ARROW'S THEOREM

444 Lecture 10

Brian Weatherson

2024-02-15

AIM

Find a function that takes individual (voter) preferences as input, and returns a group preference as output.

THEOREM

There is no function that meets five natural constraints.

1. Unanimity
2. Non-Dictatorship
3. Unrestricted Domain
4. Social Ordering
5. Independence of Irrelevant Alternatives

UNANIMITY

If everyone prefers X to Y, the social order prefers X to Y.

NON-DICTATORSHIP

For any person, there is some set of preferences the other people might have such that the group ordering and that person's ordering might disagree.

UNRESTRICTED DOMAIN

The function can take *any* individual preferences as input.

SOCIAL ORDERING

- For any X, Y, the output says either X > Y, or Y > X, or X = Y.
- If in the output, X > Y, and Y > Z, then X > Z.
- If in the output, X = Y, and Y = Z, then X = Z.

INDEPENDENCE OF IRRELEVANT ALTERNATIVES

For any two possible inputs which don't involve any change of any voters ranking of X and Y, the group ranking of X and Y is the same.

ARROW AND VOTING SYSTEMS

INDEPENDENCE

Basically every real world voting system violates Independence of Irrelevant Alternatives.

INDEPENDENCE FAILURES

Find two inputs, I1 and I2 such that:

1. Each voter has the same view about X and Y in I1 and I2. That is, if they have X > Y in I1, they do in I2 as well. And same for Y > X, and for X = Y.
2. The social ranking of X and Y is different in the two outputs (O1 and O2).

PLURALITY

Voters	Ranking
3	XYZ
2	ZXY

(a) Input I1

Voters	Ranking
3	YXZ
2	ZXY

(b) Input I2

Table 1: Plurality and Independence Violation

- On the left, the ranking is X > Z > Y.
- On the right, the ranking is Y > Z > X.
- Y and Z flip, but no voter flipped them.

RUNOFF

If we had these preferences and used either runoff system, we'd get the same result.

BORDA

Voters	Ranking
3	XYZ
2	ZXY

(a) Input I1

Voters	Ranking
3	YXZ
2	ZXY

(b) Input I2

Table 2: Borda and Independence Violation

- Use 2 points for first, 1 for second.
- On left, X(8) > Z(4) > Y(3).
- On right, Y(6) > X(5) > Z(4).
- Y and Z flip, but no voter flipped them.

RANGE

Range voting is not a function.

We can see this in just a two voter, two option case.

RANGE

Voter	X	Y
A	10	0
B	5	6

(a) *Input I1*

Voter	X	Y
A	6	5
B	0	10

(b) *Input I2*

Table 3: Range and Functionality

- On both sides, A has X > Y, B has Y > X.
- On the left, X wins 15-6.
- On the right, Y wins 15-6.

RANGE

- This shows that range voting does not merely take the preferences of voters as inputs.
- The way Independence is normally stated, it's also an independence violation.
- After all, no voter flipped their preference between left side and right side.

OLIGARCHY

WEAKENING THE CONSTRAINTS

The next few slides are about two ways to weaken the Arrow constraints to find a system that works.

We're going to start with a model that drops one of the transitivity principles.

OLIGARCHY

Designate some people, more than 1, as **oligarchs**.

The social order ranks X above Y if and only if every oligarch ranks X above Y.

That's the end of the theory.

ARROW CRITERIA

1. Unanimity
2. Non-Dictatorship
3. Unrestricted Domain
4. Social Ordering
5. Independence of Irrelevant Alternatives

It meets every one except **4**.

TRANSITIVITY

Even on 4, it gets 2/3 of the way there.

If you add that if not X > Y, and not Y > X, then X = Y, it is complete.

And > is transitive.

But = is not transitive.

EXAMPLE

There are two oligarchs, A and B, and three options, X, Y, Z.

- A ranks X > Z > Y.
- B ranks Y > X > Z
- Social ordering has X = Y, Y = Z, and X > Z.

OLIGARCHY

Note that you could have a version of this where **everyone** is an oligarch, so the view doesn't have to be inegalitarian.

And we had an argument last time that if = is defined this way, you should not expect it to be transitive.

GIBBARD

The idea for this comes in part from Allan Gibbard, who until his recent retirement was in the philosophy department here at UM.

He and a colleague proved that it was the only system that met these slightly relaxed constraints.

SINGLE-PEAKED

BASIC MODEL

Assume the options vary along one dimension.

And also assume that each voter has a 'bliss-point' on that dimension, and they prefer any change that moves towards the bliss-point.

Then say X > Y if and only if a majority of voters prefer X to Y.

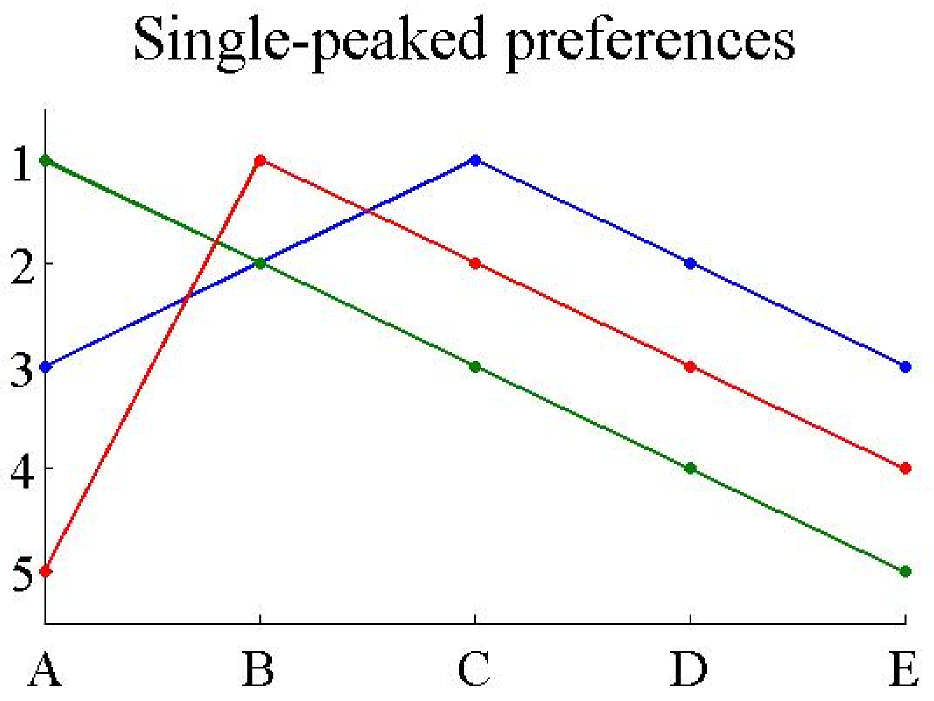
Given these assumptions, that will work - it will produce a social order.

CONSTRAINTS

1. Unanimity
2. Non-Dictatorship
3. Unrestricted Domain
4. Social Ordering
5. Independence of Irrelevant Alternatives

It meets every one except **3**. I'm assuming this doesn't actually issue an output if the voters do not meet this characteristic.

SINGLE-PEAKED



Three examples of single peaked preferences, from Wikipedia

SINGLE-PEAKED

Note that we don't require the close to the bliss point the better.

- It could be that moving away in one direction is *much worse* than moving away in another direction.
- It's just that moving towards the bliss point is always better.

POLITICS

This may be politically important. If…

1. Everyone agrees on the left-to-right arrangement of the candidates; and
2. Everyone has a 'bliss-point' on this spectrum; and
3. Everyone prefers X to Y if X is between Y and the 'bliss-point'

Then majority preferences will lead to a consistent outcome.

Is this realistic? Not in general, but maybe it's close.

OTHER COMBINATIONS

APPLICATIONS

One use of Arrow's Theorem is to put constraints on when we can sensibly talk about the general will, or what the people (collectively) want.

But it has become a very big deal across a bunch of disciplines because the basic model can apply to many things beyond voting.

MULTI-DIMENSIONALITY

Say a good is **multi-dimensional** if it has different attributes, and the goodness of the whole is a function of how good the different parts are.

Put this broadly, most goods are multi-dimensional.

E.g., in a movie we might care about the script, the acting, the cinematography, and any number of other things.

MULTI-DIMENSIONALITY

We often talk about which of two multi-dimensional goods is better.

This could be because we are giving out awards, like the Oscars.

Or it could be because we are trying to choose which one to select, under a budget constraint.

MULTI-DIMENSIONALITY

In these circumstances, we might care about the following question:

- What is the relationship between quality along each dimension, and overall quality.

Arrow's Theorem applies here.

- Consider each dimension a voter, and the overall quality the winning candidate.
- There is no way to merge attributes that satisfies Arrow's five constraints.

BETTER MOVIES

In practice we make these comparisons all the time. We choose to go to one movie rather than another, or we vote for one movie rather than another in end-of-year awards.

How do we do it?

BETTER MOVIES

Three options jump out from what we've discussed so far.

1. We don't just have comparisons along each dimension, but numerical measures of quality.
2. We assume that there are enough constraints on the voters/attributes to use majoritarian voting.
3. We allow for incomparability, and use oligarchical voting.

BETTER MOVIES

Oligarchy doesn't seem plausible.

- Saying A is better than B doesn't mean A is better than B alone *every* dimension we care about.
- One can say that Oppenheimer is better than the latest Marvel movie even if (a) the Marvel movie had better special effects, and (b) special effects are important.

BETTER MOVIES

And single-peakedness is definitely out.

- The whole point of having multiple attributes is that they are different; no reason to think they will all agree on some left-to-right ordering.

RANGE VOTING

But we can do multi-dimensional comparisons if we

1. Can give each dimension a numerical grade; and
2. The value of the whole is in some sense the sum of these grades.

For movies 2 is a little absurd - but let's set that aside. I want to briefly end with some comments about 1.

COMPARISONS

The thing about giving numerical grades to the acting, the script, the cinematography etc is we need some kind of fixed scale between the different dimensions.

We need to be able to say what being 1 point better is along each of these dimensions, and it needs in some sense to be the same amount for each dimension.

COMPARISONS

Perhaps needless to say, this is really hard to do, and arguably doesn't make sense.

This is why comparing goods with multiple attributes like this is very hard.

But it's also something we have to do every day, assuming we want to buy some things and have a budget constraint.

MORAL CHOICES

(I don't expect we'll get to this, but it's interesting enough to put on a note.)

Some moral theories say that there are many different things we should compare about morally.

A simple theory like this says we should both care about making the world a better place, and about whether we are respecting people's rights.

MORAL CHOICES

I think such theories are intuitively very plausible, but they face a big challenge.

If we care about multiple things morally, how do we ever come up with a decision about what to do.

How do we 'trade-off' a gain on one dimension against a loss on another.

MORAL CHOICES

In this context it would be really good if we could just say “Well, here is the ordinal ranking of actions according to values V1, V2, …, Vn, and here’s a way to merge those rankings into an overall moral ranking.”

But what Arrow’s Theorem seems to show is that that won’t work.

We need something like numerical values.

I don’t have any solutions here; this is just a very hard problem.

NEXT WEEK

FOR NEXT TIME

Next week the focus is on the economist-philosopher Amartya Sen.

Sen's work across a range of topics is fantastic, and he's well worth reading.

We're going to start with his very surprising argument that sometimes we should socially prefer X to Y even though *every person* prefers Y to X.