ARROW’S THEOREM

444 Lecture 9

Brian Weatherson

2024-02-11

BACKGROUND

KENNETH ARROW

- Prominent economist.
- Worked on both theoretical questions (especially general equilibrium) and practical questions (health economics, economic growth).
- Interacted a lot with philosophy.

Kenneth Arrow (1921-2017)

COMBINING WANTS

Intuitively our problem here is to decide what the group wants given what the individuals in the group want.

- This is particularly important in democratic decision making.
- But it has other implications as well.

MID-20TH CENTURY THEORY

The notion of a *want* was not particularly fashionable in mid-20th century academic work.

- Lots of writers either implicitly or explicitly endorsed *behaviorism*, and hence thought that we should replace mental state attributions with references to behavior.
- And a want, like wanting world peace, doesn't have an immediate behavioral pair.

CHOICE

Instead of looking at spooky things like wants, theorists then wanted to look at concrete things.

- A standard move was to focus not on wants, but on *choices*.
- What I want is hard to tell, and maybe meaningless.
- But what I choose from a group of options can be observed, and that seems like a scientific notion.

PREFERENCE

With some assumptions, we can get something stronger than choice.

- If I choose A, when my options include B, that is a sign that I prefer A to B.
- Strictly speaking, it only shows that I either prefer A to B, or I'm indifferent between them.
- You also need some other assumptions about my consistency, and the issues here get complicated.

PREFERENCE

But this line of reasoning was why it was common to start with preferences in mid-century work.

- And Arrow's work is definitely mid-century.
- The theorem was first publicly stated in his 1951 book *Social Choice and Individual Values*, though it came out of his earlier PhD work.

INTER-PERSONAL VALUE

One part of the tradition Arrow is working in here is the view that what are called *inter-personal value comparisons* are impossible. Here's what that means for a case with two people (A, B), and two options (X, Y)

1. A likes X a lot more than Y; B likes Y a little more than X.
2. A likes X a little more than Y; B likes Y a lot more than X.

The view is that these two situations are, for political economy purposes, the same. One person prefers one option, the other the other option.

THE TASK

GENERATE A FUNCTION

Inputs
: Individual preference rankings over options.

Output
: A group preference ranking over options.

PREFERENCE RANKING

For any two options either

1. X is (strictly) better than Y; or
2. Y is (strictly) better than X; or
3. X and Y are equally good.

TRANSITIVITY

All three of these things are meant to be transitive.

- If X is better than Y, and Y is better than Z, then X is better than Z.
- If X is as good as Y, and Y is as good as Z, then X is as good as Z.

SOCIAL CHOICE

There is something a bit funny about this as an output constraint.

- Intuitively we want to make a *social choice*.
- That just requires saying which option is best, it doesn't require saying what the social rankings are over all the options.
- We will come back to this.

THE PROJECT

1. Find some constraints on this function.
2. Find a function that meets the constraints.

SAMPLE CONSTRAINTS

Unanimity
: If everyone prefers X to Y, then the group prefers X to Y.

Non-Dictatorship
: For any person, and any pair X, Y, it is possible that they prefer X to Y, but the group does not.

STRONGER CONSTRAINTS

Option Invariance
: The output does not have any built-in difference between the options; it just matters how they are ranked.

Voter Invariance
: The output does not have any built-in difference between the voters; it just matters which preferences are held, not who holds them.

These are both strengthenings of the constraints on the previous slide.

ARROW'S 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

We've touched on the first two.

UNRESTRICTED DOMAIN

Whatever preferences the voters have, the function will produce an output.

- Intuitively, it will never say “Oh, if that’s what you think, good luck, you’re on your own.”
- This is very important in a voting system for example.

SOCIAL ORDERING

The output is **complete**, and **transitive**. That is

1. For any two options, X and Y, the output says either X is better, or Y is better, or they are equally good.
2. In the output, better than is transitive.
3. In the output, equally good is transitive.

SOCIAL ORDERING

This rules out a possibility that you might want to take seriously.

- I'll introduce it via an example.

HOLIDAY RANKING

Imagine that you have a bunch of money for summer vacation, and you can choose one of two very nice options. (These will be expensive, but you can afford it, in the hypothetical.)

1. You can spend time on some islands in the Carribean, lying on the beach, surfing, snorkeling, etc.
2. You can go to some European cities, visit museums, see incredible art, eat at fancy restaurants, etc.

For some of you, one of these will be clearly preferable. But imagine you think they are both pretty good options

HOLIDAY RANKING

Is the following possible for a rational choosder?

1. Someone thinks that each holiday would be good; they don't prefer one over the other.
2. If the price of one of them fell by $50, they would still think each was good, and wouldn't prefer one over the other.

a. Yes

b. No

COMPLETENESS AND TRANSITIVITY

If you think Yes (and I kind of think Yes), then you have to give up one of the conditions on social ordering, either completeness or transitivity. (That is, you have to give up their analogues for individual preference.)

1. For any two options, X and Y, either X is better, or Y is better, or they are equally good.
2. Better than is transitive.
3. Equally good is transitive.

COMPLETENESS AND TRANSITIVITY

Let C be the Carribean vacation, E the European vacation, and C+ the cheaper Carribean vacation after prices fall by $50.

- Since having $50 is better than not having it, C+ is better than C.
- If preferences are complete, and neither C nor C+ is better than or worse than E, then C+ is as good as E, and C is as good as E.
- If as good as is transitive, C+ is as good as C.

SWEETENINGS

What these constraints mean in practice is that if neither option is better than another, then any ‘sweetening’ of one option by making it better (e.g., by a fall in airfares) will make it the better option.

- If neither is better than the other, they are tied, and any improvement breaks a tie.
- If you think that options can be still impossible to choose between after sweetening, one of these constraints is off.

INDEPENDENCE

This is the big constraint, and I'm going to spend a bit of time explaining it.

The short version is that the social ordering of two options X and Y is solely a function of the individual orderings of X and Y.

That is, a change in individual rankings can only change the social ranking of X and Y if it changes at least one person's ranking of X and Y.

INDEPENDENCE

But it’s hard to get an intuitive grip on that, and even Arrow himself got it wrong when trying to state it intuitively.

I find it easiest to see what’s happening with some illustrations.

INDEPENDENCE AND PLURALITY

Here's a simple social choice function.

1. For each voter who likes X more than all the other options, give X 1 point.
2. If a voter has N options tied for best, give each of them 1/N points.
3. Order the options by the number of points they have.

INDEPENDENCE AND PLURALITY

Imagine there are five voters, and three options: X, Y and Z.

- Three voters have Y > X > Z; so from them Y gets 3 points.
- Two voters have Z > X > Y; so from them, Z gets 2 points.

So the social ordering is Y > Z > X, because Y has 3 points, Z has 2 points, X has 0 points.

INDEPENDENCE AND PLURALITY

Now imagine the three voters change their mind about X and Y, so we hvae:

- Three voters have X > Y > Z; so from them Z gets 3 points.
- Two voters have Z > X > Y; so from them, Z gets 2 points.

So the social ordering is Y > Z > X, because X has 3 points, Z has 2 points, Y has 0 points.

INDEPENDENCE AND PLURALITY

Between the last two slides, two things happened:

1. The social ranking of X and Z flipped. Originally Z > X, now X > Z.
2. No voter changed their ranking of X and Z. It's just that three voters were sure Z was worst, flipped their view on which of the other two options was better.

INDEPENDENCE

That's what Arrow's Independence condition rules out.

- It's only the individual ranking of X and Y that matter to the group ranking.

INDEPENDENCE AND CONDORCET

There's a similar point here to Condorcet's view about the importance of pairwise majorities.

It's easy to have the intuition that in an election with two big parties, all that should matter is the voter preference between those two.

And that's what motivates various runoff voting systems.

But this is a *really* strong constraint.

THE 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

PROOFS

I’m really not going to go over the proofs, of which there are several.

The ones I find most intuitive involve showing that if you meet the other four conditions, you can prove there must be a dictator.

ONE PROOF SKETCH

- First, show that if the conditions are met, then if everyone has an extreme view of X (i.e., has it either first or last), the group must have an extreme view (i.e., have it first or last).
- If everyone has it first, the group must have it first, and if everyone has it last, the group must have it last.
- Now imagine starting with everyone having it first, and changing the votes one voter at a time to make it last.
- At some point it flips, and the person who causes the flip must be a dictator.

FOR NEXT TIME

BIG TOPICS

1. Which conditions do real world voting systems end up violating? (It's basically always independence.)
2. What weakenings of the constraints lead to interesting views. (Completeness will be important here.)
3. How can we use Arrow's Theorem to think about other kinds of aggregation. (Hint: Think of GPA as a way of aggregating your various professor's views about your work.)