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 3, 9/27/2021. Scribed by Drew Gao.

4 Strategy Proofness

In last weeks lecture we discussed the notion of strategy proofness and gave intuitive explanations for why certain voting mechanisms are not strategy proof. One very strong formal definition of strategy-proofness comes from the notion of a dominant strategy. In voting we can define a dominant strategy as follows:

Definition 4.1 *A voter X has a* ***dominant strategy*** *if there exists a preference ranking S such that for every preference ranking T, T can produce an outcome no better for X than S regardless of the profiles of the other players. Furthermore there must exist at least one situation in which playing S is strictly better for X than playing T*

If truth telling is a dominant strategy then we can say a voting mechanism is strategy proof. We may equivalently describe strategy proof voting mechanisms as incentive compatible.

Example 4.1 *An example of a incentive compatible voting mechanisms is that where a random voter is chosen to be the dictator. In this situation voters cannot benefit from reporting anything other than their true preference.*

Example 4.2 *A non-voting example of a strategy proof mechanism can be seen in the context of an auction. In this situation we have one item which will be sold to the "winner" of the auction. However, we do not want the winner to be able to strategically manipulate the price they end up paying. One mechanism for determining the winner is known as the 2nd price auction (a specific case of the more general VCG auction). This auction mechanism has participants placing bids according to how much they would be willing to pay. The winner is the participant with the highest bid, but they pay the price of the second highest bid.*

In general, auction mechanisms in which the winner chooses their own price cannot be strategy proof as you can always benefit from bidding as low as possible while still beating other competitors.

4.1 Strategy Proofness of SCFs

It is natural to ask if more general statements can be made about when a voting mechanism is strategy proof. In the case of social choice functions, the Gibbard Satterthwaite Theorem gives us the surprising answer to this question:

Theorem 4.3 *Gibbard Satterthwaite: Any social choice function which is resolute and can output at least three different alternatives as a winner and is strategy proof must be dictatorial.*

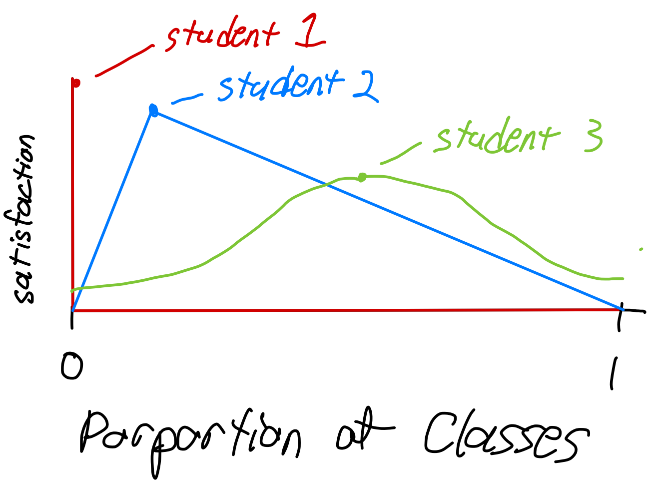
It is important to note that this theorem relies on only a few conditions (resoluteness, non-dictatorality, and the ability to output at least three different winners) to show that a SCF cannot be strategy proof. Additional conditions which are often put on voting mechanisms such as neutrality and pareto-optimality are not required. Thus we see that very few voting rules are truely strategy proof. Much of this course/the field of social choice theory is an attempt to reconcile this problem.

We can use the Gibbard Satterthwaite Theorem to conclude that plurality voting (with three or more candidates) is not strategy proof. For a concrete example consider the following situation:

Example 4.4 *Suppose we have* 5 *voters* $x_1, \dots, x_5$ *and* 3 *candidates* A, B, C *with voter* x_1 *having true preference* $C \succ B \succ A$. *Now suppose* x_2 *and* x_3 *rank candidate* A *as their top choice, while* x_3 *and* x_4 *rank candidate* B *as their top choice. In this situation, since candidate* C *has no chance of winning, voter* x_1 *would benefit from deviating from their actual preference and ranking candidate* B *as their top choice to ensure their preferred candidate between* A *and* B *wins.*

4.2 Voting on a Line

One way that we can design a strategy proof voting rule is by "voting on a line." Consider a motivating example in which three students are asked what fraction of classes should have a queue based speaking system. Each student reports their level of satisfaction for each potential decision on a continuous interval as pictured below.



We can see that each student has a distinct proportion of classes which use a queue based speaking system which maximizes their satisfaction, with satisfaction decreasing (non-increasing) for proportions further from this point. We call such a preference a **single peak preference**. Now, in order to create a voting rule for this situation we first need an elicitation rule and a aggregation rule. The elicitation rule tells us what data we need to gather about a voter's preferences (ex. most preferred alternative, a ranking of all alternatives, a utility function describing satisfaction with each alternative etc.) while the aggregation rule tells us how the preferences of all voters can be combined to obtain a winner(s). In a situation where we have voters with single peak preferences we can design a strategy-proof voting mechanism. This mechanism asks each voter to report their most preferred point as an elicitation rule, the aggregation rule then picks the median as the winner.

Note that the aggregation rule uses the median rather than the mean, since using the mean would result in strategic behavior (voters not already in the middle are incentives to go towards the extremes to achieve an outcome closer to their true preference). This is given by Black's Theorem.

It turns out that Black's Theorem also generalizes to two (or more) dimensions when using the l^1 norm for distance and when voter preferences are single peak in every dimension.