MS&E 336/CS 366: Computational Social Choice. Winter 2019-20
Course URL: `http://www.stanford.edu/~ashishg/msande336.html`.
Instructor: Ashish Goel, Stanford University.

Lecture 4, 1/22/2020. Scribed by Aitan Grossman.

4 The Gibbard-Satterthwaite Theorem

4.1 Introduction

The majority of this lecture is dedicated to providing a high-level proof sketch of the **Gibbard-Satterthwaite theorem**, which will be the basis of much of the discussion throughout the quarter. The full proof[1], which can be found on the course website, consists of three steps, with focus today on the first.

An equivalent statement to 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.*

The Gibbard-Satterthwaite Theorem is an impossibility result, which can seem counter-intuitive, because whereas most proofs concern things that do happen, this one concerns things that cannot happen. The temptation when going through the proof might be to rely on examples of Social Choice Functions (SCFs) to visualize certain steps in the proof, but most SCFs we might think of do not satisfy strategy-proofness, so it's best to follow the proof from a theoretical perspective.

As an exercise, we can show why the plurality scheme where ties are broken alphabetically is not strategy-proof. Figure 1 represents a profile where ballots are cast truthfully. However, V_5 has an incentive to swap alternatives B and C in his ranking, because this would elect C rather than A, which makes this scheme not strategy-proof.

V_1	V_2	V_3	V_4	V_5
A	A	C	C	B $\downarrow$
B	B	B	B	C $\uparrow$
C	C	A	A	A

Figure 1: Example of a profile that is not strategy-proof

4.2 A Rough Proof of the Gibbard-Satterthwaite Theorem

To start, we acknowledge a fact (Fact 1 in the full proof) that **monotonicity** is a consequence of strategy-proofness. Namely, suppose we have a strategy-proof SCF, and suppose that alternative A is selected given some profile. Modify the profile by raising some alternative X in individual i's ranking[1], holding everything else fixed. Then either A or X is now selected.

[1]Much of this proof involves starting with an arbitrary profile and changing, raising, or lowering alternatives in individuals' rankings. Note that all of the SCFs this proof concerns are strategy-proof, so these changes are meant

1	...	**r - 1**	**r**	**r + 1**	...	**N**
B	B	B	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	B	B	B	B

1	...	**r - 1**	**r**	**r + 1**	...	**N**
B	B	B	B	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	B	B	B

Figure 2: Illustration of the pivotal voter r

To prove this fact, suppose we have a strategy-proof SCF and a profile that has alternative A as the winner. Suppose, for the sake of contradiction, if individual i swaps alternative X up by one ranking, then alternative C is elected, where C $\neq$ A and C $\neq$ X. In the case that i ranks C higher than A, i would have the incentive to falsely report his modified ballot from the reference point of his original ballot. In the case that i ranks C lower than A, i would have the incentive to falsely report his original ballot from the reference point of his modified ballot. In either case, the SCF is not strategy-proof.

If V_3 considered raising alternative X in her ranking (and lowering C) thus causing a preferred third alternative (A or B) to win, V_3 would falsely do so, contradicting strategy-proofness. If V_4 considered raising alternative X in her ranking (and lowering B) thus causing a non-preferred third alternative (C) to win, V_4 would not have an incentive to do so, and thus, such a swap would also contradict strategy-proofness.

The rest of the proof can be broken down into three steps, with the focus of today's lecture being on the first step, which introduces the idea of a pivotal voter.

We begin by taking an arbitrary SCF for at least three alternatives that is strategy-proof and unanimous and showing that this function must be dictatorial. Consider a profile where everyone ranks B last. Alternative B cannot win in this scenario, because if individuals iteratively ranked some alternative A first, they would eventually elect A by unanimity, which would be preferable, thus contradicting strategy-proofness. Now consider a profile where everyone ranks B first. Alternative B must win in this scenario by unanimity.

If we come up with an arbitrary ordering of individuals and consider an arbitrary profile where everyone ranks B last, we can imagine voters incrementally raising B to the top of their ballots until B starts to win. This is illustrated in Figure 2, where the profile on the left has B losing and the profile on the right has B winning. The first voter whose switch causes B to win is defined as the **pivotal voter** for B, also known as the **pivotal individual**, the **pivotal** $\boldsymbol{r}$, or simply $\boldsymbol{r}$**(B)**.

The outline for the proof is as follows:

1. Show that as long as the first r(B) voters rank B first, B will win, and as long as last $N - r(\text{B}) + 1$ are ranked last, B will not win.

2. Show that r(B) is a dictator in the special case when everyone ranks B last.

3a. Show that if r(B) chooses K then either K or B wins.

to reflect changes in the individuals' actual preferences, in conjunction with changes in these individuals' ballots.

3b. Show that the pivotal voter $r(\text{X})$ is the same for all X. (This finishes the proof.)

4.3 A Note On Dictatorships

It's worthwhile to discuss what it really means for a voting rule to be a dictatorship. Formally, a **dictatorship** is an SCF where a single voter is selected prior to considering the voting profile, and the outcome is that single voter's top ranked candidate.

Clearly, by this definition, dictatorships are not **ex post fair**, meaning that they do not seem fair to voters after the dictator and/or outcome is determined. A **randomized dictatorship**, when a random voter is selected to choose the winner, is an example of a unanimous, strategy-proof voting mechanism that is **ex ante fair**, meaning that it seems fair to voters *before* the dictator and/or outcome is determined.

4.4 A Quick Overview of Kalai-Smorodinsky Bargaining

The **Kalai-Smorodinsky bargaining solution** is a solution to the bargaining problem, and it will be the topic of a future homework problem. Whereas the Nash solution satisfies independence of irrelevant alternatives, the Kalai-Smorodinsky solution satisfies **monontonicity**, meaning that if, for every utility level that player 1 may demand, the maximum feasible utility level that player 2 can simultaneously reach is increased, then the utility level assigned to player 2 according to the solution should also be increased.

The Kalai-Smorodinsy solution is for both players to get the same fraction of their maximum utility, assuming that the outside alternative yields zero utility to both players.

4.5 More Rules and Definitions

A **tournament graph** (see Figure 3) is a graph with nodes representing the alternatives and directed edges XY if X is preferred to Y in more than 50% of ballots in the profile. A **weighted tournament graph** has edge weights for each edge XY corresponding to the proportion of ballots in the profile where X is preferred to Y.

A **tournament rule** (also known as a **type 1 rule**) is a rule that only depends on the unweighted tournament graph. The Copeland rule is the only rule we have seen so far that is a tournament rule, where the node with the maximum out degree wins. A **weighted tournament rule** (also known as a **type 2 rule**), accordingly, is a rule that only depends on the weighted tournament graph. Borda is one example of a weighted tournament rule we have seen so far.

A **type 3 rule** is simply a voting rule that is not type 1 or type 2.

Ranked pairs is a powerful Social Welfare Function that takes advantage of a tournament graph. Edges are added greedily in order of decreasing weight and any edge that causes a cycle is discarded. This results in a directed, acyclic graph which is completely ordered (i.e., between any two nodes, at least one edge is present). Repeatedly removing the source node givess an ordering which is called the **ranked pairs ordering**.

Single Transferable Vote (also known as **Instant Runoff**) is a SCF where candidates with the least number of first-place votes are iteratively eliminated until one candidate remains.

1	1	1	1	1
A	A	C	C	B
B	C	A	B	A
C	B	D	A	C
D	D	B	D	D

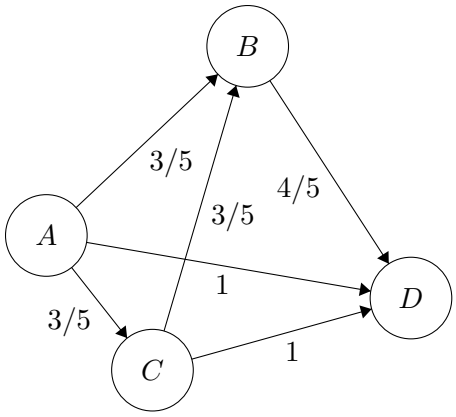

Figure 3: Example of a tournament graph

References

[1] J.-P. Benoît *The Gibbard–Satterthwaite theorem: a simple proof.* Economic Letters 69, 2000.