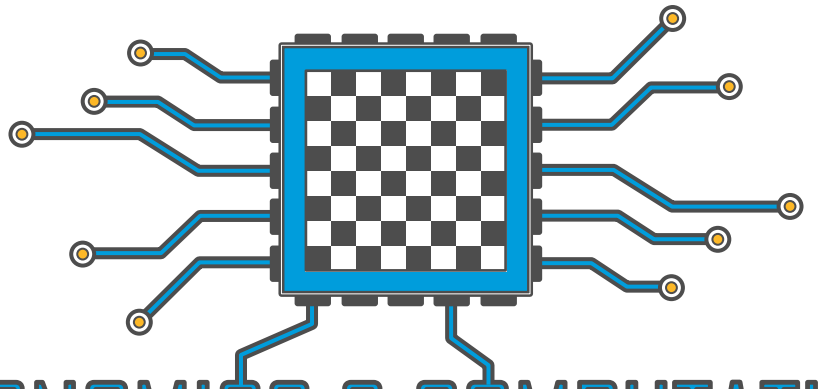


Spring 2025 | Lecture 22

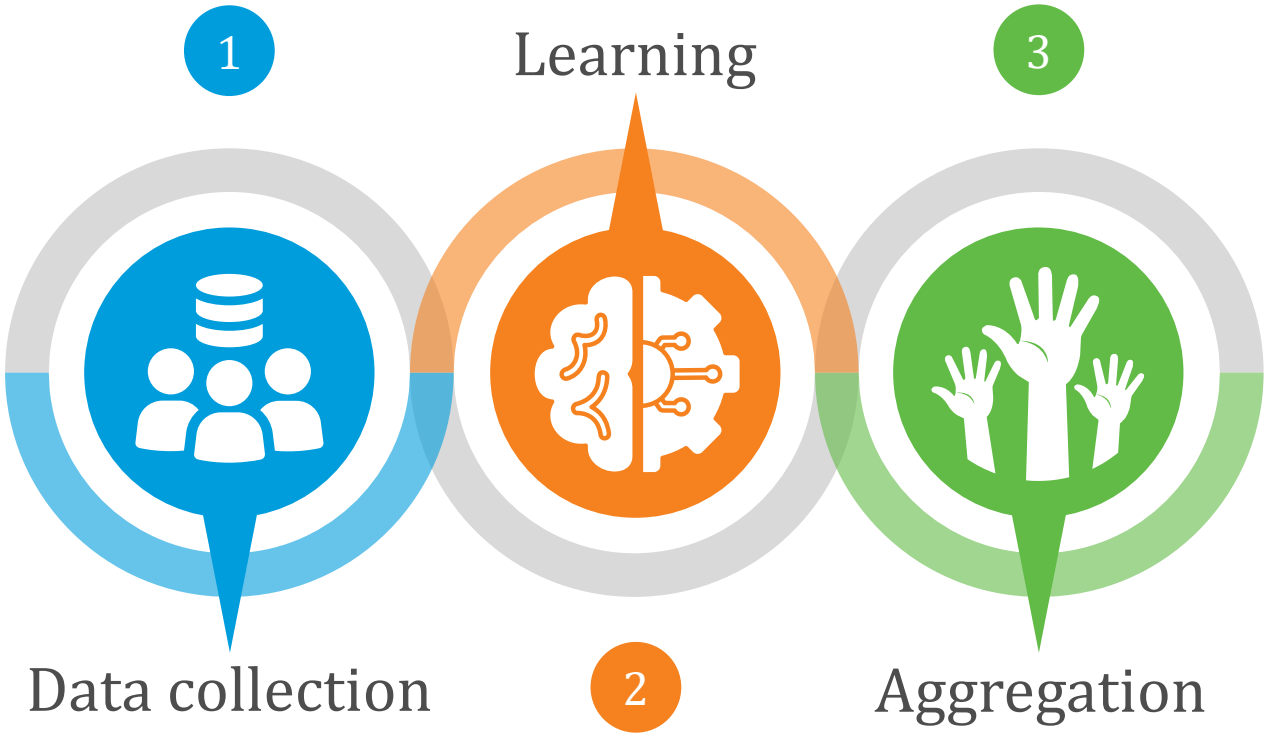
Virtual Democracy

Ariel Procaccia | Harvard University

SOCIAL CHOICE AND AI

- How should we align AI with human values? That's the trillion-dollar question
- Social choice is playing a bigger and bigger role in answering it, especially when it comes to the current practice of fine-tuning LLMs using reinforcement learning from human feedback
- We'll talk about a related approach to automating decisions through social choice and machine learning

VIRTUAL DEMOCRACY FRAMEWORK

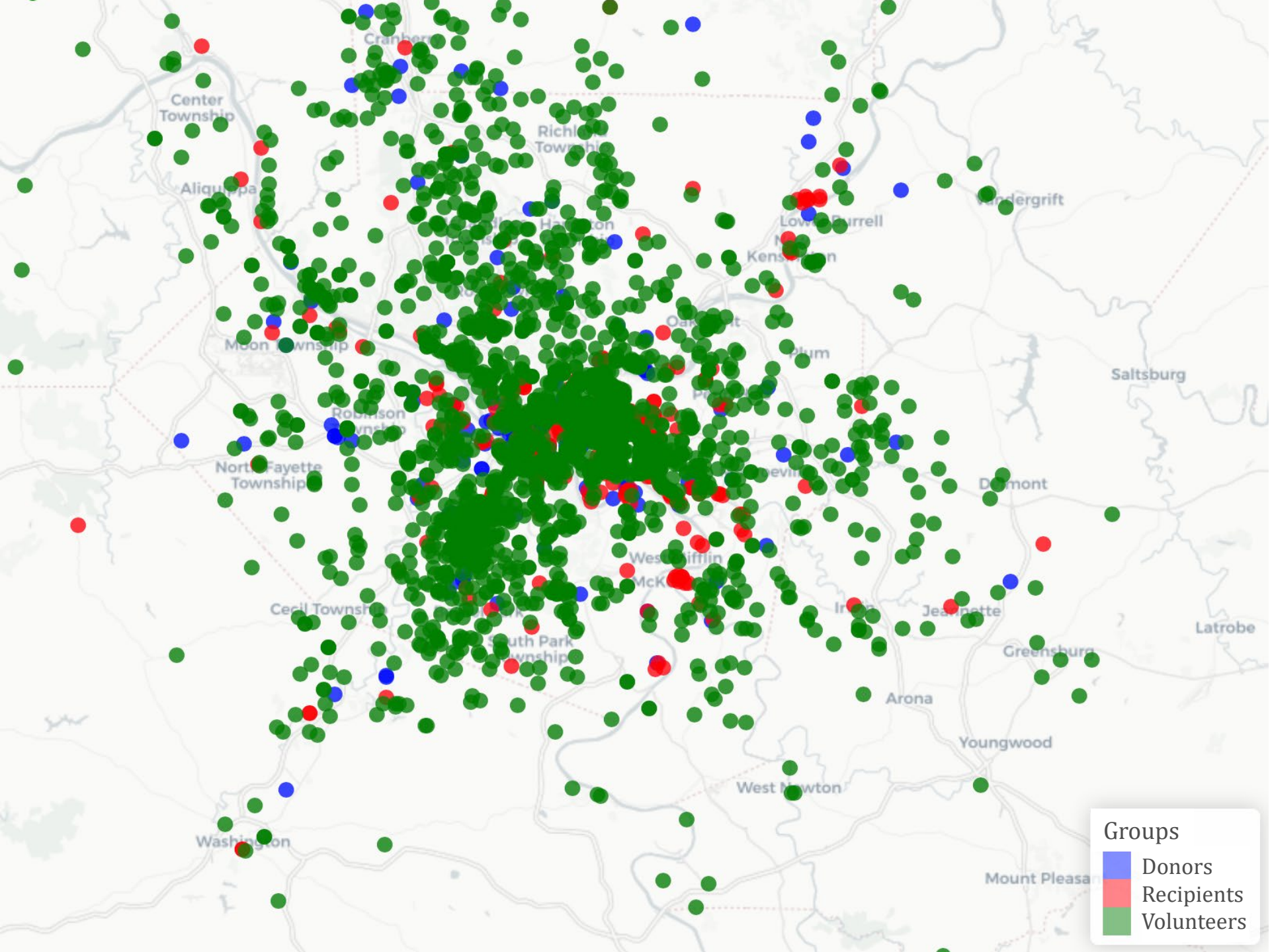


FOOD RESCUE

- We'll instantiate the virtual democracy framework in the context of food rescue
- The goal is to design a recommendation system that suggests which recipient organization should receive each incoming food donation
- All of the details of the instantiation and empirical results are from a paper by Lee et al. (2019)

FOOD RESCUE


Donors
ALDI
TRADER JOE'S
TARGET
WHOLE FOODS MARKET
EST. 1983
BRUEGGER'S BAGELS
AUTHENTIC NEW YORK STYLE
412 • FOOD
RESCUE
Recipients
YMCA
MEALS on WHEELS
THE SALVATION ARMY


Center Township
Moon Township
Robinson
North Fayette Township
Cecil Township
South Park Township
Saltsburg
Latrobe
Greensburg
Arona
Youngwood
West Newton
Washington
Groups
Donors
Recipients
Volunteers

DATA COLLECTION

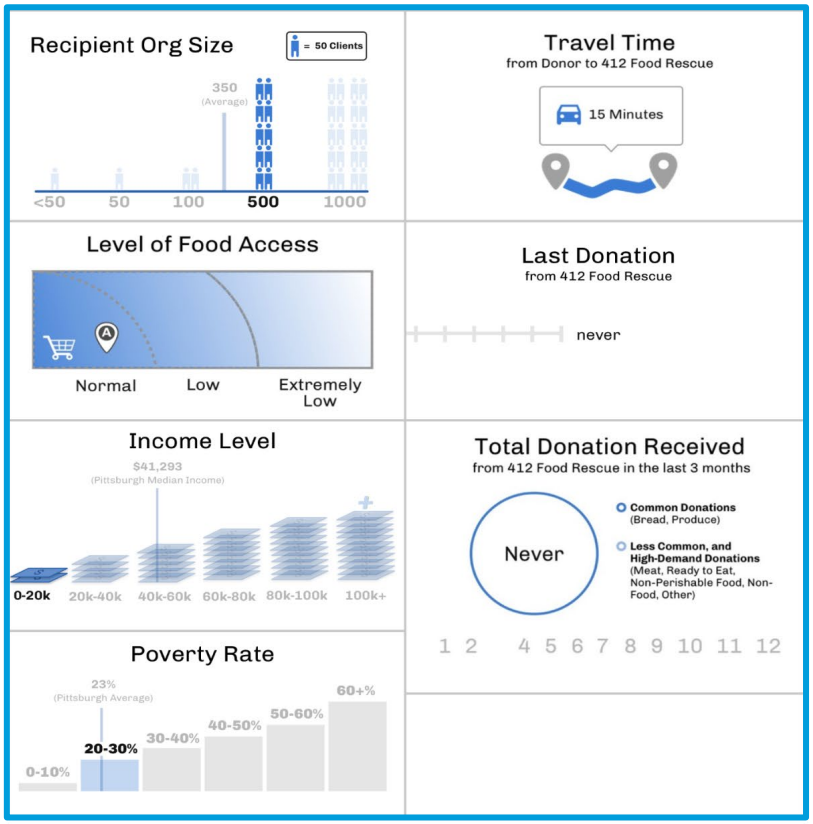
Employees
3

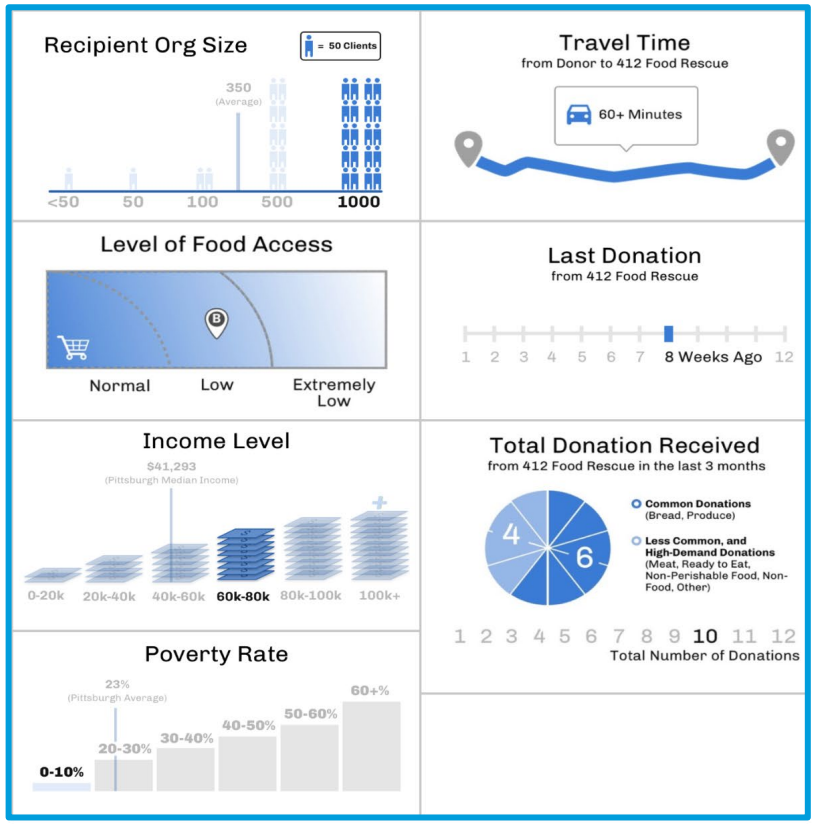
Donors
6

Recipients
8

Volunteers
6

DATA COLLECTION





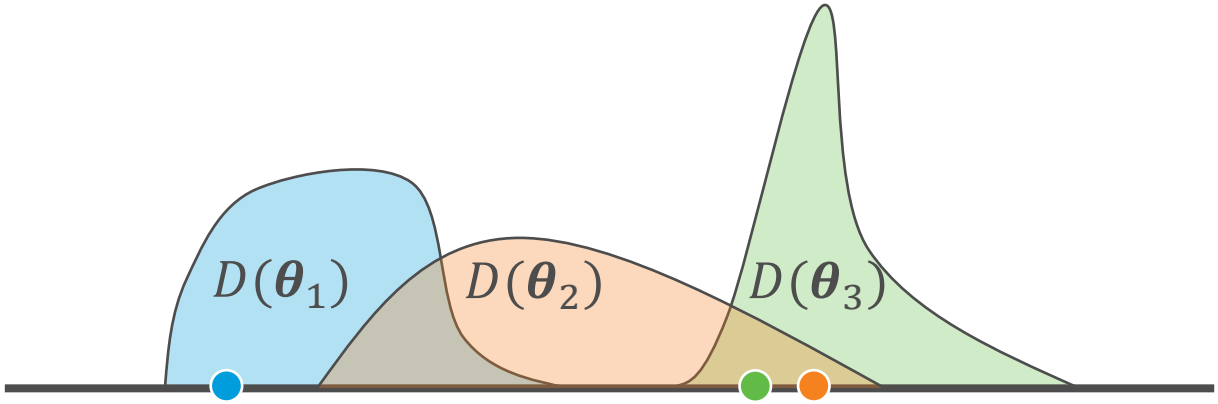
What should 412 Food Rescue do?

INTERLUDE: RANDOM UTILITY MODELS

- Parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)$
  - $m =$ number of alternatives
  - Each alternative x_j modeled by utility distribution $D(\boldsymbol{\theta}_j)$
- A voter's utility U_j for alternative x_j is drawn independently from $D(\boldsymbol{\theta}_j)$
- Voters rank alternatives by $U_1, \ldots, U_m$:

$$\Pr[x_2 \succ x_1 \succ x_3 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3] = \Pr_{U_j \sim D(\boldsymbol{\theta}_j)}[U_2 > U_1 > U_3]$$

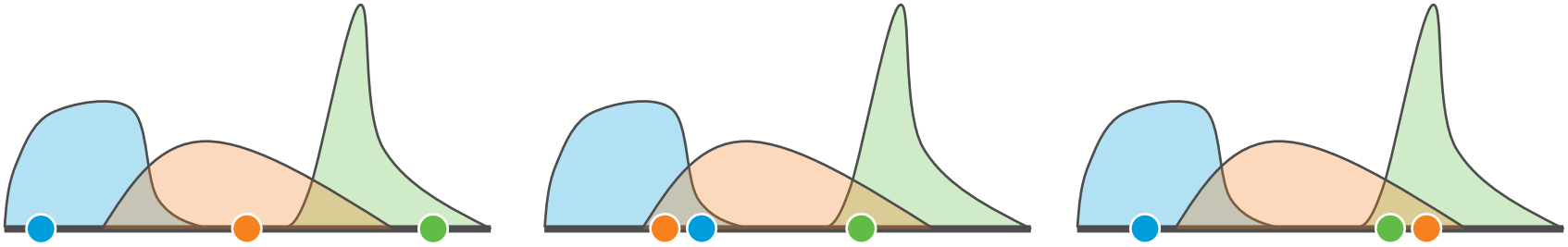
INTERLUDE: RANDOM UTILITY MODELS



Generating a single vote

$$x_2 \succ x_3 \succ x_1$$

INTERLUDE: RANDOM UTILITY MODELS



Voter 1
$x_3 \succ_{\sigma_1} x_2 \succ_{\sigma_1} x_1$

Voter 2
$x_3 \succ_{\sigma_2} x_1 \succ_{\sigma_2} x_2$

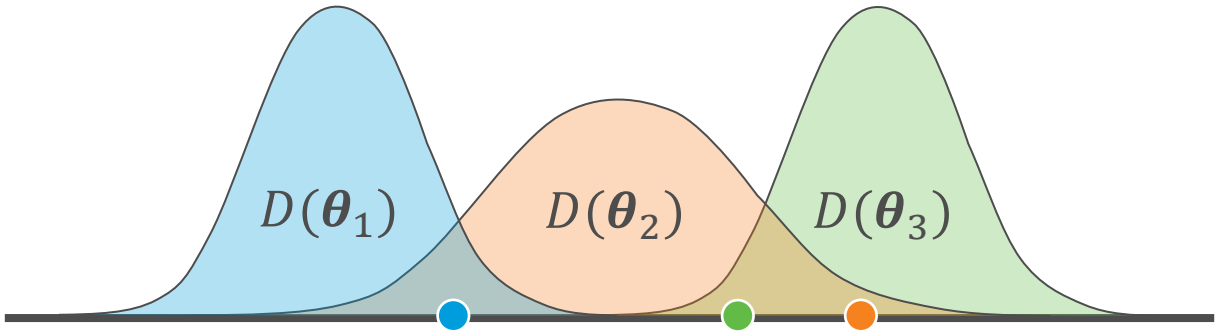
Voter 3
$x_2 \succ_{\sigma_3} x_3 \succ_{\sigma_3} x_1$

Generating a preference profile

$$\Pr[\boldsymbol{\sigma} \mid \boldsymbol{\theta}] = \prod_{i \in N} \Pr[\sigma_i \mid \boldsymbol{\theta}]$$

INTERLUDE: RANDOM UTILITY MODELS

The Thurstone-Mosteller Model is defined by a normal distribution: For each x_j, $\boldsymbol{\theta}_j = (\mu_j, \nu_j)$ and $D(\boldsymbol{\theta}_j) = \mathcal{N}(\mu_j, \nu_j^2)$



LEARNING VIA RUMS

- Assume that each alternative $\boldsymbol{x}_j$ is represented as a vector of features
- The preferences of a single voter i are represented as a parameter vector $\boldsymbol{\beta}_i$ such that $\mu_j = \boldsymbol{\beta}_i \cdot \boldsymbol{x}_j$
- Assume that $\nu_j^2 = 1/2$ for all j
- The problem is to learn, for each voter i, a maximum likelihood $\boldsymbol{\beta}_i$ given pairwise comparisons

AGGREGATION

True Profile

Noisy profile

Voting rule should be robust to noise:
Its output ranking from the true profile should coincide with the output ranking from the noisy profile

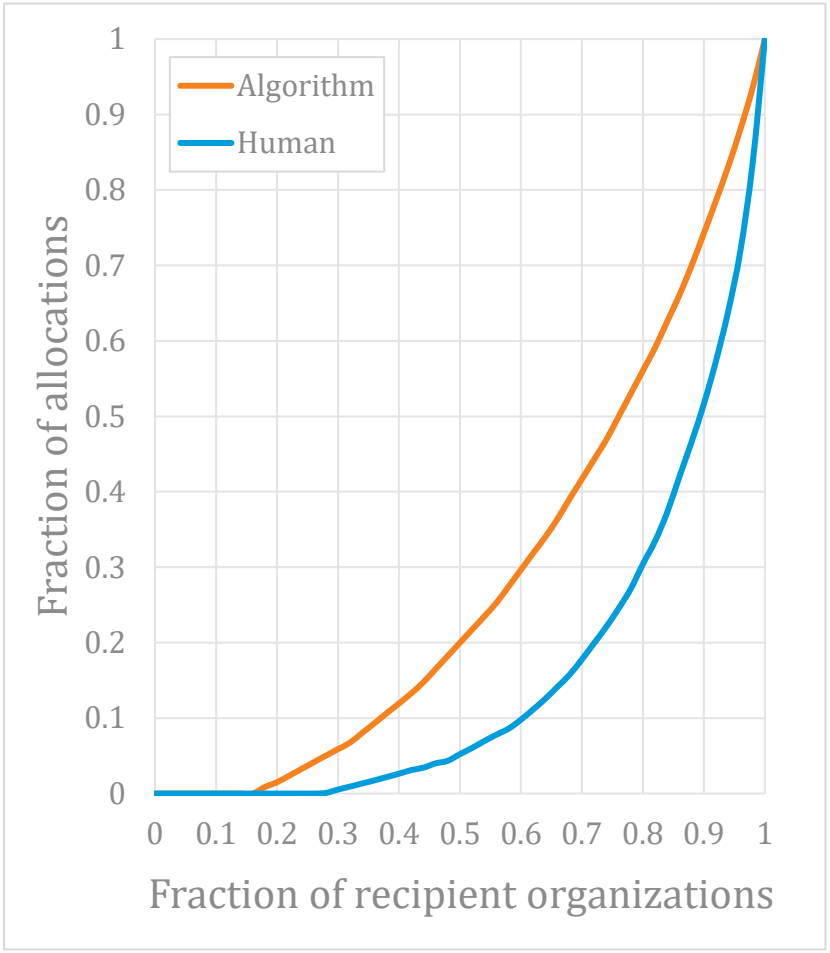
AGGREGATION

- Recall that the Mallows model is defined by parameter $\phi \in (0,1]$, and the probability of a voter having the ranking σ given true ranking π is

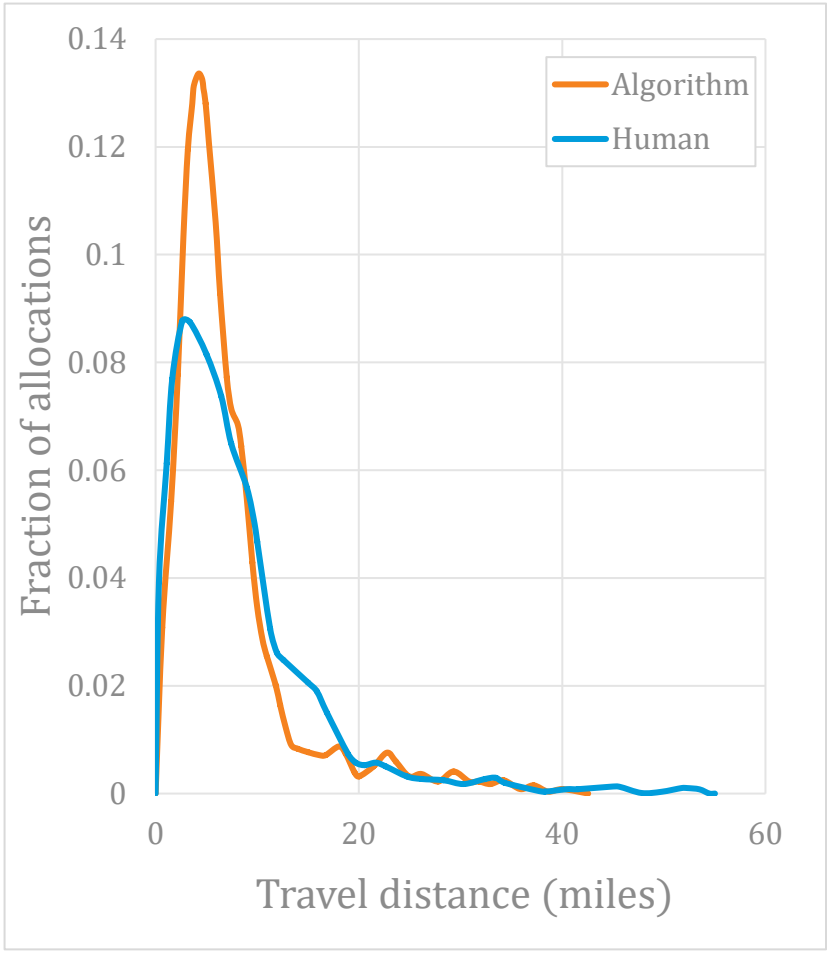
$$\Pr[\sigma|\pi] = \frac{\phi^{d_{KT}(\sigma,\pi)}}{\sum_{\tau} \phi^{d_{KT}(\tau,\pi)}}$$

- To model noisy prediction, each voter has a true ranking π_i and we predict a ranking σ_i drawn from Mallows
- Theorem (very informal): If the Borda scores of two alternatives under the true profile are "sufficiently" well separated then it's "unlikely" their Borda positions would be swapped under the noisy profile

PERFORMANCE ON HISTORICAL DATA



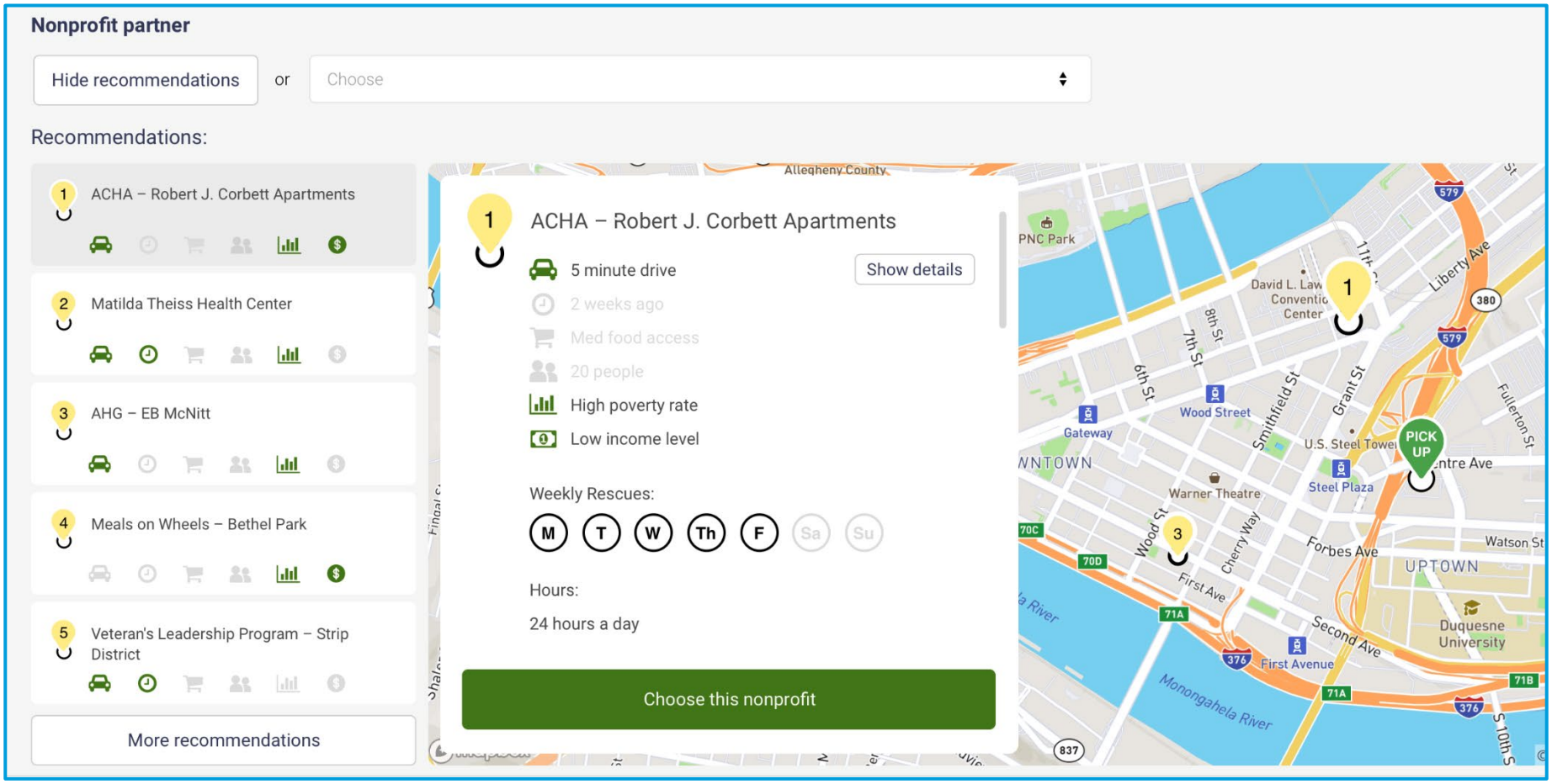
Diversity of allocations



Efficiency of allocations

INTERFACE

Designed as a decision support tool



PARTICIPANT FEEDBACK

Seeing how the algorithm's construction was broken down "into steps [...] and just taking each one at a time" made it attainable.

"This seems quite [a bit] better. If organizations are literally getting forgot[ten] about [...] this is huge."

"No matter what group or individuals we're feeding, [we] have the same regard for the food and the individuals we're serving."

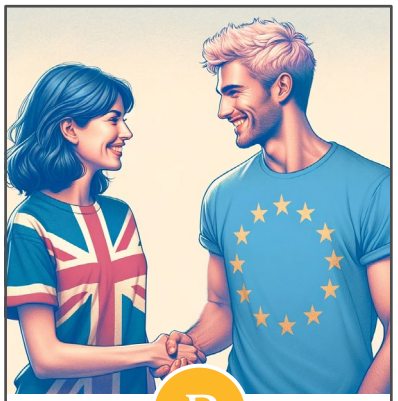
"Certainly more fair than somebody sitting at a desk trying to figure it out on their own. [...] it should be the most fair you could get."

BONUS: GENERATIVE SOCIAL CHOICE

It's 2016. Which policy would best address the UK's deepest problems?

A

Leave the European Union



B

Stay in the European Union

C

Ditch British cuisine for French cuisine

D

Exile the royal family to California

BONUS: GENERATIVE SOCIAL CHOICE

Unforeseen Alternatives

Unknown preferences

BONUS: GENERATIVE SOCIAL CHOICE

BONUS: GENERATIVE SOCIAL CHOICE

Discriminative Query

- A participant, represented by their survey response
- A textual statement

Output

Given participant's level of satisfaction for the given statement

Generative Query

- Subset of participants, represented by their survey responses
- An integer r

Output

Statement that maximizes r-highest level of satisfaction among members of given subset

BONUS: GENERATIVE SOCIAL CHOICE

Generative AI offers new building blocks for democratic systems

This technology can be misused to subvert democratic processes

But if we employ it responsibly, it can revitalize democracy