Game Theory

7. Social Choice Theory

7.1. Introduction and Examples

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

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

UNI FREIBURG

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.

- 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, spoiler effect, 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")

- 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.

$\rightsquigarrow$ Is there any voting system without such problems?
Or is there some deeper underlying reason for all those problems?

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:

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)

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 with runoff:

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 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)

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:

UNI FREIBURG

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.

UNI FREIBURG

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:

UNI FREIBURG

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.

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:

UNI FREIBURG

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
c	1	0	–	1	1
d	1	0	0	–	0
e	0	0	0	1	–

$\longleftarrow$ candidate b wins.

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!

Summary

- Multitude of possible social welfare functions (plurality voting with or without runoff, instant runoff voting, Borda count, ...).
- Tactical voting seems to be possible in all of them.
- May lead to different winners.
- Strategic choice of voting system.

Game Theory

7. Social Choice Theory

7.2. Condorcet Methods

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

Example: voters 1, 2, 3; candidates a, b, 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 returns a Condorcet winner, if one exists.

One particular Condorcet method: the Schulze method.

Relatively new: proposed in 1997

Already many users: Debian, Ubuntu, Pirate Parties, Associated Student Government at Uni Freiburg (Studierendenrat, StuRa), ...

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 $j = 1, \ldots, n-1$ such that $d(C_j, C_{j+1}) = z$

Example: path between a and d of strength 5:

$$a \xrightarrow{8} b \xrightarrow{5} c \xrightarrow{6} d$$

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.

# 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	–

$\rightsquigarrow$ No!

# 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	–

Schulze Method

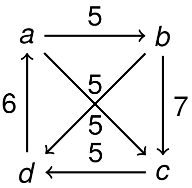
Example

# 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:



Schulze Method

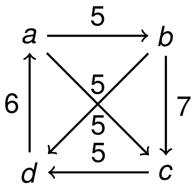
Example

# 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 $p(X, Y)$:

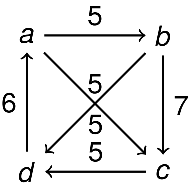
	a	*b*	*c*	*d*
a	–	5	5	5
b	5	–	7	5
c	5	5	–	5
d	6	5	5	–

# 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.

Schulze Method

Theorem

There is always at least one potential winner.

Proof.

Homework. □

Summary

- Condorcet paradox: cyclical social preferences
  $\rightsquigarrow$ Condorcet winner may not exist
- Condorcet methods produce Condorcet winner if it exists
- Example: Schulze method
  (satisfies many desirable criteria, see
  https://en.wikipedia.org/wiki/Schulze_method#Satisfied_criteria)

Game Theory

7. Social Choice Theory

7.3. Arrow’s Impossibility Theorem

7.3.1. Properties of Social Welfare Functions

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

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!

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.

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 independence of irrelevant alternatives (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 $\prec' := F(\prec'_1, \dots, \prec'_n)$,

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

UNI FREIBURG

Lemma

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

Proof.

Consider any $\prec_1, \ldots, \prec_n \in L$ with $a \prec_i b$ for all voters i.

To show: $a \prec b$, where $\prec := F(\prec_1, \ldots, \prec_n)$.

Define $\prec'_1, \ldots, \prec'_n$ with $\prec'_i := \prec_1$ for each voter i.

By total unanimity, $\prec' := F(\prec'_1, \ldots, \prec'_n) = F(\prec_1, \ldots, \prec_1) = \prec_1$.

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. □

Pairwise Neutrality

Neutrality $\approx$ candidates are treated symmetrically
(i. e., no bias, “names” of the candidates do not matter)

Definition (pairwise neutrality)

A social welfare function F satisfies pairwise neutrality if, for any two preference profiles $(\prec_1, \dots, \prec_n)$ and $(\prec'_1, \dots, \prec'_n)$,

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

where $\prec := F(\prec_1, \dots, \prec_n)$ and $\prec' := F(\prec'_1, \dots, \prec'_n)$.

Pairwise Neutrality

Lemma

(Total or partial) unanimity and independence of irrelevant alternatives together imply pairwise neutrality.

Proof sketch.

Assume that a, b, c, d are pairwise different. WLOG, $a \prec b$.

Construct a new preference profile $(\prec''_1, \dots, \prec''_n)$, where $c \prec''_i a$ and $b \prec''_i d$ for all $i = 1, \dots, n$, the order of the pairs (a, b) is taken 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 that $c \prec' d$.

The proof for the opposite direction is similar.

[Technical details if a, b, c, d not pairwise different omitted.] □

Summary

- Relevant properties of social welfare functions:
  - unanimity (total or partial)
  - non-dictatorship
  - independence of irrelevant alternatives (IIA)
  - pairwise neutrality
- Given IIA, total and partial unanimity are the same.
- Unanimity and IIA imply pairwise neutrality.

Game Theory

7. Social Choice Theory

7.3. Arrow's Impossibility Theorem

7.3.2. Proof

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

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 $a, b \in A$ with $a \neq b$ 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 $j \leq i$:

…

UNI FREIBURG

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:

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

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

...

UNI FREIBURG

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 $e \prec' d$.

...

UNI FREIBURG

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. $\square$

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).

Summary

- All social welfare functions for more than two alternatives suffer from Arrow's Impossibility Theorem:

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

- Typical handling of this issue: use unanimous, non-dictatorial social welfare functions – violate independence of irrelevant alternatives

  $\rightsquigarrow$ strategic voting inevitable

Game Theory

7. Social Choice Theory

7.4. Gibbard-Satterthwaite Theorem

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

Gibbard-Satterthwaite Theorem

UNI FREIBURG

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

UNI FREIBURG

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

UNI FREIBURG

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.

We are going to prove the theorem of Gibbard and Satterthwaite:

Every incentive compatible and surjective social choice function with three or more alternatives is necessarily a dictatorship.

Approach:

- We prove the result 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 $\prec^S$.

Example

Let $d \prec a \prec c \prec b \prec e$, and $S = \{a, b\}$.
Then $d \prec^S c \prec^S e \prec^S a \prec^S b$.

UNI FREIBURG

Lemma (top preference)

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

Proof.

Let $a \in S$.

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

Now, sequentially, for $i = 1, \ldots, n$, 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. $\square$

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.

…

UNI FREIBURG

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. $\square$

Gibbard-Satterthwaite Theorem

Extension Lemma

UNI FREIBURG

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 $a \prec b$.
- ...

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. $\square$

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 (Sections 8.1–8.3)
- Restriction of admissible preference relations (Sections 7.5.2, 8.4)

Summary

- Result corresponding to Arrow's theorem for social choice functions (Gibbard-Satterthwaite):

  Every incentive compatible and surjective social choice function with three or more alternatives is necessarily a dictatorship.

- Proof: reduction to Arrow's theorem
- Outlook (not further discussed here): score vs. ranked voting systems?

Game Theory

7. Social Choice Theory

7.5. Some Positive Results

7.5.1 May's Theorem

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel and Robert Mattmüller

Summer semester 2020

May's Theorem

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

Question: any 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

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. $\square$

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

Summary

- With only two alternatives, there is a positive result.
- May's theorem:
  A voting method for two alternatives satisfies anonymity, neutrality, and monotonicity if and only if it is the plurality method.
- Note:
  $|A| = 2 \Rightarrow$ plurality = plurality+runoff = IRV = Borda = ...

Game Theory

7. Social Choice Theory

7.5. Some Positive Results

7.5.2 Single-Peaked Preferences

Albert-Ludwigs-Universität Freiburg

Bernhard Nebel and Robert Mattmüller

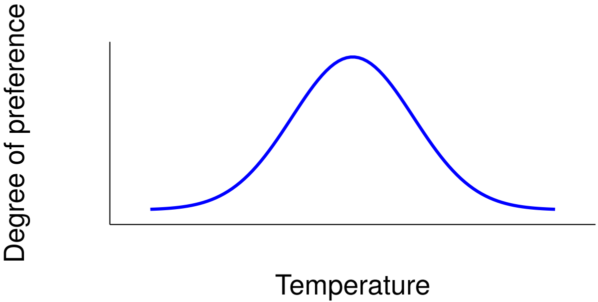
Summer semester 2020

UNI FREIBURG

Single-Peaked Preferences

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

Second special case: restrictions on preference orders



Single-Peaked Preferences

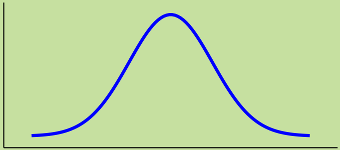
Definition (single-peaked preference)

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 $\lambda \in [0,1)$,

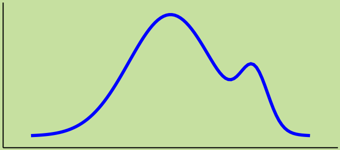
$$x \prec_i \lambda x + (1-\lambda) p_i .$$

Example

Single-peaked:



Not single-peaked:



Single-Peaked Preferences

First idea: Use arithmetic mean of all peak values.

Example

Preferred room temperatures:

- Voter 1: $10\,^{\circ}\mathrm{C}$
- Voter 2: $20\,^{\circ}\mathrm{C}$
- Voter 3: $21\,^{\circ}\mathrm{C}$

Arithmetic mean: $17\,^{\circ}\mathrm{C}$. Is this incentive compatible?

Single-Peaked Preferences

First idea: Use arithmetic mean of all peak values.

Example

Preferred room temperatures:

- Voter 1: $10\,^{\circ}\mathrm{C}$
- Voter 2: $20\,^{\circ}\mathrm{C}$
- Voter 3: $21\,^{\circ}\mathrm{C}$

Arithmetic mean: $17\,^{\circ}\mathrm{C}$. Is this incentive compatible?

No! Voter 1 can misrepresent his peak value as, e.g., $-11\,^{\circ}\mathrm{C}$.

Then the mean is $10\,^{\circ}\mathrm{C}$, 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 $p_1, \ldots, p_n$ be the peaks for the preferences $\prec_1, \ldots, \prec_n$ ordered such that we have $p_1 \leq p_2 \leq \cdots \leq p_n$. Then the median rule is the social choice function f with

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

Example

Preferred room temperatures:

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

Median: 20 °C.

Is this incentive compatible?

Median Rule

Theorem

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

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 ($\rightsquigarrow$ does not help), and reporting a higher value can only increase the median ($\rightsquigarrow$ does not help, either). Similarly, if p_i is above the median.
- Anonymous: Is implicit in the rule.
- Non-dictatorial: Follows from anonymity. $\square$

Summary

- With restricted type of preferences, there is a positive result.
- The median rule returns the median value among the reported peaks (of single-peaked preferences).
- The median rule is surjective, incentive compatible, anonymous, and non-dictatorial.