14.770-Fall 2017
Recitation 1 Notes

Arda Gitmez

September 8, 2017

Today:

- A short review of the first lecture and concepts.
- A close cousin of Arrow's Impossibility result: the Gibbard-Satterthwaite theorem.

What Happened in the First Lecture

In many ways, it's a pretty standard "Introduction to Political Economy" lecture.

- Politics: the art of aggregating preferences.[1]
- Of course, the first question is: is it possible to aggregate preferences?
  - (Spoiler: it isn't.)
- Let's formalize this question a little bit.
  - Assume a group of people, $\mathcal{H}$, have well-defined (complete, reflexive, transitive) preferences over policies.[2]
  - Can we create an "aggregation machine" to which we put these preferences, and it spits out one well-defined (transitive) preference profile over policies (reflecting the preferences for the whole group)?
  - Such an aggregation machine is also often named a "Social Welfare Function".
  - Some features we would expect from the aggregation machine (aka axioms):
    * **Weakly Paretian.** If everyone prefers one policy over the other, the society should too.
    * **Independence from Irrelevant Alternatives (IIA).** The comparisons can be reduced to *pairwise* ones.
  - **Arrow's Impossibility Theorem:** The only such aggregation machine is the *dictatorial rule* – which, come to think of it, is not an "aggregation machine" at all.
    * Is this a commentary on the difficulty of group decision-making, or a commentary on how demanding the features we expect are (in particular, IIA)?
    * A depressing result at first, but also hugely influential: gave rise to 30 years of literature on social choice.

[1] A very loose definition, I know. ¯_(ツ)_/¯

[2] These preferences can come from anywhere (see L. Notes slides 4 and 5 for a treatment), but the crucial thing is that the preference domain is *unrestricted* – i.e. they can be anything.

    * Most importantly: it gave rise to the class you're taking now. If such an aggregation machine could be found, there would be no need for a full-semester class on political economy!

- The other depressing result we covered: **Condorcet Paradox**.
  - In words: "Voting isn't a good way to aggregate preferences, as it may lead to *cycles*."
  - Not really surprising given Arrow: indeed, this is a corollary of Arrow's Impossibility Theorem. (Why?)
- Next question: OK, but can we get around Arrow? When can we have a meaningful prediction on group decision making?
  - Obviously, we should revisit the model. One feature of the model is: the preference domain is unrestricted. This is usually a feature that the impossibility results heavily rely on, because those proofs always go like: "But what happens when people have *these* preferences?"
  - Let's drop the unrestrictedness assumption, and place some restriction/structure on preferences.
  - Two such restrictions: (i) single-peakedness and (ii) single-crossing.
  - Reasonable? Sometimes, but we're making these assumptions mostly because they are what work mathematically.
- Following this restriction, next big result of the lecture: **Median Voter Theorem (MVT)**.
  - When preferences are single-peaked, there are no Condorcet cycles. Consequently, voting works!
  - Works with sincere voting, but also with strategic voting, i.e. we should not worry about people lying. (Subject to some caveats we covered in the lecture.)
- Another variant of the MVT: **Downsian Convergence Theorem**.
  - Same idea. When preferences are single-peaked, the median voter's most preferred policy is implemented.

What You Should Remember About the First Lecture

It is a brief but valuable intro: try to keep in mind the intellectual flow of the lecture. "Group decision making is hard, but if people have sufficiently structured preferences, voting is a pretty good tool for decision making."

The concepts you should remember:

1. Arrow's Impossibility Theorem (and how IIA is crucial and controversial)
2. Condorcet Paradox
3. Median Voter Theorem and Downsian Convergence Theorem

The next few lectures really build on the theoretical ideas we developed here: we will discuss whether people have really single-peaked preferences, whether people vote strategically, whether Downsian convergence really occurs etc.

Some Extra: Gibbard-Satterthwaite Theorem

You may have the following reaction upon seeing Arrow's Impossibility Theorem:

> "OK, but why are we insisting on finding a whole preference profile (i.e. a ranking over all possible policies)? Don't we just need to choose *one* policy?"

In other words, why don't we have a "choice machine" rather than an "aggregation machine"? This seems promising, but unfortunately it wouldn't work either: Gibbard-Satterthwaite (G-S) tells us that the only "choice machine" which satisfies certain desirable properties is again dictatorial.

Two main differences between G-S and Arrow:

1. As hinted above, G-S is about "choice machines" (i.e. Social Choice Functions) rather than "aggregation machines" (i.e. Social Welfare Functions).
2. G-S is about incomplete information, i.e. it takes into account that people may lie about their preferences (much like the second version of MVT we covered in the lecture).

Let's formulate these ideas and the theorem mathematically.

- Let $\mathcal{H}$ be the (finite) set of people, and let $\mathcal{P}$ be the (finite) set of policies.
- As in the lecture, let $\mathfrak{R}$ be the set of all weak orders on $\mathcal{P}$.
  - Note: as in Arrow, preference domain is again unrestricted.
- Each individual $i \in \mathcal{H}$ has preferences $R_i \in \mathfrak{R}$. Consequently, a *preference profile* has the form $\rho = (R_1, \ldots, R_{|\mathcal{H}|}) \in \mathfrak{R}^{|\mathcal{H}|}$.

Definition 1. *A* ***Social Choice Function*** f *is a function*

$$f : \mathfrak{R}^{|\mathcal{H}|} \to \mathcal{P}$$

That is, it takes a preference profile ρ as its input and chooses an policy $f(\rho) \in \mathcal{P}$.

What are some desirable features of a Social Choice Function? We will state two: Pareto efficiency and Strategy-Proofness. Let's start with Pareto efficiency.

Definition 2. *Given a preference profile $\rho \in \mathfrak{R}^{|\mathcal{H}|}$, a policy $p \in \mathcal{P}$ is* ***Pareto optimal*** *if there is no other policy $p' \in \mathcal{P}$ such that*

$$p' \succeq_i p \quad \text{for all } i \in \mathcal{H}$$

and

$$p' \succ_j p \quad \text{for some } j \in \mathcal{H}.$$

This is the very standard notion of Pareto optimality: there exists no other policy that everybody weakly prefers and at least one agent strictly prefers.

Definition 3. *A Social Choice Function f is* ***Pareto efficient*** *if, for any preference profile $\rho \in \mathfrak{R}^{|\mathcal{H}|}$, $f(\rho)$ is Pareto optimal.*

Pretty reasonable, huh? The other desirable feature we have, strategy-proofness, is also quite reasonable.

Definition 4. *A Social Choice Function f is* ***strategy-proof*** *if, for any preference profile $\rho = (R_1, \ldots, R_{|\mathcal{H}|}) \in \mathfrak{R}^{|\mathcal{H}|}$, any individual $i \in \mathcal{H}$ and any preference $R'_i \in \mathfrak{R}$,*

$$f(\rho) \succeq_i f(R_1, \ldots, R_{i-1}, R'_i, R_{i+1}, \ldots, R_{|\mathcal{H}|})$$

That is, individual i cannot strictly benefit from pretending that her preferences are R'_i.

This is also a pretty standard requirement: indeed, whenever there is incomplete information, some kind of "truth-telling" requirement needs to be imposed.[3]

[3] There's a huge story behind this I can't cover within this lecture – you should take 14.281 or 14.125 for the treatment it deserves!

Definition 5. *A Social Choice Function f is* ***dictatorial*** *if there exists an agent $i \in \mathcal{H}$ such that, for any $\rho = (R_1, \ldots, R_{|\mathcal{H}|}) \in \mathfrak{R}^{|\mathcal{H}|}$, $f(\rho)$ is the most preferred policy for R_i.*

Now we are ready to state the Gibbard-Satterthwaite theorem.

Theorem 1. ***(Gibbard-Satterthwaite).*** *If $|\mathcal{P}| \geq 3$, then any Pareto efficient and strategy-proof Social Choice Function is necessarily dictatorial.*

Some commentary on this theorem:

- Why do we need $|\mathcal{P}| \geq 3$? Can you find a Pareto efficient, strategy-proof and nondictatorial Social Choice Function when $|\mathcal{P}| = 2$?
- Once again: is this a commentary on the difficulty of group decision-making, or a commentary on how demanding strategy-proofness is?
- It should be apparent that this result is closely connected to Arrow's Impossibility Theorem. Their proofs are really similar too! Indeed, there is an illuminating paper by Phil Reny (Economics Letters, 2001) titled "Arrow's Theorem and the Gibbard-Satterthwaite Theorem: A Unified Approach" which gives the proofs side-by-side.

If you feel like you need a resource to revisit the issues discussed today, Mas-Collel, Whinston and Green's Chapter 21 and (first half of) Chapter 23 are good places to look at.