# No Representation Without Randomization!

Tradeoffs Between Randomness, Representation, Manipulation-Robustness, and Other Desiderata for Sortition Algorithms

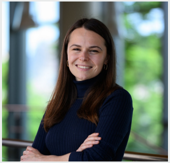

Adithya Bhaskara

University of Colorado Boulder

adithya@colorado.edu


November 7, 2025

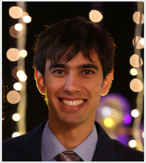
Undergraduate Thesis Proposal



Bailey Flanigan

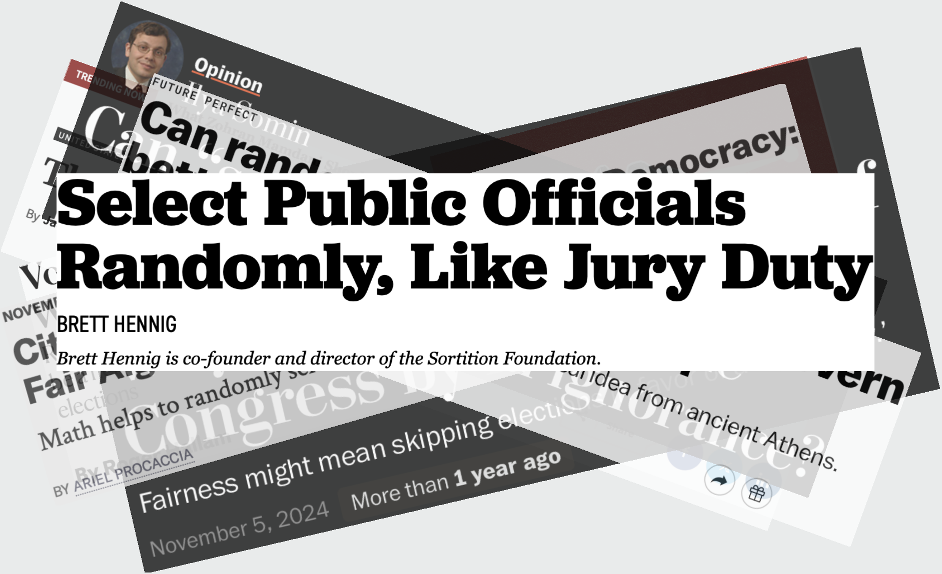
Rafael Frongillo

Joshua Strayhorn



Bo Waggoner

Other acknowledgments: Stephen Becker, Sriram Sankaranarayanan, Sherry Sarkar


Opinion
FUTURE PERFECT
Democracy:
Select Public Officials Randomly, Like Jury Duty
BRETT HENNIG
Brett Hennig is co-founder and director of the Sortition Foundation.
BY ARIEL PROCACCIA
November 5, 2024
More than 1 year ago

# Okay... But What is Sortition?

Sortition: The process of selecting citizens' assemblies at random, subject to representation quotas.

origins rooted in Athenian democracy

A lot of recent work: [FGG$^+$21, FKP21, FLPW24, BF24, ABFP25], etc.

**Question**

*Given a pool of citizens that is not representative of the population, how can we select a subset of the pool that is representative in a "good" way?*

There are several variations on this question!

Fundamentally, sortition is a random selection process.

# The Rest of This Talk

1. Model & Motivation

2. A Detour Into Optimization

3. Extensions: Manipulation-Robust Alternate Selection

# Model & Motivation

# Model

An instance $\mathcal{I}$ of a sortition problem is

- $N$, a pool of $n$ agents — $\sharp$ of CSCI majors
- $k$, a panel size — need 20 people on an advisory board
- $F$, a set of binary features — likes theory, older than 21, etc.
- $\ell_{f,v}$, $u_{f,v}$, lower and upper quota requirements for each $(f, v) \in F \times \{0, 1\}$ — need $\geq 2$ and $\leq 5$ people that like theory

Want

$$K \in \mathcal{K} = \big\{K \subseteq N : |K| = k \wedge \\ \ell_{f,v} \leq |\{i \in K : f(i) = v\}| \leq u_{f,v}, \\ \forall (f, v) \in F \times \{0, 1\}|\big\}.$$

# Randomizing Over Valid Integral Panels

Given instance $\mathcal{I}$ with valid panels $\mathcal{K}$, $\Delta_{\mathcal{K}}$ is the set of all possible randomizations over valid panels.

Define $\boldsymbol{\pi} \in [0,1]^n$ as a vector of selection probabilities over $N$ agents that can be realized by a randomization over solutions $\boldsymbol{\lambda} \in \Delta_{\mathcal{K}}$, i.e.,

$$\exists \boldsymbol{\lambda} \quad \pi_i = \sum_{K \in \mathcal{K}} \lambda_K \mathbf{1}_{i \in K} \ \wedge \ \sum_{K \in \mathcal{K}} \lambda_K = 1 \ \wedge \ \lambda_K \geq 0, \quad \forall i \in N, \ \forall K \in \mathcal{K}$$

$\pi_i$ is the probability we select agent $i$ to be on a (valid) panel

Let $\Pi(\mathcal{I})$ be the set of all possible lotteries $\boldsymbol{\pi}$ implied by $\Delta_{\mathcal{K}}$.

# Equality Objectives

Given instance $\mathcal{I}$ and selection probabilities $\boldsymbol{\pi} \in \Pi(\mathcal{I})$, to what extent does each agent have an equal chance to participate on the panel?

An equality objective $f : [0,1]^n \to \mathbb{R}$ determines how unequal a vector of selection probabilities $\boldsymbol{\pi}$ is;

We always minimize $f$

$\Pi^\star(\mathcal{I}) = \operatorname{arginf}_{\boldsymbol{\pi} \in \Pi(\mathcal{I})} f(\boldsymbol{\pi})$ is the set of all maximally equal selection probabilities.

**Example**

1. $f(\boldsymbol{\pi}) = -\min(\boldsymbol{\pi})$ *maximize minimum probability of selection*
2. $f(\boldsymbol{\pi}) = \max(\boldsymbol{\pi})$ *minimize maximum probability of selection*

# A Problem

**Question**

*What if our representation quotas are too restrictive?*

We may not have enough valid panels to randomize over!

Randomization is inherent to what sortition is.

Let's relax. Perhaps allowing flexibility in our quotas ($\pm \sim 1$) will result in better lotteries?

Figuring this out requires optimizing simultaneously over the equality objective $f$, and the flexibility of our quotas!

Hard to do when our panels are integral

Easier when they are allowed to be continuous

**No Representation Without Randomization!**

# No Free Lunch[1] ☹

In sortition, we want to optimize several properties of our panel.

Unfortunately, some of these properties are in inherent conflict with one another.

Making sense of these tradeoffs is difficult when we require our algorithm to take “whole” people, but the analysis might be nicer if we allow fractional panels.

Recall $f$ captures how unequal selection probabilities are; let $g$ capture how bad a panel is. e.g. unrepresentative, vulnerable to manipulation

[1] Props. LL & MM passed in Colorado, though! ☺

# A Detour Into Optimization

# Integral & Continuous Panels

The set of valid integer panels is

$$\mathcal{K} = \Big\{K \subseteq N : |K| = k \wedge \\ \ell_{f,v} \leq |\{i \in K : f(i) = v\}| \leq u_{f,v}, \\ \forall (f,v) \in F \times \{0,1\}|\Big\}.$$

$\Pi(\mathcal{I})$ is the set of lotteries $\boldsymbol{\pi}$ over valid integer panels implied by $\Delta_{\mathcal{K}}$.

The set of valid continuous panels is

$$\tilde{\Pi}(\mathcal{I}) = \Big\{\tilde{\boldsymbol{\pi}} \in [0,1]^n : \sum_{i \in N} \tilde{\pi}_i = k \wedge \\ \ell_{f,v} \leq \sum_{i \in N,\, f(i)=v} \tilde{\pi}_i \leq u_{f,v} \quad \forall (f,v) \in F \times \{0,1\}\Big\}.$$

Elements of $\Pi(\mathcal{I})$ are integer solutions, and elements of $\tilde{\Pi}(\mathcal{I})$ are continuous solutions.

# Integrality Gaps: Convergence in the Limit? Intuition

Given an instance $\mathcal{I} = (N, k, \boldsymbol{\ell}, \boldsymbol{u})$. What if we scaled $\mathcal{I}$ up, i.e., $\mathcal{I}_c = (cN, ck, c\boldsymbol{\ell}, c\boldsymbol{u})$? Does $\lim_{c\to\infty} \Pi(\mathcal{I}_c) = \tilde{\Pi}(\mathcal{I})$? ✓

Intuition: scaling up an instance makes people more divisible, i.e., makes the feasible region more "dense."

**Example**

| $\mathbb{Q}$ | $\mathbb{N}$ | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $\sharp$ | $\mathbb{Q}$ | $\mathbb{N}$ | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $\sharp$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $2.\overline{3}$ | 3 | 0 | 1 | 1 | 1 | 4 | $4.\overline{6}$ | 5 | 0 | 1 | 1 | 1 | 8 |
| $2.\overline{3}$ | 3 | 1 | 0 | 1 | 1 | 4 | $4.\overline{6}$ | 5 | 1 | 0 | 1 | 1 | 8 |
| $2.\overline{3}$ | 2 | 1 | 1 | 0 | 1 | 4 | $4.\overline{6}$ | 5 | 1 | 1 | 0 | 1 | 8 |
| $2.\overline{3}$ | 2 | 1 | 1 | 1 | 0 | 4 | $4.\overline{6}$ | 4 | 1 | 1 | 1 | 0 | 8 |
| $\frac{2}{3}$ | 0 | 0 | 0 | 0 | 0 | 1 | $\frac{4}{3}$ | 1 | 0 | 0 | 0 | 0 | 2 |

$k = 10$ $\qquad$ $k = 20$

$\boldsymbol{\ell}_{\cdot,1} = 7$ $\qquad$ $\boldsymbol{\ell}_{\cdot,1} = 14$

*Take $\frac{1}{2}$ a person from each of* 0111, 1011, *and redistribute.*

# Integrality Gaps: Convergence in the Limit? Formalize

**Theorem 1**

*Given $\mathcal{I} = (N, k, \boldsymbol{\ell}, \boldsymbol{u})$ and $\mathcal{I}_c(cN, ck, c\boldsymbol{\ell}, c\boldsymbol{u})$, $\lim_{c\to\infty} \Pi(\mathcal{I}_c) = \tilde{\Pi}(\mathcal{I})$ with convergence rate $O(\frac{\sqrt{n}}{c})$.*

Let $\delta_{p|c} = \left\{ \frac{k}{p} : k \in \{0, \ldots, p\}; p \in \mathbb{N},\ p|c \right\}$.

The set of valid restricted continuous panels is

$$\tilde{\Pi}^{\delta}(\mathcal{I}) = \Big\{ \tilde{\boldsymbol{\pi}} \in \delta^{n}_{p|c} : \sum_{i\in N} \tilde{\pi}_i = k \wedge$$

$$\ell_{f,v} \leq \sum_{i\in N,\, f(i)=v} \tilde{\pi}_i \leq u_{f,v} \quad \forall (f, v) \in F \times \{0, 1\} \Big\}.$$

Need to show: $\lim_{c\to\infty} \delta^{n}_{p|c} = [0, 1]^n$. technically $(\mathbb{Q} \cap [0, 1])^n$

**Theorem 1**

*Given* $\mathcal{I} = (N, k, \boldsymbol{\ell}, \boldsymbol{u})$ *and* $\mathcal{I}_c(cN, ck, c\boldsymbol{\ell}, c\boldsymbol{u})$, $\lim_{c\to\infty} \Pi(\mathcal{I}_c) = \tilde{\Pi}(\mathcal{I})$ *with convergence rate* $O(\frac{\sqrt{n}}{c})$.

**Proof (Sketch) of Theorem 1.**

*We first show* $\lim_{c\to\infty} \delta_{p|c} = [0, 1]$. *Consider*

$$\delta_{p|c} = \bigcup_{p,\, p|c} \left\{ \frac{k}{p} : k \in \{0, \ldots, p\} \right\} = \{0\} \cup \bigcup_{p,\, p|c} \bigcup_{k=0}^{p} \left\{ \frac{k}{p} \right\}$$

*Note* $\delta_{p|c} \subseteq \delta_{p|c^m}$ *since if* $p|c$, *then* $p|c^m$. *So,*

$$\lim_{c\to\infty} \delta_{P(c)} = \{0\} \cup \bigcup_{c=1}^{\infty} \bigcup_{p,\, p|c} \bigcup_{k=0}^{p} \left\{ \frac{k}{p} \right\}$$
$$= \mathbb{Q} \cap [0, 1].$$

□

# Integrality Gaps: Convergence in the Limit? Proof Sketch

**Theorem 1**

*Given* $\mathcal{I} = (N, k, \boldsymbol{\ell}, \boldsymbol{u})$ *and* $\mathcal{I}_c(cN, ck, c\boldsymbol{\ell}, c\boldsymbol{u})$, $\lim_{c\to\infty} \Pi(\mathcal{I}_c) = \tilde{\Pi}(\mathcal{I})$ *with convergence rate* $O(\frac{\sqrt{n}}{c})$.

**Proof (Sketch) of Theorem 1, Continued.**

*The Hausdorff distance between* $\delta_{p|c}$ *and* $\mathbb{Q} \cap [0,1]$ *is*

$$d_{\mathrm{H}}(\delta_{p|c}, \mathbb{Q} \cap [0,1]) = \sup_{q \in (\mathbb{Q}\cap[0,1])^n} \inf_{x \in \delta^n_{p|c}} ||x - q||_2$$
$$= \left|\left|\left(\frac{1}{2c}, \ldots, \frac{1}{2c}\right)\right|\right|_2 = \frac{\sqrt{n}}{2c}.$$

*The idea is that the furthest* $q \in (\mathbb{Q} \cap [0,1])^n$ *is at the centroid of the* $n$*-hypercube with sidelengths* $\frac{1}{c}$. □

**There's a problem! The rate should decrease with *n*!**

We'll figure this out...

# Integrality Gaps: Minimal Instances

Great! So as we scale up instances, the integral solutions converge to the continuous ones.

**Question**

*But what can we say about a fixed instance size, say one where there are $|F|$ features to deal with?*

# Integrality Gaps: Minimal Instances, A Lower Bound

## Example

| $\mathbb{Q}$ | $\mathbb{N}$ | $f_1$ | $f_2$ | $f_3$ | $\sharp$ |
|---|---|---|---|---|---|
| 1.5 | 2 | 0 | 1 | 1 | 3 |
| 1.5 | 2 | 1 | 0 | 1 | 3 |
| 1.5 | 1 | 1 | 1 | 0 | 3 |
| 0.5 | 0 | 0 | 0 | 0 | 1 |

$k = 5$ $\qquad$ $\boldsymbol{\ell}_{\cdot,1} = 3$

*In the integral case, we are prohibited from taking* 0000*. In the fractional case, we can take* 0.5 *of* 0000*. So, the integrality gap is*

$$\min(\tilde{\boldsymbol{\pi}}) - \min(\boldsymbol{\pi}) = \frac{1}{2} - 0 = \frac{1}{2}.$$

# Integrality Gaps: Minimal Instances, Generalized Lower Bounds

**Theorem 2**

*For all $F$ with $|F| \geq 3$, there exists an instance $\mathcal{I} = (N, k, \boldsymbol{\ell}, \boldsymbol{u})$ such that*

$$\min(\tilde{\boldsymbol{\pi}}) - \min(\boldsymbol{\pi}) = \frac{|F|-2}{|F|-1}.$$

**Proof (Sketch).**

*The construction is as follows.*

| $\mathbb{Q}$ | $\mathbb{N}$ | $f_1$ | $f_2$ | $\cdots$ | $f_{\lvert F\rvert-1}$ | $f_{\lvert F\rvert}$ | $\sharp$ |
|---|---|---|---|---|---|---|---|
| $\lvert F\rvert + \frac{1}{\lvert F\rvert-1} - 2$ | $\lvert F\rvert - 1$ | 0 | 1 | $\cdots$ | 1 | 1 | $\lvert F\rvert$ |
| $\lvert F\rvert + \frac{1}{\lvert F\rvert-1} - 2$ | $\lvert F\rvert - 1$ | 1 | 0 | $\cdots$ | 1 | 1 | $\lvert F\rvert$ |
| $\lvert F\rvert + \frac{1}{\lvert F\rvert-1} - 2$ | $\lvert F\rvert - 2$ | rest | of | the | feature | vectors | $\lvert F\rvert$ |
| $\frac{\lvert F\rvert-2}{\lvert F\rvert-1}$ | 0 | 0 | 0 | $\cdots$ | 0 | 0 | 1 |

$k = 2(|F|-1) + (|F|-2)^2$, $\boldsymbol{\ell}_{\cdot,1} = \boldsymbol{u}_{\cdot,1} = (|F|-1)^2 + 1.$ $\square$

**Theorem 2**

*For all $F$ with $|F| \geq 3$, there exists an instance $\mathcal{I} = (N, k, \boldsymbol{\ell}, \boldsymbol{u})$ such that*

$$\min(\tilde{\boldsymbol{\pi}}) - \min(\boldsymbol{\pi}) = \frac{|F| - 2}{|F| - 1}.$$

**Proof (Sketch), Continued.**

*First,* $\min(\boldsymbol{\pi}) = 0$, *since* $0 \cdots 0$ *is never chosen. If we didn't abide by our selections above (up to symmetry), we'd not satisfy the lower quotas. We'd have*

$$\begin{aligned} |\{(f, v) \text{ in panel} : v = 1\}| &= \max\{|F| + |F| - 1 + (|F| - 2)^2, 2(|F| - 1) \\ &\qquad + (|F| - 1)(|F| - 2)\} \\ &< 2(|F| - 1) + (|F| - 2)^2 = \ell_{f,1}. \end{aligned}$$

*The argument for* $\min(\tilde{\boldsymbol{\pi}})$ *follows accordingly.* □

# Extensions: Manipulation-Robust Alternate Selection

## Selecting Alternates

After a panel has been selected, panelists may "drop out" due to other commitments.

**Question**

*How can we select a slate of alternates $A \subseteq N$ to ensure the panel can still be representative, before seeing which panelists drop out?*

[ABFP25] give an ERM-based algorithm to select $A \subseteq N$ given estimates of panelists' dropout probabilities.

# Manipulation-Robust Sortition

Ideally, agents shouldn't be able to manipulate the selection algorithm by misreporting their features.

But, the algorithm should still be fair, i.e., has a high minimum selection probability.

- $\max(\boldsymbol{\pi})$ is robust to manipulation, but is unfair — minimax
- $-\min(\boldsymbol{\pi})$ is optimally fair, but easy to manipulate — leximin

**Question**

*Is there an optimal balance?*

[BF24] give an equality objective that achieves the best bounds that we can hope for — goldilocks

# Selecting Alternates & Manipulation-Robust Sortition?

Given a two-stage selection process, members of the pool can become panelists by being chosen as a panelist directly, or being chosen as an alternate and being called in as a replacement.

How can agents misreport their features in this setting? Incentives are very different!

# What could $g$ be?

Detour: Originally wanted to simultaneously optimize our equality objective $f$ and another objective $g$, describing some other property of the panel.

Possible choices for $g$:

- robustness to dropouts; $g(K) = \mathbb{E}_{D \sim \hat{D}} \mathcal{L}(K \setminus D)$
- diversity; $g(K)$ is the number of unique vectors in $K$
- robustness to dropouts with alternates; $g(A) = \mathbb{E}_{D \sim \hat{D}} \mathcal{L}(K \setminus D \cup R)$
  $R \subseteq A \subseteq N$
- closeness to representation targets
- anything you want

Would love to use our optimization framework in this setting!

There are some nice connections to $|F|$-dimensional matching. Can approximation results/integrality gaps/bounds from there be applied here?

This work started with how to make alternate-selection manipulation robust! How can we apply our optimization ideas/results?

As always, we'd like to improve our bounds, convergence rates, etc. And we think we certainly can.

adithya@colorado.edu

Angelos Assos, Carmel Baharav, Bailey Flanigan, and Ariel Procaccia.
**Alternates, Assemble! Selecting Optimal Alternates for Citizens' Assemblies, page 719–738.**
Association for Computing Machinery, New York, NY, USA, 2025.

Carmel Baharav and Bailey Flanigan.
**Fair, manipulation-robust, and transparent sortition.**
In *Proceedings of the 25th ACM Conference on Economics and Computation*, EC '24, page 756–775, New York, NY, USA, 2024. Association for Computing Machinery.

Bailey Flanigan, Paul Gölz, Anupam Gupta, Brett Hennig, and Ariel D. Procaccia.
**Fair algorithms for selecting citizens' assemblies.**
*Nature*, 596(7873):548–552, August 2021.

Bailey Flanigan, Gregory Kehne, and Ariel D Procaccia.
**Fair sortition made transparent.**
In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 25720–25731. Curran Associates, Inc., 2021.

Bailey Flanigan, Jennifer Liang, Ariel D. Procaccia, and Sven Wang.
**Manipulation-robust selection of citizens' assemblies.**
In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024.