# Social Choice

CMPUT 654: Modelling Human Strategic Behaviour

S&LB §9.1-9.4

# Recap: Bayesian Games

- **Epistemic types** are a profile of signals that **parameterize the utility functions** of each agent
    - Possibly correlated
    - Each agent observes only their own type
- Three notions of **expected utility**:
    - ***ex-ante:*** before observing type
    - ***ex-interim:*** after observing own type
    - ***ex-post:*** full type profile is known
- Solution concepts:
    - **Bayes-Nash equilibrium:** equilibrium of induced normal form of ***ex-ante* utilities**
    - ***Ex-post* equilibrium:** Agents are best-responding at **every type profile** (not just in expectation)

# Lecture Outline

1. Paper Bidding & Recap
2. Aggregating Preferences
3. Voting Paradoxes
4. Arrow's Theorem

# 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
- How should we choose the outcome?
- **More formally:** Can we construct a **social choice function** that maps the profile of preference orderings to an outcome?
- **More generally:** Can we construct a **social welfare function** that maps the profile of preference orderings to an **aggregated preference ordering**?

# Inter-Agent Preference Comparisons

- Utility theory converts an **ordinal** **preference relation** into a **cardinal** **utility function**
    - Can compare the **strength** of a single agent's preferences
- **Problem:** "Units" of an agent's utility function are not fixed
- **Question:** How can we compare the strength of **two** agents' preferences?
- **Question:** Why can't we just convert both agents' utilities to **dollars** (or Euros or Bitcoin or...) and compare *those*?

# Formal Model

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

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

**Definition:** A social welfare function is a function $C : L^n \to L$, where $N$, $O$, and $L$ are as above.

**Notation:**
We will denote $i$'s preference order as $\succeq_i \in L$, and a profile of preference orders as $[\succeq] \in L^n$.

# Non-ranking Voting Schemes

Voters **need not** submit a full preference ordering:

1. **Plurality voting:** Everyone votes for favourite outcome, choose the outcome with the most votes

2. **Cumulative voting:** Everyone is given $k$ votes to distribute among candidates as they like; choose the outcome with the most votes

3. **Approval voting:** Each agent casts a single vote for each of the outcomes that are "acceptable"; choose the outcome with the most votes.

# Ranking Voting Schemes

Every agent expresses their **full preference ordering**:

1. **Plurality with elimination**
   - Everyone votes for favourite outcome
   - Outcome with least votes is eliminated
   - Repeat until one outcome remains
2. **Borda**
   - Everyone assigns scores to outcome: Most-preferred gets $m - 1$, next-most-preferred gets $m - 2$, etc. Least-preferred outcome gets 0.
   - Outcome with highest sum of scores is chosen
3. **Pairwise Elimination**
   - Define a schedule over the order in which pairs of outcomes will be compared
   - For each pair, everyone chooses their favourite; least-preferred is eliminated
   - Continue to next pair of non-eliminated outcomes until only one outcome remains

$m = |O|$

# Condorcet Condition

**Definition:**
An outcome $o \in O$ is a **Condorcet winner** if $\forall o' \in O$,

$$|i \in N : o \succ_i o'| > |i \in N : o' \succ_i o| .$$

**Definition:**
A social choice function satisfies the **Condorcet condition** if it always selects a Condorcet winner when one exists.

- If there's one outcome that would win a pairwise vote against every other possible outcome, then perhaps we want our social choice rule to pick it
- Unfortunately, such an outcome does not always exist
    - There can be cycles where A would beat B, B would beat C, C would beat A

# Paradox: Condorcet Winner

499 agents: $a \succ b \succ c$

3 agents: $b \succ c \succ a$

498 agents: $c \succ b \succ a$

- **Question:** Who is the **Condorcet winner**?
- **Question:** Who wins a **plurality election**?
- **Question:** Who wins under **plurality with elimination**?

# 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:** Who wins under **Borda**?
- **Question:** Now **drop** $c$. Who wins under **plurality**?
- **Question:** After dropping $c$, who wins under **Borda**?

# Paradox:
# Sensitivity to Agenda Setter

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

- **Question:** Who wins under **pairwise elimination** with order ***a,b,c***?
- **Question:** Who wins with ordering ***a,c,b***?
- **Question:** Who wins with ordering ***b,c,a***?
- The person who sets the comparison order can cause **any** of the three outcomes to be picked!

# Paradox: Sensitivity to Agenda Setter

1 agent: $b \succ d \succ c \succ a$

1 agent: $a \succ b \succ d \succ c$

1 agent: $c \succ a \succ b \succ d$

- **Question:** Who wins with ordering ***a,b,c,d***?
- **Question:** What is **wrong** with that?

# Arrow's Theorem

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

**Notation:**

- For this section we switch to **strict total orderings** $L$
- Preference ordering selected by social welfare function $W$ is $\succ_W$.

# Pareto Efficiency

**Definition:**
$W$ is **Pareto efficient** if for any $o_1, o_2 \in O$,

$$(\forall i \in N : o_1 \succ_i o_2) \implies (o_1 \succ_W o_2).$$

- If **everyone agrees** that $o_1$ is better than $o_2$, then the aggregated preference order should reflect that.

# Independence of Irrelevant Alternatives

**Definition:**

$W$ is **independent of irrelevant alternatives** if, for any $o_1, o_2 \in O$ and any two preference profiles $[\succ'], [\succ''] \in L$,

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

- If every agent has the **same ordering** between **two particular outcomes** in two different preference profiles, then the social welfare function on those two profiles must order those two outcomes the same way

- The ordering between two outcomes should only depend on the agents' orderings between **those outcomes**, not on where any other outcomes are in the agents' orderings

# Non-Dictatorship

**Definition:**
*W* does not have a **dictator** if

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

- No single agent determines the social ordering

# Arrow's Theorem

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

- Unfortunately, restricting to social choice functions instead of full social welfare functions doesn't help.

**Theorem:** (Muller-Satterthwaite, 1977)
If $|O| > 2$, any social choice function that is weakly Pareto efficient and monotonic is dictatorial.

# Summary

- Social choice is the study of **aggregating the true preferences** of a group of agents
  - **Social choice function**: Chooses a **single outcome** based on preference profile
  - **Social welfare function**: Chooses a **full preference order** over outcomes based on preference profile
- Well-known voting rules all lead to **unfair or undesirable** outcomes
  - **Arrow's Theorem:** This is unavoidable
  - **Muller-Satterthwaite Theorem:** ... even restricting to social choice functions