Are democracies fair ?

Arrow’s impossibility theorem or the mathematical impossibility of fair voting

By Marion Agbo-Bloua

“The voting system [...] is omnipresent and omniscient. And it is fallible. Yet, when a structure is broken, we are fools if we simply ignore the defect [...]. Our obligation is to understand where the problem is, find a solution, and make the broken whole again.”
— **Stacey Abrams**

I/ Introduction: What is mathematics doing in politics?

I am an avid fan of both geopolitics and mathematics. At first glance, these two fields, like two parallel lines, seem destined to never cross paths. However, mathematics is present and plays a critical role in shaping decisions, notably through computational modelling, game theory and data analysis. Having said that, it might come to your surprise (or not) that this theorem was created by a scholar who was both a mathematician and a political theorist (although most commonly referred to as an economist): Dr Kenneth Arrow.

Dr Kenneth Arrow first developed this theory in his PhD thesis, before publishing it in his book *Social Choice and Individual Values* in 1951. This theorem, among other major contributions, earned him the Nobel Memorial Prize in Economic Sciences in 1972. All this to say that Arrow’s impossibility theorem is a keystone in social choice theory (a branch of mathematics closely related to game theory) defined as: “*the study of collective decision procedures and mechanisms.”*

But for all its importance, it is often misinterpreted. It can be thought of as a critique or even a suggestion that a democratic system is in itself inherently impossible; which in turn marks its presence in fallacious arguments, fearmongering, or even essays or videos with clickbait titles (my essay title is an example but, as the saying goes “a fault confessed is half redressed”).

What the theory attempts to prove is not an opinion on political systems but rather that ranked voting, which we will discuss later, can not take into account all fairness criteria simultaneously.

II/ A brief history of social choice theory and is ranked voting the solution?

In a democracy, every vote counts, but the way you count these votes changes the final outcome.

Let's examine three different voting systems, in the same given situation, and the generated results. These examples are not the theorem, rather an initial illustration of the impossibility of "fairness":

- 4 candidates A, B, C and D present themselves to their local election
- A and B are ideologically similar, C is neutral, and D is opposite to all three.
- We assume that voters who prefer:
  - A rank B above C or D and vice versa.
  - C rank A and B above D.
  - if their first choice is absent their votes will transfer to their second
- Let $T(X)$ be the number of votes received by candidate X.
- Let $T(A) < T(B) < T(C) < T(D)$, $T(A) + T(B) > T(C)$ and $T(A) + T(B) + T(C) > T(D)$

Let's start with first-past-the-post-voting, which consists of one voting round where whoever has the most votes is elected.

Since $T(A) < T(B) < T(C) < T(D)$, candidate D would win.
However, the majority did not vote for candidate D as $T(A) + T(B) + T(C) > T(D)$.

A candidate can thus win without majority support.

This gave rise to Duverger's law, a law that states: "*Plurality voting tends to produce two-party systems.*" Or the "spoiler effect" wherein voters will strategically place their votes on parties similar to their preferred party but with a higher chance of winning.

Since $T(A) < T(B) < T(C) < T(D)$ and $T(A) + T(B) > T(C)$ we have
$T(C) < T(A) + T(B) < T(D)$
With D still winning.

If voters know that A and B are less likely to win than C or D but that A and B prefer C to D, A's and B's votes will transfer to C's.

This would mean:
$T(C)' = T(A) + T(B) + T(C)$
Since $T(A) + T(B) + T(C) > T(D)$ we have $T(C)' > T(D)$

Giving victory to C.

Strategic voting can completely change the outcome.

Instant runoff voting. Three or more parties go through multiple rounds and each round the party with the least amount of votes gets eliminated. The last one wins.

Since $T(A) < T(B) < T(C) < T(D)$.
A is eliminated in the first round.

Due to ideological similarity, A’s votes will transfer to B’s, such that:
$T(B)' = T(A) + T(B)$, since $T(A) + T(B) > T(C)$, and $T(B) < T(D)$
we have $T(C) < T(B)' < T(D)$.
C is eliminated in the second round.

Since C prefers B to D, C’s votes will transfer to B’s, such that:
$T(B)'' = T(A) + T(B) + T(C)$ since $T(A) + T(B) + T(C) > T(D)$
we have $T(B)'' > T(D)$
Giving victory to B.

In each of these systems, a different outcome arises. This shows both that the outcome depends heavily on the voting system and that none of them can simultaneously reflect all voter preferences in a consistent way.

To attempt to solve this last issue, another type of voting system was proposed: ranked voting.

Ranked voting is a system (re)discovered by the Marquis de Condorcet, an 18th century French political philosopher and mathematician. To be “Condorcet’s winner” you had to win through a ranking, with each candidate competing “*against every other candidate in a one-on-one election, [to] win overall.*”

Let there be n voters and a set of candidates:
Let $P(A, B) = \frac{number\ of\ voters\ who\ prefer\ A\ over\ B}{n}$

A is a condorcet winner if: $P(A, X) > \frac{1}{2}$ for all candidates $X \neq A$.

This implies that if A is elected, the strict majority prefers A over any other alternative.
If another candidate $Y \neq A$ is elected instead, since $P(A, Y) > \frac{1}{2}$ a majority of voters prefer A over Y, and electing Y or any other candidate would contradict the majority preference.

The Condorcet method is appealing as it takes into account all possible pairwise preferences. It’s advantageous compared to first-past-the-post that only considers first preferences and Instant runoff that can eliminate the Condorcet winner in intermediate rounds.

However, this system has one downside: the Condorcet’s or voting paradox.

III/ The proof or the actual mathematical problem

We have seen different examples of voting systems each with distinct results, pros and cons. But are any of them *really* fair? And if not, is a perfectly fair voting system possible?

Condorcet’s paradox

Let there be a set of alternatives such that $S = \{A, B, C\}$ and a set of voters V_1, V_2 and V_3. Each voter has a preference ordering, such that:

V_1: $A > B > C$
V_2: $B > C > A$
V_3: $C > A > B$

Dr Kenneth Arrow proposes 5 conditions that a fair voting system should satisfy.

1. **Unanimity or Pareto efficiency**: if everyone in the group chooses one option over another, the outcome should reflect this decision.
2. **Non-dictatorship:** no single person's vote should override the preferences of others.
3. **Unrestricted domain:** The system must accept all logically possible preferences.
4. **Transitivity**: The voting system should be logically consistent. if $A > B$ and $B > C$ then $A > C$.
5. **Independence of irrelevant alternatives (IIA):** The ranking between A and B depends only on how voters rank A and B, no other alternative should affect this ranking.

These conditions seem obvious enough, however, if we apply them to Condorcet's method, we have:

For A and B:
V_1 and V_3 prefer $A > B$, so the majority prefers $A > B$
Hence $P(A, B) = \frac{2}{3} > \frac{1}{2}$

For B and C:
V_1 and V_2 prefer $B > C$, so the majority prefers $B > C$
Hence $P(B, A) = \frac{2}{3} > \frac{1}{2}$

For A and C:
V_2 and V_3 prefer $C > A$, so the majority prefers $C > A$
Hence $P(C, A) = \frac{2}{3} > \frac{1}{2}$

This means we have three Condorcet's winners, which should be impossible.

Furthermore, according to transitivity: $A > B$ and $B > C$ therefore, $A > C$ however, $C > A$. This inequality creates a cycle with no overall winner: $A > B > C > A$.

It is not logically consistent, and the ranking between two variables is not fixed (IIA). meaning this voting system does not meet all of Arrow's requirements, even with rational preferences following majority rule.

Condorcet's paradox is one of many examples that illustrate Arrow's impossibility theorem. "*No social choice function can satisfy all five [conditions] simultaneously for three or more alternatives.*" However the paradox is but an example, not the rule. So how do we prove this law?

Arrow's proof

Let there be a set of alternatives such that $S = \{A, B, C\}$ and a set of voters V_1, V_2, V_3, V_4 and V_5.

Let's assume every voter prefers A over B over C, therefore, by unanimity, the social ranking is $A > B > C$.

Because of IIA, the overall ranking between A and B depends only on how voters rank A versus B. We can therefore study A and B independently of C.

Let's say A loses favour, and its ranking decreases every round. We shall change voter preference one by one from $A > B$ to $B > A$.

If $A > B$ we will write A, if $B > A$ we will write B.

Voter	round 0	round 1	round 2	round 3	round 4	round 5
V_1	A	B	B	B	B	B
V_2	A	A	B	B	B	B
V_3	A	A	A	B	B	B
V_4	A	A	A	A	B	B
V_5	A	A	A	A	A	B
majority	A	A	A	B	B	B

Since $A > B$ becomes $B > A$, there comes a round where the previous decision of the majority is overturned. Therefore, there must exist a first voter, V_D (for decisive voter), whose change causes this reversal, wherein whenever they rank $A > B$, the overall ranking is $A > B$. (Here $V_D = V_3$).

Because of IIA, the rankings of two candidates depend only on voter preference. If a voter is decisive for A over B then it implies that, by similar reasoning, this same voter will also be decisive for A over C, and B over C. Using the same method, we will rearrange the voters' preference in regards to C.

Voter	round 0	round 3
V_1	A	C
V_2	A	C
V_D	A	C
V_4	A	A
V_5	A	A
majority	A	C

Voter	round 0	round 3
V_1	B	C
V_2	B	C
V_D	B	C
V_4	B	B
V_5	B	B
majority	B	C

We find that V_D changes once again the positions such that $C > B$, $B > A$ and $C > A$. Therefore, according to transitivity $C > B > A$. This means that due to V_D's votes, we move from $A > B > C$ to $C > B > A$.

If we repeat this argument for all possible pairs, a voter who is decisive for a pair must, under IIA, be decisive for all. Consequently, the majority follows V_D's preference, which dictates the overall decision. Since this power over the collective lies in only one voter, V_D is therefore a “dictator”, which goes against the non-dictatorship rule and so against one of Arrow’s five conditions.

IV/ Conclusion:

Arrow’s impossibility theorem is thus a two-dimensional theory: a political one and a mathematical one. It does not prove that democracy is flawed beyond repair, but rather that a perfectly fair voting system is (mathematically) impossible. Democracy becomes no longer about perfection but rather compromise and what we should prioritise. Nonetheless, we should not overlook that it is mathematics that manages to prove these limits and create this reflection on how to better our society. Because, after all, mathematics and geopolitics go hand in hand.

Bibliography

- *Social choice theory, **Stanford encyclopedia of philosophy**, 2nd revision, 14 october 2022*

- *Arrow's impossibility theorem, **Wikipedia**, last updated 29 march 2026*

- *Condorcet's paradox **Wikipedia**,last updated 18 february 2026*

- ***Kulkarni Abhishek**, Arrow's impossibility theorem: Can we have a fair voting system? **Science ABC**, last updated 19 oct 2023*

- ***Veritasium**, Why democracy is mathematically impossible, **Youtube**, 27 august 2024*

- Sandle Gen, What's wrong with first past the post?, **Electoral reform society**, 7 may 2025

- *Voting system quotes **Goodreads**:*

- Political science, bad mathematics, maths **Reddit**

- Lutzen Jesper, How mathematical impossibility changed welfare economics: A history of Arrow's impossibility theorem, **Science direct**, 18 february 2019

- Arrow Kenneth, A Difficulty in the concept of social welfare, **Maths university of chicago**