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")?

June 14th, 2016

B. Nebel, R. Mattmüller – Game Theory

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

$$F : L^n \rightarrow L$$

is called a social welfare function. A function

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

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 $a$. 

June 14th, 2016

B. Nebel, R. Mattmüller – Game Theory
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.

- 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.

- 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”).

---

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

<table>
<thead>
<tr>
<th># voters</th>
<th>8</th>
<th>6</th>
<th>4</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>e</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>2nd</td>
<td>d</td>
<td>b</td>
<td>c</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>3rd</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>4th</td>
<td>c</td>
<td>e</td>
<td>a</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>5th</td>
<td>a</td>
<td>d</td>
<td>e</td>
<td>e</td>
<td>a</td>
</tr>
</tbody>
</table>

- 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.

<table>
<thead>
<tr>
<th># voters</th>
<th>8</th>
<th>6</th>
<th>4</th>
<th>3</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>e</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td>2nd</td>
<td>d</td>
<td>b</td>
<td>c</td>
<td>b</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>3rd</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>4th</td>
<td>a</td>
<td>e</td>
<td>c</td>
<td>a</td>
<td>b</td>
<td>e</td>
</tr>
<tr>
<td>5th</td>
<td>a</td>
<td>d</td>
<td>c</td>
<td>a</td>
<td>e</td>
<td>e</td>
</tr>
</tbody>
</table>

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


- **Condorcet winner:** Ex.: a \succ b 16 times, b \succ a 7 times

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **Borda count:**
  - Cand. a: 8 + 6 + 4 + 6 = 24 pts
  - Cand. b: 8 + 6 + 4 + 4 = 22 pts
  - Cand. c: 8 + 6 + 1 + 3 + 1 = 18 pts
  - Cand. d: 8 + 6 + 3 + 2 + 1 + 3 = 22 pts
  - Cand. e: 8 + 4 + 6 + 1 + 0 + 1 = 16 pts

- **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 $c$.

$$
egin{align*}
  a &<_1 b <_1 c \\
  b &<_2 c <_2 a \\
  c &<_3 a <_3 b
\end{align*}
$$

Then we have cyclical preferences.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

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

### 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
- $d(C_i, C_{j+1}) \geq z$ 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.

$$
\begin{align*}
  a &\rightarrow b \rightarrow c \rightarrow d
\end{align*}
$$
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].

2 Arrow’s Impossibility Theorem

- Motivation
- Properties of Social Welfare Functions
- Main Theorem
**Properties of Social Welfare Functions**

**Desirable properties of social welfare functions:**

**Definition (Unanimity)**
A social welfare function satisfies
- total unanimity if for all \( \prec \in L \), \( F(\prec_1, \ldots, \prec_n) = \prec \).
- partial unanimity if for all \( \prec_1, \prec_2, \ldots, \prec_n \in L \), \( a, b \in A \),
  \[ a \prec_i b \text{ for each } i = 1, \ldots, n \implies a \prec b \]
where \( \prec := F(\prec_1, \ldots, \prec_n) \).

**Remark**
Partial unanimity implies total unanimity, but not vice versa.

**Properties of Social Welfare Functions**

**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: \( a \prec b \) (with \( \prec := F(\prec_1, \ldots, \prec_n) \)).

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, \( a \prec_i b \) if \( a \prec'_i b \), for all voters \( i \).

By IIA, it follows \( a \prec b \) if \( a \prec' b \).

From \( a \prec' b \) we conclude that \( a \prec b \) must hold.
Pairwise Neutrality

Lemma (pairwise neutrality)
Let $F$ be a social welfare function satisfying (total or partial) unanimity and independence of irrelevant alternatives. Let $(\succ,\ldots,\succ_n)$ and $(\succ',\ldots,\succ'_n)$ be two preference profiles, $\succ := F(\prec,\ldots,\prec_n)$ and $\succ' := F(\prec',\ldots,\prec'_n)$. Then,

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

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 $a \prec'''_i c$ iff $b \prec''_i c$ and $b \prec'''_i c$ iff $b \prec''_i a$. Comparing $\prec'''_i$ 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.

The missed case

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 $(\pi'_i)_{i=0, \ldots,n}$ of preference profiles such that in $\pi'_i$ 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.)

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

Independence of irrelevant alternatives implies $c \not\prec' d$ if $c \not\prec d$.

$$
\begin{array}{c|c|c}
\pi'^{-1} & (\prec'_i)_{i=1}^n & \pi'^{+} \\
\hline
1: & a \prec b & \pi'^{+} \\
\vdots & \vdots & \vdots \\
i^*: & a \prec b & \pi'^{+} \\
\vdots & \vdots & \vdots \\
\n: & a \prec b & \pi'^{+} \\
\end{array}
$$

For $(e, c)$ we have the same preferences in $\prec'_1, \ldots, \prec'_n$ as for $(a, b)$ in $\pi'^{+}$. Pairwise neutrality implies $c \not\prec e$.

For $(e, d)$ we have the same preferences in $\prec'_1, \ldots, \prec'_n$ as for $(a, b)$ in $\pi'^{+}$. Pairwise neutrality implies $e \not\prec d$.

...
**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**

**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).

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

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 $a, b \not\in S$: $a \prec^S b$ iff $a \prec b$.
- for $a \not\in S, b \in S$: $a \prec^S b$.

These conditions uniquely define $\prec^S$.

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

Proof
Let $a \in S$.
Since $f$ is surjective, there are $\prec_1, \dotsc, \prec_n \in L$ such that $f(\prec_1, \dotsc, \prec_n) = a$.

Now, sequentially, for $i = 1, \dotsc, n$, change the relation $\prec_i$ to $\prec_i^S$. At no point during this sequence of changes will $f$ output any candidate $b \not\in S$, because $f$ is monotone.
Gibbard-Satterthwaite Theorem

Extension of a Social Choice Function

Definition (Extension of a Social Choice Function)
The function $F : L^n \rightarrow L$ that extends the social choice function $f$ is defined as $F(\prec_1, \ldots, \prec_n) = \prec_i$, 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 $\prec_i$ 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 $\prec_i$ is not transitive, i.e., $a \prec_i b$ and $b \prec_i c$, but not $a \prec_i c$, for some $a, b$ and $c$. Because of totality, $c \prec_i a$. Consider $S = \{a, b, c\}$ and WLOG $f(\prec_1^{a,b}, \ldots, \prec_n^{a,b}) = a$. Due to monotonicity of $f$, we get $f(\prec_1^{a,c}, \ldots, \prec_n^{a,b}) = a$ by successively changing $\prec_i^{a,b}$ to $\prec_i^{a,c}$. Thus, we get $b \prec a$ in contradiction to our assumption.

Gibbard-Satterthwaite Theorem

Extension Lemma

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.

Independence of irrelevant alternatives: If for all $i$, $a \prec_i b$ iff $a \prec_i^{a,b}$, then $f(\prec_1^{a,b}, \ldots, \prec_n^{a,b}) = f(\prec_1^{a}, \ldots, \prec_n^{a})$ must hold, since due to monotonicity the result does not change when $\prec_i^{a,b}$ is successively replaced by $\prec_i^{a,c}$.

Non-dictatorship: Obvious.
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

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.
⇐: Obvious.
⇒: 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 $|A| = |B| + 1$.

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

The results by Arrow and Gibbard-Satterthwaite only apply if there are no restrictions on the preference orders.
Second special case: Let us now consider some special cases such as temperature or volume settings.

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 $\lambda \in [0, 1)$,
$$x \prec_i \lambda x + (1 - \lambda)p_i.$$
**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 \( p_1, \ldots, p_n \) 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

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

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