Strategic Manipulation

Paul Gölz

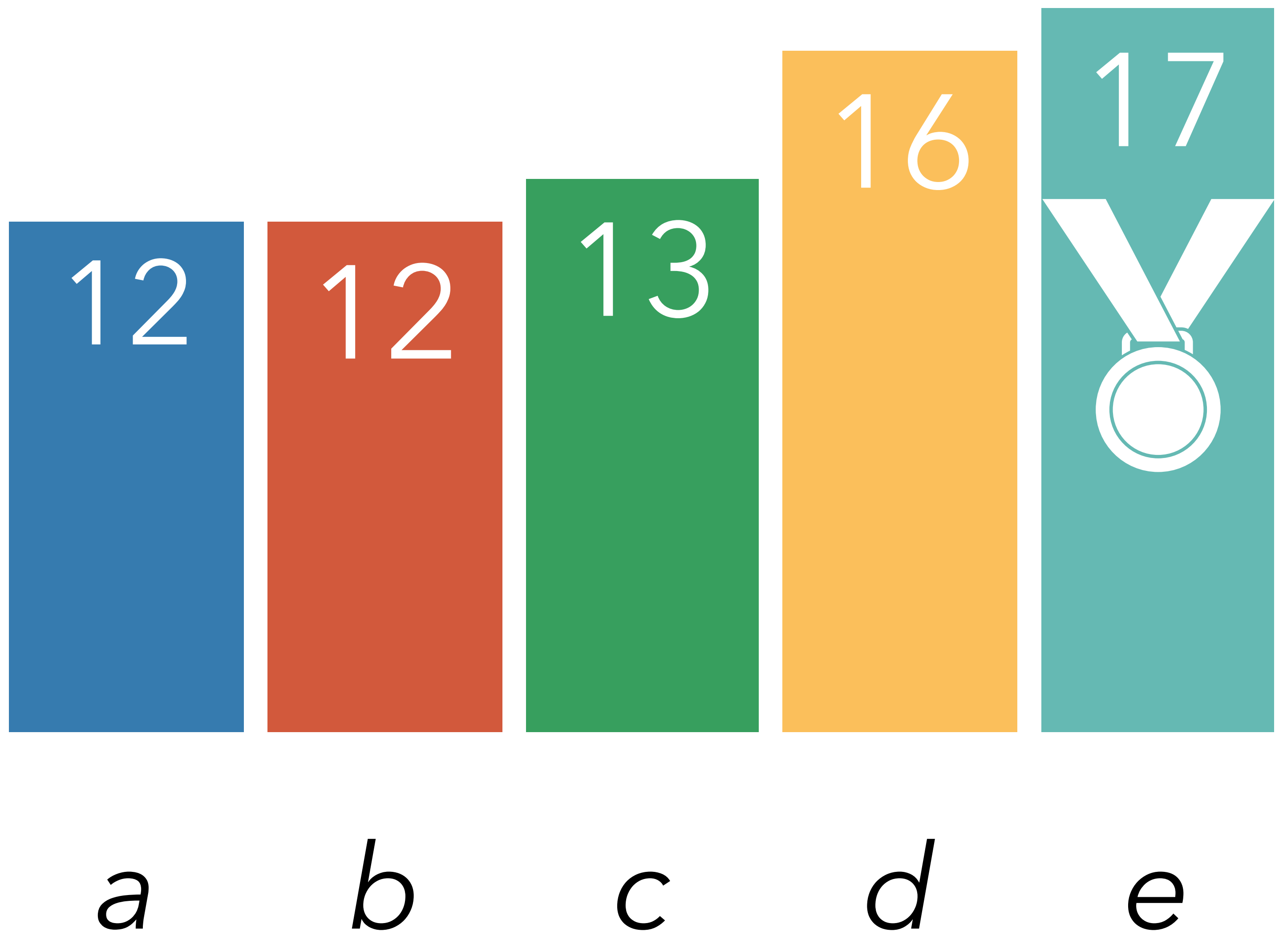
Lecture 3, ORIE 4360/5360, Fall ’26

How Would You Vote?

you

1x	1x	3x	2x
a	a	d	e
b	b	e	c
c	c	b	a
d	d	c	d
e	e	a	b

Borda count

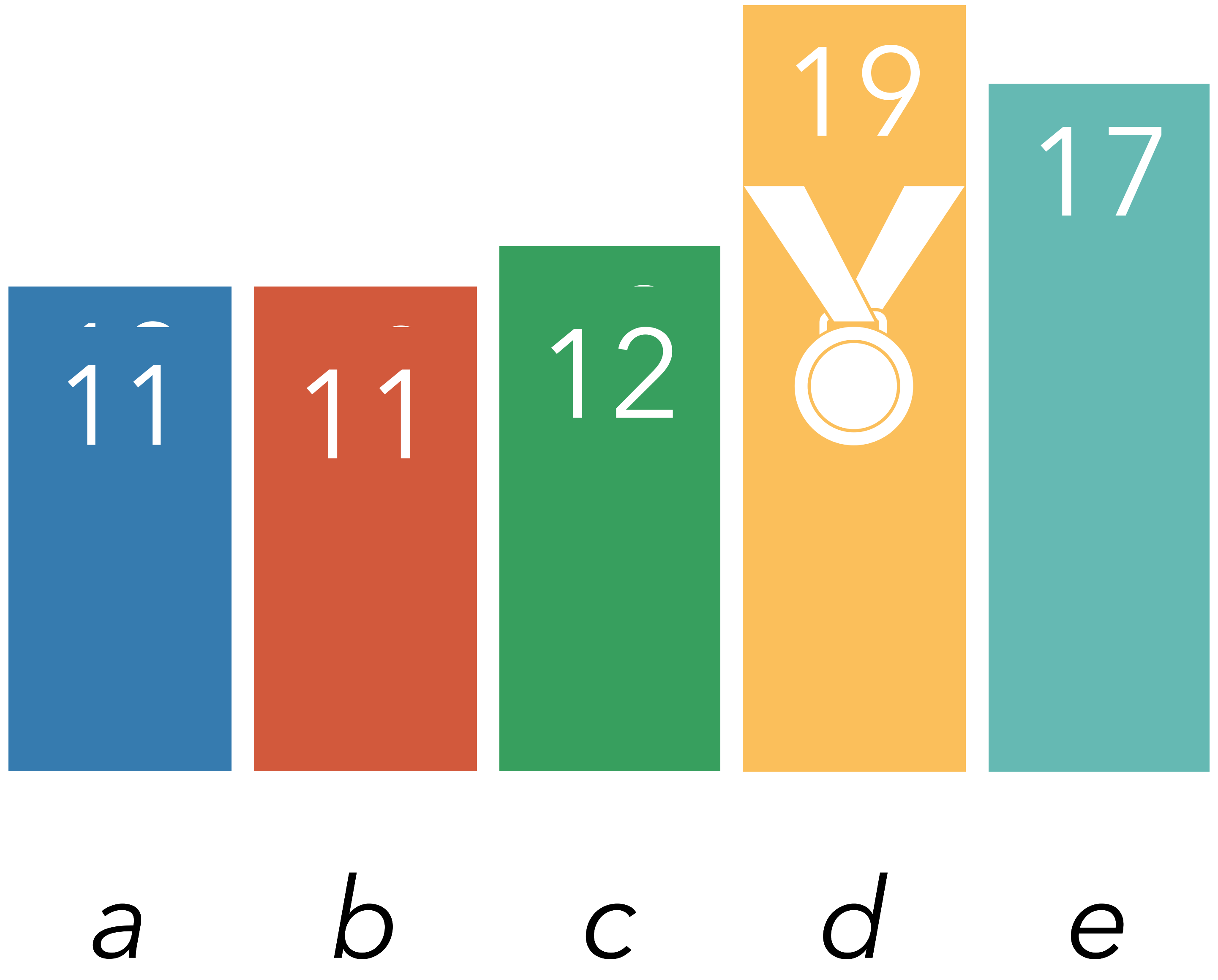


How Would You Vote?

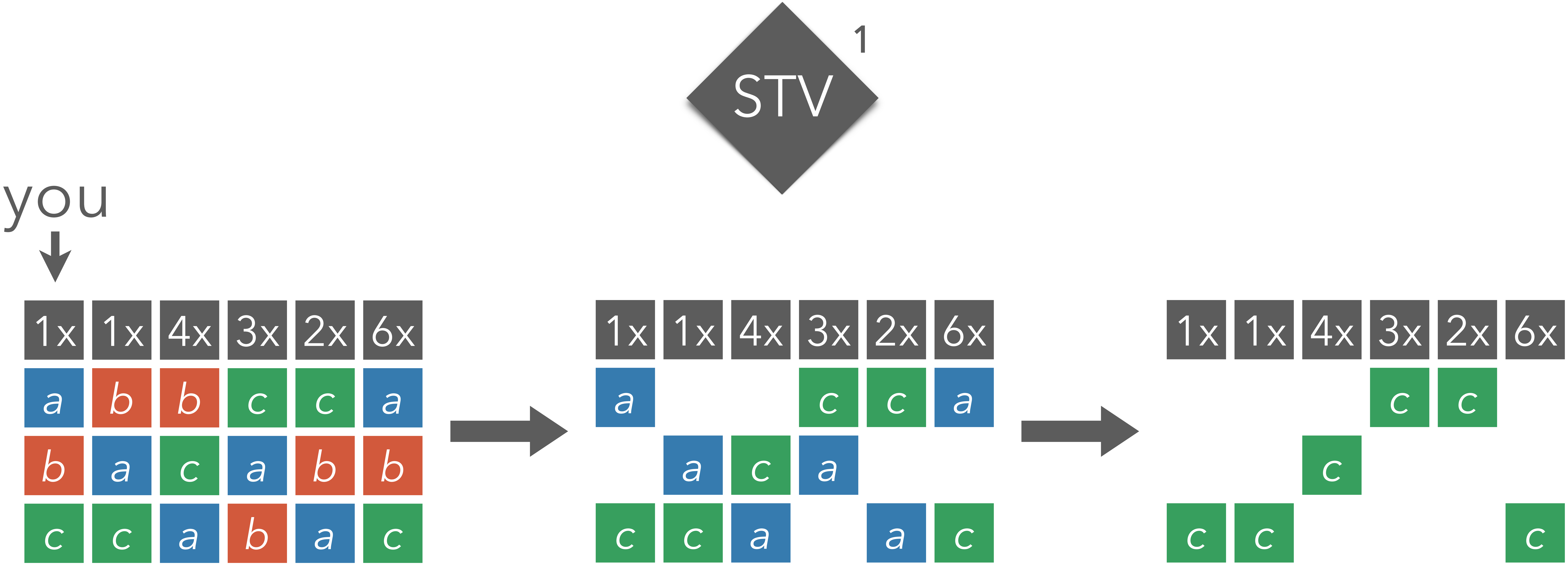
you

1x	1x	3x	2x
d	a	d	e
a	b	e	c
b	c	b	a
c	d	c	d
e	e	a	b

Borda count



How Would You Vote?



[1] Breaking ties: eliminating *a* before *b* before *c*.

How Would You Vote?

STV[1]

you 

1x	1x	4x	3x	2x	6x
b	*b*	*b*	*c*	*c*	*a*
a	*a*	*c*	*a*	*b*	*b*
c	*c*	*a*	*b*	*a*	*c*

→

1x	1x	4x	3x	2x	6x
b	*b*	*b*			*a*
a	*a*		*a*	*b*	*b*
		a	*b*	*a*	

→

1x	1x	4x	3x	2x	6x
					a
a	*a*		*a*		
		a		*a*	

[1] Breaking ties: eliminating *a* before *b* before *c*.

How To Deal With Manipulation?

My scheme is intended only for honest men!

Jean-Charles de Borda
(1733–1799)

Can the voting rule incentivize telling the truth?

Recap: Voting Model

Voters $N = \{1,2,\ldots,n\}$.

Alternatives $M = \{a, b, c, \ldots\}, |M| = m$.

Each voter i has a ranking $\sigma_i \in \mathscr{L}$ over the alternatives; $a \succ_{\sigma_i} b$ means that i prefers a to b.

A preference profile $\boldsymbol{\sigma} \in \mathscr{L}^n$ is a collection of all voters' rankings.

A social choice function is a function $f : \mathscr{L}^n \to M$; a social welfare function is $f : \mathscr{L}^n \to \mathscr{L}$.

Definitions

Denote $\boldsymbol{\sigma}_{-i} := (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$.

A social choice function f is strategyproof (SP) if a voter can never benefit from lying about their preferences:

$$\forall \boldsymbol{\sigma} \in \mathscr{L}^n, i \in N, \sigma_i' \in \mathscr{L}.\ \ f(\boldsymbol{\sigma}) \succcurlyeq_{\sigma_i} f(\sigma_i', \boldsymbol{\sigma}_{-i}).$$

A social choice function f is onto if each alternatives wins for some profile.

"$x \succcurlyeq y$" means that $x \succ y$ or $x = y$

Poll

A social choice function f is strategyproof (SP) if a voter can never benefit from lying about their preferences:

$$\forall \boldsymbol{\sigma} \in \mathscr{L}^n, i \in N, \sigma_i' \in \mathscr{L}.\ f(\boldsymbol{\sigma}) \succcurlyeq_{\sigma_i} f(\sigma_i', \boldsymbol{\sigma}_{-i}).$$

Largest m for which plurality is SP?

- $m = 1$
- $m = 2$
- $m = 4$
- $m = \infty$

The Gibbard-Satterthwaite Theorem

Theorem [Gibbard 1973, Satterthwaite 1975]**.** For $m \geq 3$, a social choice function is SP and onto if and only if f is a dictatorship.

Q: For $m \geq 3$, are there rules that are SP and nondictatorial? …onto and nondictatorial?

Proof Sketch

There is a proof that is very similar to the proof we've seen of Arrow's theorem:

We will see different proof (following Svensson & Reffgen'14), with some simplifying assumptions.

Proof Sketch

Lemmas (prove in HW):

- If f is SP, it satisfies strong monotonicity: whenever $f(\boldsymbol{\sigma}) = x$ for some $\boldsymbol{\sigma}, x$, then $f(\boldsymbol{\sigma}') = x$ for all $\boldsymbol{\sigma}'$ where $\forall i, y \,.\, x \succ_{\sigma_i} y \Rightarrow x \succ_{\sigma_i'} y$.
- If f is SP and onto, it is unanimous: if all agents prefer x over y in $\boldsymbol{\sigma}$, then $f(\boldsymbol{\sigma}) \neq y$.

Proof Sketch

To make proof easier, we will also assume:

- neutrality: for any permutation $\pi : M \to M$ of alternatives and any $\boldsymbol{\sigma}$, $f(\pi(\boldsymbol{\sigma})) = \pi(f(\boldsymbol{\sigma}))$.
- for concreteness, let $n = 4$, $m = 5$ (idea easily generalizes to $m \geq n$).

Proof Sketch

Let f be SP and onto. Consider following profile:

$\boldsymbol{\sigma}$:

1	2	3	4
a	b	c	d
b	c	d	a
c	d	a	b
d	a	b	c
e	e	e	e

By unanimity, $f(\boldsymbol{\sigma}) \neq e$.

By relabeling the voters and alternatives, can assume that $f(\boldsymbol{\sigma}) = a$.

Proof Sketch

$\boldsymbol{\sigma}$:

1	2	3	4
a	b	c	d
b	c	d	a
c	d	a	b
d	a	b	c
e	e	e	e

$f(\boldsymbol{\sigma}) = a$

by strong monotonicity

$\longrightarrow$

$\boldsymbol{\sigma}^1$:

1	2	3	4
a	d	d	d
d	a	a	a
b	b	b	b
c	c	c	c
e	e	e	e

$f(\boldsymbol{\sigma}^1) = a$

Proof Sketch

$\boldsymbol{\sigma}^1$:

1	2	3	4
a	*d*	*d*	*d*
d	*a*	*a*	*a*
b	*b*	*b*	*b*
c	*c*	*c*	*c*
e	*e*	*e*	*e*

$\boldsymbol{\sigma}^2$:

1	2	3	4
a	*d*	*d*	*d*
d	*b*	*a*	*a*
b	*c*	*b*	*b*
c	*e*	*c*	*c*
e	*a*	*e*	*e*

$$f(\boldsymbol{\sigma}^1) = a$$

Proof Sketch

$\boldsymbol{\sigma}^1$:

1	2	3	4
a	d	d	d
d	a	a	a
b	b	b	b
c	c	c	c
e	e	e	e

$f(\boldsymbol{\sigma}^1) = a$

$\boldsymbol{\sigma}^2$:

1	2	3	4
a	d	d	d
d	b	a	a
b	c	b	b
c	e	c	c
e	a	e	e

$f(\boldsymbol{\sigma}^2) = a$

Q: Why can't it be that $f(\boldsymbol{\sigma}^2)$ is in $\{b, c, e\}$?

Q: Why can't $f(\boldsymbol{\sigma}^2) = d$?

Proof Sketch

σ^2 :

1	2	3	4
a	d	d	d
d	b	a	a
b	c	b	b
c	e	c	c
e	a	e	e

$$f(\sigma^2) = a$$

σ^3 :

1	2	3	4
a	d	d	d
d	b	b	a
b	c	c	b
c	e	e	c
e	a	a	e

$$f(\sigma^3) = a$$

σ^4 :

1	2	3	4
a	d	d	d
d	b	b	b
b	c	c	c
c	e	e	e
e	a	a	a

$$f(\sigma^4) = a$$

Proof Sketch

$\boldsymbol{\sigma}^4$:

1	2	3	4
a	d	d	d
d	b	b	b
b	c	c	c
c	e	e	e
e	a	a	a

$f(\boldsymbol{\sigma}^4) = a$

Strong monotonicity implies that $f(\boldsymbol{\sigma}) = a$ whenever agent 1 ranks a first in $\boldsymbol{\sigma}$.

By neutrality, f always returns agent 1's first choice, i.e., 1 is the dictator.

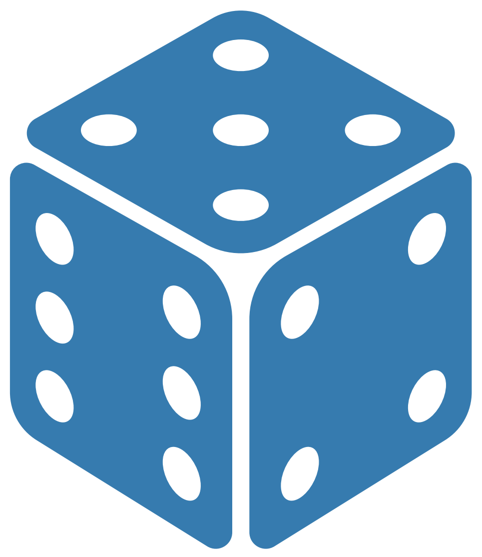
$\square$

The Gibbard-Satterthwaite Theorem

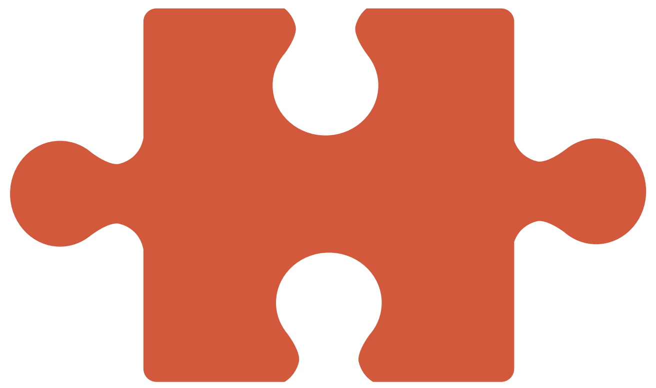
Theorem [Gibbard 1973, Satterthwaite 1975]**.** For $m \geq 3$, a social choice function is SP and onto if and only if f is a dictatorship.

Escape Routes

Randomization?



Computational hardness?



Restricted preferences.

Randomization

Consider randomized social choice function $f : \mathscr{L}^n \to \Delta(M)$. (Need to define preferences over distributions.)

Q: Can you think of a randomized voting rule that is strategyproof, anonymous, and neutral?

Randomization

Consider randomized social choice function $f : \mathscr{L}^n \to \Delta(M)$. (Need to define preferences over distributions.)

Example 1: Randomly choose a voter, return their first choice ("randomized dictatorship").

Example 2: Randomly choose two alternatives, return the winner of the head-to-head election.

Gibbard (1977): all randomized SP rules are mixtures of single-voter and two-alternative rules.

Computational Hardness

If it's too hard to find a manipulation, maybe voters will give up and be honest?

Bartholdi & Orlin (1991): STV manipulation is "NP-hard".

- That is, if there was an efficient algorithm for finding manipulations, we would get efficient algorithms for notoriously hard problems. So there probably isn't one.

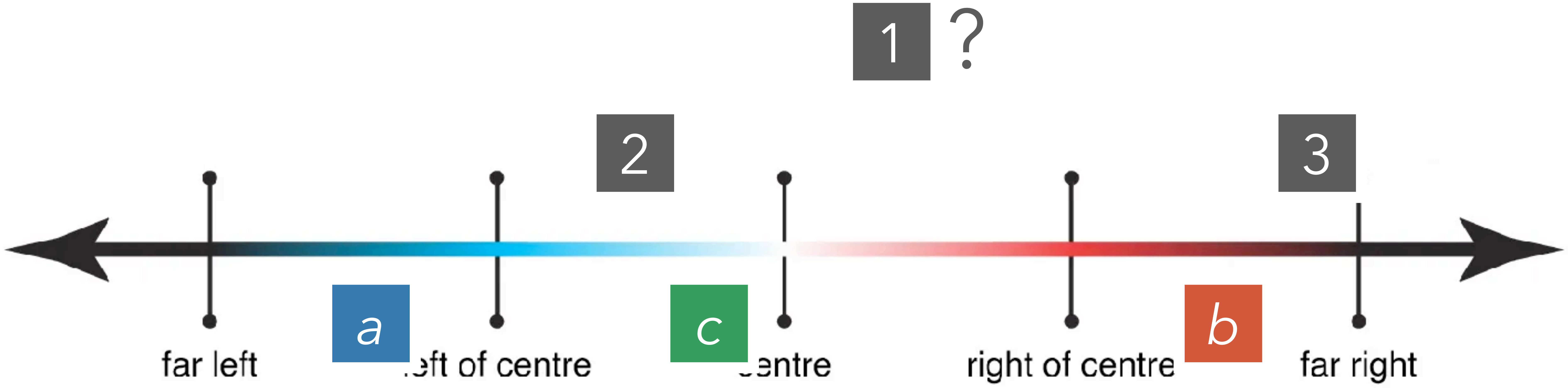
Problem: Most manipulations might still be easy to find!

Next Time: Restricted Preferences

Problem were cyclical preferences:

1	2	3
a	*c*	*b*
b	*a*	*c*
c	*b*	*a*

But, if voters and alternatives are arranged on a left-right spectrum, this is a bit surprising:



Bibliography

Our proof sketch came from

Section 5 of Svensson & Reffgen: *The proof of the Gibbard–Satterthwaite theorem revisited*. Journal of Mathematical Economics, 2014.

Discussion of manipulation and ways around it:

Chapter 6.2–6.4 in Brandt et al. (2016): *Handbook of computational social choice*.

Bibliography

Characterizaiton of randomized strategy-proof voting rules:

Gibbard: *Manipulation of schemes that mix voting with chance*. Econometrica, 1977.

Computational hardness of STV:

Bartholdi & Orlin: *Single transferable vote resists strategic voting*. Social Choice and Welfare, 1991.

Side-by-side proof of Gibbard–Satterthwaite and Arrow:

Reny: *Arrow's theorem and the Gibbard-Satterthwaite theorem*. Economics Letters, 2001.