$. Define $\sigma=\pi\circ(\overline{\pi}_S)^{-1}$, $k=\pi^\Delta_S(s)$ and $k'=\pi^v_S(s)=\overline{\pi}_S(k)$. Value Δ_k is the largest number in the set $\{\Delta_i\colon i\in S\}$, and $v_{k'}$ is the largest number in the set $\{v_i\colon i\in S\}$. Define $\sigma'\in\Pi(S)$ by $\sigma'(k')=k'$, $\sigma'(\sigma^{-1}(k'))=\sigma(k')$, and $\sigma'(i)=\sigma(i)$ for all $i\in S\setminus\{k',\sigma^{-1}(k')\}$ (see Fig. 2). Set $\pi'=\sigma'\circ\overline{\pi}_S\in\Pi(S)$.

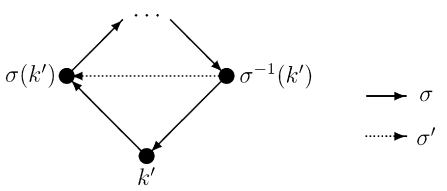

Fig. 2. Description of σ and σ'.

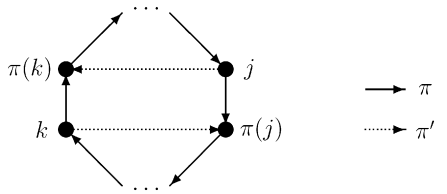

Fig. 3. Description of π and π'.

One has

$$H^S_{\pi\circ v} = \sum_{i\in S} v_{\pi(i)}\Delta_i = \sum_{i\in S} v_{\sigma\circ\overline{\pi}_S(i)}\Delta_i = \sum_{i\in\{1,\ldots,s\}} v_{\sigma\circ\pi^v_S(i)}\Delta_{\pi^\Delta_S(i)}.$$

Let $i_1 = (\pi^v_S)^{-1}(k') = s$ and $i_2 = (\pi^v_S)^{-1}(\sigma^{-1}(k'))$. Hence

$$\begin{aligned} H^S_{\pi'\circ v} - H^S_{\pi\circ v} &= v_{\sigma'\circ\pi^v_S(i_1)}\Delta_{\pi^\Delta_S(i_1)} + v_{\sigma'\circ\pi^v_S(i_2)}\Delta_{\pi^\Delta_S(i_2)} - v_{\sigma\circ\pi^v_S(i_1)}\Delta_{\pi^\Delta_S(i_1)} - v_{\sigma\circ\pi^v_S(i_2)}\Delta_{\pi^\Delta_S(i_2)} \\ &= (v_{k'} - v_{\sigma(k')}) \times (\Delta_k - \Delta_{\pi^\Delta_S(i_2)}) \geqslant 0 \end{aligned}$$

by definition of k and k'.

Moreover, $H^S_{\overline{\pi}_S\circ v}$ and $H^S_{\pi'\circ v}$ have the same value $v_{k'}\Delta_k$ on criterion k. Hence $H^S_{\overline{\pi}_S\circ v} - H^S_{\pi'\circ v} = H^{S\setminus\{k\}}_{\overline{\pi}_S\circ v} - H^{S\setminus\{k\}}_{\pi'\circ v}$. By the induction assumption, $H^{S\setminus\{k\}}_{\overline{\pi}_S\circ v} \geqslant H^{S\setminus\{k\}}_{\pi'\circ v}$. Hence we have shown that

$$H^S_{\overline{\pi}_S\circ v} \geqslant H^S_{\pi'\circ v} \geqslant H^S_{\pi\circ v}.$$

Similarly, one can show that $H^S_{\underline{\pi}_S\circ v} \leqslant H^S_{\pi\circ v}$. Hence the induction assumption holds at s. □

Proof of Proposition 9. Let $\mathcal{A} \in Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{IVT}})$ be minimal in the sense of $\sqsubseteq$. Let $\pi \in \Pi(N)$ such that $x \succeq^{\mathrm{EU}}_{(\pi\circ v_{\overline{\mathcal{A}}}, v_{N\setminus\overline{\mathcal{A}}})} y$ and $\mathcal{A} \subseteq \mathcal{A}(\pi)$. Since $x \succeq^{\mathrm{EU}}_{(\pi\circ v_{\overline{\mathcal{A}}}, v_{N\setminus\overline{\mathcal{A}}})} y$, one can choose π in such a way that for all $i \in N \setminus \overline{\mathcal{A}}$, $\pi(i) = i$. Hence $(\pi \circ v_{\overline{\mathcal{A}}}, v_{N\setminus\overline{\mathcal{A}}}) = \pi \circ v$.

Let $S \in \mathcal{A}$. By Proposition 8, we have $|S| \geqslant 2$.

Let $k, j \in S$ with $k \neq j$. Let $\pi' \in \Pi(N)$ be defined by $\pi'(j) = \pi(k)$, $\pi'(k) = \pi(j)$ and $\pi'(l) = \pi(l)$ for $l \in N \setminus \{k, j\}$ (see Fig. 3).

Clearly, $\mathcal{A}(\pi') \sqsubset \mathcal{A}(\pi)$. Let $\mathcal{A}' = (\mathcal{A} \setminus \{S\}) \cup \{S', S''\}$ where $S' = \{j, \pi'(j), \pi' \circ \pi'(j), \ldots\}$ and $S'' = \{k, \pi'(k), \pi' \circ \pi'(k), \ldots\}$ are the cycles of π' containing j and k respectively. One clearly has $\mathcal{A}' \sqsubset \mathcal{A}$ and $\mathcal{A}' \subseteq \mathcal{A}(\pi')$. Hence in order that $\mathcal{A}$ is a minimal element, one shall have $\mathcal{A}' \notin Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{IVT}})$. Hence, we have either $y \succ^{\mathrm{EU}}_{\pi'\circ v} x$ or $y \succ^{\mathrm{EU}}_{(\pi'\circ v_{\overline{\mathcal{A}'}}, v_{N\setminus\overline{\mathcal{A}'}})} x$. Since $\overline{\mathcal{A}} = \overline{\mathcal{A}'}$ and $\pi'(l) = l$ for all $l \in N \setminus \overline{\mathcal{A}} = N \setminus \overline{\mathcal{A}'}$, we have $\pi' \circ v = (\pi' \circ v_{\overline{\mathcal{A}'}}, v_{N\setminus\overline{\mathcal{A}'}})$. Hence in both case, we have $y \succ^{\mathrm{EU}}_{\pi'\circ v} x$. To sum-up, we have

$$x \succeq^{\mathrm{EU}}_{\pi\circ v} y \quad \text{and} \quad y \succ^{\mathrm{EU}}_{\pi'\circ v} x. \tag{16}$$

We obtain from (16)

$$0 < H^{\mathrm{EU}}_{\pi'\circ v}(y, x) - H^{\mathrm{EU}}_{\pi\circ v}(y, x) = (v_{\pi(j)} - v_{\pi(k)}) \times (\Delta_k - \Delta_j).$$

Therefore (4) holds.

From (4), $\Delta_{\pi^\Delta_S(l)} \neq \Delta_{\pi^\Delta_S(l+1)}$, for all $l \in \{1, \ldots, |S| - 1\}$. Hence $\Delta_{\pi^\Delta_S(l)} < \Delta_{\pi^\Delta_S(l+1)}$. By (4), the previous relation implies that $v_{\pi\circ\pi^\Delta_S(l)} > v_{\pi\circ\pi^\Delta_S(l+1)}$, for all $l \in \{1, \ldots, |S| - 1\}$. Hence (5) holds. We have $v_{\pi^v_S(1)} \leqslant \cdots \leqslant v_{\pi^v_S(|S|)}$. We conclude that $\pi(j) = \underline{\pi}_S(j)$ for all $j \in S$.

Finally, the last inequality in the proposition follows from relation (5) and Lemma 3. □

Proof of Proposition 10. Assume by contradiction that the algorithm ends without returning an explanation set. Hence $\mathcal{B} = \emptyset$ at every step of the search. This implies that the search tree is never pruned (step "*Bounding*") and thus all the subsets of elements of $\mathcal{S}$ are explored in the search. This means that there does not exist $\mathcal{A} \in Ex$ composed of elements of $\mathcal{S}$ such that $\sum_{S\in\mathcal{A}} D_S^{\mathrm{EU}} \geqslant H_v^{\mathrm{EU}}(y,x)$. Hence there does not exist $\pi \in \Pi(N)$ such that $H_{\pi\circ v}^{\mathrm{EU}}(y,x) \leqslant 0$. This contradicts the basic assumption made at the beginning of Section 6.

When the bounding condition is triggered in **Algo-**EU, we have $\mathcal{A}\cup\{T_i\} \not\sqsupset_{\mathrm{discri}} \mathcal{B}$. Since $\sqsubseteq_{\mathrm{discri}}$ is a complete order, we conclude that $\mathcal{B} \sqsubseteq_{\mathrm{discri}} \mathcal{A}\cup\{T_i\}$. In the combinatorial structure Ex endowed with $\sqsubseteq_{\mathrm{discri}}$, the whole sub-tree under the node $\mathcal{A}\cup\{T_i\}$ in the search (in $Algo(\mathcal{A},\mathcal{B},k)$ after iteration i) can indeed be pruned since the elements of Ex in the sub-tree are of the form $\mathcal{A}\cup\{T_{i_1}\}\cup\cdots\cup\{T_{i_m}\}$, with $m \geqslant 1$ and $T_i \trianglelefteq_{\mathrm{discri}} T_{i_1}$. By the relation

$$\mathcal{B} \sqsubseteq_{\mathrm{discri}} \mathcal{A}\cup\{T_i\} \sqsubset_{\mathrm{discri}} \mathcal{A}\cup\{T_i\}\cup\{T_{i_1}\}\cup\cdots\cup\{T_{i_m}\}$$

one sees that there is no way a strictly better argumentation set in Ex can be found under the node $\mathcal{A}\cup\{T_i\}$. We have shown that the bounding condition is justified in the algorithm.

Assume that the algorithm terminates and returns $\mathcal{B}$. We define $\pi \in \Pi(N)$ as $\pi(i) = \underline{\pi}_S(i)$ for all $i \in S$ and all $S \in \mathcal{B}$, and $\pi(i) = i$ for all $i \in N \setminus \overline{\mathcal{B}}$. We have

$$H_{\pi\circ v}^{\mathrm{EU}}(y,x) = H_v^{\mathrm{EU}}(y,x) - \sum_{S\in\mathcal{B}} D_S^{\mathrm{EU}} \leqslant 0.$$

Hence $\mathcal{B} \in Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{IVT}})$. All the elements of Ex that are strictly better than $\mathcal{B}$ are explored by Algorithm **Algo-**EU at the previous steps. Since the algorithm did not terminate earlier, this means that there is no $\mathcal{C} \in Ex(x, y, v, \mathrm{EU}, \psi_{\mathrm{IVT}})$ with $\mathcal{C} \sqsubset_{\mathrm{discri}} \mathcal{B}$. Hence $\mathcal{B}$ is minimal in the sense of $\sqsubset_{\mathrm{discri}}$. The explanation set $\mathcal{B}$ is also minimal in the sense of $\sqsubset$ since $\sqsubset_{\mathrm{discri}}$ is a refinement of $\sqsubset$. □

Proof of Proposition 11. The proof follows that of Proposition 9. In particular, from $k, j \in S$ with $k \neq j$, we define π'. Hence

$$0 < H_{\pi'\circ v}^{\mathrm{Maj}}(y,x) - H_{\pi\circ v}^{\mathrm{Maj}}(y,x) = (v_{\pi(j)} - v_{\pi(k)}) \times (\mathrm{sgn}_k - \mathrm{sgn}_j).$$

One necessarily has $\mathrm{sgn}_k \neq \mathrm{sgn}_j$. Hence k and j cannot belong to the same set $A^+(y,x)$, $A^=(y,x)$ or $A^-(y,x)$. There is at most one element of S in each of the three sets $A^+(y,x)$, $A^=(y,x)$ or $A^-(y,x)$. Hence $|S| \leqslant 3$. The relations between $v_{\pi(k)}$ and $v_{\pi(j)}$ in the statement of the proposition follow from the previous inequality. □

Proof of Proposition 12. We proceed as in the proof of Proposition 9. In particular, from $k, j \in S$ with $k \neq j$, we define π'. Let

$$a := h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(y) \quad \text{and} \quad b := h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(x).$$

Proceeding as in the proof of Proposition 9, we obtain the relations $y \succ_{\pi'\circ v}^{\mathrm{Pess}} x$ and $y \preceq_{\pi\circ v}^{\mathrm{Pess}} x$ (see (16)). Hence

$$\begin{aligned}
\alpha_1 &:= a \wedge \big(y_j \vee (1 - v_{\pi(k)})\big) \wedge \big(y_k \vee (1 - v_{\pi(j)})\big) \\
&> b \wedge \big(x_j \vee (1 - v_{\pi(k)})\big) \wedge \big(x_k \vee (1 - v_{\pi(j)})\big) =: \beta_1, \\
\alpha_2 &:= a \wedge \big(y_j \vee (1 - v_{\pi(j)})\big) \wedge \big(y_k \vee (1 - v_{\pi(k)})\big) \\
&\leqslant b \wedge \big(x_j \vee (1 - v_{\pi(j)})\big) \wedge \big(x_k \vee (1 - v_{\pi(k)})\big) =: \beta_2.
\end{aligned}$$

We are interested only to the case where $k \in A^+(y,x) \cup A^=(y,x)$ and $j \in A^-(y,x)$. We have two cases.

The first case is when $1 - v_{\pi(k)} > 1 - v_{\pi(j)}$. The larger weight among $v_{\pi(k)}$ and $v_{\pi(j)}$ is assigned to the negative argument, and the smaller one is assigned to the positive argument.

The second case is when $1 - v_{\pi(k)} \leqslant 1 - v_{\pi(j)}$. We proceed exactly as in the proof of Proposition 5. We perform a systematic check of all possible cases of comparison between the variables $\{a, b, x_k, x_j, y_k, y_j, 1 - v_{\pi(j)}, 1 - v_{\pi(k)}\}$ to determine whether some comparisons between some of these variables always hold. We obtain

$$\begin{aligned}
&x_k \leqslant y_k \leqslant h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(x), && h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(x) < y_j < x_j, \\
&1 - v_{\pi(k)} \leqslant h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(x), && h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(x) < h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(y), \\
& && h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(x) < 1 - v_{\pi(j)}.
\end{aligned} \tag{17}$$

Let K_S and J_S be the sets of the indices k and j respectively, satisfying (17). Let us show that $\sum_{S\in\mathcal{A}} |K_S| \leqslant 1$. Assume by contradiction that there exist two pairs (k,j) and (k',j') in $(A^+(y,x) \cup A^=(y,x)) \times A^-(y,x)$ with $(k,j) \in K_S \times J_S$ and $(k',j') \in K_{S'} \times J_{S'}$ (with $S, S' \in \mathcal{A}$) satisfying (17), with $k \neq k'$. From (17), $y_k \vee (1 - v_{\pi(k)}) < h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(y)$. Since $k' \in N \setminus \{j,k\}$, we obtain $h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j,k\}}(y) \leqslant y_{k'} \vee (1 - v_{\pi(k')})$ and thus $y_k \vee (1 - v_{\pi(k)}) < y_{k'} \vee (1 - v_{\pi(k')})$. Similarly, we have $y_{k'} \vee (1 - v_{\pi(k')}) < h_{\pi\circ v}^{\mathrm{Pess},N\setminus\{j',k'\}}(y) \leqslant y_k \vee (1 - v_{\pi(k)})$. Hence a contradiction is raised.

We assume that $K_S \neq \emptyset$. We set $K_S = \{k\}$. Assume that $S = N$. Assume by contradiction that $J_S = N \setminus \{k\}$. Let $j \in J_S$ such that $h^{\text{Pess},J_S}_{\pi\circ v}(x) = x_j \vee (1 - v_{\pi(j)})$. Let $i \in J_S \setminus \{j\}$ such that $h^{\text{Pess},J_S\setminus\{j\}}_{\pi\circ v}(x) = x_i \vee (1 - v_{\pi(i)})$. From (17) applied on the pair (k, j), we have $x_i \vee (1 - v_{\pi(i)}) = h^{\text{Pess},N\setminus\{k,j\}}_{\pi\circ v}(x) < 1 - v_{\pi(j)}$ and thus

$$1 - v_{\pi(i)} < 1 - v_{\pi(j)}.$$

From (17) applied on the pair (k, i), we have $x_j \vee (1 - v_{\pi(j)}) = h^{\text{Pess},N\setminus\{k,i\}}_{\pi\circ v}(x) < 1 - v_{\pi(i)}$ and thus

$$1 - v_{\pi(j)} < 1 - v_{\pi(i)}.$$

The previous two relations are clearly contradictory. Hence $J_S \neq N \setminus \{k\}$.

Using previous argument, one can easily show in all cases that $h^{\text{Pess},N\setminus\{k\}}_{\pi\circ v}(x) = x_i \vee (1 - v_{\pi(i)})$ for some $i \in N \setminus (J_S \cup \{k\})$. Hence

$$h^{\text{Pess},N\setminus\{k\}}_{\pi\circ v}(x) = h^{\text{Pess},N\setminus(J_S\cup\{k\})}_{\pi\circ v}(x).$$

Let $S \in \mathcal{A}$. To sum-up, we have $v_{\pi(j)} > v_{\pi(i)}$ if $i \in (A^+(y,x) \cup A^=(y,x)) \setminus K_S$ and $j \in A^-(y,x)$, or if $i \in K_S$ and $j \in A^-(y,x) \setminus J_S$. Moreover, $v_{\pi(j)} \leqslant v_{\pi(i)}$ if $i \in K_S$ and $j \in J_S$. The weights assigned to the positive arguments are the smallest ones, except for criterion K_S:

$$\begin{aligned}\left\{v_{\pi(i)}\colon i \in \left(A^-(y,x) \cap S\right) \setminus J_S\right\} &> \{v_{\pi(k)}\colon k \in K_S\} \geqslant \{v_{\pi(j)}\colon j \in J_S\} \\ &> \left\{v_{\pi(i)}\colon i \in \left(\left(A^+(y,x) \cup A^=(y,x)\right) \cap S\right) \setminus K_S\right\}. \qquad \square\end{aligned}$$

10.4. Proofs of Section 7

Proof of Lemma 5. Clearly, for every $z \in [0,1]^n$

$$\frac{1}{n!}\sum_{\pi\in\Pi(N)} h^{\text{EU}}_{\pi\circ v}(z) = \sum_{i=1}^n \left(\frac{1}{n}\sum_{k\in N} v_k\right) z_i = \sum_{i=1}^n \frac{1}{n} z_i = h^{\text{EU}}_{w^{\text{EU}}}(z). \qquad \square$$

Proof of Proposition 13. By definition of $\underline{\pi}_N$, for all permutations $\pi \in \Pi(N)$,

$$H^{\text{EU}}_{\underline{\pi}_N\circ v}(y,x) \leqslant H^{\text{EU}}_{\pi\circ v}(y,x). \tag{18}$$

Hence from Lemma 5, $H^{\text{EU}}_{w^{\text{EU}}}(y,x) \geqslant \frac{1}{n!}\sum_{\pi\in\Pi(N)} H^{\text{EU}}_{\underline{\pi}_N\circ v}(y,x) = H^{\text{EU}}_{\underline{\pi}_N\circ v}(y,x)$. Consequently, the first part of the lemma is proved.

The *if* part of the second part of the lemma is obvious. Let us show the *only if* part. Assume thus that $H^{\text{EU}}_{\underline{\pi}_N\circ v}(y,x) = H^{\text{EU}}_{w^{\text{EU}}}(y,x)$. From (18) and Lemma 5, the previous relation implies for every permutation π, $H^{\text{EU}}_{\underline{\pi}_N\circ v}(y,x) = H^{\text{EU}}_{\pi\circ v}(y,x)$. Thus for every permutation $\pi \in \Pi(N)$,

$$H^{\text{EU}}_v(y,x) = H^{\text{EU}}_{\pi\circ v}(y,x).$$

Let $i \neq j$ in N, and $\pi \in \Pi(N)$ the permutation permuting i and j and leaving the other elements of N. Previous relation applied on π gives $v_i\Delta_i + v_j\Delta_j = v_j\Delta_i + v_i\Delta_j$. Hence

$$(v_i - v_j) \times (\Delta_i - \Delta_j) = 0. \tag{19}$$

Assume that (7) holds. Let $k \in N$, and $A_k = \{i \in N,\ \Delta_i = \Delta_k\}$. Thus $N \setminus A_k \neq \emptyset$. For all $i \in A_k$ and $j \in N \setminus A_k$, one has $\Delta_i \neq \Delta_j$ and thus $v_i = v_j$ by (19). We conclude that all v_i have the same value. Since the weights are normalized, we obtained the wished result. $\square$

Proof of Proposition 14. Let

$$\mathcal{H}^{\text{EU}}_1 := \inf_{x\in\mathbb{R}^n,\, y\in\mathbb{R}^n,\, v\in\mathbb{R}^n_+\colon v_1+\cdots+v_n=1} \frac{H^{\text{EU}}_{w^{\text{EU}}}(y,x) - \min_{\pi\in\Pi(N)} H^{\text{EU}}_{\pi\circ v}(y,x)}{\max_{\pi\in\Pi(N)} H^{\text{EU}}_{\pi\circ v}(y,x) - \min_{\pi\in\Pi(N)} H^{\text{EU}}_{\pi\circ v}(y,x)}.$$

Since $H^{\text{EU}}_v(y,x) = h^{\text{EU}}_v(\Delta)$, we obtain

$$\mathcal{H}^{\text{EU}}_1 = \inf_{\Delta\in\mathbb{R}^n,\, v\in\mathbb{R}^n_+\colon v_1+\cdots+v_n=1} \frac{h^{\text{EU}}_{w^{\text{EU}}}(\Delta) - \min_{\pi\in\Pi(N)} h^{\text{EU}}_{\pi\circ v}(\Delta)}{\max_{\pi\in\Pi(N)} h^{\text{EU}}_{\pi\circ v}(\Delta) - \min_{\pi\in\Pi(N)} h^{\text{EU}}_{\pi\circ v}(\Delta)}.$$

Let $\Delta \in \mathbb{R}^n$, and

$$\alpha = \frac{1}{\max_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) - \min_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta)},$$
$$\beta = -\alpha \min_{i\in N} \Delta_i.$$

Let $\Delta' := \alpha\Delta + \beta$. By definition of β, $\Delta' \in \mathbb{R}^n_+$. Since h^{EU}_w is stable under affine transformations (i.e. $h^{\mathrm{EU}}_w(\alpha\Delta+\beta) = \alpha h^{\mathrm{EU}}_w(\Delta) + \beta$) [31], $\max_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta') - \min_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta') = 1$. Hence

$$\mathcal{H}^{\mathrm{EU}}_1 = \inf_{\Delta\in\mathbb{R}^n_+,\, v\in\mathbb{R}^n_+ :\, v_1+\cdots+v_n=1,\, \max_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) - \min_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta)=1} \left(h^{\mathrm{EU}}_{w^{\mathrm{EU}}}(\Delta) - \min_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) \right).$$

Without loss of generality, one can assume that

$$v_1 \leqslant \cdots \leqslant v_n,$$
$$\Delta_1 \leqslant \cdots \leqslant \Delta_n.$$

Then by Lemma 3

$$\min_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) = \sum_{i=1}^n v_{n-i+1}\Delta_i,$$
$$\max_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) = \sum_{i=1}^n v_i\Delta_i$$

and

$$\mathcal{H}^{\mathrm{EU}}_1 = \inf_{\Delta\in\mathbb{R}^n_+,\, v\in\mathbb{R}^n_+ :\, v_1\leqslant\cdots\leqslant v_n,\, \Delta_1\leqslant\cdots\leqslant\Delta_n,\, v_1+\cdots+v_n=1,\, \sum_{i=1}^n (v_i - v_{n-i+1})\Delta_i = 1} \sum_{i=1}^n \left(\frac{1}{n} - v_{n-i+1} \right)\Delta_i.$$

From Assertions 1 and 2 below, $\mathcal{H}^{\mathrm{EU}}_1 = \frac{1}{n}$.

Assertion 1. *For $v \in \overline{\mathcal{W}}(\mathrm{EU}) \setminus \{w^{\mathrm{EU}}\}$ such that $v_1 \leqslant \cdots \leqslant v_n$,*

$$\inf_{\Delta\in\mathbb{R}^n_+ :\, \Delta_1\leqslant\cdots\leqslant\Delta_n,\, \sum_{i=1}^n (v_i - v_{n-i+1})\Delta_i = 1} \sum_{i=1}^n \left(\frac{1}{n} - v_{n-i+1} \right)\Delta_i = \min_{k\in\{2,\ldots,n\}} \frac{\sum_{i=k}^n (\frac{1}{n} - v_{n-i+1})}{\sum_{i=k}^n (v_i - v_{n-i+1})}$$

where the numerator and the denominator are positive.

Proof. Let

$$U = \left\{ \Delta \in \mathbb{R}^n,\ 0 \leqslant \Delta_1,\ \Delta_1 \leqslant \Delta_2,\ \ldots,\ \Delta_{n-1} \leqslant \Delta_n \text{ and } \sum_{i=1}^n (v_i - v_{n-i+1})\Delta_i = 1 \right\}.$$

From [16, Theorem 18.1], Δ is a vertex of U iff $n-1$ inequalities among the n inequalities of U are transformed into equalities. Let $k \in \{1,\ldots,n\}$ be the index of the inequality that is not transformed into an equality. One cannot have $k = 1$ since $\sum_{i=1}^n (v_i - v_{n-i+1}) = 0$. Hence $k \in \{2,\ldots,n\}$. Then $0 = \Delta_1 = \cdots = \Delta_{k-1} < \Delta_k = \cdots = \Delta_n =: \alpha_k$. One has

$$1 = \sum_{i=1}^n (v_i - v_{n-i+1})\Delta_i = \alpha_k \sum_{i=k}^n (v_i - v_{n-i+1}).$$

From the normalization condition, we obtain

$$\alpha_k = \frac{1}{\sum_{i=k}^n (v_i - v_{n-i+1})}.$$

Hence the relation of the assertion is proved.

We have

$$h^{\mathrm{EU}}_{w^{\mathrm{EU}}}(\Delta) - \min_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) = \sum_{i=k}^n \left(\frac{1}{n} - v_{n-i+1} \right),$$
$$\max_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) - \min_{\pi\in\Pi(N)} h^{\mathrm{EU}}_{\pi\circ v}(\Delta) = \sum_{i=k}^n (v_i - v_{n-i+1})$$

where Δ is defined above. The signs of numerator and denominator follow from Proposition 13. □

Assertion 2.

$$\inf_{v\in\overline{\mathcal{W}}(\mathrm{EU})\setminus\{w^{\mathrm{EU}}\}:\ v_1\leqslant\cdots\leqslant v_n}\ \min_{k\in\{2,\ldots,n\}}\frac{\sum_{i=k}^{n}(\frac{1}{n}-v_{n-i+1})}{\sum_{i=k}^{n}(v_i-v_{n-i+1})}=\frac{1}{n}.$$

Proof. Let $\mathcal{H}_2^{\mathrm{EU}}$ be the left hand side in the expression of the lemma. Then

$$\mathcal{H}_2^{\mathrm{EU}}=\inf_{v\in\mathbb{R}_+^n:\ v_1\leqslant\cdots\leqslant v_n,\ v_1+\cdots+v_n=1}\ \min_{k\in\{2,\ldots,n\}}\frac{\sum_{i=k}^{n}(\frac{1}{n}\sum_{j\in N}v_j-v_{n-i+1})}{\sum_{i=k}^{n}(v_i-v_{n-i+1})}.$$

We notice that the ratio is homogeneous in v in the previous relation. Hence constraint $v_1+\cdots+v_n=1$ can be removed:

$$\mathcal{H}_2^{\mathrm{EU}}=\inf_{v\in\mathbb{R}_+^n:\ v_1\leqslant\cdots\leqslant v_n}\ \min_{k\in\{2,\ldots,n\}}\frac{\sum_{i=k}^{n}(\frac{1}{n}\sum_{j\in N}v_j-v_{n-i+1})}{\sum_{i=k}^{n}(v_i-v_{n-i+1})}.$$

We are interested in v different from the arithmetic mean. From Lemma 1 the denominator is strictly positive in this case. Hence one can arbitrarily set it to value 1:

$$\mathcal{H}_2^{\mathrm{EU}}=\inf_{v\in\mathbb{R}_+^n:\ v_1\leqslant\cdots\leqslant v_n,\ \sum_{i=k}^{n}(v_i-v_{n-i+1})=1}\ \min_{k\in\{2,\ldots,n\}}\sum_{i=k}^{n}\left(\frac{1}{n}\sum_{j\in N}v_j-v_{n-i+1}\right).$$

One has

$$\sum_{i=k}^{n}(v_i-v_{n-i+1})=\sum_{i=k'}^{n}(v_i-v_{n-i+1})$$

with $k'=\max(k,n-k+1)$.

Let

$$U=\left\{v\in\mathbb{R}^n,\ 0\leqslant v_1,\ v_1\leqslant v_2,\ \ldots,\ v_{n-1}\leqslant v_n \text{ and } \sum_{i=k'}^{n}(v_i-v_{n-i+1})=1\right\}.$$

From [16, Theorem 18.1], v is a vertex of U iff $n-1$ inequalities of U are transformed into equalities. Let $p\in\{1,\ldots,n\}$ be the index of the inequality that is not transformed into an equality. One cannot have $p=1$ since, otherwise, the equality constraint in U would not be satisfied. Hence $p\in\{2,\ldots,n\}$. Then $0=v_1=\cdots=v_{p-1}<v_p=\cdots=v_n=:\alpha_p$. One has

$$\sum_{i=k}^{n}\left(\frac{1}{n}\sum_{j\in N}v_j-v_{n-i+1}\right)=\sum_{i=1}^{n-k+1}\left(\frac{n-p+1}{n}\alpha_p-v_i\right)=:F.$$

There are two cases:

- $p\geqslant n-k+1$. Hence the functional F is

  $$F=\frac{(n-p+1)(n-k+1)}{n}\alpha_p$$

  where

  $$1=\sum_{i=k'}^{n}(v_i-v_{n-i+1})=\sum_{i=k'}^{n}v_i=\big(n-\max\big(p,k'\big)+1\big)\times\alpha_p.$$

  Hence F is always greater or equal to $\frac{1}{n}$. The minimal value $\frac{1}{n}$ is attained for $p=k=n$.
- $p<n-k+1$. The functional F is

  $$F=\frac{(n-p+1)(n-k+1)}{n}\alpha_p-\big[(n-k+1)-p+1\big]\alpha_p$$

  with

  $$1=\sum_{i=k'}^{n}(v_i-v_{n-i+1})=\big(n-k'+1\big)\alpha_p-\big[\big(n-k'+1\big)-p\big]\alpha_p=p\alpha_p.$$

  Hence the functional is

$$F = \frac{p}{n}\big[(n-p+1)(n-k+1) - n(n-k-p+2)\big] = \frac{p}{n}\big[(p-1)k+1\big] \geqslant \frac{1}{n}.$$

The minimal value is $\frac{1}{n}$. □

The proof of Proposition 14 is now completed. □

Proof of Proposition 15. Let $\lambda^{\mathrm{EU}} := H^{\mathrm{EU}}_{\overline{\pi}_N \circ v}(y,x) - H^{\mathrm{EU}}_{\underline{\pi}_N \circ v}(y,x)$. For every $\pi, \pi' \in \Pi(N)$

$$\big|H^{\mathrm{EU}}_{\pi\circ v}(y,x) - H^{\mathrm{EU}}_{\pi'\circ v}(y,x)\big| \leqslant \lambda^{\mathrm{EU}}. \tag{20}$$

Let $i_\star, i^\star, j_\star, j^\star \in N$ such that $\min_{i\in N}\Delta_i = \Delta_{i_\star}$, $\max_{i\in N}\Delta_i = \Delta_{i^\star}$, $\min_{j\in N} v_j = v_{j_\star}$ and $\max_{j\in N} v_j = v_{j^\star}$. Define $\pi, \pi' \in \Pi(N)$ such that $\pi(i_\star) = j_\star$, $\pi(i^\star) = j^\star$, $\pi'(i_\star) = j^\star$, $\pi'(i^\star) = j_\star$ and $\pi(i) = \pi'(i)$ for all $i \in N \setminus \{i_\star, i^\star\}$. From (20), we obtain

$$\big|(v_{j^\star}\Delta_{i^\star} + v_{j_\star}\Delta_{i_\star}) - (v_{j_\star}\Delta_{i^\star} + v_{j^\star}\Delta_{i_\star})\big| \leqslant \lambda^{\mathrm{EU}}.$$

Hence

$$|v_{j^\star} - v_{j_\star}||\Delta_{i^\star} - \Delta_{i_\star}| \leqslant \lambda^{\mathrm{EU}}.$$

By the definition of $i_\star$ and $i^\star$, $\Delta_{i^\star} - \Delta_{i_\star} = \delta$. This proves that

$$|v_{j^\star} - v_{j_\star}| \leqslant \frac{\lambda^{\mathrm{EU}}}{\delta}.$$

By the definition of $j_\star$ and $j^\star$, we have for all $i, j \in N$

$$|v_i - v_j| \leqslant |v_{j^\star} - v_{j_\star}| \leqslant \frac{\lambda^{\mathrm{EU}}}{\delta}.$$

Finally for any $i \in N$

$$\left|v_i - \frac{1}{n}\right| = \left|v_i - \frac{1}{n}\sum_{j\in N} v_j\right| = \frac{1}{n}\left|\sum_{j\in N\setminus\{i\}} (v_j - v_i)\right| \leqslant \frac{n-1}{n}\frac{\lambda^{\mathrm{EU}}}{\delta}.$$

By Proposition 14, $\lambda^{\mathrm{EU}} \leqslant n\chi^{\mathrm{EU}}$. Hence the two inequalities in Proposition 15 are proved. These inequalities are reached with $x = (0,\ldots,0)$, $y = (0,\ldots,0,1)$, $v = (0,\ldots,0,1)$ since we obtain $v = \frac{n-1}{n}$, $\delta = 1$, $\lambda^{\mathrm{EU}} = 1$ and $\chi^{\mathrm{EU}} = \frac{1}{n}$. □

Proof of Lemma 6. One has

$$\begin{aligned}\frac{1}{n!}\sum_{\pi\in\Pi(N)} H^{\mathrm{Maj}}_{\pi\circ v}(y,x) &= \sum_{i\in A^+(y,x)}\frac{1}{n!}\sum_{\pi\in\Pi(N)} v_{\pi(i)} - \sum_{i\in A^-(y,x)}\frac{1}{n!}\sum_{\pi\in\Pi(N)} v_{\pi(i)}\\ &= \sum_{i\in A^+(y,x)}\frac{1}{n} - \sum_{i\in A^-(y,x)}\frac{1}{n} = H^{\mathrm{Maj}}_{w^{\mathrm{Maj}}}(y,x). \qquad \square\end{aligned}$$

Proof of Proposition 16. For all $\pi \in \Pi(N)$, $H^{\mathrm{Maj}}_{\pi\circ v}(y,x) \geqslant H^{\mathrm{Maj}}_{\underline{\pi}_N\circ v}(y,x)$. Hence, by Lemma 6, $H^{\mathrm{Maj}}_{w^{\mathrm{Maj}}}(y,x) \geqslant H^{\mathrm{Maj}}_{\underline{\pi}_N\circ v}(y,x)$.

The *if* part of the second part of the lemma is obvious. Let us show the *only if* part. Assume thus that $H^{\mathrm{Maj}}_{\underline{\pi}_N\circ v}(y,x) = H^{\mathrm{Maj}}_{w^{\mathrm{Maj}}}(y,x)$. Then for all $\pi \in \Pi(N)$, $H^{\mathrm{Maj}}_{\pi\circ v}(y,x) \geqslant H^{\mathrm{Maj}}_{v}(y,x)$. Let $i \neq j$ in N, and let π be the permutation permuting i and j and leaving the other criteria. We have

$$H^{\mathrm{Maj}}_{v}(y,x) - H^{\mathrm{Maj}}_{\pi\circ v}(y,x) = \begin{cases} 0 & \text{if } i,j \in A^+(y,x) \text{ or } i,j \in A^-(y,x) \text{ or } i,j \in A^=(y,x),\\ 2(v_i - v_j) & \text{if } i \in A^+(y,x),\ j \in A^-(y,x),\\ 2(v_j - v_i) & \text{if } i \in A^-(y,x),\ j \in A^+(y,x),\\ v_i - v_j & \text{if } [i \in A^+(y,x),\ j \in A^=(y,x)] \text{ or } [i \in A^=(y,x),\ j \in A^-(y,x)],\\ v_j - v_i & \text{if } [i \in A^=(y,x),\ j \in A^+(y,x)] \text{ or } [i \in A^-(y,x),\ j \in A^=(y,x)]. \end{cases}$$

By (8), $A^+(y,x) \cup A^-(y,x) \neq \emptyset$. If $A^+(y,x) \neq \emptyset$, then we obtain $v_i = v_j$ for all $i \in A^+(y,x)$ and all $j \in A^-(y,x) \cup A^=(y,x)$. Hence $v_i = v_j$ for all $i, j \in N$. Now if $A^-(y,x) \neq \emptyset$, then we obtain $v_i = v_j$ for all $i \in A^-(y,x)$ and all $j \in A^+(y,x) \cup A^=(y,x)$. Hence $v_i = v_j$ for all $i, j \in N$. □

Proof of Proposition 17. We want to have a lower bound of

$$\mathcal{H}^{\text{Maj}} := \min_{x,y:\ (8)} \min_{v\in\mathcal{W}(\text{Maj})\setminus w^{\text{Maj}}} \frac{H^{\text{Maj}}_{w^{\text{Maj}}}(y,x) - \min_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x)}{\max_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x) - \min_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x)}.$$

One has that

$$\max_{x,y:\ (8)} \max_{v\in\mathcal{W}(\text{Maj})\setminus w^{\text{Maj}}} \max_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x) = 1,$$

$$\min_{x,y:\ (8)} \min_{v\in\mathcal{W}(\text{Maj})\setminus w^{\text{Maj}}} \min_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x) = -1$$

are attained with $v_1 = 1$ and $v_i = 0$ for $i \neq 1$, and $A^+(y,x) = \{1\}$ (for the first relation) and $A^-(y,x) = \{1\}$ (for the second relation). Hence

$$\max_{x,y:\ (8)} \max_{v\in\mathcal{W}(\text{Maj})\setminus w^{\text{Maj}}} \Big(\max_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x) - \min_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x) \Big) \leqslant 2,$$

and

$$\mathcal{H}^{\text{Maj}} \geqslant \frac{1}{2} \min_{x,y:\ (8)} \min_{v\in\mathcal{W}(\text{Maj})\setminus w^{\text{Maj}}} \Big(H^{\text{Maj}}_{w^{\text{Maj}}}(y,x) - \min_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x) \Big).$$

Setting $L_v(A^+, A^-) = \sum_{i\in A^+} v_i - \sum_{i\in A^-} v_i$, we have

$$\mathcal{H}^{\text{Maj}} \geqslant \frac{1}{2} \min_{(A^+,A^-)\in\mathcal{Q}(N):\ (8)} \min_{v\in\mathcal{W}(\text{Maj})\setminus w^{\text{Maj}}} \Big(L_{w^{\text{Maj}}}\big(A^+, A^-\big) - \min_{\pi\in\Pi(N)} L_{\pi\circ v}\big(A^+, A^-\big) \Big)$$

where $\mathcal{Q}(N) = \{(A^+, A^-) \in 2^N \times 2^N :\ A^+ \cap A^- = \emptyset\}$. Without loss of generality, one can assume that

$$v_1 \leqslant \cdots \leqslant v_n,$$
$$A^+ = \{1, \ldots, p\} \quad \text{and} \quad A^- = \{q, \ldots, n\}$$

where $p < q$ are in $\{0, \ldots, n+1\}$. Then the following two relations hold

$$L_{w^{\text{Maj}}}\big(A^+, A^-\big) = \frac{p - (n - q + 1)}{n},$$
$$\min_{\pi\in\Pi(N)} L_{\pi\circ v}\big(A^+, A^-\big) = (v_1 + \cdots + v_p) - (v_q + \cdots + v_n).$$

Condition (8) becomes

$$p < q, \quad p < n, \quad q > 1 \quad \text{and} \quad [p \geqslant 1 \text{ or } q \leqslant n]. \tag{21}$$

Hence

$$\mathcal{H}^{\text{Maj}} \geqslant \frac{1}{2} \min_{p,q:\ (21)} \min_{v\in\mathcal{W}(\text{Maj})\setminus w^{\text{Maj}}:\ v_1\leqslant\cdots\leqslant v_n} F$$

where $F := \frac{p-(n-q+1)}{n} - [(v_1 + \cdots + v_p) - (v_q + \cdots + v_n)]$. We write

$$\mathcal{H}^{\text{Maj}} \geqslant \frac{1}{2} \min_{p,q:\ (21)} \min_{v:\ 0\leqslant v_1\leqslant\cdots\leqslant v_n \text{ and } v_1+\cdots+v_n=1} F.$$

Proceeding exactly as in the proof of Proposition 14, the minimum is necessarily attained on a vertex of the polytope

$$U = \left\{ v \in \mathbb{R}^n,\ 0 \leqslant v_1 \leqslant \cdots \leqslant v_n \text{ and } \sum_{i=1}^n v_i = 1 \right\}.$$

A vector $v \in U$ is a vertex of U iff there exists $r \in \{1, \ldots, n\}$ such that

$$v_1 = \cdots = v_{r-1} = 0, \qquad v_r = \cdots = v_n =: \alpha.$$

Since v is normalized, one has $\alpha = \frac{1}{n-r+1}$. One has $r \geqslant 2$ since v is different from w^{Maj}. One has

$$v_1 + \cdots + v_p = \begin{cases} 0 & \text{if } p < r, \\ (p-r+1)\alpha & \text{if } p \geqslant r, \end{cases}$$
$$v_q + \cdots + v_n = \begin{cases} (n-q+1)\alpha & \text{if } q \geqslant r, \\ (n-r+1)\alpha & \text{if } q < r. \end{cases}$$

We have the following cases:

1. $p < r$, $q \geqslant r$. We have
$$F = \frac{p-(n-q+1)}{n} + \frac{n-q+1}{n-r+1} = \frac{p}{n} + \frac{(n-q+1)(r-1)}{n(n-r+1)} \geqslant \frac{1}{n(n-1)}$$
for $p = 0$, $r = 2$ and $q = n$.
2. $p < r$, $q < r$. We have
$$F = \frac{p-(n-q+1)}{n} + \frac{n-r+1}{n-r+1} = \frac{p+q-1}{n} \geqslant \frac{1}{n}.$$
3. $p \geqslant r$, $q \geqslant r$. We have
$$F = \frac{p-(n-q+1)}{n} - \frac{(p-r+1)-(n-q+1)}{n-r+1} = \frac{(r-1)(n-p-q+2)}{n-r+1} \geqslant \frac{1}{n}.$$
4. $p \geqslant r$, $q < r$. This situation is impossible since $q > r$.

We have shown that

$$\frac{\chi^{\text{Maj}}}{\max_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x) - \min_{\pi\in\Pi(N)} H^{\text{Maj}}_{\pi\circ v}(y,x)} \geqslant \theta := \frac{1}{2n(n-1)}.$$

Then $|H^{\text{Maj}}_{\pi\circ v}(y,x) - H^{\text{Maj}}_{\pi'\circ v}(y,x)| \leqslant \frac{\chi^{\text{Maj}}}{\theta}$ for every $\pi, \pi' \in \Pi(N)$. Let $i, j \in N$. Define π and π' as in the proof of Proposition 15. If i, j are not in the same set among $A^+(y,x)$, $A^-(y,x)$ and $A^=(y,x)$, then we obtain $|v_i - v_j| \leqslant \frac{\chi^{\text{Maj}}}{\theta}$. Hence for all $i, j \in N$, $|v_i - v_j| \leqslant \frac{2\chi^{\text{Maj}}}{\theta}$. Therefore

$$\left| v_i - \frac{1}{n} \right| = \frac{1}{n} \left| \sum_{j\neq i} (v_j - v_i) \right| \leqslant \frac{n-1}{n} \frac{2\chi^{\text{Maj}}}{\theta} \leqslant \frac{2\chi^{\text{Maj}}}{\theta}. \qquad \square$$

For the proof of Proposition 18, we need the following result.

Lemma 7. *Let $p \in \mathbb{N}$ and $a, b \in \mathbb{R}^p$. If $a_i > b_i$ for all $i \in \{1, \ldots, p\}$, then*

$$\bigwedge_{i=1}^{p} a_i > \bigwedge_{i=1}^{p} b_i.$$

Proof. By associativity of the operator $\wedge$, it is enough to prove the lemma for $p = 2$. One has then $a_1 > b_1 \geqslant b_1 \wedge b_2$ and $a_2 > b_2 \geqslant b_1 \wedge b_2$. Hence $a_1 \wedge a_2 > b_1 \wedge b_2$, which proves the result. $\square$

Proof of Proposition 18. One has for any $\pi \in \Pi(N)$

$$\bigwedge_{i=1}^{n} \big(y_i \vee (1 - v_{\pi(i)})\big) > \bigwedge_{i=1}^{n} \big(x_i \vee (1 - v_{\pi(i)})\big).$$

By Lemma 7,

$$\bigwedge_{\pi\in\Pi(N)} \left[\bigwedge_{i=1}^{n} \big(y_i \vee (1 - v_{\pi(i)})\big) \right] > \bigwedge_{\pi\in\Pi(N)} \left[\bigwedge_{i=1}^{n} \big(x_i \vee (1 - v_{\pi(i)})\big) \right].$$

Hence

$$\bigwedge_{i=1}^{n} \left[\bigwedge_{\pi\in\Pi(N)} \big(y_i \vee (1 - v_{\pi(i)})\big) \right] > \bigwedge_{i=1}^{n} \left[\bigwedge_{\pi\in\Pi(N)} \big(x_i \vee (1 - v_{\pi(i)})\big) \right]$$

and

$$\bigwedge_{i=1}^{n}\left[\bigwedge_{j=1}^{n}\big(y_i \vee (1-v_j)\big)\right] > \bigwedge_{i=1}^{n}\left[\bigwedge_{j=1}^{n}\big(y_i \vee (1-v_j)\big)\right].$$

Since there exists k such that $v_k = 1$, one has

$$\bigwedge_{j=1}^{n}\big(y_i \vee (1-v_j)\big) = y_i.$$

This gives

$$\bigwedge_{i=1}^{n} y_i > \bigwedge_{i=1}^{n} x_i. \qquad \square$$

Proof of Proposition 19. Let $t \in E$ with $t > e$. The cardinality of the set E is $2|A^+(y,x)|+1$. There are at most $|A^+(y,x)|-1$ elements of $E \setminus \{1\}$ greater than or equal to t. Hence

$$\big|\big\{i \in A^+(y,x)\colon x_i \geqslant t\big\}\big| + \big|\{i \in N\colon 1-v_i \geqslant t\}\big| = \big|\big\{\alpha \in E \setminus \{1\}\colon \alpha \geqslant t\big\}\big| < \big|A^+(y,x)\big|.$$

Therefore there does not exist $\pi \in \Pi(N)$ such that

$$h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x) \geqslant t.$$

Hence (9) holds.

Only if part of the proposition: We assume that $H^{\mathrm{Pess}}_{\pi\circ v}(y,x) > 0$ for every $\pi \in \Pi(N)$.

By the definition of $A^+(y,x)$, $A^=(y,x)$ and $A^-(y,x)$, we have $h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(y) \geqslant h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x)$, $h^{\mathrm{Pess},A^=(y,x)}_{\pi\circ v}(y) = h^{\mathrm{Pess},A^=(y,x)}_{\pi\circ v}(x)$ and $h^{\mathrm{Pess},A^-(y,x)}_{\pi\circ v}(y) \leqslant h^{\mathrm{Pess},A^-(y,x)}_{\pi\circ v}(x)$. If we assume that $h^{\mathrm{Pess}}_{\pi\circ v}(x) = h^{\mathrm{Pess},A^-(y,x)}_{\pi\circ v}(x) \wedge h^{\mathrm{Pess},A^=(y,x)}_{\pi\circ v}(x)$, then $h^{\mathrm{Pess}}_{\pi\circ v}(y) \leqslant h^{\mathrm{Pess},A^-(y,x)}_{\pi\circ v}(y) \wedge h^{\mathrm{Pess},A^=(y,x)}_{\pi\circ v}(y) \leqslant h^{\mathrm{Pess}}_{\pi\circ v}(x)$, which contradicts $H^{\mathrm{Pess}}_{\pi\circ v}(y,x) > 0$. Hence

$$h^{\mathrm{Pess}}_{\pi\circ v}(x) = h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x),$$
$$h^{\mathrm{Pess}}_{\pi\circ v}(x) < h^{\mathrm{Pess},A^-(y,x)}_{\pi\circ v}(x) \wedge h^{\mathrm{Pess},A^=(y,x)}_{\pi\circ v}(x).$$

Hence (iii) holds. If we assume that $h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(y) = h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x)$, then $H^{\mathrm{Pess}}_{\pi\circ v}(y,x) \leqslant 0$ and a contradiction is raised. Hence (i) holds.

Let $L' := \{i \in A^-(y,x) \cup A^=(y,x)\colon\ y_i \leqslant e\}$. Assume by contradiction that $L' \neq \emptyset$. Since v is normalized, there exists $m \in N$ such that $v_m = 1$. Consider then $\pi_2^\star \in \Pi(N)$ such that $\pi_2^\star(k) = m$ for some $k \in L'$, and $\pi_2^\star(i) = \pi_N^v \circ (\pi^x_{A^+(y,x)})^{-1}(i)$ for all $i \in A^+(y,x)$. In the permutation $\pi_2^\star$, the largest values of $1-v$ are assigned to the smallest values of x and vice versa. More precisely, for every $i \in A^+(y,x)$, either $x_i \geqslant e$ or $1 - v_{\pi_2^\star(i)} \geqslant e$. Hence

$$\max_{\pi\in\Pi(N)} h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x) = h^{\mathrm{Pess},A^+(y,x)}_{\pi_2^\star\circ v}(x) = e.$$

Since $\pi_2^\star(k) = m$ for some $k \in L'$, we have

$$h^{\mathrm{Pess},A^-(y,x)}_{\pi_2^\star\circ v}(y) \wedge h^{\mathrm{Pess},A^=(y,x)}_{\pi_2^\star\circ v}(y) \leqslant e = h^{\mathrm{Pess},A^+(y,x)}_{\pi_2^\star\circ v}(x) = h^{\mathrm{Pess}}_{\pi_2^\star\circ v}(x).$$

This contradicts $H^{\mathrm{Pess}}_{\pi_2^\star\circ v}(y,x) > 0$. Hence $L' = \emptyset$.

If part of the proof: Assume that (i), (ii) and (iii) hold. Let $\pi \in \Pi(N)$. Then $h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(y) > h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x) = h^{\mathrm{Pess}}_{\pi\circ v}(x)$ and $e \geqslant h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x) = h^{\mathrm{Pess}}_{\pi\circ v}(x)$ (by (9) and (iii)). By (ii), $h^{\mathrm{Pess},A^-(y,x)}_{\pi\circ v}(y) \wedge h^{\mathrm{Pess},A^=(y,x)}_{\pi\circ v}(y) > e$. Hence $h^{\mathrm{Pess}}_{\pi\circ v}(y) > h^{\mathrm{Pess}}_{\pi\circ v}(x)$. $\square$

Proof of Proposition 20. Assume that either $V_{i^+} = \emptyset$ or (10) holds. Let $\pi \in \Pi(N)$. When $\pi(i^+) \notin V_{i^+}$, we have $h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(y) = y_{i^+} > x_{i^+} \vee (1 - v_{\pi(i^+)}) \geqslant h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ v}(x)$ since $y_{i^+} > x_{i^+}$ and $y_{i^+} > 1 - v_{\pi(i^+)}$.

Suppose now that $\pi(i^+) \in V_{i^+}$ (hence $V_{i^+} \neq \emptyset$). Then, by (10),

$$L_\pi := \big\{j \in A^+(y,x)\colon x_j < 1 - v_{k^+} \text{ and } 1 - v_{\pi(j)} < y_{i^+}\big\} \neq \emptyset.$$

For all $j \in L_\pi$, $1 - v_{\pi(j)} < y_{i^+} \leqslant y_j$ and thus $y_j \vee (1 - v_{\pi(j)}) = y_j > x_j \vee (1 - v_{\pi(j)})$. Moreover, for all $j \in L_\pi$, $x_j \vee (1 - v_{\pi(j)}) < 1 - v_{k^+}$ since $x_j < 1 - v_{k^+}$ and $1 - v_{\pi(j)} < y_{i^+} \leqslant 1 - v_{k^+}$. Hence

$$h^{\mathrm{Pess},L_\pi}_{\pi\circ\nu}(y) > h^{\mathrm{Pess},L_\pi}_{\pi\circ\nu}(x) \quad \text{and} \quad h^{\mathrm{Pess},L_\pi}_{\pi\circ\nu}(x) < 1-\nu_{k^+}.$$

On the other hand, for all $j \in A^+(y,x)\setminus L_\pi$, either $y_j > x_j \geqslant 1-\nu_{k^+}$ or $1-\nu_{\pi(j)} \geqslant y_{i^+}$ (and thus $1-\nu_{\pi(j)} \geqslant 1-\nu_{k^+}$). Hence $h^{\mathrm{Pess},A^+(y,x)\setminus L_\pi}_{\pi\circ\nu}(y) \geqslant 1-\nu_{k^+}$ and

$$h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ\nu}(y) \geqslant (1-\nu_{k^+}) \wedge h^{\mathrm{Pess},L_\pi}_{\pi\circ\nu}(y) > h^{\mathrm{Pess},L_\pi}_{\pi\circ\nu}(x) \geqslant h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ\nu}(x).$$

Conversely, assume that $V_{i^+} \neq \emptyset$ and that (10) is not satisfied. We have $i^+ \in L := \{j \in A^+(y,x)\colon\ x_j < 1-\nu_{k^+}\}$ since $x_{i^+} < y_{i^+} \leqslant 1-\nu_{k^+}$. Since $|V_{i^+}| \geqslant |L|$, there exists $\pi_1^\star \in \Pi(N)$, such that $\pi_1^\star(i^+) = k^+$ and for all $i \in L$, $1-\nu_{\pi_1^\star(i)} \geqslant y_{i^+}$ (and thus $1-\nu_{\pi_1^\star(i)} \geqslant 1-\nu_{k^+}$). Since $\pi_1^\star(i^+) = k^+$, $h^{\mathrm{Pess},L}_{\pi_1^\star\circ\nu}(y) = 1-\nu_{k^+}$ and $h^{\mathrm{Pess},L}_{\pi_1^\star\circ\nu}(x) = 1-\nu_{k^+}$. Since $y_i > x_i \geqslant 1-\nu_{k^+}$ for all $i \in A^+(y,x)\setminus L$, we have $h^{\mathrm{Pess},A^+(y,x)\setminus L}_{\pi_1^\star\circ\nu}(y) \geqslant 1-\nu_{k^+}$ and $h^{\mathrm{Pess},A^+(y,x)\setminus L}_{\pi_1^\star\circ\nu}(x) \geqslant 1-\nu_{k^+}$. Hence $h^{\mathrm{Pess},A^+(y,x)}_{\pi_1^\star\circ\nu}(y) = h^{\mathrm{Pess},A^+(y,x)}_{\pi_1^\star\circ\nu}(x)$. □

Proof of Proposition 21.

- Assume that $I = \emptyset$. Since ν is normalized, there exists $m \in N$ such that $\nu_m = 1$. Let $\pi_3^\star \in \Pi(N)$ such that $\pi_3^\star(j^+) = m$. Since $\bigwedge_{i\in N} x_i = x_{j^+}$, we get $h^{\mathrm{Pess}}_{\pi_3^\star\circ\nu}(x) = x_{j^+}$, which contradicts (11).
- Assume that $I \neq \emptyset$ but $|\{k \in N\colon\ 1-\nu_k \geqslant x_{j^+}\}| \geqslant |I|$. Then there exists $\pi_4^\star \in \Pi(N)$ such that for all $i \in I$, $1-\nu_{\pi_4^\star(i)} \geqslant x_{j^+}$ and $\pi_4^\star(j^+) = m$. We obtain $h^{\mathrm{Pess},A^+(y,x)}_{\pi_4^\star\circ\nu}(x) \geqslant x_{j^+}$ and $h^{\mathrm{Pess},A^-(y,x)}_{\pi_4^\star\circ\nu}(x) \wedge h^{\mathrm{Pess},A^=(y,x)}_{\pi_4^\star\circ\nu}(x) = x_{j^+}$. This contradicts (11).
- Assume that (12) holds. Then for all $\pi \in \Pi(N)$, there exists $i \in I$ such that $1-\nu_{\pi(i)} < x_{j^+}$. Hence $h^{\mathrm{Pess},A^+(y,x)}_{\pi\circ\nu}(x) \leqslant x_i \vee (1-\nu_{\pi(i)}) < x_{j^+}$ and $h^{\mathrm{Pess},A^-(y,x)}_{\pi\circ\nu}(x) \wedge h^{\mathrm{Pess},A^=(y,x)}_{\pi\circ\nu}(x) \geqslant x_{j^+}$. Hence (11) holds. □

References

[1] L. Amgoud, S. Belabbes, H. Prade, Towards a formal framework for the search of a consensus between autonomous agents, in: 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), Utrecht, 2005, pp. 537–543.
[2] L. Amgoud, J.-F. Bonnefon, H. Prade, An argumentation-based approach to multiple criteria decision, in: 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'2005), Barcelona, 2005, pp. 269–280.
[3] L. Amgoud, H. Prade, Using arguments for making and explaining decisions, Artificial Intelligence 173 (2009) 413–436.
[4] G.E. Andrews, The Theory of Partitions, Encyclopedia of Mathematics and Its Applications, 2nd edition, Addison-Wesley, 1976.
[5] K. Arrow, Social Choice and Individual Values, 2nd edition, Wiley, 1963.
[6] K.J. Arrow, A.K. Sen, K. Suzumura, Handbook of Social Choice and Welfare, Handbooks in Economics, Elsevier, 2002.
[7] C.A. Bana e Costa, J.C. Vansnick, A theoretical framework for Measuring Attractiveness by a Categorical Based Evaluation TecHnique (MACBETH), in: Proc. XIth Int. Conf. on MultiCriteria Decision Making, Coimbra, Portugal, August 1994, pp. 15–24.
[8] F. Barbaresco, J.C. Deltour, G. Desodt, B. Durand, T. Guenais, Ch. Labreuche, Intelligent M3R radar time resources management: Advanced cognition, agility & autonomy capabilities, in: International Radar Conference, Bordeaux, France, October 12–16, 2009.
[9] C. Boutilier, R. Brafman, C. Domshlak, H. Hoos, D. Poole, CP-nets: a tool for representing and reasoning with conditional Ceteris Paribus preference statements, Journal of Artificial Intelligence Research 21 (2004) 135–191.
[10] D. Bouyssou, T. Marchant, M. Pirlot, A. Tsoukiàs, Ph. Vincke, Evaluation and Decision Models with Multiple Criteria: Stepping Stones for the Analyst, International Series in Operations Research and Management Science, Springer, 2006.
[11] R. Brafman, Intentions, Plans and Practical Reason, Harvard University Press, Massachusetts, 1987.
[12] R. Brafman, M. Tennenholtz, An axiomatic treatment of three qualitative decision criteria, J. ACM 47 (2000) 452–482.
[13] C.B. Callaway, J.C. Lester, Narrative prose generation, Artificial Intelligence 139 (2002) 213–252.
[14] G. Carenini, J.D. Moore, Generating and evaluating evaluative arguments, Artificial Intelligence 170 (2006) 925–952.
[15] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1953) 131–295.
[16] V. Chvatal, Linear Programming, W.H. Freeman and Company, New York, 1983.
[17] E.P. Corbett, R.J. Connors, Classical Rhetoric for the Modern Student, Oxford University Press, Oxford, 1999.
[18] Y. Dimopoulos, P. Moraitis, L. Amgoud, Theoretical and computational properties of preference-based argumentation, in: 18th European Conference on Artificial Intelligence (ECAI'08), Patras, Greece, 2008, pp. 463–467.
[19] J. Doyle, R. Thomason, Background to qualitative decision theory, The AI Magazine 20 (1999) 55–68.
[20] D. Dubois, H. Fargier, P. Perny, Qualitative decision theory with preference relations and comparative uncertainty: An axiomatic approach, Artificial Intelligence 148 (2003) 219–260.
[21] D. Dubois, H. Fargier, H. Prade, Refinements of the maximin approach to decision-making in a fuzzy environment, Fuzzy Sets and Systems 81 (1996) 103–122.
[22] D. Dubois, H. Prade, Weighted minimum and maximum operations in fuzzy set theory, Information Sciences 39 (1986) 205–210.
[23] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.
[24] D. Dubois, H. Prade, Possibility theory as a basis for qualitative decision theory, in: Proc. Int. Joint Conf. in AI (IJCAI'95), Montreal, Canada, August 1995, pp. 19–25.
[25] P. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (1995) 321–357.
[26] W. Edwards, J.R. Newman, Multiattribute Evaluation, Sage Publications, Cambridge, 1983.
[27] M. Elhadad, Using argumentation in text generation, Journal of Pragmatics 24 (1995) 189–220.
[28] J. Figueira, S. Greco, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Kluwer Academic Publishers, 2005.
[29] P. Fishburn, The Theory of Social Choice, Princeton University Press, 1973.
[30] P. Fishburn, Semiorders and choice functions, Econometrica 43 (1975) 975–977.
[31] J. Fodor, M. Roubens, Fuzzy Preference Modelling and Multi-Criteria Decision Aid, Kluwer Academic Publishers, 1994.
[32] J.B. Grize, Matériaux pour une logique naturelle, Technical report, Travaux du Centre de Recherche Semiologique, No. 29, Université de Neuchâtel, Suisse, 1976.

[33] C.L. Hamblin, Fallacies, Methuen, London, 1970.
[34] R.L. Keeney, H. Raiffa, Decision with Multiple Objectives, Wiley, New York, 1976.
[35] D.A. Klein, Decision Analytic Intelligent Systems: Automated Explanation and Knowledge Acquisition, Lawrence Erlbaum Associates, 1994.
[36] F.H. Knight, Risk, Uncertainty, and Profit, Houghton Mifflin, Boston, New York, 1921.
[37] D.H. Krantz, R.D. Luce, P. Suppes, A. Tversky, Foundations of Measurement, vol. 1, Additive and Polynomial Representations, Academic Press, 1971.
[38] Ch. Labreuche, Argumentation of the results of a multi-criteria evaluation model in individual and group decision aiding, in: Int. Conf. of the Euro Society for Fuzzy Logic and Technology (EUSFLAT), Barcelona, Spain, September 7–9, 2005, pp. 482–487.
[39] Ch. Labreuche, Argumentation of the decision made by several aggregation operators based on weights, in: Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), Paris, France, July 2–7, 2006, pp. 683–691.
[40] R.D. Luce, H. Raiffa, Games and Decisions, Wiley, New York, 1957.
[41] Th. Marchant, Towards a theory of MCDM: Stepping away from social choice theory, Mathematical Social Choice 45 (2003) 343–363.
[42] S. Modgil, Reasoning about preferences in argumentation frameworks, Artificial Intelligence 173 (2009) 901–934.
[43] J. Montmain, G. Mauris, A. Akharraz, Elucidation and decisional risk in a multi criteria decision based on a Choquet integral aggregation: A cybernetic framework, International Journal of Multi-Criteria Decision Analysis 13 (2005) 239–258.
[44] H. Moulin, Axioms of Cooperative Decision Making, Cambridge University Press, 1988.
[45] S.D. Parsons, N.R. Jennings, Negotiation through argumentation – a preliminary report, in: Proc. 2nd Int. Conf. on Multi-Agent Systems (ICMAS), Kyoto, Japan, 1996, pp. 267–274.
[46] C. Perelman, L. Olbrechts-Tyteca, Traité de l'Argumentation, PUF, Paris, 1958.
[47] J.P. Pignon, Ch. Labreuche, A methodological approach for operational and technical experimentation based evaluation of systems of systems architectures, in: Int. Conference on Software & Systems Engineering and Their Applications (ICSSEA), Paris, France, December 4–6, 2007.
[48] A. Rapoport, Decision Theory and Decision Behaviour, Kluwer Academic Publishers, Dordrecht, 1989.
[49] B. Roy, Multicriteria Methodology for Decision Aiding, Kluwer Academic Publishers, Dordrecht, 1996.
[50] T.L. Saaty, A scaling method for priorities in hierarchical structures, J. Math. Psychology 15 (1977) 234–281.
[51] L.J. Savage, The Foundations of Statistics, 2nd edition, Dover, 1972.
[52] M. Schroeder, R. Schweimeier, Notions of attack and justified arguments for extended logic programs, in: F. van Harmelen (Ed.), Proceedings of the European Conference on Artificial Intelligence (ECAI02), Amsterdam, 2002, pp. 536–540.
[53] M. Sugeno, Theory of fuzzy integrals and its applications, PhD thesis, Tokyo Institute of Technology, 1974.
[54] J. Von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, 1944.
[55] W.A. Wagenaar, P.J. van Koppen, H.F.M. Crombag, Anchored Narratives. The Psychology of Criminal Evidence, St. Martin's Press, New York, 1993.
[56] A. Wald, Statistical Decision Functions, Wiley, New York, 1950.
[57] D. Walton, Argumentation Schemes for Presumptive Reasoning, Erlbaum, Mahwah, NJ, 1996.