

Properties of Aggregation Operators Relevant for Ethical Decision Making in Artificial Intelligence

Federico Fioravanti (Universidad Nacional del Sur/CONICET)

Iyad Rahwan (Max Planck Institute for Human Development)

Fernando Tohmé (Universidad Nacional del Sur/CONICET)

DOCUMENTO DE TRABAJO N° 177

Septiembre de 2022

Los documentos de trabajo de la RedNIE se difunden con el propósito de generar comentarios y debate, no habiendo estado sujetos a revisión de pares. Las opiniones expresadas en este trabajo son de los autores y no necesariamente representan las opiniones de la RedNIE o su Comisión Directiva.

The RedNIE working papers are disseminated for the purpose of generating comments and debate, and have not been subjected to peer review. The opinions expressed in this paper are exclusively those of the authors and do not necessarily represent the opinions of the RedNIE or its Board of Directors.

Citar como:

Fioravanti, Federico, Iyad Rahwan y Fernando Tohmé (2022). Properties of Aggregation Operators Relevant for Ethical Decision Making in Artificial Intelligence. *Documento de trabajo RedNIE N°177.*

Properties of Aggregation Operators Relevant for Ethical Decision Making in Artificial Intelligence*

Federico Fioravanti^{†1,2}, Iyad Rahwan⁴, and Fernando Tohmé^{1,3}

¹*Instituto de Matemática de Bahía Blanca, CONICET - UNS,
Bahía Blanca, Argentina*

²*Departamento de Matemática, Universidad Nacional del Sur,
Bahía Blanca, Argentina*

³*Departamento de Economía, Universidad Nacional del Sur,
Bahía Blanca, Argentina*

⁴*Center for Humans and Machines, Max Planck Institute for
Human Development, Berlin, Germany*

Abstract

We present an axiomatic study of *aggregation operators* that could be applied to ethical AI decision making. The information is given here by different preferences over the decisions to be made by automated systems. We consider two different but very intuitive notions of preference of an alternative over another one, namely *pairwise majority* and *position* dominance. Preferences are represented by permutation processes over alternatives and aggregation rules are applied to obtain results that are socially considered to be ethically correct. We address the problem of the stability of the aggregation process, which is important when the information is variable. In this setting we find many aggregation rules that satisfy desirable properties for an autonomous system.

Keywords: Aggregation Operators; Permutation Process; Decision Analysis.

*This work began during Fioravanti's research stay at the MPI for Human Development. Thanks are due to the funding of the German Academic Exchange Service (DAAD).

[†]federico.fioravanti@uns.edu.ar (corresponding author)

1 Introduction

Many relevant scenarios arising from the use of Artificial Intelligent (AI) systems require making ethical decisions. So for instance, consider the case of an autonomous vehicle suffering a brake failure in the midst of heavy urban traffic. Should it swerve its course and hit a wall, killing its passengers? Or stay its way and kill two pedestrians who are crossing the street? Of course, a definite answer depends on the availability of several pieces of information that are missing in this story, such as the age and gender of both the passengers and pedestrians. But even with access to the complete information about the situation, it is difficult from an ethical point of view to make a decision. This kind of ethical decision-making problems, have constituted, for a long time, a great challenge for Artificial Intelligence [40]. Even when ground-truth ethical principles were available, the lack of corresponding formal specifications make it difficult to solve these problems.

Some authors suggest that “an approximation as agreed upon society” must be applied when ethical principles are not available [13]. Conitzer et al. [11] discuss a game theoretic and a machine learning approach to the development of a general framework, instead of appealing to the use of ad-hoc rules at each scenario. Furthermore, it would also be convenient to automatize the decision-making process, aggregating the diverse opinions held in the society on ethical dilemmas [21, 11]. In this sense, Rossi et al. [36] discuss related problems where preferences (not necessarily on moral issues) are aggregated by morally-ranking all the alternatives.

Noothigattu et al. [30] propose and implement a concrete approach to ethical decision-making based on tools drawn from Computational Social Choice [5, 6]. It follows four steps:

- i. Data Collection:* human voters are asked to compare pairs of alternatives,
- ii. Learning:* a model of each individual preference is generated,
- iii. Summarization:* a single model incorporating those preferences is created, and
- iv. Aggregation:* a voting rule aggregates the individual preferences into a collective decision.

Notice that these steps involve the definition of an *aggregation operator*, i.e. a mathematical procedure to combine information, in this case about the preferences over decisions to be made over alternative decisions on ethical settings. We seek to define aggregation processes yielding solutions as close to the ground-truth ethical principles as possible. We intend to provide a sound foundation for the automatizing of ethical decision-making in situations like that of the case of the failure of the braking system of an autonomous car [32], or the automatic assignation of priorities in kidney exchange programs [17] or even in the assistance of juridical deliberation [1, 3].

This paper presents, essentially, a ‘technical’ contribution to the discipline of *Machine Ethics*, the field concerned with the ‘ethical behavior’ of autonomous intelligent systems operating in social environments [41, 12, 27]. Following Bjorgen et al. [4], who propose estimating the ethical performance of an autonomous system on the basis of the responses to some ethical dilemmas. The results presented in this work may help to set up these ‘benchmarks’.

To address our main research question, we need to establish a model of preferences for each voter. There exists a vast literature that uses random utility models for the elicitation of preferences [22, 31, 2]. That is, an agent’s preferences are modeled by drawing, for each alternative, a real-value from a parameterized distribution, yielding a ranking of the alternatives according to their scores. In our setting, since an autonomous system may face different scenarios, each of which with a finite set of alternatives that may be obtained from an infinite set, we need a more general notion [9]. That is why we consider general models known as permutation processes [30].

Once we have a model for each voter, as well as a summary of those models (Step *iii*), we need a voting rule¹ to select the best alternatives. We consider two different conceptions of what it means that an alternative is the ‘best’ one, adapting to this setting the ideas of domination among alternatives of Caragiannis et al. [8]. We present some properties that a good voting rule should verify in terms of these dominance notions. The results obtained can be considered more or less positive for each dominance notion, depending on how many voting rules satisfy some desirable properties. We consider the different domains that may arise depending on the definition of the models of the individuals.

This work intends to contribute to the theory of aggregation of permutation

¹More formally known as *Social Choice correspondence*.

processes [34, 30, 24]. In particular, we focus on the very highly relevant property of *stability*, which ensures the consistency of decisions across different scenarios of variable information. This property matters for the characterization of *information aggregation operators*, since they are susceptible to information losses [7, 19].

The plan of the paper is as follows. In Section 2, we introduce a general model of the aggregation of permutation processes. Sections 3 and 4 introduce different dominance definitions and important properties that voting rules should verify. In Section 5 we consider the setting in which the aggregation of permutations yields positive results. In Section 6 we introduce the concept of stability and find some rules satisfying it. Finally, Section 7 concludes with a brief discussion about some future lines of research.

2 Model

Following the line of [34, 30] we consider a potentially infinite set of alternatives X . Let $A \subseteq X$ be a finite subset of size n . A total ordering over A can be characterized by a bijection $\sigma : A \rightarrow \{1, \dots, n\}$ such that $\sigma(a) = j$ indicates that the position of alternative a in σ is j . By a slight abuse of language we call σ a *permutation* of A .² Let S_A be the set of all permutations of A . For each $\sigma \in S_A$, $\sigma(a) < \sigma(b)$ is interpreted as indicating that a is more preferred than b in the order characterized by σ . We denote this as $a >_\sigma b$.

With $\sigma|_B$ we denote the restriction of σ on $B \subseteq A$. Given a probability distribution P over S_A and $B \subseteq A$, we have that

$$P_B(\sigma') = \sum_{\sigma \in S_A: \sigma|_B = \sigma'} P(\sigma)$$

where $\sigma' \in S_B$.

A *permutation process* Π is a collection of probability distributions over sets of permutations S_B such that $B \subseteq A$, with $\Pi(A)$ being a distribution over S_A . A permutation process is *consistent* if $\Pi(A)|_B = \Pi(B)$ for any finite subset $B \subseteq A$, that is, if the distribution over S_B is obtained by marginalizing the distribution over S_A on the extra alternatives in $A \setminus B$.

²We will also refer to permutations over sets of alternatives as votes or rankings.

Consistent permutation processes can be naturally represented in terms of utilities [24]. That is, given a stochastic process U , indexed by X , such that for any $A = \{x_1, x_2, \dots, x_n\} \subseteq X$, we can define the probability of $\sigma \in S_A$ as the probability that $\text{sort}(U_{x_1}, U_{x_2}, \dots, U_{x_n})$ coincides with that of $>_\sigma$, where sort is an operation that yields a linear ordering of $\{U_{x_j}\}_{j=1}^n$. Thus, $U : A \rightarrow \mathbb{R}$ can be conceived as a stochastic utility function.

We follow the line of Parkes et al. [31] and focus our attention on permutation processes in which the random utilities are independently drawn from distributions in the exponential family, like the Normal, Poisson, Gamma, Binomial and Negative Binomial distributions [28].

Consider a finite set of N voters. The preferences of a voter i over a finite set of alternatives A , is denoted by $\sigma_i \in S_A$. A preference profile is defined as a collection of N individual rankings, $\bar{\sigma} = (\sigma_1, \dots, \sigma_N)$. In our setting, the identity of the voters is not important. So we can consider an anonymous profile $\pi \in [0, 1]^{|A|!}$, where, for each $\sigma \in S_A$, $\pi(\sigma) \in [0, 1]$ is the probability that a random voter has the ranking σ .

The relation between an anonymous preference profile and a permutation process over a finite subset $A \subseteq X$, is that for $\sigma \in S_A$, $\pi(\sigma)$ is the probability of σ in $\Pi(A)$ ³.

We want to analyze some axiomatic aspects of the aggregation of preferences induced by a permutation process, defined by the selection of the winning alternatives. We define a *social choice correspondence* (SCC) as a function f that maps an anonymous preference profile defined over a finite subset $A \subseteq X$ into a nonempty subset of A . Examples of SCC are the *plurality* rule, that selects the alternatives that are on the top of the largest number of individual preferences, $\text{argmax}_{a \in A} \sum_{\sigma \in S_A: \sigma(a)=1} \pi(\sigma)$, or the *antiplurality* rule, that selects the alternatives that are the least preferred by the smallest number of individual preferences, $\text{argmin}_{a \in A} \sum_{\sigma \in S_A: \sigma(a)=N} \pi(\sigma)$.

3 Pairwise Majority Dominance

We can define what it means for an alternative to be better than another one in aggregate terms, based on the concept of *majoritarian power*. That is, an alternative a is better than an alternative b , if a is preferred to b by a

³Because of this, we can refer to $\Pi(A)$ as an anonymous preference profile.

majority of the individuals, i.e., a group of cardinality larger or equal than $\lfloor \frac{N}{2} \rfloor$. We have the following notion of dominance:

Definition 1. *An alternative $a \in A$ pairwise majority-dominates (PM-dominates) another alternative $b \in A$ in an anonymous preference profile π over A , denoted $a \triangleright_{\pi}^{pm} b$, if $|\{\sigma \in S_A : a >_{\sigma} b\}| \geq |\{\sigma \in S_A : b >_{\sigma} a\}|$.*

This binary relation is complete but not transitive, due to the possible existence of Condorcet cycles.⁴

The following is a desirable property to be satisfied by a social choice correspondence. Intuitively, it means that the voting rule is ‘consistent’ with the dominance relation: if it selects an alternative b , then it must also select the alternatives that are better than b .

Definition 2. *An anonymous SCC f is said to be pairwise majority-dominance-efficient (PMD-efficient) if for every anonymous preference profile π and any two alternatives $a, b \in A$, if $a \triangleright_{\pi}^{pm} b$, then $b \in f(\pi)$ implies that $a \in f(\pi)$.*

A stronger requirement states, roughly, that dominated alternatives should not be selected, unless they also dominate the alternatives that dominate them.

Definition 3. *An anonymous SCC f is said to be strongly PMD-efficient if for every anonymous preference profile π over A , and any two alternatives $a, b \in A$ such that $a \triangleright_{\pi}^{pm} b$, then:*

-if $b \not\triangleright_{\pi}^{pm} a$, then $b \notin f(\pi)$,

-if $b \triangleright_{\pi}^{pm} a$, then $b \in f(\pi)$ if and only if $a \in f(\pi)$.

It is clear that strong-efficiency implies efficiency. This notion sets the bar too high on SCCs, since there are no rules satisfying it on unrestricted domains. This problem arises because Condorcet cycles can arise in those domains. But even efficiency is a very demanding condition:⁵

Proposition 1. *Black, Dodgson, Young, Kemeny, Nanson, Minimax and Fishburn rules are not PMD-efficient voting rules.*

This means that this notion of efficiency is not satisfied by many (important and well studied) voting rules. An alternative intuition is that an “efficient” rule should select the ‘most’ dominating alternatives, or at least, the ones that are the ‘less’ dominated. The *Schwartz set* captures this idea [38]:

⁴A Condorcet cycle or ‘Condorcet Paradox’ occurs when $a \triangleright_{\pi}^{pm} b$, $b \triangleright_{\pi}^{pm} c$ and $c \triangleright_{\pi}^{pm} a$.

⁵The definitions of the voting rules as well as the proof of Proposition 1 can be found in the Appendix. The definitions may also be found in [16] and [14].

Definition 4. *The Schwartz set of a profile of preferences π , denoted $Sc(\pi)$, is the minimal set $Sc(\pi) \subseteq A$ such that $Sc(\pi) \neq \emptyset$ verifying that for any $a \in Sc(\pi)$ there is no $b \in A \setminus Sc(\pi)$ such that $b \succ_{\pi}^{pm} a$.*

This set is nonempty for any profile of preferences. The *Schwartz voting rule* is defined as the SCC that selects as winning alternatives the entire Schwartz set. The following result follows immediately from the definition of a Schwartz set:

Proposition 2. *The Schwartz voting rule is PMD-efficient.*

4 Position Dominance

In this section, we consider a different notion of dominance arising when an alternative a is positioned higher in more rankings than another one, b . Formally:⁶

Definition 5. *Given an anonymous preference profile π on A , an alternative $a \in A$, $i \in N$ and $j \in \{1, \dots, |A|\}$, let $s_j(a) = |\{i : \sigma_i(a) \leq j\}|$. That is, $s_j(a)$ is the number of voters that rank alternative a at a position j or lower. For $a, b \in A$, we say that a position-dominates (Pos-dominates) b , denoted by $a \succ_{\pi}^{pos} b$, if $s_j(a) \geq s_j(b)$ for every $j \in \{1, \dots, |A|\}$.*

This binary relation is obviously not complete, since there may exist alternatives that are not comparable under a given profile. But Pos-dominance is transitive, ruling out the possibility of cycles.

The following example shows that neither one of the dominance relations presented above (PM and Pos) implies the other.

Example 1. *In the preference profile $\{a \succ_1 b \succ_1 c, a \succ_2 b \succ_2 c, b \succ_3 c \succ_3 a\}$ we have that $a \succ_{\pi}^{pm} b$ and $a \not\succeq_{\pi}^{pos} b$, while in the profile $\{a \succ_1 c \succ_1 d \succ_1 b, b \succ_2 a \succ_2 c \succ_2 d, c \succ_3 d \succ_3 b \succ_3 a\}$ we have $a \succ_{\pi}^{pos} b$ and $a \not\succeq_{\pi}^{pm} b$.*

The properties of Position Dominance-efficient (PosD-efficient) SCC and strongly Position Dominance-efficient (strongly PosD-efficient) SCC are analogous to those of PMD-efficient and strongly PMD-efficient SCC. We have that:⁷

⁶We adapt the definition introduced by Caragiannis et al. [8], allowing ties.

⁷The definitions of the voting rules as well as the proofs of the claims in this section can be found in the Appendix.

Proposition 3. *The Bucklin SCC is PosD-efficient but not strongly PosD-efficient.*

Moreover, there exists an entire family of PosD-efficient rules:

Proposition 4. *All positional scoring rules are PosD-efficient.*⁸

In turn, the strong PosD-efficiency of scoring rules depends on their associated weights:

Proposition 5. *Scoring rules with associated decreasing weights vectors are strongly PosD-efficient.*

5 Aggregation of Permutation Processes

So far, we introduced two different notions of dominance, rather limited in their efficiency. For richer results, we can turn our attention to the *aggregation* of permutation processes.

Definition 6. *An alternative $a \in X$ PM-dominates (respectively Pos-dominates) an alternative $b \in X$ in the permutation process Π , denoted $a \triangleright_{\Pi}^{pm} b$ ($a \triangleright_{\Pi}^{pos} b$), if for every finite set of alternatives $A \subseteq X$ such that $a, b \in A$, we have that a PM-dominates (Pos-dominates) b in the anonymous preference profile $\Pi(A)$.*

Definition 7. *A permutation process Π over X is said to be PM-compatible (Pos-compatible) if for every $A \subseteq X$, the binary relation $\triangleright_{\Pi}^{pm}|_A$ ($\triangleright_{\Pi}^{pos}|_A$) is a total preorder over A .*⁹

When a permutation process is PM-compatible, the total preorder is consistent across all the different finite subsets of X :

Lemma 1. *Let Π be a consistent permutation process that is PM-compatible. Then, for any finite subset of alternatives $A \subseteq X$, $(\triangleright_{\Pi(A)}^{pm}) = (\triangleright_{\Pi}^{pm}|_A)$*

Proof. First, by definition, we have that $a \triangleright_{\Pi}^{pm} b$ implies $a \triangleright_{\Pi(A)}^{pm} b$ for every subset $A \subseteq X$ such that $a, b \in A$. Then $(\triangleright_{\Pi}^{pm}|_A) \subseteq (\triangleright_{\Pi(A)}^{pm})$.

For the other inclusion, we must prove that $a \triangleright_{\Pi(A)}^{pm} b$ implies that $a \triangleright_{\Pi}^{pm} b$ for any finite subset C that contains a and b . Let $a \triangleright_{\Pi(A)}^{pm} b$ and suppose that

⁸If scores are normalized, positional scoring rules can be seen as a family of weighted arithmetic mean operators, one for each alternative [25, 20].

⁹A total preorder is a transitive and complete binary relation.

$a \not\triangleright_{\Pi(C)}^{pm} b$ for C . Since $\triangleright_{\Pi}^{pm}$ is PM-compatible, it is a total preorder for every C . Then it must be that $b \triangleright_{\Pi(C)}^{pm} a$. Then

$$|\{\sigma \in S_A : a >_{\sigma} b\}| \geq |\{\sigma \in S_A : b >_{\sigma} a\}| \quad (1)$$

and

$$|\{\sigma \in S_C : a >_{\sigma} b\}| < |\{\sigma \in S_C : b >_{\sigma} a\}| \quad (2)$$

Since the permutation process is consistent, when we restrict the domain to $\{a, b\}$, the relative positions of a and b do not change. So, (1) leads to $|\{\sigma \in S_{\{a,b\}} : a >_{\sigma} b\}| \geq |\{\sigma \in S_{\{a,b\}} : b >_{\sigma} a\}|$ and (2) leads to $|\{\sigma \in S_{\{a,b\}} : a >_{\sigma} b\}| < |\{\sigma \in S_{\{a,b\}} : b >_{\sigma} a\}|$. This is a contradiction, derived from the assumption that $a \not\triangleright_{\Pi(C)}^{pm} b$ for any finite subset C containing a and b . Then $a \triangleright_{\Pi(C)}^{pm} b$ for any C , and thus $a \triangleright_{\Pi}^{pm} b$ verifies that $(\triangleright_{\Pi(A)}) \subseteq (\triangleright_{\Pi|_A})$. \square

This result allows us to write Π instead of $\Pi(A)$ whenever we are referring to PM-dominance relations. The same is not true for a permutation process that is Pos-compatible. The dominance order of the alternatives may not be consistent between different subsets of X , as shown in the following example.

Example 2. Consider the set of alternatives $A = \{a, b, c\}$ and the profile of preferences where one voter chooses $a > b > c$ and another one $b > c > a$. Then we have that the only possible order is $b \triangleright_{\Pi(A)}^{pos} a \triangleright_{\Pi(A)}^{pos} c$. If we now consider $B = \{a, b\}$, we obtain that $b \triangleright_{\Pi(A)}^{pos} a$ and $a \triangleright_{\Pi(A)}^{pos} b$.

We can introduce a new definition of compatibility, according to which the orders are consistent over all the subsets.

Definition 8. A permutation process Π over X is said to be strongly compatible if it is compatible and the preorder \triangleright_{Π} is consistent across all subsets $A \subseteq X$.

It is clear from Lemma 1 that if a permutation process is PM-compatible it is also strongly PM-compatible.

The next result states that if the permutation process is compatible, then under an efficient voting rule one of the winning alternatives is the one that dominates the rest of the alternatives.

Theorem 1. Let f be an anonymous PMD-efficient SCC, and let Π be a PM-compatible permutation process. Then, for any finite subset of alternatives A , there exists an $a \in A$ such that $a \triangleright_{\Pi}^{pm} b$ for all $b \in A$. Moreover, $a \in f(\Pi(A))$.

Let f be an anonymous PosD-efficient SCC, and let Π be a Pos-compatible permutation process. Then, for any finite subset of alternatives A , there exists an $a \in A$ such that $a \succ_{\Pi(A)}^{\text{pos}} b$ for all $b \in A$. Moreover, $a \in f(\Pi(A))$.

If Π is strongly Pos-compatible, there exists an $a \in A$ such that $a \succ_{\Pi}^{\text{pos}} b$ for all $b \in A$. Moreover, $a \in f(\Pi(A))$

Proof. We prove the theorem for pairwise majority-dominance, since the proof is analogous for position-dominance.

Since Π is PM-compatible, the relation \succ_{Π}^{pm} restricted to A is a total pre-order. Therefore, there exists an alternative $a \in A$ such that $a \succ_{\Pi(A)}^{\text{pm}} b$ for all $b \in A$.

The definition of a SCC states that the outcome is non-empty. Assume that there is a $b \neq a$ such that $b \in f(\Pi(A))$. Since $a \succ_{\Pi(A)}^{\text{pm}} b$ and f is PM-efficient, it must be that $a \in f(\Pi(A))$. \square

In the light of Proposition 1, it may seem that Theorem 1 is not very relevant since there are not too many PM-efficient voting rules. But this is no longer the case when we consider PM-compatible permutation processes:

Theorem 2. *Let Π be a PM-compatible and consistent permutation process. Then the Black, Nanson, Dodgson, Young, Minimax, Kemeny, Fishburn and Schwartz rules select $f(\Pi(A)) = \{a \in A : a \succ_{\Pi}^{\text{pm}} b \text{ for all } b \in A\}$. Moreover, these rules are strongly PM-efficient.*

Proof. Since Π is PM-compatible, $\succ_{\Pi(A)}$ is a total preorder. Therefore there are no ‘cycles’ and by Theorem 1 we have that there exists at least one alternative $a \in A$ that PM-dominates the other alternatives. According to the results in Fishburn [16] and Felsenthal and Tideman [14], we know that Black, Dodgson, Young, Minimax and Fishburn under the absence of cycles are Strict Condorcet Consistent. That is, they select the set of candidates that beat or tie with all other alternatives.¹⁰ It is also clear that the Schwartz rule is Weak Condorcet Consistent. Therefore we have that $f(\Pi(A)) = \{a \in A : a \succ_{\Pi}^{\text{pm}} b \text{ for all } b \in A\}$ for these rules. It is straightforward from the definition of Weak Condorcet winners that these rules are strongly PM-efficient. \square

For position dominance, we obtain an analogous result, but under the proviso that the order of alternatives may not be consistent among subsets of X :

¹⁰The undefeated alternatives are called in the literature as Weak Condorcet winners or Quasi-Condorcet candidates.

Theorem 3. *Let Π be a Pos-compatible and consistent permutation process, f a scoring rule with an associated decreasing weights vector and A a finite set of alternatives. Then $f(\Pi(A)) = \{a \in A : a \succ_{\Pi(A)}^{pos} b \text{ for all } b \in A\}$.*

Proof. According to Theorem 1 there exists an alternative $a \in A$ such that $a \succ_{\Pi(A)}^{pos} b$ for all $b \in A$ and $a \in f(\Pi(A))$. Since Π is Pos-compatible, any other alternative $b \in A$ is comparable with a . By Proposition 5, all the scoring rules with associated decreasing weights are strongly PD-efficient. If $b \succ_{\Pi(A)}^{pos} a$, it must be that $b \in f(\Pi(A))$, and $b \succ_{\Pi(A)}^{pos} c$ for all $c \in A$ (because of transitivity). If $b \not\succ_{\Pi(A)}^{pos} a$, then $b \notin f(\Pi(A))$. \square

As a consequence of Theorem 3, we get a positive result with strong compatibility, since the orders are consistent across different subsets containing the same alternatives.

Corollary 1. *Let Π be a strongly Pos-compatible and consistent permutation process and f a scoring rule with an associated decreasing weighted vector. Then for any finite set of alternatives $A \subseteq X$, $f(\Pi(A)) = \{a \in A : a \succ_{\Pi}^{pos} b \text{ for all } b \in A\}$.*

Theorems 2, 3 and Corollary 1 are particular cases of a stronger result:

Theorem 4. *Let Π be a strongly compatible and consistent permutation process and f a strongly efficient SCC. Then for any finite set of alternatives $A \subseteq X$, $f(\Pi(A)) = \{a \in A : a \succ_{\Pi} b \text{ for all } b \in A\}$.¹¹*

Proof. Let A be a finite set of alternatives. By Theorem 1 and the fact that strong efficiency implies efficiency, we have that $\{a \in A : a \succ_{\Pi} b \text{ for all } b \in A\} \subseteq f(\Pi(A))$.

Now let $a \in f(\Pi(A))$ and assume that there exists $b \in A$ such that $a \not\succ_{\Pi} b$. Since Π is strongly compatible, $\Pi(A)$ has the same total preorder, for every $A \subseteq X$. Then $b \succ_{\Pi(A)} a$ and $a \not\succ_{\Pi(A)} b$. Since f is strongly efficient, it must be that $a \notin f(\Pi(A))$, a contradiction. Then $f(\Pi(A)) \subseteq \{a \in A : a \succ_{\Pi} b \text{ for all } b \in A\}$. \square

Then, if the permutation process is compatible, an efficient social choice correspondence should select the alternatives that beat or tie up with every other alternative. The following example, due to Caragiannis et al. [8] shows that even when the permutation process is compatible, the ‘best’ alternative may not coincide under the two dominance relations.

¹¹Whenever the type of dominance relation is not explicitly specified, the result must be valid for both PM and Pos dominance.

Example 3. Consider Π , the consistent permutation process which given the alternatives a, b and c , gives the following profile: $(a >_1 b >_1 c)$, with weight $\frac{4}{11}$; $(b >_2 a >_2 c)$, with weight $\frac{2}{11}$; $(b >_3 c >_3 a)$ with weight $\frac{3}{11}$ and $(c >_4 a >_4 b)$ weighted $\frac{2}{11}$. Then we have that $a \triangleright_{\Pi}^{pm} b \triangleright_{\Pi}^{pm} c$ and $b \triangleright_{\Pi}^{pos} a \triangleright_{\Pi}^{pos} c$.

While the two dominance relations may not yield the same outcome, we seek to find out when a permutation process is compatible. First, let us recall that a permutation process can be interpreted in terms of utilities, which allows us to introduce the following definition.

Definition 9. Alternative $a \in X$ dominates $b \in X$ in the utility process U if for every finite subset of alternatives containing $x_1 = a$ and $x_2 = b$, $\{x_1, x_2, \dots, x_m\} \subseteq X$, and every vector of utilities (u_1, u_2, \dots, u_m) with $u_1 \geq u_2$ we have that

$$p(u_1, u_2, \dots, u_m) \geq p(u_2, u_1, \dots, u_m)$$

where p is the probability function.

The following two lemmas state that ‘natural’ utility processes are indeed compatible.

Lemma 2. Let Π be a consistent permutation process and U be its corresponding utility process. If alternative a dominates b in U , then $a \triangleright_{\Pi} b$.

Proof. Noothigattu et al [30] prove in Lemma 4.9 that if a dominates b in U , then a swap-dominates b . Swap dominance is a stronger notion that implies PM-dominance and Pos-dominance. Then both conditions are satisfied. \square

We consider consistent permutation processes U such that given a set of alternatives $\{x_1, \dots, x_m\}$, the utilities $(U_{x_1}, \dots, U_{x_m})$ are independent and have a distribution drawn from the Exponential Family [28]:

$$p_{U_a}(u_1) = \exp(\eta(\mu_a)T(u_1) - A(\mu_a) + B(u_1))$$

Examples of this are the Gaussian or Normal (Thurstone-Mosteller Process [39, 29]), the Gumbel (Plackett-Luce Process [33, 23]), Poisson, Gamma, Binomial and Negative Binomial distributions. Then:

Lemma 3. Under an utility process with a distribution belonging to the Exponential Family, the alternative a dominates alternative b if $\eta(\mu_a) > \eta(\mu_b)$ and $T(u_1) > T(u_2)$.¹²

¹²For the Normal ($N(\mu_{x_i}, \frac{1}{2})$) and the Gumbel ($G(\mu_{x_i}, \gamma)$) processes this means that $\mu_a \geq \mu_b$. For a Poisson process ($P(\lambda_{x_i})$), it means that $\lambda_a \geq \lambda_b$. For a Gamma process with fixed shape ($\Gamma(r, \lambda_{x_i})$), the implication is that $\lambda_b \geq \lambda_a$. For a Binomial process ($B(n_{x_i}, p_{x_i})$) with fixed n , it means that $p_a \geq p_b$. If p is fixed and n is variable, it corresponds to $n_a \geq n_b$. For a Negative Binomial process ($NB(r_{x_i}, p_{x_i})$) with fixed r , it means that $p_b \geq p_a$ while if p is fixed and r is variable, it is $r_a \geq r_b$.

Proof. We have to find the conditions according to which $p(u_1, u_2, \dots, u_m) \geq p(u_2, u_1, \dots, u_m)$ when $u_1 \geq u_2$. Since utilities are sampled independently, this implies checking when $p_{U_a}(u_1)p_{U_b}(u_2) \geq p_{U_a}(u_2)p_{U_b}(u_1)$. That is,

$$\begin{aligned} & \exp(\eta(\mu_a)T(u_1) - A(\mu_a) + B(u_1)) \exp(\eta(\mu_b)T(u_2) - A(\mu_b) + B(u_2)) \geq \\ & \exp(\eta(\mu_a)T(u_2) - A(\mu_a) + B(u_2)) \exp(\eta(\mu_b)T(u_1) - A(\mu_b) + B(u_1)) \end{aligned}$$

and thus

$$\begin{aligned} & \eta(\mu_a)T(u_1) - A(\mu_a) + B(u_1) + \eta(\mu_b)T(u_2) - A(\mu_b) + B(u_2) \geq \\ & \eta(\mu_a)T(u_2) - A(\mu_a) + B(u_2) + \eta(\mu_b)T(u_1) - A(\mu_b) + B(u_1) \end{aligned}$$

Then

$$(\eta(\mu_a) - \eta(\mu_b))T(u_1) + (\eta(\mu_b) - \eta(\mu_a))T(u_2) \geq 0$$

and

$$(\eta(\mu_a) - \eta(\mu_b))(T(u_1) - T(u_2)) \geq 0$$

Since $u_1 \geq u_2$, it follows that $T(u_1) \geq T(u_2)$ and thus $\eta(\mu_a) - \eta(\mu_b) \geq 0$ and $\eta(\mu_a) \geq \eta(\mu_b)$.

The proof for each particular probability distribution is analogous to the proof of the general case. \square

Lemma 3 implies that a consistent permutation process with an utility process drawn from the Exponential Family (which includes some of the best known distributions), is compatible under both notions of dominance. This ensues from the fact that dominance only depends on the parameters of the distributions, all of which have real values (and thus are element of a complete preorder). Moreover, Lemma 2 states that the preorders are the same under both dominance notions satisfying strong compatibility. This allows us to get rid of a fixed A even for Pos-dominance. The downside of this, is that we may rule out permutation processes that could be compatible under one dominance notion but not under the other.

According to Theorem 2, Corollary 1 and Lemma 3, in many cases of interest we only need to find the most dominant alternatives, which, depending on the permutation process, are the alternatives with a maximum defined parameter.

6 Stability

In this section we introduce another property that we expect that a SCC should verify.

Definition 10. *Let f be an anonymous SCC and Π a permutation process over X . We say that f is stable if for any non empty and finite subset of alternatives A and B such that $B \subseteq A \subseteq X$, $f(\Pi(A)) \cap B = f(\Pi(B))$ whenever $f(\Pi(A)) \cap B \neq \emptyset$.*

Stability is a highly relevant property of *aggregation operators* meaning that, in rough terms, ‘small input errors’ do not lead to ‘big output errors’ [42, 7, 26, 35]. In our setting this means that voting rules must yield consistent choices on restricted sets of alternatives. For example, we do not want the decision of an autonomous vehicle with three passengers to be to go straight in the case that no pedestrians are in front of it, but swerve and crush a wall in the same situation when it transports only two passengers.

The stability of a SCC f depends on how both f and the permutation process are defined.

Theorem 5. *Let f be a strongly efficient SCC and let Π be a strongly compatible permutation process. Then f is stable.*

Proof. Let $B \subseteq A \subseteq X$ and assume that $f(\Pi(A)) \cap B \neq \emptyset$. Let $a \in f(\Pi(A)) \cap B$. By Theorem 4 we have that $a \triangleright_{\Pi} b$ for all $b \in A$, and thus, for all $b \in B$. Then $a \in f(\Pi(B))$.

Now consider $a \in f(\Pi(B))$. Then $a \in B$ and $a \triangleright_{\Pi} b$ for all $b \in B$. Suppose that $a \notin f(\Pi(A))$. Then there is a $c \in A$ such that $a \not\triangleright_{\Pi} c$. Since Π is strongly compatible, we have that $c \triangleright_{\Pi(A)} a$. Since we assume that $f(\Pi(A)) \cap B \neq \emptyset$, there is an alternative $d \in f(\Pi(A)) \cap B$. Then $d \triangleright_{\Pi(A)} c$ and by transitivity we have that $a \not\triangleright_{\Pi(A)} d$. Since $d \in B$ and Π is strongly compatible, we have that $a \not\triangleright_{\Pi(B)} d$. But f is strongly efficient, so we have that $a \notin f(\Pi(B))$, which is a contradiction. \square

This is a very encouraging result, since according to Proposition 5 and Theorem 2 there exist many well known voting rules that are strongly efficient. Thus, these rules must also verify stability.

7 Conclusions

We have presented an axiomatic study of the method developed by Noothigattu et al. [30] for automating ethical AI decision making under different

notions of dominance. According to one of those notions, an alternative is better than another if a majority prefers it. According to the other notion, an alternative is dominant if it is better positioned than the other alternatives in most rankings. At first sight, the dominance relations here introduced, seem more intuitive and natural than the dominance relation used in [30], namely swap-dominance. It seems that PM and Pos dominance have clearer meanings and are better suited for the design of an autonomous system.

We showed that, depending on how we learn the preferences of the voters, there are many well known voting rules that behave well, in the sense that they select the most dominating alternatives. Moreover, it is only necessary to find the alternative with a maximum defined parameter (depending on the distribution used to learn the model). When we use ‘natural’ distributions for the permutation process we do not need to distinguish between dominance notions. Another important property is stability, which amounts to the consistence of choices across different sets of alternatives. For example, an autonomous vehicle is likely to face potentially infinite scenarios, and we want decisions to be similar in similar scenarios. We found again that are many voting rules verifying this property.

Our work yields stronger results than those in [30]. While the alternatives selected are the same using theirs and our dominance notions, we obtain them under voting rules satisfying different sets of properties. This is useful since this unifies the arguments that can be used to justify the choice of a given alternative.

We consider this paper as a contribution to the axiomatization of theory of the aggregation of permutations. Its full development in this light may help to overcome classical problems of aggregation theory.

A possible future line of research consists in analyzing the strategic aspects of the aggregation of permutation processes. It is important to guarantee the non-manipulability of this decision making process. For example, there should not exist advantages from lying about preferences (this property is known in the literature as strategy-proofness [18, 37]) or from voting several times (known in the literature as false name-proofness [10, 15]). The results on manipulation avoidance are rather negative, since there exist very few rules satisfying non-manipulability, when considering unrestricted domains. We believe that in the case of the aggregation of permutation processes, new rules can be found overcoming these problems.

Acknowledgements

The authors are especially grateful to Ariel Procaccia and Felix Brandt for comments and suggestions that helped to improve the paper.

Conflict of interest

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

References

- [1] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiuc-Pietro, and Vasileios Lamos. Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science*, 2:e93, 2016.
- [2] Hossein Azari Soufiani, Hansheng Diao, Zhenyu Lai, and David C Parkes. Generalized random utility models with multiple types. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
- [3] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1):3–44, 2021.
- [4] Edvard P Bjørgen, Simen Madsen, Therese S Bjørknes, Fredrik V Heimsæter, Robin Håvik, Morten Linderud, Per-Niklas Longberg, Louise A Dennis, and Marija Slavkovic. Cake, death, and trolleys: dilemmas as benchmarks of ethical decision-making. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 23–29, 2018.
- [5] Felix Brandt, Vincent Conitzer, and Ulle Endriss. Computational social choice. *Multiagent Systems*, pages 213–283, 2012.
- [6] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

- [7] Tomasa Calvo, Gaspar Mayor, and Radko Mesiar. *Aggregation operators: New trends and applications*, volume 97. Springer Science & Business Media, 2002.
- [8] Ioannis Caragiannis, Ariel D Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? *ACM Transactions on Economics and Computation (TEAC)*, 4(3):1–30, 2016.
- [9] François Caron and Yee Whye Teh. Bayesian nonparametric models for ranked data. *arXiv preprint arXiv:1211.4321*, 2012.
- [10] Vincent Conitzer. Anonymity-proof voting rules. In *Proceeding of WINE*, pages 295–306, 2008.
- [11] Vincent Conitzer, Walter Sinnott-Armstrong, Jana Schaich Borg, Yuan Deng, and Max Kramer. Moral decision making frameworks for artificial intelligence. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [12] Louise Abigail Dennis, Michael Fisher, and Alan Winfield. Towards verifiably ethical robot behaviour. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [13] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226, 2012.
- [14] Dan S Felsenthal and Nicolaus Tideman. Weak condorcet winner (s) revisited. *Public Choice*, 160(3-4):313–326, 2014.
- [15] Federico Fioravanti and Jordi Massó. False-name-proof and strategy-proof voting rules under separable preferences. *Available at SSRN 4175113*, 2022.
- [16] Peter C Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.
- [17] Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John P Dickerson, and Vincent Conitzer. Adapting a kidney exchange algorithm to align with human values. *Artificial Intelligence*, 283:103261, 2020.
- [18] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41, 587–601, 1973.

- [19] Daniel Gómez, Karina Rojas, Javier Montero, Juan Tinguaro Rodríguez, and Gleb Beliakov. Consistency and stability in aggregation operators: An application to missing data problems. *International Journal of Computational Intelligence Systems*, 7(3):595–604, 2014.
- [20] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. Aggregation functions: Means. *Information Sciences*, 181(1):1–22, 2011.
- [21] Joshua Greene, Francesca Rossi, John Tasioulas, Kristen Brent Venable, and Brian Williams. Embedding ethical principles in collective decision support systems. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [22] Joel L Horowitz, Denis Bolduc, Suresh Divakar, John Geweke, Füsün Gönül, Vassilis Hajivassiliou, Frank S Koppelman, Michael Keane, Rosa Matzkin, Peter Rossi, et al. Advances in random utility models report of the workshop on advances in random utility models duke invitational symposium on choice modeling behavior. *Marketing Letters*, 5(4):311–322, 1994.
- [23] R Duncan Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.
- [24] John I Marden. *Analyzing and modeling rank data*. Chapman and Hall/CRC, 2019.
- [25] Jean-Luc Marichal, Pierre Mathonet, and Eric Tousset. Characterization of some aggregation functions stable for positive linear transformations. *Fuzzy Sets and Systems*, 102(2):293–314, 1999.
- [26] Javier Montero, Victoria López, and Daniel Gómez. The role of fuzziness in decision making. In *Fuzzy Logic*, pages 337–349. Springer, 2007.
- [27] James H Moor. The mature, importance, and difficulty of machine ethics. In *Machine Ethics and Robot Ethics*, pages 233–236. Routledge, 2020.
- [28] Carl N Morris. Natural exponential families with quadratic variance functions. *The Annals of Statistics*, pages 65–80, 1982.
- [29] Frederick Mosteller. Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. In *Selected Papers of Frederick Mosteller*, pages 157–162. Springer, 2006.

- [30] Ritesh Noothigattu, Snehal Kumar Gaikwad, Edmond Awad, Sohan Dsouza, Iyad Rahwan, Pradeep Ravikumar, and Ariel Procaccia. A voting-based system for ethical decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [31] David C Parkes, Houssein Azari Soufiani, and Lirong Xia. Random utility theory for social choice. In *Proceedings of the 25th Annual Conference on Neural Information Processing Systems*. Curran Associates, Inc., 2012.
- [32] Samuele Lo Piano. Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, 7(1):1–7, 2020.
- [33] Robin L Plackett. The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(2):193–202, 1975.
- [34] Adarsh Prasad, Harsh Pareek, and Pradeep Ravikumar. Distributional rank aggregation, and an axiomatic analysis. In *International Conference on Machine Learning*, pages 2104–2112. PMLR, 2015.
- [35] Karina Rojas, Daniel Gomez, Javier Montero, and J Tinguaro Rodríguez. Strictly stable families of aggregation operators. *Fuzzy Sets and Systems*, 228:44–63, 2013.
- [36] Francesca Rossi. Moral preferences. In *The 10th Workshop on Advances in Preference Handling (MPREF)*, 2016.
- [37] Mark Allen Satterthwaite. Strategy-proofness and Arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10, 187–217, 1975.
- [38] Thomas Schwartz. Rationality and the myth of the maximum. *Noûs*, pages 97–117, 1972.
- [39] Louis L Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273, 1927.
- [40] Wendell Wallach and Colin Allen. *Moral machines: Teaching robots right from wrong*. Oxford University Press, 2008.
- [41] Alan FT Winfield, Christian Blum, and Wenguo Liu. Towards an ethical robot: internal models, consequences and ethical action selection. In *Advances in Autonomous Robotics Systems*, pages 85–96. Springer, 2014.

- [42] Ronald R Yager and Alexander Rybalov. Noncommutative self-identity aggregation. *Fuzzy Sets and Systems*, 85(1):73–82, 1997.

A Voting rules

Black’s rule:¹³ this rule is applied in two stages. In the first stage, the candidates that beat or tie with every other candidate are selected. If there are not such winners, each candidate receives a score depending on the positions on every ranking (the better the position the higher the score). The alternative with the largest overall score is elected.

Dodgson’s rule: a Condorcet winner is a candidate that beats every other alternative. The rule computes the minimum number of times that it is necessary to swap two adjacent alternatives on some rankings in order to make each candidate a Condorcet winner. The alternative that requires the minimum steps is considered the winner.

Young’s rule: this rule deletes candidates in order to make an alternative the Condorcet winner. The candidate that requires less deletions is declared the winner.

Kemeny’s rule: this rule selects the most preferred alternative of the rankings that minimize the number of pairs of candidates that are ranked opposite by all the voters.

Nanson’s rule: as in Black’s rule, a score is given to the candidates according to their position. Every candidate with a score below the average score is deleted. The process is repeated with the candidates left, and so on. The rule selects the undeleted candidates.

Minimax rule: the candidate whose maximum losses in the paired comparisons are the least is declared winner.

Fishburn’s rule: this rule checks out whether everything that beats x also beats y under simple majority, and if x beats or ties something that beats y , then x is ‘better than’ y under simple majority. The rule selects the best alternatives under this notion.

¹³The Black and Dodgson rules described are called Revised Black and Simplified Dodgson rules in [14].

Bucklin's rule: the Bucklin score of an alternative a , $B(a)$, is the minimum k such that a is among the first k positions in the majority of input votes. The Bucklin rule selects the alternatives with the least Bucklin scores.

Scoring rules: these rules have an associated weights vector $(\alpha_1, \alpha_2, \dots, \alpha_{|A|})$, where $\alpha_i \geq \alpha_{i+1}$ for $i = 1, \dots, |A| - 1$. An alternative a in a profile σ_j gets α_i points if $\sigma_j(a) = i$. The score of an alternative is the sum of all the points across all voters. The candidate with the most overall points is declared the winner. The Plurality and the Antiplurality rules are examples of scoring rules with associated weights vector $(1, 0, \dots, 0)$ and $(1, \dots, 1, 0)$ respectively. We say that a scoring rule has an associated decreasing weights vector if $\alpha_i > \alpha_{i+1}$ for $i = 1, \dots, |A| - 1$.

B Proofs

Proof. Proposition 1

Consider the profile $\sigma = (a >_1 b >_1 c >_1 d >_1 e, e >_2 d >_2 a >_2 c >_2 b, b >_3 c >_3 d >_3 e >_3 a)$. We have that $a \triangleright_{\pi}^{pm} b$.

Black's rule: there is no alternative that beats or tie every other alternative. Then we use a scoring rule with an associated decreasing weighted vector, for example $(5, 4, 3, 2, 1)$. We obtain that $score(a) = 9$ and $score(b) = 10$, with b getting the largest score. Thus we have that $b \in Black(\pi)$ and $a \notin Black(\pi)$.

Dodgson's rule: alternative b needs 1 swap in the first vote to become the Condorcet winner, while all the other alternatives require 2 or more. Then $b \in Dodgson(\pi)$ and $a \notin Dodgson(\pi)$.

Young's rule: alternative b needs alternative a to be eliminated in order to become the Condorcet winner, while all the other alternatives need 2 or more alternatives to be eliminated. Then $b \in Young(\pi)$ and $a \notin Young(\pi)$.

Kemeny's rule: the ranking that minimizes the number of candidates that are ranked opposite by all the voters is $b > c > d > a$. Then $b \in Kemeny(\pi)$ and $a \notin Kemeny(\pi)$.

Consider the profile $\sigma = (e >_1 a >_1 b >_1 c >_1 d, b >_2 c >_2 d >_2 e >_2 a)$. We have that $a \triangleright_{\pi}^{pm} b$.

Nanson's rule: we use a scoring rule with an associated decreasing weighted vector, for example $(5, 4, 3, 2, 1)$. During the first round, a and d are eliminated. In the second round c is eliminated. Finally we have that b and e are selected. Then $b \in Nanson(\pi)$ and $a \notin Nanson(\pi)$.

Consider the profile $\sigma = (d >_1 c >_1 a >_1 b, b >_2 c >_2 d >_2 a, d >_3 c >_3 a >_3 b, a >_4 b >_4 c >_4 d, b >_5 c >_5 d >_5 a)$. We have that $a \triangleright_{\pi}^{pm} b$.

Minimax rule: alternative b , c and d have 2 as a maximum loss while a has 3. Then $b \in Minimax(\pi)$ and $a \notin Minimax(\pi)$.

Consider the profile $\sigma = (a >_1 b >_1 c >_1 d, b >_2 c >_2 d >_2 a)$. We have that $a \triangleright_{\pi}^{pm} b$.

Fishburn's rule: in this profile b beats c and d , ties with a and beat all the alternative that beat or tie with a . Then $b \in Fishburn(\pi)$ and $a \notin Fishburn(\pi)$. \square

Proof. Proposition 3

Let $a \triangleright_{\pi}^{pd} b$ and $b \in Bucklin(\pi)$. Let $B(b) = t$, that is, $s_t(b) = \alpha \geq \lceil \frac{n+1}{2} \rceil$. Because of position dominance, we have that $s_t(a) \geq s_t(b) = \alpha$. Then, $B(a) \leq t$. So $a \in Bucklin(\pi)$.

The following example shows that it is not strong-pd-eff. Let π be the preference profile such that $\sigma = (a >_1 b >_1 c, a >_2 b >_2 c, b >_3 a >_3 c, c >_4 a >_4 b)$. Then $B(a) = B(b) = 2$, $B(c) = 3$, $s_1(a) = 2$, $s_1(b) = 2$, $s_2(a) = 4$ and $s_2(b) = 3$. So, we have that $a \triangleright_{\pi}^{pd} b$ and $b \not\triangleright_{\pi}^{pd} a$ but $Bucklin(\pi) = \{a, b\}$. \square

Proof. Proposition 4

For ease of demonstration we assume, without loss of generality, that $A = \{a, b, c\}$. The score of alternative a is given by $Score(a) = (s_1(a), s_2(a) - s_1(b), 3 - s_2(a)) \cdot (\alpha_1, \alpha_2, \alpha_3) = (\alpha_1 - \alpha_2)s_1(a) + (\alpha_2 - \alpha_3)s_2(a) + 3\alpha_3$ with $\alpha_i - \alpha_{i+1} \geq 0$ for $i = 1, 2$. From the fact that $a \triangleright_{\pi}^{pd} b$ we have that $s_j(a) \geq s_j(b)$ for $j = 1, 2$. Then $Score(a) \geq Score(b)$. If b is such that $b \in f(\pi)$, then $a \in f(\pi)$ as it has more or equal points than b . \square

Proof. Proposition 5

We assume again that $A = \{a, b, c\}$. And consider the same argument as in the proof of Proposition 4. The main difference is that $\alpha_i - \alpha_{i+1} > 0$ for

$i = 1, 2$. Then if $a \triangleright_{\pi}^{pd} b$ and $b \not\triangleright_{\pi}^{pd} a$, we have that $Score(a) > Score(b)$ and then $b \notin f(\pi)$. If instead $b \triangleright_{\pi}^{pd} a$, then $Score(a) = Score(b)$ and both alternatives are selected if they have the most overall points. \square