Computational Social Choice: Spring 2017

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

Plan for Today

It is not always in the best interest of voters to truthfully reveal their preferences when voting. This is called *strategic manipulation*.

We are going to see two theorems that show that this can't be avoided:

- *Gibbard-Satterthwaite Theorem* (1973/1975)
- *Duggan-Schwartz Theorem* (2000)

The latter generalises the former by considering irresolute voting rules, where voters have to strategise w.r.t. *sets* of winners.

Example

Recall that under the *plurality rule* the candidate ranked first most often wins the election.

Assume the preferences of the people in, say, Florida are as follows:

49%:	Bush $\succ$ Gore $\succ$ Nader
20%:	Gore $\succ$ Nader $\succ$ Bush
20%:	Gore $\succ$ Bush $\succ$ Nader
11%:	Nader $\succ$ Gore $\succ$ Bush

So even if nobody is cheating, Bush will win this election.

- It would have been in the interest of the Nader supporters to *manipulate*, i.e., to misrepresent their preferences.

Is there a better voting rule that avoids this problem?

Truthfulness, Manipulation, Strategyproofness

For now, we only deal with *resolute* voting rules $F : \mathcal{L}(X)^n \to X$.

Unlike for all earlier results discussed, we now have to distinguish:

- the *ballot* a voter reports
- her actual *preference* order

Both are elements of $\mathcal{L}(X)$. If they coincide, then the voter is *truthful*.

F is *strategyproof* (or *immune to manipulation*) if for no individual $i \in N$ there exist a profile $\boldsymbol{R}$ (including the "truthful preference" R_i of i) and a linear order R'_i (representing the "untruthful" ballot of i) such that $F(R'_i, \boldsymbol{R}_{-i})$ is ranked above $F(\boldsymbol{R})$ according to R_i.

In other words: under a strategyproof voting rule no voter will ever have an incentive to misrepresent her preferences.

Notation: $(R'_i, \boldsymbol{R}_{-i})$ is the profile obtained by replacing R_i in $\boldsymbol{R}$ by R'_i.

Importance of Strategyproofness

Why do we want voting rules to be strategyproof?

- Thou shalt not bear false witness against thy neighbour.
- Voters should not have to waste resources pondering over what other voters will do and trying to figure out how best to respond.
- If everyone strategises (and makes mistakes when guessing how others will vote), then the final ballot profile will be very far from the electorate's true preferences and thus the election winner may not be representative of their wishes at all.

The Full-Information Assumption

Here, as in most work on the topic, we make the assumption that the manipulator has *full information* about the ballots of the other voters.

Is this always realistic? No. But:

- In *small committees* (e.g., members of a department voting on who to hire) the full-information assumption is fairly realistic.
- Even in large political elections *poll information* may be accurate enough to allow groups of voters (though not individuals) to perform similar acts of manipulation as discussed here.
- When looking for *protection against manipulation*, we should assume the *worst case*, where the manipulator has full information.

The Gibbard-Satterthwaite Theorem

Recall: a resolute SCF F is *surjective* if for every alternative $x \in X$ there exists a profile $\boldsymbol{R}$ such that $F(\boldsymbol{R}) = x$.

Gibbard (1973) and Satterthwaite (1975) independently proved:

Theorem 1 (Gibbard-Satterthwaite) *Any resolute SCF for $\geqslant 3$ alternatives that is surjective and strategyproof is a dictatorship.*

Remarks:

- a *surprising* result + not applicable in case of *two* alternatives
- The opposite direction is clear: *dictatorial* ⇒ *strategyproof*
- *Random* procedures don't count (but might be "strategyproof").

A. Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica*, 41(4):587–601, 1973.

M.A. Satterthwaite. Strategy-proofness and Arrow's Conditions. *Journal of Economic Theory*, 10:187–217, 1975.

Proof

We shall prove the Gibbard-Satterthwaite Theorem to be a corollary of the Muller-Satterthwaite Theorem (even if, historically, G-S came first).

Recall the *Muller-Satterthwaite Theorem:*

- Any *resolute* SCF for $\geqslant 3$ alternatives that is *surjective* and *strongly monotonic* must be a *dictatorship*.

We shall prove a lemma showing that strategyproofness implies strong monotonicity (and we'll be done). ✓ (Details are in my review paper.)

For other short proofs of G-S, see also Barberà (1983) and Benoît (2000).

S. Barberà. Strategy-Proofness and Pivotal Voters: A Direct Proof the Gibbard-Satterthwaite Theorem. *International Economic Review*, 24(2):413–417, 1983.

J.-P. Benoît. The Gibbard-Satterthwaite Theorem: A Simple Proof. *Economic Letters*, 69(3):319–322, 2000.

U. Endriss. Logic and Social Choice Theory. In A. Gupta and J. van Benthem (eds.), *Logic and Philosophy Today*, College Publications, 2011.

Strategyproofness implies Strong Monotonicity

Lemma 2 *Any resolute SCF that is strategyproof (SP) must also be strongly monotonic (SM).*

- SP: no incentive to vote untruthfully
- SM: $F(\boldsymbol{R}) = x \;\Rightarrow\; F(\boldsymbol{R'}) = x$ if $\forall y : N^{\boldsymbol{R}}_{x \succ y} \subseteq N^{\boldsymbol{R'}}_{x \succ y}$

Proof: We'll prove the contrapositive. So assume F is *not* SM. So there exist $x, x' \in X$ with $x \neq x'$ and profiles $\boldsymbol{R}, \boldsymbol{R'}$ such that:

- $N^{\boldsymbol{R}}_{x \succ y} \subseteq N^{\boldsymbol{R'}}_{x \succ y}$ for all alternatives y, including x' $(\star)$
- $F(\boldsymbol{R}) = x$ and $F(\boldsymbol{R'}) = x'$

Moving from $\boldsymbol{R}$ to $\boldsymbol{R'}$, there must be a *first* voter affecting the winner. So w.l.o.g., assume $\boldsymbol{R}$ and $\boldsymbol{R'}$ differ only w.r.t. voter i. Two cases:

- $i \in N^{\boldsymbol{R'}}_{x \succ x'}$: if i's true preferences are as in $\boldsymbol{R'}$, she can benefit from voting instead as in $\boldsymbol{R}$ $\Rightarrow$ ↯ [SP]
- $i \notin N^{\boldsymbol{R'}}_{x \succ x'} \overset{(\star)}{\Rightarrow} i \notin N^{\boldsymbol{R}}_{x \succ x'} \Rightarrow i \in N^{\boldsymbol{R}}_{x' \succ x}$: if i's true preferences are as in $\boldsymbol{R}$, she can benefit from voting as in $\boldsymbol{R'}$ $\Rightarrow$ ↯ [SP]

Remark

Note that we can strengthen the Gibbard-Satterthwaite Theorem (and the Muller-Satterthwaite Theorem) by replacing

- F being surjective and being defined for $\geqslant 3$ alternatives

by the slightly weaker requirement of

- F having a range of $\geqslant 3$ outcomes:

$$\#\{x \in X \mid F(\boldsymbol{R}) = x \text{ for some } \boldsymbol{R} \in \mathcal{L}(X)^n\} \geqslant 3$$

The Bigger Picture

We have by now seen three impossibility theorems for *resolute* SCF's, all of which apply in case there are at least *three alternatives:*

Gibbard-Satterthwaite Theorem
[surjective + strategyproof ⇒ dictatorial]

⇑

Muller-Satterthwaite Theorem
[surjective + strongly monotonic ⇒ dictatorial]

⇑

Arrow's Theorem
[Paretian + independent ⇒ dictatorial]

We proved Arrow's Theorem by analysing when a coalition can force a pairwise ranking. The other two results followed by comparing axioms.

Shortcomings of Resolute Voting Rules

The Gibbard-Satterthwaite Theorem only applies to *resolute* rules.
But the restriction to resolute rules is problematic:

- No "natural" voting rule is resolute (w/o tie-breaking rule).
- We can get very basic impossibilities for resolute rules:

  We've seen already that *no resolute* voting rule for *two voters* and *two alternatives* can be both *anonymous* and *neutral*.

We therefore should really be analysing *irresolute* voting rules ...

Manipulability w.r.t. Psychological Assumptions

To analyse manipulability when we might get a set of winners, we need to make assumptions on how voters rank *sets of alternatives*, e.g.:

- A voter is an *optimist* if she prefers X over Y whenever she prefers her favourite $x \in X$ over her favourite $y \in Y$.
- A voter is a *pessimist* if she prefers X over Y whenever she prefers her least preferred $x \in X$ over her least preferred $y \in Y$.

Now we can speak about manipulability by certain types of voters:

- F is called *immune to manipulation by optimistic voters* if no optimistic voter can ever benefit from voting untruthfully.
- F is called *immune to manipulation by pessimistic voters* if no pessimistic voter can ever benefit from voting untruthfully.

Axiom: Nonimposition

Let F be an *irresolute* voting rule/SCF $F : \mathcal{L}(X)^n \to 2^X \setminus \{\emptyset\}$.

- F is *nonimposed* if for every alternative x there exists a profile $\boldsymbol{R}$ under which x is the unique winner: $F(\boldsymbol{R}) = \{x\}$.

For comparison, *surjectivity* means that for every element in the range of F there is an input producing that element. Thus:

$$\text{resolute} \Rightarrow (\text{nonimposed} = \text{surjective})$$

Dictatorships for Irresolute Rules

Let F be an *irresolute* voting rule/SCF $F : \mathcal{L}(X)^n \to 2^X \setminus \{\emptyset\}$.

There are two natural notions of dictatorship for such rules:

- Voter $i \in N$ is called a (strong) *dictator* if $F(\boldsymbol{R}) = \{\text{top}(R_i)\}$ for every profile $\boldsymbol{R} \in \mathcal{L}(X)^n$.
- Voter $i \in N$ is called a *weak dictator* if $\text{top}(R_i) \in F(\boldsymbol{R})$ for every profile $\boldsymbol{R} \in \mathcal{L}(X)^n$. (Such a voter is also called a *nominator*.)

F is called *weakly dictatorial* if it has a weak dictator. Otherwise F is called *strongly nondictatorial*.

The Duggan-Schwartz Theorem

There are several extensions of the Gibbard-Satterthwaite Theorem for irresolute voting rules. The Duggan-Schwartz Theorem is usually regarded as the strongest of these results.

Our statement of the theorem follows Taylor (2002):

Theorem 3 (Duggan and Schwartz, 2000) *Any voting rule for $\geqslant 3$ alternatives that is nonimposed and immune to manipulation by both optimistic and pessimistic voters is weakly dictatorial.*

Proof: Omitted.

Note that the Gibbard-Satterthwaite Theorem is a direct corollary.

J. Duggan and T. Schwartz. Strategic Manipulation w/o Resoluteness or Shared Beliefs: Gibbard-Satterthwaite Generalized. *Soc. Choice Welf.*, 17(1):85–93, 2000.

A.D. Taylor. The Manipulability of Voting Systems. *The American Mathematical Monthly*, 109(4)321–337, 2002.

Alternative Models

Today's results apply to the standard model of voting theory, where preferences and ballots are linear orders over the alternatives.

Some extra mileage is possible when we change the model:

- Under the system of *majority judgment*, no voter can strategically *manipulate the grade* assigned to an alternative. But strategic manipulation can still affect which alternative wins.
- For *approval voting* (with ballots $\in 2^X$ and preferences $\in \mathcal{L}(X)$), under certain conditions, we can ensure that no voter has an incentive to vote *insincerely* (weak variant of strategyproofness).

But care needs to be taken with how to interpret such results.

M. Balinski and R. Laraki. A Theory of Measuring, Electing, and Ranking. *PNAS*, 104(21):8720–8725, 2007.

U. Endriss. Sincerity and Manipulation under Approval Voting. *Theory and Decision*. 74(3):335–355, 2013.

Summary

We have seen that *strategic manipulation* is a major problem in voting:

- *Gibbard-Satterthwaite:* only dictatorships are strategyproof amongst the resolute and surjective voting rules
- *Duggan-Schwartz:* dropping the resoluteness requirement does not provide a clear way out of this impossibility

The study of strategic manipulation is very much at the intersection of social choice theory with *game theory* and *mechanism design*.

Other forms of strategic behaviour that may occur in the context of elections include *bribery* and *gerrymandering*.

What next? Ways of coping with the problem of strategic manipulation.