A. Proofs of results

A.1. Proofs from Section 3.2

A.1.1. Proof of Proposition 1

Proof. Consider a distribution which assigns equal weight to the following permutations: $x < y < z, y < z < x, z < x < y$. Then, $P_{x<y} = 2/3, P_{y<z} = 2/3, P_{z<x} = 2/3$. If the ϵ-Strong Condorcet condition holds, transitivity would be violated if universality were satisfied. Since transitivity is implicit in the distributional framework, universality i.e. the condition that a unique permutation should be returned must be violated. □

A.1.2. Proof of Proposition 2

Proof.

ϵ-Strong Condorcet

$\Rightarrow$ if $P : P_{x<y} \geq \frac{1}{2} + \epsilon\,; \epsilon > 0$, then $\sigma^*_P(x) < \sigma^*_P(y)$

$\Rightarrow$ for all $Q_1, Q_2 \in \{P | P_{x<y} = \frac{1}{2} + \epsilon\}, \epsilon > 0$, then

$sign\left[\sigma^*_{Q_1}(x) - \sigma^*_{Q_1}(y)\right] = sign\left[\sigma^*_{Q_2}(x) - \sigma^*_{Q_2}(y)\right] = -1$

This last statement is ϵ-IIA. □

A.2. Proofs from Section 3.3

A.2.1. Proof of Proposition 3

Proof. Consider a social welfare procedure which decides $\tilde{\sigma}_P$ for a distribution P. Then, let $g(\sigma, P) = \delta_{\tilde{\sigma}_P}(\sigma)$ in (2), where $\delta_{\sigma'}(\sigma) = 0$ if $\sigma = \sigma'$ and ∞ otherwise. This will return $\sigma^*_P = \tilde{\sigma}_P$ and thus using g is equivalent to the social welfare procedure. □

A.2.2. Proof of Proposition 4

Proof. By definition, Pareto-efficiency requires that for all $P \in \{P | \forall\, \sigma :\ P(\sigma) > 0; \sigma(x) < \sigma(y)\}, \sigma^*_P(x) < \sigma^*_P(y)$. The statement can be rewritten as: $\forall\, P : P_{x<y} = 1$, i.e. the argument($\sigma^*_P$) minimizing the objective in Equation 2, has $\sigma^*_P(x) < \sigma^*_P(y)$, which is equivalent to having:

$$\min_{\sigma:\sigma(x)<\sigma(y)} g_P(\sigma) < \min_{\sigma:\sigma(x)>\sigma(y)} g_P(\sigma)$$

i.e.. the minimizer $\sigma^*_P \in \{\sigma : \sigma(x) < \sigma(y)\}$. Also, this should hold for all distributions $P : P_{x<y} = 1$, which is true iff $U_1(g; x, y) < 0$. □

A.2.3. Proof of Proposition 5

Proof. By definition, IIA requires that for all distributions $P, Q \in \{P | P_{x<y} = \gamma\}$, $sign\left[\sigma^*_P(x) - \sigma^*_P(y)\right] = sign\left[\sigma^*_Q(x) - \sigma^*_Q(y)\right], \forall x, y \in X$. The statement can be rewritten as: for all $P \in \{P | P_{x<y} = \gamma\}$, $sign[\sigma^*_P(x) - \sigma^*_P(y)]$ should remain constant. i.e. for a fixed γ, for all $P : P_{x<y} = \gamma$, the argument(σ^*_P) minimizing the objective in Equation 2, should always belong to either $\sigma : \sigma(x) > \sigma(y)$ or $\sigma : \sigma(x) < \sigma(y)$. Also, we know that:

$$\sigma^*_P(x) < \sigma^*_P(y) \text{ iff } \min_{\sigma:\sigma(x)<\sigma(y)} g_P(\sigma) < \min_{\sigma:\sigma(x)>\sigma(y)} g_P(\sigma)$$

$$\sigma^*_P(x) > \sigma^*_P(y) \text{ iff } \min_{\sigma:\sigma(x)<\sigma(y)} g_P(\sigma) > \min_{\sigma:\sigma(x)>\sigma(y)} g_P(\sigma)$$

which is equivalent to,

$$sign[\sigma^*_P(x) - \sigma^*_P(y)] = sign\left[\min_{\sigma:\sigma(x)<\sigma(y)} g_P(\sigma) - \min_{\sigma:\sigma(x)>\sigma(y)} g_P(\sigma)\right]$$

This means that for g to satisfy IIA, $sign\left[\min_{\sigma:\sigma(x)<\sigma(y)} g_P(\sigma) - \min_{\sigma:\sigma(x)>\sigma(y)} g_P(\sigma)\right]$ is constant for all $P \in \{P | P_{x<y} = \gamma\}$. Now, any function has a constant sign, iff it's minimum and maximum value have the same sign.
Thus, g satisfies IIA iff $L_\gamma(g; x, y), U_\gamma(g; x, y) > 0$, $\forall \gamma :\ 0 \leq \gamma \leq 1$ and $\forall\, x, y \in \mathcal{X}$. □

A.2.4. Proof of Proposition 6

Proof. By definition, Condorcet Criteria requires that for all P such that if $P_{x<y} > \frac{1}{2}\ \forall\, y \in \mathcal{X} \backslash \{x\}$ then $\sigma^*_P(x) = 1$. This requirement can be rewritten as: Let $P \in \{P | P_{x<y} = \gamma_y > \frac{1}{2}\ \forall\, y \in \mathcal{X} \backslash \{x\}\}$, then $\sigma^*_P(x) < \sigma^*_P(y)$ for all $y \in \mathcal{X} \backslash \{x\}$. We know that:

$$\sigma^*_P(x) < \sigma^*_P(y) \text{ iff } \min_{\sigma:\sigma(x)<\sigma(y)} g_P(\sigma) > \min_{\sigma:\sigma(x)>\sigma(y)} g_P(\sigma)$$

Following the argument in proof of Proposition 4, we recover the required statement. □

A.2.5. Proof of Proposition 7

Proof. By definition, majority rule is satisfied iff for all $P \in \{P | P^1(x) > \frac{1}{2}\}$, then $\sigma^*_P(x) = 1$. The statement can be rewritten as: $\forall\, P : P^1(x) = 1$, i.e. the argument(σ^*_P) minimizing the objective in Equation 2, has $\sigma^*_P(x) = 1$, which is equivalent to having:

$$\min_{\sigma:\sigma(x)=1} g(\sigma, P) < \min_{\sigma:\sigma(x)\neq 1} g(\sigma, P)$$

i.e.. the minimizer $\sigma^*_P \in \{\sigma : \sigma(x) = 1$. Also, this should hold for all distributions $P : P^1(x) > \frac{1}{2}$, which is true iff $U^1_\gamma(g; y) < 0$ for all $\gamma > \frac{1}{2}$. □

A.3. Proofs from Section 3.4

A.3.1. Proof of Theorem 2

Proof. P is a distribution over S_n which has finitely many elements and hence can be represented as a finite real valued vector, thus $g(\sigma_0, P)$ for a fixed σ_0 is a multivariate function of this vector and the continuity of $g(\cdot, P)$ is a well-defined notion.

Now, pick any pair of alternatives x and y. Consider the quantity

$$h_{xy}(P) = \min_{\sigma:\sigma(x)<\sigma(y)} g(\sigma, P) - \min_{\sigma:\sigma(x)>\sigma(y)} g(\sigma, P) \quad (5)$$

Then, $h_{xy}(P) < 0 \iff \sigma^*_P(x) < \sigma^*_P(y)$, as the minimum σ^*_P will be chosen from $\{\sigma : \sigma(x) < \sigma(y)\}$. Now, pick two permutations σ_1 and σ_2 such that $\sigma_1(x) < \sigma_1(y)$ and $\sigma_2(x) > \sigma_2(y)$.

Consider a family of distributions indexed by $0 \leq \lambda \leq 1$. Let $P_\lambda = \lambda\delta_{\sigma_1} + (1-\lambda)\delta_{\sigma_2}$, where δ is such that $\delta_{\tilde{\sigma}}(\sigma) = 1$ if $\tilde{\sigma} = \sigma$ and 0 otherwise. Note the following:

- **Pareto** implies, since $P_0 \equiv \delta_{\sigma_2}$, $h_{xy}(P_0) > 0$ and similarly, $P_1 \equiv \delta_{\sigma_1}$ and $h_{xy}(P_1) < 0$
- $g(\sigma, P)$ is continuous in P for each fixed σ. Now, each of the two terms on the RHS of (5) are minimizations over a finite set of σ. The minimization of a finite number of continuous functions in P yields a continuous function and so does the subtraction of these two terms. Thus, h_{xy} is continuous in P and hence in λ.
- The above two statements together imply that for some $0 < \lambda^* < 1$, $h_{xy}(\lambda^*) = 0$. However, this means that for some pair σ^*_1 and σ^*_2, such that $\sigma^*_1(x) < \sigma^*_1(y)$ and $\sigma^*_2(x) > \sigma^*_1(y)$, $g(\sigma^*_1, P^*_\lambda) = g(\sigma^*_2, P^*_\lambda)$ and this is also the minimum value of g, so that $\sigma^*_1, \sigma^*_2 \in \operatorname{argmin}_\sigma g(\sigma, P)$. Since they differ atleast in the relative position of x and y in the ranking, the ranking procedure does not return a unique minimizer and thus does not admit **Universality**.

□

A.4. Proofs from Section 4

A.4.1. Proof of Proposition 8

Proof. The result is a direct consequence of using the rearrangement inequality on Equation (4). □

A.5. Proofs from Section 4.1

A.5.1. Proof of Proposition 9

Proof. Consider the quantity,

$$f_P(x) - f_P(y) = \sum_{\sigma:\sigma(x)<\sigma(y)} P(\sigma)(h(\sigma(x)) - h(\sigma(y))) - \sum_{\sigma:\sigma(x)>\sigma(y)} P(\sigma)(h(\sigma(y)) - h(\sigma(y)))$$

IIA requires that, for all distributions $P \in M_\gamma(x, y) = \{P|P_{x<y} = \gamma\}$, $sign(\sigma^*_{P,\ell}(x) - \sigma^*_{P,\ell}(y))$ should remain constant.

Then, any $P \in M_\gamma(x, y)$, $f_P(x) - f_P(y)$ can be lower and upper bounded by the following, using the definition of Ω_h and ω_h:

$$\gamma\omega_h - (1-\gamma)\Omega_h \leq f_P(x) - f_P(y) \leq \gamma\Omega_h - (1-\gamma)\omega_h \quad (6)$$

Both inequalities are tight i.e. There exist distributions Q_1 and Q_2 such that, $Q_{1_{x<y}} = Q_{2_{x<y}} = \gamma$, and $f_{Q_1}(x) - f_{Q_1}(y) = \gamma\Omega_h - (1-\gamma)\omega_h$, $f_{Q_2}(x) - f_{Q_2}(y) = \gamma\omega_h - (1-\gamma)\Omega_h$.

Now, $sign(f(x) - f(y))$ is constant

$$\iff (\gamma\omega_h - (1-\gamma)\Omega_h) \times (\gamma\Omega_h - (1-\gamma)\omega_h) > 0$$
$$\iff S(\gamma) = \gamma^2(\Omega_h + \omega_h)^2 - \gamma(\Omega_h + \omega_h)^2 + \Omega_h\omega_h > 0$$

This must hold $\forall\ \gamma,\ 0 \leq \gamma \leq 1$. Also observe that $S(0) = S(1) = \Omega_h\omega_h$.

$S(\gamma)$ is convex (in particular, quadratic) in γ, with its minimum occurring at:

$$\gamma^* = \operatorname*{argmin}_{\gamma\in[0,1]} S(\gamma) = \frac{1}{2} \quad (7)$$

and

$$S(\gamma^*) = -\frac{1}{4}(\Omega_h - \omega_h)^2 \leq 0 \quad (8)$$

Thus, $S(\gamma^*) \leq 0$, and hence no ℓ_h can satisfy IIA exactly. □

A.5.2. Proof of Proposition 10

Proof. Continuing from the proof of Proposition 9, Since $S(\gamma)$ is convex, quadratic, we need that $S(\frac{1}{2} + \epsilon) = S(\frac{1}{2} - \epsilon) > 0$. Plugging $\gamma = \frac{1}{2} + \epsilon$ in defintion of $S(\gamma)$ leads to a recovery of the result. □

A.5.3. Proof of Proposition 11

Proof. For, Pareto-efficiency, for all $\{P|P_{x<y} = 1\}$, we need that $\sigma^*_{P,\ell_h}(x) < \sigma^*_{P,\ell_h}(y)$, which is equivalent to having $f(x) - f(y) > 0$. Plugging $\gamma = 1$ in Equation (6), we get that $f(x) - f(y) > 0$ iff $\omega_h > 0$, which is equivalent to $h : [n] \mapsto \mathbb{R}$ being strictly monotonically decreasing. □

A.5.4. Proof of Proposition 12

Proof. The score for alternative x is given by $f_P(x) = \sum_{\sigma' \in S_n} P(\sigma')h(\sigma'(x))$. Suppose the position of x goes up such that probability mass $\delta > 0$ shifts from σ_1 to σ_2 from P_1 to P_2, where σ_1 and σ_2 differ only in that x is higher up in σ_2 than in σ_1, i.e. $\sigma_2(x) < \sigma_1(x)$ and $\sigma_2(y) \geq \sigma_1(y) \forall y \in \mathcal{X}\backslash\{x\}$,

$$f_{P_2}(x) - f_{P_1}(x) = \delta(h(\sigma_2(x)) - h(\sigma_1(x)))$$

Then, $f_P(x)$ will decrease or remain the same iff h is non-increasing. □

A.5.5. Proof of Proposition 13

Proof. Given distributions P and Q such that $\sigma^*_{P,\ell_h}(x) = 1$ and $\sigma^*_{Q,\ell_h}(x) = 1$, which implies that $f_P(x) = \mathbb{E}_{\sigma' \sim P}[h(\sigma'(x))] > f_P(y),\ \forall y \in \mathcal{X}\backslash\{x\}$. Similarly, $f_Q(x) = \mathbb{E}_{\sigma' \sim Q}[h(\sigma'(x))] > f_Q(y),\ \forall y \in \mathcal{X}\backslash\{x\}$. $f_{\frac{P+Q}{2}}(x) = \mathbb{E}_{\sigma' \sim \frac{P+Q}{2}}[h(\sigma'(x))] = (f_P(x) + f_Q(x))/2 > f_{\frac{P+Q}{2}}(y),\ \forall y \in \mathcal{X}\backslash\{x\}$. □

A.5.6. Proof of Proposition 14

Proof. Let $\omega^1_h = \min_{j \in [n] | j > 1} h(1) - h(j)$ and let $P^1(x) = \gamma = \frac{1}{2} + \epsilon; \epsilon > 0$ Then for all $y \in \mathcal{X}\backslash\{x\}$:

$$
\begin{aligned}
f(x) - f(y) &= \sum_{\sigma:\sigma(x)=1} P(\sigma)\left(h(1) - h(y)\right) \\
&\quad - \sum_{\sigma:\sigma(x)\neq 1} P(\sigma)\left(h(y) - h(x)\right) \\
&\geq \gamma(\omega^1_h) - (1-\gamma)\max(\Omega_h, -\omega_h) \\
&\equiv \epsilon\left(\omega^1_h + \max(\Omega_h, -\omega_h)\right) - \tfrac{1}{2}\left(\max(\Omega_h, -\omega_h) - \omega^1_h\right)
\end{aligned}
$$

For majority rule, we need that, $f(x) - f(y) > 0$, which is equivalent to

$$
\begin{aligned}
&\epsilon(\omega^1_h + \max(\Omega_h, -\omega_h)) - \tfrac{1}{2}(\max(\Omega_h, -\omega_h) - \omega^1_h) > 0 \\
&\frac{\max(\Omega_h, -\omega_h) + \omega^1_h}{\max(\Omega_h, -\omega_h) - \omega^1_h} > \frac{1}{2\epsilon}
\end{aligned}
$$

So, for any positional loss function ℓ_h, we can find an ϵ(and hence a distribution P), for which majority criteria is not met. □

A.5.7. Proof of Proposition 15

Proof. The argument follows directly from the proof of proposition 14. □

A.5.8. Proof of Proposition 16

Proof. For Condorcet analysis, if $P_{x<y} = \gamma = \frac{1}{2} + \epsilon,\ \epsilon > 0$, then $f(x) - f(y) > 0$ iff $\epsilon(\Omega_h + \omega_h) > \frac{1}{2}(\Omega_h - \omega_h)$. Taking $\lim_{\epsilon \to 0^+}$, shows that positional scoring rules don't satisfy Condorcet criteria in general i.e. For any positional loss ℓ_h, we can find an ϵ(and hence a distribution P), for which Condorcet criteria is not met. □

A.5.9. Proof of Proposition 17

Proof. The proof follows directly from the proof of proposition 16. □

A.6. Proofs from Section 4.2

A.6.1. Proof of Theorem 3

Proof. We show this constructively. Table 1 shows the ϵs for relaxed axioms for various ranking procedures. In particular, note that Borda Count satisfies Pareto and Monotonicity exactly and satisfies the relaxed Strong Condorcet, IIA and Majority Rule axioms with some $0 < \epsilon, \epsilon' < \frac{1}{2}$, as required. □

A.6.2. Proof of Lemma 1

Proof. Since ℓ_h satisfies Pareto, h must be strictly decreasing. Then,

$$
\boldsymbol{\omega_h} = \min_{\{i,j \in [n] | i<j\}} h(i) - h(j) = \min_{i \in [n]} \left(h(i) - h(i+1)\right) \tag{9}
$$

and

$$
\begin{aligned}
\boldsymbol{\Omega_h} &= \max_{\{i,j \in [n] | i<j\}} h(i) - h(j) = h(1) - h(n) && (10) \\
&= (h(1) - h(2)) + \cdots + (h(n-1) - h(n)) && (11) \\
&\geq (n-1)\boldsymbol{\omega_h} && (12)
\end{aligned}
$$

where the last line follows because each term in (11) if larger than their minimum given by (9), as required. □

A.6.3. Proof of Theorem 4

Proof. Let $t = \frac{\boldsymbol{\Omega_h}}{\boldsymbol{\omega_h}} \geq 1$. Then, from Proposition 10,

$$
\epsilon^* = \frac{1}{2}\left(\frac{\Omega_h - \omega_h}{\Omega_h + \omega_h}\right) = \frac{1}{2}\left(\frac{t-1}{t+1}\right) \tag{13}
$$

ϵ^* is then monotonically increasing in t for $t \geq 1$ and is thus minimized when t is minimum. From Lemma 1, this is minimized when $t = n-1$, which is achieved by Borda (see Table 1). □

B. Experimental details

In this section, we discuss the motivation and key-takeaways from our Experiments in great detail. In Section 4, we saw that for positional scoring losses, all axiomatic properties(both exact and approximate) depend on three quantities Ω_h,ω_h and ω^1_h. Recall from previous definitions that $\Omega_h = \max_{\{i,j \in [n] | i<j\}} h(i) - h(j)$; $\omega_h = \min_{\{i,j \in [n] | i<j\}} h(i) - h(j)$ and ω^1_h= $\min_{j \in [n] | j>1} h(1) - h(j)$. Moroever, all axioms are functions of ratios of these three quantities. So, any positive linear scaling or translation of h will not change the behavior of ℓ_h towards any axioms.i.e. Axioms only depend on the behavior of the discrete-derivative of the function h. We state some key properties:

- For any strictly decreasing convex function, it's discrete-derivative is decreasing in nature i.e. The sharpest change in function value is observed at the top of the list, while the slowest change is observed at the bottom. To quantify this, for any strictly decreasing convex $h, \omega_h = h(n) - h(n-1)$
- On the contrary, for any strictly decreasing concave function, it's discrete-derivative is increasing in nature. i.e. The sharpest change in function value is observed at the bottom of the list, while the slowest change is observed at the top. To quantify this, for any strictly decreasing concave $h, \omega_h = h(2) - h(1)$.
- For Borda Count, which is a linear function, it's discrete derivatives are constant.
- In our analysis of IIA, we observed that for any two alternatives x and y, for a distribution $P$with the marginal probability $P_{x<y} = \gamma$, the difference of their scores $f_P(x) - f_P(y)$ are upper and lower bounded by $\gamma\omega_h - (1-\gamma)\Omega_h \leq f(x) - f(y) \leq \gamma\Omega - (1-\gamma)\omega$.
  The distribution $\hat{P}$ over which the lower bound is attained is the one, which places x over y at the indices corresponding to $\operatorname{argmin}_{\{i,j\in[n] | i<j\}} h(i) - h(j)$ and places y over x at the indices corresponding to $\operatorname{argmax}_{\{i,j\in[n] | i<j\}} h(i) - h(j)$.
- For convex losses, $\hat{P}$ places x and y at the bottom position, whereas for concave losses $\hat{P}$ places them at the top.
- Motivated by this intuition, we define two experiments where we choose the centers for the mixture of Mallows so that we can simulate the $\hat{P}$ for convex and concave independently. Both the experiments have Z_2 to be the same, which places B at the top and A at the bottom.(corresponding to Ω).
- In Experiment 1, Z_1 puts $\{A, B\}$ at the bottom, and as we increase the weight on Z_1, we essentially simulate $\hat{P}$ for a convex function, which is why we see that the convex losses perform poorly.
- Similarly Experiment 2, Z_1 puts $\{A, B\}$ at the top, and as we increase the weight on Z_1, we essentially simulate $\hat{P}$ for a concave function, which is why we see that the concave losses perform poorly.
- Since for Borda, the discrete-derivative is constant, it is not prone to such distributions and hence never performs the worst. Hence, even empirically we can see that in a distribution agnostic setting, Borda Count would be optimal.

C. Directions for Future Work

The study of social choice theory was revolutionized by Arrow (1951) with the axiomatic approach to social choice and welfare. However, the interpretation of these axioms has been a common complaint among practitioners, in part due to the characterization of axioms using words or logical statements instead of a quantitative description. We hope that our framework has shed light on the characteristics of as well as similarities between the various axioms.

Within the utility maximization framework, we have analyzed Positional Scoring rules in detail but distance measures also fall in this class and can be analyzed with the same machinery. Their performance with regard to the relaxed axioms is an important direction for future work. In the future, we hope to tackle the more interesting problem of obtaining a representation theory for all losses which satisfy a certain axiom.

We have exhaustively characterized the set of positional scoring rules that satisfy each axiom (including our relaxed variants) and shown that Borda Count is optimal in a sense. It remains to obtain such results for broader classes of loss functionals (for instance, distance measures).

A common use of rank aggregation techniques in machine learning is to obtain rankings from a number of algorithms and combine them to produce a higher score on some other metric, such as the NDCG in learning to rank. Investigating such metrics w.r.t. the axioms and designing procedures which are guaranteed to improve performance on these metrics w.r.t. the base rankings is an important direction for future work. In many such settings, it suffices to obtain only the top k elements in the aggregation in contrast to an entire ranking over the set of all alternatives: extending our analysis to such settings is another key direction for future work.

Another important area of study in social welfare theory are game theoretic considerations such as the susceptibility of a voting procedure to collusion among the voters. A celebrated result in this vein is the Gibbard-Satterthwaite theorem regarding the strategyproofness of theorems. Analyzing distributional rank aggregation procedures under this lens is an open problem.