Independence of Irrelevant Alternatives under the Lens of Pairwise Distortion

Théo Delemazure[1], **Jérôme Lang**[1], **Grzegorz Pierczyński**[2]

[1] CNRS, Paris Dauphine University, PSL
[2] University of Warsaw

Abstract

We give a quantitative analysis of the *independence of irrelevant alternatives* (IIA) axiom. IIA says that the society's preference between x and y should depend only on individual preferences between x and y: we show that, in several contexts, if the individuals express their preferences about additional (or "irrelevant") alternatives, this information helps to estimate better which of x and y has higher social welfare. Our contribution is threefold: (1) we provide a new tool to measure the impact of IIA on social welfare (*pairwise distortion*), based on the well-established notion of voting distortion, (2) we study the average impact of IIA in both general and metric settings, with experiments on synthetic data, and its impact with real datasets; and (3) we study the worst-case impact of IIA in the 1D-Euclidean metric space.

Introduction

Independence of irrelevant alternatives (IIA) states that a society's preference between two alternatives x and y depends only on how its individual members compare x and y (Arrow 1950), as opposed to taking into account the positions they give to other alternatives. IIA is the key axiom of Arrow's impossibility theorem: in ordinal settings, with at least three alternatives, IIA is incompatible with the unrestricted domain assumption, Pareto-efficiency, and non-dictatorship—three properties that are hard to give up.

Arrow's theorem has had, until today, a tremendous importance in social choice. It is often seen as negative, since there are many compelling arguments in favor of IIA (see for instance (Maskin 2020)). Still, IIA has been criticized for not taking into account preference intensities, and this may result in a loss of social welfare. To understand this argument, consider the following example:

$$51\% \text{ of voters: } x \succ y \succ z_1 \succ \ldots \succ z_{100}$$
$$49\% \text{ of voters: } y \succ z_1 \succ \ldots \succ z_{100} \succ x$$

What should be the collective preference between x and y here? If we exclude "irrelevant" alternatives $z_1,\ldots, z_{100}$, then the rational choice seems to use majority and conclude $x \succ y$.[1] On the other hand, we may have a strong intuition that it is not the right choice, because the additional alternatives provide us some implicit information: the preference $x \succ y$ of the first group of voters is most probably much *weaker* than the preference $y \succ x$ of the second group.

However this argument lacked until now a *quantitative* analysis. We argue that a recent research trend, *ordinal-cardinal voting distortion*, that aims at measuring the loss of social welfare (sum of individual utilities) caused by the use of ordinal instead of cardinal information, provides a suitable framework for quantifying the impact of IIA on social welfare. In our example above, if we had access to information about strength of preference, expressed as cardinal utilities, we would probably find that y has a higher social welfare than x. By denying access to information about strength of preference, IIA makes us choose x against y, which leads to what we call a *pairwise distortion* relative to x and y.

This being said, there are various ways of exploiting the additional information given by the irrelevant alternatives, and the question of which ones actually help reducing the loss of social welfare is not trivial. On the latter example, using plurality scores to choose between x and y still leads to choosing x; and using Copeland scores too. But using Borda scores leads this time to choosing y.

For choosing between two alternatives given a profile, we define *pairwise voting rules*. They take as input two alternatives (say, x and y) and a preference profile on a set of alternatives containing x and y, and output either x or y. Now the most natural pairwise voting rule satisfying IIA is the *pairwise majority rule*, that outputs the result of majority voting between x and y for each x and y (note that this is the only pairwise rule satisfying IIA and extending the majority rule for elections with two alternatives). Of course, applying pairwise majority to all pairs of alternatives sometimes returns a nontransitive relation over alternatives.

On the other hand, any social welfare function mapping a preference profile to a collective ranking of alternatives induces a (transitive) pairwise voting rule, where the choice between x and y is made by projecting the collective ranking on $\{x, y\}$. Thus, pairwise voting rules are a common frame-

[1] As a consequence of May's theorem (May 1952), IIA, together with a set of very weak conditions (neutrality, anonymity, and positive responsiveness), implies that the collective preference between x and y must be determined by majority voting between x and y.

work capturing both pairwise majority, which satisfies IIA, and (nondictatorial) social welfare functions, that do not.

Next, we define *pairwise distortion* of a pairwise voting rule f relative to x, y and a profile, as the loss of social welfare caused by choosing the alternative among $\{x, y\}$ determined by f. The details of the definition may vary—we focus on the two following settings: (1) in the first one, we consider *average pairwise distortion*, where the utilities or costs of the alternatives follow a given distribution, or are drawn from real datasets, (2) in the second one, we consider *worst-case pairwise distortion* in metric domains (voters and alternatives are located in a metric space and voters' preferences decrease with distance).

Now, the main question is the following: which pairwise rules give a lower pairwise distortion than pairwise majority, and how can we compare these rules according to their pairwise distortion? Our main findings are:

- When considering *average* distortion of synthetic data, as well as *empirical* distortion of real datasets, the overall picture is that (i) pairwise rules based on Borda and Copeland perform much better than pairwise majority, and their pairwise distortion decreases with the number of alternatives m; and (ii) the pairwise rule based on plurality scoring performs worse than pairwise majority, and its pairwise distortion increases with m.
- When considering *worst-case*, *metric* distortion, pairwise majority has distortion 3; when additional alternatives are placed by the election designer *cooperatively*, in the 1-dimensional (1D) Euclidean space, the plurality and Copeland pairwise rules perform just like pairwise majority, while the Borda pairwise rule and one of its variants do much better.

In the full version of this paper, we consider more rules and experiments. It also contains the proof of all our results.

Outline of the paper After discussing related work we give the necessary background and notations. Then, building on the classical literature on distortion, we define pairwise distortion of pairwise voting rules. We start by exploring the average pairwise distortion with experiments on real and synthetic data, then we look into bounds of worst-case pairwise distortion for the 1D-Euclidean metric space. We conclude by discussing further issues.

Related Work

Distortion

Distortion has been introduced by Procaccia and Rosenschein (2006) as a means to evaluate whether it is reasonable to make a collective decision after eliciting only ordinal preferences. Assuming that cardinal preferences are represented by utilities, the social welfare of an alternative is the sum of the utilities it provides to the agents. The distortion of a voting rule f for a given profile is then defined as the ratio between the maximum social welfare of an alternative, and the social welfare of the alternative selected by f; and the distortion of f is the maximum, over all profiles, of the distortion of f for that profile. *Metric* distortion (Anshelevich et al. 2018) aims at minimizing social cost instead of maximizing social welfare: voters and alternatives belong to a metric space, and the cost of an alternative to a voter is the distance between them. Voting distortion (metric or non-metric) has been the topic of a significant number of papers, too many for us to cite them all (and most of them are only moderately related to our concerns). See (Anshelevich et al. 2021) for an extensive survey of the literature until 2021.

Average-case analyses of distortion are far less common than worst-case analyses. For single-winner voting, Boutilier et al. (2015) show that the Borda rule is optimal for the uniform distribution, and Gonczarowski et al. (2023) show that a suitable positional scoring rule (*binomial voting*) performs well for all distributions. Caragiannis et al. (2017) consider average distortion for multi-winner rules, Filos-Ratsikas, Micha, and Voudouris (2019) for district-based elections and Benadè, Procaccia, and Qiao (2019) for social welfare functions.

Independence of Irrelevant Alternatives

The primary reason why Arrow imposed IIA was to prevent the implicit use of interpersonal comparisons (Arrow 1950). However, it also prevents the use of information about *intensities of preferences between two alternatives* revealed by the positions of thesealternatives with respect to *other* ("irrelevant") alternatives. This has been previously discussed(Coakley 2016; Pearce 2021; Maskin 2020; Sen 1970; Osborne 1976; Hillinger 2005; Lehtinen 2011; Maskin 2023), and examples such as the one presented in our introduction highlight its practical negative implications.

Because Arrow's theorem ruled out the existence of a social welfare function under "reasonable" conditions, it has had a negative impact on welfare economics (Fleurbaey and Mongin 2005; Pearce 2021; Igersheim 2019). As other properties stated in Arrow's theorem can hardly be given up, IIA is the most debatable of the conditions of Arrow's theorem, and is actually given up *de facto* when defining voting rules. Still, IIA is considered attractive for several reasons, such as avoiding vote splitting (Maskin 2020).

Pairwise Voting Rules

Let V be a set of n voters and A a set of m alternatives. A ranking $\succ$ of A is a linear order (irreflexive, antisymmetric, transitive and connected relation) of A. $\mathcal{L}(A)$ denotes the set of all rankings over A. A preference profile is a collection of rankings $P = (\succ_1, \ldots, \succ_n)$. For a ranking $\succ_i$, we denote by σ_i the corresponding rank function: for each alternative $x \in A$, $\sigma_i(x) = |\{y \in A \mid y \succ_i x\}| + 1$ the rank of x in $\succ_i$.

A *pairwise (voting) rule* is a function f that, given a preference profile P over A and two alternatives $x, y \in A$, outputs $f(P \mid x, y) \in \{x, y\}$. Equivalently, f associates with every preference profile P a tournament (an irreflexive, antisymmetric and connected relation, but not necessarily transitive).[2] A pairwise rule f satisfies IIA if $f(P \mid x, y) = f(P' \mid x, y)$ for all $P = (\succ_1, \ldots, \succ_n)$ and $P' = (\succ'_1, \ldots, \succ'_n)$ such that for all voters $i \in V$, $x \succ_i y$ if and only if $x \succ'_i y$.

[2]Yet another interpretation: $f(P|.,.)$ is the restriction of a deterministic choice function (Sen 1971) on pairs of alternatives.

Among pairwise rules that satisfy IIA, the canonical one is the *pairwise majority* rule: $f_{maj}(P \mid x, y) = x$ (resp. y) if a majority of voters prefer x to y (resp. y to x). In case of a tie, we use a tie-breaking priority relation over alternatives. *We will use such a tie-breaking mechanism more generally for all pairwise rules.* Note that except for the treatment of ties, the graph induced from pairwise majority by $x \to y$ if $f_{maj}(P \mid x, y) = x$ is the majority graph associated with P.

Another prominent family of pairwise rules consists of those that output transitive tournaments, that is, if $f(P \mid x, y) = x$ and $f(P \mid y, z) = y$ then $f(P \mid x, z) = x$. In this case, f corresponds to a social welfare function g mapping every profile P to a ranking $g(P) \in \mathcal{L}(A)$ defined by $x \succ_{g(P)} y$ if and only if $f(P \mid x, y) = x$. Conversely, any social welfare function g induces a pairwise rule $g_{\mathbf{PW}}$.

Among pairwise rules of this class, we will mostly focus on those that are based on a score function Sc that maps every profile P and alternative x to a score $Sc(x, P)$. The pairwise rule f_{Sc} is then defined by $f_{Sc}(P|x, y) = \mathrm{argmax}(Sc(x, P), Sc(y, P))$.

We will make use of the following pairwise rules, all based on some scoring function Sc. We include three positional scoring rules: Plurality because of its simplicity and wide usage, Borda because of its central role in voting and its optimal average distortion in some settings (Caragiannis and Procaccia 2011), Half-approval as it was shown to have a good average (classical) distortion for a large class of distributions (Gonczarowski et al. 2023); and one prominent Condorcet rule (Copeland), which is known to have good metric distortion guarantees.

Positional scoring rules Let $\vec{s} = (s_1, s_2, \ldots, s_m)$ be a non-increasing vector. It is *normalized* if $s_1 = 1$ and $s_m = 0$. Each $x \in A$ gets s_j points from each voter $i \in V$ who ranks x at position j. The score of $x \in A$ for profile P is:

$$Sc(x, P) = \sum_{j=1}^{m} s_j \, |\{i \in V | \sigma_i(x) = j\}|$$

We consider the following pairwise rules: $\mathbf{Borda_{PW}}$ ($\vec{s} = (m-1, m-2, \ldots, 1, 0)$), $\mathbf{\textit{k}\text{-}approval_{PW}}$ ($\vec{s} = (1, \ldots, 1, 0, \ldots, 0)$ with 1 in the first k positions), with as special cases $\mathbf{Plurality_{PW}}$ ($k = 1$) and $\mathbf{Half\text{-}approval_{PW}}$ ($k = \lceil m/2 \rceil$).

$\mathbf{Copeland_{PW}}$ For $x, y \in A$, we say that x *dominates* y if a majority of voters prefer x to y. The Copeland score $Sc(x)$ is the number of alternatives y dominated by x.[3]

Pairwise Distortion

We now formally define *pairwise distortion*. Similarly as for standard distortion, we consider *unconstrained* distortion in which we assume voters gain unconstrained cardinal utilities from alternatives, and *metric* distortion, in which voters and alternatives are embedded in a metric space, and the cost of an alternative for a voter is the distance between them.

[3]The version of Copeland we use here is Copeland^0, where pairwise ties don't give any points.

Unconstrained Pairwise Distortion

In the unconstrained distortion setting, every voter $i \in V$ receives a utility $U_i(x) \in \mathbb{R}_{\geq 0}$ from alternative $x \in A$. A *utility profile* U is a collection $U = (U_i)_{i \in V}$.[4] We say that a preference profile P and a utility profile U are consistent with each other if for all $x, y \in A$ and all voters i, if $U_i(x) > U_i(y)$, then $x \succ_i y$ in P and we denote it $P \approx U$. The *social welfare* of an alternative $x \in A$ is $SW(x) = \sum_{i \in V} U_i(x)$.

The *pairwise distortion* of a pairwise rule f on a utility profile U for two alternatives $x, y \in A$ is the worst-case ratio over all $P \approx U$ between the social welfare of the optimal alternative and that of $f(P \mid x, y)$:

$$\mathrm{dist}(f, U \mid x, y) = \max_{P: P \approx U} \frac{\max(SW_U(x), SW_U(y))}{SW_U(f(P \mid x, y))}$$

Metric Pairwise Distortion

In the *metric* distortion setting, we assume that both voters and alternatives are points in some pseudometric space $(V \cup A, d)$ with $d : (V \cup A)^2 \to \mathbb{R}_{\geq 0}$ a distance function, where $d(i, x)$ represents the cost of alternative x for voter i. The *social cost* of an alternative $x \in A$ for the pseudometric d is $SC_d(x) = \sum_{i \in V} d(i, x)$. As in the general setting, we can naturally induce a preference profile P based on d. We denote $P \approx d$ if for all $x, y \in A$ and all voters i, if $d(x, i) < d(y, i)$ then $x \succ_i y$ in P.

In the metric setting, the pairwise distortion of a pairwise rule f on a metric d for two alternatives $x, y \in A$ is the worst-case ratio over all $P \approx d$ between the social cost of $f(P \mid x, y)$ and that of the optimal alternative:

$$\mathrm{dist}(f, d \mid x, y) = \max_{P: P \approx d} \frac{SC_d(f(P \mid x, y))}{\min(SC_d(x), SC_d(y))}$$

Average and Empirical Pairwise Distortion

We first focus on the average-case scenario, and define the *average pairwise distortion* given a probability distribution $\mathcal{D}$ over utility profiles U (for the unconstrained setting) or over pseudometrics d (for the metric setting). When the distribution is sampled based on a real dataset, we refer to it as *empirical distortion*.

Given a utility profile or a pseudometric, we obtain a pairwise distortion for each pair of alternatives, which we have then to aggregate; for this we consider two possibilities: taking the *maximum* or the *average* over all pairs.

The average pairwise distortion for $\delta \in \{\mathbf{avg}, \mathbf{max}\}$ and $\Delta \in \{U, d\}$ (respectively the unconstrained and the metric settings) is defined as:

$$\text{avg-dist}(f, \delta, \mathcal{D}) = \underset{\Delta \sim \mathcal{D}}{\mathbf{avg}} \ \underset{(x,y) \in A}{\delta} \mathrm{dist}(f, \Delta \mid x, y)$$

Note that we use here two ways of averaging, which should not be confused: averaging over profiles sampled with $\mathcal{D}$ (for defining *average* pairwise distortion), and averaging over pairs of alternatives ($\delta = \mathbf{avg}$).

[4]In the literature, it is often assumed that utilities are normalized (for all $i \in V$, $\sum_{x \in A} U_i(x) = 1$), as otherwise the worst-case distortion is infinitely large. However, for the analysis of average distortion, such an assumption is not required, as we are already restricting ourselves to a probability distribution over utilities.

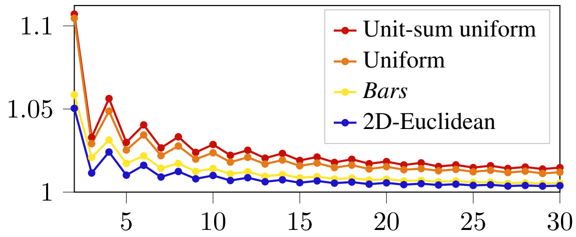

Figure 1: Average pairwise distortion (over 100,000 profiles) for $m = 2$ and several distributions. The x-axis corresponds to the number of voters n.

Distributions

Our study of average pairwise distortion relies on experiments. We first have to choose which probability distributions to use for generating profiles, with the aim of observing the behaviour of the rules when the parameters n and m vary. We choose the following distributions, two standard synthetic ones and one sampled from a real-world dataset.

Uniform In the unconstrained setting, we sample profiles according to the uniform distribution of utilities over $[0, 1]$.

2D-Euclidean uniform In the metric setting, we sample positions of voters and alternatives uniformly at random in the 2-dimensions (2D) Euclidean space.

Bars In the unconstrained setting, we investigate the empirical distortion of the *Bars* dataset (Lesser et al. 2017) from the *Preflib* database (Mattei and Walsh 2013). It contains ratings $r \in \{1, 2, 3, 4, 5\}$ of bars. We interpret these ratings as utilities, adding some small random noise to the ratings in order to remove ties. Then, for a given pair (n, m), profiles are sampled by selecting randomly n voters and m alternatives from the dataset (which originally contains 95 voters and 16 alternatives).

Moreover, for the analysis of the case with $m = 2$ alternatives, we also consider the **unit-sum uniform** distribution in the unconstrained setting: for each voter $i \in V$, the utility of x is selected uniformly at random in $[0, 1]$, and the utility of y is $U_i(y) = 1 - U_i(x)$.

Two Alternatives

In this first experiment, we focus on the case of two alternatives and investigate how the average distortion varies when we increase the number of voters n. This case is of particular interest, because it corresponds to the average distortion of the pairwise majority rule (for which the presence of additional alternatives has no influence). Note that when $m = 2$, the average pairwise distortion for $\delta = \mathbf{avg}$ and $\delta = \mathbf{max}$ are the same as there is only one pair of alternatives.

Figure 1 shows the pairwise distortion for the different distributions. Interestingly, the average pairwise distortion is very close to 1. It reach its highest value for $n = 2$ and the unit-sum uniform distribution, with an average (pairwise) distortion of $5/2 - ln(4) \approx 1.11$. We also observe that pairwise distortion gets asymptotically smaller (with some parity effect due to tie-breaking) when the number of voters n increases. This is not surprising: because of Hoeffding's inequality, the social welfare (or the social cost) of the two alternatives get closer to each other when n increases, thus reducing distortion. These observations suggest that the average distortion of the pairwise majority rule is usually very far from its worst-case distortion (3 in the metric setting and the normalized unconstrained setting)

Increasing the Number of Alternatives

In this section, we investigate how the average pairwise distortion varies with the number of alternatives m. For all experiments, we use profiles of 30 voters and up to 15 alternatives. We compare distortion for pairwise majority (which satisfies IIA) and four transitive pairwise rules: Plurality$_{\mathbf{PW}}$, Half-approval$_{\mathbf{PW}}$, Borda$_{\mathbf{PW}}$, and Copeland$_{\mathbf{PW}}$.

The first row of Figure 2 shows the average *mean* pairwise distortion ($\delta = \mathbf{avg}$). For the pairwise majority rule, it remains constant with the number of alternatives m, as it satisfies IIA. The transitive pairwise rules considered here behave differently: the average pairwise distortion of Plurality$_{\mathbf{PW}}$ increases with m, as each voter gives information about only one of the m alternatives (its preferred one). On the opposite, Borda$_{\mathbf{PW}}$ and Copeland$_{\mathbf{PW}}$ both seem to really take advantage of the extra information brought by the additional alternatives, as their average pairwise distortion decreases with m (and particularly quickly in the unconstrained setting);[5]; Half-approval$_{\mathbf{PW}}$ lays in the middle, and its distortion is not far from that of pairwise majority. Note that the average pairwise distortion of pairwise majority varies from one distribution to another. In particular, it is higher for distributions in the unconstrained setting than in the metric setting. This suggests that in the unconstrained setting, using other pairwise rules already have more potential to do better than pairwise majority.

Our conclusion is that using information about additional alternatives can help a lot, provided that the way to use it is carefully chosen, and that Borda$_{\mathbf{PW}}$ and Copeland$_{\mathbf{PW}}$ seem both to be good choices.

The second row of Figure 2 shows the variation with m of the average *max* pairwise distortion ($\delta = \mathbf{max}$). It increases with m for all distributions and all pairwise rules, even pairwise majority. It is an intuitively expected behavior: we consider all pairwise distortion values obtained for each pair of alternatives, *and keep only the largest value*. The more alternatives there are, the more pairs to be considered, and the more likely a *bad* pair is sampled. However, the relative order between the curves is the same as for $\delta = \mathbf{avg}$. In particular, Borda$_{\mathbf{PW}}$ and Copeland$_{\mathbf{PW}}$ always show a better average max pairwise distortion than pairwise majority, and the gap is wider than for $\delta = \mathbf{avg}$: that is, the average loss of social welfare caused by IIA is even larger if we only look at the worst pair of alternatives in the profile. These results align with our theoretical findings of the next section.

[5]There is an exception however: with the 1D-Euclidean uniform distribution, Borda$_{\mathbf{PW}}$ has higher average pairwise distortion than pairwise majority. This can be explained by the specific structure of 1D preferences. (Still, average distortion decreases with the number of alternatives.)

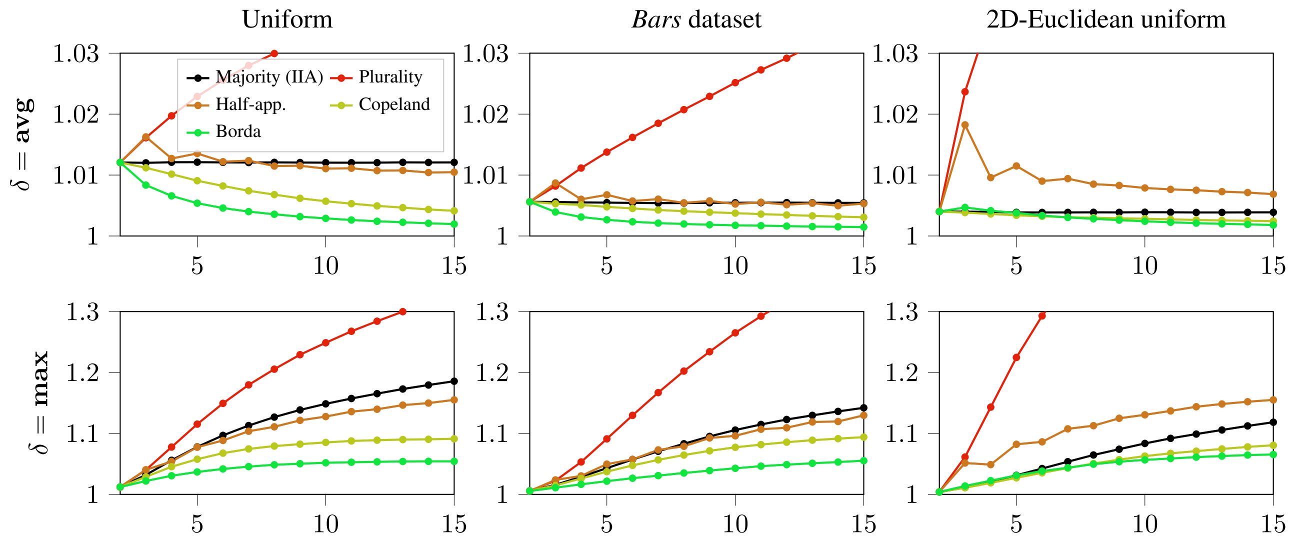

Figure 2: Average *mean* pairwise distortion ($\delta = \mathbf{avg}$, first row) and average *max* pairwise distortion ($\delta = \mathbf{max}$, second row) over 10,000 random profiles. The x-axis corresponds to the number of alternatives m.

Worst-Case Metric Pairwise Distortion

In this section we only consider the metric setting, since the research on standard distortion suggests that it leads to more positive results than the unconstrained setting (see, *e.g.*, (Anshelevich et al. 2021)). We consider *worst-case* pairwise distortion, by assuming that voters are placed in the metric space so as to maximize the pairwise distortion of a specific pair of alternatives (x, y), given the positions of all the alternatives.

A key question is how to choose the positions of the other alternatives when determining the worst-case pairwise distortion of a pair (x, y). Our aim is to compute tight lower and upper bounds of the worst-case distortion for each pairwise rule. This problem can be seen as a game: a first agent selects the positions of the alternatives, and a second agent responds in an adversarial manner by choosing the positions of the voters that maximize pairwise distortion. A *cooperative* (resp. *adversarial*) first agent that places the alternatives so as to minimize (resp. maximize) the worst-case pairwise distortion gives us a lower (resp. upper) bound.[6]

The intuition behind the lower bound with a cooperative agent is that the designer of the game can choose the positions of alternatives so as to maximize the gain of information obtained from voters' rankings, and thus ease preference elicitation. For instance, in a facility location context, if the designer wants to know which of x and y is a better collective choice (perhaps because they are the only possible choices), they can ask the voters to rank x, y, as well as other carefully chosen additional *fake* alternatives, used as reference points. The adversarial case may have less intuitive appeal but is in line with standard worst-case assumptions made when defining distortion in various settings.

It is known that the classical worst-case (pairwise) distortion for $m = 2$ alternatives $\{x, y\}$ is 3, which can be seen on this well-known instance: half of the voters prefer x and the other half y. Assume the tie is broken in favor of x. The y-voters are all located at the same position as y; the x-voters are half-way between x and y. The (pairwise) distortion is $\frac{n/2 \cdot 1 + n/2 \cdot 1/2}{n/2 \cdot 0 + n/2 \cdot 1/2} = 3$. In this instance, adding alternatives looks promising: if we add an alternative exactly between x and y, we will be able to see that the voters who prefer x to y only have a slight preference, while the others have a strong preference for y. Now, how much can we improve pairwise distortion from 3 (obtained with $m = 2$) when we increase the number of alternatives with a cooperative agent? And how bad can it get against an adversarial agent?

Clearly, for any pairwise rule satisfying IIA, pairwise distortion remains constant, independently from the number and positions of other alternatives. In particular, the worst-case pairwise distortion of pairwise majority is always 3.

Without loss of generality, fix two alternatives $x, y \in A$. Denoting $d|_A$ the restriction of a pseudometric d to the set of alternatives A, for $\gamma \in \{\inf, \sup\}$ (respectively the cooperative and adversarial cases) we define

$$\mathrm{dist}(f, \gamma, m) = \gamma_{d|_A} \sup_{d} \mathrm{dist}(f, d \mid x, y)$$

which we call γ-pairwise distortion for m alternatives.

From now on, *we focus on the case of a 1D-Euclidean space*: voters and alternatives are associated with positions on a line. We make this choice as it is a natural setting to start this study, and it is known to have a lot of practical interpretations, such as facility location; we leave the study of general metric spaces for further research.[7] We denote

[6] The terminology 'cooperative/adversarial' is consistent with our game-theoretic abstract interpretation, but we do not mean that the additional alternatives are chosen strategically. Instead we could say 'optimistic/pessimistic'; this decision-theoretic terminology is not perfect either since it suggests a behaviour towards risk.

[7] Still, Theorems 5 and 6 can be generalized to any metric space.

	IIA (majority$_{\mathbf{PW}}$)	Borda$_{\mathbf{PW}}$	OddBorda$_{\mathbf{PW}}$	k-Approval$_{\mathbf{PW}}$	Plurality$_{\mathbf{PW}}$
inf-pairwise distortion	3	$\frac{m+1}{m-1}$	$\frac{2m-1}{2m-3}$	2	3
sup-pairwise distortion	3	$2m-1$	$4m-5$	∞	∞

Table 1: γ-pairwise distortions of several pairwise voting rules in the 1d-Euclidean metric space.

$p(e) \in \mathbb{R}$ the position of $e \in V \cup A$ on the line. We will assume without loss of generality that the positions of the alternatives x and y with respect to which we study pairwise distortion are $p(x) = 0$ and $p(y) = 1$.

With only two alternatives ($A = \{x, y\}$), all the rules that we consider boil down to the majority rule, whose pairwise distortion is 3. From the well-known fact that the pairwise majority rule outputs a transitive order of alternatives in a 1D-Euclidean space (because of single-peakedness), in this setting Copeland$_{PW}$ is equivalent to pairwise majority. Therefore, its inf-pairwise and sup-pairwise distortions are 3 for any $m \geq 2$ in the 1D-Euclidean space.

In the remaining of this section, we compute the inf and sup-pairwise distortion of various pairwise rules. Table 1 summarize our results.

Lower Bound

We first focus on the cooperative case: alternatives are positioned in order to minimize worst-case pairwise distortion (however, recall that voters will still be positioned in an adversarial way). The question is here whether the lower bound of 3 obtained when we impose IIA can decrease if we use some other pairwise rules.

Without much surprise, for Plurality$_{\mathbf{PW}}$ we cannot obtain a better pairwise distortion than 3, even if we can choose the position of the alternatives.

Theorem 1. *The* inf*-pairwise distortion of Plurality*$_{\mathbf{PW}}$ *in the 1D-Euclidean metric space is* 3 *for any* $m \geq 2$.

For k-approval with $k \notin \{1, m-1\}$, we get a better inf-pairwise distortion, as we can reduce it from 3 to 2.

Theorem 2. *For any* $m \geq 4$ *and* $2 \leq k \leq m-2$*, the* inf*-pairwise distortion of* k*-approval*$_{\mathbf{PW}}$ *in the 1D-Euclidean metric space is* 2.

The result for Borda$_{\mathbf{PW}}$ is even more positive: the inf-pairwise distortion quickly tends to 1 as we add alternatives.

Theorem 3. *The* inf*-pairwise distortion of the Borda*$_{\mathbf{PW}}$ *pairwise rule in the 1D-Euclidean metric space is* ${}^{m+1}/_{m-1}$ *for any* $m \geq 2$.

To prove this, we first show that the inf-pairwise distortion of most positional scoring rule for given positions of alternatives *between* x and y can be computed easily. In the following proposition, we assume that all alternatives $z_j \in A$ are between x and y. We denote $A = \{z_1, \ldots, z_m\}$ such that $p(z_1) \leq p(z_2) \leq \cdots \leq p(z_{m-1}) \leq p(z_m)$. We naturally have $x = z_1, y = z_m$. For simplicity, the positions of the alternatives are noted $p_j = p(z_j) = d(x, z_j)$ with $p_1 = 0$ and $p_m = 1$.

Lemma 1. *For the 1D-Euclidean metric space, the* inf*-pairwise distortion of a positional scoring rule associated with the normalized scoring vector* $s = (s_1, \ldots, s_m)$ *with* $s_1 = 1$ *and* $s_{m-1} > s_m = 0$*, for fixed positions of the alternatives* $p_1, \ldots, p_m$ *between* x *and* y*, is equal to*

$$\max\left(\max_{1 \leq i,j < m} \frac{K_{i,j}}{1-K_{i,j}}, \max_{1 \leq i,j < m} \frac{K'_{i,j}}{1-K'_{i,j}}\right)$$

where:

$$K_{i,j} = \frac{p_{i+1} \cdot s_j + (1 + p_{m+1-j}) \cdot s_i}{2(s_i + s_j)}$$

$$K'_{i,j} = \frac{(1 - p_{m+1-(i+1)}) \cdot s_j + (2 - p_j) \cdot s_i}{2(s_i + s_j)}$$

Note that when the positions of the alternatives are symmetrical (for all j, $p_j = 1 - p_{m+1-j}$), we have $K_{i,j} = K'_{i,j}$.

We now give a proof sketch of Theorem 3.

Proof sketch of Theorem 3. The normalized Borda scoring vector is defined by $s_j = \frac{m-j}{m-1}$ for all j. We first show that there exist positions of the alternatives that achieve distortion $\frac{m+1}{m-1}$. We place alternatives at equal distance from each other, *i.e.*, such that $p_j = \frac{j-1}{m-1}$. Note that these positions are symmetrical (for all j, $p_j = 1 - p_{m+1-j}$), thus, $K'_{i,j} = K_{i,j}$. Using values of the Borda vector and these positions of the alternatives with Lemma 1, we can compute the value of $K_{i,j}$ for all i, j. Consider now the function h defined by $h(i,j) = K_{i,j}$. By straightforward calculations we can show that the partial derivative of h with respect to i is always non-negative, while the one with respect to j is always non-positive. It implies that h reaches its maximum for $i = m-1$ and $j = 1$, which corresponds to the case $K_{m-1,1}$, giving a distortion of $\frac{m+1}{m-1}$.

Let us now prove that we cannot obtain a better worst-case pairwise distortion. For this we consider the following profile: a fraction $\frac{m-1}{m}$ of voters are at distance $\frac{1}{2}$ of x and y and prefer x to y ($x \succ_i y$), and the remaining $\frac{1}{m}$ fraction of voters are at the same position as y, and obviously prefer y to x. Now, let $m_{(1,2]}$ be the number of alternatives $z_j \in A$ with $1 < p(z_j) \leq 2$ (they are closer to y than x is), $m_{>2}$ the number of alternatives with $p(z_j) > 2$ (x is closer to y than they are), and $m_{<0}$ the number of alternatives with $p(z_j) < 0$. The normalized score of x (i.e., its score divided by the number of voters) is $Sc(x) = \frac{1}{m} \cdot \frac{m_{>2}+m_{<0}}{m-1} + \frac{m-1}{m} \cdot \frac{m_{(1,2]}+m_{>2}+m_{<0}+1}{m-1}$ and the one of y is $Sc(y) = \frac{1}{m} \cdot 1 + \frac{m-1}{m} \cdot \frac{m_{(1,2]}+m_{>2}+m_{<0}}{m-1}$. Thus $Sc(x) - Sc(y) = \frac{m_{>2}+m_{<0}}{m(m-1)} \geq 0$, meaning w.l.o.g. that Borda$_{\mathbf{PW}}$ selects x. However, y is the better alternative in this profile, which gives a distortion $\frac{{}^{m-1}/_{m} \cdot {}^{1}/_{2} + {}^{1}/_{m} \cdot 1}{{}^{m-1}/_{m} \cdot {}^{1}/_{2}} = \frac{m+1}{m-1}$. This shows that with this profile, any positions of the alternatives leads to a worst-case distortion $\geq \frac{m+1}{m-1}$. $\square$

One may wonder whether Borda$_{\mathbf{PW}}$ is the optimal rule for inf-pairwise distortion. It appears that a better bound is obtained by a slightly different, yet very similar scoring rule, defined by the scoring vector $\vec{s} = (\dots, 9, 7, 5, 3, 1, 0)$. We call it *Odd Borda*$_{\mathbf{PW}}$, because (except for the last position) it contains consecutive odd numbers. We prove that in the 1D-Euclidean metric space, with m alternatives, the OddBorda$_{\mathbf{PW}}$ rule has a inf-pairwise distortion of ${}^{2m-1}/_{2m-3}$. We also show that if we add the additional assumption that all alternatives can be placed only between x and y, this bound is tight among all positional scoring rules and achievable only by OddBorda$_{\mathbf{PW}}$.

Theorem 4. *The* inf*-pairwise distortion of OddBorda*$_{\mathbf{PW}}$ *in the 1D-Euclidean metric space is* ${}^{2m-1}/_{2m-3}$ *for any* $m \geq 2$*. This is the lowest distortion among all positional scoring rules if we assume that all alternatives are between* x *and* y*.*

We conjecture that this bound remains tight without the assumption that all alternatives are between x and y. Another question is whether this bound applies to rules besides scoring rules. This can already be partly answered: for given positions of the alternatives between x and y, we can compute a lower bound on the worst-case pairwise distortion, independently of the rule used. We found that for $m < 6$, over $1,000,000$ different positions of the alternatives, the lower bound is always $\geq \frac{2m-1}{2m-3}$, suggesting that OddBorda$_{\mathbf{PW}}$ is optimal when all alternatives are between x and y. However, when $m \geq 6$, we found that OddBorda$_{\mathbf{PW}}$ is not the optimal pairwise rule for the 1D-Euclidean metric space.

Upper Bound

We now focus on the pessimistic case, in which both alternatives and voters are placed in an adversarial way.

For Plurality$_{\mathbf{PW}}$, and more generally k-approval$_{\mathbf{PW}}$ for any k, this upper bound is infinitely large when $m > 2$. We saw earlier that these rules (especially Plurality$_{\mathbf{PW}}$) do not use enough information to take advantage of the additional alternatives, we now know that they might actually lose all important information if we add alternatives.

Theorem 5. *The* sup*-pairwise distortion of Plurality*$_{\mathbf{PW}}$ *and more generally* k*-approval*$_{\mathbf{PW}}$ *for all* $k \leq n$*, in the 1D-Euclidean metric space, is* $+\infty$ *for any* $m \geq 3$*.*

The results are slightly better for Borda$_{\mathbf{PW}}$ and OddBorda$_{\mathbf{PW}}$.

Theorem 6. *The* sup*-pairwise distortion of Borda*$_{\mathbf{PW}}$ *(resp. OddBorda*$_{\mathbf{PW}}$*) in the 1D-Euclidean metric space is equal to* $2m - 1$ *(resp.* $4m - 5$*) for all* $m \geq 2$*.*

Proof. We show the bound for Borda$_{\mathbf{PW}}$. We divide voters into two groups: those who prefer x to y, and those who prefer y to x. We assume without loss of generality that Borda$_{\mathbf{PW}}$ selects y but x has a lower social cost. Observe that every voter who prefers x to y gives at least one more point to x than to y, and every voter who prefers y to x gives at most $m - 1$ more points to y than to x. Therefore, if we denote $\alpha \in [0, 1]$ the proportion of voters who prefer x to y, we have $Sc(x) \geq n\alpha$ and $Sc(y) \leq n(1 - \alpha)(m - 1)$. To have $Sc(y) \geq Sc(x)$ (as y is preferred), we need $\alpha \leq (1-\alpha)(m-1)$, which implies $\alpha \leq \frac{m-1}{m}$. Now, observe that voters preferring x to y maximize distortion by being at the same position as x if they could, and voters preferring y to x can maximize distortion by being exactly between x and y (${}^{1}/_{2}$ of both). Therefore, distortion cannot be higher than

$$\frac{\alpha + (1-\alpha)\frac{1}{2}}{(1-\alpha)\frac{1}{2}} = \frac{1+\alpha}{1-\alpha} \leq \frac{1 + \frac{m-1}{m}}{1 - \frac{m-1}{m}} = 2m - 1$$

Moreover, this bound is reached. Let ${}^{1}/_{4} > \varepsilon > 0$ and consider the profile in which all alternatives $z_j \neq x, y$ are at position $p(z_j) = \varepsilon$. Set $m - 1$ voters at position 2ε, ranking y last and x second last, and one voter at position ${}^{1}/_{2} + \varepsilon$, ranking y first and x last. In this profile $Sc(x) = Sc(y) = m - 1$ and assume w.l.o.g. that ties are broken in favor of y. The distortion for this profile is $\frac{(m-1)(1-2\varepsilon)+{}^{1}/_{2}-\varepsilon}{{}^{1}/_{2}+\varepsilon}$ and it tends to $2m - 1$ when ε tends to 0. $\square$

This implies that for these two rules, if distortion quickly gets close to 1 in the best case, it also quickly becomes very large in the worst case. However, by comparing the proofs of the inf-pairwise and sup-pairwise distortions, we notice that the positions of alternatives in the best-case scenario (equidistant alternatives) seems more natural than the positions of alternatives in the worst-case scenario (almost all alternatives at the same position). In particular, the equidistant positions of alternatives are their expected positions if they are uniformly distributed on the line. This uniformity of the positions partly explains why average distortion decreases with the number of alternatives, but also with the dimension. This intuition that pairwise distortion is in average closer to the best case and decreases with the number of alternatives is also supported by the experiments reported in the previous section.

Conclusion

We have introduced pairwise distortion as a tool for the quantitative analysis of the impact on social welfare of the Independence of Irrelevant Alternatives (IIA) axiom. Our conclusions are mixed:

- using information about additional alternatives may help reducing average distortion, but it crucially depends on the choice of the pairwise voting rule used. We found out that — among the rules we studied — the Copeland and Borda pairwise rules are particularly good at decreasing average distortion, but the Plurality pairwise rule has the opposite effect and leads to a larger distortion than sticking to IIA and using pairwise majority.
- when it comes to worst-case distortion, a crucial parameter is the origin of additional alternatives. If they are chosen by the election designer, then the Borda pairwise rule is quite good, and its variant OddBorda (a rule that may be interesting on its own) is even better. However, if they are chosen adversarially, then better stick to IIA.

Among further issues, it is worth looking at average pairwise distortion under distributions in which votes are correlated (such as Mallows or mixtures thereof), and proving or disproving our conjecture about the optimality of OddBorda in the 1D-Euclidean space.

Acknowledgements

Grzegorz Pierczyński was supported by Poland's National Science Center grant no. 2022/45/N/ST6/00271. Théo Delemazure and Jérôme Lang were supported by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). We thank Dominik Peters and the reviewers for helpful comments.

References

Anshelevich, E.; Bhardwaj, O.; Elkind, E.; Postl, J.; and Skowron, P. 2018. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264: 27–51.

Anshelevich, E.; Filos-Ratsikas, A.; Shah, N.; and Voudouris, A. A. 2021. Distortion in Social Choice Problems: The First 15 Years and Beyond. In *IJCAI*, 4294–4301. ijcai.org.

Arrow, K. J. 1950. A difficulty in the concept of social welfare. *Journal of political economy*, 58(4): 328–346.

Benadè, G.; Procaccia, A. D.; and Qiao, M. 2019. Low-Distortion Social Welfare Functions. In *AAAI*, 1788–1795. AAAI Press.

Boutilier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2015. Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, 227: 190–213.

Caragiannis, I.; Nath, S.; Procaccia, A. D.; and Shah, N. 2017. Subset Selection via Implicit Utilitarian Voting. *J. Artif. Int. Res.*, 58(1): 123–152.

Caragiannis, I.; and Procaccia, A. D. 2011. Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence*, 175(9-10): 1655–1671.

Coakley, M. 2016. Interpersonal comparisons of the good: Epistemic not impossible. *Utilitas*, 28(3): 288–313.

Filos-Ratsikas, A.; Micha, E.; and Voudouris, A. A. 2019. The distortion of distributed voting.

Fleurbaey, M.; and Mongin, P. 2005. The news of the death of welfare economics is greatly exaggerated. *Soc. Choice Welf.*, 25(2-3): 381–418.

Gonczarowski, Y. A.; Kehne, G.; Procaccia, A. D.; Schiffer, B.; and Zhang, S. 2023. The Distortion of Binomial Voting Defies Expectation. *CoRR*, abs/2306.15657.

Hillinger, C. 2005. The Case for Utilitarian Voting. *Homo Oeconomicus*, 23: 295–321.

Igersheim, H. 2019. The Death of Welfare Economics: History of a Controversy. *History of Political Economy*, 51(5): 827–865.

Lehtinen, A. 2011. A welfarist critique of social choice theory. *Journal of Theoretical Politics*, 23(3): 359–381.

Lesser, O.; Dery, L.; Kalech, M.; and Elovici, Y. 2017. Group Decision Support for Leisure Activities Using Voting and Social Networks. *Group Decision and Negotiation*, 26.

Maskin, E. 2020. A modified version of Arrow's IIA condition. *Soc. Choice Welf.*, 54(2-3): 203–209.

Maskin, E. 2023. Arrow's Theorem, May's Axiom, Borda's Rule.

Mattei, N.; and Walsh, T. 2013. PrefLib: A Library of Preference Data HTTP://PREFLIB.ORG. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT 2013)*, Lecture Notes in Artificial Intelligence. Springer.

May, K. O. 1952. A set of independent necessary and sufficient conditions for simple majority decisions. *Econometrica*, 20(4): 680–684.

Osborne, D. K. 1976. Irrelevant Alternatives and Social Welfare. *Econometrica*, 44(5): 1001–1015.

Pearce, D. 2021. Individual and social welfare: a bayesian perspective. In *Frisch Lecture delivered to the World Congress of the Econometric Society*.

Procaccia, A. D.; and Rosenschein, J. S. 2006. The distortion of cardinal preferences in voting. In *International Workshop on Cooperative Information Agents*, 317–331. Springer.

Sen, A. 1970. *Collective Choice and Social Welfare*. Harvard University Press.

Sen, A. K. 1971. Choice Functions and Revealed Preference. *The Review of Economic Studies*, 38(3): 307–317.