Algorithmic Game Theory

Uri Feige　　Robi Krauthgamer　　Moni Naor

Lecture 9: Social Choice

Lecturer: Moni Naor

Social choice or **Preference Aggregation**

- Collectively choosing among outcomes
  - Elections,
  - Choice of Restaurant
  - Rating of movies
  - Who is assigned what job
  - Goods allocation
  - Should we build a bridge?

- Participants have **preferences** over outcomes

- **Social choice function** aggregates those preferences and **picks and outcome**

Voting

If there are **two** options and an odd number of voters

- Each having a clear preference between the options

Natural choice: **majority voting**

- Sincere/Truthful
- Order of queries has no significance
  - trivial

When there are more than two options:

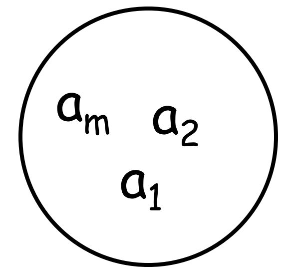
If we start pairing the alternatives:

- Order may matter

$a_{10}, a_1, \ldots, a_8$

Assumption: n voters give their complete ranking on set A of alternatives

- L – the set of **linear orders** on A (permutation).
- Each voter i provides $<_i$ in L
  - Input to the aggregator/voting rule is $(<_1, <_2, \ldots, <_n)$



A

Goal

A function $f: L^n \mapsto A$ is called a **social choice function**

- Aggregates voters preferences and selects a **winner**

A function $W: L^n \mapsto L$, is called a **social welfare function**

- Aggregates voters preference into a **common order**

Example voting rules

Scoring rules: defined by a vector $(a_1, a_2, \ldots, a_m)$

Being ranked ith in a vote gives the candidate a_i points

- **Plurality**: defined by $(1, 0, 0, \ldots, 0)$
  - Winner is candidate that is **ranked first** most often
- **Veto:** is defined by $(1, 1, \ldots, 1, 0)$
  - Winner is candidate that is **ranked last** the least often
- **Borda:** defined by $(m-1, m-2, \ldots, 0)$

Jean-Charles de Borda 1770

Plurality with (2-candidate) runoff: top two candidates in terms of plurality score proceed to runoff.

Single Transferable Vote (STV, aka. Instant Runoff): candidate with lowest plurality score drops out; for voters who voted for that candidate: the vote is transferred to the next (live) candidate

Repeat until only one candidate remains

Marquis de Condorcet

Marie Jean Antoine Nicolas de Caritat, marquis de Condorcet

1743-1794

- **There is something wrong with Borda! [1785]**

Condorcet criterion

- A candidate is the **Condorcet winner** if it wins all of its pairwise elections
- Does not always exist…

Condorcet paradox: there can be **cycles**

- Three voters and candidates:

  a > b > c, b > c > a, c > a > b

- a defeats b, b defeats c, c defeats a

Many rules do not satisfy the criterion

- For instance: **plurality**:
  - b > a > c > d
  - c > a > b > d
  - d > a > b > c
- a is the Condorcet winner, but not the plurality winner

• Candidates a and b:

• Comparing how often a is ranked above b, to how often b is ranked above a

Also Borda:

a > b > c > d > e

a > b > c > d > e

c > b > d > e > a

Even more voting rules…

- **Kemeny**:
  - Consider all pairwise comparisons.
  - Graph representation: edge from winner to loser
  - Create an overall ranking of the candidates that has as few disagreements as possible with the pairwise comparisons.
    - Delete as few edges as possible so as to make the directed comparison graph acyclic

•Honor societies
•General Secretary of the UN

- **Approval** [not a ranking-based rule]: every voter labels each candidate as **approved** or **disapproved.** Candidate with the most approvals wins

How do we choose one rule from all of these rules?

- How do we know that there does not exist another, “perfect” rule?
- We will list some criteria that we would like our voting rule to satisfy

Arrow's Impossibility Theorem

Skip to the 20th Centrury

Kenneth Arrow, an economist. In his PhD thesis, 1950, he:

- Listed desirable properties of voting scheme
- Showed that no rule can satisfy all of them.

Properties

- Unanimity
- Independence of irrelevant alternatives
- Not Dictatorial

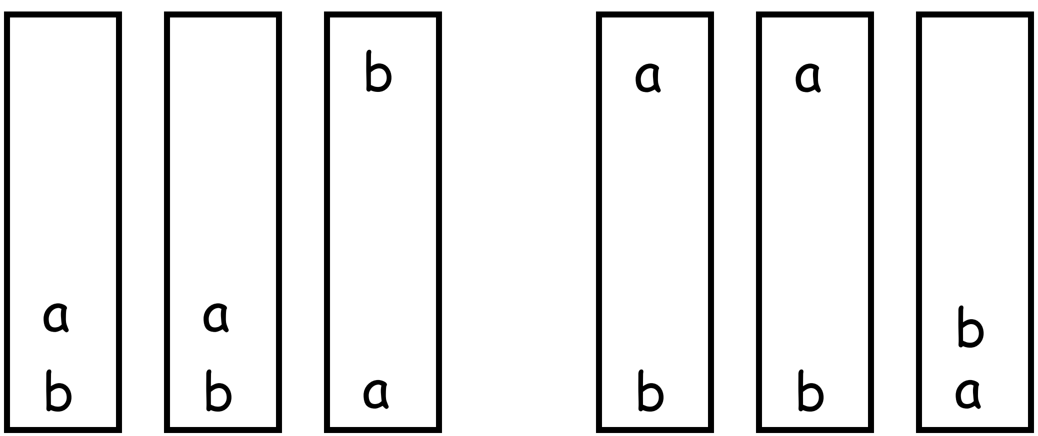
Kenneth Arrow

1921-

Independence of irrelevant alternatives

- Independence of irrelevant alternatives criterion: if
  - the rule ranks a above b for the current votes,
  - we then change the votes but do not change which is ahead between a and b in each vote

  then a should still be ranked ahead of b.
- None of our rules satisfy this property
  - Should they?



Arrow’s Impossibility Theorem

Every **Social Welfare Function** W over a set A of at least 3 candidates:

- If it satisfies
  - Unanimity (if all voters agree on $\prec$ on the result is $\prec$)

    $$W(\prec, \prec, \ldots, \prec) = \prec$$

    for all $\prec$ in L
  - Independence of irrelevant alternatives

  Then it is dictatorial : there exists a voter i where

  $$W(\prec_1, \prec_2, \ldots, \prec_n) = \prec_i$$

  for all $\prec_1, \prec_2, \ldots, \prec_n$ in L

Is there hope for the truth?

- At the very least would like our voting system to encourage voters to tell there true preferences

Strategic Manipulations

- A **social choice function** f can be manipulated by voter i if for some $<_1, <_2, \ldots, <_n$ and $<'_i$ and we have $a = f(<_1, \ldots <_i, \ldots, <_n)$ and $a' = f(<_1, \ldots, <'_i, \ldots, <_n)$ but $a <_i a'$

voter i prefers a' over a and can get it by changing his vote

f is called **incentive compatible** if it cannot be manipulated

Gibbard-Satterthwaite Impossibility Theorem

- Suppose there are at least 3 alternatives
- There exists no **social choice function** f that is simultaneously:
  - Onto
    - for every candidate, there are some votes that make the candidate win
  - Nondictatorial
  - Incentive compatible

Implication of Gibbard-Satterthwaite Impossibility Theorem

- All mechanism design problems can be modeled as a a social choice problem.
- This theorem seems to quash any hope for designing incentive compatible social choice functions.
- The whole field of Mechanism Design is trying to escape from this impossibility results.
- Introducing “money” is one way to achieve this.

Proof of Arrow's Impossibility Theorem

Claim(Pairwise Unanimity): Every **Social Welfare Function** W over a set A of at least 3 candidates

- If it satisfies
  - Unanimity (if all voters agree on $<$ on the result is $<$)

$$W(<, <, \ldots, <) = <$$

  for all $<$ in L

  - Independence of irrelevant alternatives

  Then it is Pareto efficient

  If $W(<_1, <_2, \ldots, <_n) = <$ and for all i $a <_i b$ then $a < b$

Proof of Arrow's Theorem

Claim (Neutrality): let

- $<_1, <_2, \ldots, <_n$ and $<'_1, <'_2, \ldots, <'_n$ be two profiles
- $< = W(<_1, <_2, \ldots, <_n)$ and $<' = W(<'_1, <'_2, \ldots, <'_n)$
- and where for all i

$$a <_i b \Leftrightarrow c <'_i d$$

Then $a < b \Leftrightarrow c <' d$

Proof: suppose $a < b$ and $c \neq b$

Create a single preference π_i from $<_i$ and $<'_i$: where c is just below a and d just above b.

Let $<_\pi = W(\pi_1, \pi_2, \ldots, \pi_n)$

Preserve the order!

We must have: (i) $a <_\pi b$ (ii) $c <_\pi a$ and (iii) $b <_\pi d$

And therefore $c <_\pi d$ and $c <' d$

Proof of Arrow's Theorem: Find the Dictator

Claim: For any a,b in A consider sets of profiles

Hybrid argument

Voters	0	1	2	…	n
1	ab	ba	ba	…	ba
2	ab	ab	ba	…	ba
	ab	ab	ab	…	ba
…	…	…	…		
n	ab	ab	ab		ba

$a < b$ **Profiles** $b < a$

Change must happen at some profile i^*

- Where voter i^* changed his opinion

Claim: this i^* is the dictator!

Proof of Arrow's Theorem: i* is the dictator

Claim: for any $<_1, <_2, \dots, <_n$ and $< = W(<_1, <_2, \dots, <_n)$ and c, d in A. If $c <_{i^*} d$ then $c < d$.

Proof: take $e \neq c, d$ and

- for $i < i^*$ move e to the bottom of $<_i$
- for $i > i^*$ move e to the top of $<_i$
- for i^* put e between c and d

For resulting preferences:

- Preferences of e and c like a and b in profile i^*-1.
- Preferences of e and d like a and b in profile i^*.

$c < e$

$e < d$

Therefore $c < d$

Gibbard-Satterthwaite Impossibility Theorem

- Suppose there are at least 3 alternatives
- There exists no **social choice function** f that is simultaneously:
  - Onto
    - for every candidate, there are some votes that make the candidate win
  - Nondictatorial
  - Incentive compatible

Proof of the Gibbard-Satterthwaite Theorem

Construct a Social Welfare function W_f based on f.

$W_f(<_1,\dots,<_n) = <$ where $a < b$ iff

$f(<_1^{\{a,b\}},\dots,<_n^{\{a,b\}}) = b$

Keep everything in order but move a and b to top

Lemma: if f is an **incentive compatible** social **choice** function which is onto A, then W_f is a **social welfare function**

- If f is non dictatorial, then W_f also satisfies Unanimity and Independence of irrelevant alternatives

Proof of the Gibbard-Satterthwaite Theorem

Claim: for all $<_1, \ldots, <_n$ and any subset S of A we have
$f(<_1^S, \ldots, <_n^S)$ in S

Keep everything in order but move elements of S to top

Take a in S. There is some $<'_1, <'_2, \ldots, <'_n$ where

$$f(<'_1, <'_2, \ldots, <'_n) = a.$$

Sequentially change $<'_i$ to $<^S_i$

- At no point does f output b not in S.
- Due to the incentive compatibility

Proof of Well Form Lemma

- Antisymmetry: implied by claim for $S=\{a,b\}$
- Transitivity: Suppose we obtained contradicting cycle $a < b < c < a$

  take $S=\{a,b,c\}$ and suppose $a = f(<_1^S,\dots, <_n^S)$

  Sequentially change $<^S_i$ to $<_i^{\{a,b\}}$

  Non manipulability implies that

  $f(<_1^{\{a,b\}},\dots, <_n^{\{a,b\}}) = a$ and $b < a$.
- Unanimity: if for all i, $b <_i a$ then

$(<_1^{\{a,b\}})^{\{a\}} = <_1^{\{a,b\}}$ and $f(<_1^{\{a,b\}},\dots, <_n^{\{a,b\}}) = a$

Will repeatedly use the claim to show properties

Proof of Well Form Lemma

- Independence of irrelevant alternatives: if there are two profiles $<_1, <_2, \ldots, <_n$ and $<'_1, <'_2, \ldots, <'_n$ where for all i $b <_i a$ iff $b <'_i a$, then

  $$f(<_1^{\{a,b\}}, \ldots, <_n^{\{a,b\}}) = f(<'^{\{a,b\}}_1, \ldots, <'^{\{a,b\}}_n)$$

  by sequentially flipping from $<_i^{\{a,b\}}$ to $<'^{\{a,b\}}_i$
- Non dictator: preserved