

Centralized Selection with Preferences in the Presence of Biases

L. Elisa Celis
Yale University

Amit Kumar
IIT Delhi

Nisheeth K. Vishnoi
Yale University

S. Andrew Xu
Yale University

Yale

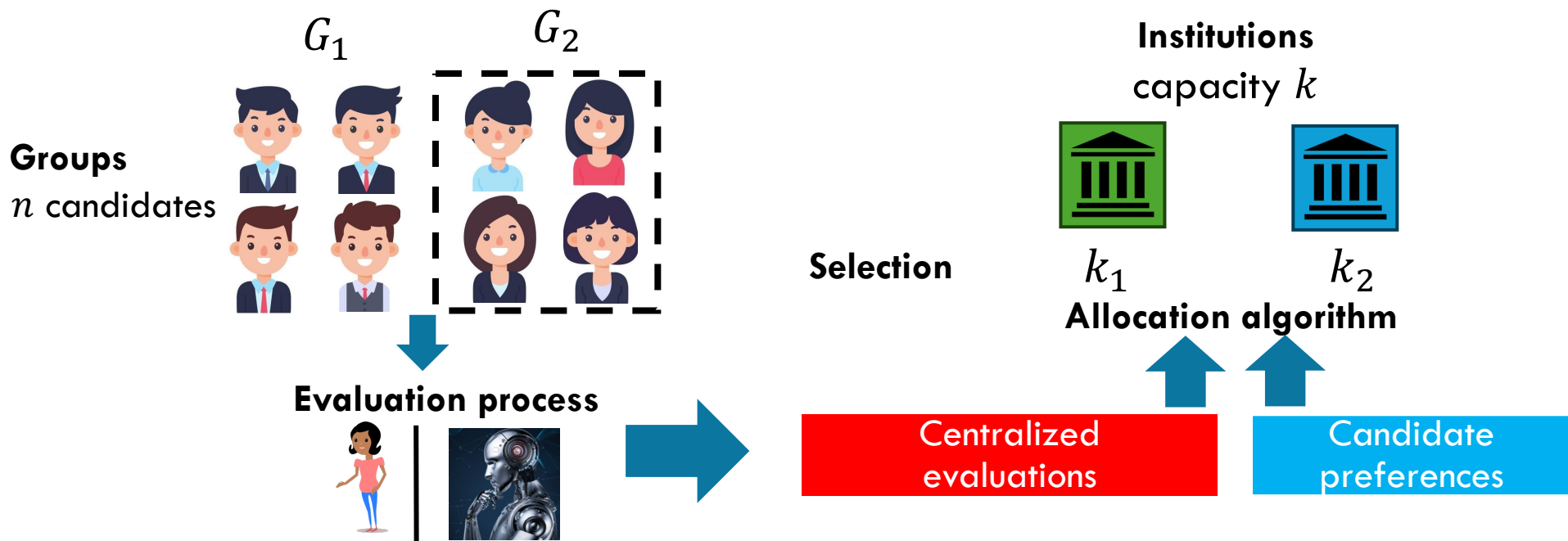


Poster

Date & Time: Wednesday, July 24, 1:30-3pm

Location: Hall C 4-9 #34807

Centralized Selection



Each candidate has a **preference list** over p institutions



Goal: Find an allocation that maximizes utility

Models: IIT JEE, China Gaokao, online labor markets, etc.

Algorithm: Can use Gale-Shapley algorithm to find utility maximizing and stable allocation (works even when institutions have different evaluations for candidates)

However, observed evaluations may be biased affecting both utility and "fairness" ...

Evaluations May Be Biased

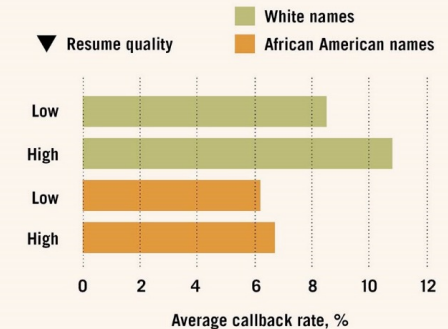
Systemic Biases: Certain groups may not be able to perform well in the evaluation processes due to unequal opportunities/information ... Or evaluation process may inadvertently benefit one group of people over another

Implicit Biases: Unconscious attribution of qualities (or lack thereof) to members of particular group: based on e.g., gender, ethnicity, or race

- Women receive lower competence scores than men in peer-reviews [Wennerås & Wold '97]
- White names receive 50% more callbacks for interviews than African-American names [Bertrand & Mullainathan '04]

Racism in a resume

Job applicants with African American-sounding names got fewer callbacks.



Source: Bertrand and Mullainathan, 2004

[Source]: "Are Emily and Greg...?"
Brooke C. Medium.com





The measured (estimated) utility may not be an accurate representation of the candidate's true (latent) utility. This can adversely affect the opportunities of candidates from disadvantaged groups while also reducing total utility for institutions. Assume there is no a priori reason that the ability of individuals depends on their socio-economic attributes. We also assume preferences are identically distributed

How do we ensure fairness and maximal utility in an assignment process with multiple institutions in this centralized setting?

Model

- n candidates apply to $p > 1$ institutions, capacity k_1, k_2, \dots, k_p
 - Candidate i has **latent** utility $u_i \geq 0$ and preference list σ_i over institutions
 - Candidates are in group G_1 (advantaged) or G_2 (disadvantaged)
- u_i drawn iid from distribution \mathcal{D} ; σ_i drawn iid from distribution \mathcal{L}

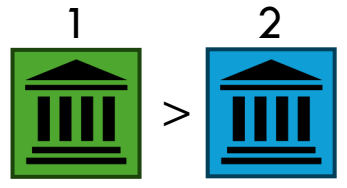
Example: $\beta = 0.5$

		\mathcal{A}_1	\mathcal{A}_2
	G	u_i	\hat{u}_i
	G_1	6	6
	G_1	8	8
	G_2	8	4
	G_2	6	3

β -Bias Model in [Kleinberg & Raghavan '18]

- A simple multiplicative model **closely models empirical data** from human-subject studies (e.g. in nepotism and sexism) and for power law distributions [Wennerås & Wold '97]
- Bias parameter β , $0 < \beta \leq 1$, and the estimated utility \hat{u}_i for agents in group G_2 is $\beta \cdot u_i$ and for agents in G_1 is u_i
 - Assumes that β is **unknown**, which aligns with the **one-round** setting of the centralized selection problems
 - Easily generalized to multiple groups (more bias parameters)
 - Assume $\beta \leq 1$ since an agent in G_2 is systemically underestimated

Uniform preferences



$$k_1 = k_2 = 2$$

 = no assignment

How do we measure fairness and utility?

Fairness and Utility Metrics

- Each metric evaluates an algorithm \mathcal{A} , results in a value in $[0,1]$ (higher is better), and generalizes to more groups
- We utilize two fairness metrics, considering both overall and top-choice representation for candidates in different groups

1 - Utility Ratio: Measures the ratio of the expected **latent** utility of a matching ($M_{\hat{u},\sigma}$) on observed utility (\hat{u}) divided by the maximum total utility (U^*)

$$U(\mathcal{A}) = \mathbb{E}_{u \sim \mathcal{D}, \sigma \sim \mathcal{L}} \frac{U(M_{\hat{u},\sigma})}{U^*(u)}$$





2 - Representational Fairness: Compares the representation of groups in **all institutions**, where ρ_j is the fraction of candidates in G_j that get selected by $M_{\hat{u},\sigma}$ [Barocas et al. '19]

$$\mathcal{R}(\mathcal{A}) = \mathbb{E}_{u \sim \mathcal{D}, \sigma \sim \mathcal{L}} \frac{\min\{\rho_1, \rho_2\}}{\max\{\rho_1, \rho_2\}}$$

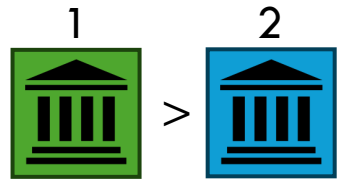
3 - Preference-based Fairness: Compares the representation of groups in **top choice institutions** by $M_{\hat{u},\sigma}$, where $\pi_j^{(\ell)}$ is the fraction of candidates in G_j that get a top- ℓ choice (e.g. $\ell = 1$)

$$\mathcal{P}^{(\ell)}(\mathcal{A}) = \mathbb{E}_{u \sim \mathcal{D}, \sigma \sim \mathcal{L}} \frac{\min\{\pi_1^{(\ell)}, \pi_2^{(\ell)}\}}{\max\{\pi_1^{(\ell)}, \pi_2^{(\ell)}\}}$$

Example: $\beta = 0.5$

		\mathcal{A}_1	\mathcal{A}_2
	G	u_i	\hat{u}_i
	G_1	6	6
	G_1	8	8
	G_2	8	4
	G_2	6	3

Uniform preferences



$$k_1 = k_2 = 2$$

- $\mathcal{R}(\mathcal{A}_1) = \mathcal{R}(\mathcal{A}_2) = 1$
- $\mathcal{P}^{(1)}(\mathcal{A}_1) = 1$
- $\mathcal{P}^{(1)}(\mathcal{A}_2) = 0$

Results 1: Traditional Methods Fail

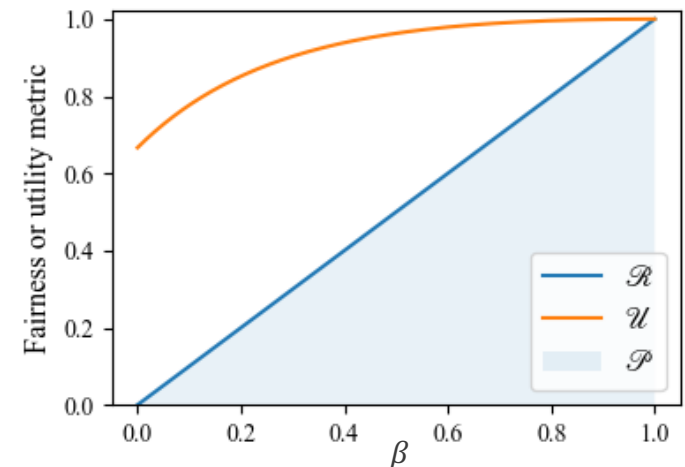
Define \mathcal{A}_{st} as the standard Gale-Shapley algorithm to construct a stable matching [Gale & Shapley '62]. Individuals are sorted by decreasing estimated utility, then they are assigned their top choice institution with available capacity

Theorem: Consider an instance where the utilities of the candidates are drawn from the uniform distribution on $[0, 1]$ and the distribution over preferences is arbitrary.

Assume $n_1 = n_2$. Then, $\mathcal{P}(\mathcal{A}_{st}) \leq \beta \pm O\left(\frac{\beta\sqrt{\log n}}{\sqrt{n}}\right)$, $\mathcal{R}(\mathcal{A}_{st}) = \beta \pm O\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right)$, and

$$\mathcal{U}(\mathcal{A}_{st}) = \frac{2}{3} + \frac{4\beta}{3(\beta+1)^2} \pm O\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right)$$

Observation: As $\beta \rightarrow 0$, $\mathcal{P}(\mathcal{A}_{st}) \rightarrow 0$, $\mathcal{R}(\mathcal{A}_{st}) \rightarrow 0$, and $\mathcal{U}(\mathcal{A}_{st}) \rightarrow \frac{2}{3}$. \mathcal{A}_{st} **cannot guarantee optimal utility or fairness.** Even beyond this setup, the results generalize to arbitrary n_1, n_2 and instances where the utilities are drawn from *any* log-concave distribution over $[0, 1]$; the given bounds generally hold in such settings



How do we implement fairness constraints and improve optimality?

Results 2: Institution-Wise Constraints Work

Define \mathcal{A}_{group} as \mathcal{A}_{st} with proportional group-wise representational constraints [Celis et al. '20]. We implement this by running \mathcal{A}_{st} on the top $|G_j|/n$ candidates in each group G_j . **Observe** that \mathcal{A}_{group} can enforce representational fairness with near optimal utility but can have low preference-based fairness

Define $\mathcal{A}_{inst-wise}$ based on \mathcal{A}_{st} . For each institution ℓ , we require it to have $k_\ell \cdot |G_j|/n$ candidates from each group G_j assigned to it. We create two instantiations of each institution with proportional capacities, then run \mathcal{A}_{st} on the respective group

Theorem: Let $\eta_1, \eta_2, \eta_3 > 0$ be parameters such that $|G_j| \geq \eta_1 n$ for $j \in \{1, 2\}$, $K = \sum_{i=1}^p k_i \geq \eta_2 n$, and $k_\ell \geq \eta_3 K$ for each $\ell \in [p]$. There is an algorithm $\mathcal{A}_{inst-wise}$ such that, for any distribution of utilities and preference lists, and bias parameter β ,

$$\mathcal{P}(\mathcal{A}_{inst-wise}) \geq 1 - O\left(\frac{p\sqrt{\log K}}{\eta_1 \eta_3 \sqrt{K}}\right), \mathcal{R}(\mathcal{A}_{st}) = 1, \text{ and } \mathcal{U}(\mathcal{A}_{st}) \geq 1 - O\left(\frac{\sqrt{\log n}}{\sqrt{\eta_2 n}}\right)$$

Observation: $\mathcal{A}_{inst-wise}$ guarantees high preference-based and representational fairness while maintaining near-optimal utility. β also does not need to be known. This algorithm is also group-wise stable, Pareto-efficient, and strategy-proof and results generalize beyond this setup

Proof Ideas

Result 1: Bounding utility and fairness guarantees of \mathcal{A}_{st}

- A Lipschitz property of the assignment given by \mathcal{A}_{st} allow us to derive tight concentration bounds on the number of selected candidates when preference lists are drawn from a distribution: *changing the preference list of only one candidate changes the assignment slightly*
- It suffices to bound the expected values of the proportion of selected people in each group that receive their top choices, bounding $\mathcal{P}(\mathcal{A}_{st})$ and $\mathcal{R}(\mathcal{A}_{st})$
- $\mathcal{U}(\mathcal{A}_{st})$ can be estimated by evaluating the expected utility of the top S_j candidates from each group G_j , where S_j is the expected number of selected candidates from G_j

Result 2: Bounding $\mathcal{A}_{inst-wise}$

- $\mathcal{R}(\mathcal{A}_{inst-wise}) = 1$ follows from the construction of the algorithm
- $\mathcal{U}(\mathcal{A}_{inst-wise}) \approx 1$ also follows because of the algorithm and iid assumptions
- Bounding $\mathcal{P}(\mathcal{A}_{inst-wise})$ requires a non-trivial proof that bounds the gap between the fraction of candidates in the two groups who receive top choices
- This relies on another Lipschitz property of \mathcal{A}_{st} : *changing the capacity of an institution by one unit changes the allocation for at most p candidates*
- This allows us to show that $\mathcal{P}(\mathcal{A}_{inst-wise}) \approx 1$

Empirical Results

- We show our algorithm's efficacy in real-world evaluation processes using data from India's centralized IIT-JEE 2009 examination
 - Using gender and birth-category as two protected attributes and varying the unknown preference distributions, we find $\mathcal{P}^{(1)}(\mathcal{A}_{inst-wise}) \geq 0.90$ while $\mathcal{P}^{(1)}(\mathcal{A}_{st}) \leq 0.25$
- We also use simulated data when groups have different preference distributions (e.g. HBCUs in the US)
 - We test on utilities distributed under Gaussian and Pareto distributions
 - We find that $\mathcal{P}^{(1)}(\mathcal{A}_{inst-wise}) \geq 0.75$ while $\mathcal{P}^{(1)}(\mathcal{A}_{st}) \leq 0.30$
- We find that our algorithm maintains high fairness and near-optimal utility and is robust when assumptions are not followed while \mathcal{A}_{st} and \mathcal{A}_{group} may fail to result in either high fairness or optimal utility
- We also test \mathcal{A}_{st} beyond the theoretical setup and find the results generally hold

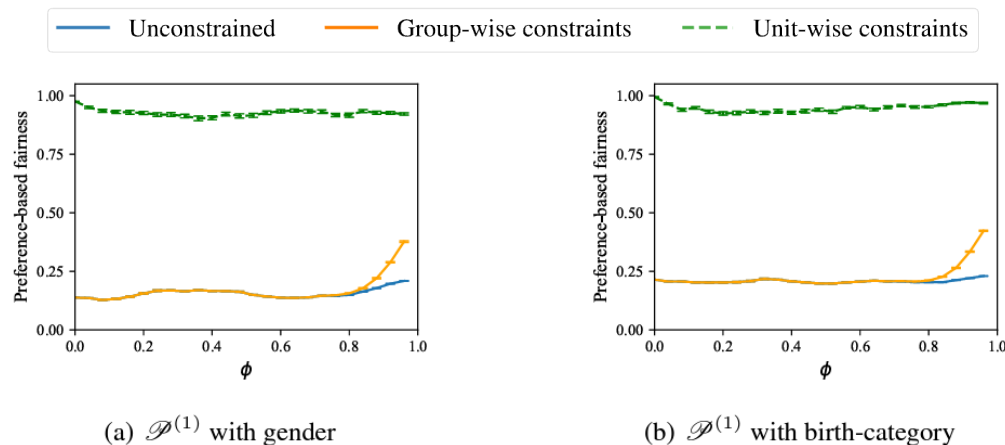


Figure: Preference-based fairness as measured by $\mathcal{P}^{(1)}$ using either gender or birth-category as the protected attribute with data from the 2009 JEE test. The x-axis denotes ϕ , the dispersion parameter of the Mallows preference distribution. Error bars denote the standard error of the mean over 50 iterations.

Conclusion, Limitations, and Future Work

- Biases in evaluation processes may lead to suboptimal results not only for candidates, **but also for institutions**
- We present a family of institution-wide constraints for the multiple institution centralized selection problem
 - They provably achieve near-optimal utility and preference-based fairness with minimal error terms
- We empirically validate our model and present an algorithm, *$A_{inst-wise}$* , that can be used under real-world data, maintaining fairness and high utility in situations where *A_{st}* fails
- Extending work beyond the iid assumption for utilities is an interesting direction
- Future work could also extend beyond centralized selection to the setting where institutes have different evaluations

Thanks!

Poster

Date & Time: Wednesday, July 24, 1:30-3pm

Location: Hall C 4-9 #34807

Paper: <https://openreview.net/pdf?id=9QRcp2ubDt>

Code: <https://github.com/sandrewxu/CentralizedSelectionwithPreferenceBias>