(Computational) Social Choice

Michal Jakob
Agent Technology Center,
Dept. of Computer Science and Engineering,
FEE, Czech Technical University

AE4M36MAS Autumn 2013 - Lecture 11

Where are We?

Agent architectures (inc. BDI architecture)

Logics for MAS

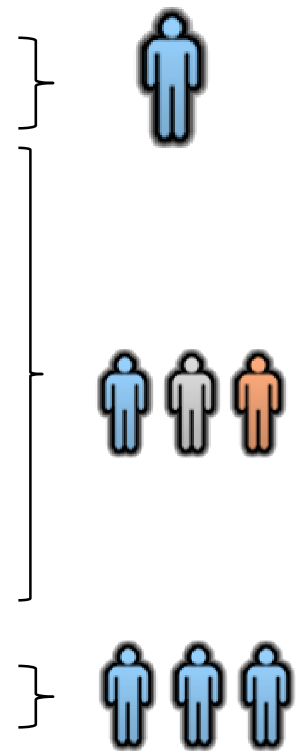
Non-cooperative game theory

Cooperative game theory

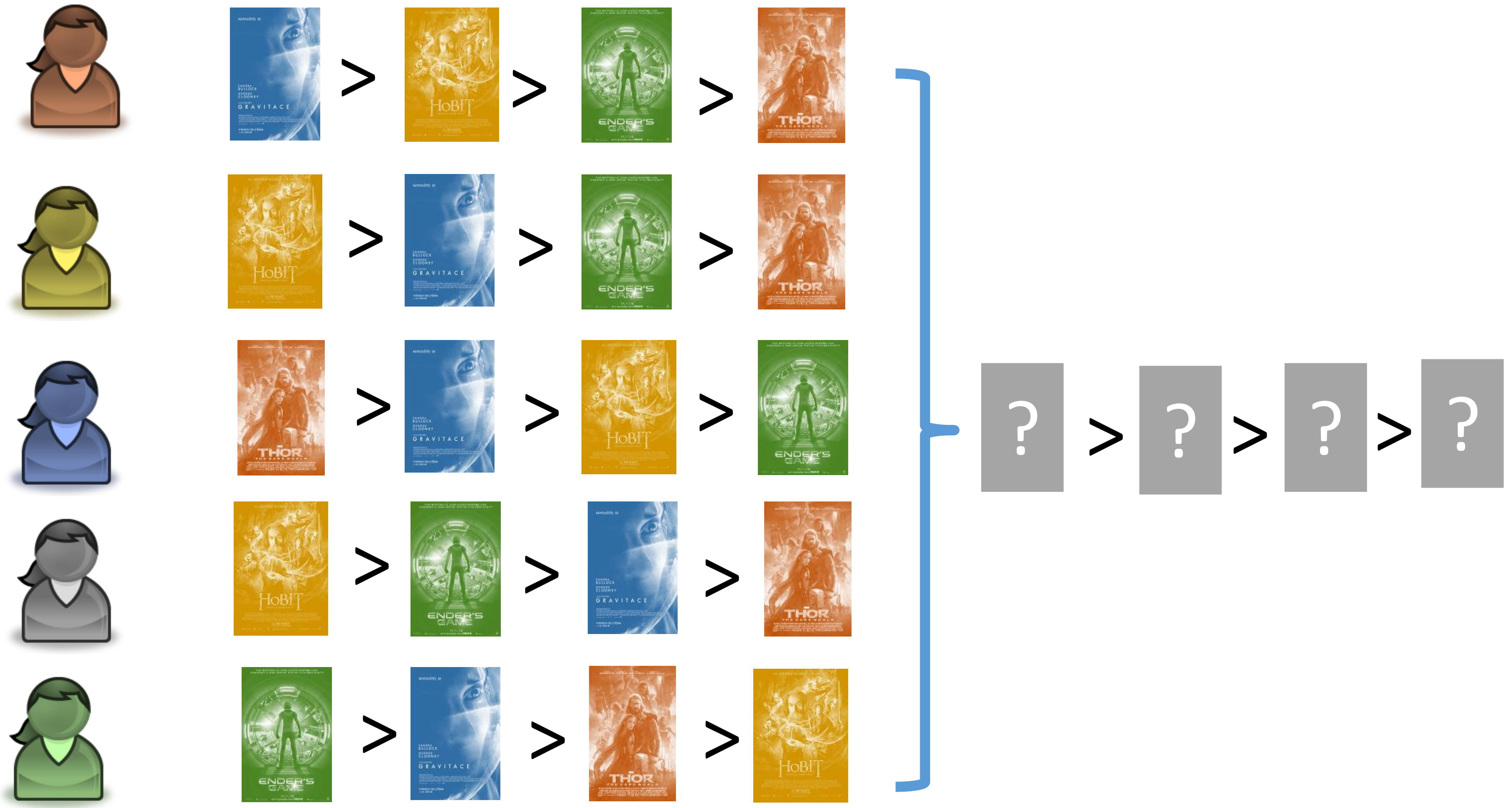
Auctions

Social choice

Distributed constraint reasoning



Motivating Example



Social Choice

Social choice theory *is a theoretical* ***framework*** *for making* ***collective decisions*** *based on the* ***preferences*** *of* ***multiple agents****.*

Key Questions

What does it mean to make **collective rational choices**?

Which **formal properties** should an **social function** satisfy?

Which of these **properties** can be **satisfied simultaneously**?

How **difficult** is it to **compute** collective **choices**?

Can voters **benefit** by **lying** about their **preferences**?

Wide Range of Applications

Elections

Joint plans (MAS)

Recommendation systems

Discussion forums

Meta-search engines

Belief merging

Human computation (crowdsourcing)

Lecture Outline

1. Basic definitions
2. Voting rules
3. Theoretical properties
4. Manipulation
5. Summary

Basic Definitions

Social Choice

Social Welfare Function

Consider

- a finite set $N = \{1, \ldots, n\}$ of at least two **agents** (sometimes called **individuals** or **voters**) and
- a finite universe U of at least two **alternatives** (sometimes called **candidates**).
- Each agent i entertains **preferences** over the alternatives in U, which are represented by a *transitive* and *complete* **preference relation** $\succcurlyeq_i$.
- The set of **all preference relations** over the universal set of alternatives U is denoted as $\mathcal{R}(U)$.
- The set of **preference profiles**, associating one preference relation with each individual agents is then given by $\mathcal{R}(U)^n$.

Definition: Social Welfare Function

A **social welfare function** (SWF) is a function $f: \mathcal{R}(U)^n \rightarrow \mathcal{R}(U)$

*A social welfare function **maps individual preference** relations to a **collective preference** relation.*

Social Choice Function

Consider

- the set of **possible feasible sets** $\mathcal{F}(U)$ defined as the set of all *non-empty* subsets of U
- a **feasible set** $A \in \mathcal{F}(U)$ (or **agenda**) defines the set of possible alternatives in a specific choice situation at hand.

Definition: Social Choice Function

A **social choice function** (SCF) is a function $f: \mathcal{R}(U)^n \times \mathcal{F}(U) \rightarrow \mathcal{R}(U)$ such that $f(R, A) \subseteq A$ for all R and A.

A social choice function ***maps individual preferences*** *and a* ***feasible*** *subset of the* ***alternatives*** *to a set of* ***socially preferred alternatives, the choice set.***

Voting Rule

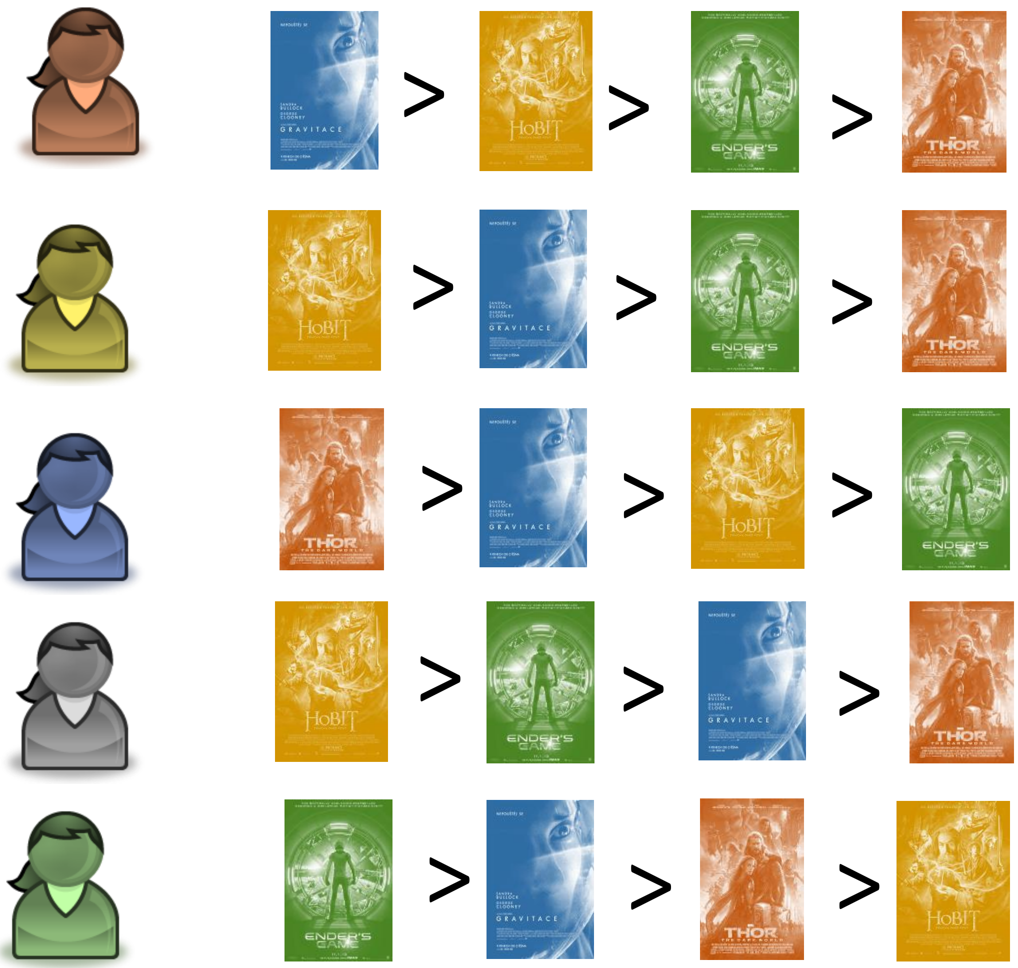
Definition: Voting Rule

A **voting rule** is a function $f: \mathcal{R}(U)^n \to \mathcal{F}(U)$.

A voting rule is **resolute** if $|f(R)| = 1$ for all preference profiles R.

Voting rules are a special case of social choice functions.

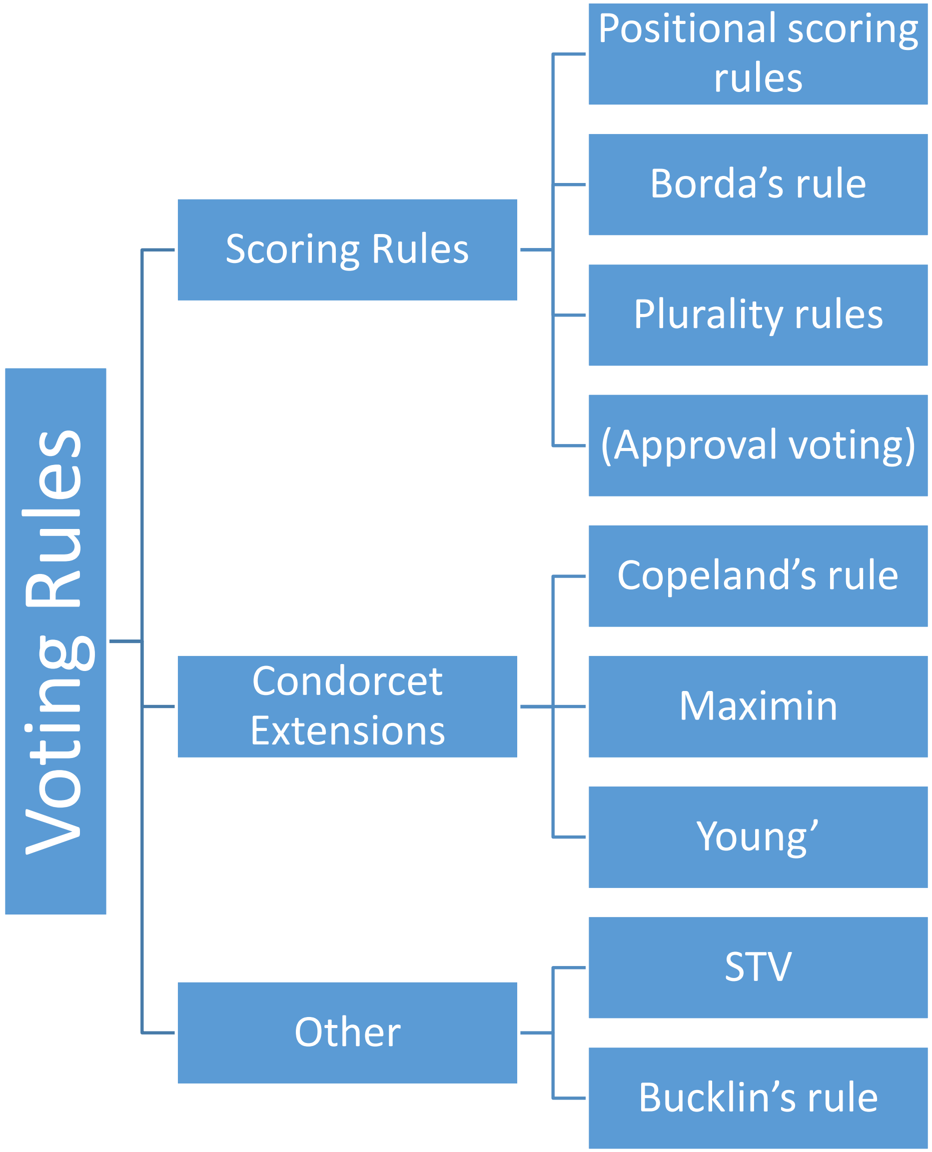
Illustration



Voting Rules

Social choice

Voting Rules



Scoring Rules

Positional scoring rules:

- assuming m alternatives, we define a **score vector** $\boldsymbol{s} = (s_1, \ldots, s_m) \in \Re^m$ such that $s_1 \geq \cdots \geq s_m$ and $s_1 > s_m$
- each time an alternative is **ranked** ith by some voter, it gets a particular **score** s_i
- the scores of each alternative are added and the alternatives with the **highest cumulative score** is selected.

Widely used in practice due to their simplicity.

Scoring Rules: Examples

Borda's rule: alternative a get k points from voter i if i prefers a to k other alternatives, i.e., the score vector is $\boldsymbol{s} = (|U| - 1, |U| - 2, \ldots, 0)$.

- chooses those alternatives with the **highest average rank** in individual rankings

Plurality rules: the score vectors is $\boldsymbol{s} = (1,0, \ldots, 0)$,i.e., the cumulative score of an alternative equals the number of voters by which it is ranked first.

- Veto / Anti-plurality rule: $\boldsymbol{s} = (1,1, \ldots, 0)$

Approval voting: every voter can approve any number of alternatives and the alternatives with the highest number of approvals win.

- not technically a rule

Condorcet Extension

An alternative a is a **Condorcet winner** if, when compared with every other candidate, is **preferred by more voters**.

- Condorcet winner is **unique** but does **not** always **exist**

Condorcet extension: a voting rule that selects Condorcet winner whenever it exists.

- **Copeland's rule**: an alternative gets a point for every pairwise majority win, and some fixed number of points between 0 and 1 (say, 1/2) for every pairwise tie. The winners are the alternatives with the greatest number of points.
- **Maximin rule**: evaluate every alternative by its worst pairwise defeat by another alternative; the winners are those who lose by the lowest margin in their worst pairwise defeats. (If there are any alternatives that have no pairwise defeats, then they win.)
- ...

Other Rules

Single transferable vote: looks for the alternatives that are ranked in first place the least often, removes them from all voters' ballots, and repeats. The alternatives removed in the last round win.

Condorcet's Paradox

agent 1: $A \succ B \succ C$
agent 2: $C \succ A \succ B$
agent 3: $B \succ C \succ A$

For every possible candidate, there is another candidate that is **preferred** by a $\frac{2}{3}$ **majority** of voters!

There are scenarios in which no matter which outcome we choose the **majority** of **voters** will be **unhappy** with the alternative chosen

Issue: Dependency on the Voting Rule

499 agents: $A \succ B \succ C$
3 agents: $B \succ C \succ A$
498 agents: $C \succ B \succ A$

What is the Condorcet winner?

B

What would win under plurality voting?

A

What would win under STV?

C

Issue: Sensitivity to Losing Candidate

35 agents: $A \succ C \succ B$
33 agents: $B \succ A \succ C$
32 agents: $C \succ B \succ A$

What candidate wins under **plurality** voting?

A

What candidate wins under **Borda** voting?

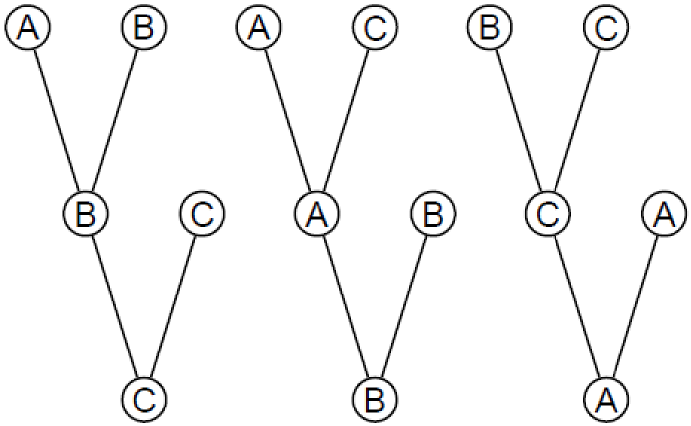
A

Now consider dropping C. Now what happens under both Borda and plurality?

B wins

Sensitivity to Agenda Setter

35 agents: $A \succ C \succ B$
33 agents: $B \succ A \succ C$
32 agents: $C \succ B \succ A$



Who wins **pairwise elimination**, with the ordering A, B, C?

C

Who wins with the ordering A, C, B?

B

Who wins with the ordering B, C, A?

A

Another Pairwise Elimination Problem

1 agent: $B \succ D \succ C \succ A$
1 agent: $\mathrm{A} \succ B \succ D \succ C$
1 agent: $C \succ A \succ B \succ D$

Who wins under pairwise elimination with the ordering A, B, C, D?

- D

What is the problem with this?

- all of the agents prefer B to D – the selected candidate is **Pareto-dominated!**

Theoretical Properties

Social Choice

Recapitulation

Definition: Social Welfare Function

A **social welfare function** (SWF) is a function $f: \mathcal{R}(U)^n \to \mathcal{R}(U)$

Definition: Social Choice Function

A **social choice function** (SCF) is a function $f: \mathcal{R}(U)^n \times \mathcal{F}(U) \to \mathcal{R}(U)$ such that $f(R, A) \subseteq A$ for all R and A.

Definition: Voting Rule

A **voting rule** is a function $f: \mathcal{R}(U)^n \to \mathcal{F}(U)$.

Pareto Efficiency

Definition: Pareto optimality (also Pareto efficiency)
A social welfare function f is **Pareto optimal** if a $\succ_i b$ for all $i \in N$ implies that $a \succ_f b$.

i.e. when all agents agree on the *strict* ordering of two alternatives, this ordering is respected in the resulting social preference relation.

Independence of Irrelevant Alternatives (IIA)

Definition: Independence of Irrelevant Alternatives (IIA)

Let R and R' be two preference profiles and a and b be two alternatives such that $R|_{\{a,b\}} = R'|_{\{a,b\}}$, i.e., the **pairwise comparisons** between a and b are **identical** in both profiles. Then, IIA requires that a and b are also **ranked identically** in $\succcurlyeq$, i.e., $\succcurlyeq_f\big|_{\{a,b\}} = \succcurlyeq'_f\big|_{\{a,b\}}$.

i.e. the social preference ordering between two alternatives depends only on the **relative orderings** they are given by the agents

IIA Example

In a Borda count election, 5 voters rank 5 alternatives [*A*, *B*, *C*, *D*, *E*]: 3 voters rank [*A*>*B*>*C*>*D*>*E*]. 1 voter ranks [*C*>*D*>*E*>*B*>*A*]. 1 voter ranks [*E*>*C*>*D*>*B*>*A*].

- Borda count: *C*=13, *A*=12, *B*=11, *D*=8, *E*=6 ➔ ***C* wins.**

Now, the voter who ranks [*C*>*D*>*E*>*B*>*A*] instead ranks [*C*>*B*>*E*>*D*>*A*]; and the voter who ranks [*E*>*C*>*D*>*B*>*A*] instead ranks [*E*>*C*>*B*>*D*>*A*]. Note that they change their preferences only over the pairs [*B*, *D*], [*B*, *E*] and [*D*, *E*].

- The new Borda count: *B*=14, *C*=13, *A*=12, *E*=6, *D*=5 ➔ ***B* wins**.

***B* now wins instead of *C*, even though no voter changed their preference over [*B*, *C*]** ➔ Borda count violates IIA

Non-dictatorship

Definition: Non-dictatorship

An SWF f is **non-dictatorial** if there is **no** agent i such that for all preference profiles R and alternatives $a, b, a \succ_i b$ implies $a \succ_f b$. We say f is **dictatorial** if it fails to satisfy this property.

i.e. there is no agent who can **dictate** a strict ranking no matter which preferences the other agents have.

Properties Summary

	Pareto optimal	Condorcet consistent	IIA	Non-dictatorship
Plurality	yes	no	no	yes
Borda	yes	no	no	yes
Sequential majority	no	yes	no	yes

Why?

Arrow’s Theorem

Theorem (Arrow, 1951)

There **exists no** social welfare function that **simultaneously satisfies** IIA, Pareto optimality, and non-dictatorship whenever $|U| \geq 3$.

Negative result: At the required **conditions** has to be **omitted** or **relaxed** in order obtain a **positive** result.

If $|U| = 2$, IIA is trivially satisfied by any SWF and reasonable SWFs (e.g. the majority rule) also satisfy remaining conditions.

Would it help if we focus on social choice functions instead?

Properties of Social Choice Functions

Reformulation of SWF properties for SCFs:

- **Pareto optimality**: $a \notin f(R, A)$ if there exists some $b \in A$ such that $b \succ_i a$ for all $i \in N$
- **Non-dictatorship**: an SCF f is non-dictatorial iff there is no agent i such that for all preference profiles R and alternatives a, $a \succ_i b$ for all $b \in A \setminus \{a\}$ implies $a \in f(R, A)$.
- **Independence of irrelevant alternatives**: an SCF satisfies IIA iff $f(R, A) = f(R', A)$ if $R|_A = R'|_A$

Definition: Weak axiom of revealed preferences (WARP)

An SCF f satisfies WARP iff for all feasible sets A and B and preference profiles R:

if $B \subseteq A$ and $f(R, A) \cap B \neq \emptyset$ then $f(R, A) \cap B = f(R, B)$.

Arrow's theorem for SCFs

Theorem (Arrow, 1951, 1959)

There **exists no** social choice function that **simultaneously satisfies** IIA, Pareto optimality, non-dictatorship, and WARP whenever $|U| \geq 3$.

Negative result: At the required **conditions** has to be **omitted** or **relaxed** in order obtain a **positive** result.

The only conditions that can be reasonably relaxed is WARP → **contraction consistency** and expansion consistency.

There are a number of appealing SCFs that satisfy all conditions if only expansion consistency is required.

Manipulation

Social Choice

Strategic Manipulation

So far, we assumed that the **true preferences** of all voters are **known**.

This is an **unrealistic assumption** because voters may be better off by **misrepresenting** their **preferences**.

Plurality winner a

- b wins if the last two voters vote for b, whom they prefer to a.

How about Borda?

- a's score: 9, b's score: 14, c's score: 13, d's score: 6
- c wins if the voters in the second column, who prefer c to b, move b to the bottom.

1	2	2	2
a	a	b	c
b	c	d	b
c	b	c	d
d	d	a	a

Manipulable Rule

Definition: Mainupulable rule

A resolute voting rule f is **manipulable** by voter i if there exist preference profiles R and R' such that $R_j = R'_j$ for all $j \neq i$ and $f(R') \succ_i f(R)$. A voting rule is **strategyproof** if it is not manipulable.

Note: we assume voters know preferences of all other voters.

Why is Manipulation Undesirable

Inefficient: Energy and resources are wasted on manipulative activities.

Unfair: Manipulative skills are not spread evenly across the population.

Erratic: Predictions or theoretical statements about election outcomes become extremely difficult.

- $\Leftarrow$ voting games can have many different equilibria

*Are there any voting methods which are **non-manipulable**, in the sense that voters can **never benefit** from **misrepresenting** preferences?*

The Gibbard-Satterthwaite Impossibility

A voting rule is **non-imposing** if its image contains all singletons of $\mathcal{F}(U)$, i.e., every single alternative is returned for some preference profile.

- technical condition weaker than Pareto optimality

Theorem (Gibbard, 1973; Satterthwaite, 1975)

Every **non-imposing**, **strategyproof**, **resolute** voting rule is **dictatorial** when $|U| \geq 3$.

Possible workarounds:

- **restricted domains**, e.g., **single-peaked** preferences
- computational **hardness of manipulation**

Computational Hardness of Manipulation

Gibbard-Satterthwaite tells us that manipulation is **possible in principle** but does not give any indication of how to misrepresent preferences.

There are voting rules that are **prone to manipulation** in principle, but where manipulation is **computationally complex**.

- E.g. Single Transferable Vote rule is NP-hard to manipulate!

Problem: NP-hardness is a **worst-case** measure.

Recent **negative result** (Isaksson et al., 2010): Essentially, for every efficiently computable, neutral voting rule, a manipulable preference profile with a corresponding manipulation can easily be found.

Summary

Social Choice

Other Topics

Combinatorial domains: preferences over combinations of base items.

→ compact preference representation languages

Fair division

- alternatives are allocations of goods to agents
- preferences are assumed to be valuation function (→ "social choice with money")

Other models: **matching**, **reputation** systems

Issues: preference elicitation, communication, ...

Conclusions

Aggregating preferences is a (surprisingly) complex problem.

All desirable properties cannot be fulfilled at once → trade-offs.

No single best social function exists

- Weight pros and cons for each particular application

Reading: F. Brandt, V. Conitzer, and U. Endriss. Computational Social Choice. In G. Weiss (ed.), *Multiagent Systems*, MIT Press, 2013; [Shoham] – 9.1 – 9.4