# Computational Social Choice: Autumn 2012

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





## Plan for Today

We have already seen that voters will sometimes have an incentive not to truthfully reveal their preferences when they vote.

Today we shall see two important theorems that show that this kind of *strategic manipulation* is impossible to avoid:

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

The latter generalises the former by considering irresolute voting rules, where voters have to strategise wrt. *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, Strategy-Proofness

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

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

- the *ballot* a voter reports
- from her actual *preference* relation.

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

$F$ is *strategy-proof* (or *immune to manipulation*) if for no individual $i \in \mathcal{N}$ 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 $F(\boldsymbol{R}_{-i}, R_i')$ is ranked above $F(\boldsymbol{R})$ according to $R_i$.

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

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

## Importance of Strategy-Proofness

Why do we want voting rules to be strategy-proof?

- 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 other 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

When studying strategy-proofness, we make the classical assumption that the manipulator has *full information* about the ballots of the other voters. Is this always realistic? No. But:

- We want possible protection against manipulation to work even in the *worst case*, where the manipulator has obtained full information.
- 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.

Aside: Recently there has been some initial research in COMSOC addressing manipulation under partial information (see references below).

V. Conitzer, T. Walsh, and L. Xia. Dominating Manipulations in Voting with Partial Information. Proc. AAAI-2011.

A. Reijngoud and U. Endriss. Voter Response to Iterated Poll Information. Proc. AAMAS-2012.





## The Gibbard-Satterthwaite Theorem

Recall: a resolute SCF/voting rule $F$ is *surjective* if for any alternative $x \in \mathcal{X}$ 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 strategy-proof is a dictatorship.*

Remarks:

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

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 strategy-proofness implies strong monotonicity (and we'll be done). ✓ (Details are in the 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.

## Strategy-Proofness implies Strong Monotonicity

**Lemma 1** *Any resolute SCF that is strategy-proof (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 $x, x' \in \mathcal{X}$ 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 wrt. 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 $\boldsymbol{R}$ $\Rightarrow$ $\lightning$ [SP]
- $i \notin N^{\boldsymbol{R'}}_{x \succ x'} \Rightarrow^{(\star)} i \notin N^{\boldsymbol{R}}_{x \succ x'} \Rightarrow i \in N^{\boldsymbol{R}}_{x' \succ x}$: if $i$'s true preferences are as in $\boldsymbol{R}$, she can benefit from voting as in $\boldsymbol{R'}$ $\Rightarrow$ $\lightning$ [SP]





## Remark

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

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

by the slightly weaker requirement of

- $F$ being a voting rule with a range of $\geqslant 3$ outcomes:

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





## Shortcomings of Resolute Voting Rules

The Gibbard-Satterthwaite Theorem only applies to *resolute* voting 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:

  Fact: *No resolute* voting rule for $2$ *voters* and $2$ *alternatives* can be both *anonymous* and *neutral*.

  Proof: Consider the case where the voters' rankings differ . . . ✓

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





## Manipulability wrt. 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 an *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.

## Aside: Ranking Sets of Objects

Optimism/pessimism is a way of *extending preferences* declared over *objects* to *sets of objects*. This is an interesting research area in its own right.

The seminal result in the field is the *Kannai-Peleg Theorem* (1984):

*For $|\mathcal{X}| \geqslant 6$, it is impossible to extend a linear order on $\mathcal{X}$ to a weak order on $2^{\mathcal{X}} \setminus \{\emptyset\}$ in a manner that satisfies:*

- *Dominance: if you (dis)prefer $x$ to every object in set $A$, then you should (dis)prefer $A \cup \{x\}$ to $A$*
- *Independence: if you prefer set $A$ to set $B$, then you should also (weakly) prefer $A \cup \{x\}$ to $B \cup \{x\}$ (for any $x$ not in $A \cap B$)*

For more on this topic, see the references cited below.

Y. Kannai and B. Peleg. A Note on the Extension of an Order on a Set to the Power Set. *Journal of Economic Theory*, 32(1):172–175, 1984.

S. Barberà, W. Bossert, and P.K. Pattanaik. Ranking sets of objects. In *Handbook of Utility Theory*, volume 2. Kluwer Academic Publishers, 2004.

C. Geist and U. Endriss. Automated Search for Impossibility Theorems in Social Choice Theory: Ranking Sets of Objects. *JAIR*, 40:143–174, 2011.





## Other Axioms

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

- Recall: a dictator can impose a unique winner. A variation:
  - A voter is a *weak dictator* (or a *nominator*) for $F$ if her top-ranked alternative is always *one of* the winners under $F$.
  - $F$ is called *weakly dictatorial* if it has a weak dictator; otherwise $F$ is called *strongly nondictatorial*.
- $F$ is *nonimposed* if for any alternative $x$ there exists a profile $\boldsymbol{R}$ under which $x$ is the unique winner: $F(\boldsymbol{R}) = \{x\}$.





## 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 2 (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.





## Summary

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

- *Gibbard-Satterthwaite:* only dictatorships are strategy-proof 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*.

## What next?

Next we will discuss how to counter the problem of strategic manipulation. The two main approaches are:

- *Domain restrictions:* If we only need our voting rule to work for certain preference profiles, then more positive results are possible.
- *Complexity as a barrier against manipulation:* The idea is that, in certain cases, manipulation maybe be possible in principle, but computationally intractable in practice.