Computational Social Choice: Autumn 2011

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 prove an important theorem that shows that this kind of *strategic manipulation* is impossible to avoid:

- The Gibbard-Satterthwaite Theorem (1973/1975)

We then discuss several ways of *circumventing* this problem, notably:

- Domain restrictions regarding voter preferences
- Computational hardness as a barrier against manipulation

Example

Recall that under the *plurality rule* the candidate receiving the highest number of votes wins.

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. But:

- 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'.

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 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 J. van Benthem and A. Gupta (eds.), *Logic and Philosophy Today*, College Publications. In press (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 $\forall y : N^{\boldsymbol{R}}_{x \succ y} \subseteq N^{\boldsymbol{R'}}_{x \succ y}$

<u>Proof:</u> 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. <u>Two cases:</u>

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

Irresolute Voting Rules

Recall that the Gibbard-Satterthwaite Theorem applies to *resolute* voting rules only. This is a limitation; most rules are irresolute.

For further reading: The best known result regarding the impossibility of designing an acceptable irresolute voting rule that is strategy-proof is the *Duggan-Schwartz Theorem* (2000).

Remark: How to *extend* a voter's preferences over individual winners to a preference relation over sets of winners (e.g., in view of her beliefs regarding the tie-breaking rule) is an interesting question its own right (to be discussed next week).

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

Circumventing Manipulation

The Gibbard-Satterthwaite Theorem tells us that there aren't any reasonable voting rules that are strategy-proof. *That's very bad!*

We will consider three possible avenues to circumvent this problem:

- Changing the formal framework a little (one slide only)
- Restricting the domain (the classical approach)
- Making strategic manipulation computationally hard

Changing the Framework

The Gibbard-Satterthwaite Theorem applies when both preferences and ballots are linear orders. The problem persists for several variations. But:

- In a framework with *money*, if preferences and ballots are modelled as (quasi-linear) *utility functions* $u : \mathcal{X} \to \mathbb{R}$, we can design strategy-proof mechanisms. Example: *Vickrey Auction*
- In the context of *approval voting* (ballots $\in 2^{\mathcal{X}}$, preferences $\in \mathcal{L}(\mathcal{X})$), under certain conditions we can ensure that no voter has an incentive to vote *insincerely* (weak variant of strategy-proofness).
- More generally, for any *preference language* and *ballot language*, we can define a notion of *sincerity* and study incentives to be sincere.

W. Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *Journal of Finance* 16(1):8–37, 1961.

U. Endriss. Vote Manipulation in the Presence of Multiple Sincere Ballots. Proc. TARK-2007.

U. Endriss, M.-S. Pini, F. Rossi, and K.B. Venable. Preference Aggregation over Restricted Ballot Languages: Sincerity and Strategy-Proofness. Proc. IJCAI-2009.

Domain Restrictions

- Note that we have made an implicit *universal domain* assumption: *any* linear order may come up as a preference or ballot.

- If we *restrict* the domain (possible ballot profiles + possible preferences), more procedures will satisfy more axioms . . .

Single-Peaked Preferences

An electorate $\mathcal{N}$ has *single-peaked* preferences if there exists a "left-to-right" ordering $\gg$ on the alternatives such that any voter prefers x to y if x is between y and her top alternative wrt. $\gg$.

The same definition can be applied to profiles of ballots.

<u>Remarks:</u>

- Quite natural: classical spectrum of political parties; decisions involving agreeing on a number (e.g., legal drinking age); ...
- But certainly not universally applicable.

Black's Median Voter Theorem

For simplicity, assume the number of voters is *odd*.

For a given left-to-right ordering $\gg$, the *median-voter rule* asks each voter for their top alternative and elects the alternative proposed by the voter corresponding to the median wrt. $\gg$.

Theorem 2 (Black's Theorem, 1948) *If an odd number of voters submit single-peaked ballots, then there exists a Condorcet winner and it will get elected by the median-voter rule.*

D. Black. On the Rationale of Group Decision-Making. *The Journal of Political Economy*, 56(1):23–34, 1948.

Proof Sketch

The candidate elected by the median-voter rule is a Condorcet winner:

> Proof: Let x be the winner and compare x to some y to, say, the left of x. As x is the median, for more than half of the voters x is between y and their favourite, so they prefer x. ✓

Note that this also implies that a Condorcet winner *exists*.

As the Condorcet winner is (always) unique, it follows that, also, every Condorcet winner is a median-voter rule election winner. ✓

Strategy-Proofness

The following result is a corollary of Black's Theorem:

Theorem 3 (Strategy-proofness) *If an odd number of voters have preferences that are single-peaked wrt. a fixed left-to-right ordering $\gg$, then the median-voter rule (wrt. $\gg$) is strategy-proof.*

Direct proof: W.l.o.g., suppose our manipulator's top alternative is to the right of the median (the winner). She has two options:

- Nominate some other alternative to the right of the current winner (or the winner itself). Then the median/winner does not change.
- Nominate an alternative to the left of the current winner. Then the new winner will be to the left of the old winner, which—by the single-peakedness assumption—is worse for our manipulator.

Thus, misrepresenting preferences has either no effect or results in a worse outcome. ✓

More on Domain Restrictions

This is a big topic in SCT. We have only scratched the surface here.

- It suffices to enforce single-peakedness for *triples* of alternatives.
- Moulin (1980) gives a *characterisation* of the class of voting rules that are strategy-proof for single-peaked domains: median-voter rule + addition of "phantom peaks"
- Sen's *triplewise value restriction* is more powerful and also guarantees Condorcet winners and strategy-proofness: for any triple of alternatives (x, y, z), there exist a $x^\star \in \{x, y, z\}$ and a value in $v^\star \in \{\text{"best"}, \text{"middle"}, \text{"worst"}\}$ such that $x^\star$ never has value $v^\star$ wrt. (x, y, z) for any voter.

H. Moulin. On Strategy-Proofness and Single Peakedness. *Public Choice*, 35(4):437–455, 1980.

A.K. Sen. A Possibility Theorem on Majority Decisions. *Econometrica*, 34(2):491–499, 1966.

Complexity as a Barrier against Manipulation

The Gibbard-Satterthwaite Theorem shows that (in the standard model) strategic manipulation can never be rule out.

Idea: So it's always *possible* to manipulate; but maybe it's also *difficult*? Tools from *complexity theory* can make this idea precise.

- If manipulation is computationally intractable for F, then F might be considered *resistant* (albeit still not *immune*) to manipulation.
- Even if standard procedures turn out to be easy to manipulate, it might still be possible to *design new ones* that are resistant.
- This approach is most interesting for voting rules for which the problem of computing election winners is tractable. At least, we want to see a *complexity gap* between manipulation (undesired behaviour) and winner determination (desired functionality).

Recap: Complexity Theory

- Given a class of problems parametrised by their "size", how hard it is to solve a problem of size n?
- Distinguish: *time*/space *worst-case*/average-case complexity
- Problems solvable in *polynomial* time (P) are considered tractable, those requiring *exponential* time (EXPTIME) not.
- Take a problem that requires searching through a tree. If you are lucky and go down the right branch at every node, you may need only polynomial time, otherwise exponential time.

  A *nondeterministic* algorithm is a (hypothetical) algorithm with an "oracle" that tells us which branch to explore next.
- NP is the class of decision problems that can be solved by such *nondeterministic* algorithms in *polynomial* time.

Recap: Complexity Theory (continued)

- Equivalent definition: NP is the class of problems for which a candidate solution can be *verified* in polynomial time.
- A decision problem is *NP-hard* iff it is at least as hard as any of the problems in NP.
- A decision problem is *NP-complete* iff it is NP-hard and in NP.
- We do not know whether $\mathrm{P} = \mathrm{NP}$, but strongly suspect $\mathrm{P} \neq \mathrm{NP}$.
- NP-complete problems are generally considered intractable. Unless $\mathrm{P} = \mathrm{NP}$, there can be no general algorithm solving NP-complete problems efficiently.
- As a rule of thumb, NP-completeness means that a naïve approach won't work, but a sophisticated algorithm may well give good results in practice.

Classical Results

The seminal paper by Bartholdi, Tovey and Trick (1989) starts by showing that manipulation is in fact *easy* for a range of commonly used voting rules, and then presents one system (a variant of the Copeland rule) for which manipulation is NP-complete. Next:

- We first present a couple of these easiness results, namely for *plurality* and for the *Borda rule*.
- We then present a result from a follow-up paper by Bartholdi and Orlin (1991): the manipulation of *STV* is *NP-complete*.

J.J. Bartholdi III, C.A. Tovey, and M.A. Trick. The Computational Difficulty of Manipulating an Election. *Soc. Choice and Welfare*, 6(3):227–241, 1989.

J.J. Bartholdi III and J.B. Orlin. Single Transferable Vote Resists Strategic Voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

Manipulability as a Decision Problem

We can cast the problem of manipulability, for a particular voting rule F, as a decision problem:

> $\textsc{Manipulability}(F)$
> **Instance:** Set of ballots for all but one voter; alternative x.
> **Question:** Is there a ballot for the final voter such that x wins?

In practice, a manipulator would have to solve $\textsc{Manipulability}(F)$ for all alternatives, in order of her preference.

If the $\textsc{Manipulability}(F)$ is computationally intractable, then manipulability may be considered less of a worry for procedure F.

Remark: We assume that the manipulator knows all the other ballots. This unrealistic assumption is intentional: if manipulation is intractable even under such favourable conditions, then all the better.

Manipulating the Plurality Rule

Recall plurality: the alternative(s) ranked first most often win(s)

The plurality rule is easy to manipulate (trivial):

- Simply vote for x, the alternative to be made winner by means of manipulation. If manipulation is possible at all, this will work. Otherwise not.

That is, we have $\textsc{Manipulability}(\mathit{plurality}) \in \mathrm{P}$.

<u>General:</u> $\textsc{Manipulability}(F) \in \mathrm{P}$ for any rule F with polynomial winner determination problem and polynomial number of ballots.

Manipulating the Borda Rule

Recall Borda: submit a ranking (super-polynomially many choices!) and give $m-1$ points to 1st ranked, $m-2$ points to 2nd ranked, etc.

The Borda rule is also easy to manipulate. Use a *greedy algorithm:*

- Place x (the alternative to be made winner through manipulation) at the top of your ballot.
- Then inductively proceed as follows: Check if any of the remaining alternatives can be put next on the ballot without preventing x from winning. If yes, do so. (If no, manipulation is impossible.)

After convincing ourselves that this algorithm is indeed correct, we also get $\textsc{Manipulability}(\mathit{Borda}) \in \mathrm{P}$.

J.J. Bartholdi III, C.A. Tovey, and M.A. Trick. The Computational Difficulty of Manipulating an Election. *Soc. Choice and Welfare*, 6(3):227–241, 1989.

Intractability of Manipulating STV

Single Transferable Vote (STV): eliminate plurality losers until an alternative is ranked first by $> 50\%$ of the voters

Theorem 4 (Bartholdi and Orlin, 1991) $\textsc{Manipulability}(STV)$ *is NP-complete.*

Proof: Omitted.

J.J. Bartholdi III and J.B. Orlin. Single Transferable Vote Resists Strategic Voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

Coalitional Manipulation

It will rarely be the case that a *single* voter can make a difference. So we should look into *manipulation by a coalition* of voters.

Variants of the problem:

- Ballots may be *weighted* or *unweighted*.

  Examples: countries in the EU; shareholders of a company

- Manipulation may be *constructive* (making alternative x a *unique* or *tied* winner) or *destructive* (ensuring x does not win).

Decision Problems

On the following slides, we will consider two decision problems, for a given voting rule F:

Constructive Manipulation(F)
Instance: Set of weighted ballots; set of weighted manipulators; $x \in \mathcal{X}$.
Question: Are there ballots for the manipulators such that x wins?

Destructive Manipulation(F)
Instance: Set of weighted ballots; set of weighted manipulators; $x \in \mathcal{X}$.
Question: Are there ballots for the manipulators such that x loses?

Constructive Manipulation under Borda

In the context of coalitional manipulation with weighted voters, we can get hardness results for elections with small numbers of alternatives:

Theorem 5 (Conitzer et al., 2007) *Under the Borda rule, the constructive coalitional manipulation problem with weighted voters is NP-complete for $\geqslant 3$ alternatives.*

Proof: We have to prove NP-membership and NP-hardness:

- NP-membership: easy (if you guess ballots for the manipulators, we can check that it works in polynomial time)
- NP-hardness: for three alternatives by reduction from PARTITION (next slide); hardness for more alternatives follows

V. Conitzer, T. Sandholm, and J. Lang. When are Elections with Few Candidates Hard to Manipulate? *Journal of the ACM*, 54(3), Article 14, 2007.

Proof of NP-hardness

We will use a reduction from the NP-complete PARTITION problem:

> PARTITION
> **Instance:** $(w_1, \ldots, w_n) \in \mathbb{N}^n$
> **Question:** Is there a set $I \subseteq \{1, \ldots, n\}$ s.t. $\sum_{i \in I} w_i = \frac{1}{2} \sum_{i=1}^{n} w_i$?

Let $K := \sum_{i=1}^{n} w_i$. Given an instance of PARTITION, we construct an election with $n + 2$ weighted voters and three alternatives:

- two voters with weight $\frac{1}{2}K - \frac{1}{4}$, voting $(x \succ y \succ z)$ and $(y \succ x \succ z)$
- a coalition of n voters with weights $w_1, \ldots, w_n$ who want z to win

Clearly, each manipulator should vote either $(z \succ x \succ y)$ or $(z \succ y \succ x)$. Suppose there does exist a partition. Then they can vote like this:

- manipulators corresponding to elements in I vote $(z \succ x \succ y)$
- manipulators corresponding to elements outside I vote $(z \succ y \succ x)$

Scores: $2K$ for z; $\frac{1}{2}K + (\frac{1}{2}K - \frac{1}{4}) \cdot (2+1) = 2K - \frac{3}{4}$ for both x and y

If there is no partition, then either x or y will get at least 1 point more.

Hence, manipulation is feasible *iff* there exists a partition. ✓

Destructive Manipulation under Borda

Theorem 6 (Conitzer et al., 2007) *Under the Borda rule, the destructive coalitional manip. problem with weighted voters is in P.*

Proof sketch: Let x be the alternative the manipulators want to lose. The following algorithm will find a manipulation, if one exists:

> For each alternative $y \neq x$, try letting all manipulators rank y first, x last, and the other alternatives in any fixed order.
>
> If x loses in one of these $m-1$ elections, then manipulation is possible; otherwise it is not.

Correctness of the algorithm follows from the fact that (a) the best we can do about x is not to give x any points and, (b) if any other alternative y has a chance of beating x, she will do so if we give y a maximal number of points. ✓

V. Conitzer, T. Sandholm, and J. Lang. When are Elections with Few Candidates Hard to Manipulate? *Journal of the ACM*, 54(3), Article 14, 2007.

Worst-Case vs. Average-Case Complexity

NP-hardness is only a *worst-case* notion. Do NP-hardness barriers provide sufficient protection against manipulation?

What about the *average complexity* of strategic manipulation?

Some recent work suggests that it might be impossible to find a voting procedure that is *usually* hard to manipulation, for a suitable definition of "usual". See Faliszewski and Procaccia (2010) for a discussion.

P. Faliszewski and A.D. Procaccia. AI's War on Manipulation: Are We Winning? *AI Magazine*, 31(4):53–64, 2010.

Controlling Elections

Strategic manipulation is not the only undesirable form of behaviour in voting we may want to contain by means of complexity barriers ...

People have studied the computational complexity of a range of different types of *control* in elections:

- Adding or removing *candidates*.
- Adding or removing *voters*.
- Redefining *districts* (if your party is likely to win district A with an 80% majority and lose district B by a small margin, you might win both districts if you carefully redraw the district borders ...).

See Faliszewski et al. (2009) for an introduction to this area.

P. Faliszewski, E. Hemaspaandra, L.A. Hemaspaandra, and J. Rothe. *A Richer Understanding of the Complexity of Election Systems*. In *Fundamental Problems in Computing*, Springer-Verlag, 2009.

Bribery in Elections

Bribery is the problem of finding $\leqslant K$ voters such that a suitable change of their ballots will make a given candidate x win.

- Connection to *manipulation:* in the (coalitional) manipulation problem the names of the voters changing ballot are part of the input, while for the bribery problem we need to choose them.
- Several *variants* of the bribery problem have been studied: when each voter has a possibly different "price"; when bribes depend on the extent of the change in the bribed voter's ballot; etc.

People have studied the complexity of several variants of the bribery problem for various voting rules (e.g., Faliszewski et al., 2009).

P. Faliszewski, E. Hemaspaandra, and L.A. Hemaspaandra. How Hard is Bribery in Elections? *Journal of Artificial Intelligence Research*, 35:485–532, 2009.

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

But we have also seen that there are several approaches that may help us to *circumvent* this problem:

- *Domain restrictions:* if we can find a natural and large class of preference profiles (+ ballot restrictions) that make strategic manipulation impossible, then that will sometimes suffice.

- *Complexity barriers:* maybe strategic manipulation will turn out to be sufficiently hard computationally to provide protection.

A related question, which we have not addressed, deals with the *frequency of manipulability*, using either empirical methods or devising formal models regarding the distribution of voter preferences.

What next?

We have briefly mentioned today that it is not clear how a voter would manipulate in the context of an irresolute voting rule, because we have not said what it means to prefer one *set* of alternatives over another.

Next week we will address this question in its own right:

- Given someone's preferences over $\mathcal{X}$, what can we say about her preferences over $2^{\mathcal{X}} \setminus \{\emptyset\}$?