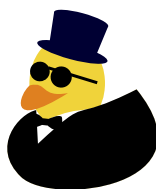


On the Structural Incompatibility of Fairness Axioms and the Quantification of Social Choice Optimality via Metric Distortion

Aarush Aggarwal

July 2025



Abstract

This paper explores the mathematical limitations of fair voting through Arrow's Impossibility Theorem and extends the analysis to approximate social optimality using metric distortion. By bridging combinatorial axioms with geometric models, we show how deterministic mechanisms fail exact fairness but approach optimality under distance-based frameworks. We synthesize structural impossibility with quantitative efficiency, offering a unified perspective on the bounds of collective rationality. *This work assumes familiarity with basic concepts in graph theory, combinatorics, and introductory game theory.*

1 Introduction

“No voting rule can transform individual judgments into a consistent—and fair—collective decision when choices exceed two.” This stark insight, first articulated by Arrow (1951), shows the profound tension at the heart of democratic systems: the aspiration for collective rationality clashes with the variety of individual preferences as choice sets grow richer.

Social choice theory formalizes this tension. The goal is to design an aggregation mechanism which satisfies normative criteria like Pareto efficiency, neutrality, non-dictatorship, and independence of irrelevant alternatives. Arrow’s theorem proves these cannot all hold together for three or more alternatives with 2 or more voters: enforcing some inevitably requires relinquishing others. The structure of preferences, and rules makes the occurrence of paradoxes mathematically inherent, rendering complete democratic coherence mathematically impossible in generality.

However, impossibility does not preclude approximate optimality. A second movement in social choice quantifies how “close” outcomes can come to utilitarian ideals when richer information is revealed. Under the *metric distortion* framework, both voters and alternatives are embedded in a metric space (X, d) , and each voter’s cost is modeled by the distance $d(x_i, a)$. While aggregation mechanisms observe only the induced rankings, their performance is evaluated in terms of worst-case *distortion*. Astonishingly, imposing only metric structure transforms an impossibility into an approximate possibility teaching us how close simple deterministic rules can get to optimal.

By combining structural impossibility results with quantitative approximation performance, we offer a unified framework: we explain not only *why* democracy cannot perfectly reconcile fairness and rationality in general, but also *how close* it can come under realistic assumptions.

2 Preliminaries and Formal Framework

We begin by establishing the mathematical foundations of social choice theory required for analyzing axiomatic impossibility theorems and metric distortion.

2.1 Preference Profiles and Aggregation Rules

Definition 2.1 (Alternatives and Voters). Let $A = \{a_1, a_2, \dots, a_m\}$ be a finite set of $m \geq 3$ alternatives, and let $N = \{1, 2, \dots, n\}$ denote the set of voters.

Definition 2.2 (Individual Preferences). Each voter $i \in N$ holds a strict linear order \succ_i over A , i.e., a binary relation on A satisfying:

- **Completeness:** For all $a, b \in A$, either $a \succ_i b$ or $b \succ_i a$.
- **Transitivity:** For all $a, b, c \in A$, if $a \succ_i b$ and $b \succ_i c$, then $a \succ_i c$.
- **Antisymmetry:** If $a \succ_i b$, then not $b \succ_i a$.

We denote by $\mathcal{L}(A)$ the set of all strict linear orders over A . Each $\succ_i \in \mathcal{L}(A)$. $\|\mathcal{L}(A)\| = m!$

Definition 2.3 (Preference Profile). A *preference profile* is a tuple $\vec{\succ} = (\succ_1, \succ_2, \dots, \succ_n) \in \mathcal{L}(A)^n$ that specifies the preferences of all n voters.

Definition 2.4 (Social Choice Function (SCF)). A social choice function is a function $f : \mathcal{L}(A)^n \rightarrow A$ that selects a single winning alternative given a preference profile.

Definition 2.5 (Social Welfare Function (SWF)). A social welfare function is a function $F : \mathcal{L}(A)^n \rightarrow \mathcal{L}(A)$ that returns a complete ranking of alternatives based on the profile.

Let \mathcal{F}_n denote the set of all social choice functions for n voters. Let $\mathcal{F}_{\text{onto}}^n \subseteq \mathcal{F}_n$ denote the set of SCFs where every alternative $a \in A$ can be elected under some profile.

2.2 Axioms of Fairness and Rationality

We now define the classical axioms that SCFs or SWFs may satisfy:

Definition 2.6 (Unanimity (Pareto Efficiency)). A rule satisfies *unanimity* if whenever all voters prefer a to b , then a is socially preferred. That is, if $a \succ_i b$ for all $i \in N$, then $f(\vec{\succ}) = a$ or $a \succ_F b$.

Definition 2.7 (Unrestricted Domain (U)). The function F must be defined for every $\vec{\succ} \in \mathcal{L}(A)^n$.

Definition 2.8 (Independence of Irrelevant Alternatives (IIA)). A rule satisfies IIA if the social preference between any two alternatives a and b depends only on the individual preferences between a and b . Changes in rankings involving other alternatives should not affect the outcome.

Definition 2.9 (Non-Dictatorship). A rule is *non-dictatorial* if there is no voter $i \in N$ such that for every profile $\vec{\succ}$, the social outcome always agrees with \succ_i .

Definition 2.10 (Strategy-Proofness). A social choice function f is *strategy-proof* if no voter can gain by misrepresenting their preferences. That is, for all $i \in N$, all $\vec{\succ}_{-i} \in \mathcal{L}(A)^{n-1}$, and all alternative preferences $\succ'_i \in \mathcal{L}(A)$:

$$f(\succ_i, \vec{\succ}_{-i}) \succeq_i f(\succ'_i, \vec{\succ}_{-i}).$$

Thus truthfulness is the dominant strategy.

2.3 Metric Preference Domains

We introduce structure by embedding voters and alternatives in a metric space:

Definition 2.11 (Metric Preference Model). Let (X, d) be a metric space. Each voter i is located at $x_i \in X$, and each alternative $a \in A$ is embedded in X . Voter i prefers alternatives that are closer:

$$a \succ_i b \iff d(x_i, a) < d(x_i, b).$$

3 The Notion of Unfairness or Inconsistency in Voting Systems

In social choice theory, a democratic voting system is formally conceived as a mechanism that aggregates individual preference rankings, where each is represented by a strict linear order $\succ_i \in \mathcal{L}(A)$, into a single coherent social ordering $\succ_F = F(\succ_1, \dots, \succ_n)$. This aggregation serves the democratic ideals of converting diverse opinions into the collective social choice of a leader, where the collective decision is both internally consistent and responsive to majority sentiment.

While multiple approaches exist, from simple plurality and runoff mechanisms to more intricate scoring and Condorcet methods, none guarantees perfect coherence. In fact, even intuitively reasonable systems do not respect unanimous preferences or produce transitive outcomes. For example, plurality voting can disregard the intensity of preference.

Consider three alternatives A, B, C and a population divided as follows:

Voters	30%	30%	40%
Ranking	$A \succ B \succ C$	$C \succ A \succ B$	$B \succ A \succ C$

Assume each voter also has the following cardinal utilitarian values for the alternatives attached to the population divided into p_1, p_2, p_3 for each distinct group preference:

Group 1 (30%): $u(A) = 3, u(B) = 2, u(C) = 1,$
Group 2 (30%): $u(C) = 3, u(A) = 2, u(B) = 1,$
Group 3 (40%): $u(B) = 3, u(A) = 2, u(C) = 1.$

Plurality outcome: Each group votes their top choice:

$$A : 30\%, \quad C : 30\%, \quad B : 40\% \quad \implies \quad \text{Winner: } B.$$

Utilitarian welfare: The sum of utilities for each alternative is:

$$U(x) = \sum_{i=1}^n p_g \cdot u_i(x).$$

If voters can be grouped into G types, with each group g having proportion p_g and common utility function $u_g(x)$ (where $\sum_{g=1}^G p_g = 1$)

In this case these Utilitarian Welfare Values are:

$$\begin{aligned} U(A) &= 0.3 \cdot 3 + 0.3 \cdot 2 + 0.4 \cdot 2 = 0.9 + 0.6 + 0.8 = 2.3, \\ U(B) &= 0.3 \cdot 2 + 0.3 \cdot 1 + 0.4 \cdot 3 = 0.6 + 0.3 + 1.2 = 2.1, \\ U(C) &= 0.3 \cdot 1 + 0.3 \cdot 3 + 0.4 \cdot 1 = 0.3 + 0.9 + 0.4 = 1.6. \end{aligned}$$

Hence A maximizes total utility (2.3), but plurality improperly chooses B . This highlights how plurality ignores preference intensity, leading to sub-optimal social welfare.

A more formally compelling example of this phenomenon is the **Condorcet paradox**, which we introduce next.

3.1 Condorcet's Paradox

Let $A = \{x_1, x_2, \dots, x_m\}$ with $m \geq 3$, and n voters each having strict rankings \succ_i . We define:

Definition 3.1 (Condorcet Winner). An alternative $x \in A$ is a *Condorcet winner* if for every $y \neq x$,

$$|\{i : x \succ_i y\}| > |\{i : y \succ_i x\}|.$$

Definition 3.2 (Condorcet Cycle). A *Condorcet cycle* (or paradox) occurs when there exists a sequence x_1, x_2, \dots, x_k in A ($k \geq 3$) such that:

$$x_1 \succ x_2 \succ \dots \succ x_k \succ x_1$$

holds in the social preference derived by majority comparisons.

The classical 3-alternative example is:


Rank	\succ_1	\succ_2	\succ_3
1	A	B	C
2	B	C	A
3	C	A	B

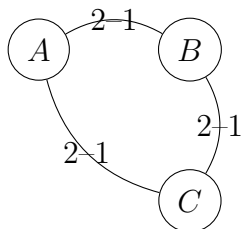
Then majority pairwise comparisons yield:

$$A \succ B \quad (2 \text{ votes}), \quad B \succ C \quad (2 \text{ votes}), \quad C \succ A \quad (2 \text{ votes}),$$

producing a cycle $A \succ B \succ C \succ A$

Theorem 3.1. *A Condorcet cycle exists \iff there is no Condorcet winner.*

Proof. If x is a Condorcet winner, it must beat every other alternative \Rightarrow no cycle can include x . Conversely, if no cycle exists, then the majority tournament is a transitive tournament, implying the existence of a maximal element x that wins against all others—i.e., a Condorcet winner 



The existence of a Condorcet cycle shows that majority rule can fail transitivity, even when all individual preferences are rational. Similarly, a larger inconsistency occurs in more general social welfare functions when we attempt to satisfy a set of seemingly reasonable fairness conditions.

4 Arrow's Impossibility Theorem

Arrow's Impossibility Theorem formalizes this contradiction: it asserts that no aggregation rule can convert individual rational preferences into a collective rational ordering while simultaneously satisfying Pareto efficiency, independence of irrelevant alternatives (IIA), and non-dictatorship. This theorem reveals a deep structural limitation in collective decision-making, showing that any such rule must sacrifice at least one of these desirable axioms when the number of alternatives is three or more.

Constructive Proof of Arrow's Impossibility Theorem

We now adopt a structural approach based on *decisive coalitions*—subsets of voters whose unanimous preferences determine social outcomes between pairs of alternatives. Under the axioms of Pareto efficiency and IIA, we establish two key results: the *Field Expansion Lemma*, which extends decisiveness across alternatives, and the *Contraction Lemma*, which reduces decisive coalitions while preserving their influence. These lemmas together imply the existence of a single globally decisive voter, violating non-dictatorship.

4.0.1 Decisive Coalitions

Definition 4.1 (Decisive Coalition). A coalition $S \subseteq N$ is *decisive for pair* (x, y) if whenever $\forall i \in S : x \succ_i y$, then $x \succ_F y$ regardless of preferences of voters in $N \setminus S$. If S is decisive for every (x, y) , we call S *globally decisive*.

By Pareto efficiency, the grand coalition N is globally decisive.

4.0.2 Field Expansion Lemma

Lemma 4.1 (Field Expansion). *If a coalition $S \subseteq N$ is decisive for a pair (x, y) , then for any third alternative z , S is also decisive for (x, z) and (z, y) .*

Proof. Assume S is decisive for (x, y) ; that is, whenever all members of S rank $x \succ_i y$, the social ordering F satisfies $x \succ_F y$ regardless of others' preferences.

We show S remains decisive for (x, z) . By Unrestricted Domain, consider the profile where:


- Every $i \in S$ ranks: $x \succ_i z \succ_i y$.
- Every $j \notin S$ ranks: $z \succ_j y \succ_j x$.

Here, S unanimously prefers x to y , so by decisiveness $x \succ_F y$. By Pareto, since all $j \notin S$ prefer $z \succ_j y$, we also have $z \succ_F y$. Then by transitivity, $x \succ_F z$. Applying IIA to ignore y , we conclude $x \succ_F z$ must follow whenever all in S prefer x to z ; hence S is decisive for (x, z) .

A symmetric argument shows S is decisive for (z, y) : construct a profile with

$$\begin{aligned} i \in S : & \quad z \succ_i x \succ_i y, \\ j \notin S : & \quad x \succ_j y \succ_j z, \end{aligned}$$

use decisiveness for (x, y) and IIA/transitivity to deduce $z \succ_F y$, and apply IIA again to drop x .

Thus S is decisive for both (x, z) and (z, y) , completing the proof. 

4.0.3 Contraction Lemma

Lemma 4.2 (Contraction). *If $S \subseteq N$ is globally decisive and $|S| \geq 2$, then there exists a proper subset $T \subset S$ that is also globally decisive.*

Proof. Let S be globally decisive with $|S| \geq 2$. Partition S arbitrarily into two non-empty, disjoint coalitions S_1, S_2 such that $S_1 \cup S_2 = S$.

Select three distinct alternatives $x, y, z \in A$. By global decisiveness of S , S decides each pair.

Construct the following profile P , specifying only the preferences of voters in S , and letting others vote arbitrarily (so as not to affect comparisons involving S):


	Voters in S_1	Voters in S_2	Others
1 st	x	y	\cdot
2 nd	y	z	\cdot
3 rd	z	x	\cdot

Here S_1 ranks $x \succ y$, and S_2 ranks $y \succ z$, so by transitivity S enforces both $x \succ_F y$ and $y \succ_F z$. Since S decides each pair, we also know $z \succ_F x$ in profile P .

There are two cases:

Case 1: $x \succ_F z$. Then S_1 alone caused $x \succ_F z$, as voters in S_2 preferred $z \succ x$. Thus by IIA and the profile's structure, S_1 is decisive at least on (x, z) . By the *Field Expansion Lemma*, S_1 is globally* decisive. Set $T = S_1 \subset S$.

Case 2: $z \succ_F x$. Then S_2 must have enforced $z \succ_F x$ (because S_1 ranked $x \succ z$). By a symmetric argument to Case 1, S_2 is globally decisive. Set $T = S_2 \subset S$.

In either case, we have found a proper, globally decisive subset T . This completes the contraction. 

Proving Arrow's Theorem By Existence of Dictator

Theorem 4.3 (Arrow, 1951). *No social welfare function $F : \mathcal{L}(A)^n \rightarrow \mathcal{L}(A)$ satisfies Unrestricted Domain, Pareto Efficiency, Independence of Irrelevant Alternatives (IIA), and Non-Dictatorship if $|A| \geq 3$.*

Proof. Assume, for the sake of contradiction, that there exists a social welfare function F satisfying all four axioms: Unrestricted Domain (U), Pareto Efficiency (P), Independence of Irrelevant Alternatives (IIA), and Non-Dictatorship (ND), with $|A| \geq 3$.

Step 1: The Grand Coalition is Globally Decisive. The Pareto condition implies that if all voters in the electorate N strictly prefer an alternative x over another alternative y , then the social welfare function must rank x above y . In other words,

$$\forall x, y \in A, \quad \text{if } x \succ_i y \text{ for all } i \in N, \text{ then } x \succ_F y.$$

This means that the entire electorate N is a *globally decisive coalition*: it can collectively determine the social ranking between any pair of alternatives. Thus, N possesses full decisive power under unanimous preference profiles.

Step 2: Recursive Application of the Contraction Lemma. Given that N is globally decisive and $|N| \geq 2$, we now apply the *Contraction Lemma*, which states that if a coalition is globally decisive and has at least two members, then there exists a strict subset of that coalition that is also globally decisive. That is, we can remove one voter and still retain the power to determine all pairwise social comparisons.

We denote the sequence of shrinking coalitions as follows:

$$\begin{aligned} S_0 &= N, \\ S_1 &\subset S_0, \\ S_2 &\subset S_1, \\ &\vdots \\ S_k &= \{i^*\}, \end{aligned}$$

where each S_j is globally decisive. Since N is finite, this recursive contraction must terminate. Eventually, we reach a singleton coalition $\{i^*\}$ that remains globally decisive.

Step 3: Singleton Decisiveness Implies Dictatorship. If a single voter i^* is globally decisive, then for any pair of alternatives $x, y \in A$, whenever $x \succ_{i^*} y$, it must follow that $x \succ_F y$, regardless of the preferences of the other voters. Formally,

$$\forall x, y \in A, \quad x \succ_{i^*} y \implies x \succ_F y.$$

This means that the social ranking produced by the function F always agrees with the individual ranking of voter i^* . Thus, i^* is a *dictator*, which contradicts the Non-Dictatorship axiom.

Step 4: Conclusion. We have shown that assuming the existence of a social welfare function that satisfies all four axioms leads to the conclusion that a dictator must exist. This contradicts the Non-Dictatorship requirement.

Therefore, no such function F exists, completing the proof.



Voter	Ranking 1	Ranking 2
1	$a \succ b \succ c$	$b \succ a \succ c$
2	$b \succ c \succ a$	$b \succ c \succ a$
3	$c \succ a \succ b$	$c \succ a \succ b$
Outcome	a	b

4.1 Gibbard–Satterthwaite Theorem

Building on Arrow’s investigation of aggregation consistency, the Gibbard–Satterthwaite theorem examines voters’ strategic incentives. It shows that any deterministic onto social choice function that selects a single winner from three or more alternatives must be dictatorial if it is strategy-proof, other words, if no voter can benefit by misrepresenting his preferences. This result starkly illustrates how the goals of fairness and strategic resistance collide in rich-choice settings.

Theorem 4.4 (Gibbard–Satterthwaite Theorem). *Formally, let A be a set of $|A| \geq 3$ alternatives and let $f : \mathcal{L}(A)^n \rightarrow A$ be a social choice function (SCF), where $\mathcal{L}(A)$ is the set of strict total orders over A . The function f is said to be:*

- **Onto:** *for every $a \in A$, there exists some profile $\vec{\succ} \in \mathcal{L}(A)^n$ such that $f(\vec{\succ}) = a$,*
- **Strategy-proof:** *see Definition 2.10*

Then the theorem asserts that f must be a dictatorship.

Proof. Intuitively this theorem can be proved by understanding the existence of a **pivotal voter**. A core idea in the proof is the concept of **monotonicity**: if $f(\vec{\succ}) = a$ and a voter raises another alternative b in his ranking while keeping the rest of the profile unchanged, then f either remains at a or switches to b ; it cannot switch to a third option without violating strategy-proofness. This idea helps identify a **pivotal voter** who determines the outcome between two alternatives.

In this example, Voter 1’s change in ballot causes the social choice to

shift from a to b , indicating that she is pivotal between a and b . Strategy-proofness ensures that Voter 1's sincere top choice among a and b must be selected whenever she is pivotal.


Repeating this argument for all alternative pairs implies that his top-ranked candidate is always the outcome, regardless of others' preferences, hence establishing dictatorship. In other words, once a voter k is identified as pivotal for a pair (a, b) , the argument extends to show that k controls the outcome between every pair of alternatives making it a dictator.

First, since f is onto, we start from a profile \succsim^0 where $f(\succsim^0) = a$, and systematically raise b across ballots one voter at a time. Monotonicity guarantees that for each intermediate profile \succsim^i , the outcome stays within $\{a, b\}$. Onto ensures that eventually some profile \succsim^n yields $f(\succsim^n) = b$. By choosing the first index k such that

$$f(\succsim^{k-1}) = a \quad \text{and} \quad f(\succsim^k) = b,$$

we establish that voter k 's single-ballot change alone flips the result—making k pivotal for (a, b) .

Once k is pivotal, a similar construction for any other alternative $c \neq a, b$ shows k is also pivotal between (a, c) . Again, monotonicity and onto produce a change in outcome exactly when k modifies his ballot. Therefore, for all $x, y \in A$, whenever k prefers x to y , the outcome is x . This satisfies the definition of a dictator.

Thus, starting from a single pivotal act, one shows that the same voter k dictates the winner for all alternative pairs. Under the assumptions of strategy-proofness and onto, the only way to avoid manipulation by misrepresenting preferences is to concentrate decision power in a single agent's truthful preference proving that f is a dictatorship. 

Interpretation of Voting Impossibility Theorems

The combined force of **Arrow’s Impossibility Theorem** and the **Gibbard–Satterthwaite Theorem** delineates the fundamental structural limits of social choice under ordinal preference aggregation. Arrow’s theorem exposes a logical inconsistency among axioms of collective rationality—*unanimity*, *independence of irrelevant alternatives (IIA)*, and *non-dictatorship*—within any social welfare function $F : \mathcal{L}(A)^n \rightarrow \mathcal{L}(A)$, where $\mathcal{L}(A)$ denotes the set of strict linear orders over a finite set of alternatives A . Gibbard–Satterthwaite complements this by showing that any deterministic, surjective, and strategy-proof social choice function $f : \mathcal{L}(A)^n \rightarrow A$ must be dictatorial when $|A| \geq 3$.

Collectively, they posit that: within unrestricted domains of ordinal preferences, no non-dictatorial mechanism can simultaneously satisfy minimal fairness and strategic robustness.

These impossibility results are not, however, an indication of the complete failure of the democratic system, but foundational basis for exploring the utility that can be extracted from them. They motivate a transition from seeking axiomatic perfection to quantitatively evaluating imperfections to make the best of what is available. By imposing additional structure—such as latent metric embeddings of preferences—one can define notions of suboptimality, such as the *distortion* of a social choice rule, which measures its worst-case efficiency loss relative to a metric social optimum.

To this end, impossibility theorems are reinterpreted not as terminal conclusions about the detrimentality of choice, but as rigorous thresholds for evaluating trade-offs between fairness, strategy-proofness, and efficiency in collective decision-making.

5 Metric Distortion in Voting Systems

In practice, most voting systems rely only on ordinal information—voters rank candidates from best to worst—without accounting for the intensity of preferences. To evaluate how this limitation affects the quality of outcomes, we adopt the framework of *metric distortion*. This model assumes voters and candidates are embedded in a latent metric space and seeks to quantify the inefficiency of a voting rule compared to the optimal outcome under that space.

5.1 Setting Up Metric Distortion

5.1.1 Spatial Preference Model

Let $P = (p_1, p_2, \dots, p_n)$ denote a preference profile for n voters over a finite candidate set $A = \{a_1, a_2, \dots, a_m\}$. We assume each voter $p_i \in R^d$ and each candidate $a_j \in R^d$ is located in a Euclidean space (or more generally, a metric space (X, d)). Voter disutility is modeled by the distance function $D(p_i, a_j)$, often taken to be Euclidean:

$$D(p_i, a_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (\text{for } d = 2).$$

5.1.2 Definition of Social Cost

Given a voting rule f that selects a winner $f(P) \in A$ based on the profile P , the **actual social cost** of the outcome is defined as:

$$D(P, f(P)) = \sum_{i=1}^n D(p_i, f(P)).$$

The **ideal social cost** is achieved by the candidate $a^* \in A$ that minimizes total distance to all voters:

$$a^* = \arg \min_{a \in A} \sum_{i=1}^n D(p_i, a), \quad \text{and} \quad D(P, a^*) = \sum_{i=1}^n D(p_i, a^*).$$

5.1.3 Definition of Metric Distortion

Metric distortion is defined as the worst-case ratio between the actual social cost incurred by the winner chosen by the voting rule and the optimal cost achievable:

$$\text{Distortion}(f) = \sup_{P,D} \frac{D(P, f(P))}{D(P, a^*)},$$

where the supremum is taken over all profiles P and all metric embeddings D consistent with the given preference orderings.

5.1.4 Interpretation

Intuitively, a distortion of δ means that in the worst case, the outcome selected by the voting rule incurs total cost up to δ times higher than the best possible candidate. A lower distortion reflects a rule that more faithfully approximates the utilitarian optimum, despite only having access to ordinal input.

5.2 A Visual Illustration of Metric Distortion

To build intuition for the concept of metric distortion, consider a simplified example where both voters and candidates are embedded in the Euclidean plane R^2 . In this spatial model, each voter prefers candidates who are geometrically closer, and disutility is measured using Euclidean distance. The social cost of a candidate is thus the sum of their distances to all voters.

Figure 1 illustrates a small instance with three voters $x_1, x_2, x_3 \in R^2$ and three candidates $a_1, a_2, a_3 \in R^2$. Dashed lines represent the disutilities (i.e., distances) from each voter to all candidates.

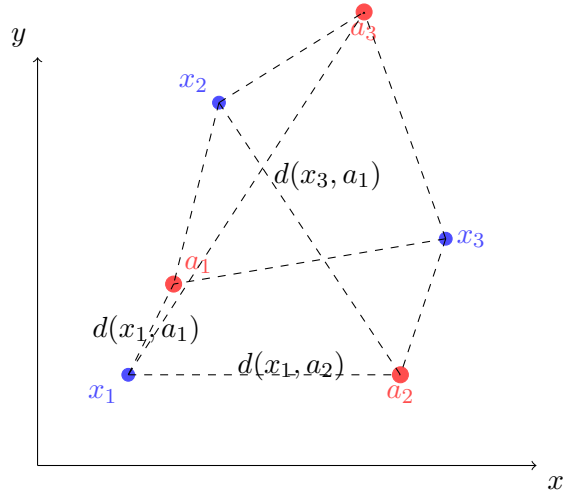


Figure 1: Voters x_1, x_2, x_3 and candidates a_1, a_2, a_3 embedded in R^2 . Dashed lines represent Euclidean disutilities.

Interpretation

Suppose a voting rule selects candidate a_2 based on first-place votes, despite candidate a_1 being, on average, geometrically closer to all voters. Then:

- The *actual social cost* is $D(P, a_2) = \sum_{i=1}^3 d(x_i, a_2)$,
- The *optimal social cost* is $D(P, a_1) = \sum_{i=1}^3 d(x_i, a_1)$,
- The *distortion* is given by the ratio $\frac{D(P, a_2)}{D(P, a_1)}$.

This simple geometric setup captures the essence of metric distortion: even though the underlying spatial structure favors candidate a_1 , a rule that only uses ordinal rankings may ignore this and select a worse candidate in terms of social cost. This inefficiency—arising from ignoring metric intensity—is what distortion seeks to quantify.

6 Ordinal vs. Cardinal Preferences

Let $A = \{a_1, a_2, \dots, a_m\}$ be a finite set of candidates and $N = \{1, 2, \dots, n\}$ the set of voters. Each voter $i \in N$ has either:

- a **cardinal utility function** $u_i : A \rightarrow R$, or
- an **ordinal preference** \succ_i , a total order on A .

Definition 6.1 (Cardinal Preferences). A cardinal utility $u_i(a)$ quantifies exactly how much voter i prefers candidate a . Under utilitarian aggregation, the socially optimal alternative is

$$a^* = \arg \min_{a \in A} \sum_{i=1}^n u_i(a).$$

Definition 6.2 (Ordinal Preferences). An ordinal preference \succ_i records only the ranking:

$$a \succ_i b \iff u_i(a) > u_i(b),$$

but not the magnitude $|u_i(a) - u_i(b)|$. Classical voting rules—Plurality, Borda, Instant-Runoff—use only these rankings.

6.1 Strategic and Practical Constraints

1. *Elicitation Cost:* Reporting full $u_i(a)$ is cognitively burdensome.
2. *Manipulation Risk:* Voters can distort cardinal reports to influence outcomes.
3. *Institutional Norms:* Ballots typically collect only rank-orders.

Thus, in practice, only the profile of orderings $\vec{\succ} = (\succ_1, \dots, \succ_n)$ is available to the voting mechanism.

6.2 How Metric Distortion Uses Ordinal Values

We posit that there exists some (unknown) metric d and embeddings x_i for voters and y_j for candidates so that

$$a \succ_i b \implies d(x_i, y_a) < d(x_i, y_b).$$

However, the mechanism observes only \succ , not the distances.

A voting rule f maps \succ to a chosen candidate $f(\succ)$. We then compare total disutilities:

$$D(P, f(\succ)) = \sum_{i=1}^n d(x_i, y_{f(\succ)}), \quad \min_{a \in A} \sum_{i=1}^n d(x_i, y_a).$$

The *metric distortion* of f is

$$(f) = \sup_{\substack{d, x_i, y_j \\ \text{consistent with } \succ}} \frac{\sum_i d(x_i, y_{f(\succ)})}{\min_{a \in A} \sum_i d(x_i, y_a)},$$

where the supremum ranges over all metrics and embeddings that induce the same ordinal profile \succ . In this way, metric distortion quantifies how much worse f can perform in terms of true (cardinal) social cost, given only ordinal inputs.

7 Worst-Case Distortion Bounds

In the metric distortion framework, we analyze how well a voting rule approximates the socially optimal candidate when only ordinal preferences (rankings) are observed. Since these rankings derive from an unknown metric space, we are interested in the *worst-case distortion*: the maximal inefficiency a rule may incur across all consistent metrics.

7.1 Formal Setup

Let $A = \{a_1, \dots, a_m\}$ be a set of candidates, and let $\succ = (\succ_1, \dots, \succ_n)$ be the profile of voter rankings, where each \succ_i is a total order over A .

Assume a hidden metric space (X, d) such that:

$$a \succ_i b \implies d(x_i, y_a) < d(x_i, y_b),$$

where $x_i \in X$ is the location of voter i , and $y_a \in X$ is the location of candidate a . These locations are unobservable.

The *social cost* of candidate a is:

$$(a) = \sum_{i=1}^n d(x_i, y_a),$$

and the optimal candidate is:

$$a^* = \arg \min_{a \in A} (a).$$

A voting rule f chooses a candidate based only on \succ , and its *worst-case distortion* is defined as:

$$(f) = \sup_{\substack{(X,d), x_i, y_a \\ \text{consistent with } \succ}} \frac{(f(\succ))}{(a^*)}.$$

This measures how far the selected winner can be from the true optimum in terms of aggregate voter disutility.

7.2 Universal Lower Bound (Deterministic)

Theorem 7.1 (Gkatzelis–Halpern–Shah, 2020). *For any deterministic voting rule f , we have:*

$$(f) \geq 3.$$

Proof Sketch. Consider a profile with three voters and three candidates. Construct rankings such that each voter ranks a different candidate first.

Then, place the voters at vertices of an equilateral triangle, and place the candidates so that the winner chosen by the rule (e.g., a fixed tie-breaking) lies close to one voter and far from the other two. In contrast, the optimal candidate lies at the centroid, equidistant from all voters. This setup achieves a distortion arbitrarily close to 3.

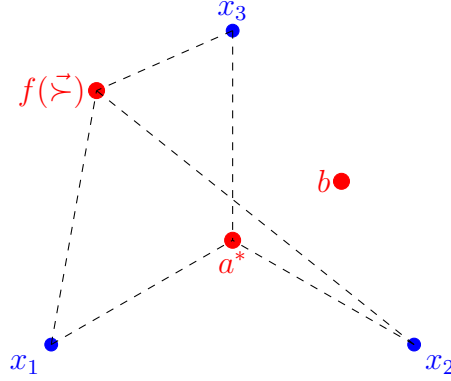


Figure 2: Worst-case distortion with 3 voters and 3 candidates.

In this example:

$$(a^*) \approx 3 \cdot r, \quad (f(\vec{\succ})) \approx r + 2R \Rightarrow \frac{r + 2R}{3r} \rightarrow 3 \text{ as } R/r \rightarrow \infty.$$

7.3 Specialized Rules

Theorem 7.2 (Anshelevich et al., 2015). *There exists a deterministic rule with distortion exactly 3.*

One such rule is *Plurality Veto* (Kizilkaya–Kempe, 2022). It selects the candidate who minimizes the number of times they are ranked last, breaking ties by fewest second-to-last rankings.

7.4 Randomized Rules

Allowing randomization can lower distortion. Let f be a probabilistic voting rule, outputting a distribution over candidates.

Theorem 7.3 (Charikar–Ramakrishnan, 2023). *There exists a randomized rule with:*

$$(f) < 2.73.$$

The best known **lower bound** for any randomized rule remains:

$$(f) \geq 2,$$

8 Real-World Implications and Broader Perspectives

While the theoretical bounds on fairness and distortion may appear abstract, they carry deep consequences for how societies design democratic institutions, allocate resources, and ensure equity in collective decisions.

The core impossibility results show that no voting system can be fully fair, non-dictatorial, and immune to manipulation. This motivates institutional humility: policymakers must accept trade-offs and clearly prioritize which values (e.g., fairness, resistance to strategic voting, or representational accuracy) matter most in their contexts.

The concept of metric distortion introduces a powerful lens for comparing electoral systems even when only ordinal ballots are available. By modeling voters and candidates in a latent metric space (e.g., political ideology, geographic proximity, or utility space), we can ask: How far from optimal was the result? This enables empirical audits of existing rules—such as plurality or ranked-choice voting—using simulations or survey-embedded spatial data.

Conclusion

This paper has shown that the dream of a perfectly fair voting system must yield to mathematical impossibility—but that this does not render all hope lost. By quantifying inefficiencies through the framework of metric distortion, we recover a new kind of possibility: an efficient, if imperfect, approximation to the ideal. This shift from the absolute to the approximate, from structure to geometry, and from proof to practice, reflects the ongoing power of mathematics to inform not only what is, but what could be.

References

- Arrow, Kenneth J. *Social Choice and Individual Values*. 2nd ed., Yale University Press, 1963.
- Conitzer, Vincent. “Voting.” *Introduction to Computational Social Choice*, edited by Felix Brandt et al., Cambridge University Press, 2016, pp. 37–60. <https://doi.org/10.1017/CB09781107446984.007>.
- Gibbard, Allan. “Manipulation of Voting Schemes: A General Result.” *Econometrica*, vol. 41, no. 4, 1973, pp. 587–601. <https://doi.org/10.2307/1914083>.
- Geanakoplos, John. “Three Brief Proofs of Arrow’s Impossibility Theorem.” *Economic Theory*, vol. 26, no. 1, 2005, pp. 211–215. <https://doi.org/10.1007/s00199-004-0507-z>.
- Kizilkaya, Fatih Erdem, and David Kempe. “Plurality Veto: A Simple Voting Rule Achieving Optimal Metric Distortion.” *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, 2022, pp. 358–364. <https://www.ijcai.org/proceedings/2022/0050.pdf>.
- Peters, Jannik. “A Note on Rules Achieving Optimal Metric Distortion.” *arXiv*, 2022. <https://arxiv.org/abs/2209.14430>.
- Charikar, Moses, et al. “Breaking the Metric Voting Distortion Barrier.” *arXiv*, 2023. <https://arxiv.org/abs/2302.03498>.
- Anshelevich, Elliot, and John Postl. “Randomized Social Choice Functions Under Metric Preferences.” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015. <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9823>.
- Gkatzelis, Vasilis, Daniel Halpern, and Nisarg Shah. “Resolving the Optimal Metric Distortion Conjecture.” *Journal of the ACM*, vol. 67, no. 6, 2020. <https://doi.org/10.1145/3366141>.
- Acemoglu, Daron. “Lecture 12: Political Economy of Institutions and

Voting.” *MIT OpenCourseWare*, 2012. https://ocw.mit.edu/courses/14-75-political-economy-and-economic-development-fall-2012/a9fd8e5ab75a325016094e6bbe625b2a_MIT14_75F12_Lec12.pdf.

Brandt, Felix, et al., editors. *Introduction to Computational Social Choice*. <https://pure.uva.nl/ws/files/163401688/introduction-to-computational-social-choice.pdf>.

List, Christian, and Philip Pettit. “Social Choice Theory.” *The Stanford Encyclopedia of Philosophy*, Fall 2020 Edition, edited by Edward N. Zalta. <https://plato.stanford.edu/entries/social-choice/#GibbSattTheo>.

Acknowledgement

The author would like to thank Addison Prairie and Simon Rubinstein-Salzedo for their helpful conversations.