MS&E 336/CS 366: Computational Social Choice. Aut 2010-21
Course URL: http://www.stanford.edu/~ashishg/msande336/index.html.
Instructor: Ashish Goel, Stanford University.

Lecture 5, 10/4/2021. Scribed by Anna Zhao.

Gibbard-Satterthwaite Theorem

This lecture gives an overview of the Gibbard-Satterthwaite Theorem, of which the full proof of can be found here[1]. These notes are meant to give increased intuition behind the formal proof, not as a substitute. The theorem is as follows:

> *Suppose there are at least three alternatives and that for each individual any strict ranking of these alternatives is permissible. Then the only unanimous, strategy-proof social choice function is a dictatorship.*

Unanimity means that if everyone ranks the same alternative as the highest, that alternative must be selected as the winner. *Strategy-proofness* means that reporting one's true preferences is the dominant strategy no matter what other voters do.

5.1 Fact 1: Monotonicity

From strategy-proofness, we can derive the following fact of *strong monotonicity*[1]

> **Fact 1:** Suppose that alternative A is selected given some preference profile. Modify the profile by raising some alternative X in individual i's ranking (holding everything else fixed). Then either A or X is now selected.

We can prove by contradiction. Let's say there are two preference profiles for an individual, profile i and profile $\widetilde{i}$ in an alternate universe where we move alternative X up in their ranking.

i	$\widetilde{i}$
B	B
A	A
⋮	⋮
⋮	X
X	⋮
⋮	⋮

[1] This is in contrast to the definition of *weak monotonicity*, which states that if alternative A is selected and we raise the position of A in someone's rankings, then A will still win. Weak monotonicity follows from strong monotonicity, if we make alternative X as A.

Suppose alternative A wins in profile i, but alternative B ($B \neq X$ and $B \neq A$) wins in profile $\widetilde{i}$. Because B is above A in individual i's true ranking, they can misreport X to profile $\widetilde{i}$ and do better. However, this violates strategy-proofness.

Now, let's say individual i prefers A to B in their true ranking instead:

i'	$\widetilde{i'}$
A	A
B	B
⋮	⋮
⋮	X
X	⋮
⋮	⋮

Like before, let's say A wins in profile i', but after moving X up in profile $\widetilde{i'}$, alternative B wins. However, because the individual prefers A to B, they would never report profile $\widetilde{i'}$ where X is promoted but rather report profile i where X is further down.

Therefore, monotonicity follows from strategy-proofness.

5.2 Step 1: Define pivotal voter r

Begin with an arbitrary profile in which all voters rank alternative B last.

B is not the winner because of strategy-proofness.

If the voters, one at a time, kept raising another alternative A while keeping everything else fixed, either A or B will be selected because of monotonicity (Fact 1). Everyone will promote A to the top because they all hate B the most, which contradicts unanimity.

When we make B jump to the top for each voter, starting with individual 1 and proceeding iteratively, there is some voter at which B will start winning. Call this voter r, the pivotal voter, also denoted $r(B)$ to indicate the pivotal voter for alternative B in particular.

1	2	...	r	...	n
B	B	B	K	...	...
A	D	⋮	⋮	⋮	⋮
C	C	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
..	...	S	B	B	B

Profile 1: Alternative B is not selected.

1	2	...	r	...	n
B	B	B	B	C	...
A	D	⋮	⋮	⋮	⋮
C	C	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
..	...	...	S	B	B

Profile 2: Alternative B is selected.

We make two observations.

B is chosen whenever the first r voters rank B first.

Considering profile 2, where B is selected, let's say we make a change for voter 1, so that their preference is $B \succ_1 C \succ_1 A \succ_1 ...$, and that C wins at this new preference profile. However, voter 1 still prefers B to C, so why would they report this? If they misreported their true preferences and instead reported the preference in profile 1 ($B \succ_1 A \succ_1 C \succ_1 ...$), they could do better. So, this violates strategy-proofness, and means that B must win even if the rest of the alternatives excluding top-choice B were changed for the first r voters. In addition, voters $r+1$ to n dislike B the most, so if they could promote B so that it doesn't win, they would. However, that goes against monotonicity, so we do not have to worry about these voters' preferences.

If voters r to n put B as their last choice, then B will not be selected.

Considering profile 1, where B is not selected, if voters 1 through $r-1$ were to change their rankings so that B could win, they would do so, which violates strategy-proofness. If voters r to n submitted different rankings but with B still ranked last, and B were to be selected, they would not honestly report that ranking as they hate B the most.

5.3 Steps 2 and 3

The takeaway for step 2 is:

The pivotal individual $r(B)$ is a dictator in the special case when everyone ranks B last.

Consider any profile of the following form, with B ranked last for everyone and K ranked first for voter r.

1	2	...	r	...	n
?	?	?	K	?	?
?	?	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
B	B	B	B	B	B

Profile 3

Alternative K will be chosen. Proof omitted, see step 2 in the full proof linked for more details.

Step 3a states that:

If $r(B)$ chooses K then either K or B wins.

From step 2, if B is at the bottom for all voters, then alternative K is chosen. Now, raise B to its original position for all voters, one at a time. Because of monotonicity, either K or B is chosen.

Next, step 3b states that:

The pivotal voter $r(X)$ is the same for all alternatives X.

Consider:

1	2	$r(C)$	$r(B)$	...	n
C	C	C	A	...	...
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	C	C	C
B	B	B	B	B	B

Profile 4

Assume that the pivotal voters for C and B, denoted $r(C)$ and $r(B)$, respectively, are not equivalent. Because the first $r(C)$ voters have alternative C first, C is selected. However, because of step 2, the dictator for $r(B)$ has A as the winner. This contradiction means that the pivotal voters $r(C)$ and $r(B)$ must be the same for all alternatives.

5.4 Other lecture questions and discussions

In class, we also talked about Arrow's Impossibility Theorem and its comparison to Gibbard-Satterthwaite. Arrow's Impossibility Theorem calls for *Pareto optimality*, not *unanimity*. For example, in the following preference profile:

1	2	3
A	C	C
B	A	A
C	B	B

Everyone prefers A to B, so the social welfare function should not have B as a winner in order to satisfy Pareto optimality. Unanimity doesn't apply in this profile because the top choice is not the same for all three voters.

We also discussed ways to "break" the theorem: by breaking the social choice function and using random choice instead, by breaking strategy-proofness and replacing it with Nash equilibrium, or by breaking by assuming some sort of preference structure like single-peakness.

The *revelation principle* was brought up during discussion, which is that a mechanism can be used to implement a social choice function in an incentive-compatible way (truth telling is the dominant strategy).

References

[1] Benoıt, Jean-Pierre. "The Gibbard–Satterthwaite theorem: a simple proof." Economics Letters 69.3 (2000): 319-322.