Mechanism Design and Social Choice

Algorithmic Game Theory
Mechanism Design

Strategic Voting and Social Choice

Impossibility Results
Mechanism Design and Social Choice

- Can selfish agents achieve an agreement on a common outcome?

- How can we set rules that determine an outcome and make sure agents find it in their interest to follow them?

- How can we motivate rational agents to cooperate?
Mechanism Design

- Mechanism Design?

Design a (publicly known) set of rules that interact with selfish agents and implement a common outcome or choice. Mechanism design is sometimes called “Implementation Theory”.

- Mechanism?

A mechanism is an institution (a function, a set of rules) that collects private information from selfish agents and determines an outcome.
Applications

- **Elections** - each voter has preferences, an outcome is the result of the election

- **Markets, E-Commerce** - each participant in a market has preferences and desires, the outcome is an allocation of goods and money.

- **Auctions** - a small market, single seller, each bidder has an amount she is willing to pay, the outcome is the identity of the winner

- **Government Policy** - each citizen has his or her preferences, government must make a single decision

- **Internet Protocols** - each user has load, protocol must make routing assignments
Mechanism Design

Strategic Voting and Social Choice

Impossibility Results
Strategic Voting and the Majority Rule with Two Candidates

Presidential Election: (O)bama, (R)omney

<table>
<thead>
<tr>
<th>Voter</th>
<th>Preference Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O R</td>
</tr>
<tr>
<td>2</td>
<td>R O</td>
</tr>
<tr>
<td>3</td>
<td>O R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Voter</th>
<th>Reported Preference Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O R</td>
</tr>
<tr>
<td>2</td>
<td>O R</td>
</tr>
<tr>
<td>3</td>
<td>O R</td>
</tr>
</tbody>
</table>

Result of Majority Rule: O, R

The Majority rule for two candidates implements many desirable properties:

- Represents the majority of preferences
- Each candidate is in the position he/she appears most often
- Strategic voting is not profitable:
  If a winning voter changes his vote, it only becomes worse off.
  A losing voter cannot change the outcome by changing his vote.
Three Candidates

Presidential Election: (O)bama, (R)omney, (S)tein

<table>
<thead>
<tr>
<th>Voter</th>
<th>Preference Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O S R</td>
</tr>
<tr>
<td>2</td>
<td>R O S</td>
</tr>
<tr>
<td>3</td>
<td>S R O</td>
</tr>
</tbody>
</table>

Majority vote yields a cycle:
2 voters prefer O over S, 2 prefer S over R, 2 prefer R over O ...

This constellation shows that the collective preference can be conflicting (cyclic, not transitive) although each individual preference is well-defined. It is called Condorcet’s Paradox and was discovered by Marquis de Condorcet around 1785.
Definitions

- Set of candidates (or outcomes, alternatives) $A$
- Set of $n$ voters (or players) $N$
- Set of possible preferences (total orders of $A$) is $L$
- Each voter $i$ has a preference (or preference order) $\succ_i \in L$ on the candidates $A$

- A social welfare function is a function $F : L^n \rightarrow L$.
- A social choice function is a function $f : L^n \rightarrow A$.

A social choice function outputs only a single winner, a social welfare function outputs a complete ranking of all candidates.
Properties of Social Welfare Functions

- **Unanimity**: For every $\succ \in L$ we have $F(\succ, \ldots, \succ) = \succ$.

- **Voter $i$ is a dictator** in a social welfare function if for all $\succ_1, \ldots, \succ_n \in L$ we have $F(\succ_1, \ldots, \succ_n) = \succ_i$. Then $F$ is called a dictatorship.

- **Independence of Irrelevant Alternatives (IIA)**: The social preference between any two candidates $a$ and $b$ depends only on the voters’ preferences between $a$ and $b$.

  Formally, for every $a, b \in A$ and every $\succ_1, \ldots, \succ_n, \succ'_1, \ldots, \succ'_n \in L$, let $\succ = F(\succ_1, \ldots, \succ_n)$ and $\succ' = F(\succ'_1, \ldots, \succ'_n)$ then $a \succ_i b \iff a \succ'_i b$ for all $i$ implies $a \succ b \iff a \succ' b$. 
Arrow’s Theorem

**Theorem (Arrow, 1950)**

*Every social welfare function over a set of $|A| \geq 3$ candidates that satisfies unanimity and IIA is a dictatorship.*

For the proof fix $F$ to be a social welfare function that satisfies the conditions.

**Lemma (Pairwise Neutrality)**

*Let $\succ_1, \ldots, \succ_n$ and $\succ'_1, \ldots, \succ'_n$ two preference profiles, and $\succ = F(\succ_1, \ldots, \succ_n)$ and $\succ' = F(\succ'_1, \ldots, \succ'_n)$. If for every player $i$ we have $a \succ_i b \Leftrightarrow c \succ'_i d$, then $a \succ b \Leftrightarrow c \succ' d$.***

**Proof:**

We first rename our elements to let $a \succ b$ and $c \neq b$ (but possibly $a = c$, and/or $b = d$).
Pairwise Neutrality

- Now we adjust \( \succ_i \) and \( \succ'_i \) to become identical w.r.t. \( a, b, c, d \) by moving \( c \) and \( d \) in \( \succ_i \) and \( a \) and \( b \) in \( \succ'_i \):

\[
\begin{align*}
\succ_1 &: \quad \ldots, a, \ldots, b, \ldots \quad \rightarrow \quad \ldots, c, a, \ldots, b, d, \ldots \\
\succ'_1 &: \quad c, \ldots, d, \ldots \quad \rightarrow \quad c, a, \ldots, b, d, \ldots \\
\succ_2 &: \quad \ldots, b, \ldots, a, \ldots \quad \rightarrow \quad \ldots, b, d, \ldots, c, a, \ldots \\
\succ'_2 &: \quad \ldots, d, c, \ldots \quad \rightarrow \quad \ldots, b, d, c, a, \ldots
\end{align*}
\]

and so on

- IIA guarantees that \( a \) and \( b \) remain in the same order in \( \succ \); \( c \) and \( d \) remain in the same order in \( \succ' \). Similarly, by IIA we can now move all other elements and assume \( \succ'_i = \succ_i \).

- By unanimity now \( c \succ a \) and \( b \succ d \), so \( c \succ d \). With \( \succ_i = \succ'_i \) for all \( i \) we also get \( c \succ' d \). \[\square\] (Lemma)
Who is the Dictator?

Pairwise neutrality implies that a social welfare function that satisfies unanimity and IIA has a general underlying approach of determining a global preference. This approach is similar for all preference orders and all pairwise comparisons of elements. This can be used to show that, in fact, the approach boils down to having one dictator determine the output.

Fix $a \neq b$ and $c \neq d$.

- If there are no players with $a \succ_i b$, then $b \succ a$.
- If there are $n$ players with $a \succ_i b$, then $a \succ b$.
- Breakpoint: $i^*$ players

<table>
<thead>
<tr>
<th></th>
<th>...</th>
<th>$i^* - 1$</th>
<th>$i^*$</th>
<th>...</th>
<th>n</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a \succ_i b$</td>
<td>$b \succ_i a$</td>
<td>$b \succ a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a \succ_i b$</td>
<td>$b \succ_i a$</td>
<td>$a \succ b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Claim: $i^*$ is the dictator!
i* is the Dictator

- i* is a dictator if c ≻_i* d ⇒ c ≻ d for all c ≠ d ∈ A.
- Consider an arbitrary set of preferences with c ≻_i* d and e ∈ A with e ≠ c and e ≠ d.
- Switch third element e s.t. it appears as below in ≻_i:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td></td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>i*</td>
<td>c</td>
<td>e</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td>e</td>
</tr>
</tbody>
</table>

- Because of IIA this does not change order of c and d in ≻.
- (c, e) appears exactly as (a, b) previously, by the lemma on pairwise neutrality we know c ≻ e. Similarly e ≻ d.
- Thus c ≻ d, and this proves Arrow’s Theorem. ☐ (Theorem)
Properties of Social Choice Functions

- \( f \) can be strategically manipulated by voter \( i \) if for some \( \succ_1, \ldots, \succ_n \) and some \( \succ'_i \) we have that \( a \succ_i b \) where \( b = f(\succ_1, \ldots, \succ_n) \) and \( a = f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n) \). \( f \) is called incentive compatible (IC) or strategyproof if it cannot be manipulated.

- \( f \) is monotone if \( f(\succ_1, \ldots, \succ_n) = a \neq b = f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n) \) implies that \( a \succ_i b \) and \( b \succ'_i a \).

- Voter \( i \) is a dictator in \( f \) if for all \( \succ_1, \ldots, \succ_n \in L \) we have that if \( a \succ_i b \) for all \( b \neq a \), then \( f(\succ_1, \ldots, \succ_n) = a \). Then \( f \) is called a dictatorship.

- \( f \) is onto \( A \) if for every candidate \( a \in A \) there is a set of preferences such that \( a \) is the winner.
Gibbard-Satterthwaite Theorem

Proposition

A social choice function is IC if and only if it is monotone.

Proof: Direct implication of definitions.

Theorem (Gibbard 1973; Satterthwaite 1975)

A social choice function $f$ onto $A$ with $|A| \geq 3$ is IC if and only if it is a dictatorship.

Proof:
We prove the non-trivial direction of the theorem by using a social choice function $f$ onto $A$ to define a social welfare function $F$ that satisfies IIA and unanimity.
Extending Social Choice Functions

For a preference order $\succ$ and a set $S \subset A$ we denote by $\succ^S$ the adjustment of moving all elements of $S$ in order to the front of $\succ$.

$$S = \{a, b, c\}, \quad A = S \cup \{d, e, f\}$$

\[
\begin{array}{cccccc}
\succ & \rightarrow & \succ^S \\
\hline
a & e & d & c & b & f \\
b & f & e & d & a & c \\
\end{array}
\]

We define $F$ as the social welfare function extending $f$ by $F(\succ_1, \ldots, \succ_n) = \succ$, where $a \succ b$ if and only if $f(\succ_1^{\{a,b\}}, \ldots, \succ_n^{\{a,b\}}) = a$. 
Reaching a Contradiction

**Lemma**

If $f$ is an incentive compatible social choice function onto $A$ then the extension $F$ is a social welfare function.

Show asymmetry and transitivity.

**Lemma**

If $f$ is an incentive compatible social choice function onto $A$ and not a dictatorship, then the extension $F$ satisfies unanimity, independence of irrelevant alternatives, and is not a dictatorship.

A contradiction follows with Arrow’s Theorem
Proof of Theorem

We prove the theorem by verifying the properties of $F$:

- **Antisymmetry**: If $a \succ b$ and $b \succ a$, then $a = b$.
- **Transitivity**: If $a \succ b$ and $b \succ c$, then $a \succ c$.
- **Unanimity**: $F(\succ, \ldots, \succ) = \succ$.
- **IIA**
- **Non-Dictatorship**
Properties

Claim

For any $\succ_1, \ldots, \succ_n$ and any $S$ the winner $f(\succ^S_1, \ldots, \succ^S_n) \in S$.

Proof:

$\triangleright$ $f$ is onto, so there is $\succ''_1, \ldots, \succ''_n$ that gives some $a \in S$ as winner.

$\triangleright$ Iteratively move elements of $S$ to the front, re-sort elements in the back, re-sort elements of $S$ in the front

$\Rightarrow$ Transformation into $\succ^S_1, \ldots, \succ^S_n$.

$\triangleright$ Monotonicity ensures that no $b \notin S$ will ever be a winner in the course of the transformation.

\hfill □  (Claim)
Properties

- Antisymmetry: If \( a \succ b \) and \( b \succ a \), then \( a = b \).

  As \( f(\succ_1^{\{a,b\}}, \ldots, \succ_n^{\{a,b\}}) \in \{a, b\} \).

- Transitivity: If \( a \succ b \) and \( b \succ c \), then \( a \succ c \).

Suppose for contradiction that \( a \succ b \succ c \succ a \). Take \( S = \{a, b, c\} \) and w.l.o.g. let \( f(\succ_1^S, \ldots, \succ_n^S) = a \). Sequential changes to \( \succ^S \) for \( S = \{a, c\} \) imply \( f(\succ_1^S, \ldots, \succ_n^S) = a \), and hence \( a \succ c \). A contradiction follows with antisymmetry.

Hence, if \( f \) is IC onto \( A \), then \( F \) is a valid social welfare function.
Properties

▶ Unanimity: \( F(\succ, \ldots, \succ) = \succ. \)

If \( a \succ_i b \) for all \( i \), then by the claim and monotonicity we have
\[ f(\succ_1\{a,b\}, \ldots, \succ_n\{a,b\}) = a. \]

▶ IIA:

Assume \( a \succ_i b \iff a \succ'_i b \). Note that
\[ f(\succ_1\{a,b\}, \ldots, \succ_n\{a,b\}) = f(\succ'_1\{a,b\}, \ldots, \succ'_n\{a,b\}) \]
because by sequential change of \( \succ_i\{a,b\} \) into \( \succ'_i\{a,b\} \) outcome does not change due to monotonicity and claim.

▶ Non-Dictatorship: Obvious. □ (Theorem)