EFFICIENT ELICITATION OF COLLECTIVE DISAGREEMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper focuses on how to measure disagreements in a population that faces alternative proposals, be them political candidates, policy issues, or projects to be funded. Disagreement measures require increasing levels of information to be elicited from individuals, from one single pairwise comparison to complete rankings over the alternatives. In this paper, we propose a stratified framework that classifies disagreement measures according to this information, allowing us in turn to design efficient elicitation protocols for their computation. Specifically, we introduce the plurality matrix, an aggregated preference representation that records, for every subset S of alternatives, the probability that each $a \in S$ ranks first in S. We define the level of a disagreement measure as the smallest subset size needed to compute it. We prove that rank variance and divisiveness require exactly level 3 and that the hierarchy is strict, with skewness and excess kurtosis at levels 4 and 5, respectively. Electoral data illustrate the contrasting rank distributions that these higher moments describe. To make our results practically actionable, we design two elicitation protocols to estimate the various levels of the plurality matrix, exploring the trade-off between the number of required participants and the cognitive load requested to each of them.

1 INTRODUCTION

Recent years have seen a rise in participatory democracy platforms aimed at deliberating and collecting preferences about policy proposals (see, e.g., Pol.is (Small et al., 2021) or Remesh.ai). While many such initiatives are aimed at spotting the most-agreed upon alternatives, some of them focused at identifying the disagreements among the group of users (Navarrete et al., 2024). Such a process is crucial for deliberation, since identifying disagreements is the first step to resolve them (Waldron, 1999). However, a solid theoretical study of disagreement measures, and how to elicit them, seems to be lacking. The closest related field is social choice theory (Arrow et al., 2010; Brandt et al., 2016b; Fürnkranz & Hüllermeier, 2011). However, a large part of this literature focuses on identifying agreements, by for instance selecting the most preferred alternative(s) from the agents' declared preferences, characterizing such rules axiomatically, and designing efficient elicitation protocols for their computation. Disagreement itself has received comparatively little attention. Participatory and deliberative platforms, on the other hand, report it explicitly: as an example, Pol.is presents the proposals with the narrowest voting margins alongside the consensual ones (Small et al., 2021; 2023).

We model this problem as a set of individuals who can be queried about their preferences over a number of alternatives, with the objective of identifying the most divisive ones. Assuming individuals can be queried with pairwise comparisons, the most common data structure to represent their aggregated preferences is the weighted majority graph, which stores, for each pair of alternatives, the relative support of one over another. However, the majority graph hides the most important disagreements. This can be seen in a simple example. Consider two populations over three alternatives a, b and c: in the first, the six possible rankings occur with equal frequency; in the second, half of the participants report $a \succ b \succ c$ and the other half report $c \succ b \succ a$. All entries of the weighted majority graph are equal to $1/2$ in both populations, so no measure defined on these proportions can tell them apart. Their structures of disagreement differ nonetheless: the rank of a is uniform over $\{1, 2, 3\}$ in the first population, whereas it takes the values 1 and 3 with equal probability in the second.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

To overcome this problem, we propose a stratified data structure that generalizes the majority graph and that is based on plurality quantities (initially defined by Block & Marschak (1959)), which store the probability that a user would choose a specific alternative in a set of fixed size. This representation allows us to classify a number of disagreement measures from the literature (and classical agreement measures too) on the level of information they require to be elicited from users. Computing a measure of level 2 in our hierarchy (which corresponds to the weighted majority graph) can be done by asking each participant a single pairwise comparison, provided the population is large enough for every pair of alternatives to be sampled by sufficiently many participants. In contrast, computing a measure over complete rankings amounts to $\Theta(m \log m)$ comparisons over m alternatives. To the best of our knowledge, disagreement measures have never been compared under the lens of their elicitation burden: our framework allows to identify and chart the tradeoff between the number of users that needs to be queried and their cognitive load, measured in number of pairwise comparisons elicited.

Contributions. Building on the choice probabilities of Block & Marschak (1959), we characterize the minimum information levels required to compute disagreement measures and connect these requirements to the effort demanded from each participant. Our main contributions are as follows:

- We organize choice probabilities into the *plurality matrix*, whose entry for a subset S and an alternative $a \in S$ records the probability that a is ranked first within S. Grouping entries by subset size defines different *degrees* of information, with the majority graph corresponding to degree 2.
- We establish that the minimum degrees required to compute the *agreement index* (Faliszewski et al., 2023), the *rank variance* (Kendall & Smith, 1939), and the divisiveness measure of Navarrete et al. (2024) are 2, 3, and 3, respectively.
- We prove that the hierarchy is strict: for every $k \geq 2$, there exist preference profiles that agree on all information up to degree k but differ at degree $k+1$. Moreover, we construct disagreement measures whose minimum required degree is exactly k, for every $k \geq 2$.
- We show that the hierarchy collapses to degree 2 under preferences that are single-peaked on a known common axis or follow the Plackett–Luce model.
- We propose two elicitation protocols and characterize their trade-off between per-voter cognitive load and the number of participants required to achieve a given estimation accuracy.

Related work. The related literature spans three complementary directions: measuring disagreement in social choice, enhancing elicitation beyond pairwise comparisons, and designing cost-efficient elicitation protocols. First, a number of works in social choice have studied how to define measures orthogonal to agreement. Existing approaches typically capture either the diversity of voters' preferences (Ammann & Puppe, 2025; Faliszewski et al., 2026; Hashemi & Endriss, 2014; Karpov, 2017) or polarization (Can et al., 2015; Faliszewski et al., 2023). These notions are generally defined at the level of full preference profiles and often rely on complete rankings. Closer to our perspective are measures defined at the level of alternatives, such as the divisiveness measure of Colley et al. (2023b); Navarrete et al. (2024), and the axiomatic analysis of conflicting pairs by Delemazure et al. (2024). Second, several works consider richer query models than pairwise comparisons for preference elicitation and are closely related to our framework. In particular, subset-wise queries, where a set of alternatives is presented and the best one is returned, have been studied in learning and ranking settings (Saha & Gopalan, 2019). Related models include top-k queries, which extract more information per interaction (Ayadi et al., 2022; Chen et al., 2018), as well as recent work on eliciting full rankings over queried subsets of alternatives (Halpern et al., 2024). Third, a complementary line of work studies elicitation protocols under cognitive constraints and relates to our goal of designing viable alternatives to full ranking elicitation. In voting, the recent work of Terzopoulou (2023) introduces the notion of voters' energy in reporting preferences, capturing the limitation on the number of alternatives that can be ranked by each voter, and studies the resulting loss in social welfare for plurality and Borda rules. More broadly, behavioral decision theory highlights the impact of cognitive load and task complexity on response quality (Griffin et al., 2005), motivating the design of hybrid or adaptive elicitation formats that balance informational richness with practical usability (Brams & Sanver, 2009; Dery, 2024).

2 The Model

This section is devoted to introducing a probabilistic model of agents' preferences over alternatives, to define the *plurality matrix*, and to define three disagreement measures as our use-cases.

Let $\mathcal{A} = \{a, b, \ldots\}$ be a finite set of $m \geq 3$ alternatives, Π_m the set of all $m!$ strict rankings of $\mathcal{A}$, and $\mathcal{S} := \{S \subseteq \mathcal{A} : 2 \leq |S| \leq m\}$ (sorted by size, then lexicographically). For a statement τ, $\mathbb{1}\{\tau\}$ denotes the indicator of τ. A **preference profile** (or voter population) is a probability distribution π on Π_m, and a **voter** corresponds to a ranking $\succ \in \Pi_m$ sampled from π.

We adopt a probabilistic view as many applications, e.g. online platforms, involve in general a large number of participants. We now introduce one of the main concept of our paper, the *plurality matrix*, which will serve as the basis for first analyzing disagreement and later designing elicitation protocols.

Definition 2.1. Let π be a preference profile. We define the **plurality matrix** $\mathcal{P}_\pi \in [0,1]^{\mathcal{S}\times\mathcal{A}}$, where, for each $S \in \mathcal{S}$ and $a \in \mathcal{A}$, the entry (S, a) is given by,

$$p^\pi_S(a) := \mathbb{P}_\pi(a \text{ is ranked first among the alternatives in } S) \cdot \mathbb{1}\{a \in S\}.$$

For simplicity, we write p^π_{xywz} instead of $p^\pi_{\{x,y,w,z\}}$. Given $k \in \{2, ..., m\}$, we refer to the **data or information of degree** k to the sub-matrix containing only the rows related to sets of size k. We say a real-value statistic Φ is of **level** k if for any preference profile π and alternative a, $\Phi(\pi, a)$ can be expressed using data of degree at most k but not at most $k-1$.

We assume throughout the article that $p^\pi_S(a) > 0$ for every profile π, every $S \in \mathcal{S}$, and $a \in S$: an alternative violating this is dominated on every voter's ranking, hence never a locus of disagreement. For simplicity, we make one exception in Example 2.3 below. We write $p^\pi_S(a) = \cdot$ whenever $a \notin S$.

Definition 2.2. Let $r_a(\succ)$ denote the **rank** of a in $\succ$, with 1 best. Its **Borda score** under π is $\mathrm{Bor}_\pi(a) := \sum_{b\in\mathcal{A}\setminus\{a\}} p^\pi_{ab}(a) = m - \mathbb{E}_\pi[r_a]$. For the subpopulation preferring x to y, write $\mathrm{Bor}_\pi(a; \mathcal{N}^{x\succ y}_\pi) := m - \mathbb{E}_\pi[r_a \mid x \succ y]$.

The weighted majority graph encodes level 2 of the plurality matrix; thus Borda scores and many voting rules are of level 2 (Brandt et al., 2016a; Fischer et al., 2016; Navarrete et al., 2024). Degree-m entries give plurality scores: the proportions of voters ranking each alternative first. We next illustrate the plurality matrix and provide initial insights.

Example 2.3. Let $\mathcal{A} = \{a, b, c\}$ and consider two preference profiles:[1] the impartial-culture profile π_{IC},[2] where all alternatives are uniformly ranked, and the antagonism profile π_{AN}, where b and c are uniformly ranked but a is ranked either first or last, each with probability $1/2$. Formally, r_a verifies,

$$\mathbb{P}_{\pi_{\mathrm{IC}}}[r_a = i] = \tfrac{1}{3}, \text{ for } i \in \{1,2,3\} \text{ and } \mathbb{P}_{\pi_{\mathrm{AN}}}[r_a = i] = \tfrac{1}{2}\cdot\mathbb{1}\{i=1\} + \tfrac{1}{2}\cdot\mathbb{1}\{i=3\}.$$

Table 1 shows the plurality matrix for the two aforementioned profiles. Notice that the two matrices

Table 1: Plurality matrices for π_{IC} and π_{AN} over $\mathcal{A} = \{a, b, c\}$

$\mathcal{P}_{\pi_{\mathrm{IC}}}$			
$\mathcal{S}\times\mathcal{A}$	a	b	c
$\{a,b\}$	$1/2$	$1/2$	$\cdot$
$\{a,c\}$	$1/2$	$\cdot$	$1/2$
$\{b,c\}$	$\cdot$	$1/2$	$1/2$
$\{a,b,c\}$	$1/3$	$1/3$	$1/3$

$\mathcal{P}_{\pi_{\mathrm{AN}}}$			
$\mathcal{S}\times\mathcal{A}$	a	b	c
$\{a,b\}$	$1/2$	$1/2$	$\cdot$
$\{a,c\}$	$1/2$	$\cdot$	$1/2$
$\{b,c\}$	$\cdot$	$1/2$	$1/2$
$\{a,b,c\}$	$1/2$	$1/2$	0

coincide on all entries of degree 2 but differ at degree 3. Any measure based solely on the weighted majority graph, therefore, will not be able to distinguish them. Yet, both profiles clearly differ in terms of disagreement.

The plurality matrix will be shown to be rich enough to express disagreement notions studied in this paper (as well as several pairwise-based social choice concepts), yet simple enough to define efficient elicitation protocols under limited cognitive load. Moreover, by working with aggregate proportions, we adopt a privacy-preserving approach that does not require storing individual preferences.

In the rest of the article, unless required, we will drop the dependence of all defined terms on the preference profile π for simplicity.

[1] Note the two profiles sit at opposite poles of the visualization compass of Szufa et al. (2025).

[2] Preference profiles in our terminology are typically called *cultures* or *distributions* in the social choice literature. In this example we consider the impartial culture. Figure 3 includes profiles from other distributions.

2.1 Three Measures of Disagreement

We introduce three measures of disagreement from the literature to be the use-cases of the article, namely, the *agreement index* of Faliszewski et al. (2023), the *rank variance* of Kendall & Smith (1939), and the *divisiveness* of Navarrete et al. (2024).

Definition 2.4. Given a profile π and an alternative $a \in \mathcal{A}$, define the **agreement index** $A(\pi)$, the **rank variance** $\mathrm{Var}_\pi(a)$, and the **divisiveness** $\mathrm{Div}_\pi(a)$ by

$$A(\pi) := \frac{1}{\binom{m}{2}} \sum\nolimits_{\{x,y\}\subseteq\mathcal{A}} \left|2 \cdot p_{xy}(x) - 1\right|, \qquad \mathrm{Var}_\pi(a) := \mathbb{E}_\pi[(r_a - \mathbb{E}_\pi[r_a])^2],$$

$$\mathrm{Div}_\pi(a) := \frac{1}{m-1} \sum\nolimits_{b\in\mathcal{A}\setminus\{a\}} \left|\mathrm{Bor}_\pi(a; \mathcal{N}_\pi^{a\succ b}) - \mathrm{Bor}_\pi(a; \mathcal{N}_\pi^{b\succ a})\right|.$$

The agreement index aggregates pairwise comparisons into a scalar ranging from 0 (all pairs perfectly split) to 1 (unanimity), while rank variance quantifies fluctuations in the rank of a across the population. Following Navarrete et al. (2024), divisiveness compares the Borda score of a between voters preferring a to b and those preferring b to a, for each $b \neq a$; other scoring functions, such as Copeland, could also be used. We illustrate these measures below.

Example 2.5. Consider an instance with $|\mathcal{A}| = 15$ alternatives and $a \in \mathcal{A}$ fixed. Consider the two preference profiles π_{IC} and π_{AN} as defined in Example 2.3, extended to 15 alternatives. In addition, consider the preferences profiles $\pi_{\min}$ and π_{sym} such that, for any $b \in \mathcal{A} \setminus \{a\}$, the rank of b is uniformly distributed, and, given $\varepsilon_1, \varepsilon_2 > 0$,

$$\mathbb{P}_{\pi_{\min}}[r_a = i] = \varepsilon_1 \cdot \mathbb{1}\{i = 1\} + (1 - \varepsilon_1) \cdot \mathbb{1}\{i = 15\},$$
$$\mathbb{P}_{\pi_{\mathrm{sym}}}[r_a = i] = \varepsilon_2 \cdot \mathbb{1}\{i = 1\} + \varepsilon_2 \cdot \mathbb{1}\{i = 15\} + (1 - 2\varepsilon_2) \cdot \mathbb{1}\{i = 8\}.$$

In words, in π_{IC} all rankings are equally likely. In π_{AN}, half of the population ranks a first while the other half ranks it last. In $\pi_{\min}$, most of the population ranks a last while a small portion ranks it first. Finally, in π_{sym}, small populations rank a respectively first and last, while most of the voters rank it in the middle. Figure 1 illustrates the four distributions.



Figure 1: Distribution of r_a for each profile represented as probability mass functions.

Table 2 shows the metrics on the four profiles. As expected, the agreement index struggles to differ a uniform profile from profiles with split populations, while the other metrics manages to do it. However, rank variance is maximized at π_{AN}, as a is equally often ranked first or last, while divisiveness is maximized for both π_{AN} and $\pi_{\min}$, showing that only the presence, not the frequency, of extreme ranks might matter. Interestingly, for π_{IC} and π_{sym}, both metrics coincide.

Table 2: Agreement index, rank variance, and divisiveness of four profiles, with $\varepsilon_1 = 0.05$ and $\varepsilon_2 = \frac{4}{21}$.

	π_{IC}	π_{AN}	$\pi_{\min}$	π_{sym}
$A(\pi)$	0	0	0.12	0
$\mathrm{Var}_\pi(a)$	$18.\bar{6}$	49	9.31	$18.\bar{6}$
$\mathrm{Div}_\pi(a)$	$5.\bar{3}$	14	14	$5.\bar{3}$

3 The Plurality Hierarchy

This section is dedicated to study the hierarchy of the different degrees information of the plurality matrix. We begin by determining the level of the disagreement measures introduced in Section 2.1, complemented by an extensive analysis in Appendix D of a large number of agreement, disagreement, and polarization measures from the literature. We then prove that the induced level hierarchy is strict and that it collapses under standard structural assumptions on voters' preferences.

3.1 Level of Disagreement Measures

Our first result is a straighforward corollary of the definition of the agreement index.

Proposition 3.1. *The Agreement Index $A(\pi)$ is a measure of level 2.*

Regarding the rank variance and the divisiveness, we prove two important results: first, they are fully captured by our framework and second, they require entries of degree strictly higher than two.

Proposition 3.2. *The rank variance is a measure of level 3. Moreover, for any preference profile π and alternative $a \in \mathcal{A}$, it holds,*

$$\mathrm{Var}(a) \;=\; \sum_{b\in\mathcal{A}\setminus\{a\}} p_{ab}(a)\cdot(1-p_{ab}(a)) \;+\; \sum_{b,c\in\mathcal{A}\setminus\{a\},b\neq c} \big(p_{abc}(a)-p_{ab}(a)\cdot p_{ac}(a)\big).$$

Proposition 3.3. *The divisiveness is a measure of level 3. Moreover, for any preference profile π and alternative $a \in \mathcal{A}$, it holds,*

$$\mathrm{Div}(a) = \frac{1}{m-1}\sum_{b\in\mathcal{A}\setminus\{a\}}\left|1+\sum_{c\in\mathcal{A}\setminus\{a,b\}}\left[p_{abc}(a)\cdot\left(\frac{1}{1-p_{ab}(a)}+\frac{1}{p_{ab}(a)}\right)-\frac{p_{ac}(a)}{1-p_{ab}(a)}\right]\right|.$$

The proofs of Propositions 3.2 and 3.3 are deferred to Appendix B. Note that, since rank variance and divisiveness can be expressed with entries of degree 2 and 3, both are of level at most 3. To show tightness, consider π_{IC} and π_{AN} as in Example 2.5: they coincide at degree 2 ($p_{xy}(x) = 1/2$ for any $x, y \in \mathcal{A}$), yet $\mathrm{Var}_{\pi_{\mathrm{IC}}}(a) \neq \mathrm{Var}_{\pi_{\mathrm{AN}}}(a)$ and $\mathrm{Div}_{\pi_{\mathrm{IC}}}(a) \neq \mathrm{Div}_{\pi_{\mathrm{AN}}}(a)$, as shown in Table 2. Hence, neither measure can be expressed using only information of degree 2.

The previous observation actually extends beyond level 3: for any integer $k \geq 2$, there exist two profiles that coincide on all degrees from 2 to k, but differ on at least one entry of degree $k+1$. In particular, although the two profiles are different, no measure of level k can distinguish them.

Proposition 3.4. *For every $k \geq 2$, there exists $m \in \mathbb{N}$, a set $\mathcal{A}$ of m alternatives, and two preference profiles π, π' such that $p^{\pi}_{S}(a) = p^{\pi'}_{S}(a)$ for any $S \subseteq \mathcal{A}$ with $|S| \leq k$ and $a \in \mathcal{A}$, while $p^{\pi}_{T}(a) \neq p^{\pi'}_{T}(a)$ for some $T \subseteq \mathcal{A}$ with $|T| = k+1$ and some $a \in \mathcal{A}$. In particular, no measure of level k distinguishes π from π'.*

The proof of Proposition 3.4 (in Appendix B) constructs the two stated profiles π and π'. However, the question of whether a measure of level k, for any $k \in \mathbb{N}$, actually exists, rises. The answer is *yes*.

Definition 3.5. Let $\mathcal{A}$ be a set of m alternatives. For any preference profile π and $k \in \{1, ..., m\}$, define the **k-central moment** as $M^{\pi}_{k}(a) := \mathbb{E}_{\pi}\big[(r_a - \mathbb{E}[r_a])^k\big]$, for any $a \in \mathcal{A}$.

The following result generalizes Proposition 3.2 to any degree. The proof is included in Appendix B.

Theorem 3.6. *The k-central moment M_k is a measure of level $k+1$. Moreover, for any profile π and alternative $a \in \mathcal{A}$, it holds,*

$$M_k(a) = (-1)^k \sum_{s=0}^{k} c_s(a)\cdot \sum_{S\subseteq\mathcal{A}\setminus\{a\},|S|=s} p_{S\cup\{a\}}(a) = (-1)^k \sum_{s=0}^{k} c_s(a)\cdot \mathbb{E}_{\pi}\left[\binom{m-r_a}{s}\right],$$

where $c_s(a) := \sum_{j=0}^{s}(-1)^{s-j}\binom{s}{j}(j-\mathrm{Bor}(a))^k$ and $p_a(a) = 1$.

The k-central moments, for different values of k, measure different properties of the distributions of the alternatives' ranks. Notably, given a preference profile π, for $k = 3$ and $k = 4$, we obtain, respectively, the **skewness** γ^{π}_{1} and the **excess kurtosis** γ^{π}_{2}, formally given by,

$$\gamma^{\pi}_{1}(a) := \frac{M^{\pi}_{3}(a)}{(M^{\pi}_{2}(a))^{3/2}} \text{ and } \gamma^{\pi}_{2}(a) := \frac{M^{\pi}_{4}(a)}{(M^{\pi}_{2}(a))^{2}} - 3.$$

The skewness measures the asymmetry of the distribution of r_a while the excess kurtosis relates to tail extremity, reflecting the tendency of the distribution of r_a to produce outliers. From Proposition 3.4 and Theorem 3.6, γ^{π}_{1} is a measure of level 4 while γ^{π}_{2} is a measure of level 5. In particular, they give a two-dimensional summary of the rank distributions that no measure of level 2 and 3 can capture.

3.2 DISAGREEMENT BEYOND LEVEL 3

To further motivate the interest of disagreement measures of level higher than 3, we present an experiment on electoral data where four real-world candidates are divisive in ways that can be differentiated only with level-4 and level-5 measures, in this case skewness and excess kurtosis. We use electoral data from the Glasgow STV city council elections (O'Neill, 2013), the New South Wales Legislative Assembly elections (State of New South Wales through NSW Electoral Commission, 2018), and the 2017 Voter Autrement survey (Baujard et al., 2025), comprising 208, 201, and 22 candidate observations, respectively. We restrict the analysis to complete rankings. Each point in Figure 2 corresponds to a candidate $x \in \mathcal{A}$ under the empirical preference profile π of the corresponding election. Its horizontal coordinate is the skewness $\gamma_1^\pi(x)$, and its vertical coordinate is the excess kurtosis $\gamma_2^\pi(x)$. These coordinates are computed from the distribution of r_x, where $r_x = 1$ means that x is ranked first. Thus, the figure compares the shapes of candidates' rank distributions across the three datasets.

Four highlighted points link skewness and excess kurtosis to individual rank distributions: Q_1 (Hamilton Ross), Q_2 (Mark Speakman), Q_3 (Marine Le Pen), and Q_4 (Jean-Luc Mélenchon). Each inset histogram shows the frequencies of ranks $j \in \{1, \ldots, |\mathcal{A}|\}$ for candidate x under π. At Q_3 and Q_4, opposite signs of $\gamma_1^\pi(x)$ reflect opposite asymmetries: tails representing minorities at favorable and unfavorable ranks, respectively, visible in the histograms and distinguished by degree-4 information. At Q_2, the histogram explicitly reveals two separated groups through two peaks, consistent with proximity to the Pearson boundary. The degree-5 quantity $\gamma_2^\pi(x)$ complements skewness in describing this distribution's shape. Further, Figure 3 in Appendix C.1 presents four synthetic profiles with identical plurality quantities of degree 3, yet distinct structures of disagreement that is detectable using skewness and excess kurtosis (measures of levels 4 and 5, respectively).

Different patterns of disagreements, as the one exemplified above, may lead to different interpretation and actions in a participatory process. To unravel or better understand disagreements in an asymmetric distribution such as Q_3 and Q_4, it is necessary to involve the various minorities forming the tail into the deliberative process, whereas in presence of two separate groups of favorable and unfavorable evaluations such as in Q_2 the debate can focus on the two main camps.

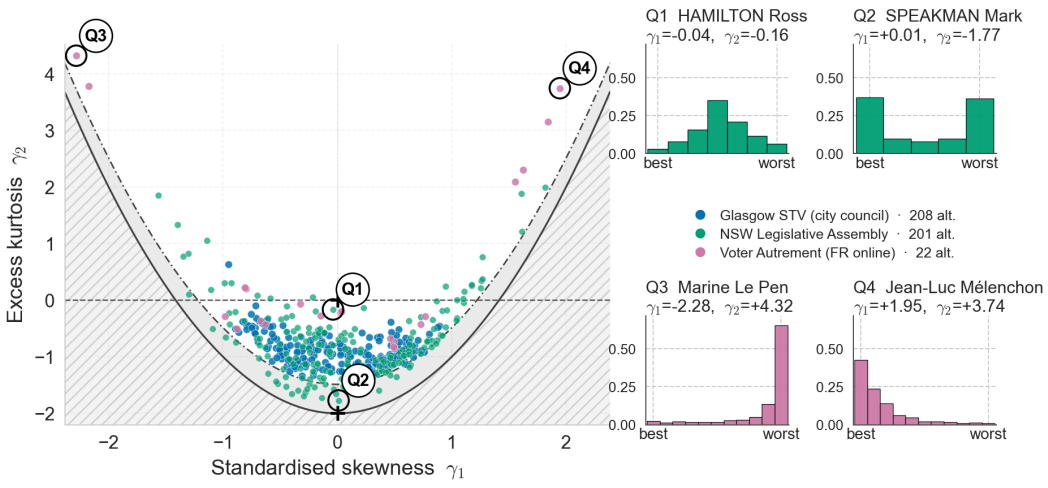


Figure 2: Pearson moment plane for three real-election datasets: Glasgow STV city council elections (208 alternatives), New South Wales Legislative Assembly (201 alternatives), and the 2017 Voter Autrement online survey (22 candidates). Solid and dash-dotted curves are the bounds from Figure 3, with the upper envelope set for the largest m. Four highlighted candidates illustrate the four regions of the plane: Q1 (Hamilton Ross, NSW), Q2 (Speakman Mark, NSW), Q3 (Marine Le Pen, Voter Autrement), Q4 (Jean-Luc Mélenchon, Voter Autrement). Inset histograms show the rank distribution of each highlighted candidate. The solid curve is the Pearson bound $\gamma_2 = \gamma_1^2 - 2$, with equality for distributions supported on exactly two ranks. The dash-dotted curve $\gamma_2 = \gamma_1^2 - 186/125$ bounds continuous unimodal distributions (Klaassen et al., 2000).

3.3 Disagreement Under Structural Assumptions

The strictness of the hierarchy (Theorem 3.6) shows that higher-degree information cannot, in general, be recovered from lower-degree observations. However, we identify two preference domains, i.e., assumptions on the structure of the individual preferences, for which the hierarchy collapses to level 2. We focus on two classical models representing two distinct types of structure: the *Plackett-Luce* model (Luce et al., 1959; Plackett, 1975), a probabilistic latent-utility model, where each alternative has a strength $v_a > 0$, and rankings are generated by sequential sampling without replacement, with probabilities proportional to these strengths; and *single-peaked preferences* (Black, 1948; Moulin, 1980), where alternatives are ordered on a fixed one-dimensional axis and each voter's preferences are monotone away from their personal peak. The proof of the following result is Appendix B.

Proposition 3.7. *Let $\mathcal{A}$ be a set of alternatives. Consider π_{PL} a preference profile following the Plackett-Luce model. Then, for any $S \subseteq \mathcal{A}$ and $a \in S$, it follows $p_S^{\pi_{PL}}(a) = v_a / \sum_{x \in S} v_x$.*

Consider π_{SP} a preference profile having in its support only preferences single-peaked over an axis $(s_1, ..., s_m)$. Then, for any $S = \{s_{i_1}, ..., s_{i_k}\} \subseteq \mathcal{A}$ sorted along the axis and $j \in \{1, ..., k\}$, it follows,

$$p_S^{\pi_{SP}}(s_{i_j}) = p_{s_{i_j} s_{i_{j+1}}}^{\pi_{SP}}(s_{i_j}) - p_{s_{i_{j-1}} s_{i_j}}^{\pi_{SP}}(s_{j-1}), \text{ where } p_{s_{i_0} s_{i_1}}^{\pi_{SP}}(s_1) := 0 \text{ and } p_{s_{i_k} s_{i_{k+1}}}^{\pi_{SP}}(s_k) := 1.$$

In particular, for each case, for any $k \geq 3$, the information of degree k can be obtained as a function of the information of degree 2.

Note that, although both structural assumptions *collapse* the hierarchy to level 2, single-peakedness requires knowing the axis in advance, which is rarely the case in practice (Conitzer, 2009). In contrast, in the Plackett–Luce model, the latent strengths $(v_a)_{a \in \mathcal{A}}$ need not to be accessed or estimated, thus, eliciting pairwise comparisons is sufficient to recover all higher-degree quantities.

4 Elicitation Protocols for the Plurality Matrix

This section is devoted to studying elicitation protocols to approximate the plurality matrix. In particular, we adopt a PAC approach, that is, to find ε-approximations of each matrix entry with probability $1 - \delta$. We focus on adaptive protocols based on pairwise comparisons, and chart the tradeoff between the number of participants and the cognitive load as a function of the required level of information in the plurality matrix.

4.1 Elicitation Primitives and Protocols

Estimating the information of degree 2 of the plurality matrix trivially reduces to querying pairwise comparisons. For $|S| \geq 3$, estimating $p_S^{\pi}(a)$ is no longer direct, and the choice of elicitation primitive becomes nontrivial. We consider two primitives, described as sequences of pairwise comparisons.[3]

Definition 4.1. Let $S = \{a_1, \ldots, a_{|S|}\} \subseteq \mathcal{A}$ be a set of alternatives.

1. We define the **S-chain** as the primitive that performs $|S| - 1$ sequential pairwise comparisons: compare a_1 and a_2, then iteratively compare the running winner with a_j for $j \in \{3, \ldots, |S|\}$.
2. We define the **S-ranking** as the primitive that elicits a full ranking of S via pairwise comparisons.

Based on these elicitation primitives, we design two elicitation *protocols*. For this, given a profile π, we denote $\mathcal{V}_n^{\pi}$ a set of n voters drawn independently from π.

Definition 4.2. Given $N \leq n$ and $k \leq m$, a **k-chain (resp. k-ranking) protocol of length N** samples uniformly with replacement N subsets $S_1, ..., S_N$ each of size k, with $S_i \subseteq \mathcal{A}$, and uniformly without replacement N voters $\{v_1, ..., v_N\}$ from $\mathcal{V}_n^{\pi}$, and for each $i \in \{1, ..., N\}$, asks voter v_i to answer the S_i-chain (respectively, S_i-ranking).

Remark 4.3. Protocols in Definition 4.2 differ in *cognitive load* since an S-chain requires $|S| - 1$ pairwise comparisons while an S-ranking requires $\Theta(|S| \log |S|)$ (e.g. when using merge sort).

With this in mind, we define two cost axes to assess our protocols.

[3] Using pairwise comparisons as elicitation primitives allows us to compare protocols on their cognitive load, since pairwise comparisons are a standard proxy for cognitive load in elicitation studies (Conitzer, 2009).

Definition 4.4. Given a protocol of length N, let c_j be the number of pairwise comparisons performed at step j. We define the **budget** B and the **maximum cognitive load** λ of the protocol as:

$$B := \sum_{j=1}^{N} c_j \text{ and } \lambda := \max_{j \in \{1,\dots,N\}} c_j.$$

Estimating the plurality matrix involves a trade-off between budget and maximum cognitive load. Queries range from single pairwise comparisons to full rankings: simpler queries require more voters for a given accuracy, while richer queries can reduce their number at the cost of greater individual effort. Under a fixed budget, this becomes a trade-off between the number of queried voters and maximum cognitive load.

4.2 Estimating the Plurality Matrix

Denote Q_k the total number of non-zero plurality matrix entries of degree k to be estimated. It follows that $Q_k = k\binom{m}{k}$. In order to estimate these values, we will look at the aggregated number of samples we require over the entries. Applying Hoeffding's inequality and a union bound (see Appendix **??**), we obtain $T_k = \ln(2Q_k/\delta)/(2\varepsilon^2)$ as a sufficient threshold on the number of independent observations per entry for simultaneous accuracy ε with probability at least $1-\delta$.

With this in mind, we can compare our two elicitation protocols with respect to the number of samples they produce. An S-chain yields $|S|-1$ samples, one per degree, on the nested prefixes $p_{a_1a_2}, p_{a_1a_2a_3}, \dots, p_{a_1\dots a_{|S|}}$. Additional comparisons may be inferred by transitivity, but naively aggregating these preference-dependent deductions can introduce selection bias (see Appendix A.1). Conversely, an S-ranking reveals the voter's preference restricted to S which, by transitivity, determines the winner of every set $T \subseteq S$, yielding $\binom{|S|}{k}$ samples at each degree $k \leq |S|$, totaling $2^{|S|} - |S| - 1$ values per voter. Clearly, this connects with Remark 4.3 regarding the cognitive load of each primitive.

The following result formally states the previous discussion. Its proof and a more detailed analysis is included in Appendix B.7.

Theorem 4.5. *Let $\mathcal{A}$ contain m alternatives, fix $2 \leq k \leq m$ and $\varepsilon, \delta \in (0,1)$, and set*

$$H_k := \frac{\ln(2Q_k/\delta)}{1 - e^{-2\varepsilon^2}}.$$

For the protocols of Definition 4.2, the following voter counts suffice to estimate all degree-k entries simultaneously within ε, with probability at least $1-\delta$:

1. *A k-chain uses $N_{chain} := \left\lceil \binom{m}{k} H_k \right\rceil$ voters and entails maximum cognitive load $\lambda = k-1$.*
2. *For any $k' \in \{k, \dots, m\}$, a k'-ranking uses*

$$N_{rank}(k') := \left\lceil \frac{\binom{m}{k}}{\binom{k'}{k}} H_k \right\rceil$$

voters and, using merge sort, entails maximum cognitive load at most $k' \lceil \log_2 k' \rceil$.

Theorem 4.5 characterizes the trade-off induced by our protocols. In particular, for fixed values of ε and δ, the range of entries of the plurality matrix that can be accurately estimated depends on the number of available voters. In applications such as online platforms, where the number of participants is typically large, designers may favor protocols with low cognitive load. Conversely, in settings with a limited number of voters who are willing to answer more demanding queries, eliciting rankings appears to be more appropriate.

By transitivity, rankings of size $k' > k$ also provide degree-k information, potentially reducing the required number of queries at the cost of greater per-voter cognitive load. Using $\lceil \log_2(s!) \rceil$ as the comparison-cost proxy, the largest ranking size under a load limit λ^* is $k^* := \max\{s \in \{1,\dots,m\} \mid \lceil \log_2(s!) \rceil \leq \lambda^*\}$.

Clearly, as already observed in Proposition 3.4, the degree of the estimated entries of the plurality matrix will never be higher than the value k of the chain or ranking used. This connects to the maximum cognitive load of the protocol, as stated in the following result proved in Appendix B.

Table 3: Trade-off between individual effort and population size across target levels ℓ. Queries involve k alternatives; λ is the maximum number of pairwise comparisons per voter, and N is the required number of voters, reported as the median over five seeds. Results use impartial culture with $m = 10$ and matrix-entry accuracy $\varepsilon = 0.05$. Chains reduce individual effort, while larger rankings reduce population size. The example measures require entries through level ℓ, including lower-degree entries.

Protocol	Quantity	Level $\ell = 2$ Mean rank / Borda	Level $\ell = 3$ Rank variance / divisiveness	Level $\ell = 4$ Skewness	Level $\ell = 5$ Excess kurtosis
Chain $k = \ell$	λ	1	2	3	4
	Median N	31 197	112 784	189 817	208 979
Ranking $k = \ell + 1$	λ	3	5	7	10
	Median N	10 116	29 507	39 590	35 101
Ranking $k = 9$	λ	19	19	19	19
	Median N	834	1 232	1 442	1 508

Proposition 4.6. *Any elicitation protocol estimating a plurality matrix entry of degree k has a maximum cognitive load λ of value at least $k - 1$. Thus, k-chains have the minimal cognitive load among all elicitation protocols for data of degree k.*

Table 3 illustrates the trade-off between individual effort and population size. Chains obtain higher-degree information through short interactions, whereas rankings exploit transitivity to reuse each response across multiple subsets. At level 5, the median voter count falls from 208 979 for a chain with four comparisons per voter to 1 508 for a nine-alternative ranking with at most 19 comparisons per voter. Chains thus suit large populations with limited individual attention, while rankings suit smaller populations able to answer more demanding queries. Our results therefore quantify how protocol choice can accommodate different participation constraints: chains suit large populations with limited individual attention, whereas rankings suit smaller populations whose participants can answer more demanding queries.

5 Conclusions

To identify the disagreements of a population over a set of alternatives we introduce the plurality matrix, a stratified framework for organizing aggregated preference data according to subset sizes. Our definitions allow us to chart an hierarchy of disagreement measures, with notable examples such as rank variance and divisiveness both lying at level 3, showing, in particular, that neither of these measures can be computed on the weighted majority graph. We further establish that the induced hierarchy is strict, by constructing preference profiles that coincide up to level k but differ at level k+1. Moreover, for every level k, we exhibit a meaningful measure that inherently requires information from level k, showing their importance in an analysis of divisive candidates from real-world electoral data. Finally, we propose two elicitation protocols to estimate the plurality matrix levels, and chart the tradeoff they induce between the cognitive load demanded to each voter and the population size.

Many of our results can be directly applied by designers of participatory campaigns interested in implementing elicitation schemes to measure the disagreements of a population over a set of alternatives. The chosen disagreement measure implies a plurality level (Section 3.1), and an elicitation scheme can then be designed depending on the expected participation and cognitive load (Section 4.2). If structured preferences are detected, the collapse of the hierarchy allows the designers to simply sample pairwise comparisons independently (Section 3.3). One of the limitations of our work is that some disagreement measures, such as the k-Kemeny measures (Faliszewski et al., 2023), do not fit the plurality matrix. However, these measures require complete rankings to be elicited and might be hard to approximate with low cognitive load on participants. Moreover, extending our framework beyond eliciting pairwise comparisons, such as querying voters using top-k lists, approval ballots, partial orders, or including Bayesian priors over response noise and cognitive load, suggest interesting directions for future work building, e.g., on recent work on identifying bridging statements with approval data, a notion that requires identifying disagreement among groups using approval data (Blair et al., 2026).

AI use disclosure. In this work, we used generative AI tools to aid and polish the written text, to retrieve potentially related articles from the literature to our work, and to assist with the Python code developed for the experiments. All the AI-assisted work was carefully reviewed. No other use of generative AI was done for this work.

REFERENCES

Jorge Alcalde-Unzu and Marc Vorsatz. Measuring the cohesiveness of preferences: an axiomatic analysis. *Social Choice and Welfare*, 41(4):965–988, 2013. doi: 10.1007/s00355-012-0716-9.

Matt Ammann and Clemens Puppe. Preference diversity. *Review of Economic Design*, 2025.

Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura. *Handbook of Social Choice and Welfare*, volume 2. Elsevier, 2010.

Manel Ayadi, Nahla Ben Amor, and Jérôme Lang. Approximating voting rules from truncated ballots. *Autonomous Agents and Multi-Agent Systems*, 36(1):24, 2022.

Antoinette Baujard, Herrade Igersheim, and Théo Delemazure. Voter Autrement 2007 - Dataset of the In Situ Experiments. Paper available at https://hal.science/hal-04986968. Data available at 10.5281/zenodo.14990025, 2025.

Duncan Black. On the rationale of group decision-making. *Journal of political economy*, 56(1): 23–34, 1948.

Carter Blair, Jakob De Raaij, Ariel D Procaccia, Maxon Rubin-Toles, Nisarg Shah, Michelle Si, and Serena Wang. The structure of bridging. In *Proceedings of the 27th International Conference on Economics and Computation (EC)*, 2026.

Henry David Block and Jacob Marschak. Random orderings and stochastic theories of response. 1959.

Niclas Boehmer, Piotr Faliszewski, Łukasz Janeczko, Andrzej Kaczmarczyk, Grzegorz Lisowski, Grzegorz Pierczyński, Simon Rey, Dariusz Stolicki, Stanisław Szufa, and Tomasz Wąs. Guide to numerical experiments on elections in computational social choice. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2024.

S. J. Brams and M. R. Sanver. *Voting Systems that Combine Approval and Preference*. 2009.

Felix Brandt, Markus Brill, and Paul Harrenstein. Tournament solutions. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (eds.), *Handbook of Computational Social Choice*, pp. 57–84. Cambridge University Press, 2016a.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, 2016b.

James W. Bucklin. The grand junction plan of city government and its results. *The Annals of the American Academy of Political and Social Science*, 38(3):87–102, 1911.

Burak Can, Ali Ihsan Ozkes, and Ton Storcken. Measuring polarization in preferences. *Mathematical Social Sciences*, 78:76–79, 2015.

Burak Can, Ali Ihsan Özkes, and Ton Storcken. Generalized measures of polarization in preferences. Technical Report AMSE Working Paper 1734, Aix-Marseille School of Economics, 2017.

Xi Chen, Yuanzhi Li, and Jieming Mao. A nearly instance optimal algorithm for top-k ranking under the multinomial logit model. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2018.

Rachael Colley, Umberto Grandi, César Hidalgo, Mariana Macedo, and Carlos Navarrete. Measuring and controlling divisiveness in rank aggregation. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, 2023a.

Rachael Colley, Umberto Grandi, César A. Hidalgo, Mariana Macedo, and Carlos Navarrete. Measuring and controlling divisiveness in rank aggregation, 2023b.

V. Conitzer. Eliciting single-peaked preferences using comparison queries. *Journal of Artificial Intelligence Research*, 35:161–191, 06 2009.

Arthur Herbert Copeland. A 'reasonable' social welfare function. Mimeographed notes, Seminar on Applications of Mathematics to the Social Sciences, University of Michigan, Ann Arbor, 1951.

Jean-Charles de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, pp. 657–665, 1781.

Théo Delemazure, Łukasz Janeczko, Andrzej Kaczmarczyk, and Stanisław Szufa. Selecting the most conflicting pair of candidates. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2024.

L. Dery. Interactive and iterative peer assessment. In *Proceedings of the 27th European Conference on Artificial Intelligence (ECAI)*, 2024.

Piotr Faliszewski, Andrzej Kaczmarczyk, Krzysztof Sornat, Stanisław Szufa, and Tomasz Wąs. Diversity, agreement, and polarization in elections. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, 2023. doi: 10.24963/ijcai.2023/299.

Piotr Faliszewski, Krzysztof Sornat, Stanisław Szufa, and Tomasz Wąs. Diversity of structured domains via k-Kemeny scores. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 40, 2026.

Felix Fischer, Olivier Hudry, and Rolf Niedermeier. Weighted tournament solutions. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (eds.), *Handbook of Computational Social Choice*, pp. 85–102. Cambridge University Press, 2016.

Johannes Fürnkranz and Eyke Hüllermeier. *Preference Learning*. Springer, 2011.

Hugo Gilbert, Tom Portoleau, and Olivier Spanjaard. Beyond pairwise comparisons in social choice: A setwise Kemeny aggregation problem. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 2020. doi: 10.1609/aaai.v34i02.5569.

D. Griffin, W. Liu, and U. Khan. A new look at constructed choice processes. *Marketing Letters*, 2005.

Daniel Halpern, Safwan Hossain, and Jamie Tucker-Foltz. Computing voting rules with elicited incomplete votes. In *Proceedings of the 25th ACM Conference on Economics and Computation (EC)*, 2024.

Vahid Hashemi and Ulle Endriss. Measuring diversity of preferences in a group. In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI)*, 2014.

Alexander Karpov. Preference diversity orderings. *Group Decision and Negotiation*, 26(4):753–774, 2017. doi: 10.1007/s10726-017-9532-z.

John G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

Maurice G Kendall and B Babington Smith. The problem of m rankings. *The annals of Mathematical Statistics*, 10(3):275–287, 1939.

Chris AJ Klaassen, Philip J Mokveld, and Bert Van Es. Squared skewness minus kurtosis bounded by 186/125 for unimodal distributions. *Statistics & probability letters*, 50(2):131–135, 2000.

Gerald H. Kramer. A dynamical model of political equilibrium. *Journal of Economic Theory*, 16(2): 310–334, 1977. doi: 10.1016/0022-0531(77)90011-4.

R Duncan Luce et al. *Individual choice behavior*, volume 4. Wiley New York, 1959.

Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

Roger B. Myerson. Incentives to cultivate favored minorities under alternative electoral systems. *American Political Science Review*, 87(4):856–869, 1993. doi: 10.2307/2938820.

Carlos Navarrete, Mariana Macedo, Rachael Colley, Jingling Zhang, Nicole Ferrada, Maria Eduarda Mello, Rodrigo Lira, Carmelo Bastos-Filho, Umberto Grandi, Jérôme Lang, et al. Understanding political divisiveness using online participation data from the 2022 French and Brazilian presidential elections. *Nature Human Behaviour*, 8(1):137–148, 2024.

Jeffrey O'Neill. Open stv. www.OpenSTV.org, 2013. Glasgow City Council data set available at https://preflib.github.io/PrefLib-Jekyll/dataset/00008.

Robin L Plackett. The analysis of permutations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 24(2):193–202, 1975.

Aadirupa Saha and Aditya Gopalan. Active ranking with subset-wise preferences. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

Markus Schulze. A new monotonic, clone-independent, reversal symmetric, and Condorcet-consistent single-winner election method. *Social Choice and Welfare*, 36(2):267–303, 2011. doi: 10.1007/s00355-010-0475-4.

Paul B. Simpson. On defining areas of voter choice: Professor Tullock on stable voting. *The Quarterly Journal of Economics*, 83(3):478–490, 1969. doi: 10.2307/1880533.

Patrick Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48(3–4):303–312, 1961. doi: 10.1093/biomet/48.3-4.303.

Christopher T. Small, Michael Bjorkegren, Timo Erkkilä, Lynette Shaw, and Colin Megill. Polis: Scaling deliberation by mapping high dimensional opinion spaces. *Recerca: Revista de Pensament i Anàlisi*, 26(2), 2021.

Christopher T Small, Ivan Vendrov, Esin Durmus, Hadjar Homaei, Elizabeth Barry, Julien Cornebise, Ted Suzman, Deep Ganguli, and Colin Megill. Opportunities and risks of LLMs for scalable deliberation with polis. *arXiv preprint arXiv:2306.11932*, 2023.

State of New South Wales through NSW Electoral Commission. NSW Electoral Commission, 2018. URL https://elections.nsw.gov.au/copyright.

Stanisław Szufa, Niclas Boehmer, Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Drawing a map of elections. *Artificial Intelligence*, 343:104332, 2025.

Zoi Terzopoulou. Voting with limited energy: A study of plurality and borda. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2023.

Nicolaus Tideman. The single transferable vote. *Journal of Economic Perspectives*, 9(1):27–38, 1995.

T. Nicolaus Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, 1987. doi: 10.1007/BF00433944.

Jeremy Waldron. *Law and disagreement*. OUP Oxford, 1999.

Appendix Table of Contents

A Examples and omitted details **13**

A.1 Bias in chain transitivity 13

B Deferred proofs **14**

B.1 Proof of Proposition 3.2 14

B.2 Proof of Proposition 3.3 15

B.3 Proof of Proposition 3.4 16

B.4 Proof of Theorem 3.6 and consequences 17

B.5 Proof of Proposition 3.7 18

B.6 Proof of Proposition 4.6 19

B.7 Proof of Theorem 4.5 19

C Additional Experiments **20**

C.1 Synthetic Data 20

C.2 Synthetic Data 21

D Inventory of measures and rules in the S-plurality hierarchy **21**

D.1 Notation 21

D.2 Aggregation rules. 24

D.3 Rank variance. 24

D.4 Navarrete et al. Divisiveness. 24

D.5 α-Divisiveness. 24

D.6 Expected rank-distance. 24

D.7 Conditional rank-distances. 25

D.8 Discrepancy balance. 25

D.9 Conflictual rule scores. 25

E Supplementary material: Interactive moment-plane visualization **25**

A Examples and omitted details

A.1 Bias in chain transitivity

No additional *unbiased* observations can be inferred by transitivity from k-chains. Indeed, the probability that a given pair is resolved varies with the underlying ranking of each voter, inducing a selection bias. As a result, empirical frequencies computed from such inferred comparison do not, in general, estimate the true pairwise proportion. We illustrate this phenomenon on the following example.

Example A.1. Let $\mathcal{A} = \{a, b, c\}$ and

$$\pi \;=\; \tfrac{1}{2}\,\sigma_A \;+\; \tfrac{1}{2}\,\sigma_B, \qquad \sigma_A : a \succ c \succ b, \qquad \sigma_B : b \succ c \succ a,$$

so that $p_{ca} = \mathbb{P}_\pi[c \succ a] = \frac{1}{2}$. Run a 3-chain on $S = \{a, b, c\}$ with ordering τ drawn uniformly from the 6 permutations, independently of the voter. For a voter σ, record

$$Y_{ca}(\sigma, \tau) = \begin{cases} \mathbb{1}[c \succ_\sigma a] & \text{if } \{a, c\} \text{ is compared directly or } w_3 \in \{a, c\} \text{ (transitive inference),} \\ \perp & \text{otherwise,} \end{cases}$$

and estimate p_{ca} by $\hat{p}_{ca} = \mathbb{P}[Y_{ca} = 1 \mid Y_{ca} \neq \perp]$. Enumerating all 2×6 cases:

	σ_A ($w_3 = a$)		σ_B ($w_3 = b$)	
τ	resolved?	Y_{ca}	resolved?	Y_{ca}
(a, b, c)	yes (trans.)	0	no	$\perp$
(a, c, b)	yes (direct)	0	yes (direct)	1
(b, a, c)	yes (trans.)	0	no	$\perp$
(b, c, a)	yes (trans.)	0	no	$\perp$
(c, a, b)	yes (direct)	0	yes (direct)	1
(c, b, a)	yes (trans.)	0	no	$\perp$
rate	$6/6$	all 0	$2/6$	all 1

Aggregating over the uniform voter-ordering distribution gives $\mathbb{P}[Y_{ca} = 0] = \frac{1}{2} \cdot \frac{6}{6} = \frac{1}{2}$, $\mathbb{P}[Y_{ca} = 1] = \frac{1}{2} \cdot \frac{2}{6} = \frac{1}{6}$, and $\mathbb{P}[Y_{ca} \neq \perp] = \frac{2}{3}$, so

$$\hat{p}_{ca} \;=\; \frac{1/6}{2/3} \;=\; \tfrac{1}{4}, \qquad \text{bias} \;=\; \hat{p}_{ca} - p_{ca} \;=\; -\tfrac{1}{4}.$$

The estimator's observation rate depends on the voter: a voter with $\text{top}_\sigma(S) \in \{a, c\}$ always resolves $\{a, c\}$ via transitivity ($6/6$ orderings), whereas a voter with $\text{top}_\sigma(S) \notin \{a, c\}$ resolves it only when a and c are placed at positions 1–2 ($2/6$ orderings). Equal-mass voter types therefore produce unequal observation counts, and the conditional distribution of Y_{ca} given a non-$\perp$ record differs from π's marginal on $\{a, c\}$. The asymmetry is between voters, not between orderings, so uniform randomization of τ does not correct it. Only the nested prefix winners $w_2 = \text{top}_\sigma(\{x_1, x_2\})$ and $w_3 = \text{top}_\sigma(S)$ yield unbiased samples, since their reference subsets are fixed by the protocol independently of σ.

B Deferred proofs

B.1 Proof of Proposition 3.2

Proposition 3.2. *The rank variance is a measure of level* 3*. Moreover, for any preference profile* π *and alternative* $a \in \mathcal{A}$*, it holds,*

$$\text{Var}(a) \;=\; \sum_{b \in \mathcal{A} \setminus \{a\}} p_{ab}(a) \cdot (1 - p_{ab}(a)) \;+\; \sum_{b, c \in \mathcal{A} \setminus \{a\}, b \neq c} \big(p_{abc}(a) - p_{ab}(a) \cdot p_{ac}(a)\big).$$

Proof. Write $X_b := \mathbb{1}[a \succ b]$ for each $b \neq a$, so that $r_a = m - \sum_{b \neq a} X_b$. Since variance is invariant under translations and sign flips,

$$\text{Var}_\pi(r_a) \;=\; \text{Var}_\pi\Big(\textstyle\sum_{b \neq a} X_b\Big) \;=\; \mathbb{E}_\pi\Big[\Big(\sum_{b \neq a} X_b\Big)^2\Big] \;-\; \Big(\mathbb{E}_\pi\Big[\sum_{b \neq a} X_b\Big]\Big)^2$$

We now compute each term separately. Expanding the first term,

$$\mathbb{E}_\pi\Big[\Big(\textstyle\sum_{b \neq a} X_b\Big)^2\Big] = \sum_{b \neq a} \mathbb{E}_\pi[X_b^2] + \sum_{\substack{b, c \neq a \\ b \neq c}} \mathbb{E}_\pi[X_b X_c]$$

Since $X_b \in \{0, 1\}$, we have $X_b^2 = X_b$, hence

$$\mathbb{E}_\pi[X_b^2] = \mathbb{E}_\pi[X_b] = p_{ab}.$$

Moreover, for $b \neq c$,

$$X_b X_b = \mathbb{1}[a \succ b \text{ and } a \succ c],$$

so that

$$\mathbb{E}_\pi[X_b X_c] = p_{abc}(a).$$

Thus,

$$\mathbb{E}_\pi\Big[\Big(\textstyle\sum_{b\neq a} X_b\Big)^2\Big] = \sum_{b\neq a} p_{ab} + + \sum_{\substack{b,c\neq a\\ b\neq c}} p_{abc}(a).$$

Regarding the second term,

$$\Big(\mathbb{E}_\pi\Big[\textstyle\sum_{b\neq a} X_b\Big]\Big)^2 = \sum_{b,c\neq a} p_{ab}\, p_{ac}.$$

Combining the both terms yields

$$\mathrm{Var}_\pi(r_a) = \sum_{b\neq a} p_{ab} + \sum_{\substack{b,c\neq a\\ b\neq c}} p_{abc}(a) - \sum_{b,c\neq a} p_{ab}\, p_{ac}.$$

Finally, separating diagonal and off-diagonal terms in the last sum gives

$$\mathrm{Var}_\pi(r_a) \;=\; \sum_{b\neq a} p_{ab}\,(1 - p_{ab}) \;+ \sum_{\substack{b,c\neq a\\ b\neq c}} \big(\, p_{abc}(a) - p_{ab}\, p_{ac}\,\big).$$

□

B.2 Proof of Proposition 3.3

Proposition 3.3. *The divisiveness is a measure of level* 3*. Moreover, for any preference profile* π *and alternative* $a \in \mathcal{A}$*, it holds,*

$$\mathrm{Div}(a) = \frac{1}{m-1} \sum_{b\in\mathcal{A}\setminus\{a\}} \left| 1 + \sum_{c\in\mathcal{A}\setminus\{a,b\}} \left[p_{abc}(a) \cdot \left(\frac{1}{1-p_{ab}(a)} + \frac{1}{p_{ab}(a)} \right) - \frac{p_{ac}(a)}{1-p_{ab}(a)} \right] \right|.$$

Proof. We recall that, by definition,

$$\mathrm{Div}_\pi(a) := \frac{1}{m-1} \sum\nolimits_{b\in\mathcal{A}\setminus\{a\}} \big| \mathrm{Bor}_\pi\big(a; \mathcal{N}_\pi^{a\succ b}\big) \;-\; \mathrm{Bor}_\pi\big(a; \mathcal{N}_\pi^{b\succ a}\big) \big|.$$

with

$$\mathrm{Bor}_\pi\big(a; \mathcal{N}_\pi^{x\succ y}\big) := \sum\nolimits_{b\in\mathcal{A}\setminus\{a\}} \mathbb{P}_\pi(a \succ b \mid x \succ y),$$

Using the identity $r_a = 1 + \sum_{c\neq a} \mathbb{1}[c \succ a]$, and by linearity of conditional expectation, we obtain

$$\mathbb{E}_\pi[r_a \mid \mathcal{E}] \;=\; 1 + \sum_{c\neq a} \mathbb{P}_\pi[c \succ a \mid \mathcal{E}]. \tag{1}$$

Therefore,

$$\mathrm{Bor}_\pi\big(a; \mathcal{N}_\pi^{x\succ y}\big) = \sum_{c\neq a} \mathbb{P}_\pi(a \succ c \mid x \succ y) = (m-1) - \sum_{c\neq a} \mathbb{P}_\pi(c \succ a \mid x \succ y) = m - \mathbb{E}_\pi[r_a \mid x \succ y].$$

Finally, the divisiveness formula rewrites as

$$\mathrm{Div}_\pi(a) = \frac{1}{m-1} \sum_{b\neq a} \big| \mathbb{E}_\pi[r_a \mid b \succ a] - \mathbb{E}_\pi[r_a \mid a \succ b] \big|.$$

We now compute these conditional expectations.

Computing $\mathbb{E}_\pi[r_a \mid b \succ a]$: By equation 1, we have

$$\mathbb{E}_\pi[r_a \mid b \succ a] = 1 + \sum_{c \neq a} \mathbb{P}_\pi[c \succ a \mid b \succ a].$$

For $c = b$, we have $\mathbb{P}_\pi[b \succ a \mid b \succ a] = 1$.

For $c \neq a, b$, inclusion–exclusion gives

$$\mathbb{P}_\pi[b \succ a \text{ and } c \succ a] \;=\; 1 - \mathbb{P}_\pi[a \succ b \text{ or } a \succ c] \;=\; 1 - p_{ab} - p_{ac} + p_{abc}(a),$$

and dividing by $\mathbb{P}_\pi[b \succ a] = 1 - p_{ab}$ yields

$$\mathbb{P}_\pi[c \succ a \mid b \succ a] \;=\; \frac{1 - p_{ab} - p_{ac} + p_{abc}(a)}{1 - p_{ab}} \;=\; 1 + \frac{p_{abc}(a) - p_{ac}}{1 - p_{ab}}.$$

Summing over all $m - 2$ terms $c \neq a, b$ and adding the $c = b$ term gives:

$$\begin{aligned}\mathbb{E}_\pi[r_a \mid b \succ a] \;&=\; 1 + 1 + (m-2) + \frac{1}{1 - p_{ab}} \sum_{c \neq a,b} \big(p_{abc}(a) - p_{ac}\big) \\ &=\; m + \frac{1}{1 - p_{ab}} \sum_{c \neq a,b} \big(p_{abc}(a) - p_{ac}\big).\end{aligned}$$

Computing $\mathbb{E}_\pi[r_a \mid b \succ a]$: By equation 1, we have

$$\mathbb{E}_\pi[r_a \mid a \succ b] = 1 + \sum_{c \neq a} \mathbb{P}_\pi[c \succ a \mid a \succ b].$$

For $c = b$, we have $\mathbb{P}_\pi[b \succ a \mid a \succ b] = 0$.

For $c \neq a, b$,

$$\mathbb{P}_\pi[c \succ a \mid a \succ b] \;=\; 1 - \mathbb{P}_\pi[a \succ c \mid a \succ b] \;=\; 1 - \frac{\mathbb{P}_\pi[a \succ b \text{ and } a \succ c]}{p_{ab}} \;=\; 1 - \frac{p_{\{a,b,c\}}(a)}{p_{ab}}.$$

Summing over all $(m - 2)$ terms $c \neq a, b$ in equation 1 gives

$$\mathbb{E}_\pi[r_a \mid a \succ b] \;=\; 1 + (m-2) - \frac{1}{p_{ab}} \sum_{c \neq a,b} p_{abc}(a) \;=\; m - 1 - \frac{1}{p_{ab}} \sum_{c \neq a,b} p_{abc}(a).$$

Plugging these both expectation terms back to the divisiveness formula yields the desired result. □

B.3 Proof of Proposition 3.4

Proposition 3.4. *For every $k \geq 2$, there exists $m \in \mathbb{N}$, a set $\mathcal{A}$ of m alternatives, and two preference profiles π, π' such that $p^\pi_S(a) = p^{\pi'}_S(a)$ for any $S \subseteq \mathcal{A}$ with $|S| \leq k$ and $a \in \mathcal{A}$, while $p^\pi_T(a) \neq p^{\pi'}_T(a)$ for some $T \subseteq \mathcal{A}$ with $|T| = k + 1$ and some $a \in \mathcal{A}$. In particular, no measure of level k distinguishes π from π'.*

Proof. Set $m = d + 2$ and fix a focal alternative $a \in \mathcal{A}$.

Let $w = (w_1, \ldots, w_m)$ be a probability distribution on rank positions of a, and define π_w as follows: the rank of a is drawn from w, and the other $m - 1$ alternatives are then placed uniformly at random among the remaining positions.

We first compute S-plurality quantities involving a. Fix $S \ni a$ with $|S| = s \geq 2$. Conditioning on the position $r_a = j$, the event that a is top within S occurs exactly when all other $s - 1$ alternatives of S are placed below position j. Since these alternatives are distributed uniformly, this happens with probability

$$\frac{\binom{m-j}{s-1}}{m - 1s - 1}.$$

Taking expectation over r_a yields

$$p_S(a)(\pi_w) \;=\; \sum_{j=1}^{m} w_j \,\frac{\binom{m-j}{s-1}}{\binom{m-1}{s-1}}. \tag{2}$$

Thus, $p_S(a)(\pi_w)$ is a linear functional of w depending on S only through its size s. For subsets $S \not\ni a$, symmetry among non-focal alternatives implies that their S-plurality quantities depend only on s, and therefore coincide across all profiles in the family $\{\pi_w\}$.

Now let $\delta = w - w' \in \mathbb{R}^m$. From equation 2, two profiles π_w and $\pi_{w'}$ agree on all degrees $s \le d$ if and only if

$$\sum_{j=1}^{m} \delta_j \binom{m-j}{s-1} \;=\; 0 \qquad \text{for } s = \{1,\dots,d\}, \tag{3}$$

a linear system of d equations in $m = d+2$ unknowns. Viewed as a function of j, the coefficient vector of an equation given a fixed s is a polynomial of degree $s-1$ for $s \in \{1,\dots,d\}$, producing polynomials of distinct degrees. The coefficient matrix thus has rank d and the solution set of equation 3 has two free parameters. The degree-$(d+1)$ separation condition $\sum_j \delta_j \binom{m-j}{d} \neq 0$ is governed by a polynomial of degree d, which cannot be a linear combination of polynomials of strictly lower degree; it is therefore not implied by equation 3, and there exists δ^* satisfying equation 3 while violating the separation equality.

Take w uniform on $\{1,\dots,m\}$ and set $w' := w + t\,\delta^*$ for $|t|$ small. Since δ^* satisfies equation 3 at $s=1$, $\sum_j \delta^*_j = 0$, so w' still sums to 1; since w has strictly positive entries, w' stays non-negative. By construction, π_w and $\pi_{w'}$ agree at every degree $\le d$ and differ at degree $d+1$.

Witness at $d=3$, $m=5$. Take $w = \left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right)$ and $w' = \left(\frac{3}{20},\frac{7}{20},\frac{1}{20},\frac{5}{20},\frac{4}{20}\right)$, differing by $\delta^* = \frac{1}{20}\cdot(-1,3,-3,1,0)$. Substituting in equation 2 gives $p_S(a)(\pi_w) = p_S(a)(\pi_{w'}) = 1/|S|$ for every $|S| \in \{2,3\}$; at $|S|=4$ the values are $1/4$ and $19/80$. □

B.4 Proof of Theorem 3.6 and consequences

Definition B.1 (Aggregate plurality)**.** For $a \in \mathcal{A}$ and $s \ge 0$,

$$P_s(a) \;:=\; \sum_{\substack{T \subseteq \mathcal{A}\setminus\{a\} \\ |T|=s}} p_{T\cup\{a\}}(a), \qquad P_0(a) := 1.$$

Lemma B.2. *For every $a \in \mathcal{A}$ and every $s \ge 0$, $\mathbb{E}_\pi\left[\binom{W_a}{s}\right] = P_s(a)$.*

Proof. For $T \subseteq \mathcal{A}\setminus\{a\}$, the event "$a \succ b$ for every $b \in T$" has probability $p_{T\cup\{a\}}(a)$ and indicator $\prod_{b\in T} \mathbb{1}[a \succ b]$. Summing over $|T| = s$ and taking expectations, $P_s(a) = \mathbb{E}_\pi\big[\sum_{|T|=s} \prod_{b\in T} \mathbb{1}[a \succ b]\big] = \mathbb{E}_\pi\big[\binom{W_a}{s}\big]$, where the second equality holds because $\sum_{|T|=s} \prod_{b\in T} \mathbb{1}[a \succ b]$ counts the size-s subsets of $\{b : a \succ b\}$. □

Theorem 3.6. *The k-central moment M_k is a measure of level $k+1$. Moreover, for any profile π and alternative $a \in \mathcal{A}$, it holds,*

$$M_k(a) = (-1)^k \sum_{s=0}^{k} \; c_s(a) \cdot \sum_{S\subseteq\mathcal{A}\setminus\{a\},|S|=s} p_{S\cup\{a\}}(a) = (-1)^k \sum_{s=0}^{k} \; c_s(a)\cdot \mathbb{E}_\pi\left[\binom{m-r_a}{s}\right],$$

where $c_s(a) := \sum_{j=0}^{s} (-1)^{s-j}\binom{s}{j}(j - \mathrm{Bor}(a))^k$ and $p_a(a) = 1$.

Proof. Write $X_b := \mathbb{1}[a \succ b]$ for $b \neq a$, so $r_a = m - \sum_{b\neq a} X_b$ and $r_a - \mu_a = -\sum_{b\neq a}(X_b - p_{ab})$. Set $B := \mathrm{Bor}(a)$ and $g(X) := \big(\sum_{b\neq a}(X_b - p_{ab})\big)^k$, giving $M_k(a) = (-1)^k \mathbb{E}_\pi[g(X)]$.

Since each $X_b \in \{0,1\}$ and g has total degree at most k in the X_b's (it is the k-th power of a linear form), g reduces on $\{0,1\}^{m-1}$ to a multilinear polynomial $g(X) = \sum_{|T|\le k} \gamma(T) \prod_{b\in T} X_b$

supported on subsets $T \subseteq \mathcal{A} \setminus \{a\}$ of size at most k. Taking expectations, $\mathbb{E}_\pi[g(X)] = \sum_T \gamma(T)\, p_{T\cup\{a\}}(a)$ (using Lemma B.2).

By Möbius inversion on the subset lattice, $\gamma(T) = \sum_{U\subseteq T}(-1)^{|T|-|U|}\, g(e_U)$, where $e_U \in \{0,1\}^{m-1}$ is the indicator vector of U. Evaluation at $X = e_U$ gives $g(e_U) = (|U| - B)^k$, which depends on U only through $|U|$, so the Möbius formula collapses to $\gamma(T) = \sum_{j=0}^{|T|}(-1)^{|T|-j}\binom{|T|}{j}(j - B)^k = c_{|T|}(a)$ with $c_s(a)$ as stated in the theorem. Collecting by $s = |T|$, $0 \le s \le k$, yields $\mathbb{E}_\pi[g(X)] = \sum_{s=0}^{k} c_s(a)\, P_s(a)$, and multiplying by $(-1)^k$ gives equation 3.6. The right side depends on a only through $B = P_1(a)$ and on π only through $P_0(a), \dots, P_k(a)$, hence on data of degree at most $k+1$. □

Corollary B.3. *For every $k \ge 2$, the k-th central moment $M_k(a)$ has level exactly $k+1$.*

Proof. The upper bound is Theorem 3.6. For the lower bound, apply Proposition 3.4 with $d = k$: the witness profiles π, π' agree at every degree $\le k$, so $P_0(a), P_1(a), \dots, P_{k-1}(a)$ coincide on the two profiles (each is determined by degree-$\le k$ data), and in particular $\mathrm{Bor}(a) = P_1(a)$ matches, making $c_0(a), \dots, c_k(a)$ identical on both profiles. Only $P_k(a)$ differs (it depends on degree-$(k+1)$ data). Since $c_k(a) = k! \neq 0$ (the k-th forward difference of a monic degree-k polynomial), equation 3.6 gives $M_k(a)(\pi) - M_k(a)(\pi') = (-1)^k\, k! \cdot (P_k(a)(\pi) - P_k(a)(\pi')) \neq 0$. So no function of degree-$\le k$ data agrees with M_k on these two profiles. □

B.5 Proof of Proposition 3.7

Proposition 3.7. *Let $\mathcal{A}$ be a set of alternatives. Consider π_{PL} a preference profile following the Plackett-Luce model. Then, for any $S \subseteq \mathcal{A}$ and $a \in S$, it follows $p_S^{\pi_{PL}}(a) = v_a / \sum_{x\in S} v_x$.*

Consider π_{SP} a preference profile having in its support only preferences single-peaked over an axis $(s_1, ..., s_m)$. Then, for any $S = \{s_{i_1}, ..., s_{i_k}\} \subseteq \mathcal{A}$ sorted along the axis and $j \in \{1, ..., k\}$, it follows,

$$p_S^{\pi_{SP}}(s_{i_j}) = p^{\pi_{SP}}_{s_{i_j} s_{i_{j+1}}}(s_{i_j}) - p^{\pi_{SP}}_{s_{i_{j-1}} s_{i_j}}(s_{j-1}), \text{ where } p^{\pi_{SP}}_{s_{i_0} s_{i_1}}(s_1) := 0 \text{ and } p^{\pi_{SP}}_{s_{i_k} s_{i_{k+1}}}(s_k) := 1.$$

In particular, for each case, for any $k \ge 3$, the information of degree k can be obtained as a function of the information of degree 2.

Proof. Under Plackett–Luce, the probability that a voter ranks a first among S is obtained by the sequential elimination rule: at each step, the next-ranked alternative is drawn with probability proportional to its strenght. Luce's Choice Axiom gives $p_S(a) = v_a / \sum_{x\in S} v_x$ directly. Since $p_{ab} = v_a/(v_a + v_b)$, the ratio $v_a/v_b = p_{ab}/(1 - p_{ab})$ is determined by pairwise data. Fixing any reference alternative r, we recover v_a/v_r for all a, and then

$$p_S(a) = \frac{v_a/v_r}{\sum_{x\in S} v_x/v_r} = \frac{p_{ar}/(1-p_{ar})}{\sum_{x\in S} p_{xr}/(1-p_{xr})},$$

which is a function of pairwise proportions alone.

Regarding the formula for single-peaked preferences, we prove it by induction on $|S|$.

Base case ($|S| = 2$). For $S = \{s_i, s_j\}$ with $i < j$: $p_S(s_i) = p_{s_i s_j}$ and $p_S(s_j) = 1 - p_{s_i s_j}$, which matches equation 3.7.

Key lemma ($|S| = 3$). For $a < b < c$ on the axis, single-peakedness implies that b is never ranked last among $\{a, b, c\}$: every voter's peak is some element x, and if $x \le b$ then $b \succ c$ (closer to peak), while if $x \ge b$ then $b \succ a$. In either case b is not last. Therefore $\mathbb{P}(a \succ b \text{ and } c \succ b) = 0$.

The three plurality scores sum to 1 and the "b is last" event has probability zero, so:

$$\begin{aligned} p_{\{a,b,c\}}(a) &= \mathbb{P}(a \succ b \text{ and } a \succ c) = \mathbb{P}(a \succ b) - \mathbb{P}(a \succ b \text{ and } c \succ a) \\ &= p_{ab} - \mathbb{P}(c \succ a \succ b) = p_{ab} - \big(p_{ab} - p_{\{a,b,c\}}(a)\big). \end{aligned}$$

Said differently: the events "a first", "b first", "c first" partition the probability space. The event "b first" decomposes as $\{b \succ a\} \cap \{b \succ c\}$. Since "$b$ last" $= \{a \succ b\} \cap \{c \succ b\}$ has probability zero,

we have :

$$\begin{aligned} p_{\{a,b,c\}}(b) &= \mathbb{P}(b \succ a \text{ and } b \succ c) \\ &= 1 - \mathbb{P}(a \succ b) - \mathbb{P}(c \succ b) + \underbrace{\mathbb{P}(a \succ b \text{ and } c \succ b)}_{=0} \\ &= 1 - p_{ab} - (1 - p_{bc}) = p_{bc} - p_{ab}. \end{aligned}$$

Then $p_{\{a,b,c\}}(a) = p_{ab}$ and $p_{\{a,b,c\}}(c) = 1 - p_{bc}$ follow from $p_{\{a,b,c\}}(a) + p_{\{a,b,c\}}(b) + p_{\{a,b,c\}}(c) = 1$ and $p_{\{a,b,c\}}(a) + p_{\{a,b,c\}}(b) = p_{ab} + (p_{bc} - p_{ab}) = p_{bc} = 1 - p_{\{a,b,c\}}(c)$.

Inductive step. Suppose the formula holds for all subsets of size $k-1$. Let $S = \{s_{i_1} < \cdots < s_{i_k}\}$ with $k \geq 4$. For the interior element s_{i_j} with $1 < j < k$:

$$p_S(s_{i_j}) = \mathbb{P}\big(s_{i_j} \succ s_{i_{j-1}} \text{ and } s_{i_j} \succ s_{i_{j+1}} \text{ and } s_{i_j} \succ \text{all others in } S\big).$$

Single-peakedness on the axis implies two monotonicity properties for any voter:

(i) if $a < b < c$ on the axis and $a \succ b$, then $a \succ c$;

(ii) if $a < b < c$ on the axis and $c \succ b$, then $c \succ a$.

By (i), $s_{i_j} \succ s_{i_{j+1}}$ implies $s_{i_j} \succ s_{i_\ell}$ for all $\ell > j+1$; by (ii), $s_{i_j} \succ s_{i_{j-1}}$ implies $s_{i_j} \succ s_{i_\ell}$ for all $\ell < j-1$. Therefore s_{i_j} is first in S if and only if s_{i_j} is first in the three-element subset $\{s_{i_{j-1}}, s_{i_j}, s_{i_{j+1}}\}$. By the base case, this probability is $p_{s_{i_j} s_{i_{j+1}}} - p_{s_{i_{j-1}} s_{i_j}}$.

For the boundary elements: s_{i_1} is first in S iff $s_{i_1} \succ s_{i_2}$ (by (i), this implies $s_{i_1} \succ s_{i_\ell}$ for all $\ell > 2$), so $p_S(s_{i_1}) = p_{s_{i_1} s_{i_2}}$. Similarly, s_{i_k} is first iff $s_{i_k} \succ s_{i_{k-1}}$, giving $p_S(s_{i_k}) = 1 - p_{s_{i_{k-1}} s_{i_k}}$. □

B.6 Proof of Proposition 4.6

Proposition 4.6. *Any elicitation protocol estimating a plurality matrix entry of degree k has a maximum cognitive load λ of value at least $k-1$. Thus, k-chains have the minimal cognitive load among all elicitation protocols for data of degree k.*

Proof. Fix a subset S with $|S| = k$ and an alternative $a \in S$. Each pairwise comparison between elements of S creates an oriented link between two alternatives, and transitivity allows information to propagate along chains of such links.

To determine whether a is the top-ranked element of S, all alternatives of S must be reachable from a through such a chain of links; otherwise, there exists some $b \in S \setminus \{a\}$ whose relation to a can not be resolved.

Any structure connecting all k elements must contain at least $k-1$ links. Therefore, any protocol must satisfy $\lambda \geq k-1$. □

B.7 Proof of Theorem 4.5

Theorem 4.5. *Let $\mathcal{A}$ contain m alternatives, fix $2 \leq k \leq m$ and $\varepsilon, \delta \in (0,1)$, and set*

$$H_k := \frac{\ln(2Q_k/\delta)}{1 - e^{-2\varepsilon^2}}.$$

For the protocols of Definition 4.2, the following voter counts suffice to estimate all degree-k entries simultaneously within ε, with probability at least $1-\delta$:

1. *A k-chain uses $N_{chain} := \lceil \binom{m}{k} H_k \rceil$ voters and entails maximum cognitive load $\lambda = k-1$.*
2. *For any $k' \in \{k, \ldots, m\}$, a k'-ranking uses*

$$N_{rank}(k') := \left\lceil \frac{\binom{m}{k}}{\binom{k'}{k}} H_k \right\rceil$$

voters and, using merge sort, entails maximum cognitive load at most $k' \lceil \log_2 k' \rceil$.

Proof. Fix $S \subseteq \mathcal{A}$ with $|S| = k$ and $a \in S$. A k-chain observes the winner of S when its queried subset equals S. A k'-ranking observes that winner when its queried subset contains S. The respective probabilities are

$$\rho_{\text{chain}} = \frac{1}{\binom{m}{k}}, \qquad \rho_{\text{rank}} = \frac{\binom{m-k}{k'-k}}{\binom{m}{k'}} = \frac{\binom{k'}{k}}{\binom{m}{k}}.$$

Write ρ for the relevant probability. Under uniform independent subset sampling, the number of observations satisfies $R_S \sim \text{Bin}(N, \rho)$.

The selected voters are distinct members of an i.i.d. sample from π, chosen independently of preferences; their rankings are therefore independent with distribution π. Conditional on $R_S = r \geq 1$, the estimator averages r independent Bernoulli observations with mean $p_S^\pi(a)$. Hoeffding's inequality bounds its conditional error probability by $2e^{-2r\varepsilon^2}$. This bound also holds for $r = 0$, since its right-hand side is then 2.

Averaging over R_S gives

$$\begin{aligned}\mathbb{P}(|\widehat{p}_S(a) - p_S^\pi(a)| > \varepsilon) &\leq 2\mathbb{E}[e^{-2\varepsilon^2 R_S}] \\ &= 2(1 - \rho + \rho e^{-2\varepsilon^2})^N \\ &\leq 2e^{-N\rho(1-e^{-2\varepsilon^2})}.\end{aligned}$$

A union bound over the Q_k entries gives failure probability at most δ whenever $N\rho \geq H_k$. Substituting the two values of ρ proves the population bounds. Independence between different entries is not required.

A k-chain uses exactly $k - 1$ comparisons. Merge sort on k' alternatives uses at most $k'\lceil \log_2 k' \rceil$ comparisons, establishing the load bounds. □

C Additional Experiments

C.1 Synthetic Data

We provide an informal explanation for the empirical observation in Figure 3 that several structured preference models (single-peaked, Plackett–Luce and Euclidean) yield unimodal rank distributions for most alternatives/

A common intuition Across these models, a similar phenomenon is observed: alternatives that occupy "central" positions in the underlying structure are more likely to be ranked highly across voters, while "extreme" alternatives are less frequently top-ranked. This induces near-unimodal distributions of ranks. We will now specify how the notions of "central" and "extreme" translates under each of the structures.

Single-peaked preferences (Walsh distribution). In the single-peaked model, alternatives are ordered along an axis $c_1 < c_2 < \ldots < c_m$. Each voter has their most preferred candidate (peak) somewhere on the axis, and their preference decrease as we move away from the peak towards each of the extremes.

Given such an axes, there are two classical ways to sample rankings:

(i) **the Conitzer distribution**, which samples the peak uniformly and then completes the ranking, and

(ii) **the Walsh distribution** (used in our experiments), which samples uniformly from all rankings consistent with the axis.

Under the Walsh distribution, the number of rankings compatible with a given peak depends strongly on its position: while there is a unique preference with c_1 (resp. c_m) as peak ($c_1 \succ c_2 \succ \ldots \succ c_m$, resp. $c_m \succ c_{m-2} \succ \ldots \succ c_1$), there are exponentially many compatible rankings with the peak corresponding to middle alternative of the axis. As a result, peaks – and more generally high-ranked alternatives– are overwhelmingly concentrated around the middle of the axis. This induces unimodal rank distribution for most alternatives.

Euclidean preferences. In the k-Euclidean model, both voters and alternatives are embedded in $\mathbb{R}^k$, and each voter ranks alternatives by increasing distance from their own position. When both voters and alternatives are drawn from a roughly symmetric distribution (e.g., uniformly over a bounded region), alternatives near the geometric center of the space are, on average, closer to more voters. Consequently, central alternatives are more likely to be top-ranked, whereas peripheral alternatives tend to appear at lower positions. This again creates a concentration of probability mass in the upper ranks, leading to unimodal rank distribution.

Plackett–Luce model. In the Plackett–Luce model, each alternative a is assigned a strength parameter v_a, and rankings are generated sequentially: the top alternative is drawn with probability proportional to these strengths, then removed, and the process repeats on the remaining set. When strengths are heterogeneous (as in our experiments), a small set of high-strength alternatives dominates top positions in the ranking, while weaker alternatives are pushed toward lower ranks. This induces a strong concentration of probability mass in high ranks for a few alternatives, again yielding unimodal rank distribution.

C.2 Synthetic Data

To better illustrate these measures, we present a numerical experiment on synthetic preference profiles from Boehmer et al. (2024). In Figure 3, we consider an instance with 256 alternatives and 7 different preference profiles, namely, Mallows ($\varphi = 0.85$), Mallows mix-2 and mix-4 ($\varphi = 0.3$), Plackett-Luce with linear strengths, Walsh single-peaked, and k-Euclidean for $k \in \{2, 10\}$. Figure 3 plots $(\gamma_1^\pi(x), \gamma_2^\pi(x))$ for each alternative $x \in \mathcal{A}$ under each profile π (represented by the colored dots). The solid parabola is the boundary of the *Pearson inequality*, which states that for any $x \in \mathcal{A}$, $\gamma_2^\pi(x) \geq (\gamma_1^\pi(x))^2 - 2$. The dashed parabola marks the transition between unimodal (above) and bimodal (below) rank distributions, the latter indicating higher disagreement. The figure has been zoomed on this transition region - in particular, some dots were left out of the image. Interestingly, ranks obtained from single-peaked, Plackett-Luce, and k-Euclidean distributions lie in the unimodal region, as they induce less disagreement among voters (see Appendix C.1 for more detailed discussion). In contrasts, Mallows mixtures shift toward the bimodal region, as alternatives ranked differently in the mixtures increase disagreement among voters.

Additionally, we plot $(\gamma_1^\pi(a), \gamma_2^\pi(a))$ for a fixed alternative $a \in \mathcal{A}$ over five preference profiles $\pi_{\mathrm{IC}}, \pi_A, \pi_B, \pi_C$ and π_D. For each of them, all alternatives $b \in \mathcal{A} \setminus \{a\}$ are uniformly ranked, as well as a in π_{IC}. The distribution of r_a under the rest of the profiles is represented on the right-hand side of Figure 3. Up to degree 3, it follows that all five distributions present the same plurality matrix. However, at degree 4, the skewness is able to distinguish π_{IC}, π_C, and π_D. Similarly, at degree 5, the excess kurtosis distinguishes π_{IC}, π_A, and π_B. In particular, for a practitioner using an elicitation model considering only information of degrees 2 and 3 will never be able to distinguish the unimodal distribution π_A from the bimodal distributions π_B, π_C, π_D, from impartial culture π_{IC}.

D Inventory of measures and rules in the S-plurality hierarchy

In the main text we showed that the rank variance $\mathrm{Var}(r_a)$, the Navarrete divisiveness, and the agreement index $A(\pi)$ admit closed forms in S-plurality coordinates, and that these forms place them at levels 3, 3, and 2 of the hierarchy respectively. Beyond these three, a number of further disagreement measures, polarization indices, and aggregation rules from the social-choice literature can be analyzed in the same way: each is either expressible as a function of $\{p_S(a)\}_{S \subseteq \mathcal{A},\, a \in S}$ at some finite level ℓ, or provably outside the hierarchy. We collect the results of this analysis here.

Table 4 catalogs every measure and rule we have placed. Subsec. D.1 fixes the notation needed to read it. Subsec. D.1 gives the long-form expressions for the entries that the table abbreviates by equation number, with derivations where the closed form is new to this paper.

D.1 Notation

We work over m alternatives $\mathcal{A}$ and a profile π on $\mathcal{A}$.

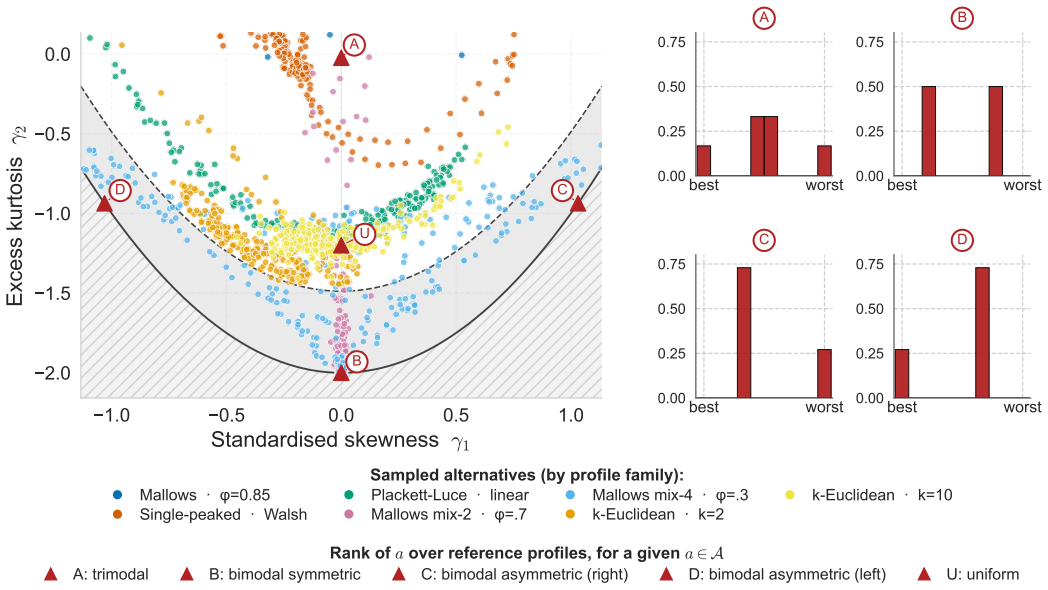


Figure 3: Skewness versus Excess kurtosis produced by seven different preferences profiles from Boehmer et al. (2024) over 256 alternatives. For each alternative $x \in \mathcal{A}$, a colored dot represents $(\gamma_1^\pi(x), \gamma_2^\pi(x))$ for the respective profile π. The solid parabola represents the boundary of the Pearson inequality $(\gamma_2^\pi(x) \geq (\gamma_1^\pi(x))^2 - 2$ for any $x \in \mathcal{A}$) while the dashed parabola marks the transition between unimodal (above) and bimodal (below) distributions. The triangles show $(\gamma_1^\pi(a), \gamma_2^\pi(a))$ for a fixed $a \in \mathcal{A}$ whose rank is either uniform (triangle U) or displayed on the right-hand side. The plurality matrix of the five distributions $\pi_{\text{IC}}, \pi_A, \pi_B, \pi_C$ and π_D are identical up to degree 3, so no measure of level 3 can detect the difference between them. An interactive version of this plot is provided as supplementary material and presented in Appendix E.

S-plurality and pairwise. The S-plurality of a on $S \ni a$ is

$$p_S(a) = \mathbb{P}_\pi\big[a \text{ ranked first in } S\big], \qquad \sum_{a \in S} p_S(a) = 1.$$

Pairwise proportions are the level-2 specialization,

$$p_{ab} = p_{\{a,b\}}(a) = \mathbb{P}_\pi[a \succ b], \qquad p_{ab} + p_{ba} = 1.$$

Borda and aggregate S-plurality.

$$\text{Bor}(a) = \sum_{b \neq a} p_{ab} = \mathbb{E}_\pi[\, m - r_a \,],$$

$$P_k(a) = \sum_{T \subseteq \mathcal{A} \setminus \{a\},\ |T| = k} p_{T \cup \{a\}}(a), \qquad P_0(a) = 1.$$

$P_k(a)$ aggregates the S-plurality of a over all S of size $k+1$ containing a. Note $P_1(a) = \text{Bor}(a)$ and $P_{m-1}(a) = p_{\mathcal{A}}(a)$.

Expected rank-distance. The expected absolute rank gap between a and b, used throughout the level-3 disagreement entries, is

$$\Delta(a,b) = \mathbb{E}_\pi\big[\, |r_a - r_b| \,\big].$$

Despite the shared symbol, $\Delta(a,b)$ is unrelated to the profile-level KT diversity $\Delta_{\text{KT}}(\pi)$ in Table 4. Its closed form in S-plurality coordinates is given as Eq. equation 10 below.

Conditional rank-distances.

$$\Delta^+(a,b) = \mathbb{E}_\pi\big[\, |r_a - r_b| \,\big|\, a \succ b \,\big], \qquad \Delta^-(a,b) = \mathbb{E}_\pi\big[\, |r_a - r_b| \,\big|\, b \succ a \,\big].$$

Closed forms are given as Eq. equation 11–equation 12.

Table 4: Aggregation rules and disagreement measures expressed in S-plurality form, grouped by purpose and sorted by informational level. Long formulas are referenced by equation number; notations are in App. D.1.

Lvl	Quantity	S-plurality form	Reference
Aggregation measures			
	Level 2: pairwise, tournament-family		
2	Borda $\mathrm{Bor}(a)$ *(score)*	$\sum_{b\neq a} p_{\{a,b\}}(a)$	(de Borda, 1781)
2	Copeland *(rule)*	$\#\{b : p_{ab} > \frac{1}{2}\} - \#\{b : p_{ab} < \frac{1}{2}\}$	(Copeland, 1951)
2	Minimax *(rule)*	$\min_{b\neq a} p_{\{a,b\}}(a)$	(Simpson, 1969; Kramer, 1977)
2	Kemeny *(rule)*	$\arg\max_{\lambda} \sum_{a\succ_\lambda b} p_{\{a,b\}}(a)$	(Kemeny, 1959)
2	Slater, Ranked Pairs, Schulze *(rule)*	computable from the tournament $\{p_{\{a,b\}}(a)\}_{a,b}$	(Slater, 1961; Tideman, 1987; Schulze, 2011)
	Level k (parametric in k)		
k	k-wise Kemeny *(rule)*	Eq. equation 4	(Gilbert et al., 2020)
	Level m: positional, runoff		
m	Plurality $\mathbb{P}[r_a = 1]$ *(rule)*	$p_{\mathcal{A}}(a)$	(de Borda, 1781)
m	Anti-plurality $1 - \mathbb{P}[r_a = m]$ *(rule)*	Eq. equation 6	(Myerson, 1993)
m	k-approval, $1 < k < m-1$ *(rule)*	$\sum_{i\leq k} \mathbb{P}[r_a = i]$, partial sum of Eq. equation 5	
m	Bucklin *(rule)*	smallest k with $\sum_{i\leq k} \mathbb{P}[r_a = i] \geq \frac{1}{2}$	(Bucklin, 1911)
m	STV / IRV *(rule)*	recursive $\arg\min_{a\in S} p_S(a)$ on shrinking S	(Tideman, 1995)
Disagreement measures			
	Level 2: pairwise		
2	Agreement $A(\pi)$, polarization $1-A(\pi)$ *(profile)*	$\frac{1}{\binom{m}{2}} \sum_{\{a,b\}} \lvert 2\, p_{\{a,b\}}(a) - 1\rvert$	(Alcalde-Unzu & Vorsatz, 2013; Can et al., 2015)
2	Generalized polarization family *(profile)*	$\sum_{\{a,b\}} f(p_{\{a,b\}}(a))$ for a class of f	(Can et al., 2017)
2	Diversity $\Delta_{\mathrm{KT}}(\pi)$ *(profile)*	$2\sum_{\{a,b\}} p_{\{a,b\}}(a)\,(1 - p_{\{a,b\}}(a))$	(Hashemi & Endriss, 2014)
2	Partitioning ratio $\alpha(a,b)$ *(pair)*	$2\min(p_{ab},\, 1 - p_{ab})$	(Delemazure et al., 2024)
	Level 3: triple, divisiveness and conflict		
3	Rank variance $\mathrm{Var}(r_a)$ *(per alt.)*	Eq. equation 7	Kendall & Smith (1939)
3	Navarrete divisiveness *(per alt.)*	Eq. equation 8	(Navarrete et al., 2024)
3	α-divisiveness, $s = \mathrm{Bor}$ *(per alt.)*	Eq. equation 9	(Colley et al., 2023a)
3	Discrepancy $\beta(a,b)$ *(pair)*	$\Delta(a,b)/(m-1)$	(Delemazure et al., 2024)
3	Group imbalance $\phi(a,b)$ *(pair)*	$\lvert \mathrm{Bor}(a) - \mathrm{Bor}(b)\rvert \,/\, \Delta(a,b)$	(Delemazure et al., 2024)
3	Discrepancy balance $\gamma(a,b)$ *(pair)*	Eq. equation 13	(Delemazure et al., 2024)
3	MaxSum (most-conflictual pair) *(rule)*	Eq. equation 14	(Delemazure et al., 2024)
3	MaxNash *(rule)*	Eq. equation 15	(Delemazure et al., 2024)
3	MaxSwap *(rule)*	Eq. equation 16	(Delemazure et al., 2024)
3	p-MaxPolar, $p > 0$ *(rule)*	Eq. equation 17	(Delemazure et al., 2024)
	Outside the hierarchy		
n/a	k-Kemeny distance κ_k ($k \geq 2$) and derived $P(E)$, $D(E)$ *(profile)*	inexpressible in $\{p_S(\cdot)\}$	(Faliszewski et al., 2023)
n/a	Spearman, Cayley distance-based diversity *(profile)*	inexpressible in $\{p_S(\cdot)\}$	(Hashemi & Endriss, 2014)
n/a	Support-based diversity (# distinct rankings) *(profile)*	inexpressible in $\{p_S(\cdot)\}$	(Hashemi & Endriss, 2014)
n/a	Karpov diversity orderings *(profile)*	inexpressible in $\{p_S(\cdot)\}$	(Karpov, 2017)

Covariance kernel.

$$K_b = \sum_{c\neq a,b} \big(p_{\{a,b,c\}}(a) - p_{ab}\, p_{ac}\big) = \sum_{c\neq a,b} \mathrm{Cov}_\pi\big(\mathbb{1}[a \succ b],\ \mathbb{1}[a \succ c]\big).$$

The expressions below correspond to the equation references in Table 4.

D.2 Aggregation rules.

With $\mathrm{top}_\lambda(S)$ denoting the alternative ranked first in S by ranking λ, the k-wise Kemeny rule of Gilbert et al. (2020) is

$$\arg\min_{\lambda} \sum_{\ell=2}^{k} \sum_{|S|=\ell} \big(1 - p_S(\mathrm{top}_\lambda(S))\big). \tag{4}$$

The level-m inversion identity expresses the rank-i frequency in terms of $\{P_k(a)\}_k$,

$$\mathbb{P}_\pi[r_a = i] = \sum_{k=m-i}^{m-1} (-1)^{k-m+i} \binom{k}{m-i} P_k(a), \tag{5}$$

from which the anti-plurality complement follows as the $i = m$ specialization,

$$1 - \mathbb{P}_\pi[r_a = m] = 1 - \sum_{k=0}^{m-1} (-1)^k P_k(a). \tag{6}$$

D.3 Rank variance.

$$\mathrm{Var}_\pi(r_a) = \mathrm{Bor}(a)\big(1 - \mathrm{Bor}(a)\big) + 2 \sum_{\{b,c\} \subseteq \mathcal{A} \setminus \{a\}} p_{\{a,b,c\}}(a). \tag{7}$$

D.4 Navarrete et al. Divisiveness.

$$\mathrm{Div}_{\mathrm{Nav}}(a) = \frac{1}{m-1} \sum_{b \neq a} \left| 1 + \frac{K_b}{p_{ab}\,(1 - p_{ab})} \right|. \tag{8}$$

D.5 α-Divisiveness.

The α-divisiveness (Colley et al., 2023a) with Borda scoring is

$$\mathrm{Div}_\alpha^{\mathrm{Bor}}(a) = \frac{1}{m-1} \sum_{b \neq a} \big(p_{ab}\,(1 - p_{ab})\big)^\alpha \, \big|\mathrm{Bor}(a \mid a \succ b) - \mathrm{Bor}(a \mid b \succ a)\big|, \tag{9}$$

where the conditional Borda scores expand as

$$\mathrm{Bor}(a \mid a \succ b) = 1 + \frac{1}{p_{ab}} \sum_{c \neq a,b} p_{\{a,b,c\}}(a),$$

$$\mathrm{Bor}(a \mid b \succ a) = \frac{1}{1 - p_{ab}} \sum_{c \neq a,b} \big(p_{ac} - p_{\{a,b,c\}}(a)\big).$$

D.6 Expected rank-distance.

$$\Delta(a,b) = 2(m-1) - \mathrm{Bor}(a) - \mathrm{Bor}(b) - 2 \sum_{c \neq a,b} p_{\{a,b,c\}}(c), \tag{10}$$

obtained from the triple-ordering identity

$$\mathbb{P}_\pi\big[c \text{ between } a \text{ and } b\big] = 2 - 2\,p_{\{a,b,c\}}(c) - p_{ac} - p_{bc},$$

summed over $c \neq a, b$, together with $\sum_{c \neq a,b} p_{ac} = \mathrm{Bor}(a) - p_{ab}$.

D.7 Conditional rank-distances.

$$\Delta^{+}(a,b) = \frac{\Delta(a,b) + \mathrm{Bor}(a) - \mathrm{Bor}(b)}{2\,p_{ab}}, \tag{11}$$

$$\Delta^{-}(a,b) = \frac{\Delta(a,b) - \mathrm{Bor}(a) + \mathrm{Bor}(b)}{2\,(1-p_{ab})}. \tag{12}$$

The first equality follows from total expectation, $\Delta(a,b) = p_{ab}\,\Delta^{+} + (1-p_{ab})\,\Delta^{-}$; the closed form combines this with the signed identity $\mathrm{Bor}(a) - \mathrm{Bor}(b) = p_{ab}\,\Delta^{+} - (1-p_{ab})\,\Delta^{-}$.

D.8 Discrepancy balance.

$$\gamma(a,b) = \min\left(\frac{\Delta^{+}(a,b)}{\Delta^{-}(a,b)},\ \frac{\Delta^{-}(a,b)}{\Delta^{+}(a,b)}\right). \tag{13}$$

D.9 Conflictual rule scores.

Substituting equation 10–equation 12 into the definitions of Delemazure et al. (2024) reduces each conflictual rule to a function of $\Delta(a,b)$, p_{ab}, $\mathrm{Bor}(a)$, and $\mathrm{Bor}(b)$. Each rule selects $\arg\max_{\{a,b\}}$ of its score:

$$\text{MaxSum}(a,b) = \Delta(a,b) + (1-2\,p_{ab})\big(\mathrm{Bor}(a) - \mathrm{Bor}(b)\big), \tag{14}$$

$$\text{MaxNash}(a,b) = \Delta(a,b)^2 - \big(\mathrm{Bor}(a) - \mathrm{Bor}(b)\big)^2, \tag{15}$$

$$\text{MaxSwap}(a,b) = \Delta(a,b) - \big|\mathrm{Bor}(a) - \mathrm{Bor}(b)\big|, \tag{16}$$

$$p\text{-MaxPolar}(a,b) = \min(p_{ab},\,1-p_{ab}) \cdot \Delta(a,b)^{p}. \tag{17}$$

All four scores depend on the level-3 data $\{p_{\{a,b,c\}}(c)\}_{c\neq a,b}$ through $\Delta(a,b)$ in equation 10, together with the level-2 data $\{p_{ab}, \mathrm{Bor}(\cdot)\}$.

E Supplementary material: Interactive moment-plane visualization

The supplementary material accompanying this paper contains a standalone HTML page, `index.html`, that reproduces every moment-plane figure of the empirical sections as an interactive visualization. The page reads its data from a single inlined JSON payload, depends only on D3 v7 loaded from a CDN, and runs in any modern browser without a local server.

The plotted families are: synthetic samples from impartial culture, Mallows and Mallows mixtures, Plackett-Luce, Walsh single-peaked, and k-Euclidean profiles at $m = 10$; the five LP-constructed witnesses (A, B, C, D, U) at $m = 256$; and three real-world PrefLib datasets (Glasgow STV, the New South Wales Legislative Assembly, and the Voter Autrement French online polling experiment), with one point per (file, candidate) pair backed by at least 50 full-rank ballots.

The page exposes per-family visibility toggles with a synthetic-density slider, free choice of axes among μ, σ^2, γ_1, γ_1^2, γ_2, γ_3, γ_4, the Pearson and Klaassen-Mokveld-van Es bound curves with the corresponding forbidden region hatched, mouse-wheel zoom and click-and-drag pan with bound curves resampled at every transform event, and a rank-distribution panel that shows the bar chart and the full set of disagreement measures for any hovered or pinned point.

The full pipeline is reproducible from source: `python run.py` samples the synthetic profiles, solves the witness linear programs, loads the real-world ballot files via `preflibtools`, computes all moments, and writes both the JSON and the rendered HTML. On a MacBook Pro with an Apple M4 Pro chip and 24 GB of unified memory, the full regeneration completes in under 5 seconds, and the rendered page is interactive at 60 frames per second.