Proceedings of the
1st Workshop on Artificial Intelligence and Formal Verification, Logics, Automata and Synthesis (OVERLAY),
Rende, Italy, November 19–20, 2019

Preference Theories on Weak Orders

Marco Faella[1] and Luigi Sauro[2]

[1]m.faella@unina.it
[2]luigi.sauro@unina.it
[1,2]University of Naples Federico II, Italy

Abstract

We study the rational preferences of an agent participating to a mechanism whose outcome is a weak order among participants. We propose a set of self-interest axioms and characterize the resulting preference theories.

1 Introduction

In many applications, a series of measurements are designed in order to obtain a ranking among a set of agents. Tournaments [3], public/recruiting competitions, elections [2], or partitioning students in groups with homogeneous level of ability [5] can all be considered in this way. However, any measurement can in principle fail to distinguish two items [1, 8, 6] and hence any ranking mechanism may end up with one or more ties (formally, it returns a weak order). When a linear ranking is absolutely necessary, various tie-breaking techniques are added to the measurement, either by randomization [4] or by resorting to some fixed order between agents [7].

In this paper, we assume that an arbitrary mechanism produces a weak order among a set of agents, and we study the preferences that those agents may exhibit over the outcomes of that mechanism. More specifically, we introduce six basic conditions reflecting different aspects of the rational behaviour of an agent and study the theories deriving from choosing a generic subset of those conditions. We show that only a relatively small number of theories are distinct and we provide a hierarchical characterization thereof. Understanding the types of preferences that a rational agent can hold is necessary to study the strategic reasoning that they apply to the mechanism, and ultimately predict or control their behaviour. This provides the theoretical underpinnings for avoiding the often arbitrary tie-breaking phase or for minimizing in sporting tournaments the occurrence of matches that are competitively irrelevant for one or both of the involved players.

This preliminary study paves to way to many further directions. For example, we aim at investigating the connection between this preference theory and mechanisms producing a numerical ranking or distributing monetary rewards.

2 Definitions

Recall the following definitions: a *pre-order* is a reflexive and transitive binary relation, a *weak order* is a total pre-order, and a *linear order* is an antisymmetric weak order. Given a pre-order $\sqsubseteq$, we denote its asymmetric part by $\sqsubset$ (that is, $a \sqsubset b$ iff $a \sqsubseteq b$ and $b \not\sqsubseteq a$), and its symmetric complement by $\lessgtr$ (that is, $a \lessgtr b$ iff $a \not\sqsubseteq b$ and $b \not\sqsubseteq a$).

We assume that an undescribed process, such as a competition or a vote, produces a weak order on a set of agents A. We want to study the ways in which the agents compare different outcomes.

Given a finite set of agents A, let $\mathcal{W}(A)$ denote the set of all weak orders on A. For an agent $a \in A$, a *preference relation for a* (in short, an a-preference) is a weak order P on $\mathcal{W}(A)$. In Section 2.2 we will qualify which preference relations are rational for an agent by introducing suitable *self-interest* axioms.

We say that a pre-order $\sqsubseteq_2$ *perfects* another pre-order $\sqsubseteq_1$ if $a \sqsubset_1 b$ implies $a \sqsubset_2 b$, and $a \lessgtr_2 b$ implies $a \lessgtr_1 b$, for all agents a and b. Intuitively, $\sqsubseteq_2$ exhibits at least the same strong preferences of $\sqsubseteq_1$, but it may distinguish agents that are equivalent for $\sqsubseteq_1$, as well as distinguishing or equating agents that are incomparable for $\sqsubseteq_1$. This notion is incomparable to the classical notion of refinement (that is, inclusion) between relations. Perfectioning is itself a pre-order, whose bottom element is the identity relation and whose top elements are the linear orders.

2.1 A Catalog of Preferences

This section presents a selection of preferences, showing that there is a wide variety of reasonable ways to compare two weak rankings from the point of view of one of the participants.

For a weak order $\sqsubseteq \in \mathcal{W}(A)$ and an agent a, let $below_a(\sqsubseteq)$, $same_a(\sqsubseteq)$, $above_a(\sqsubseteq)$ be the partition of A into the agents that are strictly below, equivalent, and strictly above a, respectively. Moreover, let $level_a^\top(\sqsubseteq)$ be the length of the longest $\sqsubset$-chain starting from a, which can be recursively defined as follows:

$$level_a^\top(\sqsubseteq) = \begin{cases} 1 & \text{if } above_a(\sqsubseteq) = \emptyset \\ \max_{b \in above_a(\sqsubseteq)} level_b^\top(\sqsubseteq) + 1 & \text{otherwise.} \end{cases}$$

One can dually define $level_a^\bot(\sqsubseteq)$ as the length of the longest $\sqsubset$-chain *ending* in a. Informally, this measures the level of a, starting from the bottom.

Next, we are going to present a list of interesting preference relations for a, using a uniform naming system, based on the following abbreviations:

Abbreviation	Meaning
$\top$	$level_a^\top(\sqsubseteq)$
$\bot$	$-level_a^\bot(\sqsubseteq)$
a	$\lvert above_a(\sqsubseteq)\rvert$
s	$\lvert same_a(\sqsubseteq)\rvert$
b	$\lvert below_a(\sqsubseteq)\rvert$

We use the above abbreviations as a superscript to indicate that a preference tries to minimise the corresponding quantity. For example, P^{a} is the preference that minimises $|above_a(\sqsubseteq)|$, i.e., $P^{\mathrm{a}}(\sqsubseteq_1, \sqsubseteq_2)$ holds iff $|above_a(\sqsubseteq_1)| \geq |above_a(\sqsubseteq_2)|$. We also write $P^{x,y}$ for the preference that tries to minimise first quantity x and then quantity y, lexicographically. Finally, we allow simple arithmetic expressions, like $P^{\mathrm{a+s}}$ for the preference that minimises the sum of $|above_a(\sqsubseteq)|$ and $|same_a(\sqsubseteq)|$.

- Level-based. $P^\top$ prefers to minimise the level of a in the weak order, i.e., $level_a^\top(\sqsubseteq)$. In particular, it is insensitive to ties.

- Level-based top-k. For a positive integer k, $P_k^\top$ prefers a to be within the first k levels in the weak order. This preference only distinguishes two classes of weak orders: those where a sits in one of the top k levels, and all the others. Any element of the first class is (strongly) preferred to any element of the second class.

- Lexicographic. $P^{\mathrm{a,s}}$ prefers having fewer agents above a; equal that, it prefers to have fewer agents tied with a. It is equivalent to minimising the vector $\big(|above_a(\sqsubseteq)|, |same_a(\sqsubseteq)|\big)$, lexicographically.

- Best linearization. P^{bst} judges a weak order the same as its linear extension where a has the best position. The canonical name for this preference is P^{a}, because it is equivalent to minimising $|above_a(\sqsubseteq)|$.
- Worst linearization. P^{wst} is the dual of P^{bst}. The canonical name for this preference is $P^{\mathrm{a+s}}$, because it is equivalent to minimising $|above_a(\sqsubseteq) \cup same_a(\sqsubseteq)|$.

The preference $P^{\top}$ may reflect a conscientious participant when a group of students is being partitioned into different levels of ability. Similarly, the preference $P_k^{\top}$ can be adopted by the candidate of a competition based on a public threshold k such that all candidates with the k highest marks are recruited. In [3] the preference $P^{\mathrm{a,s}}$ has been used to model the fact that a team participating in a tournament aims at prevailing over the opponents. Notice that $P^{\mathrm{a,s}}$ perfects P^{bst}, and $P^{\top,\mathrm{s}}$ perfects $P^{\top}$. Preferences P^{bst} and P^{wst} can be adopted whenever an unknown rule is used to resolve ties. In particular, they reflect an optimistic or pessimistic attitude, respectively.

2.2 Self-Interest Axioms

Given two weak orders $\sqsubseteq_1, \sqsubseteq_2 \in \mathcal{W}(A)$ and an agent $a \in A$, define the following properties:

Same-context. The restrictions of $\sqsubseteq_1$ and $\sqsubseteq_2$ to $(A \setminus \{a\})^2$ coincide. If we need to emphasize the arguments, this property can also be denoted by the extended notation Same-context$(a, \sqsubseteq_1, \sqsubseteq_2)$.

Dominance. For all $b \in A$, if $b \sqsubset_1 a$ then $b \sqsubset_2 a$, and if $b \equiv_1 a$ then $b \sqsubseteq_2 a$. Extended notation: Dominance$(a, \sqsubseteq_1, \sqsubseteq_2)$.

Improvement. There exists $b \in A$ such that either $a \equiv_1 b$ and $b \sqsubset_2 a$, or $a \sqsubset_1 b$ and $b \sqsubseteq_2 a$. Extended notation: Improvement$(a, \sqsubseteq_1, \sqsubseteq_2)$.

Swap. There exists $b \in A$ such that $a \sqsubset_1 b$ and $\sqsubseteq_2$ is obtained from $\sqsubseteq_1$ by switching a and b.

The properties above allow us to describe three notions of a basic improvement for agent a:

- a trades place with an agent that is higher up in the order (Swap);
- a moves up in the order, and all other agents stay put (Same-context and Improvement);
- a moves up in the order, and all other agents do not cross the position of a (Dominance and Improvement).

In any of these three cases, one may think that a rational agent should prefer the new situation to the old one, at least weakly. In fact, in the following we show that there are reasonable scenarios where some of the above are undesired.

First, we observe that the Dominance property has a close relationship with the idea of splitting the agents into three tiers (*below*, *same*, and *above*), relative to a specified agent a. Say that two weak orders $\sqsubseteq_1, \sqsubseteq_2$ are *3-tier equivalent for* a if $below_a(\sqsubseteq_1) = below_a(\sqsubseteq_2)$, $same_a(\sqsubseteq_1) = same_a(\sqsubseteq_2)$, and $above_a(\sqsubseteq_1) = above_a(\sqsubseteq_2)$. The following equivalence states that two weak orders are equivalent according to Dominance iff they are 3-tier equivalent. The proofs of this result and of all subsequent ones are omitted due to space limitations.

Lemma 1. *For all agents a and weak orders $\sqsubseteq_1, \sqsubseteq_2 \in \mathcal{W}(A)$, the following two properties are equivalent:*

1. *Dominance$(a, \sqsubseteq_1, \sqsubseteq_2)$ and Dominance$(a, \sqsubseteq_2, \sqsubseteq_1)$;*
2. *$\sqsubseteq_1$ and $\sqsubseteq_2$ are 3-tier equivalent for a.*

We now introduce six *self-interest axioms* on a given preference relation P, based on the above three notions of basic improvement. For each notion, we propose a weak and a strong axiom, depending on whether they prescribe weak or strong preference. We break the symmetry only for Dominance $\wedge$ Improvement. Specifically, we follow classical game theory where the notion of dominance alone implies weak preference and we reserve strong preference to the tighter condition Dominance $\wedge$ Improvement.

$$\text{Same-context} \land \text{Improvement} \implies P(\sqsubseteq_1, \sqsubseteq_2) \quad \text{(SC)}$$
$$\text{Same-context} \land \text{Improvement} \implies P(\sqsubseteq_1, \sqsubseteq_2) \land \neg P(\sqsubseteq_2, \sqsubseteq_1) \quad \text{(SSC)}$$
$$\text{Dominance} \implies P(\sqsubseteq_1, \sqsubseteq_2) \quad \text{(Dom)}$$
$$\text{Dominance} \land \text{Improvement} \implies P(\sqsubseteq_1, \sqsubseteq_2) \land \neg P(\sqsubseteq_2, \sqsubseteq_1) \quad \text{(SDom)}$$
$$\text{Swap} \implies P(\sqsubseteq_1, \sqsubseteq_2) \quad \text{(Sw)}$$
$$\text{Swap} \implies P(\sqsubseteq_1, \sqsubseteq_2) \land \neg P(\sqsubseteq_2, \sqsubseteq_1) \quad \text{(SSw)}$$

Let $\mathcal{T}$ denote the set of the above six axioms, the following result characterizes the implications holding between the self-interest axioms in $\mathcal{T}$.

Theorem 1. *The only implications holding among the self-interest axioms in $\mathcal{T}$ are displayed in the Hasse diagram in Figure 1.*

For a set of axioms $T \subseteq \mathcal{T}$, we denote by $\mathit{Pref}(T)$ the set of preferences satisfying those axioms (a.k.a. the *preference theory* of T). In the next section we investigate the relationships between the preference theories induced by different subsets of axioms.

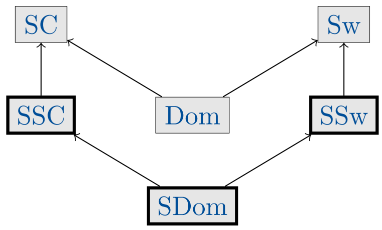

Figure 1: Hasse diagram for implication order between self-interest axioms. Lower axioms imply higher ones. Thick boxes denote axioms that prescribe strong preference.

3 Self-Interest Theories

We claim that axioms *SC* and *Sw* are primitive and should be satisfied by any a-preference. This claim is supported by the observation that all "plausible" preferences that we have been able to define satisfy those two axioms, starting with the preferences listed in Section 2.1. Therefore, in this section we study the set of all preference theories that refine $\mathit{Pref}(SC, Sw)$.

Syntactically, there are 32 subsets of $\mathcal{T}$ that contain both *SC* and *Sw*. Those theories obviously refine $\mathit{Pref}(SC, Sw)$. Figure 1 implies that all theories containing either *Dom* or *SDom* also refine $\mathit{Pref}(SC, Sw)$. This adds 27 more sets to the count. As a matter of fact, it is easier to list the only five subsets of $\mathcal{T}$ that do not refine $\mathit{Pref}(SC, Sw)$: they are $\emptyset$, $\{SC\}$, $\{Sw\}$, $\{SSC\}$, and $\{SSw\}$. Of the remaining 59 theories, we find that only nine of them are actually distinct, as stated by the following result and depicted in Figure 2.

Theorem 2. *For all $T \subseteq \mathcal{T}$, if $\mathit{Pref}(T) \subseteq \mathit{Pref}(SC, Sw)$ then $\mathit{Pref}(T)$ is equal to one of the nine canonical theories in Figure 2. Moreover, all of those theories are distinct and non-empty.*

Figure 2 also completely characterizes the containments holding between the nine canonical theories.

Theorem 3. *Let T_1, T_2 be two of the nine theories in Figure 2. It holds $\mathit{Pref}(T_1) \subset \mathit{Pref}(T_2)$ if and only if there is a sequence of arrows going from T_1 to T_2 in Figure 2.*

Define P^{dom} as the pre-order on $\mathcal{W}(A)$ such that $P^{dom}(\sqsubseteq_1, \sqsubseteq_2)$ iff Dominance holds. The following theorem characterizes the preferences in the strongest theory in terms of P^{dom}.

Theorem 4. *A preference belongs to $\mathit{Pref}(Dom, SDom)$ iff it refines and perfects P^{dom}.*

To appreciate the relevance of the previous theorem, notice that Lemma 1 tells us that all preferences satisfying the axiom *Dom* are 3-tier (that is, they do not distinguish between 3-tier equivalent weak orders). However, if Dominance establishes a *strong* preference between two weak orders, a preference satisfying Dom may instead equate those orders. For an extreme example, the degenerate preference $P^{\equiv}$ equates all weak orders, but it still satisfies *Dom*. Theorem 4 states that adding *SDom* to the picture forces preferences to uphold those strong preferences as well. Indeed, since preferences are total orders, a preference refining and perfecting P^{dom} can only (and must) establish a preference among weak orders that are incomparable for P^{dom}.

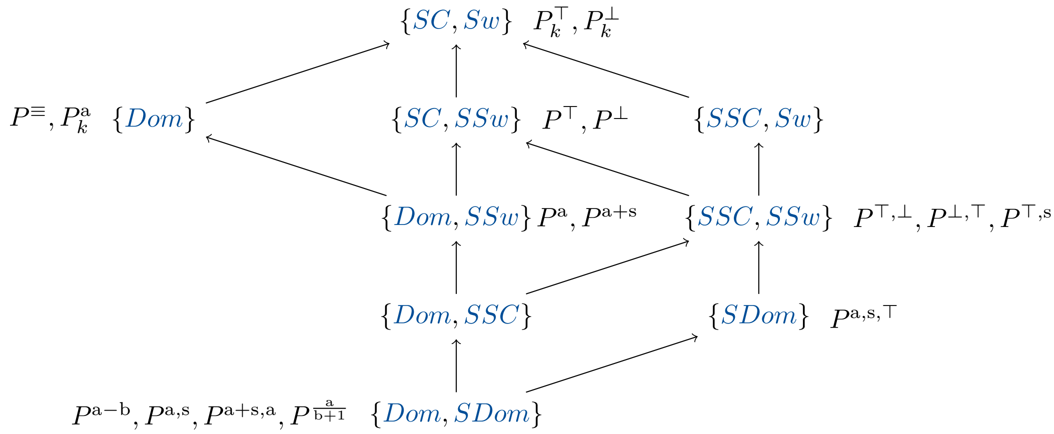

Figure 2: Hasse diagram for containment between self-interest theories. Lower theories are contained in higher ones. The preferences listed next to a theory belong to that theory and do not belong to any stronger theory in the diagram. $P^\equiv$ is the degenerate preference that equates all weak orders.

4 Conclusions

This paper initiates an investigation about the rational behavior of self-interested agents participating in a competition whose outcome is a weak order among participants. Our preliminary results show an intriguing landscape of self-interest theories, most of which include simple preferences that are easily motivated by practical scenarios.

From the point of view of the competition designer, being able to pinpoint the likely preference theory adopted by the participants is valuable information because it allows the designer to predict the participants' behavior during the competition. Generally speaking, the designer will want participants to act competitively, rather than indifferently or even cooperatively.

How to direct participants toward a specific theory or set of theories is a matter of future investigation, but some simple observations are already apparent. For example, the designer may announce that ties in the weak order will be resolved by a random linearization. In that case, cautious (that is, pessimistic) participants are likely to adopt preference $P^{\mathrm{wst}} = P^{\mathrm{a}}$, whereas optimistic ones may adopt $P^{\mathrm{bst}} = P^{\mathrm{a+s}}$. Both preferences lie in the theory $\{Dom, SSw\}$, and the presence of the axiom SSw ensures that an agent will act competitively whenever its actions may result in swapping its position with another agent that would otherwise finish higher up in the final ranking.

We also plan to investigate the case when the competition designer is able to assign monetary rewards to the participants, based on the weak order. Even if the assigned rewards do not break the ties in the order, they may still be able to influence the agents' preferences simply by tuning the *distance* between different levels in the weak order.

References

[1] E. W. Adams. Elements of a theory of inexact measurement. *Philosophy of science*, 32(3/4):205–228, 1965.

[2] E. Elkind and P. Faliszewski. Properties of multiwinner voting rules. *Social Choice and Welfare*, 48(3):599–632, 2017.

[3] M. Faella and L. Sauro. Do all tournaments admit irrelevant matches? In *Proc. of the 17th Int. Conf. on Autonomous Agents and MultiAgent Systems (AAMAS18)*, pages 982–989, 2018.

[4] R. Freeman, M. Brill, and V. Conitzer. General tiebreaking schemes for computational social choice. In *Proc. of the 17th Int. Conf. on Autonomous Agents and MultiAgent Systems (AAMAS15), Istanbul, Turkey, May 4-8*, pages 1401–1409, 2015.

[5] J. Lang, J. Monnot, and A. Slinko. Beyond electing and ranking: Collective dominating chains, dominating subsets and dichotomies. In *Proc. of the International conference on Autonomous Agents and Multi-Agent Systems (AAMAS17), Sao Paulo, Brazil, May 8–12*, 2017.

[6] R. Myerson. Axiomatic derivation of scoring rules without the ordering assumption. *Social Choice and Welfare*, 21(12):59–74, 1995.

[7] M. R. Sanver and W. S. Zwicker. Monotonicity properties and their adaptation to irresolute social choice rules. *Social Choice and Welfare*, 39(2):371–398, Jul 2012.

[8] L. Sauro. On the hierarchical nature of partial preferences over lotteries. *Autonomous Agents and Multi-Agent Systems*, 31(6):1467–1505, 2017.