Game Theory

8. Social Choice Theory

Albert-Ludwigs-Universität Freiburg

Bernhard Nebel and Robert Mattmüller

June 14th, 2016

1 Social Choice Theory

- Introduction
- Social Choice Functions
- Condorcet Methods

Social Choice Theory

Motivation: Aggregation of individual preferences

Examples:

- political elections
- council decisions
- Eurovision Song Contest

Question: If voters' preferences are private, then how to implement aggregation rules such that voters vote truthfully (no "strategic voting")?

Social Choice Theory

Definition (Social Welfare and Social Choice Function)

Let A be a set of alternatives (candidates) and L be the set of all linear orders on A. For n voters, a function

$$F : L^n \to L$$

is called a social welfare function. A function

$$f : L^n \to A$$

is called a social choice function.

Notation: Linear orders $\prec \in L$ express preference relations.
$a \prec_i b$: voter i prefers candidate b over candidate a.
$a \prec b$: candidate b socially preferred over candidate a.

Social Choice Functions
Examples

- Plurality voting (aka first-past-the-post or winner-takes-all):
  - only top preferences taken into account
  - candidate with most top preferences wins

  Drawback: Wasted votes, compromising, winner only preferred by minority
- Plurality voting with runoff:
  - First round: two candidates with most top votes proceed to second round (unless absolute majority)
  - Second round: runoff

  Drawback: still, tactical voting and strategic nomination possible.

Social Choice Functions
Examples

- Instant runoff voting:
  - each voter submits his preference order
  - iteratively candidates with fewest top preferences are eliminated until one candidate has absolute majority

  Drawback: Tactical voting still possible.
- Borda count:
  - each voter submits his preference order over the m candidates
  - if a candidate is in position j of a voter's list, he gets $m - j$ points from that voter
  - points from all voters are added
  - candidate with most points wins

  Drawback: Tactical voting still possible ("Voting opponent down").

Social Choice Functions
Examples

- Condorcet winner:
  - each voter submits his preference order
  - perform pairwise comparisons between candidates
  - if one candidate wins all his pairwise comparisons, he is the Condorcet winner

  Drawback: Condorcet winner does not always exist.

Social Choice Functions
Examples

23 voters, candidates a, b, c, d, e.

# voters	8	6	4	3	1	1
1st	e	a	b	c	d	d
2nd	d	b	c	b	c	c
3rd	b	c	d	d	a	b
4th	c	e	a	a	b	e
5th	a	d	e	e	e	a

- Plurality voting: candidate e wins (8 votes)
- Plurality voting with runoff:
  - first round: candidates e (8 votes) and a (6 votes) proceed
  - second round: candidate a $(6 + 4 + 3 + 1 = 14$ votes) beats candidate e $(8 + 1 = 9$ votes)

Social Choice Functions

Examples

23 voters, candidates a, b, c, d, e.

# voters	8	6	4	3	1	1
1st	e	a	b	c	d	d
2nd	d	b	c	b	c	c
3rd	b	c	d	d	a	b
4th	c	e	a	a	b	e
5th	a	d	e	e	e	a

- Instant runoff voting:
  First elimination: d
  Second elimination: b
  Third elimination: a
  Now c has absolute majority and wins.

Social Choice Functions

Examples

Social Choice Theory
Introduction
Social Choice Functions
Condorcet Methods
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

23 voters, candidates a, b, c, d, e.

# voters	8	6	4	3	1	1	
1st	e	a	b	c	d	d	4 points
2nd	d	b	c	b	c	c	3 points
3rd	b	c	d	d	a	b	2 points
4th	c	e	a	a	b	e	1 point
5th	a	d	e	e	e	a	0 points

- Borda count:
  - Cand. a: $8\cdot 0+6\cdot 4+4\cdot 1+3\cdot 1+1\cdot 2+1\cdot 0=33$ pts
  - Cand. b: $8\cdot 2+6\cdot 3+4\cdot 4+3\cdot 3+1\cdot 1+1\cdot 2=62$ pts
  - Cand. c: $8\cdot 1+6\cdot 2+4\cdot 3+3\cdot 4+1\cdot 3+1\cdot 3=50$ pts
  - Cand. d: $8\cdot 3+6\cdot 0+4\cdot 2+3\cdot 2+1\cdot 4+1\cdot 4=46$ pts
  - Cand. e: $8\cdot 4+6\cdot 1+4\cdot 0+3\cdot 0+1\cdot 0+1\cdot 1=39$ pts

$\rightsquigarrow$ Candidate b wins.

Social Choice Functions

Examples

Social Choice Theory
Introduction
Social Choice Functions
Condorcet Methods
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

23 voters, candidates a, b, c, d, e.

# voters	8	6	4	3	1	1
1st	e	a	b	c	d	d
2nd	d	b	c	b	c	c
3rd	b	c	d	d	a	b
4th	c	e	a	a	b	e
5th	a	d	e	e	e	a

- Condorcet winner: Ex.: a $\prec_i$ b 16 times, b $\prec_i$ a 7 times

	a	b	c	d	e	
a	–	0	0	0	1	
b	1	–	1	1	1	$\longleftarrow$ candidate b wins.
c	1	0	–	1	1	
d	1	0	0	–	0	
e	0	0	0	1	–	

Social Choice Functions

Examples

Social Choice Theory
Introduction
Social Choice Functions
Condorcet Methods
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

23 voters, candidates a, b, c, d, e.

# voters	8	6	4	3	1	1
1st	e	a	b	c	d	d
2nd	d	b	c	b	c	c
3rd	b	c	d	d	a	b
4th	c	e	a	a	b	e
5th	a	d	e	e	e	a

- Plurality voting: candidate e wins.
- Plurality voting with runoff: candidate a wins.
- Instant runoff voting: candidate c wins.
- Borda count / Condorcet winner: candidate b wins.

- Different winners for different voting systems.
- Which voting system to prefer? Can even strategically choose voting system!

Condorcet Paradox

Why Condorcet Winner not Always Exists

Example: Preferences of voters 1, 2 and 3 on candidates a, b and c.

$$a \prec_1 b \prec_1 c$$
$$b \prec_2 c \prec_2 a$$
$$c \prec_3 a \prec_3 b$$

Then we have cyclical preferences.

	a	b	c
a	–	0	1
b	1	–	0
c	0	1	–

$a \prec b$, $b \prec c$, $c \prec a$: violates transitivity of linear order consistent with these preferences.

Condorcet Methods

Definition

A Condorcet method return a Condorcet winner, if one exists.

One particular Condorcet method: the Schulze method.

Relatively new: Proposed in 1997

Already many users: Debian, Ubuntu, Pirate Party, ...

Schulze Method

Notation: $d(X, Y)$ = number of pairwise comparisons won by X against Y

Definition

For candidates X and Y, there exists a path $C_1, \ldots, C_n$ between X and Y of strength z if

- $C_1 = X$,
- $C_n = Y$,
- $d(C_i, C_{i+1}) > d(C_{i+1}, C_i)$ for all $i = 1, \ldots, n-1$, and
- $d(C_i, C_{i+1}) \geq z$ for all $i = 1, \ldots, n-1$ and there exists $j = 1, \ldots, n-1$ s.t. $d(C_j, C_{j+1}) = z$

Example: path of strength 3.

$$a \xrightarrow{8} b \xrightarrow{5} c \xrightarrow{3} d$$

Schulze Method

Definition

Let $p(X, Y)$ be the maximal value z such that there exists a path of strength z from X to Y, and $p(X, Y) = 0$ if no such path exists.

Then, the Schulze winner is the Condorcet winner, if it exists. Otherwise, a potential winner is a candidate a such that $p(a, X) \geq p(X, a)$ for all $X \neq a$.

Tie-Breaking is used between potential winners.

Schulze Method
Example

# voters	3	2	2	2
1st	a	d	d	c
2nd	b	a	b	b
3rd	c	b	c	d
4th	d	c	a	a

Is there a Condorcet winner?

	a	b	c	d
a	–	1	1	0
b	0	–	1	1
c	0	0	–	1
d	1	0	0	–

⇝ No!

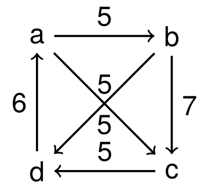
Schulze Method
Example

# voters	3	2	2	2
1st	a	d	d	c
2nd	b	a	b	b
3rd	c	b	c	d
4th	d	c	a	a

Weights $d(X, Y)$:

	a	b	c	d
a	–	5	5	3
b	4	–	7	5
c	4	2	–	5
d	6	4	4	–

As a graph:



Path strengths $p(X, Y)$:

	a	b	c	d
a	–	5	5	5
b	5	–	7	5
c	5	5	–	5
d	6	5	5	–

Potential winners: b and d.

Social Choice Theory
Introduction
Social Choice Functions
Condorcet Methods
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

Schulze Method
Why Use the Schulze Method?

According to Wikipedia (`http://en.wikipedia.org/wiki/Schulze_method`), the method satisfies a large number of desirable criteria:

Unrestricted domain, non-imposition, non-dictatorship, Pareto criterion, monotonicity criterion, majority criterion, majority loser criterion, Condorcet criterion, Condorcet loser criterion, Schwartz criterion, Smith criterion, independence of Smith-dominated alternatives, mutual majority criterion, independence of clones, reversal symmetry, mono-append, mono-add-plump, resolvability criterion, polynomial runtime, prudence, MinMax sets, Woodall's plurality criterion if winning votes are used for d[X,Y], symmetric-completion if margins are used for d[X,Y].

Social Choice Theory
Introduction
Social Choice Functions
Condorcet Methods
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

2 Arrow's Impossibility Theorem

- Motivation
- Properties of Social Welfare Functions
- Main Theorem

Arrow's Impossibility Theorem
Motivation

Motivation: It appears as if all considered voting systems encourage strategic voting.

Question: Can this be avoided or is it a fundamental problem?

Answer (simplified): It is a fundamental problem!

Properties of Social Welfare Functions

Desirable properties of social welfare functions:

Definition (Unanimity)

A social welfare function satisfies

- total unanimity if for all $\prec \in L$, $F(\prec, \dots, \prec) = \prec$.
- partial unanimity if for all $\prec_1, \prec_2, \dots, \prec_n \in L$, $a, b \in A$,

$$a \prec_i b \text{ for each } i = 1, \dots, n \implies a \prec b$$

  where $\prec := F(\prec_1, \dots, \prec_n)$.

Remark

Partial unanimity implies total unanimity, but not vice versa.

Social Choice Theory
Arrow's Impossibility Theorem
Motivation
Properties of Social Welfare Functions
Main Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

Properties of Social Welfare Functions

Desirable properties of social welfare functions:

Definition (Non-Dictatorship)

A voter i is called a dictator for F, if $F(\prec_1, \dots, \prec_i, \dots, \prec_n) = \prec_i$ for all orders $\prec_1, \dots, \prec_n \in L$.
F is called non-dictatorial if there is no dictator for F.

Definition (Independence of Irrelevant Alternatives, IIA)

F satisfies IIA if for all alternatives a, b the social preference between a and b depends only on the preferences of the voters between a and b.
Formally, for all $(\prec_1, \dots, \prec_n), (\prec'_1, \dots, \prec'_n) \in L^n$, $\prec := F(\prec_1, \dots, \prec_n)$, and $\prec' := F(\prec'_1, \dots, \prec'_n)$,

$$a \prec_i b \text{ iff } a \prec'_i b, \text{ for each } i = 1, \dots, n \implies a \prec b \text{ iff } a \prec' b.$$

Social Choice Theory
Arrow's Impossibility Theorem
Motivation
Properties of Social Welfare Functions
Main Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

Properties of Social Welfare Functions

Lemma

Total unanimity and independence of irrelevant alternatives together imply partial unanimity.

Proof

Consider any $\prec_1, \dots, \prec_n \in L$ with $a \prec_i b$ for all voters i.
To show: $a \prec b$ (with $\prec := F(\prec_1, \dots, \prec_n)$).
Define $\prec'_1, \dots, \prec'_n$ with $\prec'_i := \prec_1$ for each voter i.
By total unanimity, $\prec' := F(\prec'_1, \dots, \prec'_n) = F(\prec_1, \dots, \prec_1) = \prec_1$.
Hence, we have $a \prec' b$.
Moreover, $a \prec_i b$ iff $a \prec'_i b$, for all voters i.
By IIA, it follows $a \prec b$ iff $a \prec' b$.
From $a \prec' b$ we conclude that $a \prec b$ must hold.

Social Choice Theory
Arrow's Impossibility Theorem
Motivation
Properties of Social Welfare Functions
Main Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
Summary

Pairwise Neutrality

Lemma (pairwise neutrality)

Let F be a social welfare function satisfying (total or partial) unanimity and independence of irrelevant alternatives.
Let $(\prec_1, \dots, \prec_n)$ and $(\prec'_1, \dots, \prec'_n)$ be two preference profiles, $\prec := F(\prec_1, \dots, \prec_n)$ and $\prec' := F(\prec'_1, \dots, \prec'_n)$.
Then,

$$a \prec_i b \text{ iff } c \prec'_i d \text{ for each } i = 1, \dots, n \implies a \prec b \text{ iff } c \prec' d.$$

Pairwise Neutrality

Proof

Wlog., $a \prec b$ (otherwise, rename a and b) and ~~$c \neq d$~~ $c \neq b$ (otherwise, rename a and c as well as b and d).
Construct a new preference profile $(\prec''_1, \dots, \prec''_n)$, where $c \prec''_i a$ (unless $c = a$) and $b \prec''_i d$ (unless $b = d$) for all $i = 1, \dots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec'_i$.

By unanimity, we get $c \prec'' a$ and $b \prec'' d$ ($\prec'' := F(\prec''_1, \dots, \prec''_n)$).
Because of IIA, we have $a \prec'' b$.
By transitivity, we obtain $c \prec'' d$.
With IIA, it follows $c \prec' d$.

The proof for the opposite direction is similar.

Turns out the proof [Nisan 2007] is incomplete [Nipkow 2009].

The missed case

Proof

Let us assume $a \prec b$ and $a = d$ and $b = c$. I.e., we want to show: $a \prec_i b$ iff $b \prec'_i a$ for each $i \implies a \prec b$ iff $b \prec' a$.
Pick c and create $\prec''_i$ from $\prec_i$ by moving c directly below b, i.e., $a \prec_i b$ iff $a \prec''_i c$. This implies $a \prec b$ iff $a \prec'' c$ (by the previous part). Construct $\prec'''_i$ from $\prec''_i$ by moving b directly below a. Construct $\prec''''_i$ from $\prec'''_i$ by moving a directly below c. It follows that $a \prec'' c$ iff $b \prec''' c$ and $b \prec''' c$ iff $b \prec'''' a$. Comparing $\prec''''$ with $\prec$, we notice: $a \prec_i b$ iff $b \prec''''_i a$, hence $a \prec'_i b$ iff $a \prec''''_i b$. By IIA, it follows, $a \prec' b$ iff $a \prec'''' b$, yielding $a \prec b$ iff $b \prec' a$ as desired.

Arrow's Impossibility Theorem

Arrow's Impossibility Theorem

Every social welfare function over more than two alternatives that satisfies unanimity and independence of irrelevant alternatives is necessarily dictatorial.

Proof

We assume unanimity and independence of irrelevant alternatives.

Consider two elements $a, b \in A$ mit $a \neq b$ and construct a sequence $(\pi^i)_{i=0,\dots,n}$ of preference profiles such that in π^i exactly the first i voters prefer b to a, i.e., $a \prec_j b$ iff $j \leq i$:

...

Arrow's Impossibility Theorem

Proof (ctd.)

	π^0	$\dots$	π^{i^*-1}	π^{i^*}	$\dots$	π^n
1:	$b \prec_1 a$	$\dots$	$a \prec_1 b$	$a \prec_1 b$	$\dots$	$a \prec_1 b$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
i^*-1:	$b \prec_{i^*-1} a$	$\dots$	$a \prec_{i^*-1} b$	$a \prec_{i^*-1} b$	$\dots$	$a \prec_{i^*-1} b$
i^*:	$b \prec_{i^*} a$	$\dots$	$b \prec_{i^*} a$	$a \prec_{i^*} b$	$\dots$	$a \prec_{i^*} b$
i^*+1:	$b \prec_{i^*+1} a$	$\dots$	$b \prec_{i^*+1} a$	$b \prec_{i^*+1} a$	$\dots$	$a \prec_{i^*+1} b$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n:	$b \prec_n a$	$\dots$	$b \prec_n a$	$b \prec_n a$	$\dots$	$a \prec_n b$
F:	$b \prec^0 a$	$\dots$	$b \prec^{i^*-1} a$	$a \prec^{i^*} b$	$\dots$	$a \prec^n b$

Unanimity $\Rightarrow b \prec^0 a$ for $\prec^0 = F(\pi^0)$, $a \prec^n b$ for $\prec^n := F(\pi^n)$.

Thus, there must exist a minimal index i^* such that $b \prec^{i^*-1} a$ and $a \prec^{i^*} b$ for $\prec^{i^*-1} := F(\pi^{i^*-1})$ and $\prec^{i^*} = F(\pi^{i^*})$.

Arrow's Impossibility Theorem

Proof (ctd.)

Show that i^* is a dictator.

Consider two alternatives $c, d \in A$ with $c \neq d$ and show that for all $(\prec_1, \dots, \prec_n) \in L^n$, $c \prec_{i^*} d$ implies $c \prec d$, where $\prec = F(\prec_1, \dots, \prec_{i^*}, \dots, \prec_n)$.

Consider $e \notin \{c, d\}$ and construct preference profile $(\prec'_1, \dots, \prec'_n)$, where:

$$\begin{array}{lll} \text{for } j < i^*: & e \prec'_j c \prec'_j d \quad \text{or} & e \prec'_j d \prec'_j c \\ \text{for } j = i^*: & c \prec'_j e \prec'_j d \quad \text{or} & d \prec'_j e \prec'_j c \\ \text{for } j > i^*: & c \prec'_j d \prec'_j e \quad \text{or} & d \prec'_j c \prec'_j e \end{array}$$

depending on whether $c \prec_j d$ or $d \prec_j c$.

...

Arrow's Impossibility Theorem

Proof (ctd.)

Let $\prec' = F(\prec'_1, \dots, \prec'_n)$.

Independence of irrelevant alternatives implies $c \prec' d$ iff $c \prec d$.

	π^{i^*-1}	$(\prec'_i)_{i=1,\dots,n}$	π^{i^*}	$(\prec'_i)_{i=1,\dots,n}$
1:	$a \prec_1 b$	$e \prec'_1 c$	$a \prec_1 b$	$e \prec'_1 d$
i^*-1:	$a \prec_{i^*-1} b$	$e \prec'_{i^*-1} c$	$a \prec_{i^*-1} b$	$e \prec'_{i^*-1} d$
i^*:	$b \prec_{i^*} a$	$c \prec'_{i^*} e$	$a \prec_{i^*} b$	$e \prec'_{i^*} d$
n:	$b \prec_n a$	$c \prec'_n e$	$b \prec_n a$	$d \prec'_n e$
F:	$b \prec^{i^*-1} a$	$c \prec' e$	$a \prec^{i^*} b$	$e \prec' d$

For (e, c) we have the same preferences in $\prec'_1, \dots, \prec'_n$ as for (a, b) in π^{i^*-1}. Pairwise neutrality implies $c \prec' e$.

For (e, d) we have the same preferences in $\prec'_1, \dots, \prec'_n$ as for (a, b) in π^{i^*}. Pairwise neutrality implies $e \prec' d$.

...

Arrow's Impossibility Theorem

Proof (ctd.)

With transitivity, we get $c \prec' d$.

By construction of $\prec'$ and independence of irrelevant alternatives, we get $c \prec d$.

Opposite direction: similar. □

Arrow's Impossibility Theorem

Remark:

Unanimity and non-dictatorship often satisfied in social welfare functions. Problem usually lies with independence of irrelevant alternatives.

Closely related to possibility of strategic voting: insert "irrelevant" candidate between favorite candidate and main competitor to help favorite candidate (only possible if independence of irrelevant alternatives is violated).

3 Gibbard-Satterthwaite Theorem

- Motivation
- Preliminaries
- Main Theorem

Gibbard-Satterthwaite Theorem

Motivation:

- Arrow's Impossibility Theorem only applies to social welfare functions.
- Can this be transferred to social choice functions?
- Yes! Intuitive result: Every "reasonable" social choice function is susceptible to manipulation (strategic voting).

Strategic Manipulation and Incentive Compatibility

Definition (Strategic Manipulation, Incentive Compatibility)

A social choice function f can be strategically manipulated by voter i if there are preferences $\prec_1, \dots, \prec_i, \dots, \prec_n, \prec'_i \in L$ such that $a \prec_i b$ for $a = f(\prec_1, \dots, \prec_i, \dots, \prec_n)$ and $b = f(\prec_1, \dots, \prec'_i, \dots, \prec_n)$.

The function f is called incentive compatible if f cannot be strategically manipulated.

Definition (Monotonicity)

A social choice function is monotone if $f(\prec_1, \dots, \prec_i, \dots, \prec_n) = a$, $f(\prec_1, \dots, \prec'_i, \dots, \prec_n) = b$ and $a \neq b$ implies $b \prec_i a$ and $a \prec'_i b$.

Incentive Compatibility and Monotonicity

Proposition

A social choice function is monotone iff it is incentive compatible.

Proof

Let f be monotone. If $f(\prec_1, \dots, \prec_i, \dots, \prec_n) = a$, $f(\prec_1, \dots, \prec'_i, \dots, \prec_n) = b$ and $a \neq b$, then also $b \prec_i a$ and $a \prec'_i b$.

Then there cannot be any $\prec_1, \dots, \prec_n, \prec'_i \in L$ such that $f(\prec_1, \dots, \prec_i, \dots, \prec_n) = a$, $f(\prec_1, \dots, \prec'_i, \dots, \prec_n) = b$ and $a \prec_i b$.

Conversely, violated monotonicity implies that there is a possibility for strategic manipulation. □

Dictatorship in Social Choice Functions

Definition (Dictatorship)

Voter i is a dictator in a social choice function f if for all $\prec_1, \dots, \prec_i, \dots, \prec_n \in L$, $f(\prec_1, \dots, \prec_i, \dots, \prec_n) = a$, where a is the unique candidate with $b \prec_i a$ for all $b \in A$ with $b \neq a$.

The function f is a dictatorship if there is a dictator in f.

Gibbard-Satterthwaite Theorem

Reduction to Arrow's Theorem

Approach:

- We prove the result by Gibbard and Satterthwaite using Arrow's Theorem.
- To that end, construct social welfare function from social choice function.

Notation:

Let $S \subseteq A$ and $\prec \in L$. By $\prec^S$ we denote the order obtained by moving all elements from S "to the top" in $\prec$, while preserving the relative orderings of the elements in S and of those in $A \setminus S$.

More formally:

- for $a, b \in S$: $a \prec^S b$ iff $a \prec b$,
- for $a, b \notin S$: $a \prec^S b$ iff $a \prec b$,
- for $a \notin S$, $b \in S$: $a \prec^S b$.

These conditions uniquely define $\prec^S$.

Gibbard-Satterthwaite Theorem

Top-Preference Lemma

Lemma (Top Preference)

Let f be an incentive compatible and surjective social choice function. Then for all $\prec_1, \dots, \prec_n \in L$ and all $\emptyset \neq S \subseteq A$, we have $f(\prec_1^S, \dots, \prec_n^S) \in S$.

Proof

Let $a \in S$.

Since f is surjective, there are $\prec'_1, \dots, \prec'_n \in L$ such that $f(\prec'_1, \dots, \prec'_n) = a$.

Now, sequentially, for $i = 1, \dots, n$, change the relation $\prec'_i$ to $\prec_i^S$. At no point during this sequence of changes will f output any candidate $b \notin S$, because f is monotone.

Definition (Extension of a Social Choice Function)

The function $F : L^n \to L$ that extends the social choice function f is defined as $F(\prec_1, \dots, \prec_n) = \prec$, where $a \prec b$ iff $f(\prec_1^{\{a,b\}}, \dots, \prec_n^{\{a,b\}}) = b$ for all $a, b \in A, a \neq b$.

Lemma

If f is an incentive compatible and surjective social choice function, then its extension F is a social welfare function.

Proof

We show that $\prec$ is a strict linear order, i.e., asymmetric, total and transitive.

...

Proof (ctd.)

- Asymmetry and Totality: Because of the Top-Preference Lemma, $f(\prec_1^{\{a,b\}}, \dots, \prec_n^{\{a,b\}})$ is either a or b, i.e., $a \prec b$ or $b \prec a$, but not both (asymmetry) and not neither (totality).
- Transitivity: We may already assume totality. Suppose that $\prec$ is not transitive, i.e., $a \prec b$ and $b \prec c$, but not $a \prec c$, for some a, b and c. Because of totality, $c \prec a$. Consider $S = \{a, b, c\}$ and WLOG $f(\prec_1^{\{a,b,c\}}, \dots, \prec_n^{\{a,b,c\}}) = a$. Due to monotonicity of f, we get $f(\prec_1^{\{a,b\}}, \dots, \prec_n^{\{a,b\}}) = a$ by successively changing $\prec_i^{\{a,b,c\}}$ to $\prec_i^{\{a,b\}}$. Thus, we get $b \prec a$ in contradiction to our assumption. □

Social Choice Theory
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Motivation
Preliminaries
Main Theorem
Some Positive Results
Summary

Lemma (Extension Lemma)

If f is an incentive compatible, surjective, and non-dictatorial social choice function, then its extension F is a social welfare function that satisfies unanimity, independence of irrelevant alternatives, and non-dictatorship.

Proof

We already know that F is a social welfare function and still have to show unanimity, independence of irrelevant alternatives, and non-dictatorship.

- Unanimity: Let $a \prec_i b$ for all i. Then $(\prec_i^{\{a,b\}})^{\{b\}} = \prec_i^{\{a,b\}}$. Because of the Top-Preference Lemma, $f(\prec_1^{\{a,b\}}, \dots, \prec_n^{\{a,b\}}) = b$, hence $a \prec b$.
- Independence of irrelevant alternatives: ...

Social Choice Theory
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Motivation
Preliminaries
Main Theorem
Some Positive Results
Summary

Proof (ctd.)

- Independence of irrelevant alternatives: If for all i, $a \prec_i b$ iff $a \prec'_i b$, then $f(\prec_1^{\{a,b\}}, \dots, \prec_n^{\{a,b\}}) = f(\prec_1'^{\{a,b\}}, \dots, \prec_n'^{\{a,b\}})$ must hold, since due to monotonicity the result does not change when $\prec_i^{\{a,b\}}$ is successively replaced by $\prec_i'^{\{a,b\}}$.
- Non-dictatorship: Obvious. □

Social Choice Theory
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Motivation
Preliminaries
Main Theorem
Some Positive Results
Summary

Gibbard-Satterthwaite Theorem

Theorem (Gibbard-Satterthwaite)

If f is an incentive compatible and surjective social choice function with three or more alternatives, then f is a dictatorship. □

The purpose of mechanism design is to alleviate the negative results of Arrow and Gibbard and Satterthwaite by changing the underlying model. The two usually investigated modifications are:

- Introduction of money
- Restriction of admissible preference relations

4 Some Positive Results

- May's Theorem
- Single-Peaked Preferences

May's Theorem

We had some negative results on social choice and welfare functions so far: Arrow, Gibbard-Satterthwaite.

Question: Are there also positive results for special cases?

First special case: Only two alternatives.

Intuition: With only two alternatives, no point in misrepresenting preferences.

May's Theorem

Axioms for voting systems:

- Neutrality: "Names" of candidates/alternatives should not be relevant.
- Anonymity: "Names" of voters should not be relevant.
- Monotonicity: If a candidate wins, he should still win if one voter ranks him higher.

May's Theorem

Theorem (May, 1958)

A voting method for two alternatives satisfies anonymity, neutrality, and monotonicity if and only if it is the plurality method.

Proof.

$\Leftarrow$: Obvious.

$\Rightarrow$: For simplicity, we assume that the number of voters is odd.

Anonymity and neutrality imply that only the numbers of votes for the candidates matter.

Let A be the set of voters that prefer candidate a, and let B be the set of voters that prefer candidate b. Consider a vote with $|A| = |B| + 1$.

May's Theorem

Social Choice Theory
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
May's Theorem
Single-Peaked Preferences
Summary

Proof (ctd.)

- Case 1: Candidate a wins. Then by monotonicity, a still wins whenever $|A| > |B|$. With neutrality, we also get that b wins whenever $|B| > |A|$. This uniquely characterizes the plurality method.
- Case 2: Candidate b wins. Assume that one voter for a changes his preference to b. Then $|A'| + 1 = |B'|$. By monotonicity, b must still win. This is completely symmetric to the original vote. Hence, by neutrality, a should win. This is a contradiction, implying that case 2 cannot occur. □

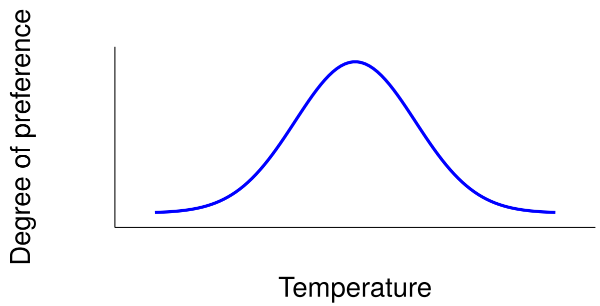
Remark: For three or more alternatives, there are no voting methods that satisfy such a small set of desirable criteria.

Single-Peaked Preferences

Social Choice Theory
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
May's Theorem
Single-Peaked Preferences
Summary

The results by Arrow and Gibbard-Satterthwaite only apply is there are no restrictions on the preference orders.

Second special case: Let us now consider some special cases such as temperature or volume settings.



Single-Peaked Preferences

Social Choice Theory
Arrow's Impossibility Theorem
Gibbard-Satterthwaite Theorem
Some Positive Results
May's Theorem
Single-Peaked Preferences
Summary

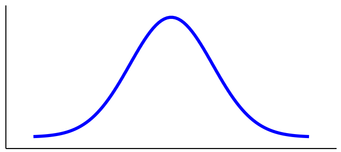
Definition (Single-peaked preference)

A preference relation $\prec_i$ over the interval $[0,1]$ is called a single-peaked preference relation if there exists a value $p_i \in [0,1]$ such that for all $x \in [0,1] \setminus p_i$ and for all $\lambda \in [0,1)$,

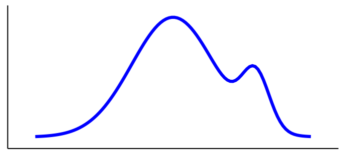
$$x \prec_i \lambda x + (1-\lambda) p_i.$$

Example

Single-peaked:



Not single-peaked:



Single-Peaked Preferences

First idea: Use arithmetic mean of all peak values.

Example

Preferred room temperatures:

- Voter 1: 10 °C
- Voter 2: 20 °C
- Voter 3: 21 °C

Arithmetic mean: 17 °C. Is this incentive compatible?

No! Voter 1 can misrepresent his peak value as, e.g., −11 °C.
Then the mean is 10 °C, his favorite value!

Question: What is a good way to design incentive compatible social choice functions for this setting?

Median Rule

Definition (Median rule)

Let $p_1, \dots, p_n$ be the peaks for the preferences $\prec_1, \dots, \prec_n$ ordered such that we have $p_1 \leq p_2 \leq \dots \leq p_n$. Then the median rule is the social choice function f with

$$f(\prec_1, \dots, \prec_n) = p_{\lceil n/2 \rceil}.$$

Theorem

The median rule is surjective, incentive compatible, anonymous, and non-dictatorial.

Median Rule

Proof.

- Surjective: Obvious, because the median rule satisfies unanimity.
- Incentive compatible: Assume that p_i is below the median. Then reporting a lower value does not change the median ($\leadsto$ does not help), and reporting a higher value can only increase the median ($\leadsto$ does not help, either). Similarly, if p_i is above the median.
- Anonymous: Is implicit in the rule.
- Non-dictatorial: Follows from anonymity. □

5 Summary

Summary

- Multitude of possible social welfare functions (plurality voting with or without runoff, instant runoff voting, Borda count, Schulze method, . . .).
- All social welfare functions for more than two alternatives suffer from Arrow's Impossibility Theorem.
- Typical handling of this issue: Use unanimous, non-dictatorial social welfare functions – violate independence of irrelevant alternatives.
- Thus: Strategic voting inevitable.
- The same holds for social choice functions (Gibbard-Satterthwaite Theorem).