HUMANS IN ONLINE JOB MARKETS

Sihem Amer-Yahia

CNRS Research Director

Laboratoire d'Informatique de Grenoble

Simon Fraser University (Webinar)

Nov. 18th, 2020

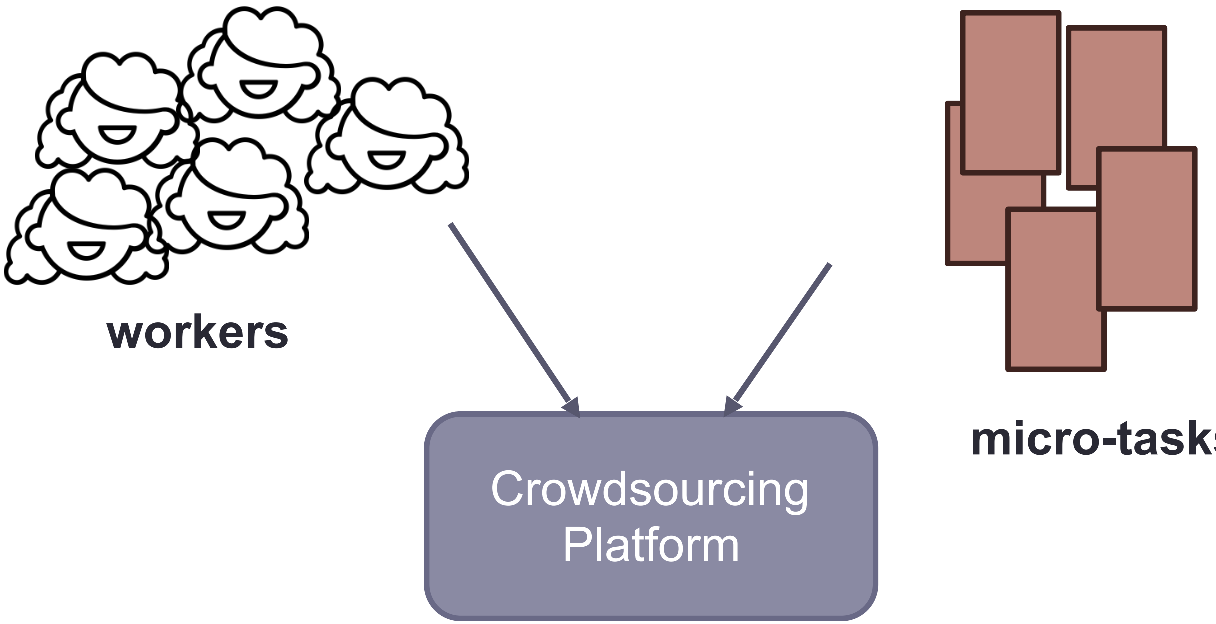
Online Job Markets

- Web destinations where one finds work
- Open calls to hire *cheap, immediate, skilled, and easily accessible* labor online

1. **Crowdsourcing platforms**
   - **micro-tasks**: image labeling, sentiment recognition
   - **collaborative tasks**: citizens for biodiversity

2. **Freelancing marketplaces**
   - **micro-gigs:** virtual/physical
   - resume preparation, website design, plumbing, assembling furniture

Data production in Crowdsourcing

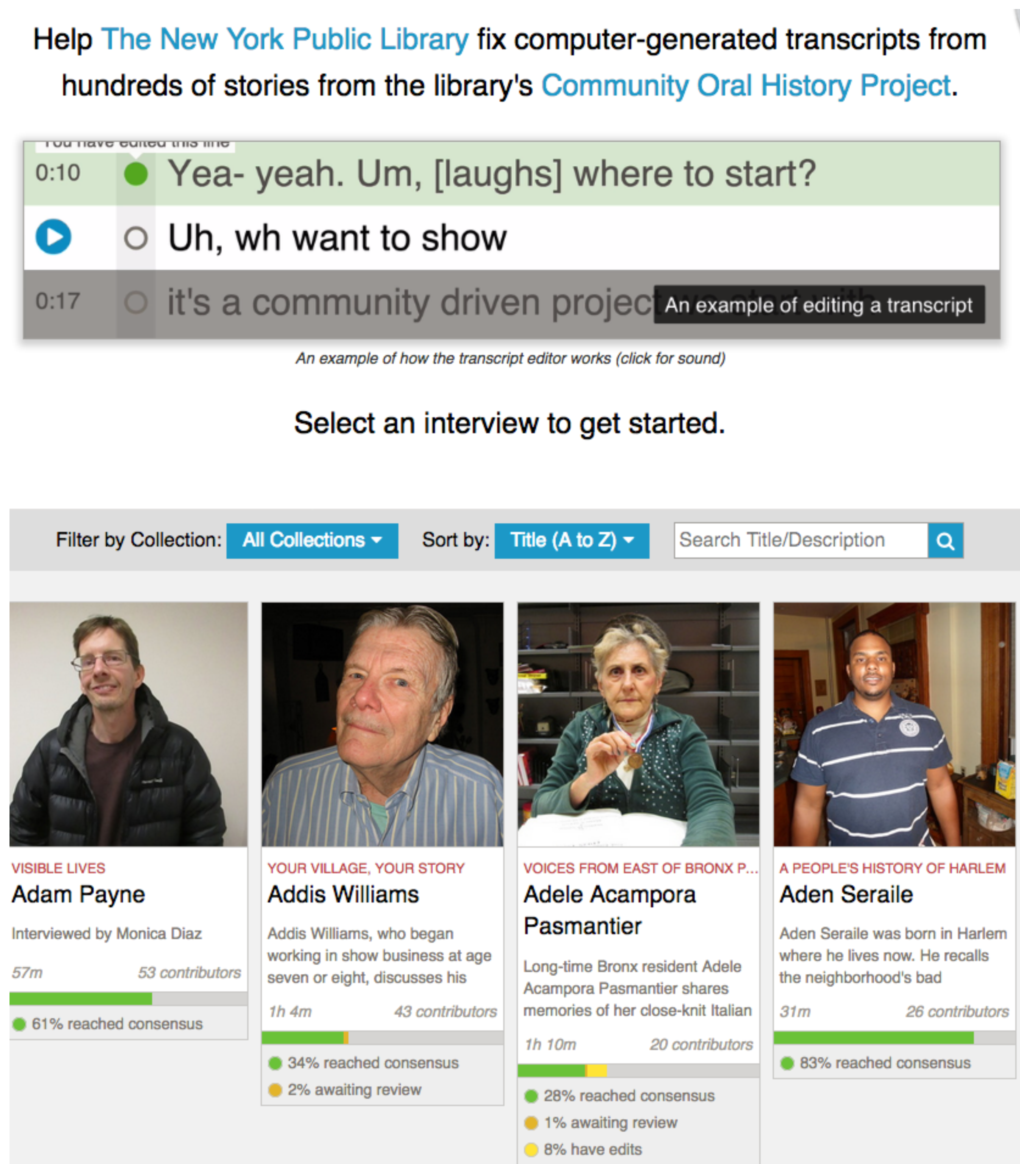


pybossa foulefactory Prolific

amazon mechanical turk Crowd4U figure eight

Content creation NYPL Lab

Together We Listen



Receipt Transcription on AMT

KEYBOARD SHORTCUTS: Scroll: Shift + up/down Open Image

Classify Receipt

Hit Reward: $0.02

Real readable original receipt | Not a receipt or not readable

The following details can often be found at the top or bottom of the receipt. Enter as much information as you can find.

Find and enter the business phone number:

Phone

Example: (888) 555-1234 or 8885551234

Find and enter the business address:

Address

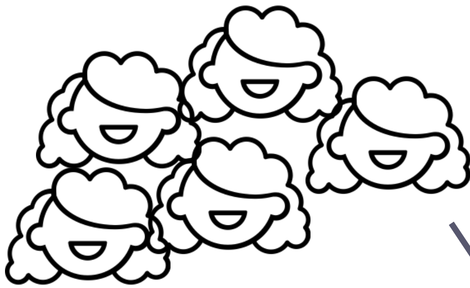
City

State | Postal code

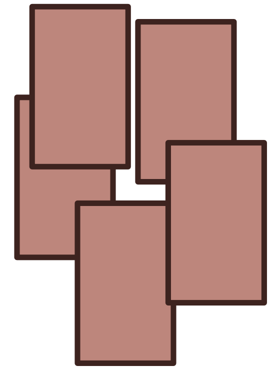
Example: 321 Fake Street, Los Angeles, CA, 90210

Next

Freelancing Marketplaces



workers

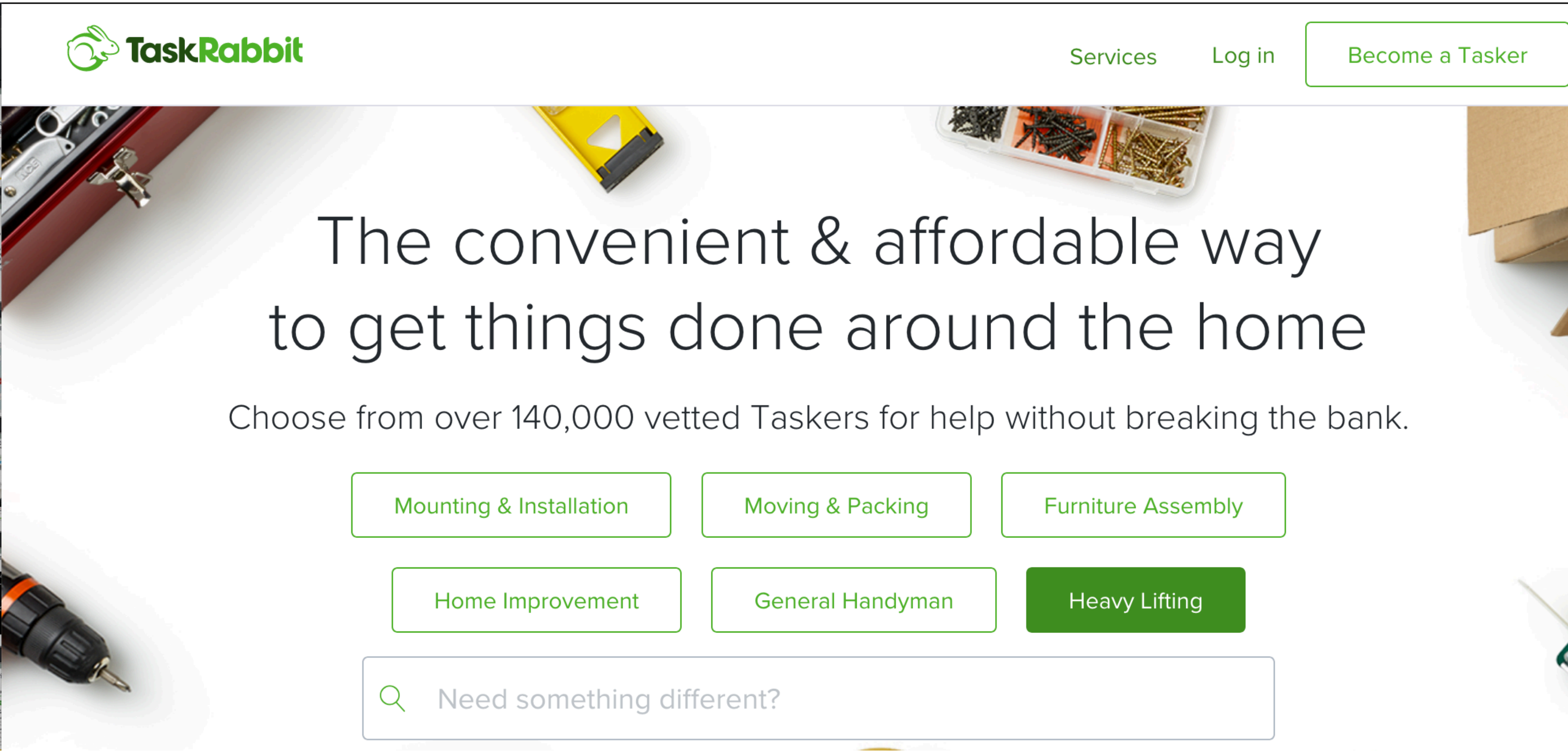


Platform

micro-gigs by requesters

- résumé preparation
- website design
- plumbing
- furniture assembly
- cleaning services

Online and offline help



FoW vision

- The current view of online labor markets tends to see humans as low-level agents, in the service of broader AI goals.
- We envision FoW as a place where humans are empowered with the ability to
  - get help from a mix of humans and AI machines
  - enhance their capabilities through skill acquisition
  - feel safe

S. Amer-Yahia, S. B. Roy, L. Chen, A. Morishima, J. Abello, P. Bourhis, F. Charoy, M. Danilevsky, G. Das, G. Demartini, A. Dubey, S. Elbassuon, D. Gross-Amblard, E. Hoareau, M. Inoguchi, J. Kenworthy, I. Kitahara, D. Lee, Y. Li, R. M. Borromeo, P. Papotti, R. Rao, S. Roy, P. Senellart, K. Tajima, S. Thirumuruganathan, M. Tommasi, K. Umemoto, A. Wiggins, K. Yoshida: **Making AI Machines Work for Humans in FoW.** SIGMOD Record, 49 (2), June 2020.

This talk's purpose and outline

Data Science cares about **data** and **insights**

- **Humans care about**
  - **how they are treated:** fairness
  - **how they are doing:** skill, feedback
  - **how they feel:** fatigue, boredom, motivation
  - **what they are learning:** capital advancement

Fairness

with **Shady Elbassuoni**, American University in Beirut

Disparity/Unfairness in Online Jobs

The unbalanced targeting of workers based on their **protected** attributes

The French Criminal Law lists 23 such attributes

gender, ethnicity, sexual orientation

http://www.allenovery.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx

A. Singh, T. Joachims: **Fairness of Exposure in Rankings**. KDD 2018

QAPA
Et voilà le travail

Inscription Connexion Accès candidat

DISTANCE

Rayon de recherche

30 km

NIVEAU D'EXPÉRIENCE

☐ Débutant (< 2 ans)	601
☐ Intermédiaire (2 à 5 ans)	559
☐ Confirmé (> 5 ans)	1 753

NIVEAU D'ÉTUDES

☐ Aucun diplôme	562
☐ CAP, BEP ou équivalent	33
☐ Niveau Bac	44
☐ Bac Obtenu	125
☐ Bac +2	652
☐ Bac +3	735
☐ Bac +4	298
☐ Bac +5	431
☐ > Bac +5	33

RECHERCHER DES CANDIDATS

web designer

Où ?

2 913 résultats correspondent à votre recherche

Thanh-van L.

Courdimanche
1 an et 9 mois d'expérience (tous métiers confondus)
Dernière formation : Communication Visuelle

Dernière connexion il y a 1 jour

Derniers postes
Assistant / Assistante chef de projet
Chef de projet
Web designer

Aurelien N.

Caluire-et-Cuire
19 ans et 6 mois d'expérience (tous métiers confondus)
Dernière formation : BTS attaché technico commercial

Dernière connexion il y a 1 jour

Derniers postes
Consultant / Consultante en recrutement
Négociateur / Négociatrice en immobilier
Délégué commercial / Déléguée commerciale en biens d'équipement auprès des entreprises

Thierry marcellin F.

Paris
6 ans et 10 mois d'expérience (tous métiers confondus)
Dernière formation : Développeur Intégrateur En

Derniers postes
Téléconseiller / Téléconseillère
Agent / Agente de planning informatique

A ranking formulation

Input: a query and a set of workers

Output: a ranking of workers

User	Gender	Country	Birth	Language	Ethnicity	Experience	Test	Approval	f(u)
U1	Female	America	2000	English	White	5	0.76	0.56	0.620
U2	Female	India	2004	English	Indian	0	0.50	0.20	0.290
U3	Male	America	1976	English	White	14	0.89	0.92	0.911
U4	Male	India	1976	Indian	White	6	0.65	0.65	0.650
U5	Male	Other	1963	Other	Indian	18	0.64	0.76	0.724
U6	Female	India	1963	Indian	Indian	21	0.85	0.90	0.885
U7	Male	America	1995	English	Black	2	0.42	0.20	0.266
U8	Female	America	1982	English	Black	16	0.95	0.98	0.971
U9	Male	Other	2008	English	Other	0	0.30	0.15	0.195
U10	Male	Other	1992	English	White	2	0.32	0.25	0.271

protected attributes

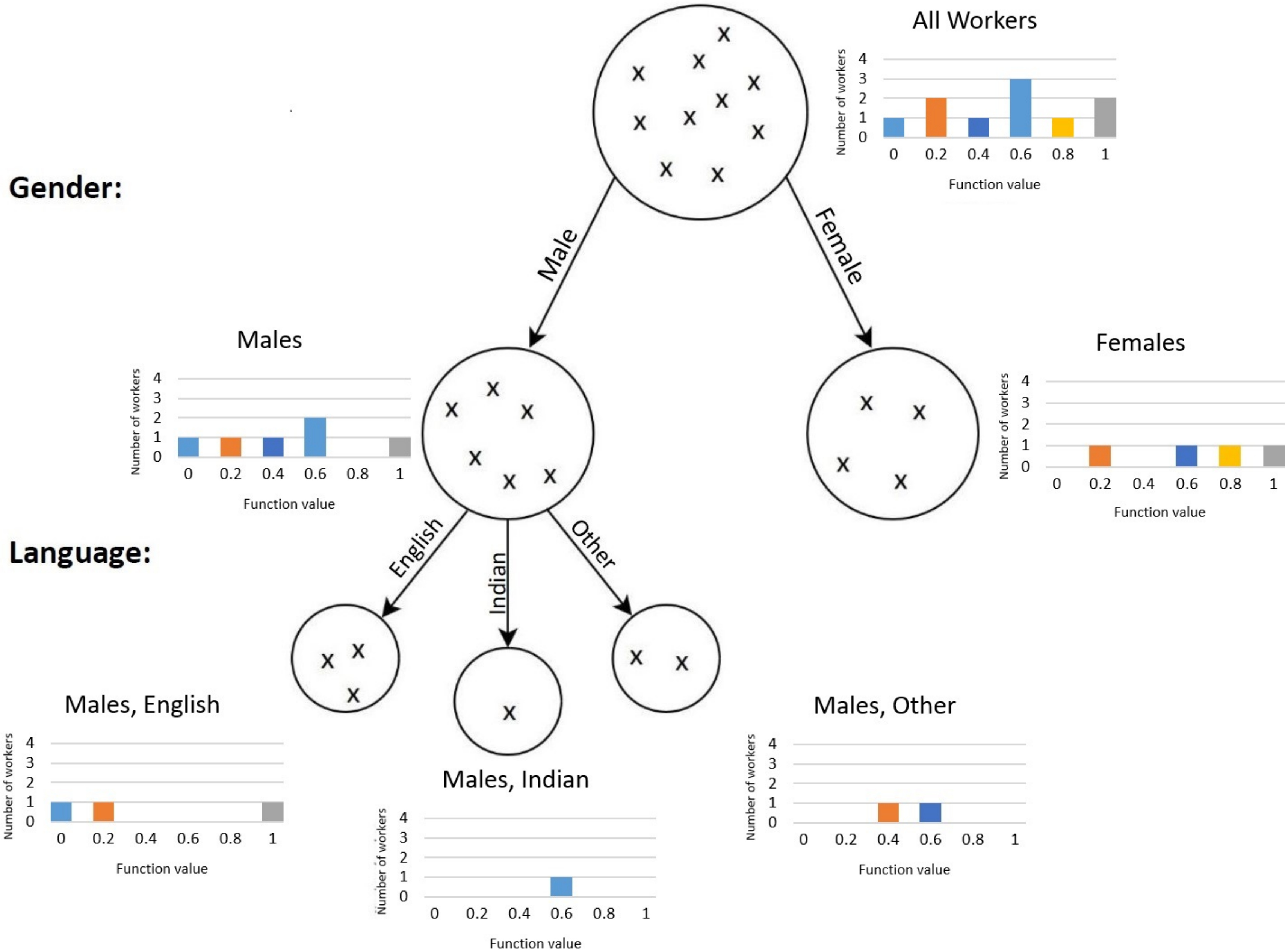
inferred attributes

$$f(u) = 0.3 \times LanguageTest(u) + 0.7 \times ApprovalRate(u)$$

All Workers:

Gender:

Language:



Most unfair partitioning problem

Given W and f, find partitioning $P = \{p_1, p_2, ..., p_k\}$ such that:

$$\begin{aligned} \underset{P}{\operatorname{argmax}} \quad & unfairness(P, f) \\ \text{subject to} \quad & \forall i, j \; p_i \bigcap p_j = \phi \\ & \bigcup_{i=1}^{k} p_i = W \end{aligned} \tag{1}$$

where

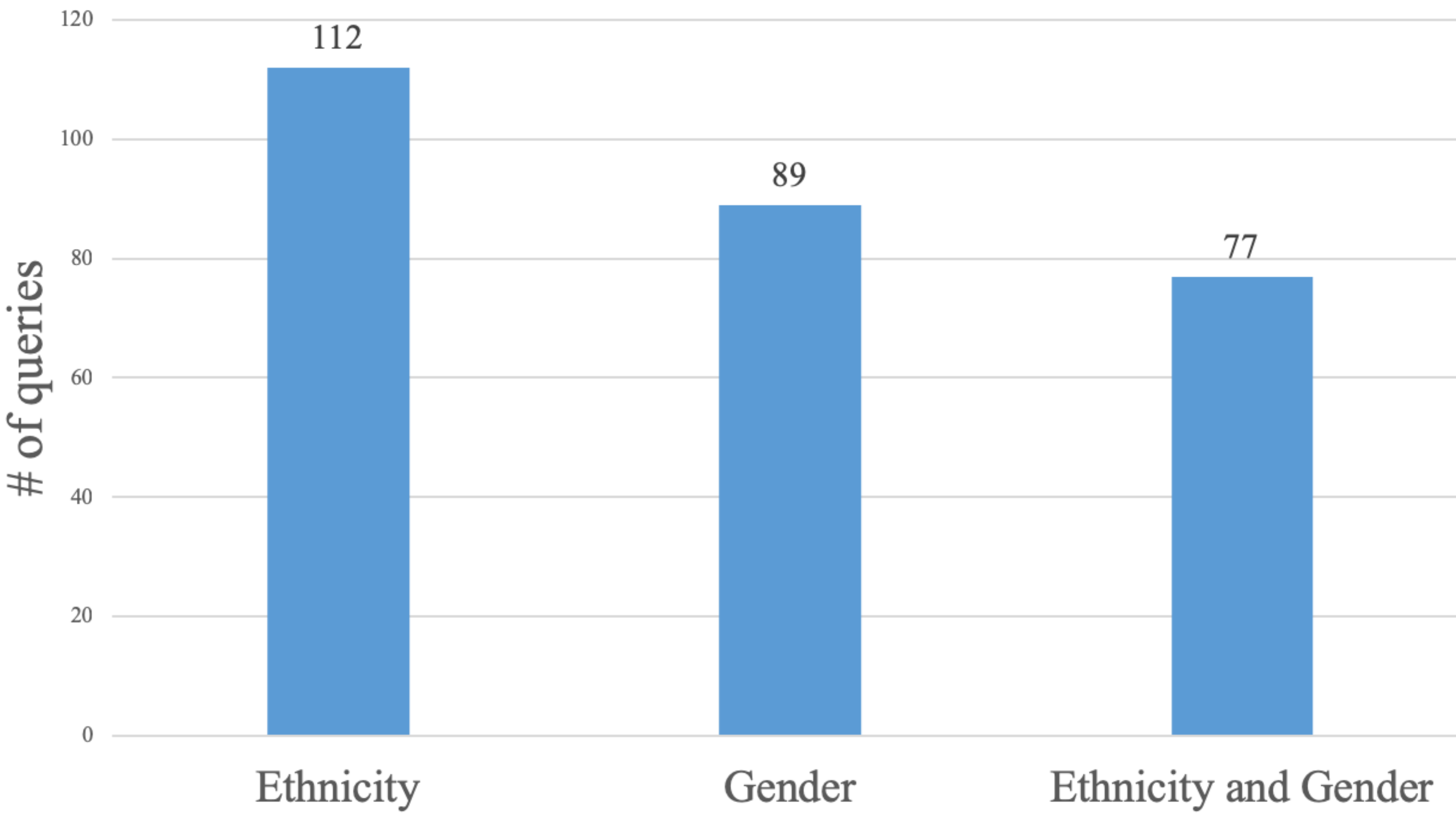
$$unfairness(P, f) = \underset{i,j}{\operatorname{avg}} \; EMD(h(p_i, f), h(p_j, f)) \tag{2}$$

and $h(p_i, f)$ is a histogram of the scores of individuals in p_i using f.

TaskRabbit

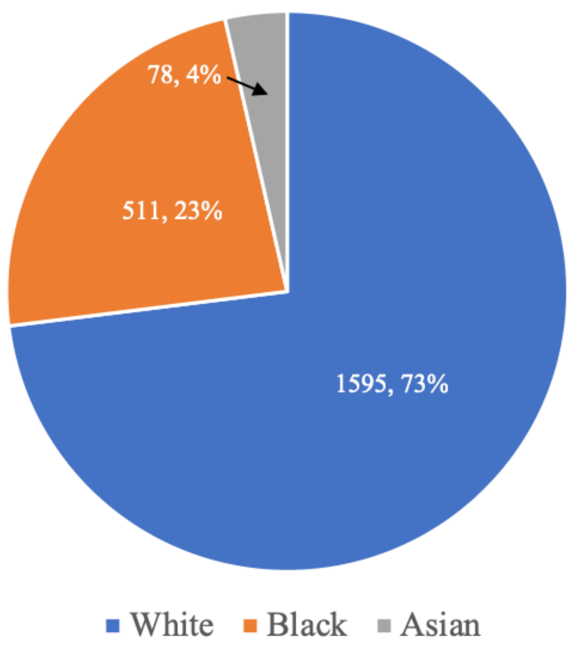
- 20 most popular services in 45 US cities
- 2,182 unique taskers, 287 queries: Home Cleaning
- Rank of each tasker
  - picture used to assign gender, ethnicity
  - not used: badge, reviews, and hourly rate

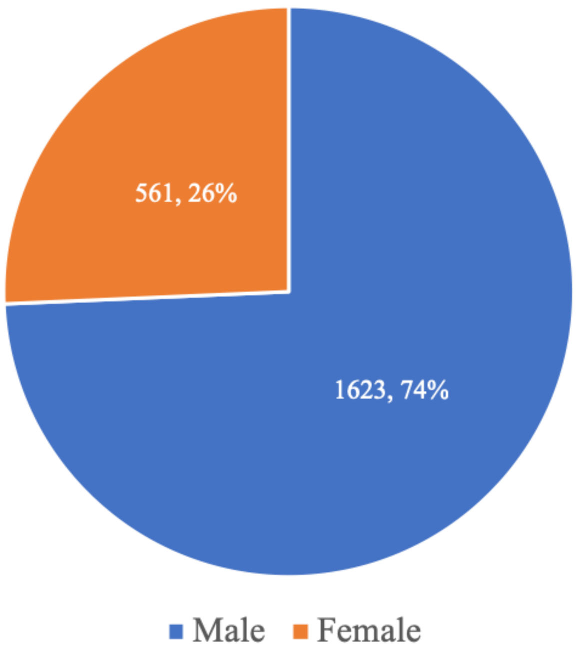
Number of queries for different partitionings

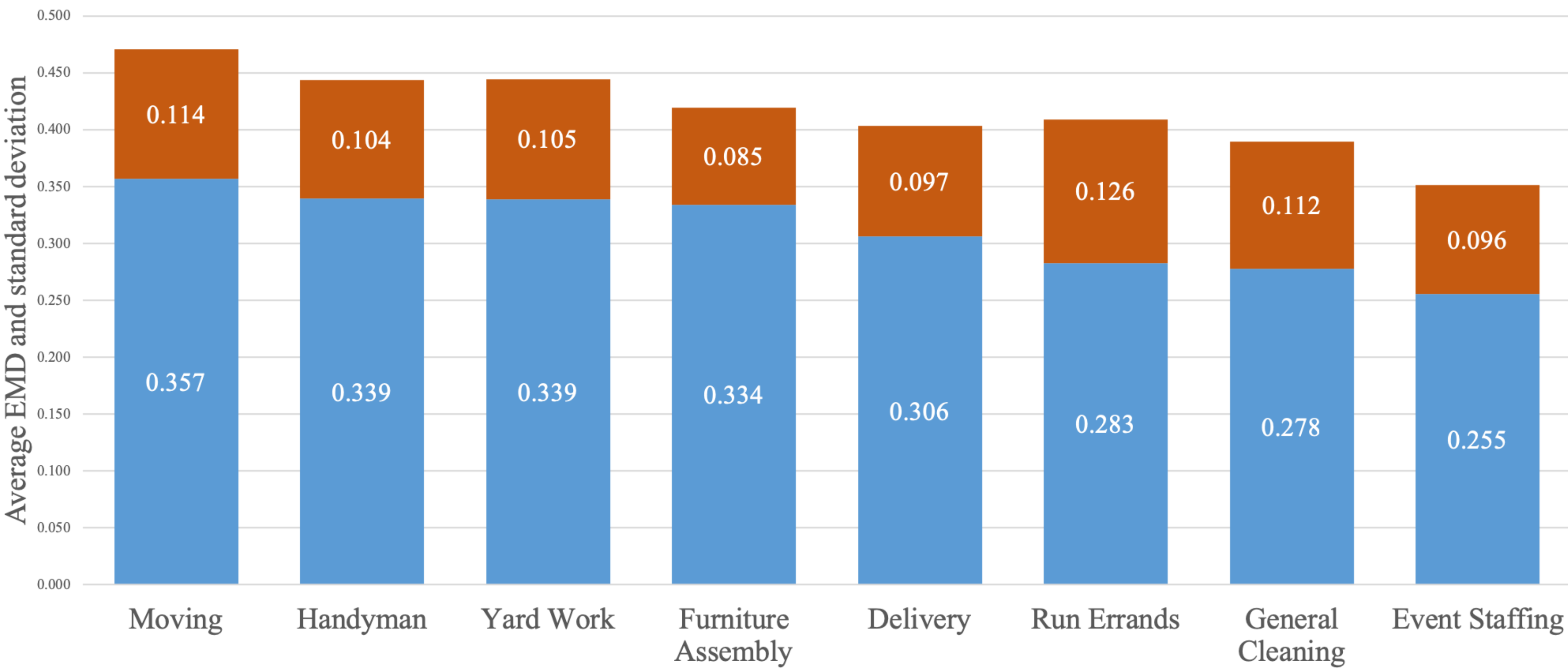


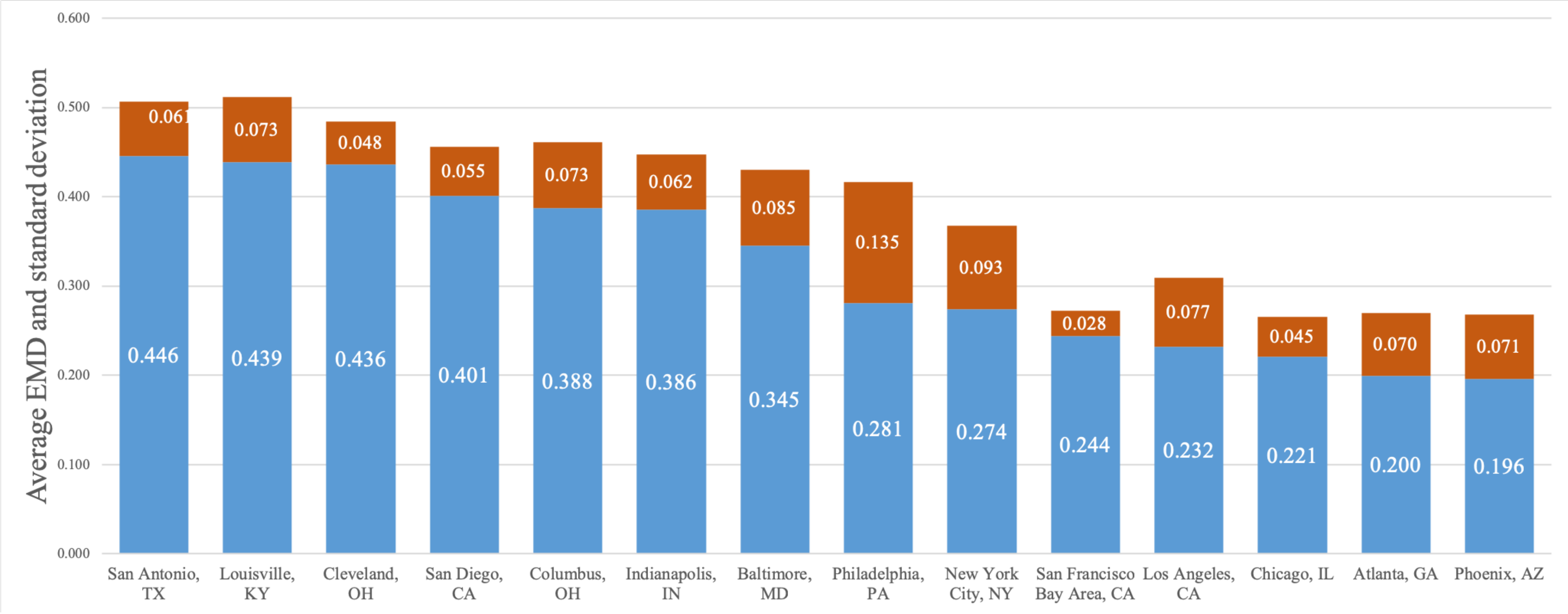
Summary of results

- All 112 queries returned white taskers in the top-50
- For the 89 queries (gender): 83% of top-50 are males






Average EMD and standard deviation
0.500
0.450
0.400
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
0.114
0.357
0.104
0.339
0.105
0.339
0.085
0.334
0.097
0.306
0.126
0.283
0.112
0.278
0.096
0.255
Moving
Handyman
Yard Work
Furniture Assembly
Delivery
Run Errands
General Cleaning
Event Staffing



Average EMD and standard deviation
0.600
0.500
0.400
0.300
0.200
0.100
0.000
0.061
0.446
0.073
0.439
0.048
0.436
0.055
0.401
0.073
0.388
0.062
0.386
0.085
0.345
0.135
0.281
0.093
0.274
0.028
0.244
0.077
0.232
0.045
0.221
0.070
0.200
0.071
0.196
San Antonio, TX
Louisville, KY
Cleveland, OH
San Diego, CA
Columbus, OH
Indianapolis, IN
Baltimore, MD
Philadelphia, PA
New York City, NY
San Francisco Bay Area, CA
Los Angeles, CA
Chicago, IL
Atlanta, GA
Phoenix, AZ

Generalization of fairness model

- On any given site, we consider:
  - a set of labeled groups **G**
  - a set of job-related queries **Q**
  - a set of locations **L**

- Each query **q** in **Q** contains a set of **keywords** such as ``**Home Cleaning**'' or ``**Logo Design**''

S. Elbassuoni, S. Amer-Yahia, A. Ghizzawi: **Fairness of Scoring in Online Job Marketplaces.** ACM Tansactions in Data Science, 2020

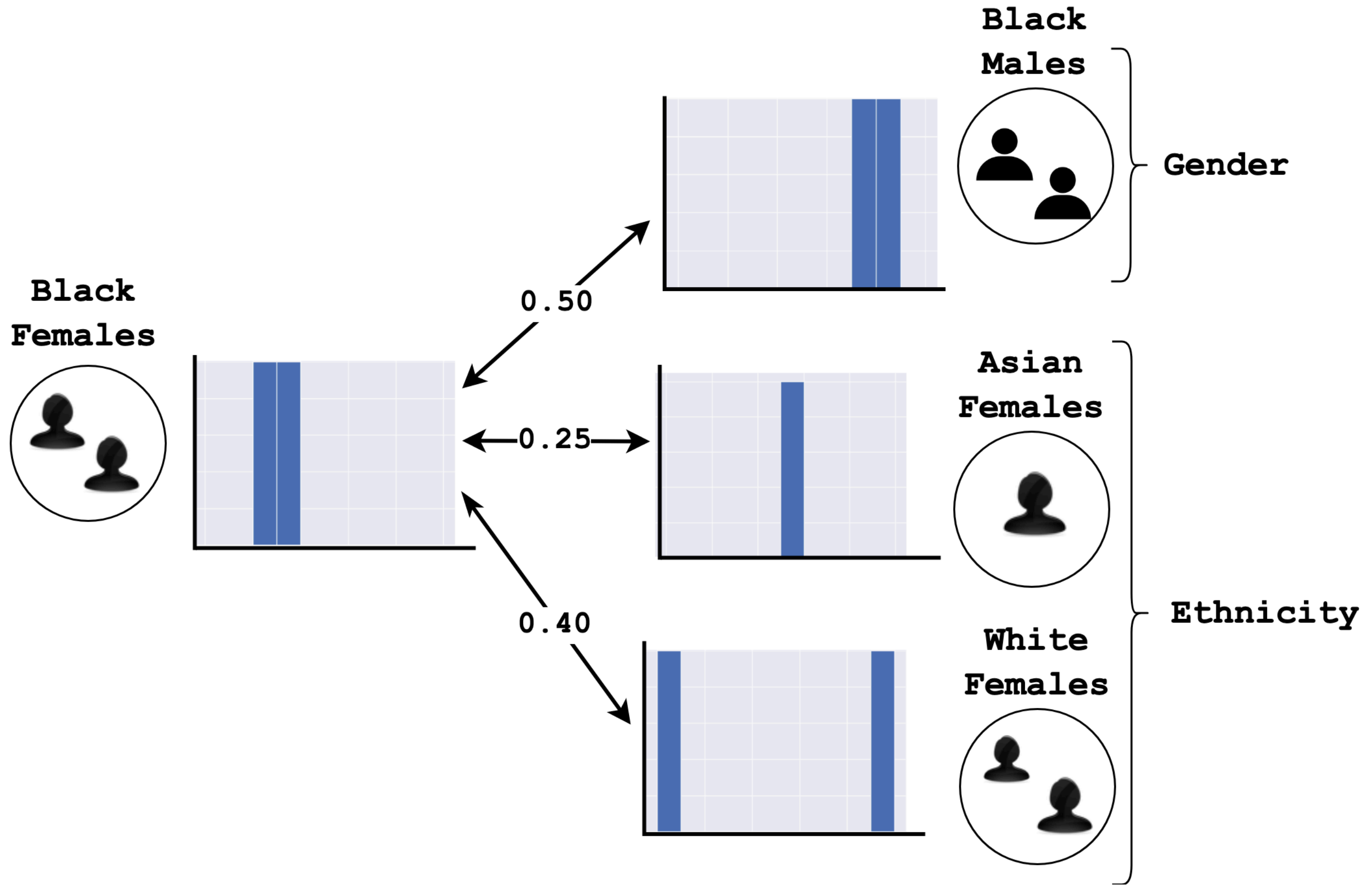
Discrimination

- The discrimination value of a triple **<g,q,l>** denotes the discrimination of **g** wrt query **q** at location **l**

$$d_{<g,q,l>} = \text{avg}_{g'}\, DIST(g, g')\ \forall g' \in \cup_{a \in A(g)} variants(g, a)$$

- ***variants(g,a)* are all groups that differ by one attribute only: e.g., for a group of *black females***, variants are ***black males, asian females, white females***.

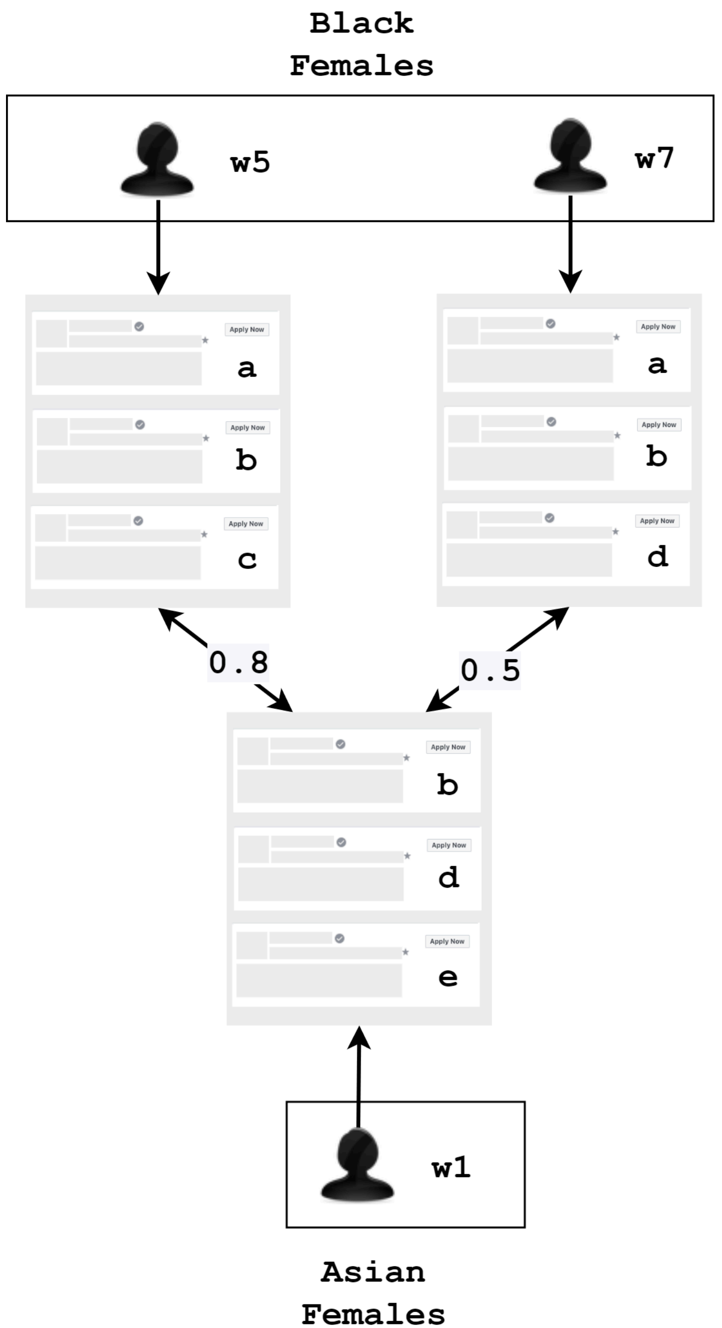
On TaskRabbit



Black Females
for Home
Cleaning in San Francisco
EMD (0.5+0.25+0.4)/3

On Google Jobs

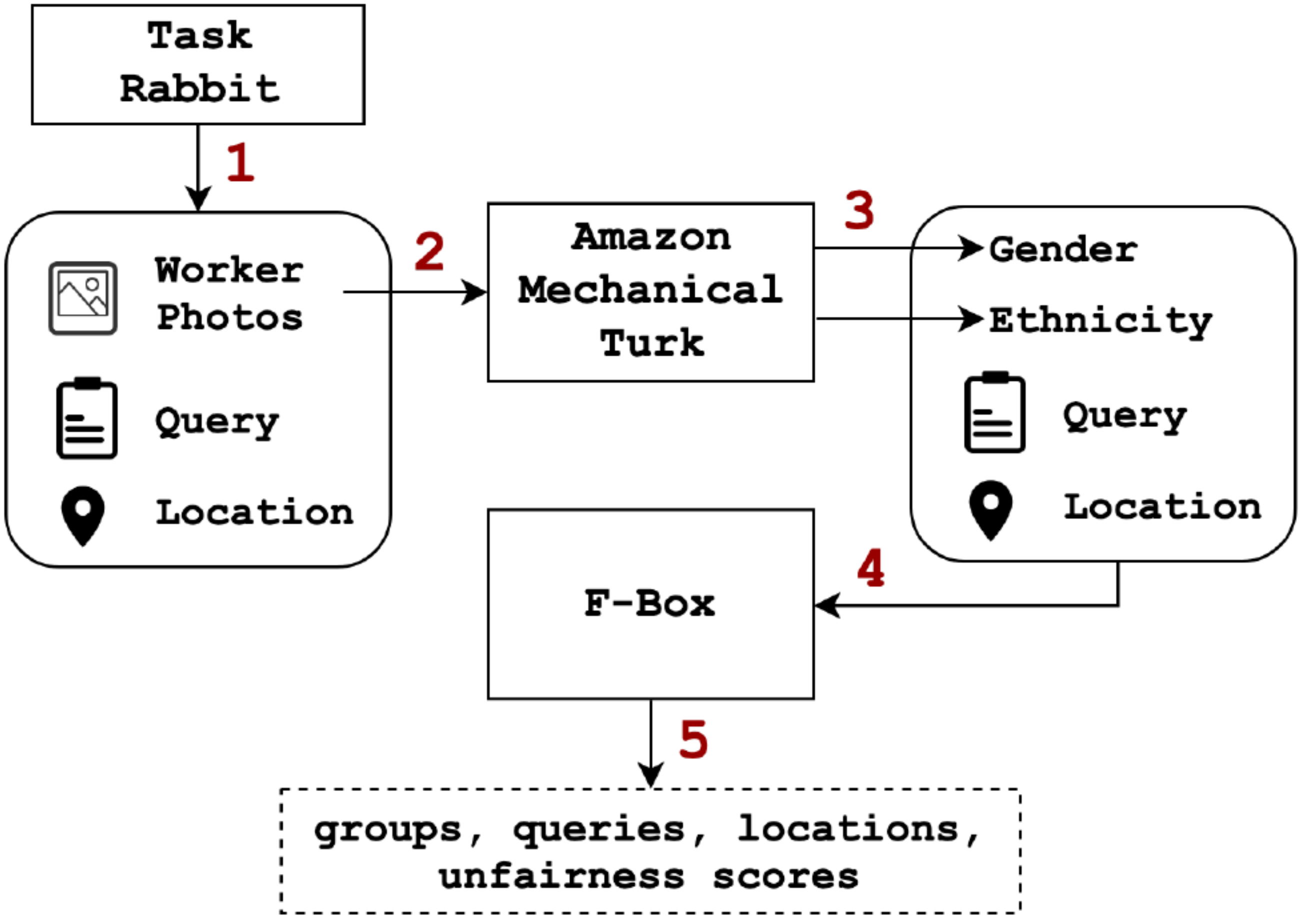
Discrimination of **Black Females** wrt one group, Asian Females:
Jaccard
(0.8+0.5)/2 = 0.65



Different Fairness aggregations

- **Worker-fairness**
  - Which 2 groups TaskRabbit is the most unfair for?
- **Query-fairness**
  - *What are the 5 least discriminating jobs for Asian males at all locations?*
- **Location-fairness**
  - *Out of NYC, Boston and Washington DC, what is the least discriminating location for women looking for an event staffing job on TaskRabbit?*

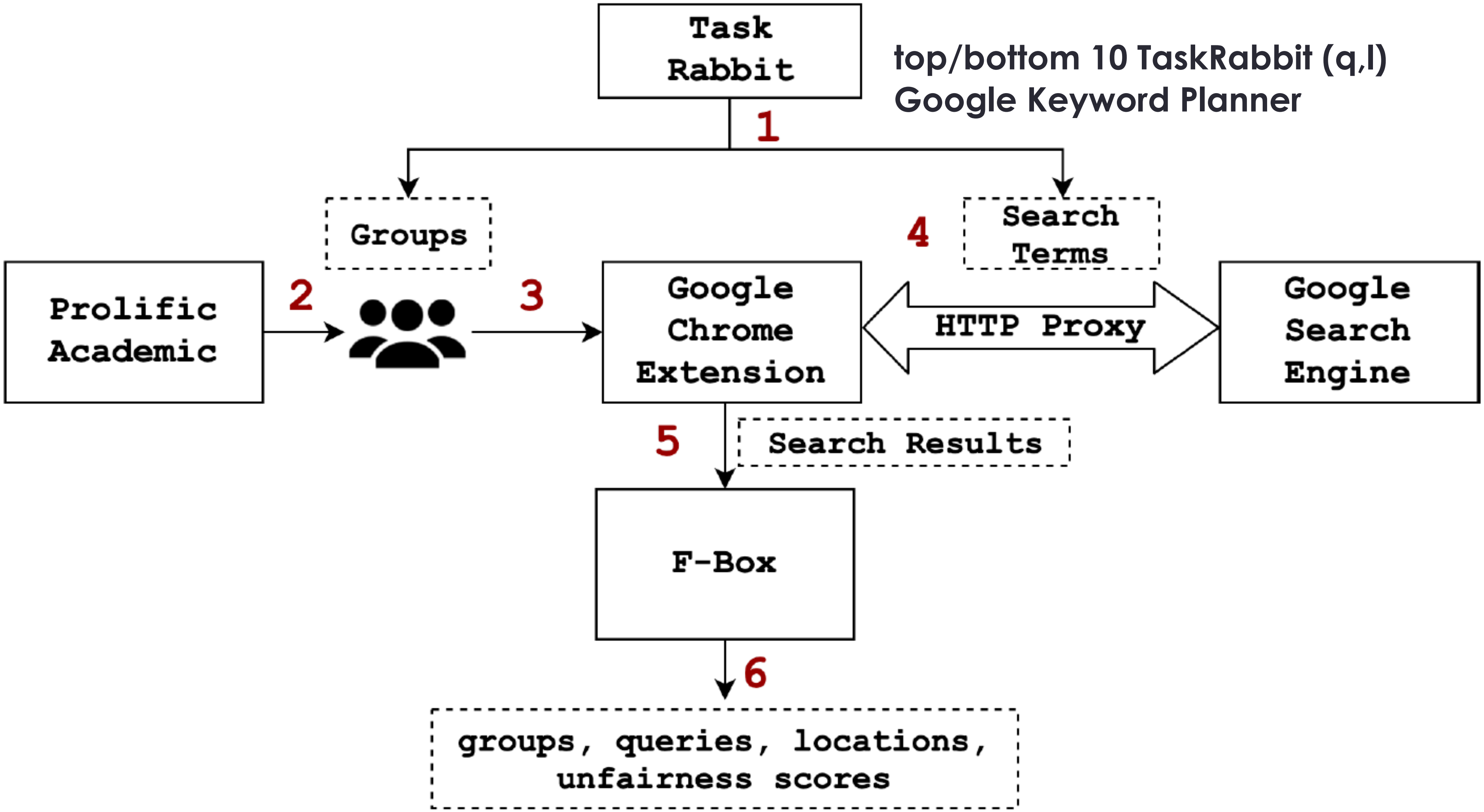
On TaskRabbit (June to August 2019)



All jobs offered in 56 cities: 3,311 workers; 5,361 query/location

On Google Job Search (June to August 2019)

top/bottom 10 TaskRabbit (q,l)
Google Keyword Planner



TaskRabbit quantification results

Group	EMD	Group	Exposure
Asian Female	0.876	Asian Female	0.821
Asian Male	0.755	Asian Male	0.662
Black Female	0.726	Black Female	0.615
Asian	0.694	Asian	0.594
Black Male	0.578	Black Male	0.413
White Female	0.542	White Female	0.359
Black	0.498	Black	0.341
Male	0.468	Female	0.299
Female	0.468	White Male	0.154
White	0.448	Male	0.117
White Male	0.421	White	0.104

Job	EMD	Job	Exposure
Handyman	0.692	Handyman	0.515
Event Staffing	0.639	Event Staffing	0.504
General Cleaning	0.611	General Cleaning	0.456
Yard Work	0.672	Yard Work	0.5
Moving	0.604	Moving	0.418
Delivery	0.499	Furniture Assembly	0.383
Furniture Assembly	0.541	Delivery	0.331
Run Errands	0.519	Run Errands	0.352

City	EMD	City	Exposure
Birmingham, UK	1	Birmingham, UK	0.926
Oklahoma City, OK	0.998	Oklahoma City, OK	0.819
Bristol, UK	0.91	Bristol, UK	0.761
Manchester, UK	0.851	Manchester, UK	0.739
New Haven, CT	0.838	New Haven, CT	0.67
Milwaukee, WI	0.824	Memphis, TN	0.668
Indianapolis, IN	0.815	Milwaukee, WI	0.668
Nashville, TN	0.808	Charlotte, NC	0.643
Detroit, MI	0.806	Nashville, TN	0.637

TaskRabbit comparison results

Group-comparison	Males	Females
All	**0.117**	**0.299**
Charlotte, NC	0.399	0.345
Chicago, IL	0.062	0.062
Nashville, TN	0.330	0.309
Norfolk, VA	0.331	0.168
San Francisco Bay Area, CA	0.084	0.084
St. Louis, MO	0.255	0.190

Job-comparison	Lawn Mowing	Event Decorating
All	**0.500**	**0.442**
Black	0.445	0.453

Location-comparison	San Francisco Bay Area, CA	Chicago, IL
All	**0.213**	**0.233**
Back To Organized	0.198	0.135
Organize & Declutter	0.224	0.191
Organize Closet	0.174	0.153

Summary of fairness

- Framework to quantify discrimination in ranking
  - can be used by workers, requesters, platform designers
  - accommodates many fairness measures
  - accommodates many optimization formulations

- Open questions
  - Explaining discrimination
  - Repairing discrimination

B. Salimi, B. Howe, D. Suciu: **Database Repair Meets Algorithmic Fairness**. SIGMOD Rec. 49(1), 2020

A. Asudeh, H. V. Jagadish, J. Stoyanovich, G. Das: **Designing Fair Ranking Schemes**. SIGMOD 2019

This talk's purpose and outline

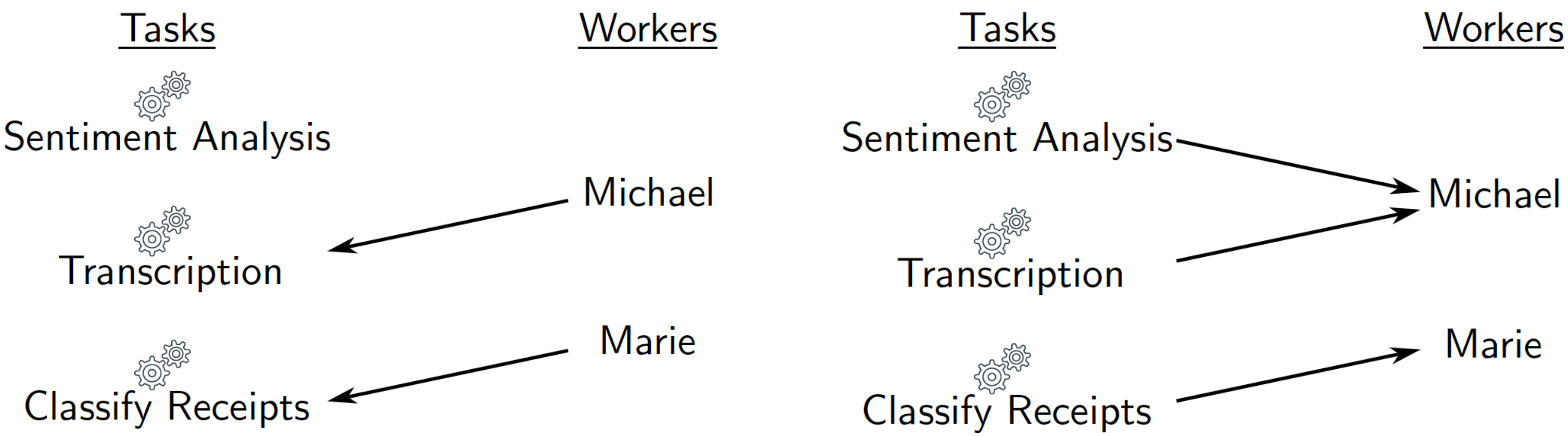
- **Humans care about**
  - **how they are treated:** fairness
  - **how they are doing:** skill, feedback
  - **how they feel:** fatigue, boredom, motivation
  - **what they are learning:** capital advancement

Skill and Motivation

with **Senjuti Basu Roy** The New Jersey Institute of Technology
and **Gautam Das** UT Arlington

Self-appointment in AMT

Algorithmic assignment



Self-Appointment

Algorithmic Assignment

Algorithmic Task Assignment

- **Input**: collaborative tasks, workers
- **Output**: one team per task

- Each task has
  - ***Budget, Required Expertise, Expected Quality***
  - *English comprehension* for audio transcription

- Each worker has human factors:
  - ***Skill, Expected wage, Acceptance ratio***

H. Rahman, S. B. Roy, S. Thirumuruganathan, S. Amer-Yahia, G. Das: **Task Assignment Optimization in Collaborative Crowdsourcing.** ICDM 2015

Goal: maximize crowd-work quality

$$\text{Maximize } \mathcal{V} = \Sigma_{\forall t \in T} v_t$$

worker skills workers' wages task quality

$$v_t = \begin{cases} \boxed{W_1 \times \Sigma_{\forall j \in \{1..m\}} q_{t_j}} + \boxed{W_2 \times (1 - \frac{w_t}{W_t})} & \text{if } q_{t_j} \geq Q_{t_j} \\ & \wedge\ w_t \leq W_t \\ 0 & \text{if } q_{t_j} < Q_{t_j} \\ & \vee\ w_t > W_t \end{cases}$$

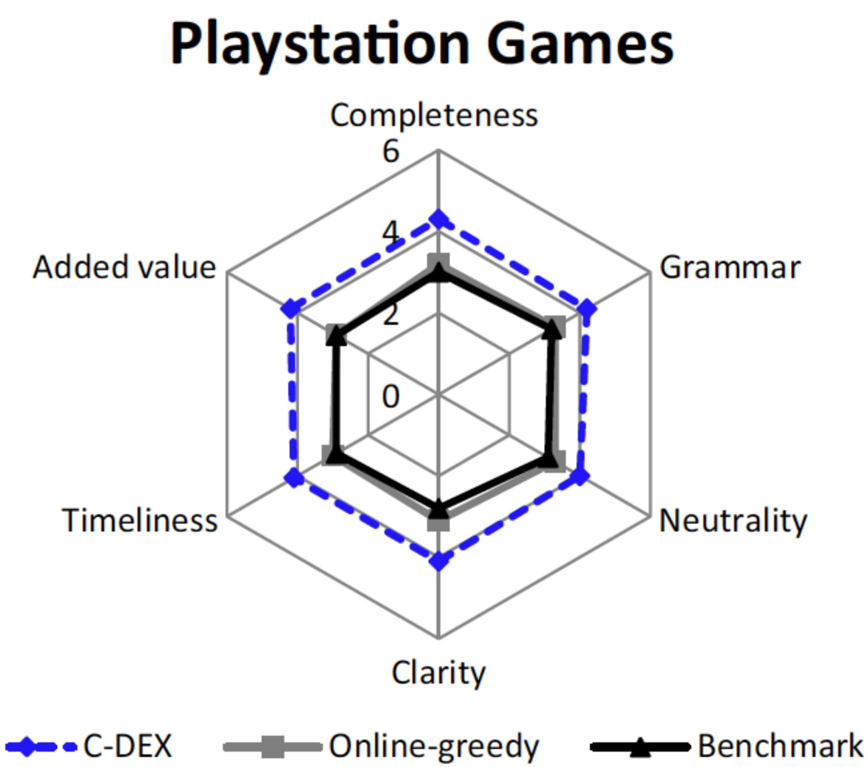
task budget

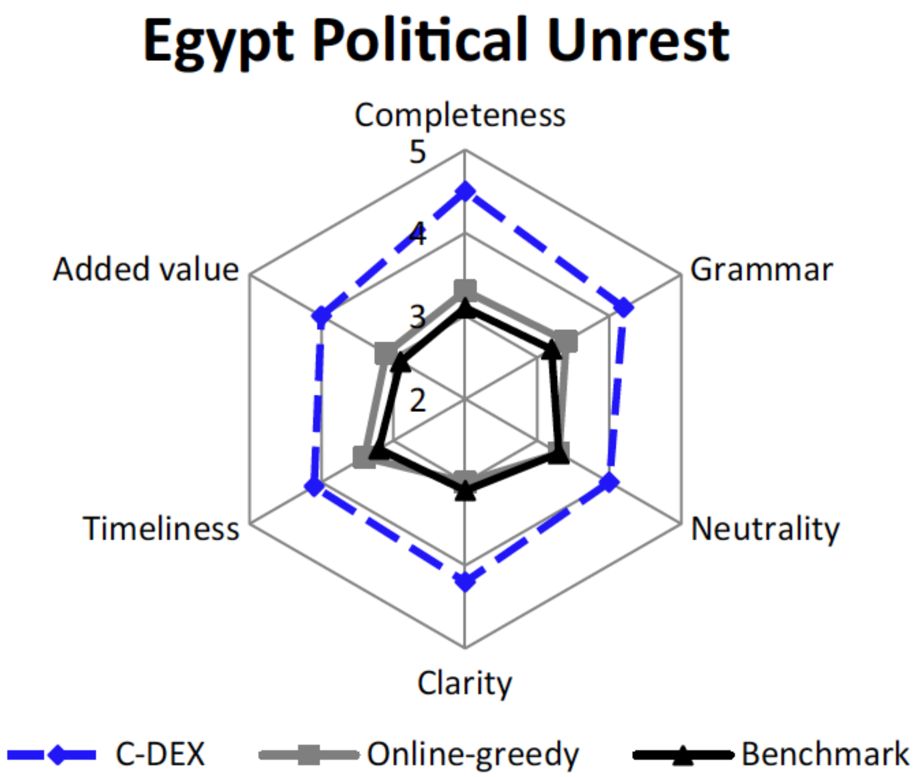
where $W_1, W_2 \geq 0$ and $W_1 + W_2 = 1$.

Quality Experiments

- **Phase 1**: 8 multi-choice questions/task, to assess skills
- **Phase 2**: Collaborative Document Editing task
  - 20 workers asked to produce reports on 5 different topics:
    1) *Political unrest in Egypt,*
    2) *NSA document leakage,*
    3) *Playstation games,*
    4) *All electric cars*
    5) *Global warming*
- **Phase 3**: Completed tasks evaluated by crowd workers
  - 150 AMT workers (selected similarly)
  - Completeness, Grammar, Neutrality, Clarity, Timeliness, Added-Value

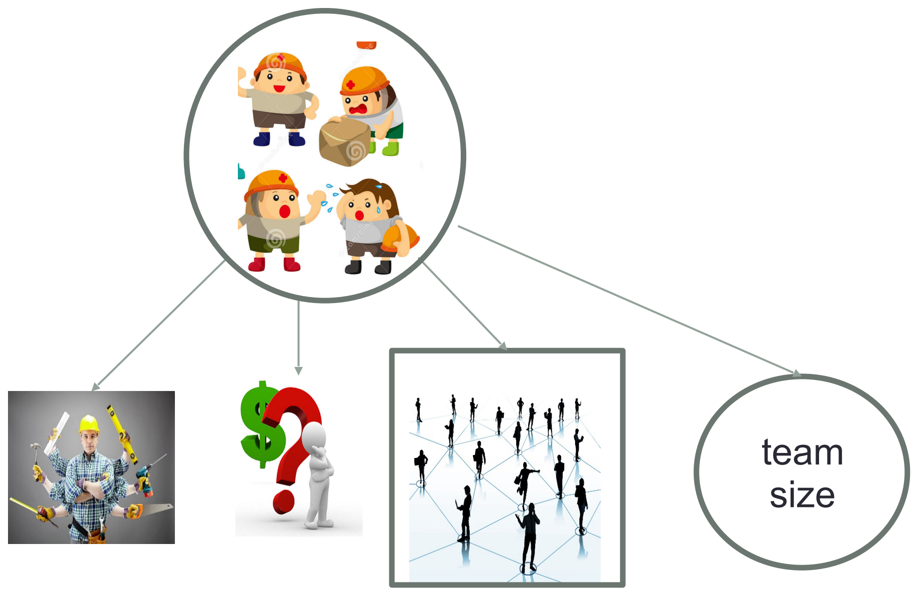
Outcome Quality





Group-level Human Factors

- In some cases, outcome quality was low
  - Conflicting opinions
  - Edit wars



Skill **Cost** **Affinity** **Critical Mass**

G. Hertel and G. Hertel: **Synergetic effects in working teams**. Journal of Managerial Psychology 2011

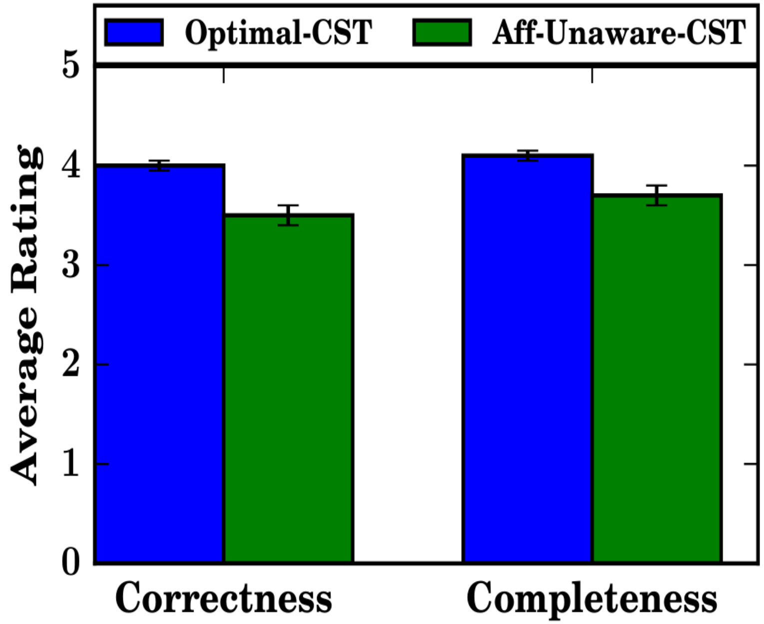
A two-stage solution

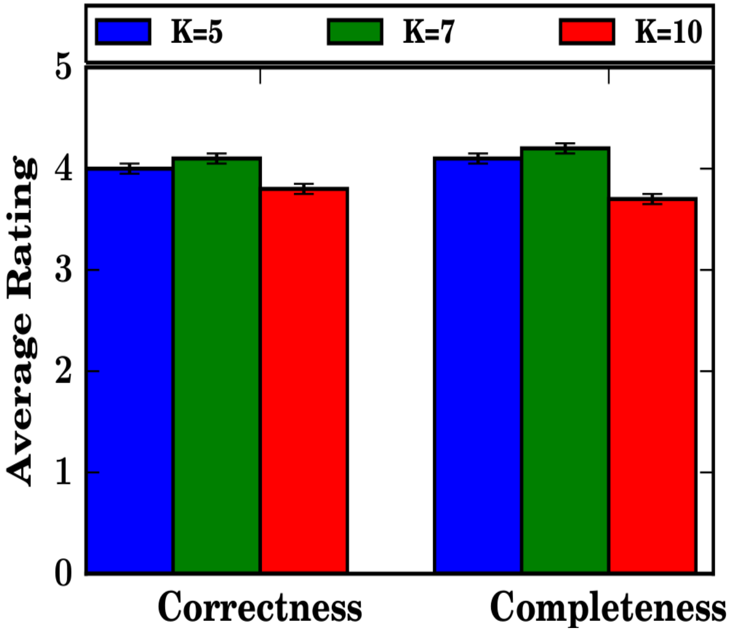
An instance optimal exact algorithm and a 2-approximation algorithm (when distance is a metric)

2. Form one team that maximizes **intra-affinity**, and satisfies **skill and cost** (*variant of Compact Location*)
3. Decompose into smaller teams, each satisfies **critical mass** and maximizes **inter-affinity** (*variant of Minimum Bisection*)

Experiments with Affinity and Critical Mass

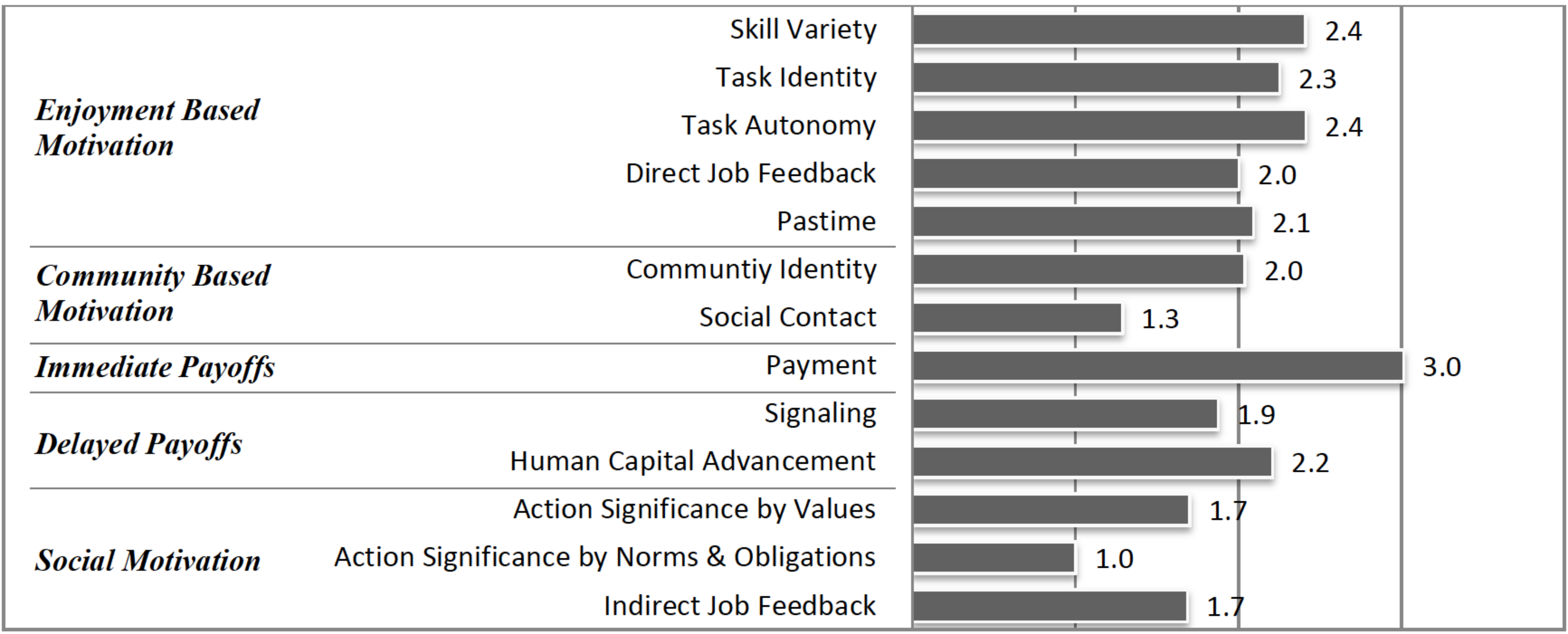
- Translation task with 120 AMT workers
- Region more effective than age/gender



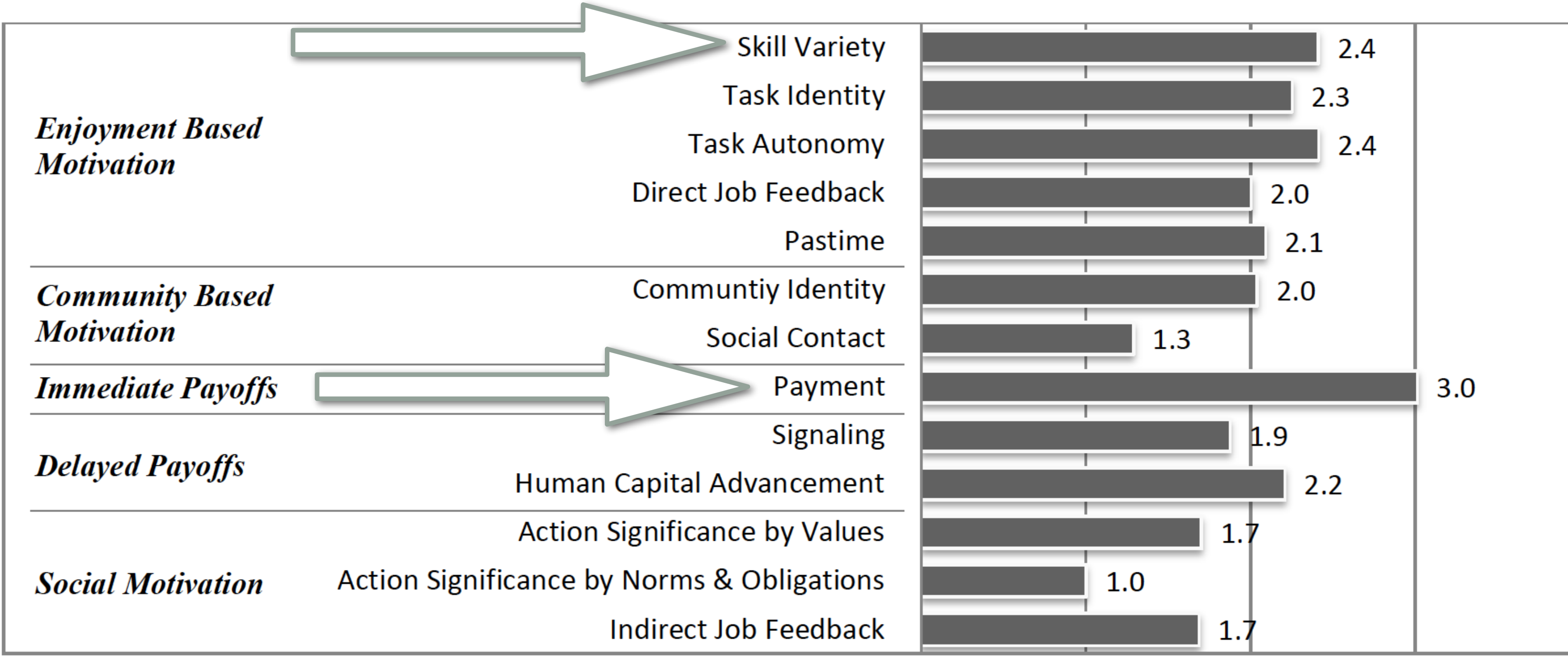


Motivation on AMT

More than fun and money. worker motivation in crowdsourcing-a study on mechanical turk.
N. Kaufmann, T. Schulze, and D. Veit. **AMCIS 2011**



What to observe?

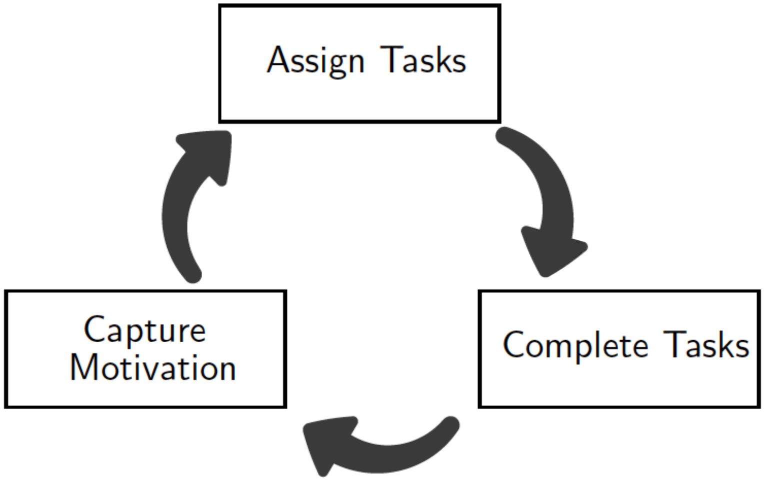


J. Pilourdault, S. Amer-Yahia, S. B. Roy, D. Lee: **Task Relevance and Diversity as Worker Motivation in Crowdsourcing.** ICDE 2018

Two Motivation Factors

intrinsic factor, ***task diversity***

extrinsic factor, ***task payment***

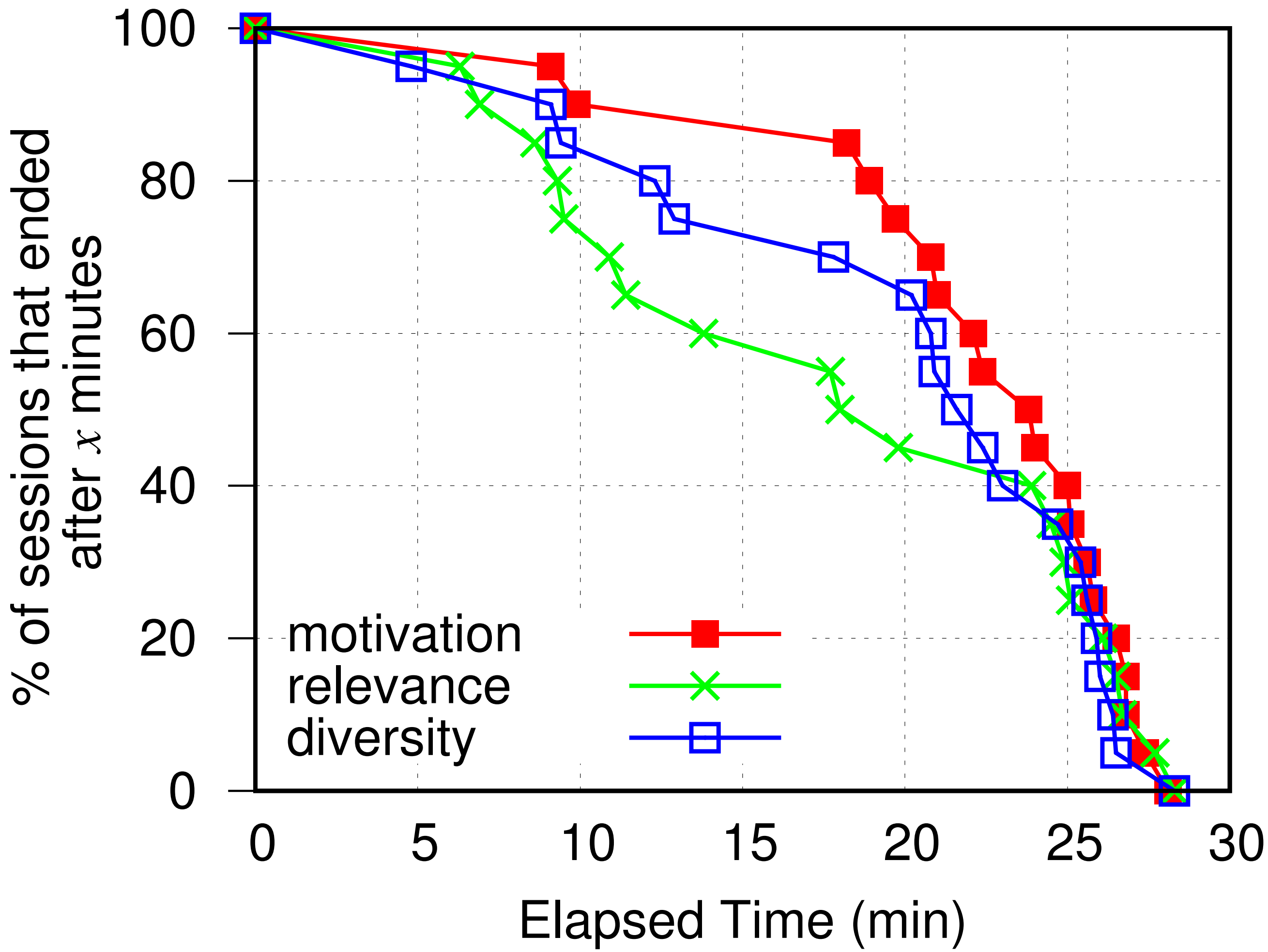


$$motiv_w^i(\mathcal{T}_w^i) = \alpha_w^i \times TD(\mathcal{T}_w^i) + (1-\alpha_w^i) \times TP(\mathcal{T}_w^i)$$

worker-specific

to be learned between iterations

Worker Retention



Summary so far

- Human factors dictate algorithm design
- Human factors must be observed
  - **Skill, affinity, critical mass yield higher quality contributions**
  - **Motivation yields better worker retention**

This talk's purpose and outline

- **Humans care about**
  - **how they are treated:** fairness
  - **how they are doing:** skill, feedback
  - **how they feel:** fatigue, boredom, motivation
  - **what they are learning:** capital advancement

Peer Learning

with **Payam Esfandiari** and **Senjuti Basu Roy (NJIT)**

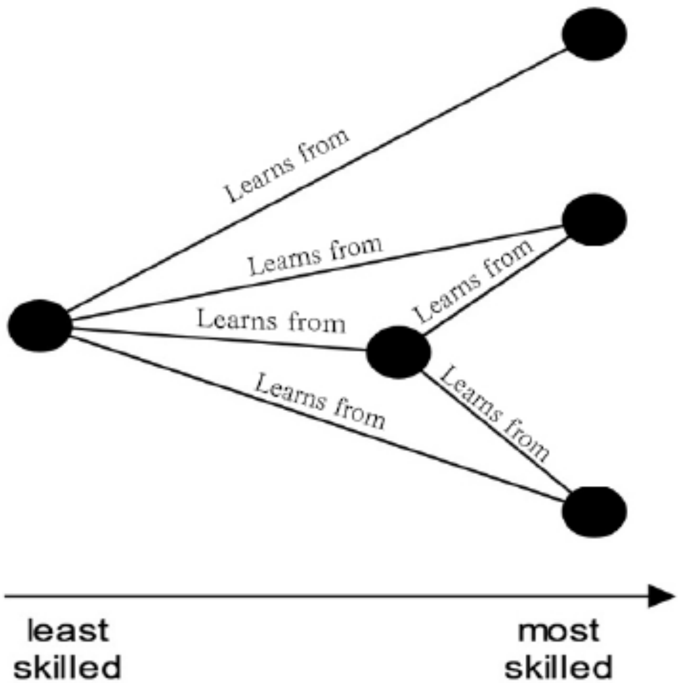
SIGKDD 2019

Peer Learning

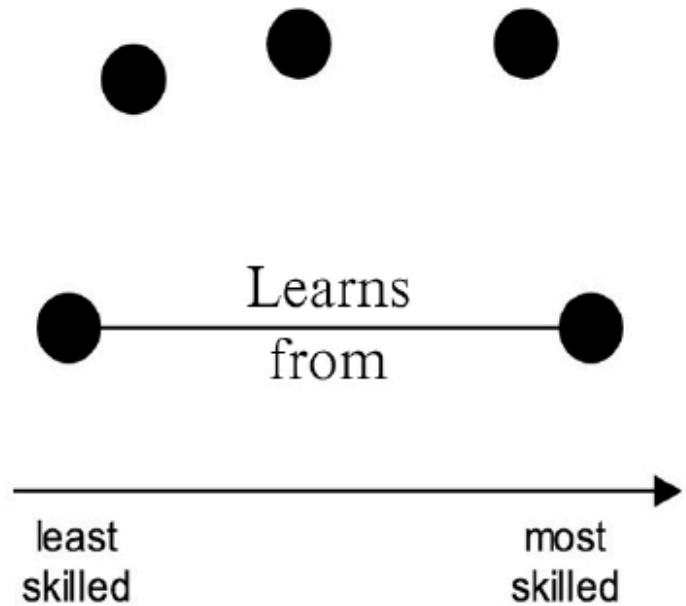
Explore how **affinity** affects **learning potential**

- Formalize Learning Potential (LP)
- Formalize Affinity structures (AFF)
- Algorithms with provable theoretical guarantees

Learning Potential

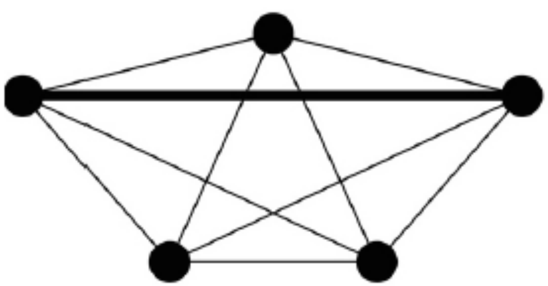


LPA: members learn from higher-skilled ones

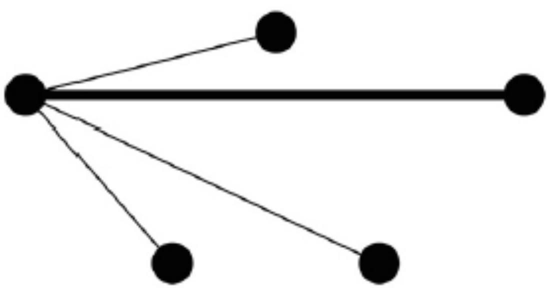


LPD: the least skilled member learns from the most skilled one

Affinity



AffD: function of affinities between all pair of members



AffC: function of affinities between one member and the rest

Example Task

The Queen does not need a passport to travel True or False ?	• Worker 1: True. All British Passports are issued in the Name of Her Majesty, The Queen. • Worker 2 : I found an article which agrees with your findings. **Fun fact: she also doesn't need a driver's license or a license plate on her car.**
Members of the royal family have to accept absolutely all gifts.	• Worker 1 : (Mostly false; Large true in practice.) While I couldn't find any law requiring the Royals to accept all gifts. • Worker 2 : I found an article which says they make a list of all gifts they receive throughout the year and release it publicly. In addition, they donate many of their gifts.

Team Formation with Affinity and Learning Potential

$$\underset{\mathcal{G}}{\text{optimize}} \quad \sum_{i=1}^{k} LP(g_i), \sum_{i=1}^{k} A\!f\!f(g_i)$$

$$\text{s.t.} \quad |\mathcal{G}| = k,\ |g_i| = \frac{n}{k}$$

Simplified Formulation

The learning potential expressions are polynomial time solvable problems, because the primary operation they require is sorting.

$$\underset{\mathcal{G}}{\text{optimize}} \quad \sum_{i=1}^{k} \mathit{Aff}(g_i)$$

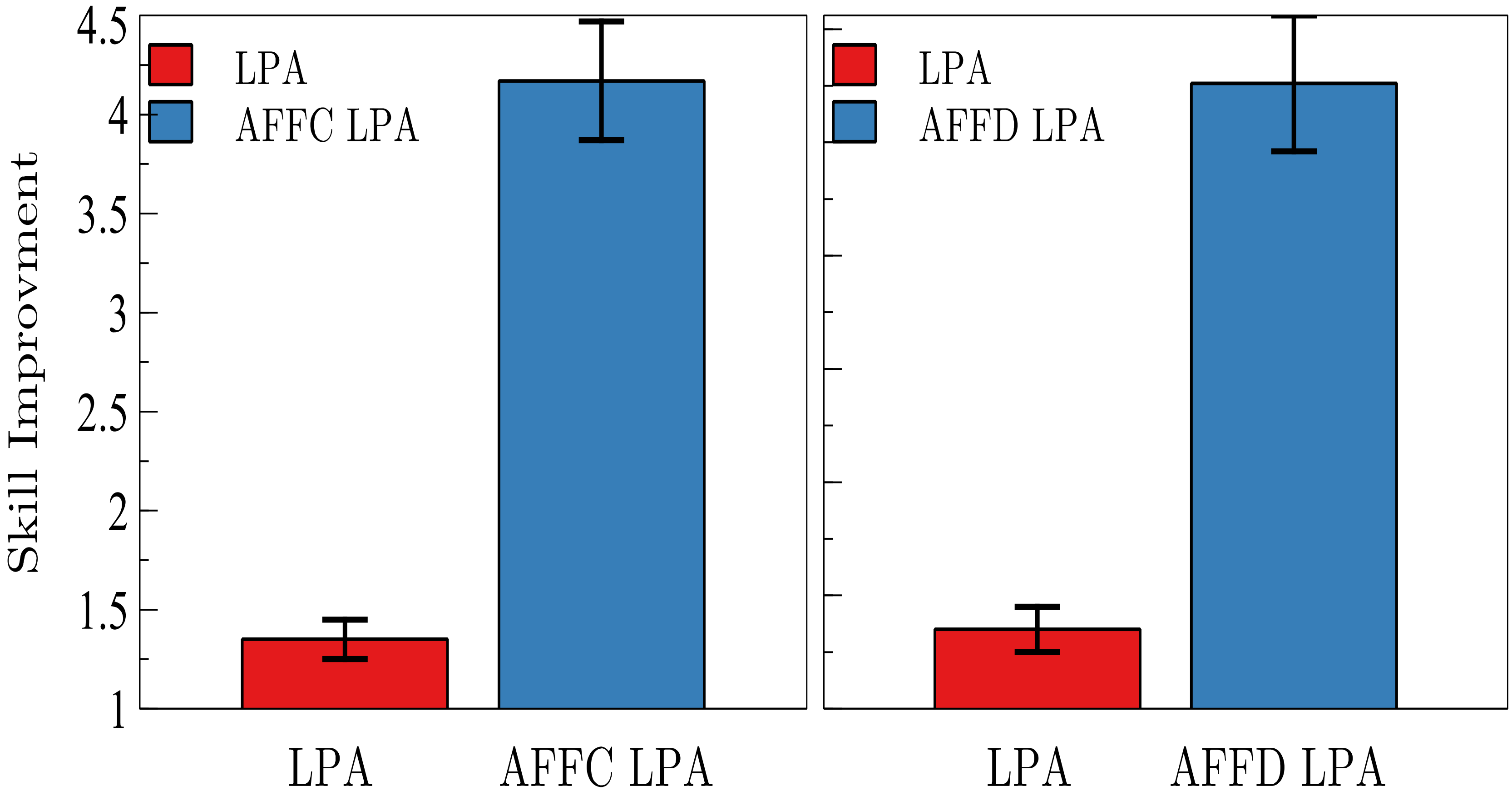
$$\text{s.t.} \quad \sum_{i=1}^{k} LP(g_i) \geq OptLP$$

$$|\mathcal{G}| = k, \; |g_i| = \frac{n}{k}$$

Problem variants and algorithms

Problem	Algo.	Approx.	Time
(AffC LpD)	GrAffC-LpD	exact LpD, 3 AffC	$O(k\log n + n\log k)$
(AffC LpA)	GrAffC-LpA	exact LpD, 3 AffD	$O(n\log n)$
(AffD LpD)	GrAffD-LpD	exact LpA, 6 AffC	$O(k\log n + n\log k)$
(AffD LpA)	GrAffD-LpA	exact LpA, 6 AffD	$O(n\log n)$

Experiments with fact checking/learning



Summary

- We, researchers, have a big role to play in
  - providing fairness assessment tools
  - helping workers find jobs that improve their skills, and account for human factors such as affinity and motivation

- Existing platforms can rethink their design to empower humans and be at the frontier of FoW.

Open challenges

- **Fairness**
  - explain and repair discrimination
- **Learning**
  - train for a new job with upskilling strategies
- **Putting it all together**
  - optimize for more than one objective
  - enable portability across platforms by building ML-enabled human data management systems