Explainable Affirmative Action

Nick Arnosti, University of Minnesota
Industrial and Systems Engineering

(Based on joint work with Carlos Bonet and Jay Sethuraman)

Selection Problems

An organization must select from a set of applicants.

In contexts we consider:

- Default ranking of applicants
- Additional considerations (“Affirmative Action”)

Examples

School Seats

- High Schools (Boston: Neighborhood. New York: Free/Reduced Lunch)
- Brazilian Universities (Public School, Low Income, Minority)

US Visas

- Diversity Visas (Country & Regional Upper Quotas)
- H1B Visas (Advanced Degree Reserves)

Govt & Civil Service Positions

- Indian Civil Service (Scheduled castes and tribes, women, disability)
- Chilean Constitutional Convention (50% women, 17 indigenous, 5% disability)

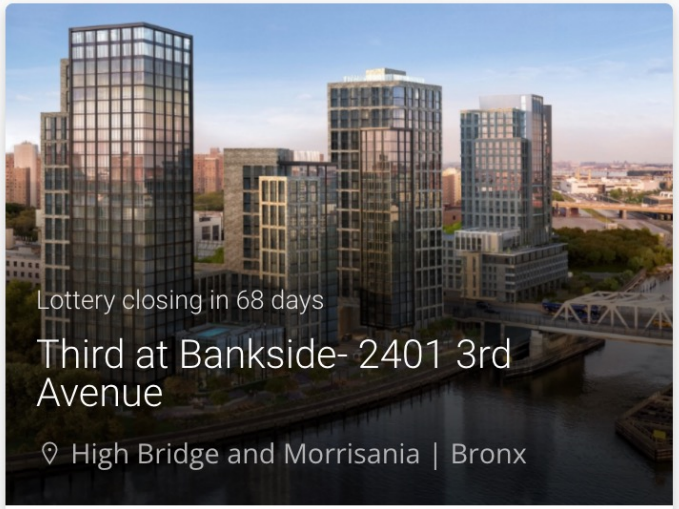
NYC Affordable Housing (Household size, income, neighborhood residence)

A Talk in Two Parts

1. Affordable housing allocation in New York City (Practice)

2. Explainable Affirmative Action (Theory)

New York Affordable Housing Lotteries

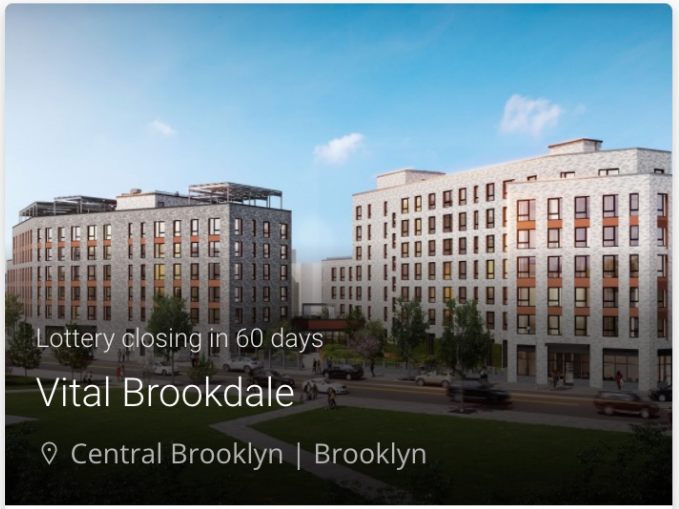


134 Units Available

Nearby Transit

2 3 4 5 6

Summary Details Map

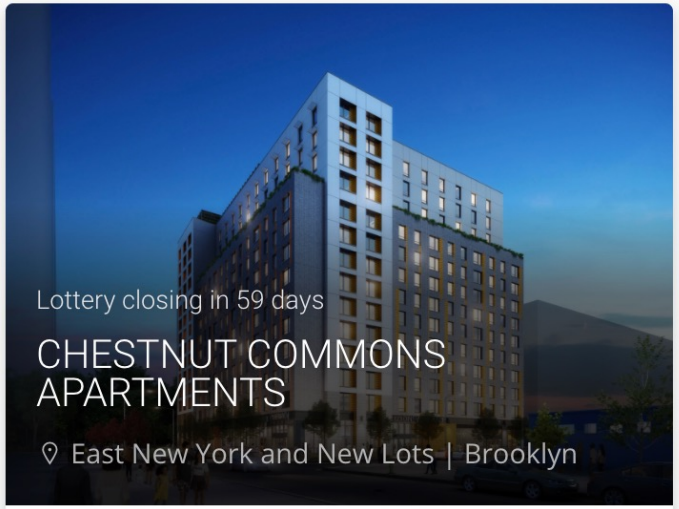


117 Units Available

Nearby Transit

A C L

Summary Details Map



219 Units Available

Nearby Transit

A J S Z

Summary Details Map

After Applications Close:

1. The city places applicants in a random order.
2. Developers screen and place applicants.

AMI %	Layout	# Units	Monthly R	Household	Household Income
40%	Studio	15 units	$625	1 person	$24,000 - $37,360
				2 people	$24,000 - $42,720
40%	1 Bedroom	11 units	$791	1 person	$30,000 - $37,360
				2 people	$30,000 - $42,720
				3 people	$30,000 - $48,040
40%	2 Bedroom	21 units	$940	2 people	$36,000 - $42,720
				3 people	$36,000 - $48,040
				4 people	$36,000 - $53,360
				5 people	$36,000 - $57,640
40%	3 Bedroom	2 units	$1,078	3 people	$41,623 - $48,040
				4 people	$41,623 - $53,360
				5 people	$41,623 - $57,640
				6 people	$41,623 - $61,920
				7 people	$41,623 - $66,200

Unit Types

Eligibility Criteria

 # bedrooms → **household size**

↓

AMI target → **household income**

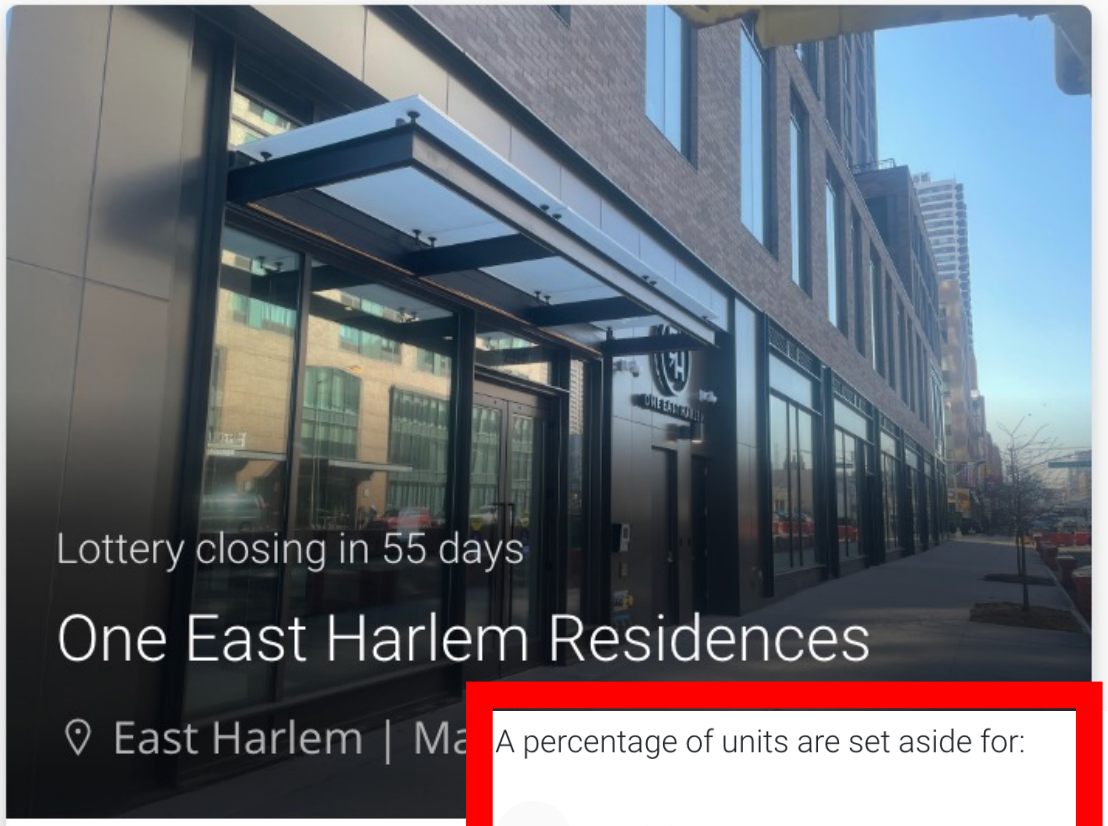
268 Units Available

Nearby Transit

2 3 4 6

Summary Details Map

AMI %	Layout	# Units	Monthly R	Household	Household Income
40%	Studio	15 units	$625	1 person	$24,000 - $37,360
				2 people	$24,000 - $42,720
40%	1 Bedroom	11 units	$791	1 person	$30,000 - $37,360
				2 people	$30,000 - $42,720
				3 people	$30,000 - $48,040
40%	2 Bedroom	21 units	$940	2 people	$36,000 - $42,720
				3 people	$36,000 - $48,040
				4 people	$36,000 - $53,360
				5 people	$36,000 - $57,640
40%	3 Bedroom	2 units	$1,078	3 people	$41,623 - $48,040
				4 people	$41,623 - $53,360
				5 people	$41,623 - $57,640
				6 people	$41,623 - $61,920
				7 people	$41,623 - $66,200
165%	Studio	23 units	$1,815	1 person	$64,800 - $154,110
				2 people	$64,800 - $176,220
165%	1 Bedroom	24 units	$2,278	1 person	$80,983 - $154,110
				2 people	$80,983 - $176,220
				3 people	$80,983 - $198,165
165%	2 Bedroom	41 units	$2,726	2 people	$97,235 - $176,220
				3 people	$97,235 - $198,165
				4 people	$97,235 - $220,110
				5 people	$97,235 - $237,765
165%	3 Bedroom	5 units	$3,142	3 people	$112,389 - $198,165
				4 people	$112,389 - $220,110
				5 people	$112,389 - $237,765
				6 people	$112,389 - $255,420
				7 people	$112,389 - $273,075



A percentage of units are set aside for:

5% Mobility

2% Vision/Hearing

Preference for a percentage of units goes t

50% Community Board Resident

5% NYC Employee

AMI %	Layout	# Units	Monthly R	Household	Household Income
40%	Studio	15 units	$625	1 person	$24,000 - $37,360
				2 people	$24,000 - $42,720
40%	1 Bedroom	11 units	$791	1 person	$30,000 - $37,360
				2 people	$30,000 - $42,720
				3 people	$30,000 - $48,040
40%	2 Bedroom	21 units	$940	2 people	$36,000 - $42,720
				3 people	$36,000 - $48,040
				4 people	$36,000 - $53,360
				5 people	$36,000 - $57,640
40%	3 Bedroom	2 units	$1,078	3 people	$41,623 - $48,040
				4 people	$41,623 - $53,360
				5 people	$41,623 - $57,640
				6 people	$41,623 - $61,920
				7 people	$41,623 - $66,200
165%	Studio	23 units	$1,815	1 person	$64,800 - $154,110
				2 people	$64,800 - $176,220
165%	1 Bedroom	24 units	$2,278	1 person	$80,983 - $154,110
				2 people	$80,983 - $176,220
				3 people	$80,983 - $198,165
165%	2 Bedroom	41 units	$2,726	2 people	$97,235 - $176,220
				3 people	$97,235 - $198,165
				4 people	$97,235 - $220,110
				5 people	$97,235 - $237,765
165%	3 Bedroom	5 units	$3,142	3 people	$112,389 - $198,165
				4 people	$112,389 - $220,110
				5 people	$112,389 - $237,765
				6 people	$112,389 - $255,420
				7 people	$112,389 - $273,075

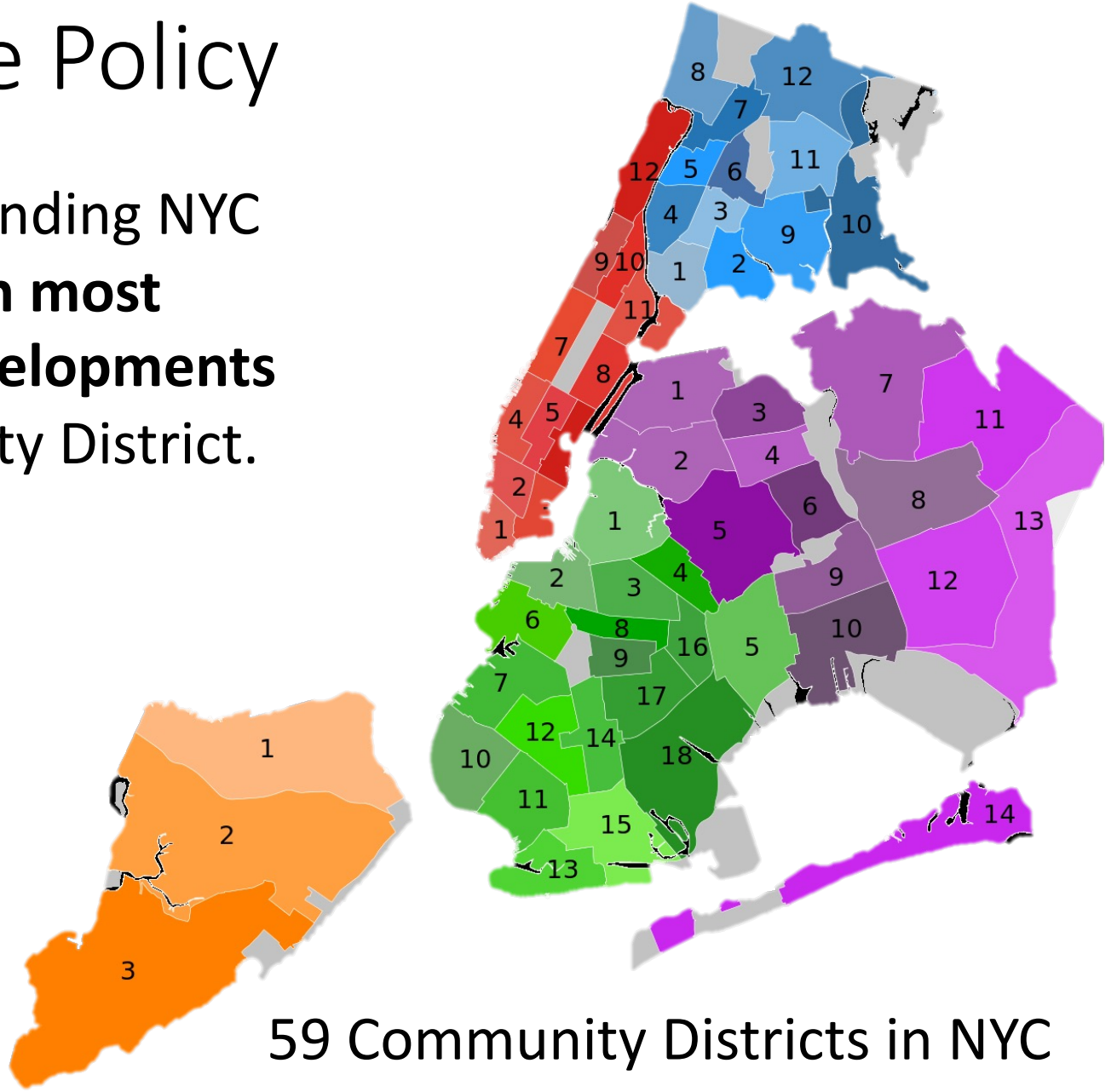
Community Preference Policy

Community Preference is a longstanding NYC policy that **reserves 50% of units in most subsidized affordable housing developments** for residents of the local Community District.

Lawsuit!

Plaintiffs claim that the policy has racially discriminatory impacts and perpetuates the harmful legacy of segregation.

Learn more at my blog:
nickarnosti.com/blog



59 Community Districts in NYC

Quick Recap

Units have different #bedrooms, AMI targets.

“Preferences” (quotas) for people with disabilities, community residents, and city employees.

Lottery numbers, household size and income for each applicant.

How do they handle this all?

NYC Developer Screening Algorithm

1. Order applicants by lottery ("log") number.
2. Satisfy preferences:
   I. Applicants eligible for **disability preference** considered in log order until 7% of units have been allocated.
   II. Applicants eligible for **community preference** considered in log order until 50% of units filled with people from community district.
   III. Applicants eligible for **municipal employee preference** considered in log order until 5% of units filled with municipal employees.
3. **Remaining applicants** considered in log order until all units filled.

Details:

Granting someone with multiple preferences a unit counts against each preference's goal.

If household is eligible for multiple available unit types, it chooses which to claim.

What are consequences of using this algorithm?

My claim:

- Most low-income units go to community residents.
- Most middle-income units go to non-residents.

Two key facts:

- Competition is highest for low-income units.
- Most applicants are not from the community district.

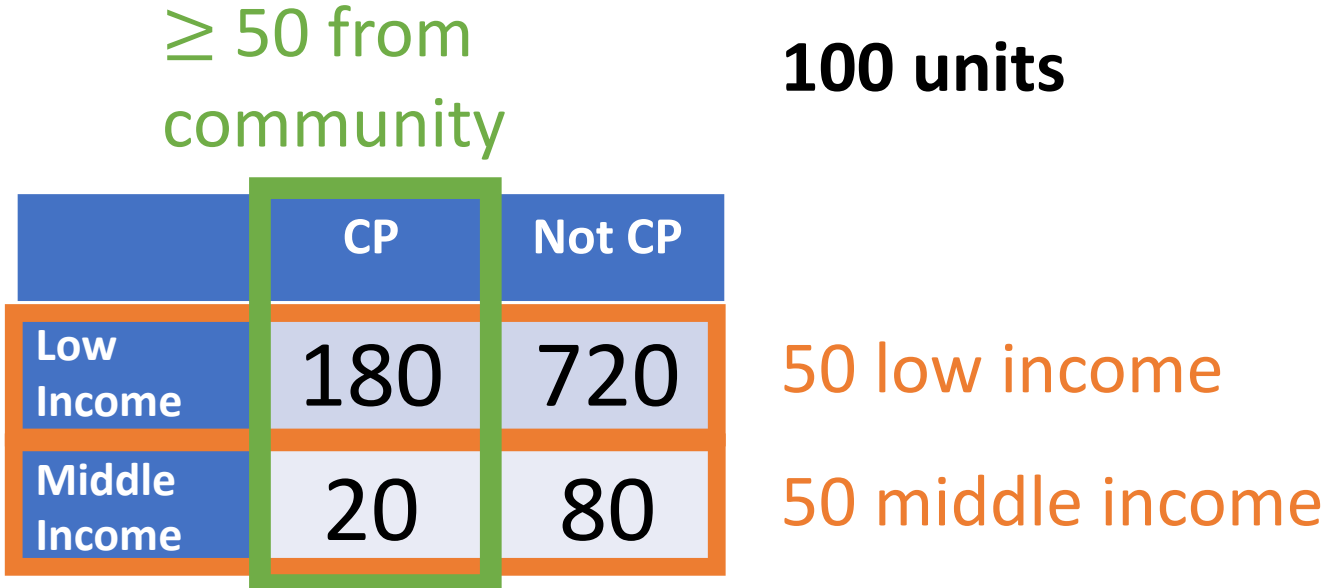
A Simple Example

Assume:

- All units have same number of bedrooms.
- Two non-overlapping target income ranges.
- No disability or municipal employee preferences
- All applicants are eligible and will accept.

1000 applicants:

- 20% from community district
- 90% low income
- Income and community preference are independent.



The effect of community preference

No Community Preference

	CP	Not CP
Low Income	10	40
Middle Income	10	40

50% Community Preference

	CP	Not CP
Low Income	45	5
Middle Income	10	40

Community members get

- **90%** of low-income seats
- **20%** of middle income seats (same as without CP policy!)

The city does **NOT** reserve 50% of each type of unit for community members.

- Community members (like other applicants) are mostly low income.
- They claim most low-income units.
- Few low-income units left for applicants from outside the community.

Why is this happening? (more detail)

First pass: take top 50 community residents.

Community Preference Satisfied

50/200 = 25%

	CP	Not CP
Low Income	45	5
Middle Income	10	40

Second pass:

Fill remaining units

5/720 = 0.7%

50/100 = 50%

Low Income Units Fill

Middle Income Units Fill

Challenge: Explain to person #6 why they were not selected.

Inevitable consequence of stated policies? NO.

At least 50 community members.
At most 50 low income.
At most 50 middle income.

	CP	Not CP
Low Income	45	5
Middle Income	10	40

Conclusion:
Processing algorithm is not carefully considered, but has a large effect!

	CP	Not CP
Low Income	30	20
Middle Income	20	30

Affirmative Action Gone Wrong

Boston Schools	50% of seats reserved for neighborhood students. Almost no effect! Algorithms may not have intended effect.
India Civil Service,	Ranked by exam score. Reservations for minorities. Declaring minority status can actually harm applicants!
Brazilian Universities	Algorithms may have *opposite* of intended effect.
NYC Affordable Housing	Community Preference primarily affects low-income applicants. Algorithms may have unintended *side effects.*

Related Literature

- ***Reserve Design: Unintended Consequences and the Demise of Boston's Walk Zones.*** Dur, Kominers, Pathak, Sonmez (2018)
- ***College Admission with Multidimensional Privileges: The Brazilian Affirmative Action Case.*** Aygun, Bo (2021)
- ***Immigration Lottery Design: Engineered and Coincidental Consequences of H-1B Reforms.*** Pathak, Rees-Jones, Sonmez (2022)
- ***Affirmative Action in India via Horizontal, Vertical, and Overlapping Reservations.*** Sonmez, Yenmez (2022)
- ***How to De-reserve Reserves: Admissions to Technical Colleges in India.*** Aygun, Turhan (Forthcoming)

My Take

Each paper is inspiring:

- Attention to institutional detail
- Interesting theoretical questions
- Could help practitioners use these systems effectively.

But collectively, they are concerning:

- Idiosyncratic models and axioms.
- So many mistakes are made!
- Feels like “whack a mole”.

We don’t want to become intuit. TurboTax !

Motivation For This Work

Instead of helping people use complicated systems, design simple ones!

My Goals:

1. **Unify the literature.**
   Describe existing and proposed algorithms using common framework.

2. **Identify simple/explainable policies.**
   Hopefully, these are less prone to user error!

Preview of Remainder of Talk

1. Axiomatic characterization of explainability: "outcome-based" rules.
   Only rules that are **monotonic, lower invariant,** and **non-bossy.**

   Selection Rule 1
   Hire top woman, then
   Hire top remaining candidate.

   Selection Rule 2
   Hire top candidate, then
   Hire top remaining woman.

2. Informal advocacy for outcome-based selection rules

3. Special Cases of interest: reserves and quotas. Unifying the literature

Model

Set of individuals I.

Selection Rule ψ: function from priority order $\succ$ on I to 2^I.

Why define selection rule in this way?

Example 1: select top k applicants.

Example 2: NYC Affordable Housing Developer Algorithm.

I. Consider applicants eligible for **disability preference** in log order until 7% of units have been allocated.
II. Consider applicants eligible for **community preference** in log order until 50% of units filled with people from community district.
III. Consider applicants eligible for **municipal employee preference** in log order until 5% of units filled with municipal employees.
IV. Consider **remaining applicants** in log order until all units filled.

What do I mean by explainable? (Intuition)

Changes to the priority list have predictable consequences.

Can explain to someone why they were not selected.

Explainability in Three Axioms

A selection rule ψ is:

Monotonic if increasing priority of i (fixing others' priority) weakly helps i.

Lower Invariant if i's outcome does not depend on the ranking of individuals with lower priority than i.

Non-bossy if any change to the priority of i that does not affect i's outcome also does not affect anybody else's outcome.

Can we get all three?

“Outcome-Based” Selection Rules

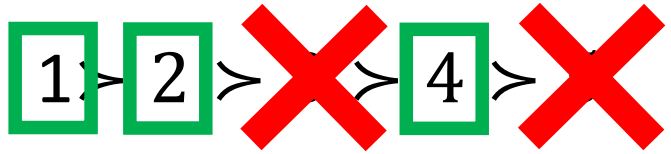
Instead of describing a selection algorithm explicitly,

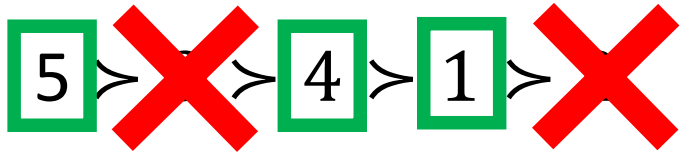
Defined (implicitly) by feasible outcomes $\mathcal{F} \subseteq 2^I$.

Feasible outcomes $\mathcal{F}$ → selection rule $\psi_{\mathcal{F}}$:

Consider individuals in priority order.

Accept i if exists $S \in \mathcal{F}$ containing $\{i\} \cup \{j \succ i : j \in \psi(\succ)\}$





Characterization Result

Theorem.

For any $\mathcal{F} \subseteq 2^S$, $\psi_{\mathcal{F}}$ is monotonic, lower invariant, and non-bossy.

If ψ is monotonic, non-bossy, and lower invariant, then $\psi = \psi_{\mathcal{F}}$ for some $\mathcal{F} \subseteq 2^S$.

What does outcome-based selection offer?

1. A criteria that rules out some current policies.

2. A new description of existing policies (*outcomes*, not *algorithms)*

Algorithmic vs Outcome-based Descriptions

	Selection Rule 1	Selection Rule 2
Algorithmic Description	Hire top woman, then Hire top remaining candidate.	Hire top candidate, then Hire top remaining woman.
Outcome Description	Selections must have 2 people, at least 1 woman.	

These are often confused:

Matching with slot-specific priorities: Theory. Kominers, Sonmez (2016)

Reversing Reserves. Pathak, Rees-Jones, Sonmez (2022)

Selection Rule 2 is Bossy

Hire top candidate, then

Hire top remaining woman.

Applicant 1 is a man. Applicants $2, 3$ are women.

- If order is $1 \succ 2 \succ 3$, we choose $\{1,2\}$.
- If order is $2 \succ 1 \succ 3$, we choose $\{2,3\}$.

Increasing 2's priority didn't affect 2, but affected 1 and 3.

Put another way: if $\{1,2\}$ is feasible, then an outcome-based selection rule still chooses it when order is $2 \succ 1 \succ 3$.

Algorithmic vs Outcome-based Descriptions

	Selection Rule 1	Selection Rule 2
Algorithmic Description	Hire top woman, then Hire top remaining candidate.	Hire top candidate, then Hire top remaining woman.
Outcome Description	Selections must have 2 people, at least 1 woman.	

Advantages of Outcome-Based Descriptions

1. **Easier to incorporate public opinion ex ante.**
   People debate priorities, discuss outcomes. Little attention to algorithms.

2. **Simpler explanation for final outcome ex post.**

Advantage #1: Incorporating Public Opinion

Metaphor: Making a Chair

I have no idea which knobs to adjust!

But I can describe desirable output:

- Seat at 18”-20”
- Supports up to 300 lbs
- Has a backrest

Advantage #2: Explaining Final Outcome

	Selection Rule 1	How to explain why someone wasn't chosen?
Algorithmic Description	Hire top woman, then Hire top remaining candidate.	"Here's the algorithm. It never chooses you."
Outcome Description	Selections must have 2 people, at least 1 woman.	"Choosing you would require violating constraints or displacing a higher-priority applicant"

Preview of Remainder of Talk

1. Axiomatic characterization of explainability: "outcome-based" rules.
   Only rules that are **monotonic, lower invariant,** and **non-bossy.**

Selection Rule 1
Hire top woman, then
Hire top remaining candidate.

Selection Rule 2
Hire top candidate, then
Hire top remaining woman.

2. Informal advocacy for outcome-based selection rules

3. Special Cases of interest: reserves and quotas. Unifying the literature

Allocative Efficiency

Values $v_i \in \mathbb{R}_+$ such that $i \succ i' \Rightarrow v_i \geq v_{i'}$.

Maximize $\sum_{i \in S} v_i$
Subject to $S \in \mathcal{F}$.

May not coincide with $\psi_{\mathcal{F}}$:

- If $\mathcal{F} = \{\{1\}, \{2,3,4\}\}$, and $1 \succ 2 \succ 3 \succ 4$, then $\psi_{\mathcal{F}}(\succ) = \{1\}$.
- If $v_i = 1$ for all i, then $\{2,3,4\}$ is higher value.

In essence $\psi_{\mathcal{F}}$ tries to solve a maximization problem greedily.

When is this guaranteed to work?

When does greedy selection maximize total value?

Matroids

Define $L(\mathcal{F}) = \{S : S \subseteq S' \in \mathcal{F}\}$.

Say that $\mathcal{F}$ **induces a matroid** if $\left(I, L(\mathcal{F})\right)$ is a matroid.

Proposition. If $\mathcal{F}$ induces a matroid, then $\psi_{\mathcal{F}}$ maximizes total value for any values consistent with $\succ$.

Two Natural (Common) Feasibility Structures

Reserves "One to One" Convention

- Each position reserved for people with certain attributes.
- Always induce a matroid.
- Widespread algorithms don't find optimal feasible selection.

Quotas "One to All" Convention

- Minimums and Maximums for different groups.
- Only induce matroid if groups are nested/laminar/hierarchical.

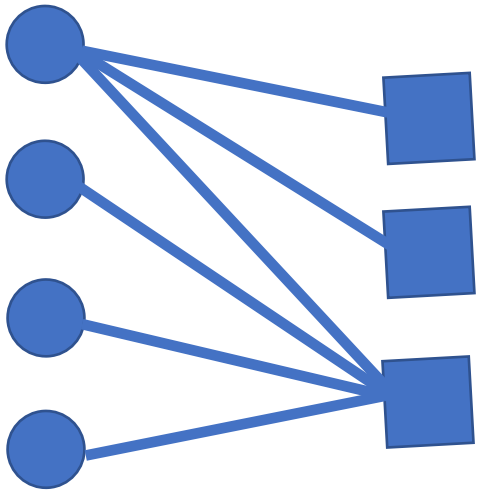
Affirmative Action with Overlapping Reserves. Sonmez, Yenmez.

Matching for the Israeli "Mechinot" Gap-Year Programs: Handling Rich Diversity Requirements. Gonczarowski, Kovalio, Nisan, Romm (2019).

Reserves

Finite set of positions P, bipartite compatibility graph $G = (I, P, E)$.

A selection S is feasible if there exists a matching $\mu : S \to P$ such that for each $i \in S$, $\big(i, \mu(i)\big) \in E$.

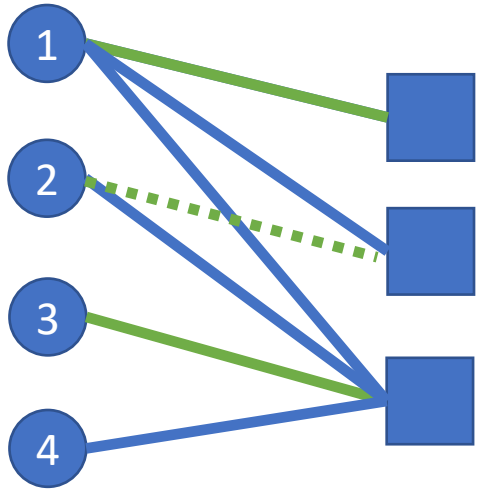


Examples:

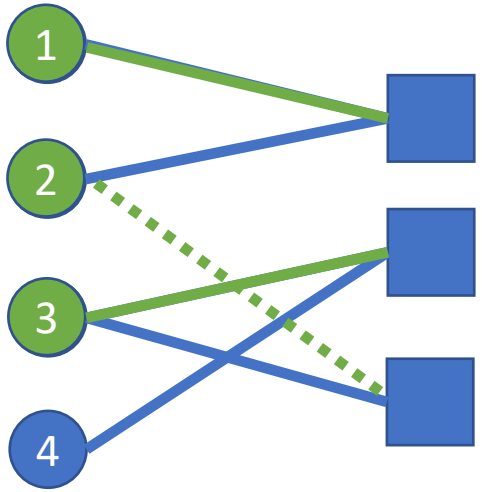
- Schools reserve seats for low-income/neighborhood students
- H1B visas reserved for people with graduate degrees from US schools
- Affordable housing reserved for households of appropriate size & income.

Theorem. For any compatibility graph G, $\mathcal{F}(G)$ induces a matroid.

Soft Reserves ("De-Reservation")



No way to select 3 people!



If we allow soft reserves, can we select applicants 1, 2, 3?

Theorem. For any graph G, $\mathcal{F}^{hard}(G) \subseteq \mathcal{F}^{maximum}(G) \subseteq \mathcal{F}^{maximal}(G)$, and all three induce a matroid.

Quotas

Finite set of traits T. For each $t \in T, I_t \subseteq I$.

Each trait t has a lower quota l_t and an upper quota u_t.

A selection S is feasible if for each $t \in T$, $l_t \leq |S \cap I_t| \leq u_t$.

People count towards quota for every category they belong to.

Traits form a **hierarchy** if for each $t, t' \in T$, either $I_t \cap I_{t'} = \emptyset$ or $I_t \subseteq I_{t'}$ or $I_{t'} \subseteq I_t$.

Also called laminar, nested.

Theorem. If traits form a hierarchy, then $\mathcal{F}(T, l_t, u_t)$ induces a matroid.

Summary

1. Affirmative action algorithms often have unintended consequences.

2. We propose **outcome-based selection rules**.
   - Only these rules are monotonic, lower invariant, and non-bossy.
   - Place focus on outcomes, not algorithms.
   - Simpler justification ex post.
   - Flexible enough to encode many considerations.
   - Question: are they a good idea if feasible selections not a matroid?

Discussing interpretability/explainability is new to me.

I want feedback on what you found convincing/not convincing!