# Aggregating Preferences

CMPUT 366: Intelligent Systems

S&LB §9.1-9.4, §10.1-10.3

# Lecture Outline

1. Logistics & Recap
2. Voting Schemes
3. Mechanism Design

# Logistics

- Labs & Assignment #4
  - Assignment #4 is due **Apr 12 (this Friday)** by midnight
  - Today's lab is from **5:00pm to 7:50pm** in **CAB 235**
    - Not mandatory
    - Opportunity to get help from the TAs
- **USRI surveys** are live until **Apr 10 (Wednesday)** at midnight
  - **Question:** Should we spend some lecture time on this?

# Recap: Zero-Sum Games

- **Maxmin** strategies maximize an agent's worst-case payoff
- **Nash equilibrium** strategies are **different** from maxmin strategies in general games
- In **zero-sum games**, they are the same thing
  - It is always **safe** to play an equilibrium strategy in a zero-sum game
  - **Alpha-beta search** computes equilibrium of zero-sum games more efficiently than backward induction

# Aggregating Preferences

- Suppose we have a collection of agents, each with **individual preferences** over some **outcomes**
    - Ignore strategic reporting issues: Either the center already knows everyone's preferences, or the agents don't lie
- **Question:** How should we choose the outcome?
    - **Informally:** What is the right way to turn a collection of individual preferences into the group's preferences?
    - **More formally:** Can we construct a **social choice function** that maps the profile of preference orderings to an outcome?

# Formal Model

**Definition:** A **social choice function** is a function $C : L^n \rightarrow O$, where

- $N=\{1,2,..,n\}$ is a set of **agents**
- $O$ is a finite set of **outcomes**
- $L$ is the set of strict total orderings over $O$.

**Notation:**
We will denote ***i*'s preference order** as $>_i \in L$

# Two Voting Schemes

1. **Plurality voting**

   - Everyone votes for favourite outcome, choose the outcome with the **most votes**
   - Voters **need not** submit a full preference ordering

2. **Borda score**

   - Everyone assigns **scores** to each outcome:
     Most-preferred gets $n$-1, next-most-preferred gets $n$-2, etc.
     Least-preferred outcome gets 0.
   - Outcome with **highest sum of scores** is chosen
   - This amounts to submitting a **full preference order**

# Paradox:
# Sensitivity to Losing Candidate

35 agents: $a \succ c \succ b$
33 agents: $b \succ a \succ c$
32 agents: $c \succ b \succ a$

- **Question:** Who wins under **plurality**?
  - **Question:** Now **drop *c***. Who wins under **plurality**?

35 agents prefer a > b
65 agents: b > a

- **Question:** Who wins under **Borda**?

a: 2*35 + 1*33 = 103
b: 2*33 + 1*32 = 98
c: 2*32 + 1*35 = 99

  - **Question:** After **dropping *c***, who wins under **Borda**?

# Arrow's Theorem

These problems are not a coincidence; they affect **every possible** voting scheme.

# Pareto Efficiency

**Definition:**
*W* is **Pareto efficient** if for any $o_1, o_2 \in O$, if **everyone agrees** that $o_1$ is better than $o_2$, then the aggregated order W should also prefer $o_1$ over $o_2$.

Formally: $(\forall i \in N : o_1 \succ o_2) \implies (o_1 \succ_W o_2)$

# Independence of Irrelevant Alternatives

**Definition:**
*W* is **independent of irrelevant alternatives** if the preference between any two alternatives $o_1, o_2 \in O$ depends only on the agents' preferences between $o_1$ and $o_2$.

- "Spoiler" candidates shouldn't matter

Formally:

$$(\forall i \in N : o_1 \succ_i' o_2 \iff o_1 \succ_i'' o_2) \implies (o_1 \succ_{W[\succ']} o_2 \iff o_1 \succ_{W[\succ'']} o_2)$$

# Non-Dictatorship

**Definition:**
*W* does not have a **dictator** if no single agent determines the social ordering.

Formally:

$$\neg i \in N : \forall [\succ] \in L^n : \forall o_1, o_2 \in O : (o_1 \succ_i o_2) \implies (o_1 \succ_W o_2)$$

# Arrow's Theorem

**Theorem:** (Arrow, 1951)
If $|O| > 2$, any social welfare function that is Pareto efficient and independent of irrelevant alternatives is dictatorial.

# Mechanism Design

- In social choice, we assume that agents' preferences are **known**
- We now allow agents to **report** their preferences **strategically**
- Which social choice functions are **implementable** in this new setting?

**Differences:**

1. Social choice function is **fixed**
2. Agents **report** preferences

# Mechanism

**Definition:**
In a setting with **agents** $N$ who have preferences over **outcomes** $O$, a **mechanism** is a pair $(A,M)$, where:

- $A = A_1 \times \ldots \times A_n$, where $A_i$ is a set of **actions** made available to the agent
- $M : A \to \Delta(O)$ maps each action profile to a distribution over outcomes

# Example Mechanism: First Price Auction

- Every agent has **value** $v_i \in \mathbb{R}$ for some object
- *Social choice function:* Give the object to the agent who values the object most
- **Question:** Can we just ask the agents how much they like it?
- *Actions:* Agents declare a value simultaneously
- *Outcomes:* Highest bidder wins, and **pays their bid**
- **Question:** Do the agents have an incentive to **tell the truth**?

# Example Mechanism: Second Price Auction

- Every agent has **value** $v_i \in \mathbb{R}$ for some object
- *Social choice function:* Give the object to the agent who values the object most
- *Actions:* Agents declare a value simultaneously
- *Outcomes:* Highest bidder wins, and pays the bid of the **next-highest bidder**
- **Question:** Do the agents have an incentive to **tell the truth**?

# Dominant Strategy Implementation

**Definition:**
A mechanism $(A,M)$ is an **implementation in dominant strategies** of a social choice function $C$ (over $N$ and $O$) if for any vector $u$ of utility functions,

1. Every agent has a dominant strategy: Regardless of the actions $a_{-i}$ of the other agents, there is at least one action $a^*_i$ such that $u_i(a^*_i, a_{-i}) \geq u_i(a'_i, a_{-i}) \quad \forall\, a'_i \in A_i$

2. In any such equilibrium $a^*$, we have $M(a^*) = C(u)$.

# Direct Mechanisms

- The space of all functions that map actions to outcomes is impossibly large to reason about

- Fortunately, we can restrict ourselves without loss of generality to the class of **truthful, direct** mechanisms

**Definition:** A **direct** mechanism is one in which $A_i = L$ for all agents $i$.

**Definition:**
A direct mechanism is **truthful** (or **incentive compatible**, or **strategy-proof**) if, for all preference profiles, it is a dominant strategy in the game induced by the mechanism for each agent to report their true preferences.

# Revelation Principle

**Theorem:** (Revelation Principle)
If there exists **any** mechanism that implements a social choice function $C$ in dominant strategies, then there exists a **direct** mechanism that implements $C$ in dominant strategies and is truthful.

# General
# Dominant-Strategy Implementation

**Theorem:** (Gibbard-Satterthwaite)
Consider any social choice function $C$ over $N$ and $O$. If

1. $|O| > 2$ (there are at least **three** outcomes),

2. $C$ is **onto**; that is, for every outcome $o \in O$ there is a preference profile such that $C([>]) = o$
(this is sometimes called **citizen sovereignty**), and

3. $C$ is dominant-strategy **truthful**,

then C is **dictatorial**.

# Hold On A Second

- Haven't we already seen an example of a dominant-strategy truthful direct mechanism?
  - Yes, the second-price auction!
- **Question:** Why is this not ruled out by Gibbard-Satterthwaite?

# Restricted Preferences

- Gibbard-Satterthwaite only applies to social choice functions that operate on **every possible** preference ordering over the outcomes
- By **restricting the set of preferences** that we operate over, we can circumvent Gibbard-Satterthwaite
- i.e., the second-price auction only considers preferences of the following form:
    1. Getting the item for **less than it's worth** to *i* is better than
    2. **Not getting** the item, which is better than
    3. Getting the item for **more than it's worth** to *i*

# Summary

- All **voting rules** lead to **unfair or undesirable** outcomes
    - **Arrow's Theorem**: this is **unavoidable**
- **Mechanism design**: Setting up a system for **strategic agents** to provide input to a **social choice function**
- **Revelation Principle** means we can restrict ourselves to **truthful direct** mechanisms without loss of generality
- Non-dictatorial dominant-strategy mechanism design is **impossible in general** (Gibbard-Satterthwaite)
    - But in practice we get around this by **restricting** the set of possible **preferences**