

Arrow

Wednesday, January 14, 2026

Overview



Figure 1: Kenneth Arrow, photo by Linda A. Cicero

- Lived 1921-2017.
- Worked primarily at Stanford.
- Won Nobel Prize in 1972.
- Could have won it twice over, once for work on equilibrium theory, and once for what we'll cover.

Arrow's Impossibility Theorem (1951) demonstrates that no voting system can simultaneously satisfy a set of seemingly reasonable conditions for democratic decision-making.

Definition: An individual preference ranking is an ordering of all available alternatives from most preferred to least preferred, with ties allowed, but not gaps.

We'll write $A \succ_i B$ to mean that person i prefers A to B , and $A \simeq_i B$ to mean that person i thinks that A and B are equally good.

Key Assumptions:

- **Completeness:** Exactly one of $A \succ_i B$, $B \succ_i A$, and $A \simeq_i B$ holds.
- **Transitivity of Better:** If $A \succ_i B$ and $B \succ_i C$, then $A \succ_i C$.
- **Transitivity of Equal:** If $A \simeq_i B$ and $B \simeq_i C$, then $A \simeq_i C$.

We'll write $A \succ_V B$ to mean that the group prefers A to B, and $A =_V B$ to mean that the group thinks they are equally good.

Same Key Assumptions

- **Completeness:** Exactly one of $A \succ_V B$, $B \succ_V A$, and $A =_V B$ holds.
- **Transitivity of Better:** If $A \succ_V B$ and $B \succ_V C$, then $A \succ_V C$.
- **Transitivity of Equal:** If $A =_V B$ and $B =_V C$, then $A =_V C$.

How do we combine individual rankings into a single social ranking?

What We Want: A systematic method (social choice function) that takes individual preference profiles and produces a group preference ranking.

Background

There are a few strange things about the setup.

- Most of us don't have views about all possible outcomes of an election.
- The combination just takes individual *rankings* into account. It doesn't look at *strength* of preference.
- The output doesn't just ask for a winner, but for a full ordering.

Say $A \succ_{\forall} B$ if more people have $A \succ_i B$ than have $B \succ_{\forall} A$, and $A =_{\forall} B$ if the numbers are the same.

This is a very simple kind of majoritarian rule.

You can probably see where this is going.

- Imagine that there are three options, A, B and C.
- And there are three voters, X, Y and Z.
- X has $A \succ_X B \succ_X C$.
- Y has $B \succ_Y C \succ_Y A$.
- Z has $C \succ_Z A \succ_Z B$.

- Since two people (X and Z) prefer A to B, we have $A \succ_V B$.
- Since two people (X and Y) prefer B to C, we have $B \succ_V C$.
- Since two people (Y and Z) prefer C to A, we have $C \succ_V A$.
- And we've violated transitivity.

Let's try a different solution.

- Ignore every part of \succ_i except who is top.
- For each option, give it 1 point for every voter for which it's the top option.
- If the voter has two options tied for top, give them 1/2 point each, and so on for more ties.
- Rank the options by how many points.

- This is guaranteed to satisfy the assumptions.
- Any approach where each option gets a number, and you rank the numbers, will produce the right kind of ordering.
- But it has a very strange result.

Imagine there are three options, A, B and C, and 100 voters. On Monday the preferences are like this:

- 40 people have $A > B > C$.
- 35 people have $B > C > A$.
- 25 people have $C > B > A$.
- So A wins, with 40 votes, and B is second, with 35.

On Tuesday, 10 of the people in the last group change their mind, just about B and C. So now we have.

- 40 people have $A > B > C$.
- 45 people have $B > C > A$.
- 15 people have $C > B > A$.
- So now B moves ahead of A in the social ranking, even though no one changed their mind about A and B.
- Is this odd?

The Theorem

Rather than going through options one at a time, let's find some principles.

- Start with a list of things we'd like the aggregation to be like.
- See what kind of things could meet all of the requirements on our list.

For a very natural set of requirements, there are **zero** aggregation rules that satisfy all of them.

- Any aggregation rule you pick will have some slightly odd feature.

Arrow identified four conditions that seem essential for any fair democratic system:

1. Unrestricted Domain (U)
2. Weak Pareto Principle (P)
3. Independence of Irrelevant Alternatives (I)
4. Non-dictatorship (D)

What it means: The voting system must work for ANY possible combination of individual preference rankings.

Why it seems reasonable: We shouldn't restrict what preferences people are allowed to have in a democracy.

What it means: If everyone prefers alternative A to alternative B, then the group ranking should also prefer A to B.

Why it seems reasonable: If there's unanimous agreement, the social choice should reflect that agreement.

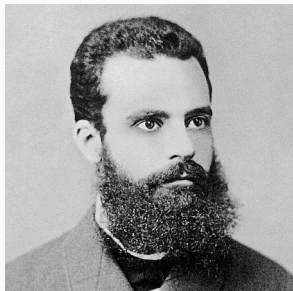


Figure 2: Vilfredo Pareto

- Lived 1848-1923.
- Trained in Turin, worked primarily at Lausanne.
- He died around the time Mussolini came to power, and the Italian fascists claimed him as an inspiration. This *may* have been a misreading; he often sounds like a classical liberal.

Weak Pareto: If for all i , $A \succ_i B$, then $A \succ_{\forall} B$.

Strong Pareto: If for some i , $A \succ_i B$, and for no i $B \succ_i A$ (but maybe there are some ties), then $A \succ_{\forall} B$.

The latter is more contentious, but it isn't needed for the proof.

(In fact, all that the proof really needs is that every outcome is possible.)

What it means: The social ranking between any two alternatives should depend *only* on how individuals rank those two alternatives, not on how they rank other alternatives.

Key Insight: If we're deciding between A and B, it shouldn't matter what people think about C.

Formal Definition: For any two preference profiles R and R' , if every individual has the same preference between alternatives A and B in both profiles, then the social choice between A and B must be the same in both profiles.

This isn't a property of how the aggregation function treats what happens on any given day; it's a property of how it treats possible changes in individual preferences.

What it means: No single individual's preferences should always determine the group ranking, regardless of what others prefer.

Why it seems reasonable: This is basic to democratic ideals - no one person should have absolute power.

Formal statement: For every individual i , there is some profile R and some pair of options A and B such that in R , $A \succ_i B$, but not $A \succ_{\forall} B$.

Note: This doesn't prevent someone from being influential, just from being decisive in every case.

You might have wanted a stronger condition, like this one. But it turns out D is enough to get the result.

Anonymity (A) If two people switch their preferences, so on Tuesday X has the views that Y had on Monday, and Y has the views that X had on Monday, that shouldn't change the answer.

Arrow's Theorem: There is no social choice function that simultaneously satisfies:

- Unrestricted Domain (U)
- Weak Pareto Principle (P)
- Independence of Irrelevant Alternatives (I)
- Non-dictatorship (D)

I'm not going to make you sit through a proof here. There are a million proofs out there if you want to find one.

The usual structure is to assume U , P and I , and infer the existence of a dictator.

The main benefit of sitting through a proof is that you see just how strong a condition I is, and we'll come back to that strength next time.

We'll look at ways out of, or around, the problem.