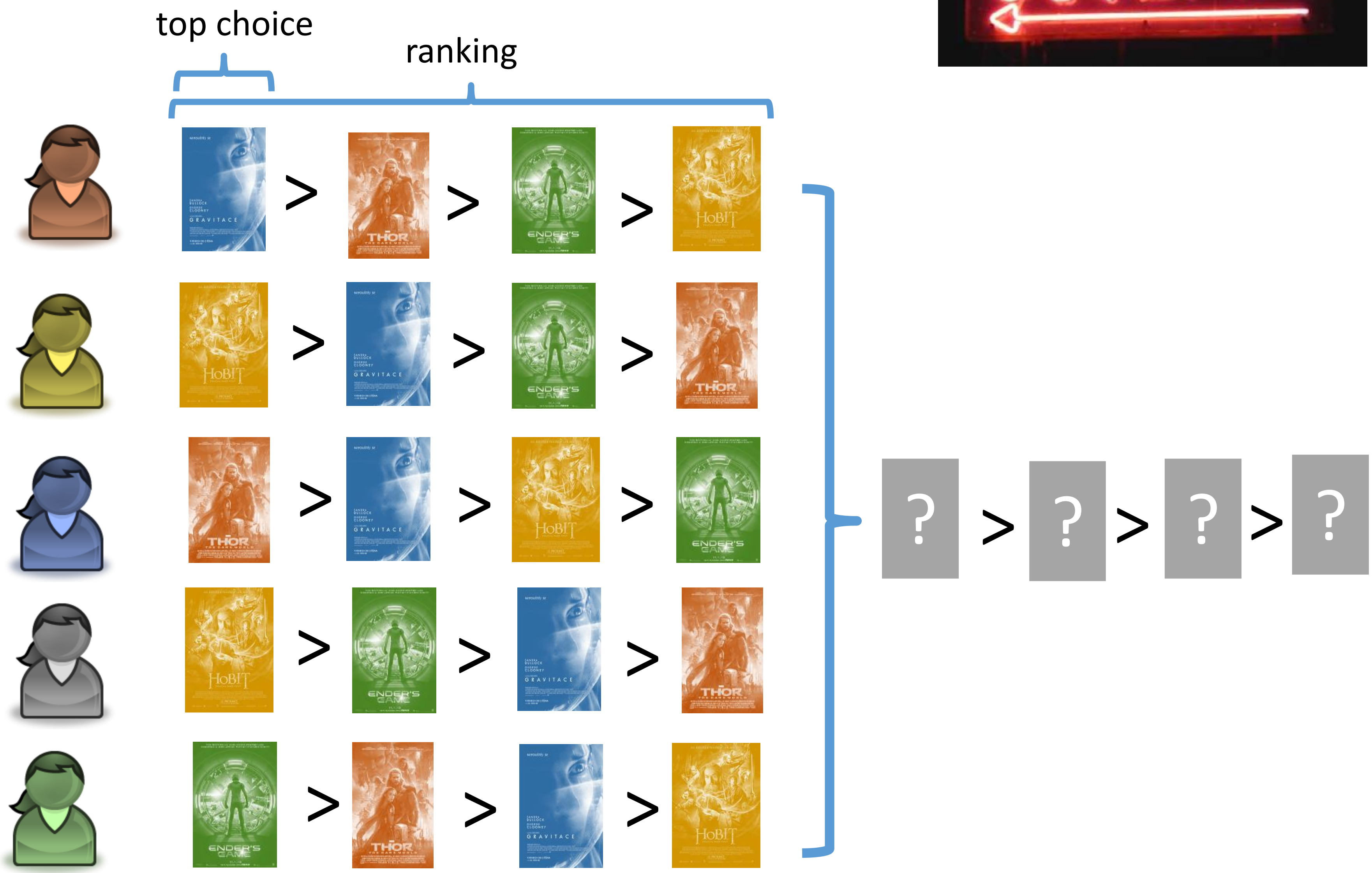
(Computational) Social Choice

Michal Jakob
Artificial Intelligence Center,
Dept. of Computer Science and Engineering,
FEE, Czech Technical University

AE4M36MAS Autumn 2018 - Lecture 11

Motivating Example



Social Choice

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

(does not consider payments: settings with payments → auctions)

Wide Range of Applications

Elections

Joint plans (MAS)

Resource allocation

Recommendation and reputation systems

Human computation (crowdsourcing)

Webpage ranking and meta-search engines

Discussion forums

Key Questions

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

Which **formal properties** should such choices 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**?

Basic Definitions

Social Choice

Social Welfare Function

Consider

- a finite set $N = \{1, \dots, n\}$ of at least two **agents** (also called **individuals** or **voters**), and
- a finite universe U of at least two **alternatives** (also called **candidates**).
- Each agent i has **preferences** over the alternatives in U, which are represented by a *transitive* and *complete* **preference relation** $\succcurlyeq_i$ (*likes at least as much*).
- The set of **all preference relations** (also called **rankings**) 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 agent is then given by $\mathcal{R}(U)^n$.

Definition: Social Welfare Function

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

*A social welfare function **maps individual preference** relations to a **collective preference** relation (also called **social ranking**)*

Social Welfare Function: Remarks

Transitivity: $a \succcurlyeq_i b \succcurlyeq_i c$ implies $a \succcurlyeq_i c$.

Completeness: For any pair of alternatives $a, b \in N$ either $a \succcurlyeq_i b$ or $a \preccurlyeq_i b$ or both

Antisymmetry in general **not** assumed / required.

if $a \succcurlyeq_i b$ and $a \preccurlyeq_i b$ then we say $a \sim_i b$ are **indifferent**.

Cardinal Voting Systems

Cardinal voting refers to social choice mechanisms which allows the voter to give each candidate an independent rating or grade.

- only practical when there is a common numeraire such as money

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 $f: \mathcal{R}(U)^n \times \mathcal{F}(U) \rightarrow \mathcal{F}(U)$ 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

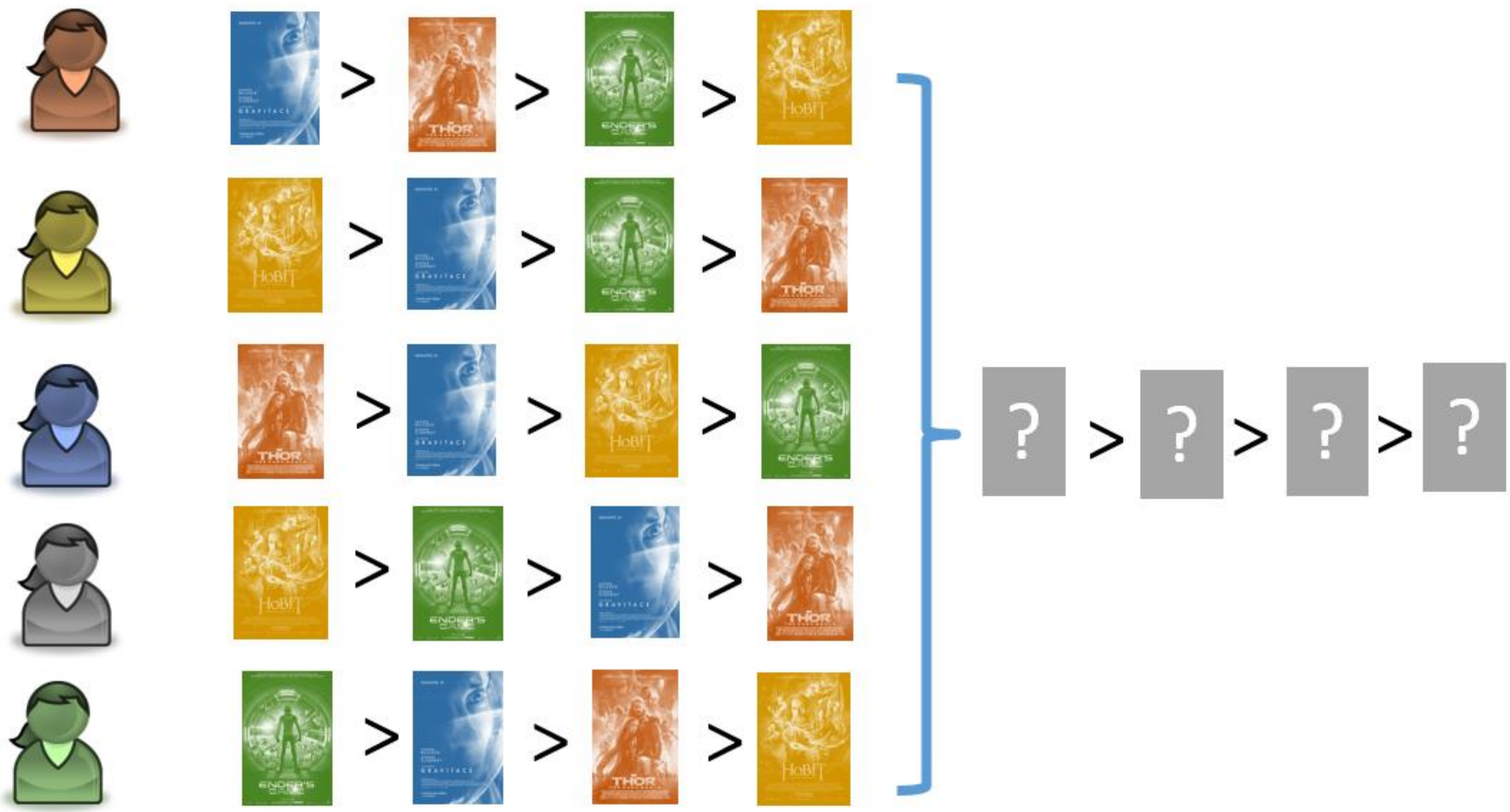
Definition: Voting Rule

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

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



SWFs and Voting Rules

Social choice

Kemeny's Rule

Kemeny's rule returns

$$\text{argmax}_{\succ} \sum_{i \in N} | \succ \cap \succ_i |$$

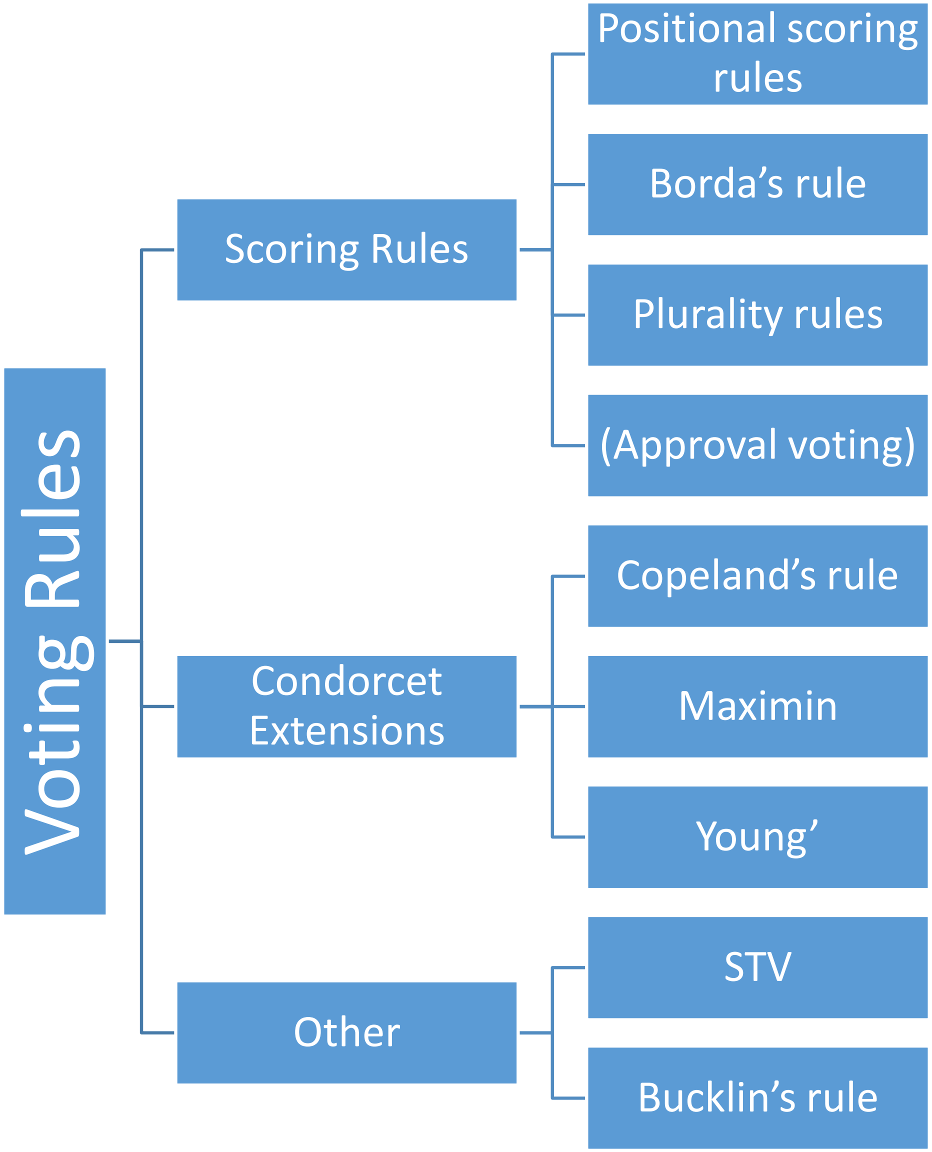
i.e. all strict rankings that **agree** with as many pairwise preferences as possible.

- there might be more than one so technically not an SWF but multi-valued SWF

Maximum likelihood interpretation: agents provide noisy estimates of a "correct" ranking.

Computation is **NP-hard**, even when there are just four voters.

Voting Rules



Positional Scoring Rules

Assuming m alternatives, we define a **score vector** $\boldsymbol{s} = (s_1, \dots, s_m) \in \Re^m$ 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.

Positional scoring rules are 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 $\boldsymbol{s} = (|U| - 1, |U| - 2, \dots, 0)$.

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

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

- Veto (anti-plurality) rule: $\boldsymbol{s} = (1,1, \dots, 0)$

Approval voting *procedure*[1]: every voter can approve any number of alternatives and the alternatives with the highest number of approvals win.

- e.g. likes on social networks

[1]not technically a voting rule

Condorcet Extensions

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

- Is it **unique**? ✓ **Yes!**
- Does it always **exist**? ✘ **No!**

→ **Condorcet extensions**: Voting rules that selects Condorcet winner whenever it exists.

- **Copeland's rule**: an alternative gets **+1** point for reach pairwise victory and **-1** point for each pairwise defeat. 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.)

Scoring rules are *not* Condorcet extensions!

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: $B \succ C \succ A$

Condorcet winner?

✘ No!

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

Collective preferences can be **cyclic**, even if the preferences of **individual** voters are **not cyclic**.

- the majority of voters agree that A is preferable to B, B to C, and C to A!

There are elections 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 would win under plurality voting?

A

What is the Condorcet winner?

B

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 (gets 35x2+33x1=103 points)

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

B wins (B gets 65 points, A only 35 points)

Theoretical Properties

Social Choice

Recap: Definitions

Definition: Social Welfare Function

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

Definition: Social Choice Function

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

Definition: Voting Rule

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

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 the resulting social ranking $\succcurlyeq$, 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: Plurality vote

In a plurality voting system 7 voters rank 3 alternatives (*A*, *B*, *C*).

- 3 voters rank $A \succ B \succ C$
- 2 voters rank $B \succ A \succ C$
- 2 voters rank $C \succ B \succ A$

Initially only *A* and *B* run in an election: *B* wins with 4 votes to A's 3.

But the entry of *C* into the race makes *A* the new winner.

➔ The relative positions of *A* and *B* are reversed by the introduction of *C*, an "irrelevant" alternative
➔ *plurality voting violates IIA.*

IIA Example: Borda Count

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.

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 least one of the desired **properties** 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 least one of the desired **properties** 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 (3 top votes)
- <u>but:</u> b wins if the last voter votes for b, whom it prefers to a (b gets 4 top votes then).

How about Borda?

- winner b (b's score: 14, c's: 13)
- <u>but:</u> c wins if the voters in the second column, who prefer c to b, move b to the bottom (b's score drops to 12).

Example

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

Manipulable Rule

Definition: Manipulable 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 **strategy-proof** 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**, **strategy-proof**, **resolute** voting rule is **dictatorial** when $|U| \geq 3$.

In other words, every "realistic" voting method is prey to strategic manipulation

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!

Summary

Social Choice

Other Topics

Combinatorial domains: preferences over combinations of base items.

→ compact preference representation languages

Fair division

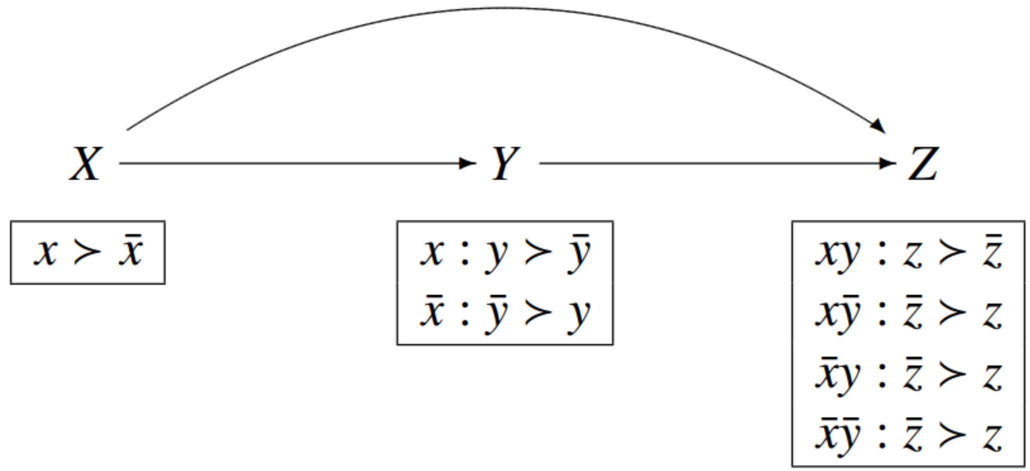
- alternatives are allocations of goods to agents
- preferences are assumed to be valuation function (→ "social choice with money")

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

Issues: preference elicitation, communication, ...

Conditional Preference Networks (CP-nets)

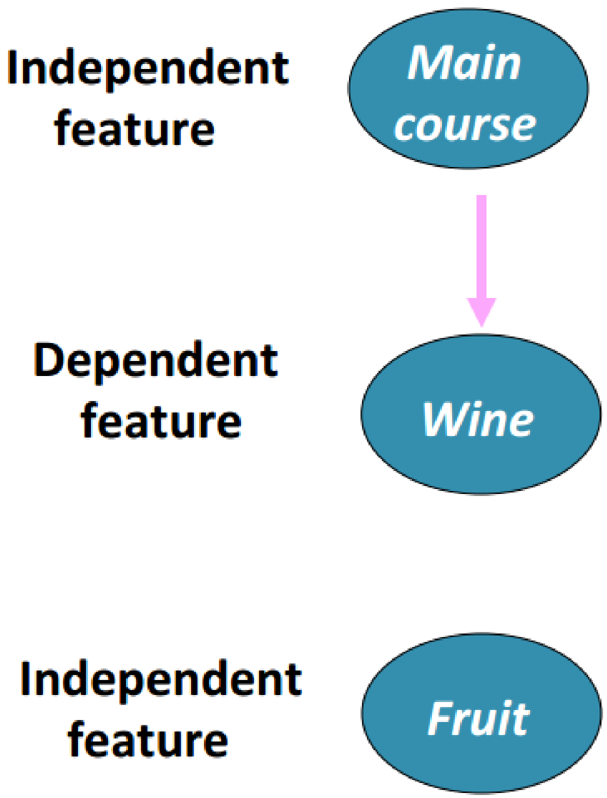
(Possibly) succinct way of representing complex preference relationships



Full partial order is the **transitive closure** of individual preference statements

$$\begin{tikzcd}
 & & & & \bar{x}y\bar{z} \arrow[dr] & \\
xyz \arrow[r] & xy\bar{z} \arrow[r] & x\bar{y}\bar{z} \arrow[r] \arrow[dr] & \bar{x}\bar{y}\bar{z} \arrow[r] \arrow[ur] & \bar{x}\bar{y}z \arrow[r] & \bar{x}yz \\
 & & & x\bar{y}z \arrow[ur] & &
\end{tikzcd}$$

CP-Net Example



fish>meat

Conditional Preference Table

Main course	*Wine*
fish	white > red
meat	red > white

peaches > strawberries

Conclusions

Aggregating preferences is a (surprisingly) complex problem.

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

No perfect 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