

भारतीय प्रौद्योगिकी संस्थान मुंबई

Indian Institute of Technology Bombay

CS 6001: Game Theory and Algorithmic Mechanism Design

Week 8 – Social Choice, Voting, and the Gibbard–Satterthwaite Theorem

Swaprava Nath

September, 2026

ज्ञानम् परमम् ध्येयम्

Knowledge is the supreme goal

Week 8 Roadmap

Lecture	Chunks (20–25 min each)	Quizzes
Lecture 15	1. Social Choice Setup & Arrow's Impossibility 2. Voting Rules & Condorcet Consistency 3. Pareto Efficiency, Unanimity, Manipulability 4. Strategyproofness $\Leftrightarrow$ Monotonicity	Quiz 1 after Chunk 2 Quiz 2 after Chunk 4
Lecture 16	1. The Gibbard–Satterthwaite Theorem 2. Proof of GS Theorem 3. Domain Restrictions & Single-Peaked Preferences 4. The Median Voter Theorem	Quiz 1 after Chunk 2 Quiz 2 after Chunk 4

Contents



- Lecture 15: The Social Choice Setup and Introduction to Voting
- Lecture 16: The Gibbard–Satterthwaite Theorem

Chunk 1

Social Choice Setup and Arrow's Impossibility

(≈ 20–25 minutes)

Choosing a Model: A Committee Problem

A company must deploy exactly one generative AI model as its customer-support assistant. A review committee shortlists several candidate models, and every member ranks them differently: one weighs benchmark accuracy, another latency and cost per token, another safety and refusal behaviour.

Choosing a Model: A Committee Problem

A company must deploy exactly one generative AI model as its customer-support assistant. A review committee shortlists several candidate models, and every member ranks them differently: one weighs benchmark accuracy, another latency and cost per token, another safety and refusal behaviour.

The question

Is there a good way to combine everyone's ranking into a single team decision?

Choosing a Model: A Committee Problem

A company must deploy exactly one generative AI model as its customer-support assistant. A review committee shortlists several candidate models, and every member ranks them differently: one weighs benchmark accuracy, another latency and cost per token, another safety and refusal behaviour.

The question

Is there a good way to combine everyone's ranking into a single team decision?

Definition (Social choice function)

Let $A = \{a_1, \ldots, a_m\}$ be a finite set of alternatives (here, candidate models), $N = \{1, \ldots, n\}$ a finite set of agents (committee members), and $\mathcal{P}$ the set of all linear (strict, complete, transitive) preference orderings over A. A **social choice function** (SCF) is

$$f : \mathcal{P}^n \to A.$$

Choosing a Model: A Committee Problem

A company must deploy exactly one generative AI model as its customer-support assistant. A review committee shortlists several candidate models, and every member ranks them differently: one weighs benchmark accuracy, another latency and cost per token, another safety and refusal behaviour.

The question

Is there a good way to combine everyone's ranking into a single team decision?

Definition (Social choice function)

Let $A = \{a_1, \ldots, a_m\}$ be a finite set of alternatives (here, candidate models), $N = \{1, \ldots, n\}$ a finite set of agents (committee members), and $\mathcal{P}$ the set of all linear (strict, complete, transitive) preference orderings over A. A **social choice function** (SCF) is

$$f : \mathcal{P}^n \to A.$$

A **profile** $P = (P_1, \ldots, P_n) \in \mathcal{P}^n$ records every member's full ranking of the models.

Recall: Three Fairness Conditions for Welfare Function



- **Weak Pareto**: if every member strictly prefers model a to b, the aggregate ranking must place a above b.

Recall: Three Fairness Conditions for Welfare Function



- **Weak Pareto**: if every member strictly prefers model a to b, the aggregate ranking must place a above b.
- **Independence of Irrelevant Alternatives (IIA)**: the aggregate comparison between a and b should depend only on how each member individually ranks a against b – not on where some third model c sits.

Recall: Three Fairness Conditions for Welfare Function



- **Weak Pareto**: if every member strictly prefers model a to b, the aggregate ranking must place a above b.
- **Independence of Irrelevant Alternatives (IIA)**: the aggregate comparison between a and b should depend only on how each member individually ranks a against b – not on where some third model c sits.
- **Non-dictatorship**: there is no single member whose top choice always becomes the team's decision, regardless of everyone else's input.

Recall: Three Fairness Conditions for Welfare Function



- **Weak Pareto**: if every member strictly prefers model a to b, the aggregate ranking must place a above b.
- **Independence of Irrelevant Alternatives (IIA)**: the aggregate comparison between a and b should depend only on how each member individually ranks a against b – not on where some third model c sits.
- **Non-dictatorship**: there is no single member whose top choice always becomes the team's decision, regardless of everyone else's input.

Each condition on its own sounds like something any reasonable committee process should satisfy.

Arrow's Impossibility Theorem (Rephrased)

Theorem (Arrow)

If $|A| \geqslant 3$, there exists no social welfare function that simultaneously satisfies Weak Pareto, IIA, and Non-dictatorship.

Arrow's Impossibility Theorem (Rephrased)

Theorem (Arrow)

If $|A| \geqslant 3$, there exists no social welfare function that simultaneously satisfies Weak Pareto, IIA, and Non-dictatorship.

What this means for our committee

The moment there are three or more candidate models, no method of combining full rankings into one social ranking can honor all three conditions at once. Something has to give.

Arrow's Impossibility Theorem (Rephrased)

Theorem (Arrow)

If $|A| \geqslant 3$, there exists no social welfare function that simultaneously satisfies Weak Pareto, IIA, and Non-dictatorship.

What this means for our committee

The moment there are three or more candidate models, no method of combining full rankings into one social ranking can honor all three conditions at once. Something has to give.

Question

Given this impossibility, how do practical systems – elections, review committees, recommender aggregation – get anything done at all?

Two Escape Routes

1. **Change the object of choice**: instead of a full social ordering, only ask for a single winner. This is exactly the social choice function $f : \mathcal{P}^n \to A$ we defined earlier – and it is the object we will study for the rest of the week.

Two Escape Routes

1. **Change the object of choice**: instead of a full social ordering, only ask for a single winner. This is exactly the social choice function $f : \mathcal{P}^n \to A$ we defined earlier – and it is the object we will study for the rest of the week.
2. **Restrict the domain of preferences**: assume members' preferences satisfy some structure (for example, single-peaked preferences, which we meet later this week). Under such restrictions, much stronger positive results become possible.

Two Escape Routes

1. **Change the object of choice**: instead of a full social ordering, only ask for a single winner. This is exactly the social choice function $f : \mathcal{P}^n \to A$ we defined earlier – and it is the object we will study for the rest of the week.
2. **Restrict the domain of preferences**: assume members' preferences satisfy some structure (for example, single-peaked preferences, which we meet later this week). Under such restrictions, much stronger positive results become possible.

Answer

Both routes are used constantly in practice – a plurality vote picks one winner rather than a full ranking (route 1); a facility-location decision assumes preferences fall off with distance (route 2).

Notation Summary and a Small Example

- $A = \{a_1, \ldots, a_m\}$: alternatives (models, candidates, policies, ...)
- $N = \{1, \ldots, n\}$: agents (committee members, voters, ...)
- $\mathcal{P}$: set of all linear orderings over A
- A profile: $P = (P_1, \ldots, P_n) \in \mathcal{P}^n$
- An SCF: $f : \mathcal{P}^n \to A$

Notation Summary and a Small Example

- $A = \{a_1, \ldots, a_m\}$: alternatives (models, candidates, policies, ...)
- $N = \{1, \ldots, n\}$: agents (committee members, voters, ...)
- $\mathcal{P}$: set of all linear orderings over A
- A profile: $P = (P_1, \ldots, P_n) \in \mathcal{P}^n$
- An SCF: $f : \mathcal{P}^n \to A$

P			
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

$\xrightarrow{f}$ $A = \{a, b, c, d\}$

Chunk 2

Voting Rules and Condorcet Consistency

(≈ 20–25 minutes)

A Model Selection Vote

The committee has shortlisted four candidate models $A = \{a, b, c, d\}$. Four members each rank all four, weighing benchmark accuracy, latency and cost, safety behaviour, and multilingual coverage differently.

$\mathbf{M}_1$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

A Model Selection Vote

The committee has shortlisted four candidate models $A = \{a, b, c, d\}$. Four members each rank all four, weighing benchmark accuracy, latency and cost, safety behaviour, and multilingual coverage differently.

$\mathbf{M}_1$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

Scoring rule

Assign a score vector $(s_1, \ldots, s_m)$, $s_i \geqslant s_{i+1}$, to ranking positions; sum each model's score across members; highest total wins.

Common Scoring Rules

- **Plurality**: only the top rank counts. Score vector $(1, 0, 0, 0)$.
- **Veto**: every position except last gets a point. Score vector $(1, 1, 1, 0)$.
- **Borda count**: named after Jean-Charles de Borda. Score vector $(m-1, m-2, \ldots, 1, 0)$; for $m = 4$: $(3, 2, 1, 0)$.
- **Harmonic rule**: score vector $\left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{m}\right)$.
- k**-Approval**: top k positions each get 1, rest get 0.

Common Scoring Rules

- **Plurality**: only the top rank counts. Score vector $(1, 0, 0, 0)$.
- **Veto**: every position except last gets a point. Score vector $(1, 1, 1, 0)$.
- **Borda count**: named after Jean-Charles de Borda. Score vector $(m-1, m-2, \ldots, 1, 0)$; for $m = 4$: $(3, 2, 1, 0)$.
- **Harmonic rule**: score vector $\left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{m}\right)$.
- k**-Approval**: top k positions each get 1, rest get 0.

Question

None of these need the member to report anything beyond a ranking – but do any two of them ever disagree on the winner for the same profile?

Non-Scoring Rules

- **Plurality with runoff** (two-round system, TRS): round 1 is plain plurality, top two models survive; round 2 is plurality between just those two. Used in the French presidential election.

Non-Scoring Rules

- **Plurality with runoff** (two-round system, TRS): round 1 is plain plurality, top two models survive; round 2 is plurality between just those two. Used in the French presidential election.
- **Maximin**: $\text{score}(a) = \min_{y \in A \setminus \{a\}} |\{i : a \succ_i y\}|$ – a model's score is its worst-case pairwise support.

Non-Scoring Rules

- **Plurality with runoff** (two-round system, TRS): round 1 is plain plurality, top two models survive; round 2 is plurality between just those two. Used in the French presidential election.
- **Maximin**: $\text{score}(a) = \min_{y \in A \setminus \{a\}} |\{i : a \succ_i y\}|$ – a model's score is its worst-case pairwise support.
- **Copeland**: $\text{score}(a) =$ number of pairwise contests a wins against other models.

Maximin on the Model Vote

$\mathbf{M}_1$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

Maximin on the Model Vote

$\mathbf{M}_1$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

$$\text{score}(a) = \min\{2(b), 2(c), 2(d)\} = 2, \qquad \text{score}(b) = \min\{2(a), 2(c), 3(d)\} = 2$$
$$\text{score}(c) = \min\{2(a), 2(b), 3(d)\} = 2, \qquad \text{score}(d) = \min\{2(a), 1(b), 1(c)\} = 1$$

Maximin on the Model Vote

$\mathbf{M}_1$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

$$\text{score}(a) = \min\{2(b), 2(c), 2(d)\} = 2, \qquad \text{score}(b) = \min\{2(a), 2(c), 3(d)\} = 2$$
$$\text{score}(c) = \min\{2(a), 2(b), 3(d)\} = 2, \qquad \text{score}(d) = \min\{2(a), 1(b), 1(c)\} = 1$$

Models a, b, c tie at 2; d trails at 1, the weakest worst-case matchup.

Copeland on the Same Profile

All six head-to-head contests on the very same four ballots:

Copeland on the Same Profile

All six head-to-head contests on the very same four ballots:

	vs b	**vs** c	**vs** d
a	2–2 tie	2–2 tie	2–2 tie
b		2–2 tie	3–1 b
c			3–1 c

Copeland on the Same Profile

All six head-to-head contests on the very same four ballots:

	vs b	**vs** c	**vs** d
a	2–2 tie	2–2 tie	2–2 tie
b		2–2 tie	3–1 b
c			3–1 c

$$\text{Copeland}(a) = 0, \quad \text{Copeland}(b) = 1, \quad \text{Copeland}(c) = 1, \quad \text{Copeland}(d) = 0$$

Copeland on the Same Profile

All six head-to-head contests on the very same four ballots:

	vs b	**vs** c	**vs** d
a	2–2 tie	2–2 tie	2–2 tie
b		2–2 tie	3–1 b
c			3–1 c

$$\text{Copeland}(a) = 0, \quad \text{Copeland}(b) = 1, \quad \text{Copeland}(c) = 1, \quad \text{Copeland}(d) = 0$$

Answer

b and c tie as Copeland winners. Maximin rated a, b, c all equal; Copeland separates them, because ties pay nothing and only b and c actually *win* a duel. Both rules agree that d is worst.

Condorcet Winners

Definition (Condorcet winner)

Model x is a **Condorcet winner** if it beats every other model in a head-to-head majority vote.

Condorcet Winners

Definition (Condorcet winner)

Model x is a **Condorcet winner** if it beats every other model in a head-to-head majority vote.

Definition (Condorcet consistency)

An SCF is **Condorcet consistent** if it always selects the Condorcet winner when one exists.

Condorcet Winners

Definition (Condorcet winner)

Model x is a **Condorcet winner** if it beats every other model in a head-to-head majority vote.

Definition (Condorcet consistency)

An SCF is **Condorcet consistent** if it always selects the Condorcet winner when one exists.

A Condorcet winner need not exist. Three members with a rock-paper-scissors style cycle:

P		
a	b	c
b	c	a
c	a	b

Condorcet Winners

Definition (Condorcet winner)

Model x is a **Condorcet winner** if it beats every other model in a head-to-head majority vote.

Definition (Condorcet consistency)

An SCF is **Condorcet consistent** if it always selects the Condorcet winner when one exists.

A Condorcet winner need not exist. Three members with a rock-paper-scissors style cycle:

P		
a	b	c
b	c	a
c	a	b

– every model loses one pairwise contest, wins another.

When a Condorcet Winner Does Exist

The org polls 100 engineers. Their rankings, as vote shares:

30%	30%	40%
a	b	c
b	a	a
c	c	b

When a Condorcet Winner Does Exist

The org polls 100 engineers. Their rankings, as vote shares:

30%	30%	40%
a	b	c
b	a	a
c	c	b

Pairwise: a beats b (70–30), a beats c (60–40), b beats c (60–40). Model a beats everyone, so it is the **Condorcet winner**.

When a Condorcet Winner Does Exist

The org polls 100 engineers. Their rankings, as vote shares:

30%	30%	40%
a	b	c
b	a	a
c	c	b

Pairwise: a beats b (70–30), a beats c (60–40), b beats c (60–40). Model a beats everyone, so it is the **Condorcet winner**.

Question

Plurality only counts top-place votes: a gets 30%, b gets 30%, c gets 40%. Who does plurality pick?

When a Condorcet Winner Does Exist

The org polls 100 engineers. Their rankings, as vote shares:

30%	30%	40%
a	b	c
b	a	a
c	c	b

Pairwise: a beats b (70–30), a beats c (60–40), b beats c (60–40). Model a beats everyone, so it is the **Condorcet winner**.

Question

Plurality only counts top-place votes: a gets 30%, b gets 30%, c gets 40%. Who does plurality pick?

Answer

Plurality picks c – even though b would beat c head-to-head, 60 to 40. No scoring rule is Condorcet consistent in general. Copeland and maximin, however, both correctly pick a here.

No Scoring Rule Is Condorcet Consistent



Theorem

For $|A| \geqslant 3$, no scoring rule is Condorcet consistent.

No Scoring Rule Is Condorcet Consistent



Theorem

For $|A| \geqslant 3$, no scoring rule is Condorcet consistent.

Enough to argue with **three** models and the normalised vector $(1, \lambda, 0)$, $\lambda \in [0, 1]$; next slide justifies both.

No Scoring Rule Is Condorcet Consistent



Theorem

For $|A| \geqslant 3$, no scoring rule is Condorcet consistent.

Enough to argue with **three** models and the normalised vector $(1, \lambda, 0)$, $\lambda \in [0, 1]$; next slide justifies both.

Case $\lambda > 0$. Take $N > 1/\lambda$ ballots $a \succ b \succ c$ and $N - 1$ ballots $b \succ c \succ a$. Then a beats b and c by N to $N - 1$, so a is the **Condorcet winner**. But

$$\text{score}(a) = N, \qquad \text{score}(b) = \lambda N + (N - 1) \; > \; N.$$

No Scoring Rule Is Condorcet Consistent

Theorem

For $|A| \geqslant 3$, no scoring rule is Condorcet consistent.

Enough to argue with **three** models and the normalised vector $(1, \lambda, 0)$, $\lambda \in [0, 1]$; next slide justifies both.

Case $\lambda > 0$. Take $N > 1/\lambda$ ballots $a \succ b \succ c$ and $N - 1$ ballots $b \succ c \succ a$. Then a beats b and c by N to $N - 1$, so a is the **Condorcet winner**. But

$$\text{score}(a) = N, \qquad \text{score}(b) = \lambda N + (N - 1) \; > \; N.$$

Case $\lambda = 0$ (plurality). Two ballots $a \succ b \succ c$, two $c \succ b \succ a$, one $b \succ a \succ c$: now b beats a and c by 3–2, yet plurality scores $a = 2$, $c = 2$, $b = 1$.

No Scoring Rule Is Condorcet Consistent

Theorem

For $|A| \geqslant 3$, no scoring rule is Condorcet consistent.

Enough to argue with **three** models and the normalised vector $(1, \lambda, 0)$, $\lambda \in [0, 1]$; next slide justifies both.

Case $\lambda > 0$. Take $N > 1/\lambda$ ballots $a \succ b \succ c$ and $N - 1$ ballots $b \succ c \succ a$. Then a beats b and c by N to $N - 1$, so a is the **Condorcet winner**. But

$$\text{score}(a) = N, \qquad \text{score}(b) = \lambda N + (N - 1) \; > \; N.$$

Case $\lambda = 0$ (plurality). Two ballots $a \succ b \succ c$, two $c \succ b \succ a$, one $b \succ a \succ c$: now b beats a and c by 3–2, yet plurality scores $a = 2$, $c = 2$, $b = 1$.

Borda $(2, 1, 0)$: $\lambda = \frac{1}{2}$, so $N = 3$. Three ballots $a \succ b \succ c$, two $b \succ c \succ a$: a is the Condorcet winner, but $a = 6$ and $b = 7$, so Borda picks b.

Why Three Models, and Why $(1, \lambda, 0)$?

Why $(1, \lambda, 0)$ is no loss of generality. Each model occupies *exactly one* position on each ballot, so:

- adding a constant c to every s_i adds exactly cn to *every* model's total,

Why Three Models, and Why $(1, \lambda, 0)$?

Why $(1, \lambda, 0)$ is no loss of generality. Each model occupies *exactly one* position on each ballot, so:

- adding a constant c to every s_i adds exactly cn to *every* model's total,
- multiplying every s_i by a constant $\gamma > 0$ scales every total by γ.

Why Three Models, and Why $(1, \lambda, 0)$?

Why $(1, \lambda, 0)$ is no loss of generality. Each model occupies *exactly one* position on each ballot, so:

- adding a constant c to every s_i adds exactly cn to *every* model's total,
- multiplying every s_i by a constant $\gamma > 0$ scales every total by γ.

Neither changes who is ranked top, so shift by $-s_3$ and divide by $s_1 - s_3$ to get $(1, \lambda, 0)$ with $\lambda = \frac{s_2 - s_3}{s_1 - s_3} \in [0, 1]$, using $s_1 \geqslant s_2 \geqslant s_3$.

Why Three Models, and Why $(1, \lambda, 0)$?

Why $(1, \lambda, 0)$ is no loss of generality. Each model occupies *exactly one* position on each ballot, so:

- adding a constant c to every s_i adds exactly cn to *every* model's total,
- multiplying every s_i by a constant $\gamma > 0$ scales every total by γ.

Neither changes who is ranked top, so shift by $-s_3$ and divide by $s_1 - s_3$ to get $(1, \lambda, 0)$ with $\lambda = \frac{s_2 - s_3}{s_1 - s_3} \in [0, 1]$, using $s_1 \geqslant s_2 \geqslant s_3$.

Degenerate case: if $s_1 = s_3$ then $s_1 = s_2 = s_3$, every model ties at every profile, and the rule ignores the ballots entirely, so it is not Condorcet consistent either.

Why Three Models, and Why $(1, \lambda, 0)$?

Why $(1, \lambda, 0)$ is no loss of generality. Each model occupies *exactly one* position on each ballot, so:

- adding a constant c to every s_i adds exactly cn to *every* model's total,
- multiplying every s_i by a constant $\gamma > 0$ scales every total by γ.

Neither changes who is ranked top, so shift by $-s_3$ and divide by $s_1 - s_3$ to get $(1, \lambda, 0)$ with $\lambda = \frac{s_2 - s_3}{s_1 - s_3} \in [0, 1]$, using $s_1 \geqslant s_2 \geqslant s_3$.

Degenerate case: if $s_1 = s_3$ then $s_1 = s_2 = s_3$, every model ties at every profile, and the rule ignores the ballots entirely, so it is not Condorcet consistent either.

Why three models. At $|A| = 2$ the theorem is false: every scoring rule (s_1, s_2) with $s_1 > s_2$ *is* majority rule, which always picks the Condorcet winner. For $m > 3$, pad every ballot with the extra models at the bottom: pairwise comparisons among $\{a, b, c\}$ are untouched, and those models sit at positions scoring $\leqslant 0$, so they cannot win.



In-Class Quiz 1

Open the Google Form now.

Questions refer only to what was just discussed –
you cannot answer them without having been in class.

Chunk 3

Pareto Efficiency, Unanimity, and Manipulability

(≈ 20–25 minutes)

Pareto Domination and Efficiency

Back to the model selection committee, $f : \mathcal{P}^n \to A$.

Pareto Domination and Efficiency



Back to the model selection committee, $f : \mathcal{P}^n \to A$.

Definition (Pareto domination)

Model a is **Pareto dominated** by b at profile P if every member strictly prefers b to a: $\forall i, b \succ_i a$.

Pareto Domination and Efficiency

Back to the model selection committee, $f : \mathcal{P}^n \to A$.

Definition (Pareto domination)

Model a is **Pareto dominated** by b at profile P if every member strictly prefers b to a: $\forall i, b \succ_i a$.

Definition (Pareto efficiency)

SCF f is **Pareto efficient** (PE) if it never outputs a Pareto dominated model.

Pareto Domination and Efficiency

Back to the model selection committee, $f : \mathcal{P}^n \to A$.

Definition (Pareto domination)

Model a is **Pareto dominated** by b at profile P if every member strictly prefers b to a: $\forall i, b \succ_i a$.

Definition (Pareto efficiency)

SCF f is **Pareto efficient** (PE) if it never outputs a Pareto dominated model.

Intuition

If every member agrees b is strictly better than a, promoting a to production would be a strange, avoidable mistake.

Pareto Domination: An Example

Four members, four models:

M1	M2	M3	M4
a	*b*	*b*	*d*
b	*a*	*d*	*b*
c	*c*	*a*	*a*
d	*d*	*c*	*c*

Pareto Domination: An Example

Four members, four models:

M1	M2	M3	M4
a	*b*	*b*	*d*
b	*a*	*d*	*b*
c	*c*	*a*	*a*
d	*d*	*c*	*c*

Which models are Pareto dominated?

Pareto Domination: An Example

Four members, four models:

M1	M2	M3	M4
a	b	b	d
b	a	d	b
c	c	a	a
d	d	c	c

Which models are Pareto dominated?

Answer

Only c: every ballot puts both a and b above it. A PE rule may output a, b or d, never c.

Pareto Domination: An Example

Four members, four models:

M1	M2	M3	M4
a	b	b	d
b	a	d	b
c	c	a	a
d	d	c	c

Which models are Pareto dominated?

Answer

Only c: every ballot puts both a and b above it. A PE rule may output a, b or d, never c.

The subtle part

Being nobody's favourite is not what kills c: a tops one ballot and survives. What kills c is that a beats it on *every* ballot.

Unanimity

Definition (Unanimity)

SCF f is **unanimous** (UN) if whenever every member's top pick is the same model a, then $f(P) = a$.

Unanimity

Definition (Unanimity)

SCF f is **unanimous** (UN) if whenever every member's top pick is the same model a, then $f(P) = a$.

Question

Which implies which – does PE imply UN, does UN imply PE, or both?

Unanimity

Definition (Unanimity)

SCF f is **unanimous** (UN) if whenever every member's top pick is the same model a, then $f(P) = a$.

Question

Which implies which – does PE imply UN, does UN imply PE, or both?

Answer

PE $\implies$ **UN**: if everyone's top choice is a, every other model is Pareto dominated by a, so PE forces $f(P) = a$.

Unanimity

Definition (Unanimity)

SCF f is **unanimous** (UN) if whenever every member's top pick is the same model a, then $f(P) = a$.

Question

Which implies which – does PE imply UN, does UN imply PE, or both?

Answer

PE $\implies$ **UN**: if everyone's top choice is a, every other model is Pareto dominated by a, so PE forces $f(P) = a$.

The converse fails. Consider $f(P) = a$ if the profile is unanimous for a, else $f(P) = b$ always. This f is unanimous by construction, but if every member prefers c to b without being unanimous for a, f still outputs b – a Pareto dominated model. **UN does not imply PE.**

Onto and Manipulability

Definition (Onto)

f is **onto** (ONTO) if every model $a \in A$ is the outcome of f for some profile.

Onto and Manipulability

Definition (Onto)

f is **onto** (ONTO) if every model $a \in A$ is the outcome of f for some profile.

UN $\implies$ ONTO: unanimously ranking any a first forces $f(P) = a$, so every a is reachable.

Onto and Manipulability

Definition (Onto)

f is **onto** (ONTO) if every model $a \in A$ is the outcome of f for some profile.

UN $\implies$ ONTO: unanimously ranking any a first forces $f(P) = a$, so every a is reachable.

Definition (Manipulability)

f is **manipulable** if some member i, by misreporting P_i' instead of their true P_i, can force an outcome they strictly prefer: $f(P_i', P_{-i}) P_i f(P_i, P_{-i})$.

Onto and Manipulability

Definition (Onto)

f is **onto** (ONTO) if every model $a \in A$ is the outcome of f for some profile.

UN $\implies$ ONTO: unanimously ranking any a first forces $f(P) = a$, so every a is reachable.

Definition (Manipulability)

f is **manipulable** if some member i, by misreporting P_i' instead of their true P_i, can force an outcome they strictly prefer: $f(P_i', P_{-i}) P_i f(P_i, P_{-i})$.

Question

Why should a committee care about manipulability at all?

Manipulability: Plurality

A city council with three candidate pricing policies $\{a, b, c\}$, fixed tie-break $a \succ b \succ c$, plurality rule. Nine council members vote sincerely:

4	4	1
a	b	c
b	a	b
c	c	a

Manipulability: Plurality

A city council with three candidate pricing policies $\{a, b, c\}$, fixed tie-break $a \succ b \succ c$, plurality rule. Nine council members vote sincerely:

4	4	1
a	b	c
b	a	b
c	c	a

$\Rightarrow$

4	4	1
a	b	b
b	a	c
c	c	a

Manipulability: Plurality

A city council with three candidate pricing policies $\{a, b, c\}$, fixed tie-break $a \succ b \succ c$, plurality rule. Nine council members vote sincerely:

4	4	1
a	b	c
b	a	b
c	c	a

$\Rightarrow$

4	4	1
a	b	b
b	a	c
c	c	a

a and b tie at 4 votes, tie-break gives a. The lone member who truly prefers c switches their ballot to $b \succ c \succ a$, pushing b to a clean majority. They never wanted b most – but they strictly prefer b to the tie-break winner a, so lying pays off. **Plurality with fixed tie-breaking is manipulable.**

Manipulability: Copeland

Three committee members, three shortlisted models $\{a, b, c\}$, Copeland rule, tie-break $a \succ b \succ c$:

1	1	1
a	b	c
b	c	a
c	a	b

Manipulability: Copeland

Three committee members, three shortlisted models $\{a, b, c\}$, Copeland rule, tie-break $a \succ b \succ c$:

1	1	1
a	b	c
b	c	a
c	a	b

$\Rightarrow$

1	1	1
a	c	c
b	b	a
c	a	b

Manipulability: Copeland

Three committee members, three shortlisted models $\{a, b, c\}$, Copeland rule, tie-break $a \succ b \succ c$:

1	1	1
a	b	c
b	c	a
c	a	b

$\Rightarrow$

1	1	1
a	c	c
b	b	a
c	a	b

Every model wins exactly one pairwise contest, a perfect cycle, so Copeland scores tie at 1 and the tie-break picks a. Member 2 misreports, changing the pairwise results and the Copeland scores. **Even Copeland with fixed tie-breaking is manipulable.**

Chunk 4

Strategyproofness and Its Equivalence to Monotonicity

(≈ 25–30 minutes)

Strategyproofness

Definition (Strategyproof)

SCF f is **strategyproof** (SP) if it is not manipulable by any member at any profile.

Strategyproofness

Definition (Strategyproof)

SCF f is **strategyproof** (SP) if it is not manipulable by any member at any profile.

Definition (Dominated set)

For model a and preference P_i, the **dominated set** is

$$D(a, P_i) := \{b \in A : aP_ib\}.$$

Strategyproofness

Definition (Strategyproof)

SCF f is **strategyproof** (SP) if it is not manipulable by any member at any profile.

Definition (Dominated set)

For model a and preference P_i, the **dominated set** is

$$D(a, P_i) := \{b \in A : aP_ib\}.$$

If member i's ranking is $b \succ a \succ c \succ d$, then $D(a, P_i) = \{c, d\}$ – everything i ranks below a.

Strategyproofness

Definition (Strategyproof)

SCF f is **strategyproof** (SP) if it is not manipulable by any member at any profile.

Definition (Dominated set)

For model a and preference P_i, the **dominated set** is

$$D(a, P_i) := \{b \in A : aP_ib\}.$$

If member i's ranking is $b \succ a \succ c \succ d$, then $D(a, P_i) = \{c, d\}$ – everything i ranks below a.

Reading $D(a, P_i)$

A bigger dominated set means a sits higher in i's ranking.

Monotonicity

Definition (Monotonicity)

f is **monotone** (MONO) if whenever $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for every member i, then $f(P') = a$.

Monotonicity

Definition (Monotonicity)

f is **monotone** (MONO) if whenever $f(P) = a$ and $D(a, P_i) \subseteq D(a, P_i')$ for every member i, then $f(P') = a$.

In words: if model a already won, and every member moves a up (or leaves it where it was) in their ranking, a must still win.

Monotonicity

Definition (Monotonicity)

f is **monotone** (MONO) if whenever $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for every member i, then $f(P') = a$.

In words: if model a already won, and every member moves a up (or leaves it where it was) in their ranking, a must still win.

P			
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

P'			
c	a	c	d
b	c	b	c
a	b	d	b
d	d	a	a

Monotonicity

Definition (Monotonicity)

f is **monotone** (MONO) if whenever $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for every member i, then $f(P') = a$.

In words: if model a already won, and every member moves a up (or leaves it where it was) in their ranking, a must still win.

P			
a	a	c	d
b	b	b	c
c	c	d	b
d	d	a	a

P'			
c	a	c	d
b	c	b	c
a	b	d	b
d	d	a	a

Model c's position only improved from P to P'. If c was the outcome at P, monotonicity says c must still be the outcome at P'.

The Key Theorem: SP $\Leftrightarrow$ MONO



Theorem

An SCF f is strategyproof if and only if it is monotone.

The Key Theorem: SP $\Leftrightarrow$ MONO

Theorem

An SCF f is strategyproof if and only if it is monotone.

Why this matters

Strategyproofness is a statement about incentives – hard to check directly, since it quantifies over every possible lie. Monotonicity is a purely structural property of the rule – often much easier to verify. The equivalence lets us check the easy one.

The Key Theorem: SP $\Leftrightarrow$ MONO

Theorem

An SCF f is strategyproof if and only if it is monotone.

Why this matters

Strategyproofness is a statement about incentives – hard to check directly, since it quantifies over every possible lie. Monotonicity is a purely structural property of the rule – often much easier to verify. The equivalence lets us check the easy one.

Proof splits into two directions: SP $\implies$ MONO, and SP $\impliedby$ MONO (equivalently, $\neg$SP $\implies$ $\neg$MONO).

Proof: SP $\implies$ MONO

Suppose f is SP. Take P, P' with $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for all i. Change one member's report at a time:

$$P^{(0)} = P \to P^{(1)} = (P'_1, P_2, \ldots, P_n) \to \cdots \to P^{(k)} = (P'_1, \ldots, P'_k, P_{k+1}, \ldots, P_n) \to \cdots \to P^{(n)} = P'$$

Proof: SP $\implies$ MONO

Suppose f is SP. Take P, P' with $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for all i. Change one member's report at a time:

$$P^{(0)} = P \to P^{(1)} = (P'_1, P_2, \ldots, P_n) \to \cdots \to P^{(k)} = (P'_1, \ldots, P'_k, P_{k+1}, \ldots, P_n) \to \cdots \to P^{(n)} = P'$$

Claim: $f(P^{(k)}) = a$ for every k. Suppose not: $f(P^{(k-1)}) = a$ but $f(P^{(k)}) = b \neq a$, where only member k's report changed, $P_k \to P'_k$.

Proof: SP $\implies$ MONO

Suppose f is SP. Take P, P' with $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for all i. Change one member's report at a time:

$$P^{(0)} = P \to P^{(1)} = (P'_1, P_2, \ldots, P_n) \to \cdots \to P^{(k)} = (P'_1, \ldots, P'_k, P_{k+1}, \ldots, P_n) \to \cdots \to P^{(n)} = P'$$

Claim: $f(P^{(k)}) = a$ for every k. Suppose not: $f(P^{(k-1)}) = a$ but $f(P^{(k)}) = b \neq a$, where only member k's report changed, $P_k \to P'_k$.

- $a \, P_k \, b$ and $a \, P'_k \, b$: member k profitably misreports $P'_k \to P_k$ at $P^{(k)}$
- $b \, P_k \, a$ and $b \, P'_k \, a$: member k profitably misreports $P_k \to P'_k$ at $P^{(k-1)}$
- $b \, P_k \, a$ and $a \, P'_k \, b$: member k has a profitable misreport in *both* profiles

Proof: SP $\Longrightarrow$ MONO



Suppose f is SP. Take P, P' with $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for all i. Change one member's report at a time:

$$P^{(0)} = P \to P^{(1)} = (P'_1, P_2, \ldots, P_n) \to \cdots \to P^{(k)} = (P'_1, \ldots, P'_k, P_{k+1}, \ldots, P_n) \to \cdots \to P^{(n)} = P'$$

Claim: $f(P^{(k)}) = a$ for every k. Suppose not: $f(P^{(k-1)}) = a$ but $f(P^{(k)}) = b \neq a$, where only member k's report changed, $P_k \to P'_k$.

- $a \, P_k \, b$ and $a \, P'_k \, b$: member k profitably misreports $P'_k \to P_k$ at $P^{(k)}$
- $b \, P_k \, a$ and $b \, P'_k \, a$: member k profitably misreports $P_k \to P'_k$ at $P^{(k-1)}$
- $b \, P_k \, a$ and $a \, P'_k \, b$: member k has a profitable misreport in *both* profiles

The one combination that cannot occur is $a \, P_k \, b$ together with $b \, P'_k \, a$: from $a \, P_k \, b$ we get $b \in D(a, P_k) \subseteq D(a, P'_k)$, which forces $a \, P'_k \, b$.

Proof: SP $\Longrightarrow$ MONO

Suppose f is SP. Take P, P' with $f(P) = a$ and $D(a, P_i) \subseteq D(a, P'_i)$ for all i. Change one member's report at a time:

$$P^{(0)} = P \to P^{(1)} = (P'_1, P_2, \ldots, P_n) \to \cdots \to P^{(k)} = (P'_1, \ldots, P'_k, P_{k+1}, \ldots, P_n) \to \cdots \to P^{(n)} = P'$$

Claim: $f(P^{(k)}) = a$ for every k. Suppose not: $f(P^{(k-1)}) = a$ but $f(P^{(k)}) = b \neq a$, where only member k's report changed, $P_k \to P'_k$.

- $a \, P_k \, b$ and $a \, P'_k \, b$: member k profitably misreports $P'_k \to P_k$ at $P^{(k)}$
- $b \, P_k \, a$ and $b \, P'_k \, a$: member k profitably misreports $P_k \to P'_k$ at $P^{(k-1)}$
- $b \, P_k \, a$ and $a \, P'_k \, b$: member k has a profitable misreport in *both* profiles

The one combination that cannot occur is $a \, P_k \, b$ together with $b \, P'_k \, a$: from $a \, P_k \, b$ we get $b \in D(a, P_k) \subseteq D(a, P'_k)$, which forces $a \, P'_k \, b$.

Every possible case contradicts SP. So $f(P^{(k)}) = a$ throughout, and $f(P') = a$.

Proof: MONO $\implies$ SP



We show the contrapositive: $\neg$SP $\implies$ $\neg$MONO.

Proof: MONO $\implies$ SP

We show the contrapositive: $\neg$SP $\implies$ $\neg$MONO.

$\neg$SP: some member i can misreport $P_i \to P_i'$ and move the outcome from $a = f(P_i, P_{-i})$ to $b = f(P_i', P_{-i})$ with bP_ia.

Proof: MONO $\implies$ SP

We show the contrapositive: $\neg$SP $\implies$ $\neg$MONO.

$\neg$SP: some member i can misreport $P_i \to P_i'$ and move the outcome from $a = f(P_i, P_{-i})$ to $b = f(P_i', P_{-i})$ with bP_ia.

Construct P_i'' with b ranked first, a ranked second. Consider two transitions to (P_i'', P_{-i}):

- From (P_i, P_{-i}): $D(a, P_i) \subseteq D(a, P_i'')$, so MONO would force $f(P_i'', P_{-i}) = a$

Proof: MONO $\implies$ SP

We show the contrapositive: $\neg$SP $\implies$ $\neg$MONO.

$\neg$SP: some member i can misreport $P_i \to P_i'$ and move the outcome from $a = f(P_i, P_{-i})$ to $b = f(P_i', P_{-i})$ with bP_ia.

Construct P_i'' with b ranked first, a ranked second. Consider two transitions to (P_i'', P_{-i}):

- From (P_i, P_{-i}): $D(a, P_i) \subseteq D(a, P_i'')$, so MONO would force $f(P_i'', P_{-i}) = a$
- From (P_i', P_{-i}): $D(b, P_i') \subseteq D(b, P_i'')$ (b is now top-ranked), so MONO would force $f(P_i'', P_{-i}) = b$

Proof: MONO $\implies$ SP

We show the contrapositive: $\neg$SP $\implies$ $\neg$MONO.

$\neg$SP: some member i can misreport $P_i \to P_i'$ and move the outcome from $a = f(P_i, P_{-i})$ to $b = f(P_i', P_{-i})$ with bP_ia.

Construct P_i'' with b ranked first, a ranked second. Consider two transitions to (P_i'', P_{-i}):

- From (P_i, P_{-i}): $D(a, P_i) \subseteq D(a, P_i'')$, so MONO would force $f(P_i'', P_{-i}) = a$
- From (P_i', P_{-i}): $D(b, P_i') \subseteq D(b, P_i'')$ (b is now top-ranked), so MONO would force $f(P_i'', P_{-i}) = b$

Both cannot hold since $a \neq b$. **Contradiction** – so f cannot be monotone.

MONO + ONTO $\Longrightarrow$ PE



Lemma

If an SCF f is MONO and ONTO, then f is PE.

MONO + ONTO $\implies$ PE

Lemma

If an SCF f is MONO and ONTO, then f is PE.

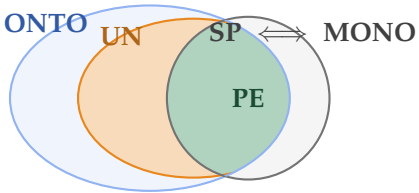


Figure: ONTO $\cap$ (SP $\Leftrightarrow$ MONO) $=$ PE: the entire overlap is Pareto efficient. UN is strictly larger than PE and reaches outside the SP circle, while staying inside ONTO.

Proof of the Lemma

Suppose f is MONO and ONTO but not PE: $\exists\, a, b, P$ with $bP_i a\ \forall i$, yet $f(P) = a$.

Proof of the Lemma

Suppose f is MONO and ONTO but not PE: $\exists\, a, b, P$ with $bP_i a\ \forall i$, yet $f(P) = a$.

- Since f is ONTO, some P' has $f(P') = b$

Proof of the Lemma

Suppose f is MONO and ONTO but not PE: $\exists\, a, b, P$ with $bP_i a\ \forall i$, yet $f(P) = a$.

- Since f is ONTO, some P' has $f(P') = b$
- Construct P'' with b ranked first, a ranked second, for every member

Proof of the Lemma

Suppose f is MONO and ONTO but not PE: $\exists\, a, b, P$ with $bP_i a\ \forall i$, yet $f(P) = a$.

- Since f is ONTO, some P' has $f(P') = b$
- Construct P'' with b ranked first, a ranked second, for every member
- $D(b, P'_i) \subseteq D(b, P''_i)\ \forall i \overset{\text{MONO}}{\Longrightarrow} f(P'') = b$

Proof of the Lemma

Suppose f is MONO and ONTO but not PE: $\exists\, a, b, P$ with $bP_i a\ \forall i$, yet $f(P) = a$.

- Since f is ONTO, some P' has $f(P') = b$
- Construct P'' with b ranked first, a ranked second, for every member
- $D(b, P'_i) \subseteq D(b, P''_i)\ \forall i \overset{\text{MONO}}{\Longrightarrow} f(P'') = b$
- $D(a, P_i) \subseteq D(a, P''_i)\ \forall i \overset{\text{MONO}}{\Longrightarrow} f(P'') = a$

Proof of the Lemma

Suppose f is MONO and ONTO but not PE: $\exists\, a, b, P$ with $bP_i a\ \forall i$, yet $f(P) = a$.

- Since f is ONTO, some P' has $f(P') = b$
- Construct P'' with b ranked first, a ranked second, for every member
- $D(b, P'_i) \subseteq D(b, P''_i)\ \forall i \overset{\text{MONO}}{\Longrightarrow} f(P'') = b$
- $D(a, P_i) \subseteq D(a, P''_i)\ \forall i \overset{\text{MONO}}{\Longrightarrow} f(P'') = a$
- Contradiction – both cannot hold

Proof of the Lemma

Suppose f is MONO and ONTO but not PE: $\exists\, a, b, P$ with $bP_i a\ \forall i$, yet $f(P) = a$.

- Since f is ONTO, some P' has $f(P') = b$
- Construct P'' with b ranked first, a ranked second, for every member
- $D(b, P'_i) \subseteq D(b, P''_i)\ \forall i \overset{\text{MONO}}{\Longrightarrow} f(P'') = b$
- $D(a, P_i) \subseteq D(a, P''_i)\ \forall i \overset{\text{MONO}}{\Longrightarrow} f(P'') = a$
- Contradiction – both cannot hold

Corollary

For a strategyproof f: PE $\iff$ UN $\iff$ ONTO.

Lecture 15: What We Built



- Social choice functions $f : \mathcal{P}^n \to A$; Arrow's impossibility motivates studying SCFs instead of full social orderings

Lecture 15: What We Built



- Social choice functions $f : \mathcal{P}^n \to A$; Arrow's impossibility motivates studying SCFs instead of full social orderings
- Scoring rules (plurality, Borda, ...) and non-scoring rules (runoff, maximin, Copeland); none is Condorcet consistent in general except Copeland and maximin

Lecture 15: What We Built

- Social choice functions $f : \mathcal{P}^n \to A$; Arrow's impossibility motivates studying SCFs instead of full social orderings
- Scoring rules (plurality, Borda, ...) and non-scoring rules (runoff, maximin, Copeland); none is Condorcet consistent in general except Copeland and maximin
- PE $\implies$ UN $\implies$ ONTO, but not conversely; plurality and Copeland are both manipulable under fixed tie-breaking

Lecture 15: What We Built

- Social choice functions $f : \mathcal{P}^n \to A$; Arrow's impossibility motivates studying SCFs instead of full social orderings
- Scoring rules (plurality, Borda, ...) and non-scoring rules (runoff, maximin, Copeland); none is Condorcet consistent in general except Copeland and maximin
- PE $\implies$ UN $\implies$ ONTO, but not conversely; plurality and Copeland are both manipulable under fixed tie-breaking
- **The big equivalence:** SP $\Leftrightarrow$ MONO, and MONO + ONTO $\implies$ PE, so under SP: PE $\iff$ UN $\iff$ ONTO

Lecture 15: What We Built

- Social choice functions $f : \mathcal{P}^n \to A$; Arrow's impossibility motivates studying SCFs instead of full social orderings
- Scoring rules (plurality, Borda, ...) and non-scoring rules (runoff, maximin, Copeland); none is Condorcet consistent in general except Copeland and maximin
- PE $\implies$ UN $\implies$ ONTO, but not conversely; plurality and Copeland are both manipulable under fixed tie-breaking
- **The big equivalence:** SP $\Leftrightarrow$ MONO, and MONO + ONTO $\implies$ PE, so under SP: PE $\iff$ UN $\iff$ ONTO
- **Next:** what happens when we demand SP and ONTO together, for $|A| \geqslant 3$?



End of Lecture 15

See you next class.

Contents

- Lecture 15: The Social Choice Setup and Introduction to Voting
- Lecture 16: The Gibbard–Satterthwaite Theorem



Chunk 1

The Gibbard–Satterthwaite Theorem

(≈ 20–25 minutes)

Can Any Rule Be Trusted?



Our model selection committee now has at least three candidate models on the table, $|A| \geqslant 3$.

Can Any Rule Be Trusted?



Our model selection committee now has at least three candidate models on the table, $|A| \geqslant 3$.

Question

Is there *any* voting rule that is simultaneously strategyproof and onto, without being a dictatorship?

Can Any Rule Be Trusted?

Our model selection committee now has at least three candidate models on the table, $|A| \geqslant 3$.

Question

Is there *any* voting rule that is simultaneously strategyproof and onto, without being a dictatorship?

Theorem (Gibbard 1973, Satterthwaite 1975)

Suppose $|A| \geqslant 3$. f is ONTO and SP if and only if f is dictatorial.

Can Any Rule Be Trusted?

Our model selection committee now has at least three candidate models on the table, $|A| \geqslant 3$.

Question

Is there *any* voting rule that is simultaneously strategyproof and onto, without being a dictatorship?

Theorem (Gibbard 1973, Satterthwaite 1975)

Suppose $|A| \geqslant 3$. f is ONTO and SP if and only if f is dictatorial.

Combined with last lecture's corollary, this is equally a statement about PE or UN in place of ONTO.

So What Just Happened?

- No reasonable voting rule is **truthful** once $|A| \geqslant 3$

So What Just Happened?



- No reasonable voting rule is **truthful** once $|A| \geqslant 3$
- Plurality, Borda, Copeland, maximin – every one of them is manipulable somewhere

So What Just Happened?

- No reasonable voting rule is **truthful** once $|A| \geqslant 3$
- Plurality, Borda, Copeland, maximin – every one of them is manipulable somewhere
- **Crucial caveat**: preferences are **unrestricted** – all $m!$ rankings are feasible reports, so a manipulator always has room to lie

So What Just Happened?

- No reasonable voting rule is **truthful** once $|A| \geqslant 3$
- Plurality, Borda, Copeland, maximin – every one of them is manipulable somewhere
- **Crucial caveat**: preferences are **unrestricted** – all $m!$ rankings are feasible reports, so a manipulator always has room to lie

Reading it for our committee

No matter how cleverly we design the model selection rule, if it always picks a winner (ONTO) and there are 3 or more models, some member somewhere has an incentive to misreport – unless one member is simply a dictator.

Points to Note Before the Proof

1. $|A| = 2$: GS does *not* hold. Plurality with a fixed tie-break is SP, ONTO, and non-dictatorial.

Points to Note Before the Proof

1. $|A| = 2$: GS does *not* hold. Plurality with a fixed tie-break is SP, ONTO, and non-dictatorial.
2. The domain is all of $\mathcal{P}$: every permutation of models is a feasible report. A restricted domain can escape GS – this is exactly where we are headed next chunk.

Points to Note Before the Proof

1. $|A| = 2$: GS does *not* hold. Plurality with a fixed tie-break is SP, ONTO, and non-dictatorial.
2. The domain is all of $\mathcal{P}$: every permutation of models is a feasible report. A restricted domain can escape GS – this is exactly where we are headed next chunk.
3. **Indifference**: in general GS fails once ties in preference are allowed, unless the specific constructions used in the proof remain possible.

Points to Note Before the Proof

1. $|A| = 2$: GS does *not* hold. Plurality with a fixed tie-break is SP, ONTO, and non-dictatorial.
2. The domain is all of $\mathcal{P}$: every permutation of models is a feasible report. A restricted domain can escape GS – this is exactly where we are headed next chunk.
3. **Indifference**: in general GS fails once ties in preference are allowed, unless the specific constructions used in the proof remain possible.
4. **Cardinalization**: GS still holds under cardinal preferences, as long as every ordinal ranking remains achievable.

Proof Strategy

- Follow a direct approach (Sen, 2001)

Proof Strategy

- Follow a direct approach (Sen, 2001)
- First prove the theorem for $n = 2$ members

Proof Strategy

- Follow a direct approach (Sen, 2001)
- First prove the theorem for $n = 2$ members
- Then extend to $n \geqslant 3$ by induction on the number of members

Proof Strategy

- Follow a direct approach (Sen, 2001)
- First prove the theorem for $n = 2$ members
- Then extend to $n \geqslant 3$ by induction on the number of members

Lemma (Top Choice)

Suppose $|A| \geqslant 3$, $N = \{1, 2\}$, *and* f *is ONTO and SP. Then for every profile* P,

$$f(P) \in \{P_1(1), P_2(1)\}.$$

Proof Strategy

- Follow a direct approach (Sen, 2001)
- First prove the theorem for $n = 2$ members
- Then extend to $n \geqslant 3$ by induction on the number of members

Lemma (Top Choice)

Suppose $|A| \geqslant 3$, $N = \{1, 2\}$, *and* f *is ONTO and SP. Then for every profile* P,

$$f(P) \in \{P_1(1), P_2(1)\}.$$

In words: with just two members, the committee's decision can never be a model that neither of them top-ranked.

Proof of the Lemma: Setting Up



Case 1: $P_1(1) = P_2(1)$. Then UN (equivalent to ONTO under SP) forces $f(P) = P_1(1)$ directly.

Proof of the Lemma: Setting Up



Case 1: $P_1(1) = P_2(1)$. Then UN (equivalent to ONTO under SP) forces $f(P) = P_1(1)$ directly.

Case 2: $P_1(1) = a \neq b = P_2(1)$. Suppose for contradiction $f(P) = c \neq a, b$ (needs $|A| \geqslant 3$).

Proof of the Lemma: Setting Up

Case 1: $P_1(1) = P_2(1)$. Then UN (equivalent to ONTO under SP) forces $f(P) = P_1(1)$ directly.

Case 2: $P_1(1) = a \neq b = P_2(1)$. Suppose for contradiction $f(P) = c \neq a, b$ (needs $|A| \geqslant 3$).

P_1	P_2	P_1	P_2'	P_1'	P_2'	P_1'	P_2
a	b	a	b	a	b	a	b
.	.	.	a	b	a	b	.
.	.	.	.	.	.	.	.

$f(P_1, P_2) = c \ (\neq a, b)$

Proof of the Lemma: The Contradiction

P_1	P_2	P_1	P_2'	P_1'	P_2'	P_1'	P_2
a	b	a	b	a	b	a	b
.	.	.	a	b	a	b	.
.	.	.	.	.	.	.	.

$$f(P_1, P_2) = c \ (\neq a, b)$$

- In (P_1, P_2'), every model except b is Pareto dominated by a, so $f(P_1, P_2') \in \{a, b\}$

Proof of the Lemma: The Contradiction

P_1	P_2	P_1	P_2'	P_1'	P_2'	P_1'	P_2
a	b	a	b	a	b	a	b
.	.	.	a	b	a	b	.
.	.	.	.	.	.	.	.

$f(P_1, P_2) = c \ (\neq a, b)$

- In (P_1, P_2'), every model except b is Pareto dominated by a, so $f(P_1, P_2') \in \{a, b\}$
- If $f(P_1, P_2') = b$, member 2 profitably manipulates from P_2 to P_2' – so $f(P_1, P_2') = a$

Proof of the Lemma: The Contradiction

P_1	P_2	P_1	P_2'	P_1'	P_2'	P_1'	P_2
a	b	a	b	a	b	a	b
.	.	.	a	b	a	b	.
.	.	.	.	.	.	.	.

$f(P_1, P_2) = c \ (\neq a, b)$

- In (P_1, P_2'), every model except b is Pareto dominated by a, so $f(P_1, P_2') \in \{a, b\}$
- If $f(P_1, P_2') = b$, member 2 profitably manipulates from P_2 to P_2' – so $f(P_1, P_2') = a$
- Symmetrically, $f(P_1', P_2) = b$

Proof of the Lemma: The Contradiction

P_1	P_2	P_1	P_2'	P_1'	P_2'	P_1'	P_2
a	b	a	b	a	b	a	b
.	.	.	a	b	a	b	.
.	.	.	.	.	.	.	.

$f(P_1, P_2) = c \ (\neq a, b)$

- In (P_1, P_2'), every model except b is Pareto dominated by a, so $f(P_1, P_2') \in \{a, b\}$
- If $f(P_1, P_2') = b$, member 2 profitably manipulates from P_2 to P_2' – so $f(P_1, P_2') = a$
- Symmetrically, $f(P_1', P_2) = b$

Now apply monotonicity:

- a's relative position weakly improved from (P_1, P_2') to $(P_1', P_2') \implies f(P_1', P_2') = a$
- b's relative position weakly improved from (P_1', P_2) to $(P_1', P_2') \implies f(P_1', P_2') = b$

Proof of the Lemma: The Contradiction

P_1	P_2	P_1	P_2'	P_1'	P_2'	P_1'	P_2
a	b	a	b	a	b	a	b
.	.	.	a	b	a	b	.
.	.	.	.	.	.	.	.

$f(P_1, P_2) = c \ (\neq a, b)$

- In (P_1, P_2'), every model except b is Pareto dominated by a, so $f(P_1, P_2') \in \{a, b\}$
- If $f(P_1, P_2') = b$, member 2 profitably manipulates from P_2 to P_2' – so $f(P_1, P_2') = a$
- Symmetrically, $f(P_1', P_2) = b$

Now apply monotonicity:

- a's relative position weakly improved from (P_1, P_2') to $(P_1', P_2') \implies f(P_1', P_2') = a$
- b's relative position weakly improved from (P_1', P_2) to $(P_1', P_2') \implies f(P_1', P_2') = b$

$a \neq b$ – contradiction. So $f(P) \in \{a, b\}$ after all, proving the lemma.

Chunk 2

Proof of the Two-Player GS Theorem

(≈ 25–30 minutes)

The Two-Player GS Theorem

Lemma

Suppose $|A| \geqslant 3$, $N = \{1, 2\}$, f *is ONTO and SP.*

- *Let* $P : P_1(1) = a \neq b = P_2(1)$, *and* $P' : P'_1(1) = c,\ P'_2(1) = d$
- *If* $f(P) = a$, *then* $f(P') = c$
- *If* $f(P) = b$, *then* $f(P') = d$

The Two-Player GS Theorem

Lemma

Suppose $|A| \geqslant 3$, $N = \{1, 2\}$, f *is ONTO and SP.*

- *Let* $P : P_1(1) = a \neq b = P_2(1)$, *and* $P' : P'_1(1) = c,\ P'_2(1) = d$
- *If* $f(P) = a$, *then* $f(P') = c$
- *If* $f(P) = b$, *then* $f(P') = d$

This says: once we know member 1's top choice wins at one profile with distinct top choices, member 1's top choice wins at *every* such profile – member 1 is a dictator.

The Two-Player GS Theorem

Lemma

Suppose $|A| \geqslant 3$, $N = \{1, 2\}$, f *is ONTO and SP.*

- *Let* $P : P_1(1) = a \neq b = P_2(1)$, *and* $P' : P'_1(1) = c,\ P'_2(1) = d$
- *If* $f(P) = a$, *then* $f(P') = c$
- *If* $f(P) = b$, *then* $f(P') = d$

This says: once we know member 1's top choice wins at one profile with distinct top choices, member 1's top choice wins at *every* such profile – member 1 is a dictator.

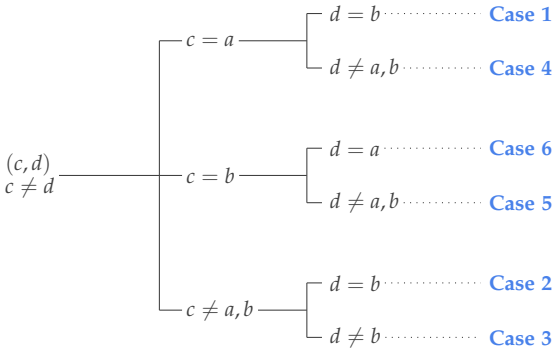
Proof: if $c = d$, unanimity settles it immediately. So assume $c \neq d$; six exhaustive cases remain, and by symmetry it suffices to prove $f(P) = a \implies f(P') = c$ in each.

The Six Cases

case	c	d
1	a	b
2	$\neq a, b$	b
3	$\neq a, b$	$\neq b$
4	a	$\neq a, b$
5	b	$\neq a, b$
6	b	a

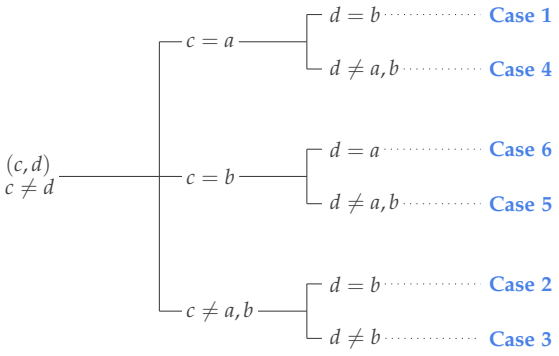
The Six Cases

case	c	d
1	a	b
2	$\neq a, b$	b
3	$\neq a, b$	$\neq b$
4	a	$\neq a, b$
5	b	$\neq a, b$
6	b	a



The Six Cases

case	c	d
1	a	b
2	$\neq a, b$	b
3	$\neq a, b$	$\neq b$
4	a	$\neq a, b$
5	b	$\neq a, b$
6	b	a



Why six and not nine: $c = d$ is already settled by unanimity, which kills $d = a$ under $c = a$ and $d = b$ under $c = b$. In the third branch all of $d = a$ and $d \neq a, b$ behave alike, so they merge into Case 3.

Case 1: $c = a, d = b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	a	b	a	b
.	.	.	.	b	a
.	.	.	.	.	.

Case 1: $c = a, d = b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	a	b	a	b
.	.	.	.	b	a
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{a, b\}$. Suppose for contradiction $f(P') = b$.

Case 1: $c = a, d = b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	a	b	a	b
.	.	.	.	b	a
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{a, b\}$. Suppose for contradiction $f(P') = b$.

- a's position unchanged, $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$: MONO $\implies f(\hat{P}) = a$
- b's position unchanged, $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$: MONO $\implies f(\hat{P}) = b$

Case 1: $c = a, d = b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	a	b	a	b
.	.	.	.	b	a
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{a, b\}$. Suppose for contradiction $f(P') = b$.

- a's position unchanged, $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$: MONO $\implies f(\hat{P}) = a$
- b's position unchanged, $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$: MONO $\implies f(\hat{P}) = b$

Contradiction. So $f(P') = a$.

Case 2: $c \neq a, b,\ d = b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	P_2
a	b	c	b	c	b
.	.	.	.	a	.
.	.	.	.	.	.

Case 2: $c \neq a, b,\ d = b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	P_2
a	b	c	b	c	b
.	.	.	.	a	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{c, b\}$. Suppose $f(P') = b$.

Case 2: $c \neq a, b,\ d = b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	P_2
a	b	c	b	c	b
.	.	.	.	a	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{c, b\}$. Suppose $f(P') = b$.

$(P'_1, P'_2) \to (\hat{P}_1, P_2)$ follows Case 1's pattern, so $f(\hat{P}_1, P_2) = b$. But then member 1 can manipulate $\hat{P}_1 \to P_1$, since $a\ \hat{P}_1\ b$ and this flips the outcome to a – contradicting SP.

Case 2: $c \neq a, b,\ d = b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	P_2
a	b	c	b	c	b
.	.	.	.	a	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{c, b\}$. Suppose $f(P') = b$.

$(P'_1, P'_2) \to (\hat{P}_1, P_2)$ follows Case 1's pattern, so $f(\hat{P}_1, P_2) = b$. But then member 1 can manipulate $\hat{P}_1 \to P_1$, since $a\ \hat{P}_1\ b$ and this flips the outcome to a – contradicting SP.

So $f(P') = c$.

Case 3: $c \neq a, b,\ d \neq b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	c	d	c	b
.	.	.	.	.	.
.	.	.	.	.	.

Case 3: $c \neq a, b,\ d \neq b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	c	d	c	b
.	.	.	.	.	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{c, d\}$. Suppose $f(P') = d$.

Case 3: $c \neq a, b,\ d \neq b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	c	d	c	b
.	.	.	.	.	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{c, d\}$. Suppose $f(P') = d$.

- $(P_1', P_2') \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 2: $f(P') = d \implies f(\hat{P}) = b$
- $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 1: $f(P) = a \implies f(\hat{P}) = c$

Case 3: $c \neq a, b,\ d \neq b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	c	d	c	b
.	.	.	.	.	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{c, d\}$. Suppose $f(P') = d$.

- $(P_1', P_2') \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 2: $f(P') = d \implies f(\hat{P}) = b$
- $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 1: $f(P) = a \implies f(\hat{P}) = c$

Contradiction ($b \neq c$). So $f(P') = c$.

Case 4: $c = a,\ d \neq a, b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	a	d	a	b
.	.	.	.	.	.
.	.	.	.	.	.

Case 4: $c = a,\ d \neq a, b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	a	d	a	b
.	.	.	.	.	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{a, d\}$. Suppose $f(P') = d$.

Case 4: $c = a,\ d \neq a, b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	a	d	a	b
.	.	.	.	.	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{a, d\}$. Suppose $f(P') = d$.

- $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 2: $f(P') = d \implies f(\hat{P}) = b$
- $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 1: $f(P) = a \implies f(\hat{P}) = a$

Case 4: $c = a,\ d \neq a, b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	a	d	a	b
.	.	.	.	.	.
.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{a, d\}$. Suppose $f(P') = d$.

- $(P_1', P_2') \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 2: $f(P') = d \implies f(\hat{P}) = b$
- $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 1: $f(P) = a \implies f(\hat{P}) = a$

Contradiction. So $f(P') = a$.

Case 5: $c = b,\ d \neq a, b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$	P_1	$\hat{P}_2$
a	b	b	d	b	d	a	d
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Case 5: $c = b,\ d \neq a, b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$	P_1	$\hat{P}_2$
a	b	b	d	b	d	a	d
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{b, d\}$. Suppose $f(P') = d$.

Case 5: $c = b,\ d \neq a, b$

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$	P_1	$\hat{P}_2$
a	b	b	d	b	d	a	d
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{b, d\}$. Suppose $f(P') = d$.

- $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 1: $f(P') = d \implies f(\hat{P}) = d$
- $(P_1, P_2) \to (P_1, \hat{P}_2)$ is Case 4: $f(P) = a \implies f(P_1, \hat{P}_2) = a$
- $(P_1, \hat{P}_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2: $f(P_1, \hat{P}_2) = a \implies f(\hat{P}) = b$

Case 5: $c = b,\ d \neq a, b$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$	P_1	$\hat{P}_2$
a	b	b	d	b	d	a	d
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

By Lemma (Top Choice), $f(P') \in \{b, d\}$. Suppose $f(P') = d$.

- $(P_1', P_2') \to (\hat{P}_1, \hat{P}_2)$ is Case 1: $f(P') = d \implies f(\hat{P}) = d$
- $(P_1, P_2) \to (P_1, \hat{P}_2)$ is Case 4: $f(P) = a \implies f(P_1, \hat{P}_2) = a$
- $(P_1, \hat{P}_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2: $f(P_1, \hat{P}_2) = a \implies f(\hat{P}) = b$

Contradiction ($b \neq d$). So $f(P') = b$.

Case 6: $c = b,\ d = a$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	b	a	e	b
.	.	.	.	.	.
.	.	.	.	.	.

where $e \neq a, b$; assume for contradiction $f(P') = a$.

Case 6: $c = b,\ d = a$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	b	a	e	b
.	.	.	.	.	.
.	.	.	.	.	.

where $e \neq a, b$; assume for contradiction $f(P') = a$.

- $(P_1', P_2') \to (\hat{P}_1, \hat{P}_2)$ is Case 5 for member 2: $f(P') = a \implies f(\hat{P}) = b$
- $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 1: $f(P) = a \implies f(\hat{P}) = e$

Case 6: $c = b,\ d = a$

P_1	P_2	P_1'	P_2'	$\hat{P}_1$	$\hat{P}_2$
a	b	b	a	e	b
.	.	.	.	.	.
.	.	.	.	.	.

where $e \neq a, b$; assume for contradiction $f(P') = a$.

- $(P_1', P_2') \to (\hat{P}_1, \hat{P}_2)$ is Case 5 for member 2: $f(P') = a \implies f(\hat{P}) = b$
- $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for member 1: $f(P) = a \implies f(\hat{P}) = e$

Contradiction. So $f(P') = b$. All six cases done – **member 1 is a dictator** whenever the two members' top choices differ. For $n \geqslant 3$, induct on the number of members (Sen, 2001, "A direct proof of the Gibbard–Satterthwaite theorem," *Economics Letters*).

In-Class Quiz 1

Open the Google Form now.

Questions refer only to what was just discussed –
you cannot answer them without having been in class.

Chunk 3

Domain Restrictions and Single-Peaked Preferences

(≈ 20–25 minutes)

Escaping GS: Restrict the Domain

GS is stated over the full domain $f : \mathcal{P}^n \to A$, where every one of the $m!$ rankings is a feasible report.

Escaping GS: Restrict the Domain

GS is stated over the full domain $f : \mathcal{P}^n \to A$, where every one of the $m!$ rankings is a feasible report.

Strategyproofness, restated: for every member i, every $P_i, P_i' \in \mathcal{P}$, every $P_{-i} \in \mathcal{P}^{n-1}$,

$$f(P_i, P_{-i}) \; P_i \, f(P_i', P_{-i}) \quad \text{or} \quad f(P_i, P_{-i}) = f(P_i', P_{-i}).$$

Escaping GS: Restrict the Domain

GS is stated over the full domain $f : \mathcal{P}^n \to A$, where every one of the $m!$ rankings is a feasible report.

Strategyproofness, restated: for every member i, every $P_i, P_i' \in \mathcal{P}$, every $P_{-i} \in \mathcal{P}^{n-1}$,

$$f(P_i, P_{-i}) \; P_i \, f(P_i', P_{-i}) \quad \text{or} \quad f(P_i, P_{-i}) = f(P_i', P_{-i}).$$

If we shrink the feasible set from $\mathcal{P}$ to $\mathcal{S} \subset \mathcal{P}$:

- any f strategyproof on $\mathcal{P}$ stays strategyproof on $\mathcal{S}$
- but on the smaller domain, new strategyproof rules can appear – manipulators simply have fewer misreports to try

Three Major Domain Restrictions

1. **Single-peaked preferences**: alternatives sit on a line, each agent's ranking rises to one peak and falls away on both sides

Three Major Domain Restrictions

1. **Single-peaked preferences**: alternatives sit on a line, each agent's ranking rises to one peak and falls away on both sides
2. **Divisible goods allocation**: preferences are over shares of a divisible resource

Three Major Domain Restrictions

1. **Single-peaked preferences**: alternatives sit on a line, each agent's ranking rises to one peak and falls away on both sides
2. **Divisible goods allocation**: preferences are over shares of a divisible resource
3. **Quasi-linear preferences**: utility is linear in money plus preferences over other attributes – payments enable truthful, non-dictatorial design

Three Major Domain Restrictions



1. **Single-peaked preferences**: alternatives sit on a line, each agent's ranking rises to one peak and falls away on both sides
2. **Divisible goods allocation**: preferences are over shares of a divisible resource
3. **Quasi-linear preferences**: utility is linear in money plus preferences over other attributes – payments enable truthful, non-dictatorial design

We focus on the first, since it has a clean, classical theory.

Example: A Shared GPU Power Cap



Several teams share one training cluster. The cluster-wide GPU power cap is set once, and everyone lives with it.

- Each team has an ideal cap t_i^*: too low throttles training, too high increases the cost

Example: A Shared GPU Power Cap



Several teams share one training cluster. The cluster-wide GPU power cap is set once, and everyone lives with it.

- Each team has an ideal cap t_i^*: too low throttles training, too high increases the cost
- A team's satisfaction falls off the further the shared cap sits from t_i^*, *in either direction*

Example: A Shared GPU Power Cap

Several teams share one training cluster. The cluster-wide GPU power cap is set once, and everyone lives with it.

- Each team has an ideal cap t_i^*: too low throttles training, too high increases the cost
- A team's satisfaction falls off the further the shared cap sits from t_i^*, *in either direction*

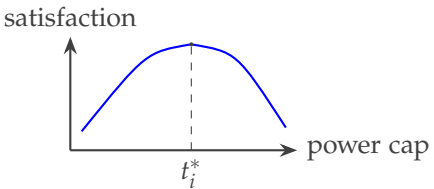


Figure: Single-peaked at t_i^*.

Other Classic Single-Peaked Examples



- **Facility location**: siting a school, hospital, or post office along a line
- **Political ideology**: positions on a left–center–right spectrum
- **Temperature control in an office**: each employee has an ideal desired temperature – cooler or hotter reduces their desirability monotonically

Other Classic Single-Peaked Examples

- **Facility location**: siting a school, hospital, or post office along a line
- **Political ideology**: positions on a left–center–right spectrum
- **Temperature control in an office**: each employee has an ideal desired temperature – cooler or hotter reduces their desirability monotonically

In every case there is one **common order** $<$ over the alternatives (think: the real line), and each agent's ranking is single-peaked with respect to it. We restrict attention to **one-dimensional** single-peakedness.

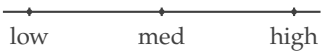
Definition (Single-peaked preference)

P_i is single-peaked w.r.t. common order $<$ if, writing $P_i(1)$ for the peak:

1. $\forall\, b, c$ with $b < c \leqslant P_i(1)$: cP_ib (left monotonicity)
2. $\forall\, b, c$ with $P_i(1) \leqslant b < c$: bP_ic (right monotonicity)

How Big Is the Restriction?

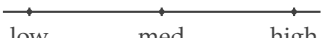
Three candidate power-cap settings, low $<$ medium $<$ high. All $3! = 6$ strict orderings:



low	med	med	high	low	high
med	low	high	med	high	low
high	high	low	low	med	med

How Big Is the Restriction?

Three candidate power-cap settings, low $<$ medium $<$ high. All $3! = 6$ strict orderings:



low	med	med	high	low	high
med	low	high	med	high	low
high	high	low	low	med	med

The two orderings in red – *low* $\succ$ *high* $\succ$ *med*, and its mirror – are not single-peaked: satisfaction rises, falls, then rises again. With m alternatives, single-peaked orderings number only 2^{m-1} out of $m!$ total – a genuine restriction.

Why 2^{m-1}?

Fix the peak at a_j. Single-peakedness forces the order *within* each side:

- left of the peak: $a_{j-1} \succ a_{j-2} \succ \cdots \succ a_1$
- right of the peak: $a_{j+1} \succ a_{j+2} \succ \cdots \succ a_m$

Why 2^{m-1}?

Fix the peak at a_j. Single-peakedness forces the order *within* each side:

- left of the peak: $a_{j-1} \succ a_{j-2} \succ \cdots \succ a_1$
- right of the peak: $a_{j+1} \succ a_{j+2} \succ \cdots \succ a_m$

Closer to the peak is always better, so neither block has any freedom left.

Why 2^{m-1}?

Fix the peak at a_j. Single-peakedness forces the order *within* each side:

- left of the peak: $a_{j-1} \succ a_{j-2} \succ \cdots \succ a_1$
- right of the peak: $a_{j+1} \succ a_{j+2} \succ \cdots \succ a_m$

Closer to the peak is always better, so neither block has any freedom left.
A ranking is then just an **interleaving** of the two blocks below the peak:

$$\binom{(j-1)+(m-j)}{j-1} = \binom{m-1}{j-1}.$$

Why 2^{m-1}?

Fix the peak at a_j. Single-peakedness forces the order *within* each side:

- left of the peak: $a_{j-1} \succ a_{j-2} \succ \cdots \succ a_1$
- right of the peak: $a_{j+1} \succ a_{j+2} \succ \cdots \succ a_m$

Closer to the peak is always better, so neither block has any freedom left.
A ranking is then just an **interleaving** of the two blocks below the peak:

$$\binom{(j-1)+(m-j)}{j-1} = \binom{m-1}{j-1}.$$

Summing over every possible peak:

$$\sum_{j=1}^{m} \binom{m-1}{j-1} = 2^{m-1}.$$

Why 2^{m-1}?

Fix the peak at a_j. Single-peakedness forces the order *within* each side:

- left of the peak: $a_{j-1} \succ a_{j-2} \succ \cdots \succ a_1$
- right of the peak: $a_{j+1} \succ a_{j+2} \succ \cdots \succ a_m$

Closer to the peak is always better, so neither block has any freedom left.
A ranking is then just an **interleaving** of the two blocks below the peak:

$$\binom{(j-1)+(m-j)}{j-1} = \binom{m-1}{j-1}.$$

Summing over every possible peak:

$$\sum_{j=1}^{m} \binom{m-1}{j-1} = 2^{m-1}.$$

m	2^{m-1}	$m!$	share
3	4	6	67%
4	8	24	33%
5	16	120	13%
6	32	720	4.4%
8	128	40,320	0.3%

Why 2^{m-1}?

Fix the peak at a_j. Single-peakedness forces the order *within* each side:

- left of the peak: $a_{j-1} \succ a_{j-2} \succ \cdots \succ a_1$
- right of the peak: $a_{j+1} \succ a_{j+2} \succ \cdots \succ a_m$

Closer to the peak is always better, so neither block has any freedom left.
A ranking is then just an **interleaving** of the two blocks below the peak:

$$\binom{(j-1)+(m-j)}{j-1} = \binom{m-1}{j-1}.$$

Summing over every possible peak:

$$\sum_{j=1}^{m} \binom{m-1}{j-1} = 2^{m-1}.$$

m	2^{m-1}	$m!$	share
3	4	6	67%
4	8	24	33%
5	16	120	13%
6	32	720	4.4%
8	128	40,320	0.3%

By $m = 8$ more than 99.7% of the domain is gone – a manipulator has very little left to lie with.

Escaping GS: Leftmost-Peak Rule

Let $\mathcal{S}$ be single-peaked preferences, $f : \mathcal{S}^n \to A$.

Escaping GS: Leftmost-Peak Rule

Let $\mathcal{S}$ be single-peaked preferences, $f : \mathcal{S}^n \to A$.

Question

How does restricting to $\mathcal{S}$ let us escape the Gibbard–Satterthwaite trap?

Escaping GS: Leftmost-Peak Rule

Let $\mathcal{S}$ be single-peaked preferences, $f : \mathcal{S}^n \to A$.

Question

How does restricting to $\mathcal{S}$ let us escape the Gibbard–Satterthwaite trap?

Answer

Let f pick the *leftmost* reported peak (the lowest requested power cap). The team with the leftmost peak already gets its top choice – no reason to lie. Any other team can only move the outcome by reporting a peak even further left than the current winner – which, by single-peakedness, is strictly worse for them.

Escaping GS: Leftmost-Peak Rule

Let $\mathcal{S}$ be single-peaked preferences, $f : \mathcal{S}^n \to A$.

Question

How does restricting to $\mathcal{S}$ let us escape the Gibbard–Satterthwaite trap?

Answer

Let f pick the *leftmost* reported peak (the lowest requested power cap). The team with the leftmost peak already gets its top choice – no reason to lie. Any other team can only move the outcome by reporting a peak even further left than the current winner – which, by single-peakedness, is strictly worse for them.

The same argument works for any fixed k-th peak from the left – so the rightmost-peak rule is strategyproof too, and so is the **median**-peak rule with $k = \lceil n/2 \rceil$.



Chunk 4

The Median Voter Theorem

(≈ 25–30 minutes)

Median Voter Social Choice Functions

Definition (Median voter SCF)

$f : \mathcal{S}^n \to A$ is a **median voter SCF** if there exist **phantom peaks** $B = \{y_1, \ldots, y_{n-1}\}$, fixed and independent of P, such that for every profile,

$$f(P) = \text{median}\big(B,\ \text{peaks}(P)\big).$$

Median Voter Social Choice Functions

Definition (Median voter SCF)

$f : \mathcal{S}^n \to A$ is a **median voter SCF** if there exist **phantom peaks** $B = \{y_1, \ldots, y_{n-1}\}$, fixed and independent of P, such that for every profile,

$$f(P) = \text{median}\big(B,\ \text{peaks}(P)\big).$$

Think of B as $n-1$ fixed reference power-caps baked permanently into the policy, alongside the n teams' actual reported peaks – median of all $2n-1$ numbers.

Median Voter Social Choice Functions

Definition (Median voter SCF)

$f : \mathcal{S}^n \to A$ is a **median voter SCF** if there exist **phantom peaks** $B = \{y_1, \ldots, y_{n-1}\}$, fixed and independent of P, such that for every profile,

$$f(P) = \text{median}\big(B,\ \text{peaks}(P)\big).$$

Think of B as $n-1$ fixed reference power-caps baked permanently into the policy, alongside the n teams' actual reported peaks – median of all $2n-1$ numbers.

- All phantoms at the leftmost value $\implies f \equiv$ leftmost-peak rule
- All phantoms at the rightmost value $\implies f \equiv$ rightmost-peak rule
- Different phantom placements sweep out the *entire* spectrum of median voter SCFs

Moulin: Median Voter SCFs Are SP

Theorem (Moulin, 1980)

Every median voter SCF is strategyproof.

Moulin: Median Voter SCFs Are SP

Theorem (Moulin, 1980)

Every median voter SCF is strategyproof.

Proof: fix profile P, outcome $f(P) = a$. Suppose team i considers misreporting peak $b \to c$, moving the outcome to a'.

Moulin: Median Voter SCFs Are SP

Theorem (Moulin, 1980)

Every median voter SCF is strategyproof.

Proof: fix profile P, outcome $f(P) = a$. Suppose team i considers misreporting peak $b \to c$, moving the outcome to a'.

- $a > b$, $a > c$: c is on the same side of a as b – the count of peaks on each side of a is unchanged, so $a' = a$

Moulin: Median Voter SCFs Are SP

Theorem (Moulin, 1980)

Every median voter SCF is strategyproof.

Proof: fix profile P, outcome $f(P) = a$. Suppose team i considers misreporting peak $b \to c$, moving the outcome to a'.

- $a > b$, $a > c$: c is on the same side of a as b – the count of peaks on each side of a is unchanged, so $a' = a$
- $a > b$, $a < c$: peaks right of a increase by one, so the median shifts right, $a < a'$ – but by single-peakedness $aP_i a'$

Moulin: Median Voter SCFs Are SP

Theorem (Moulin, 1980)

Every median voter SCF is strategyproof.

Proof: fix profile P, outcome $f(P) = a$. Suppose team i considers misreporting peak $b \to c$, moving the outcome to a'.

- $a > b$, $a > c$: c is on the same side of a as b – the count of peaks on each side of a is unchanged, so $a' = a$
- $a > b$, $a < c$: peaks right of a increase by one, so the median shifts right, $a < a'$ – but by single-peakedness $aP_i a'$
- $a < b$, $a < c$ and $a < b$, $a > c$: symmetric to the two cases above

Moulin: Median Voter SCFs Are SP

Theorem (Moulin, 1980)

Every median voter SCF is strategyproof.

Proof: fix profile P, outcome $f(P) = a$. Suppose team i considers misreporting peak $b \to c$, moving the outcome to a'.

- $a > b$, $a > c$: c is on the same side of a as b – the count of peaks on each side of a is unchanged, so $a' = a$
- $a > b$, $a < c$: peaks right of a increase by one, so the median shifts right, $a < a'$ – but by single-peakedness $aP_i a'$
- $a < b$, $a < c$ and $a < b$, $a > c$: symmetric to the two cases above

In every case, misreporting either leaves the outcome unchanged or strictly hurts team i. Note: the *mean* does not have this property.

When Is a Median Voter SCF PE?

Proposition

Let $p_{\min}, p_{\max}$ be the leftmost and rightmost reported peaks. Then f is PE if and only if $f(P) \in [p_{\min}, p_{\max}]$.

When Is a Median Voter SCF PE?

Proposition

Let $p_{\min}, p_{\max}$ be the leftmost and rightmost reported peaks. Then f is PE if and only if $f(P) \in [p_{\min}, p_{\max}]$.

Proof ($\Longrightarrow$): if $f(P) < p_{\min}$, every team prefers $p_{\min}$ over $f(P)$ (both lie on the same side of every team's peak), so $f(P)$ is Pareto dominated – contradiction. Symmetric argument if $f(P) > p_{\max}$.

When Is a Median Voter SCF PE?

Proposition

Let $p_{\min}, p_{\max}$ be the leftmost and rightmost reported peaks. Then f is PE if and only if $f(P) \in [p_{\min}, p_{\max}]$.

Proof ($\Longrightarrow$): if $f(P) < p_{\min}$, every team prefers $p_{\min}$ over $f(P)$ (both lie on the same side of every team's peak), so $f(P)$ is Pareto dominated – contradiction. Symmetric argument if $f(P) > p_{\max}$.

Proof ($\Longleftarrow$): if $f(P) \in [p_{\min}, p_{\max}]$, let $aP_i f(P)$ for all i for contradiction. If $a < f(P)$, the team at $p_{\max}$ prefers $f(P)$ to a (same side of its peak) – contradiction; symmetric if $a > f(P)$.

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Proof: PE $\implies$ UN $\implies$ ONTO always holds; we need ONTO $\implies$ PE under SP.

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Proof: PE $\implies$ UN $\implies$ ONTO always holds; we need ONTO $\implies$ PE under SP.

- Suppose SP, ONTO, not PE: $\exists\, a, b$ with $aP_i b\ \forall i$, yet $f(P) = b$

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Proof: PE $\implies$ UN $\implies$ ONTO always holds; we need ONTO $\implies$ PE under SP.

- Suppose SP, ONTO, not PE: $\exists\, a, b$ with $aP_i b\ \forall i$, yet $f(P) = b$
- Single-peakedness gives a neighbour c of b with $cP_i b\ \forall i$ (possibly $c = a$)

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Proof: PE $\implies$ UN $\implies$ ONTO always holds; we need ONTO $\implies$ PE under SP.

- Suppose SP, ONTO, not PE: $\exists\, a, b$ with $aP_ib\ \forall i$, yet $f(P) = b$
- Single-peakedness gives a neighbour c of b with $cP_ib\ \forall i$ (possibly $c = a$)
- ONTO $\implies \exists P'$ with $f(P') = c$

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Proof: PE $\implies$ UN $\implies$ ONTO always holds; we need ONTO $\implies$ PE under SP.

- Suppose SP, ONTO, not PE: $\exists\, a, b$ with $aP_ib\ \forall i$, yet $f(P) = b$
- Single-peakedness gives a neighbour c of b with $cP_ib\ \forall i$ (possibly $c = a$)
- ONTO $\implies \exists P'$ with $f(P') = c$
- Construct P'': peak at c, second choice b, for every team

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Proof: PE $\implies$ UN $\implies$ ONTO always holds; we need ONTO $\implies$ PE under SP.

- Suppose SP, ONTO, not PE: $\exists\, a, b$ with $aP_i b\ \forall i$, yet $f(P) = b$
- Single-peakedness gives a neighbour c of b with $cP_i b\ \forall i$ (possibly $c = a$)
- ONTO $\implies$ $\exists P'$ with $f(P') = c$
- Construct P'': peak at c, second choice b, for every team
- $D(b, P_i) \subseteq D(b, P_i'') \overset{\text{MONO}}{\implies} f(P'') = b$; $D(c, P_i') \subseteq D(c, P_i'') \overset{\text{MONO}}{\implies} f(P'') = c$

Equivalence Under Single-Peakedness

Theorem

Let $f : \mathcal{S}^n \to A$ be SP. Then ONTO $\iff$ UN $\iff$ PE.

Proof: PE $\implies$ UN $\implies$ ONTO always holds; we need ONTO $\implies$ PE under SP.

- Suppose SP, ONTO, not PE: $\exists\, a, b$ with $aP_i b\ \forall i$, yet $f(P) = b$
- Single-peakedness gives a neighbour c of b with $cP_i b\ \forall i$ (possibly $c = a$)
- ONTO $\implies$ $\exists P'$ with $f(P') = c$
- Construct P'': peak at c, second choice b, for every team
- $D(b, P_i) \subseteq D(b, P_i'') \overset{\text{MONO}}{\implies} f(P'') = b$; $D(c, P_i') \subseteq D(c, P_i'') \overset{\text{MONO}}{\implies} f(P'') = c$
- Contradiction, completing the proof

Anonymity



We now want *non-dictatorial* strategyproof rules, so we need one more property.

Anonymity

We now want *non-dictatorial* strategyproof rules, so we need one more property.

Definition (Anonymity)

f is **anonymous** (ANON) if for every profile P and every permutation σ of the agents, $f(P^\sigma) = f(P)$ – the outcome does not depend on *who* reported which preference.

Anonymity

We now want *non-dictatorial* strategyproof rules, so we need one more property.

Definition (Anonymity)

f is **anonymous** (ANON) if for every profile P and every permutation σ of the agents, $f(P^\sigma) = f(P)$ – the outcome does not depend on *who* reported which preference.

Example: three teams sharing the cluster, $N = \{1, 2, 3\}$, $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$:

P_1	P_2	P_3	P_1^σ	P_2^σ	P_3^σ
a	b	b	b	a	b
b	a	c	c	b	a
c	c	a	a	c	c

Anonymity

We now want *non-dictatorial* strategyproof rules, so we need one more property.

Definition (Anonymity)

f is **anonymous** (ANON) if for every profile P and every permutation σ of the agents, $f(P^\sigma) = f(P)$ – the outcome does not depend on *who* reported which preference.

Example: three teams sharing the cluster, $N = \{1, 2, 3\}$, $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$:

P_1	P_2	P_3	P_1^σ	P_2^σ	P_3^σ
a	b	b	b	a	b
b	a	c	c	b	a
c	c	a	a	c	c

Dictatorships are not anonymous: relabel the dictator and the outcome tracks a different team.

The Median Voter Theorem

Theorem (Moulin, 1980)

A strategyproof SCF f is ONTO and anonymous if and only if it is a median voter SCF.

The Median Voter Theorem

Theorem (Moulin, 1980)

A strategyproof SCF f is ONTO and anonymous if and only if it is a median voter SCF.

Will discuss the proof next week.

Lecture 16 and Week 8: What We Built

- Gibbard–Satterthwaite: for $|A| \geqslant 3$, ONTO + SP $\iff$ dictatorial – proved directly for $n = 2$ via six exhaustive cases, then by induction for $n \geqslant 3$

Lecture 16 and Week 8: What We Built

- Gibbard–Satterthwaite: for $|A| \geqslant 3$, ONTO + SP $\iff$ dictatorial – proved directly for $n = 2$ via six exhaustive cases, then by induction for $n \geqslant 3$
- The unrestricted domain $\mathcal{P}$ is what drives the impossibility; restricting to single-peaked preferences $\mathcal{S}$ reopens the door

Lecture 16 and Week 8: What We Built

- Gibbard–Satterthwaite: for $|A| \geqslant 3$, ONTO + SP $\iff$ dictatorial – proved directly for $n = 2$ via six exhaustive cases, then by induction for $n \geqslant 3$
- The unrestricted domain $\mathcal{P}$ is what drives the impossibility; restricting to single-peaked preferences $\mathcal{S}$ reopens the door
- Leftmost-, rightmost-, and median-peak rules are all strategyproof on $\mathcal{S}$

Lecture 16 and Week 8: What We Built

- Gibbard–Satterthwaite: for $|A| \geqslant 3$, ONTO + SP $\iff$ dictatorial – proved directly for $n = 2$ via six exhaustive cases, then by induction for $n \geqslant 3$
- The unrestricted domain $\mathcal{P}$ is what drives the impossibility; restricting to single-peaked preferences $\mathcal{S}$ reopens the door
- Leftmost-, rightmost-, and median-peak rules are all strategyproof on $\mathcal{S}$
- Median voter SCFs (Moulin, 1980): strategyproof by construction, PE exactly on $[p_{\min}, p_{\max}]$, and – together with ONTO and ANON – they are *the complete characterization* of non-dictatorial strategyproof rules on the single-peaked domain

Lecture 16 and Week 8: What We Built

- Gibbard–Satterthwaite: for $|A| \geqslant 3$, ONTO + SP $\iff$ dictatorial – proved directly for $n = 2$ via six exhaustive cases, then by induction for $n \geqslant 3$
- The unrestricted domain $\mathcal{P}$ is what drives the impossibility; restricting to single-peaked preferences $\mathcal{S}$ reopens the door
- Leftmost-, rightmost-, and median-peak rules are all strategyproof on $\mathcal{S}$
- Median voter SCFs (Moulin, 1980): strategyproof by construction, PE exactly on $[p_{\min}, p_{\max}]$, and – together with ONTO and ANON – they are *the complete characterization* of non-dictatorial strategyproof rules on the single-peaked domain
- **Next week:** onward to mechanism design with payments



In-Class Quiz 2

Open the Google Form now.

Questions refer only to what was just discussed –
you cannot answer them without having been in class.



End of Lecture 16

See you next class.



भारतीय प्रौद्योगिकी संस्थान मुंबई

Indian Institute of Technology Bombay