Consequences of Arrow

2026-01-21

The Theorem

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

Arrow focussed on four requirements you'd like a system to have.

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

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)

Consequences

You'll often read people saying that Arrow's Theorem is a particular problem for **democratic** systems.

It seems to show that there is no system of aggregating votes that meets some really plausible conditions.

So is that bad news for voting systems?

- Yes, but...

Let's say that instead of democracy your preferred social model is benevolent dictatorship, where an all powerful ruler acts in the best interests of the people collectively.

- Aside one: If this was a history of social thought class, we'd walk through how popular a model this was among reformers through the 19th century.
- Aside two: For things where the government has to act although approximately zero votes will turn on how they act, this is our actual system. Think about things like the city council deciding the rules for a particular intersection.

If our dictator genuinely cares about the people, then what they do should be sensitive to their interests collectively. And that presumably means:

1. They have to act no matter what those interests are (U)
2. They should do A rather than B if it's in everyone's interests (W)
3. Whether they do A or B should only turn on which of those two is better for the people (I)
4. They shouldn't act just in the interests of one person (D)

Note that the condition D does not just say there shouldn't be one guy making all the decisions.

It says that the decisions shouldn't be made just sensitive to what one person wants, or in the interests of one person.

So really all the conditions might be met, even for a benevolent dictator.

Well maybe you don't like dictators; I don't particularly either.

Perhaps you think a market system of free trade between free people will produce a better outcome for the people collectively.

That's an interesting claim, and actually one that Arrow himself did some important work on formalising.

But we still need to define what it is for something to be a better outcome.

And we'll now have the same problem the benevolent dictator faced.

- Our theory of better outcomes should work no matter what's good or bad for individuals (U)
- What's better for everyone should be better overall (W)
- Whether A or B is better overall should just turn on which of them is better for individuals (I)
- Whether A or B is better overall should not just turn on what's better for one person (D)

Arrow's Theorem does raise a challenge to democratic systems.

But it raises more or less the same challenge for non-democratic systems as well.

What it really challenges is the idea of a public good, and that's something that democrats, benevolent dictators, and free marketeers, all care about.

Two Partial Ways Out

I won't spend much time on this, but it turns out that there are some pretty interesting results on how far you can go if you restrict U.

The most significant result concerns what are called **single-peaked preferences**.

Imagine that there is some linear scale you can place the options on, and for every voter, the options satisfy the following condition.

- The voter has a preferred spot on the scale.
- If A is between B and that preferred spot, then the voter prefers A to B.
- E.g., all the candidates fall on a left-right spectrum, and so do all the voters, and a voter will never prefer to move away from their preferred spot on the spectrum.

Given this constraint, majority rule works. You never get Condorcet style cycles.

- Is this realistic?
- It sometimes is! If you think in advance that your voter pool meets this condition, simple majority rule is a perfectly good rule.
- In a different class I'd work through a more precise statement of what single-peaked means, because it's actually a little more generous a condition than it looks, but we'll leave it go here unless people have particular questions.

Amartya Sen, who we'll talk more about in future lectures, argued for a while that we should give up the constraint that the output should be an ordering.

In particular, he thought that we should give up the constraint that the output should be complete.

So let's say that we want an output to be a pair $>_\forall$ and $=_\forall$, and our rule is unanimity.

If everyone prefers A to B, then A $>_\forall$ B, and if everyone ranks them equally, then A $=_\forall$ B.

We still insist that at most one of A $>_\forall$ B, B $>_\forall$ A, and A $=_\forall$ B holds, but we don't insist that at least one holds.

This rule satisfies all our constraints, except for completeness of the social ordering.

- It gives a result whatever the input, satisfying U.
- If everyone prefers A to B, so does the social ordering, satisfying W.
- The social ordering of A and B only turns on what people think about those options, satisfying I.
- And anyone might not get their way on A and B, satisfying D.

The problem is that this is an incredibly weak rule.

- It just says, choose Pareto dominant options.
- It doesn't say what to do when none of them exist.
- Moreover, dropping completeness essentially only gets you this extra rule.
- Allan Gibbard (a long time member of the philosophy department at UM) proved that the only rule satisfying all four constraints but not completeness is unanimity among 'oligarchs', where membership of the oligarchy could be anything.

Independence of Irrelevant Alternatives

Here's a research topic that at one time I thought would be interesting.

- Look through actual existing voting systems.
- Since they exist, they must be possible.
- And so they must violate one of the Arrow constraints.
- Question: For each system, which one do they violate?

- They violate Independence of Irrelevant Alternatives (I).
- They **all** violate I.
- It wasn't an interesting research topic because that's always the answer.

When you see weird failures of plurality, like when two similar candidates lose because of vote splitting, it's tempting to say that the problem is that plurality voting violates I.

- That can't be right on its own; all systems violate I.
- What could be a good complaint is that plurality leads to *frequent* violations of I in *realistic* situations, but the existence of violations can't be the problem.

Here's one difference between IIA and real-life voting systems.

- Voting systems tend to care a lot about which option a voter ranks
  first. (Plurality says this is all that matters, which is extreme, but all
  real life systems have this feature>0
- The spirit behind IIA is that this shouldn't matter.
- If I prefer A to B, that should count the same whether or not A is my
  first choice.

So one question to think about is basically, does being a voter's first choice matter, or is this just another spot on the preference ranking.

I think there's a reasonable case that being first should matter.

- Think about football fandom (to use a timely but non-political example).
- It matters a lot more to most people whether their most preferred team (i.e., the team they're a fan of) wins, than whether it's their 12th or 13th favorite team that wins.
- That's the intuition behind voting systems, and I think it's an OK one.

Numerical Measures

Another way aggregation systems get around Arrow is by using numerical scales.

- Amazon reviews average out numbers from 1-5.
- Traffic planners try to minimise everyone's time spent at an intersection (conditional on the intersection being safe).
- How does this interact with the Arrow principles?

The answer is that it violates one of the very first assumptions we made.

- Arrow says that the aggregation system should take the preferences of the voters as an input, and return a social ordering as the output.
- But numerical measures of how good an option is for a citizen/voter, can't be deduced from their preferences.

In the two tables below, Alex and Charlie are 'voters', and the numbers show how they rate (on a 1-5 scale), blickets and widgets. The left table is what they think on Monday, the right table is Tuesday.

Table 1: Monday's ratings

Voter	Blicket	Widget
Alex	5	1
Charlie	2	3
Average	3.5	2

Table 2: Tuesday's ratings

Voter	Blicket	Widget
Alex	3	2
Charlie	1	5
Average	2	3.5

Table 3: Monday's ratings

Voter	Blicket	Widget
Alex	5	1
Charlie	2	3
Average	3.5	2

Table 4: Tuesday's ratings

Voter	Blicket	Widget
Alex	3	2
Charlie	1	5
Average	2	3.5

- On Monday, blickets have a higher average score than widgets.
- On Tuesday, it's the other way around.
- But no one's preferences changed.

Technically, this is a violation of I.

But more importantly, it's a violation of functionality.

The social ordering is not a function, in the mathematical sense, of the preferences of the voters.

Any numerical system will be like this.

So why isn't this the solution? Why don't we just use numerical measures of how good an option is for each person.

- One reason is one we discussed in voting systems; it really doesn't work in voting because the incentives to vote 'strategically' are too high and too easy to spot.
- But another reason comes from philosophy.

If the benevolent dictator wants to use numbers not rankings, they have to generate what are called "Interpersonal Utility Comparisons".

The standard view in mid-century philosophy/economics/social science was that these were impossible.

The canonical argument for this is Lionel Robbins's argument in his 1938 paper "Interpersonal Comparisons of Utility: A Comment", which turned on the difficulty of measuring comparisons between people.

For the benevolent dictator, or the free market exponent, this turns out to be the big question.

If you can say not just that A is better than B for citizen C, but you can measure how much better it is, then you should use that quantity in making social decisions.

If not, you're back in Arrow's framework, and you have to think about which condition to give up.

We'll look at a surprising argument that sometimes the right condition to give up is the Pareto optimality condition.

Some plausible looking systems of social choice end up violating it, and perhaps that's a good thing.