Game Theory

16. Social Choice Theory

Albert-Ludwigs-Universität Freiburg

Bernhard Nebel, Robert Mattmüller,
Stefan Wölfl, Christian Becker-Asano

Research Group Foundations of Artificial Intelligence

June 24, 2013

Social Choice Theory

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")?

Definition (Social Welfare and Social Choice Function)

Let A be a set of alternatives (candidates) and L be a set of linear orders on A. For n voters, $F : L^n \to L$ denotes a social welfare function and $f : L^n \to A$ a social choice function.

Notation: A linear order $\prec \in L$ is called a preference relation. For voter i, $\prec_i$ is the preference relation of that voter. E.g., $a \prec_i b$ means that voter i prefers candidate b 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:
- Plurality voting with runoff:
  - first round:
  - second round:

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:
  - second round:

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:

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:

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

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

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

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

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

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.

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

Example: path of strength 3.

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

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	–

$\rightsquigarrow$ 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	–

Schulze Method

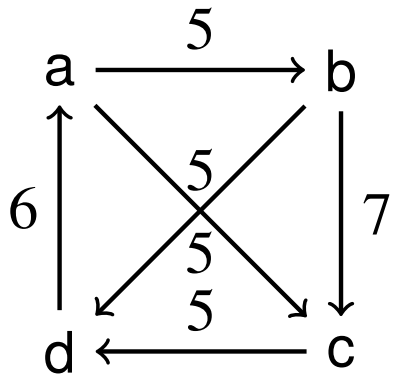
Example

# 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

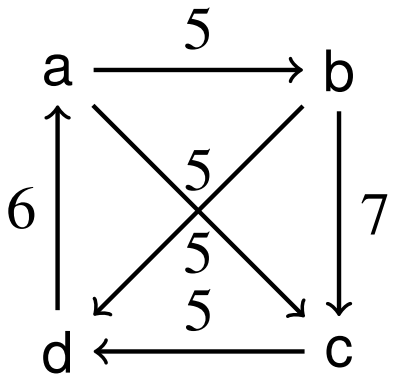
Example

# 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	–

Schulze Method

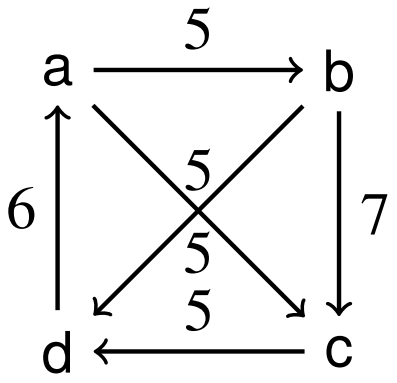
Example

# 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.

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].

Arrow's Impossibility 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 (Total Unanimity)

For all $\prec \in L$, $F(\prec, \ldots, \prec) = \prec$.

Definition (Partial Unanimity)

For all $\prec_1, \prec_2, \ldots, \prec_n, \prec \in L$ with $F(\prec_1, \ldots, \prec_n) = \prec$, if $a \prec_i b$ for all $i = 1, \ldots, n$, then also $a \prec b$.

Remark

Partial unanimity implies total unanimity, but not vice versa.

Properties of Social Welfare Functions

Desirable properties of social welfare functions:

Definition (Non-Dictatorship)

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

Definition (Independence of Irrelevant Alternatives, IIA)

Whether $a \prec b$ only depends on the preferences of the voters between a and b, i.e., for all $\prec_1, \ldots, \prec_n, \prec'_1, \ldots, \prec'_n \in L$ the following has to hold: if $\prec = F(\prec_1, \ldots, \prec_n)$ and $\prec' = F(\prec'_1, \ldots, \prec'_n)$ and $a \prec_i b$ iff $a \prec'_i b$ for all $i = 1, \ldots, n$, then this implies that $a \prec b$ iff $a \prec' b$.

Properties of Social Welfare Functions

Lemma

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

Proof

Let $\prec_1, \ldots, \prec_n \in L$ with $a \prec_i b$ for all voters i. Let $\prec = F(\prec_1, \ldots, \prec_n)$. Consider $\prec'_1, \ldots, \prec'_n$ with $\prec'_i = \prec_1$ for all voters i. Obviously, with total unanimity we get $\prec' = F(\prec'_1, \ldots, \prec'_n) = F(\prec_1, \ldots, \prec_1) = \prec_1$. Hence, we have $a \prec' b$. Since $a \prec_i b$ iff $a \prec'_i b$, with independence of irrelevant alternatives this implies that also $a \prec b$ iff $a \prec' b$. Because we know that $a \prec' b$ holds, also $a \prec b$ must hold.

Properties of Social Welfare Functions

Lemma

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

Proof

Let $\prec_1, \ldots, \prec_n \in L$ with $a \prec_i b$ for all voters i. Let $\prec = F(\prec_1, \ldots, \prec_n)$. Consider $\prec'_1, \ldots, \prec'_n$ with $\prec'_i = \prec_1$ for all voters i. Obviously, with total unanimity we get $\prec' = F(\prec'_1, \ldots, \prec'_n) = F(\prec_1, \ldots, \prec_1) = \prec_1$. Hence, we have $a \prec' b$. Since $a \prec_i b$ iff $a \prec'_i b$, with independence of irrelevant alternatives this implies that also $a \prec b$ iff $a \prec' b$. Because we know that $a \prec' b$ holds, also $a \prec b$ must hold.

Properties of Social Welfare Functions

Lemma

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

Proof

Let $\prec_1, \ldots, \prec_n \in L$ with $a \prec_i b$ for all voters i. Let $\prec = F(\prec_1, \ldots, \prec_n)$. Consider $\prec'_1, \ldots, \prec'_n$ with $\prec'_i = \prec_1$ for all voters i. Obviously, with total unanimity we get $\prec' = F(\prec'_1, \ldots, \prec'_n) = F(\prec_1, \ldots, \prec_1) = \prec_1$. Hence, we have $a \prec' b$. Since $a \prec_i b$ iff $a \prec'_i b$, with independence of irrelevant alternatives this implies that also $a \prec b$ iff $a \prec' b$. Because we know that $a \prec' b$ holds, also $a \prec b$ must hold.

Properties of Social Welfare Functions

Lemma

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

Proof

Let $\prec_1, \ldots, \prec_n \in L$ with $a \prec_i b$ for all voters i. Let $\prec = F(\prec_1, \ldots, \prec_n)$. Consider $\prec'_1, \ldots, \prec'_n$ with $\prec'_i = \prec_1$ for all voters i. Obviously, with total unanimity we get $\prec' = F(\prec'_1, \ldots, \prec'_n) = F(\prec_1, \ldots, \prec_1) = \prec_1$. Hence, we have $a \prec' b$. Since $a \prec_i b$ iff $a \prec'_i b$, with independence of irrelevant alternatives this implies that also $a \prec b$ iff $a \prec' b$. Because we know that $a \prec' b$ holds, also $a \prec b$ must hold.

Properties of Social Welfare Functions

Lemma

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

Proof

Let $\prec_1, \ldots, \prec_n \in L$ with $a \prec_i b$ for all voters i. Let $\prec = F(\prec_1, \ldots, \prec_n)$. Consider $\prec'_1, \ldots, \prec'_n$ with $\prec'_i = \prec_1$ for all voters i. Obviously, with total unanimity we get $\prec' = F(\prec'_1, \ldots, \prec'_n) = F(\prec_1, \ldots, \prec_1) = \prec_1$. Hence, we have $a \prec' b$. Since $a \prec_i b$ iff $a \prec'_i b$, with independence of irrelevant alternatives this implies that also $a \prec b$ iff $a \prec' b$. Because we know that $a \prec' b$ holds, also $a \prec b$ must hold.

Properties of Social Welfare Functions

Lemma

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

Proof

Let $\prec_1, \ldots, \prec_n \in L$ with $a \prec_i b$ for all voters i. Let $\prec = F(\prec_1, \ldots, \prec_n)$. Consider $\prec'_1, \ldots, \prec'_n$ with $\prec'_i = \prec_1$ for all voters i. Obviously, with total unanimity we get $\prec' = F(\prec'_1, \ldots, \prec'_n) = F(\prec_1, \ldots, \prec_1) = \prec_1$. Hence, we have $a \prec' b$. Since $a \prec_i b$ iff $a \prec'_i b$, with independence of irrelevant alternatives this implies that also $a \prec b$ iff $a \prec' b$. Because we know that $a \prec' b$ holds, also $a \prec b$ must hold.

Pairwise Neutrality

Lemma (pairwise neutrality)

Assume F satisfies (total or partial) unanimity and independence of irrelevant alternatives. Let $\prec_1, \ldots, \prec_n$ and $\prec'_1, \ldots, \prec'_n$ be two preference profiles with $a \prec_i b$ iff $c \prec'_i d$ for all $i = 1, \ldots, n$. Then $a \prec b$ iff $c \prec' d$, if $\prec = F(\prec_1, \ldots, \prec_n)$ and $\prec' = F(\prec'_1, \ldots, \prec'_n)$.

Pairwise Neutrality

Proof

WLOG, $a \prec b$ (otherwise, rename a and b) and $b \neq c$ (otherwise, rename a and c as well as b and d). We construct a new preference profile $\prec_1'', \ldots, \prec_n''$, where $c \prec_i'' a$ and $b \prec_i'' d$ for all $i = 1, \ldots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec_i'$.

Due to the unanimity we get $c \prec'' a$ and $b \prec'' d$ for $\prec'' = F(\prec_1'', \ldots, \prec_n'')$. Because of the independence of irrelevant alternatives, we have $a \prec'' b$. Then with transitivity, we obtain $c \prec'' d$. With independence of irrelevant alternatives, we eventually get $c \prec' d$.

The proof for the opposite direction is similar.

Pairwise Neutrality

Proof

WLOG, $a \prec b$ (otherwise, rename a and b) and $b \neq c$ (otherwise, rename a and c as well as b and d). We construct a new preference profile $\prec_1'', \ldots, \prec_n''$, where $c \prec_i'' a$ and $b \prec_i'' d$ for all $i = 1, \ldots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec_i'$.

Due to the unanimity we get $c \prec'' a$ and $b \prec'' d$ for $\prec'' = F(\prec_1'', \ldots, \prec_n'')$. Because of the independence of irrelevant alternatives, we have $a \prec'' b$. Then with transitivity, we obtain $c \prec'' d$. With independence of irrelevant alternatives, we eventually get $c \prec' d$.

The proof for the opposite direction is similar.

Pairwise Neutrality

Proof

WLOG, $a \prec b$ (otherwise, rename a and b) and $b \neq c$ (otherwise, rename a and c as well as b and d). We construct a new preference profile $\prec_1'', \ldots, \prec_n''$, where $c \prec_i'' a$ and $b \prec_i'' d$ for all $i = 1, \ldots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec_i'$.

Due to the unanimity we get $c \prec'' a$ and $b \prec'' d$ for $\prec'' = F(\prec_1'', \ldots, \prec_n'')$. Because of the independence of irrelevant alternatives, we have $a \prec'' b$. Then with transitivity, we obtain $c \prec'' d$. With independence of irrelevant alternatives, we eventually get $c \prec' d$.

The proof for the opposite direction is similar.

Pairwise Neutrality

Proof

WLOG, $a \prec b$ (otherwise, rename a and b) and $b \neq c$ (otherwise, rename a and c as well as b and d). We construct a new preference profile $\prec_1'', \ldots, \prec_n''$, where $c \prec_i'' a$ and $b \prec_i'' d$ for all $i = 1, \ldots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec_i'$.

Due to the unanimity we get $c \prec'' a$ and $b \prec'' d$ for $\prec'' = F(\prec_1'', \ldots, \prec_n'')$. Because of the independence of irrelevant alternatives, we have $a \prec'' b$. Then with transitivity, we obtain $c \prec'' d$. With independence of irrelevant alternatives, we eventually get $c \prec' d$.

The proof for the opposite direction is similar.

Pairwise Neutrality

Proof

WLOG, $a \prec b$ (otherwise, rename a and b) and $b \neq c$ (otherwise, rename a and c as well as b and d). We construct a new preference profile $\prec_1'', \ldots, \prec_n''$, where $c \prec_i'' a$ and $b \prec_i'' d$ for all $i = 1, \ldots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec_i'$.

Due to the unanimity we get $c \prec'' a$ and $b \prec'' d$ for $\prec'' = F(\prec_1'', \ldots, \prec_n'')$. Because of the independence of irrelevant alternatives, we have $a \prec'' b$. Then with transitivity, we obtain $c \prec'' d$. With independence of irrelevant alternatives, we eventually get $c \prec' d$.

The proof for the opposite direction is similar.

Pairwise Neutrality

Proof

WLOG, $a \prec b$ (otherwise, rename a and b) and $b \neq c$ (otherwise, rename a and c as well as b and d). We construct a new preference profile $\prec_1'', \ldots, \prec_n''$, where $c \prec_i'' a$ and $b \prec_i'' d$ for all $i = 1, \ldots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec_i'$.

Due to the unanimity we get $c \prec'' a$ and $b \prec'' d$ for $\prec'' = F(\prec_1'', \ldots, \prec_n'')$. Because of the independence of irrelevant alternatives, we have $a \prec'' b$. Then with transitivity, we obtain $c \prec'' d$. With independence of irrelevant alternatives, we eventually get $c \prec' d$.

The proof for the opposite direction is similar.

Pairwise Neutrality

Proof

WLOG, $a \prec b$ (otherwise, rename a and b) and $b \neq c$ (otherwise, rename a and c as well as b and d). We construct a new preference profile $\prec_1'', \ldots, \prec_n''$, where $c \prec_i'' a$ and $b \prec_i'' d$ for all $i = 1, \ldots, n$, whereas the order of the pairs (a, b) is copied from $\prec_i$ and the order of the pairs (c, d) is taken from $\prec_i'$.

Due to the unanimity we get $c \prec'' a$ and $b \prec'' d$ for $\prec'' = F(\prec_1'', \ldots, \prec_n'')$. Because of the independence of irrelevant alternatives, we have $a \prec'' b$. Then with transitivity, we obtain $c \prec'' d$. With independence of irrelevant alternatives, we eventually get $c \prec' d$.

The proof for the opposite direction is similar.

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 $(\pi^i)_{i=0,\ldots,n}$ 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

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 $(\pi^i)_{i=0,\ldots,n}$ 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.)

	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 $\prec_1, \ldots, \prec_n \in L$, $c \prec_{i^*} d$ implies $c \prec d$, where $\prec = F(\prec_1, \ldots, \prec_{i^*}, \ldots, \prec_n)$.

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

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

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

...

Arrow's Impossibility Theorem

Proof (ctd.)

Show that i^* is a dictator.

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

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

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

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

...

Arrow's Impossibility Theorem

Proof (ctd.)

Show that i^* is a dictator.

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

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

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

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

...

Proof (ctd.)

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

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

	π^{i^*-1}	$(\prec'_i)_{i=1,\ldots,n}$	π^{i^*}	$(\prec'_i)_{i=1,\ldots,n}$
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 $\prec'_1, \ldots, \prec'_n$ as for (a, b) in π^{i^*-1}. Pairwise neutrality implies $c \prec' e$.

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

...

Proof (ctd.)

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

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

	π^{i^*-1}	$(\prec'_i)_{i=1,\ldots,n}$	π^{i^*}	$(\prec'_i)_{i=1,\ldots,n}$
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 $\prec'_1, \ldots, \prec'_n$ as for (a,b) in π^{i^*-1}. Pairwise neutrality implies $c \prec' e$.

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

...

Proof (ctd.)

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

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

	π^{i^*-1}	$(\prec'_i)_{i=1,\ldots,n}$	π^{i^*}	$(\prec'_i)_{i=1,\ldots,n}$
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 $\prec'_1, \ldots, \prec'_n$ as for (a,b) in π^{i^*-1}. Pairwise neutrality implies $c \prec' e$.

For (e,d) we have the same preferences in $\prec'_1, \ldots, \prec'_n$ 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

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

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. $\square$

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).

Gibbard-Satterthwaite 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 $\prec_1, \ldots, \prec_i, \ldots, \prec_n, \prec_i' \in L$ such that $a \prec_i b$ for $a = f(\prec_1, \ldots, \prec_i, \ldots, \prec_n)$ and $b = f(\prec_1, \ldots, \prec_i', \ldots, \prec_n)$.

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

Definition (Monotonicity)

A social choice function is monotone if $f(\prec_1, \ldots, \prec_i, \ldots, \prec_n) = a, f(\prec_1, \ldots, \prec_i', \ldots, \prec_n) = b$ 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 $f(\prec_1, \ldots, \prec_i, \ldots, \prec_n) = a$, $f(\prec_1, \ldots, \prec_i', \ldots, \prec_n) = b$ and $a \neq b$, then also $b \prec_i a$ and $a \prec_i' b$.

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

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

Incentive Compatibility and Monotonicity

Proposition

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

Proof

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

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

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

Incentive Compatibility and Monotonicity

Proposition

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

Proof

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

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

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.

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.

Gibbard-Satterthwaite Theorem

Extension of a Social Choice Function

Definition (Extension of a Social Choice Function)

The function $F : L^n \to L$ that extends the social choice function f is defined as $F(\prec_1, \ldots, \prec_n) = \prec$, where $a \prec b$ iff $f(\prec_1^{\{a,b\}}, \ldots, \prec_n^{\{a,b\}}) = b$ 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, $f(\prec_1^{\{a,b\}}, \ldots, \prec_n^{\{a,b\}})$ 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 $f(\prec_1^{\{a,b,c\}}, \ldots, \prec_n^{\{a,b,c\}}) = a$. Due to monotonicity of f, we get $f(\prec_1^{\{a,b\}}, \ldots, \prec_n^{\{a,b\}}) = a$ by successively changing $\prec_i^{\{a,b,c\}}$ to $\prec_i^{\{a,b\}}$. Thus, we get $b \prec a$ in contradiction to our assumption. $\square$

Proof (ctd.)

- Asymmetry and Totality: Because of the Top-Preference Lemma, $f(\prec_1^{\{a,b\}}, \ldots, \prec_n^{\{a,b\}})$ 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 $f(\prec_1^{\{a,b,c\}}, \ldots, \prec_n^{\{a,b,c\}}) = a$. Due to monotonicity of f, we get $f(\prec_1^{\{a,b\}}, \ldots, \prec_n^{\{a,b\}}) = a$ by successively changing $\prec_i^{\{a,b,c\}}$ to $\prec_i^{\{a,b\}}$. Thus, we get $b \prec a$ in contradiction to our assumption. $\square$

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, $f(\prec_1^{\{a,b\}}, \ldots, \prec_n^{\{a,b\}}) = b$, hence $a \prec b$.
- Independence of irrelevant alternatives: ...

Proof (ctd.)

- Independence of irrelevant alternatives: If for all i, $a \prec_i b$ iff $a \prec'_i b$, then $f(\prec_1^{\{a,b\}}, \ldots, \prec_n^{\{a,b\}}) = f(\prec_1'^{\{a,b\}}, \ldots, \prec_n'^{\{a,b\}})$ 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
- Restriction of admissible preference relations

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).