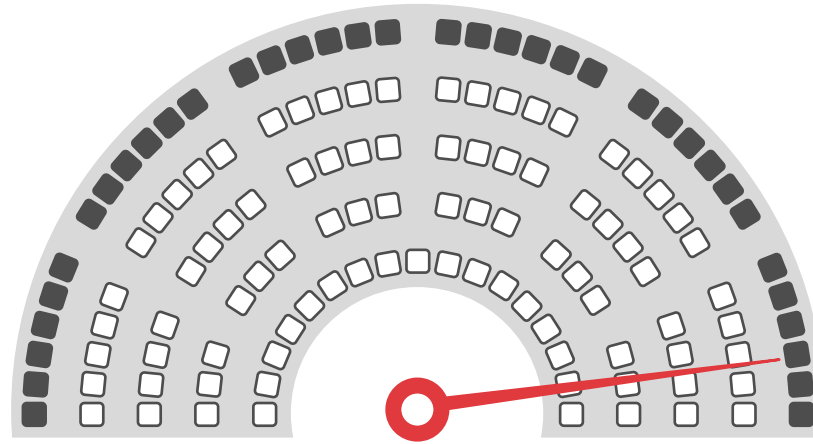




# *Optimized Democracy*

Fall 2026 | Lecture 5

**Strategic Manipulation**

Ariel Procaccia | Harvard University

# REMINDER: THE VOTING MODEL

- Set of **voters**  $N = \{1, \dots, n\}$  (assume  $n \geq 2$ )
- Set of alternatives  $A$ ; denote  $|A| = m$
- Each voter has a **ranking**  $\sigma_i \in \mathcal{L}$  over the alternatives;  $x \succ_{\sigma_i} y$  means that voter  $i$  prefers  $x$  to  $y$
- A **preference profile**  $\sigma \in \mathcal{L}^n$  is a collection of all voters' rankings
- A **social choice function** is a function  $f: \mathcal{L}^n \rightarrow A$

# MANIPULATION



So far the voters were honest!

# MANIPULATION

- Using Borda count
- Top profile:  $b$  wins
- Bottom profile:  $a$  wins
- By changing their vote, voter 3 achieves a better outcome!

| 1   | 2   | 3   |
|-----|-----|-----|
| $b$ | $b$ | $a$ |
| $a$ | $a$ | $b$ |
| $c$ | $c$ | $c$ |
| $d$ | $d$ | $d$ |

| 1   | 2   | 3   |
|-----|-----|-----|
| $b$ | $b$ | $a$ |
| $a$ | $a$ | $c$ |
| $c$ | $c$ | $d$ |
| $d$ | $d$ | $b$ |



Jean-Charles de Borda

1733–1799

“My rule is intended for honest men!”



# STRATEGYPROOFNESS

- Denote  $\sigma_{-i} = (\sigma_1, \dots, \sigma_{i-1}, \sigma_{i+1}, \dots, \sigma_n)$
- A social choice function  $f$  is **strategyproof (SP)** if a voter can never benefit from lying about their preferences:

$$\forall \sigma \in \mathcal{L}^n, \forall i \in N, \forall \sigma'_i \in \mathcal{L}, f(\sigma) \succsim_{\sigma_i} f(\sigma'_i, \sigma_{-i})$$

## Poll 1

Max  $m$  for which plurality is SP?

- $m = 2$
- $m = 3$
- $m = 4$
- $m = \infty$



# THE G-S THEOREM

- Theorem [Gibbard 1973, Satterthwaite 1975]: Let  $m \geq 3$ , then a social choice function  $f$  is SP and onto  $A$  (any alternative can win) if and only if  $f$  is dictatorial
- In other words, any voting rule that is onto and nondictatorial is manipulable

## Question

For  $m \geq 3$ , all common rules are onto and nondictatorial. What about SP and nondictatorial?



# PROOF SKETCH OF G-S

- Lemmas (prove in Pset 2):
  - **Strong monotonicity:** If  $f$  is SP function,  $\sigma$  profile,  $f(\sigma) = a$ , then  $f(\sigma') = a$  for all profiles  $\sigma'$  s.t.  $\forall x \in A, i \in N: [a \succ_{\sigma_i} x \Rightarrow a \succ_{\sigma'_i} x]$
  - **Unanimity:** If  $f$  is SP and onto function,  $\sigma$  profile, then  $[\forall i \in N, a \succ_{\sigma_i} b] \Rightarrow f(\sigma) \neq b$
- Let us assume that  $m \geq n$ , and **neutrality:**  
 $f(\pi(\sigma)) = \pi(f(\sigma))$  for all  $\pi: A \rightarrow A$

# PROOF SKETCH OF G-S

- Say  $n = 4$  and  $A = \{a, b, c, d, e\}$
- Consider the following profile

$\sigma =$

| 1   | 2   | 3   | 4   |
|-----|-----|-----|-----|
| $a$ | $b$ | $c$ | $d$ |
| $b$ | $c$ | $d$ | $a$ |
| $c$ | $d$ | $a$ | $b$ |
| $d$ | $a$ | $b$ | $c$ |
| $e$ | $e$ | $e$ | $e$ |

- Unanimity  $\Rightarrow e$  is not the winner
- Suppose  $f(\sigma) = a$

# PROOF SKETCH OF G-S

| 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|
| <i>a</i> | <i>b</i> | <i>c</i> | <i>d</i> |
| <i>b</i> | <i>c</i> | <i>d</i> | <i>a</i> |
| <i>c</i> | <i>d</i> | <i>a</i> | <i>b</i> |
| <i>d</i> | <i>a</i> | <i>b</i> | <i>c</i> |
| <i>e</i> | <i>e</i> | <i>e</i> | <i>e</i> |

$\sigma$

| 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|
| <i>a</i> | <i>d</i> | <i>d</i> | <i>d</i> |
| <i>d</i> | <i>a</i> | <i>a</i> | <i>a</i> |
| <i>b</i> | <i>b</i> | <i>b</i> | <i>b</i> |
| <i>c</i> | <i>c</i> | <i>c</i> | <i>c</i> |
| <i>e</i> | <i>e</i> | <i>e</i> | <i>e</i> |

$\sigma^1$

- Strong monotonicity  $\Rightarrow f(\sigma^1) = a$

# PROOF SKETCH OF G-S

| 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|
| <i>a</i> | <i>d</i> | <i>d</i> | <i>d</i> |
| <i>d</i> | <i>a</i> | <i>a</i> | <i>a</i> |
| <i>b</i> | <i>b</i> | <i>b</i> | <i>b</i> |
| <i>c</i> | <i>c</i> | <i>c</i> | <i>c</i> |
| <i>e</i> | <i>e</i> | <i>e</i> | <i>e</i> |

$\sigma^1$

| 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|
| <i>a</i> | <i>d</i> | <i>d</i> | <i>d</i> |
| <i>d</i> | <i>b</i> | <i>a</i> | <i>a</i> |
| <i>b</i> | <i>c</i> | <i>b</i> | <i>b</i> |
| <i>c</i> | <i>e</i> | <i>c</i> | <i>c</i> |
| <i>e</i> | <i>a</i> | <i>e</i> | <i>e</i> |

$\sigma^2$

## Poll 2

How many options are there for  $f(\sigma^2)$ ?

- 1 option
- 2 options
- 3 options
- 4 options



# PROOF SKETCH OF G-S

| 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|
| <i>a</i> | <i>d</i> | <i>d</i> | <i>d</i> |
| <i>d</i> | <i>b</i> | <i>a</i> | <i>a</i> |
| <i>b</i> | <i>c</i> | <i>b</i> | <i>b</i> |
| <i>c</i> | <i>e</i> | <i>c</i> | <i>c</i> |
| <i>e</i> | <i>a</i> | <i>e</i> | <i>e</i> |

$\sigma^2$

| 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|
| <i>a</i> | <i>d</i> | <i>d</i> | <i>d</i> |
| <i>d</i> | <i>b</i> | <i>b</i> | <i>a</i> |
| <i>b</i> | <i>c</i> | <i>c</i> | <i>b</i> |
| <i>c</i> | <i>e</i> | <i>e</i> | <i>c</i> |
| <i>e</i> | <i>a</i> | <i>a</i> | <i>e</i> |

$\sigma^3$

| 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|
| <i>a</i> | <i>d</i> | <i>d</i> | <i>d</i> |
| <i>d</i> | <i>b</i> | <i>b</i> | <i>b</i> |
| <i>b</i> | <i>c</i> | <i>c</i> | <i>c</i> |
| <i>c</i> | <i>e</i> | <i>e</i> | <i>e</i> |
| <i>e</i> | <i>a</i> | <i>a</i> | <i>a</i> |

$\sigma^4$

- Unanimity  $\Rightarrow f(\sigma^j) \notin \{b, c, e\}$
- $[\text{SP} \Rightarrow f(\sigma^j) \neq d] \Rightarrow f(\sigma^j) = a$
- Strong monotonicity  $\Rightarrow f(\sigma') = a$  for every  $\sigma'$  where 1 ranks *a* first
- Neutrality  $\Rightarrow$  1 is a dictator ■

# HARDNESS OF MANIPULATION



Manipulation may be unavoidable in theory, but can we design “reasonable” voting rules where manipulation is computationally hard?

# THE COMPUTATIONAL PROBLEM

- $f$ -MANIPULATION problem:
  - Given votes of nonmanipulators and a preferred alternative  $p$
  - Can manipulator cast vote that makes  $p$  **uniquely** win under  $f$ ?
- Example: Borda,  $p = a$

| 1   | 2   | 3 |
|-----|-----|---|
| $b$ | $b$ |   |
| $a$ | $a$ |   |
| $c$ | $c$ |   |
| $d$ | $d$ |   |

| 1   | 2   | 3   |
|-----|-----|-----|
| $b$ | $b$ | $a$ |
| $a$ | $a$ | $c$ |
| $c$ | $c$ | $d$ |
| $d$ | $d$ | $b$ |

# A GREEDY ALGORITHM

- Rank  $p$  in first place
- While there are unranked alternatives:
  - If there is an alternative that can be placed in next spot without preventing  $p$  from winning, place this alternative
  - Otherwise return false

# EXAMPLE: BORDA

| 1        | 2        | 3        | 1        | 2        | 3        | 1        | 2        | 3        |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| <i>b</i> | <i>b</i> | <i>a</i> | <i>b</i> | <i>b</i> | <i>a</i> | <i>b</i> | <i>b</i> | <i>a</i> |
| <i>a</i> | <i>a</i> |          | <i>a</i> | <i>a</i> | <i>b</i> | <i>a</i> | <i>a</i> | <i>c</i> |
| <i>c</i> | <i>c</i> |          | <i>c</i> | <i>c</i> |          | <i>c</i> | <i>c</i> |          |
| <i>d</i> | <i>d</i> |          | <i>d</i> | <i>d</i> |          | <i>d</i> | <i>d</i> |          |

| 1        | 2        | 3        | 1        | 2        | 3        | 1        | 2        | 3        |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| <i>b</i> | <i>b</i> | <i>a</i> | <i>b</i> | <i>b</i> | <i>a</i> | <i>b</i> | <i>b</i> | <i>a</i> |
| <i>a</i> | <i>a</i> | <i>c</i> | <i>a</i> | <i>a</i> | <i>c</i> | <i>a</i> | <i>a</i> | <i>c</i> |
| <i>c</i> | <i>c</i> | <i>b</i> | <i>c</i> | <i>c</i> | <i>d</i> | <i>c</i> | <i>c</i> | <i>d</i> |
| <i>d</i> | <i>d</i> |          | <i>d</i> | <i>d</i> |          | <i>d</i> | <i>d</i> | <i>b</i> |

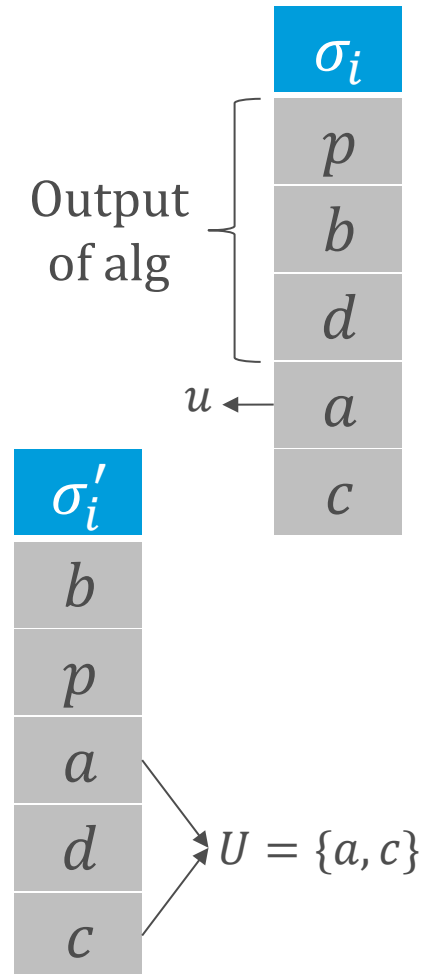
# WHEN DOES THE ALG WORK?

- **Theorem:** Fix  $i \in N$  and the votes of other voters. Let  $f$  be a rule s.t.  $\exists$  function  $s(\sigma_i, x)$  such that:
  1. For every  $\sigma_i$ ,  $f$  chooses an alternative that **uniquely** maximizes  $s(\sigma_i, x)$
  2. If  $\{y: y \prec_{\sigma_i} x\} \subseteq \{y: y \prec_{\sigma'_i} x\}$  then  $s(\sigma_i, x) \leq s(\sigma'_i, x)$

Then the greedy algorithm decides the  $f$ -MANIPULATION problem correctly

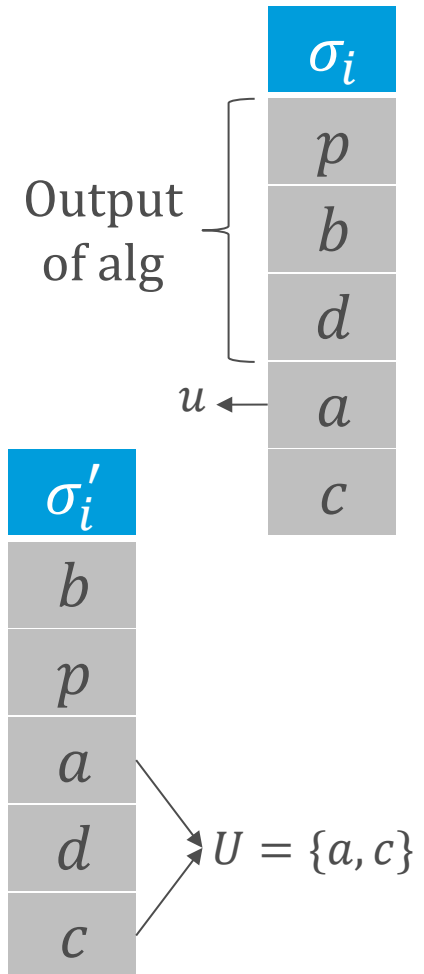
# PROOF OF THEOREM

- Suppose the algorithm failed, producing a partial ranking  $\sigma_i$
- Assume for contradiction  $\sigma'_i$  makes  $p$  win
- $U \leftarrow$  alternatives not ranked in  $\sigma_i$
- $u \leftarrow$  highest ranked alternative in  $U$  according to  $\sigma'_i$
- Complete  $\sigma_i$  by adding  $u$  first, then others arbitrarily



# PROOF OF THEOREM

- Property 2  $\Rightarrow s(\sigma_i, p) \geq s(\sigma'_i, p)$
- Property 1 and  $\sigma'_i$  makes  $p$  the winner  $\Rightarrow s(\sigma'_i, p) > s(\sigma'_i, u)$
- Property 2  $\Rightarrow s(\sigma'_i, u) \geq s(\sigma_i, u)$
- Conclusion:  $s(\sigma_i, p) > s(\sigma_i, u)$ , so the alg could have inserted  $u$  next ■



# HARD-TO-MANIPULATE RULES



IRV and Lull (with second-order tie breaking)  
are NP-hard to manipulate!

# HARD-TO-MANIPULATE RULES



But worst-case hardness isn't necessarily an obstacle to manipulation in the average case!

# BIBLIOGRAPHY

A. Gibbard. **Manipulation of Voting Schemes: A General Result.** Econometrica, 1973.

M. A. Satterthwaite. **Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions.** Journal of Economic Theory, 1975.

J. Bartholdi, III, C. A. Tovey, and M. A. Trick. **The Computational Difficulty of Manipulating an Election.** Social Choice and Welfare, 1989.

