

Introduction to Social Choice Theory

Mehdi Dastani
BBL-521
M.M.Dastani@uu.nl

What Is Social Choice Theory Trying to Accomplish?

Goal: Given the individual preferences of players, how to aggregate these so as to obtain a social preference.

Agent A_1 :	<i>football</i>	$>_1$	<i>tennis</i>	$>_1$	<i>hockey</i>
Agent A_2 :	<i>tennis</i>	$>_2$	<i>hockey</i>	$>_2$	<i>football</i>
Agent A_3 :	<i>hockey</i>	$>_3$	<i>football</i>	$>_3$	<i>tennis</i>

Social preference: ?

- ▶ Prime example: voting
- ▶ Impossibility results (Arrow, Muller-Satterthwaite)

Artificial Intelligence & Social Choice

Social choice theory has many applications in artificial intelligence:

- ▶ **Search Engines:** ranking documents based on links
- ▶ **Recommendation systems:** recommending an item to a user based on the popularity of the item among similar users.
- ▶ **Belief and Preference aggregation:** selecting most acceptable actions (most acceptable answers)
- ▶ **Algorithms:** developing complex algorithms for voting procedures to avoid manipulation

Voting Mechanisms

3	5	7	6
<i>a</i>	<i>a</i>	<i>b</i>	<i>c</i>
<i>b</i>	<i>c</i>	<i>d</i>	<i>b</i>
<i>c</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>d</i>	<i>d</i>	<i>a</i>	<i>a</i>

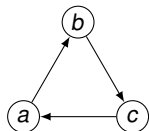
Definition:

- ▶ **Plurality voting:** The candidate which is top ranked by most voters is selected.
- ▶ **Majority voting:** The candidate which is top ranked by majority of voters is selected.
- ▶ **Cumulative voting:** Each voter has k votes which can be cast arbitrarily (e.g., all on one candidate). The candidate with most votes is selected.
- ▶ **Approval voting:** Each voter can cast one single vote for as many of the candidates. The candidate with most votes is selected (voters cannot rank candidates).

Condorcet Paradox

(Weak) Condorcet winner: A candidate is (weak) Condorcet winner if she (ties) beats with every other candidate in a pairwise election.

8	6	7
<hr/>		
a	c	b
b	a	c
c	b	a



The Condorcet Paradox: A Condorcet winner does not always exist.

Condorcet Paradox

		number of players						
		3	5	7	9	11	$\rightarrow \infty$	
number of alternatives	3	.056	.069	.075	.078	.080088
	4	.111	.139	.150	.156	.160176
	5	.160	.200	.215	.230	.251251
	6	.202	.255	.258	.284	.294315
	7	.239	.299	.305	.342	.243369
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots		\vdots
	$\rightarrow \infty$	1	1	1	1	1	...	1

Probability $p(x, y)$ of no Condorcet winner for x alternatives and y players.

(from: Moulin (1988), page 230)

The Borda Rule

Definition (*The Borda Rule*): Given a finite set of alternatives X and strict individual preferences, for each ballot, each alternative is given one point for every other alternative it is ranked below. The alternatives are then ranked proportional to the number of points they aggregate.

1	5	5	3	2
<hr/>				
a	a	c	b	b
d	d	b	a	d
b	c	a	d	c
c	b	d	c	a

8	6	7
<hr/>		
a	c	b
b	a	c
c	b	a

Exercise 1: Check for the left case whether the Borda rule selects the Condorcet winner.

Judgment Aggregation: Doctrinal paradox

- ▶ **p**: The defendant was contractually obliged not to do a particular action.
- ▶ **q**: The defendant did that action.
- ▶ **r**: The defendant is liable for breach of contract.

	p	q	r	Consistent relative to legal doctrine?
Judge 1	true	true	true	yes
Judge 2	false	true	false	yes
Judge 3	true	false	false	yes
Majority voting	true	true	false	no

Voting and Manipulation: Strategy-Proof

Suppose we have the following preferences in the Netherlands political landscape:

45%: $VVD > PvdA > CDA$

25%: $PvdA > CDA > VVD$

15%: $CDA > PvdA > VVD$

15%: $PvdA > VVD > CDA$

Plurality: VVD

Voting and Manipulation: Strategy-Proof

Suppose we have the following preferences in the Netherlands political landscape:

45%: $VVD > PvdA > CDA$

25%: $PvdA > CDA > VVD$

15%: $CDA > PvdA > VVD$ \implies $PvdA > CDA > VVD$

15%: $PvdA > VVD > CDA$

Plurality: $VVD \implies PvdA$

Voting and Manipulation: Strategy-Proof

Suppose we have the following preferences in the Netherlands political landscape:

45%: $VVD \succ PvdA \succ CDA$

25%: $PvdA \succ CDA \succ VVD$

15%: $CDA \succ PvdA \succ VVD \quad \implies \quad PvdA \succ CDA \succ VVD$

15%: $PvdA \succ VVD \succ CDA$

Plurality: $VVD \implies PvdA$

Note: A voting rule is **strategy-proof** if no voter has incentive to misrepresent its true preferences.

Example: A dictatorial voting rule is strategy-proof.

Social Choice Rules: A Formal Setting

Definition: Be $N = \{1, 2, \dots, n\}$ a set of players, O a set of alternatives, and L a class of preference relations over O .

- *Social Choice Function (scf):* $C: L^N \rightarrow O$
- *Social Choice Correspondence (scc):* $C: L^N \rightarrow 2^O$
- *Social Welfare Function (swf):* $W: L^N \rightarrow L$
- *Social Welfare Correspondence (swc):* $W: L^N \rightarrow 2^L$

Social Choice Rules: A Formal Setting

Definition: Be $N = \{1, 2, \dots, n\}$ a set of players, O a set of alternatives, and L a class of preference relations over O .

- *Social Choice Function (scf):* $C: L^N \rightarrow O$
- *Social Choice Correspondence (scc):* $C: L^N \rightarrow 2^O$
- *Social Welfare Function (swf):* $W: L^N \rightarrow L$
- *Social Welfare Correspondence (swc):* $W: L^N \rightarrow 2^L$

Definition (Condorcet Condition): An outcome $o \in O$ is a Condorcet winner if $\forall o' \in O: \#(o > o') \geq \#(o' > o)$.

A social choice function that always choose the Condorcet winner, when there is one, is said to have Condorcet consistency.

Social Choice Rules: A Formal Setting

Definition: Be $N = \{1, 2, \dots, n\}$ a set of players, O a set of alternatives, and L a class of preference relations over O .

- *Social Choice Function (scf):* $C: L^N \rightarrow O$
- *Social Choice Correspondence (scc):* $C: L^N \rightarrow 2^O$
- *Social Welfare Function (swf):* $W: L^N \rightarrow L$
- *Social Welfare Correspondence (swc):* $W: L^N \rightarrow 2^L$

Definition (Smith set): The Smith set is the smallest non-empty set $S \subseteq O$ such that $\forall o \in S, \forall o' \notin S: \#(o > o') \geq \#(o' > o)$, i.e., each member of S beats every other candidate outside S in a pairwise election.

Note:

- ▶ The Smith set exists always.
- ▶ When the Condorcet winner exists, then the Smith set is a singleton consisting of the Condorcet winner.

Exercise 2: Construct an example to show that Condorcet winner is the singleton Smith set.

Borda cannot always select one single winner

Example:

\succsim_1	\succsim_2	\succsim_3
a	b	c
b	c	a
c	a	b

Question: Who is the Borda winner?

Borda cannot always select one single winner

Example:

\succsim_1	\succsim_2	\succsim_3
a	b	c
b	c	a
c	a	b

x	count
a	$2 + 0 + 1 = 3$
b	$1 + 2 + 0 = 3$
c	$0 + 1 + 2 = 3$

Question: Who is the Borda winner?

Remark: The Borda rule does not always provide a social choice function, but a social choice correspondence.

Other Voting Methods

Definition (*Plurality with Elimination*): Each voter casts a single vote for their most-preferred candidate. The candidate with the fewest votes is eliminated. Each voter who cast a vote for the eliminated candidate cast a new vote for the candidate he most prefers among the remaining candidates. This process is eliminated until only one candidate remains.

Definition (*Pairwise Elimination*): Voters are given in advance a schedule for the order in which pairs of candidates will be compared. Given two candidates (and based on each voter's preference ordering) determine the candidate that each voter prefers. The candidate who is preferred by a minority of voters is eliminated, and the next pair of non-eliminated candidates in the schedule is considered. Continue until only one candidate remains.

Sensitivity of Voting Methods

8	6	7
<i>a</i>	<i>c</i>	<i>b</i>
<i>b</i>	<i>a</i>	<i>c</i>
<i>c</i>	<i>b</i>	<i>a</i>

Exercise 3:

- ▶ Sensitivity to a losing candidate: remove *c* and check Majority, Borda, and condorcet methods.
- ▶ Sensitivity to the agenda setter: compare *a.b.c* and *a.c.b* agenda's in pairwise elimination method.

Social Welfare Functions

Definition: Let W be a social welfare function, $o_1, o_2 \in O$, $(\succ_1, \dots, \succ_n) \in L^N$.

Social Welfare Functions

Definition: Let W be a social welfare function, $o_1, o_2 \in O$, $(\succ_1, \dots, \succ_n) \in L^N$.

- ▶ W has the *Pareto property* if

$$o_1 \succ_i o_2 \text{ for all } i \in N \text{ implies } o_1 \succ_{W(\succ_1, \dots, \succ_n)} o_2$$

Intuition: If alternative o_1 is unanimously preferred to alternative o_2 , o_1 should be ranked higher than o_2 in the social ordering.

Social Welfare Functions

Definition: Let W be a social welfare function, $o_1, o_2 \in O$, $(\succ_1, \dots, \succ_n) \in L^N$.

- ▶ W has the *Pareto property* if

$$o_1 \succ_i o_2 \text{ for all } i \in N \text{ implies } o_1 \succ_{W(\succ_1, \dots, \succ_n)} o_2$$

Intuition: If alternative o_1 is unanimously preferred to alternative o_2 , o_1 should be ranked higher than o_2 in the social ordering.

- ▶ W is *dictatorial* if there is some $i \in N$ such that for all preference profiles \succ :

$$o_1 \succ_i o_2 \text{ implies } o_1 \succ_{W(\succ_1, \dots, \succ_n)} o_2$$

Intuition: There is some player whose preferences determine the strict preferences of the social ordering.

Social Welfare Functions

Definition: Let W be a social welfare function, $o_1, o_2 \in O$, $(\succ_1, \dots, \succ_n) \in L^N$.

- ▶ W has the *Pareto property* if

$$o_1 \succ_i o_2 \text{ for all } i \in N \text{ implies } o_1 \succ_{W(\succ_1, \dots, \succ_n)} o_2$$

Intuition: If alternative o_1 is unanimously preferred to alternative o_2 , o_1 should be ranked higher than o_2 in the social ordering.

- ▶ W is *dictatorial* if there is some $i \in N$ such that for all preference profiles \succ :

$$o_1 \succ_i o_2 \text{ implies } o_1 \succ_{W(\succ_1, \dots, \succ_n)} o_2$$

Intuition: There is some player whose preferences determine the strict preferences of the social ordering.

- ▶ A social welfare function has an *unrestricted domain* if it defines a social ordering for *all* preference profiles.

Independence of Irrelevant Alternatives

Definition (Independence of Irrelevant Alternatives): For preference profiles $\succ, \succ' \in L^N$ and alternatives $o_1, o_2 \in O$:

for all $i \in N$ $o_1 \succ_i o_2 \Leftrightarrow o_1 \succ'_i o_2$ implies $o_1 \succ_{W(\succ)} o_2 \Leftrightarrow o_1 \succ_{W(\succ')} o_2$,

Intuition: The social preference of two alternatives only depends on the *relative ordering* of these two alternatives in the individual preference relations.

Independence of Irrelevant Alternatives

Ordering according to Borda scores

	1	5	5	3	2
<i>a</i>	<i>a</i>	<i>a</i>	<i>c</i>	<i>b</i>	<i>b</i>
<i>d</i>	<i>d</i>	<i>d</i>	<i>b</i>	<i>a</i>	<i>d</i>
<i>b</i>	<i>c</i>	<i>c</i>	<i>a</i>	<i>d</i>	<i>c</i>
<i>c</i>	<i>b</i>	<i>b</i>	<i>d</i>	<i>c</i>	<i>a</i>

$a \succ_{\text{Borda}} b$

	1	5	5	3	2
<i>a</i>	<i>a</i>	<i>a</i>	<i>c</i>	<i>b</i>	<i>b</i>
<i>b</i>	<i>c</i>	<i>c</i>	<i>b</i>	<i>a</i>	<i>c</i>
<i>c</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>c</i>	<i>a</i>
<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>

$b \succ_{\text{Borda}} a$

Independence of Irrelevant Alternatives

Definition (*Independence of Irrelevant Alternatives*): For preference profiles $\succ, \succ' \in L^N$ and alternatives $o_1, o_2 \in O$:

$$o_1 \succ_{w(\succ)} o_2 \Leftrightarrow o_1 \succ_{w(\succ')} o_2, \text{ whenever } o_1 \succ_i o_2 \Leftrightarrow o_1 \succ'_i o_2, \text{ for all } i \in N$$

Intuition: The social preference of two alternatives only depends on the *relative ordering* of these two alternatives in the individual preference relations.

Remark: IIA captures a consistency property of social choice rules. Lack of such consistency enables strategic manipulation.

Arrow's Impossibility Theorem

Theorem (Arrow, 1951): *For $|O| \geq 3$, any social welfare function with unrestricted domain satisfying the Pareto property and Independence of Irrelevant Alternatives is dictatorial.*



Arrow's Impossibility Theorem

Theorem (Arrow, 1951): *For $|O| \geq 3$, any social welfare function with unrestricted domain satisfying the Pareto property and Independence of Irrelevant Alternatives is dictatorial.*



Remark: No hope for general social welfare functions, but what about social choice functions?

Social Choice Function

Definition (Weak Pareto efficiency): A social choice function C is weakly Pareto efficient if, for any preference profile $\succ \in L^n$, if there exists a pair of outcomes o_1 and o_2 such that $\forall i \in N : o_1 \succ_i o_2$, then $C(\succ) \neq o_2$.

Intuition: Social choice function does not select any outcome that is dominated by other outcomes for all players.

Definition (Monotonicity): A social choice function C is *monotonic* if for all $o \in O$ and for any preference profile $\succ \in L^N$ with $C(\succ) = o$, then for any other preference profile \succ' with the property that

$$\forall i \in N, \forall o' \in O : \text{if } o \succ_i o' \text{ then } o \succ'_i o', \text{ it must be that } C(\succ') = o$$

Intuition on monotonicity: The winner keeps its winning position when the support for him/her increases.

Social Choice Function

Theorem (Muller-Satterthwaite, 1977): For $|O| \geq 3$, any social choice function C that is weakly Pareto efficient and monotonic is dictatorial.

Consider Plurality scores which is not dictatorial and satisfies weak Pareto efficiency.

3	2	2
<hr/>		
a	b	c
b	c	b
c	a	a

$$C(>_{\text{plurality}}) = a$$

3	2	2
<hr/>		
a	b	b
b	c	a
c	a	c

$$C(>_{\text{plurality}}) = b$$

Social Choice Function and Strategy-Proofness

Let \succ_i^* be the true preference of voter $i \in N$. A voting rule $F : L^N \rightarrow O$ is **strategy-proof** if for any preference profile $\succ = (\succ_1, \dots, \succ_n) \in L^N$ and for voters $j \in N$, we have $F(\succ_1, \dots, \succ_j^*, \dots, \succ_n) \succ_j^* F(\succ)$.

Intuition: A voting rule is **strategy-proof** if no voter has incentive to misrepresent its true preferences.

Theorem (*Gibbard-Satterthwaite Theorem, 1973-1975*): For $|O| \geq 3$, any voting rule $F : L^N \rightarrow O$ that is strategy-proof is dictatorial.

Any hope for voting rules without manipulation?

Restrictions

Definition: An alternative a is a *weak Condorcet winner* if there is no other alternative b for which there are more voters that prefer b to a , i.e., alternative a beats or ties with every other alternative b in a pairwise election.

Restrictions

Definition: An alternative a is a *weak Condorcet winner* if there is no other alternative b for which there are more voters that prefer b to a , i.e., alternative a beats or ties with every other alternative b in a pairwise election.

Theorems: In case of the following restrictions, the existence of a weak Condorcet winner is guaranteed.

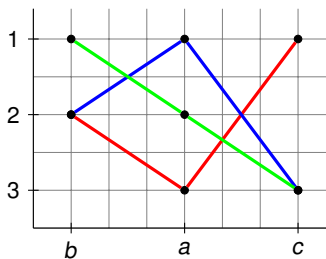
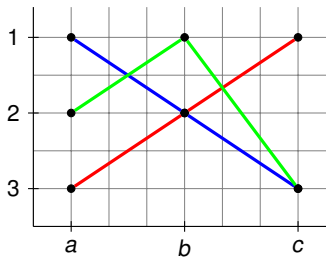
- ▶ Single-peaked preferences (Black, 1948)
- ▶ Dichotomous preferences (Inada, 1964)

Single-peaked Preferences

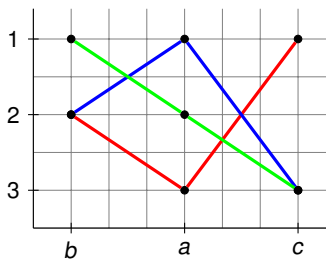
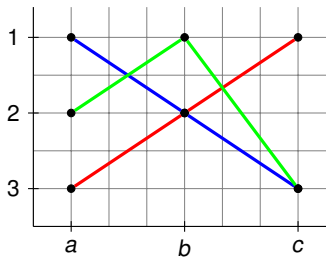
Definition: Given a predetermined linear ordering of the alternative set O , a preference relation \succ_i is *single-peaked* if there exists a point $p \in O$ (its peak) such that for all $o_1, o_2 \in O$ such that $p \geq o_1 > o_2$ or $o_2 > o_1 \geq p$ then $o_1 \succ_i o_2$.

Intuition: Each player prefers points closer to their peaks over points that are more distant.

Single-peaked Preferences



Single-peaked Preferences



\succ_1	\succ_2	\succ_3
c	a	b
b	b	a
a	c	c

Single-peaked Preferences

Definition: Given a predetermined linear ordering of the alternative set O , a preference relation \succ_i is *single-peaked* if there exists a point $p \in O$ (its peak) such that for all $o_1, o_2 \in O$ such that $p \geq o_1 > o_2$ or $o_2 > o_1 \geq p$ then $o_1 \succ_i o_2$.

Intuition: Each player prefers points closer to their peaks over points that are more distant.

Definition: Given a profile of single-peaked preferences \succ_1, \dots, \succ_n with peaks p_1, \dots, p_n (with respect to a predetermined linear ordering on alternatives), the *median voting rule* selects the median among the peaks (the choice of the median voter).

Theorem (Black, 1948): *The median voter rule selects a weak Condorcet winner.*

Dichotomous Preferences and Approval Voting

Preferences \succ_i over O are dichotomous if there is a non-empty set $O' \subseteq O$ such that

- ▶ $o' \succ_i o$, for all $o' \in O'$ and $o \in O \setminus O'$
- ▶ $o'_1 \sim_i o'_2$, for all $o'_1, o'_2 \in O'$
- ▶ $o_1 \sim_i o_2$, for all $o_1, o_2 \in O \setminus O'$



Intuition: Agents with dichotomous preferences only distinguish between “good” and “bad” alternatives

Theorem (Inada, 1964): *If all players have dichotomous preferences, a weak Condorcet winner is guaranteed to exist.*

Remark: Approval voting is a voting rule where each voter can submit zero or one vote to each candidate. The candidates with most votes win.

Exercise 4: Check that the Approval voting rule always selects a weak Condorcet winner.