

# Lecture 6:

From the Other Side:  
Who are the “Humans” in Human-in-the-Loop Systems

Chien-Ju (CJ) Ho

# Logistics: Assignment 1

- Due: Feb 9 (Friday)
- For majority voting, we mean simple majority voting
  - Break ties randomly when there are equal votes for each label
    - Equivalent (in expectation) to set the error to 0.5 in this case
- EM (only assuming unknown worker skills):
  - How to choose the weights during the step of weighted majority voting?
  - The MLE weight  $\log(p/(1-p))$  is very sensitive to noise when  $p$  approaches 0 or 1
  - You can do either one of the following:
    - Crop the probability using ideas in Section 4 of “Who Moderates the Moderators”
    - Use the  $(2p-1)$  weight for workers with correct probability  $p$ 
      - technically not leading to MLE, but it give the smallest theoretical error bounds (see lecture 3 and 4)
- SVD
  - As a sanity check, you should get an error 0 when you simulate a complete matrix
    - Let every worker label every task according to ground truth and worker skills

# Logistics: Presentation

- The schedule is posted on Piazza

Date	Topic	Presenters
Feb 12	Incentive Design: Financial Incentives	CJ
Feb 14	Incentive Design: Badges and Attention	CJ
Feb 26	Workflow Design for Crowdsourcing	CJ
Feb 28	Learning in the Presence of Disagreements	Kai Ang, Sangwook Suh
Mar 4	LLM as a Proxy for Humans	Yifan Yuan, Ilan Barr, Nancy Patel
Mar 6	Adapting Crowdsourcing Techniques for LLM	Oscar Ortiz, Zhiyuan He
Mar 18	Fairness in AI	CJ
Mar 20	Human Perceptions of Fairness	Oen McKinley, Kyle Stein
Mar 25	Ethical Decision Making and Participatory Design	Garrett Kearney, Jake Valentine, Leib Malina
Mar 27	Human Trust in AI-Assisted Decision Making	Vincent Siu, Arpit Jain
Apr 1	Designing AI for AI-Assisted Decision Making	Joshua Tang, Sunny Yuan, Meichuan Yin
Apr 3	Explainable Machine Learning	Kaitlin Day, Micah Benson, Victoria Black
Apr 10	Human-Centered Explainable Machine Learning	Tory Farmer, Stuart Aldrich
Apr 15	Designing Collaborative AI in Human-AI Teams	CJ

# Logistics: Presentation

- For presenters:
  - Give a **55~60 min** presentation
    - Based on the **required reading** and **N-optional reading** for N-person groups
    - The papers are the “backbone” of the presentation
  - Prepare **2 reading questions** for the required reading
  - Prepare around **~2 discussion sessions**
  - Lead the discussion for the discussion sessions
  - Template format (if you are not sure what to do):
    - Discussion on the required reading (10~15 min)
    - Discussion session (5~10 min)
    - Discussion on the optional reading (15 min)
    - Another discussion session (5~10 min)
    - Another optional reading + summary (15 min)
    - Feel free to be creative and include materials outside of the papers

# Logistics: Presentation

- For presenters:
  - You do not need to submit the review for the lecture of your presentation
  - Talk to me **one week before your presentation**
    - Default time: talk to me after class
  - You need to be ready for the following before meeting with me
    - Finish reading the papers
    - A structure of your presentation (no need to show me the completed slides)
    - Topics for the discussion sessions
    - Two reading questions for the required reading

# Today's Lecture

# Assignment 1 in Past Years

- Be a crowd worker and make at least \$0.25
  - Motivated by this post: [My MTurk \(half\) Workday](#). Jeff Bigham
- How much students earn per hour?

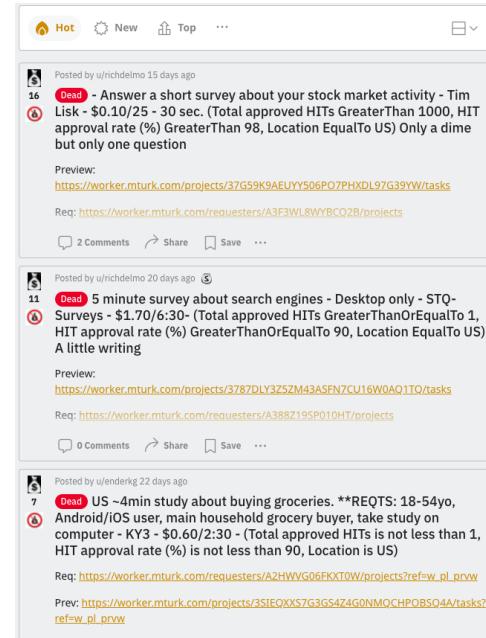
12.30	5.51	2.92	1.80	1.00	0.75
10.71	5.40	2.88	1.51	1.00	0.75
10.00	5.00	2.74	1.43	0.96	0.50
9.00	5.00	2.48	1.43	0.81	0.50
8.00	5.00	2.40	1.38	0.80	0.41
6.00	4.35	2.08	1.15	0.77	0.40
6.00	4.15	2.00	1.12	0.76	0.31

Mean: \$3.18

Median: \$1.90

# Some Attempts to Improve Work Environment

- Enable information transparency
  - Online forums, requester reputation tools



- Alternative platforms that enforce minimum wage

**Prolific**

Ethical rewards

On Prolific, we endorse the principle of 'ethical rewards'. We believe that fair pay leads to better data quality. We therefore recommend you pay participants at least £9.00 / \$12.00 per hour, while the minimum pay allowed is £6.00 / \$8.00 per hour.

# Crowdsourcing (Requester's Perspective)



Input



Black Box

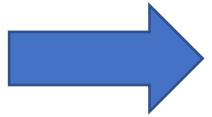


Flower  
Dog  
Cute

...

Output

# The Crowd is Made of People



Flower  
Dog  
Cute

...

Who are these people?

# Why Should We Care?

- So we know how to better utilize and interact with the crowd
  - CS tools let us reason about machine programs (runtime, scalability, correctness, ...)
  - Need to develop **models of human behavior** when humans are in the loop
  - Most studies so far make strong assumptions about human behavior
- Remind us crowd workers are humans like us (sounds obvious, but...)
  - Human-centered research has been receiving great attention
  - Ethical-related issues (fairness, transparency, and privacy) are important
- **Relatively under-explored.** Potential interesting topics for your projects...
  - They might be quite challenging though

# Who Are These Workers?

- Demographic information
- Are workers working independently?
- How many workers are there?
- ...?

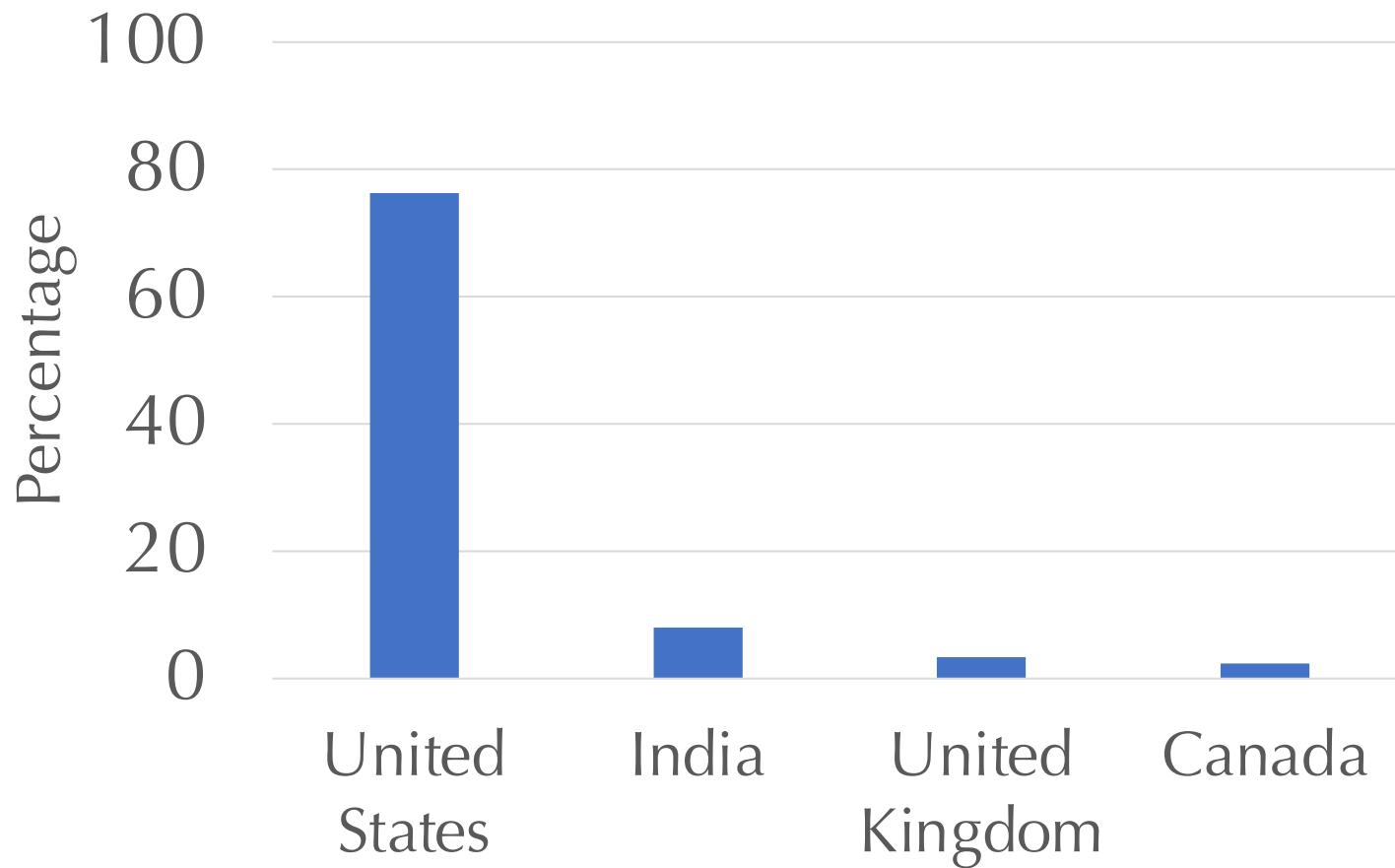
# The information is not trivial to get

- There is no official census data about crowd workers
- Potential solutions:
  - Conducting surveys
  - Ethnographic analysis of digital trace
  - What else?

# The (Old) Demographic of MTurk

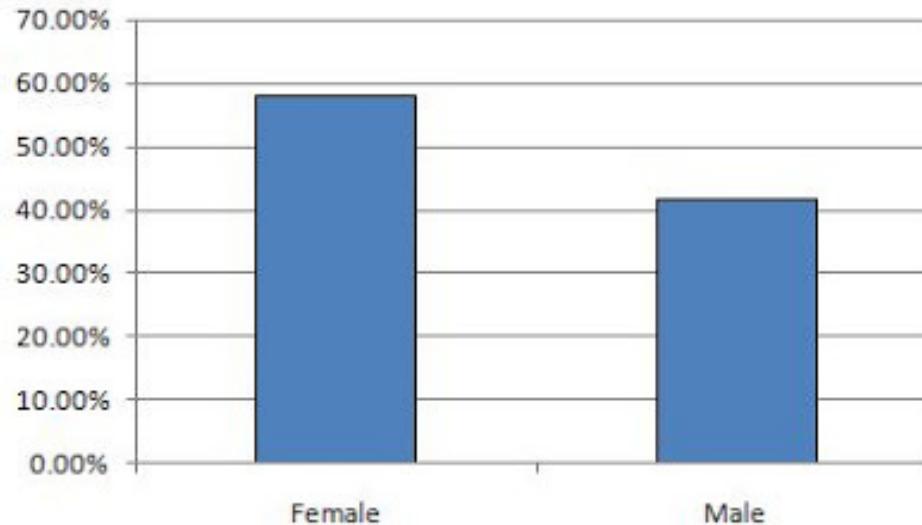
- Survey on 1,000 Turkers
  - Conduct the survey twice (Dec. 2008 and Oct. 2008)
  - Consistent statistics
  - Blog Post:
    - A Computer Scientist in a Business School
- One of the early attempts to understand MTurk demographics

# MTurk Worker Demographics (2008): Country

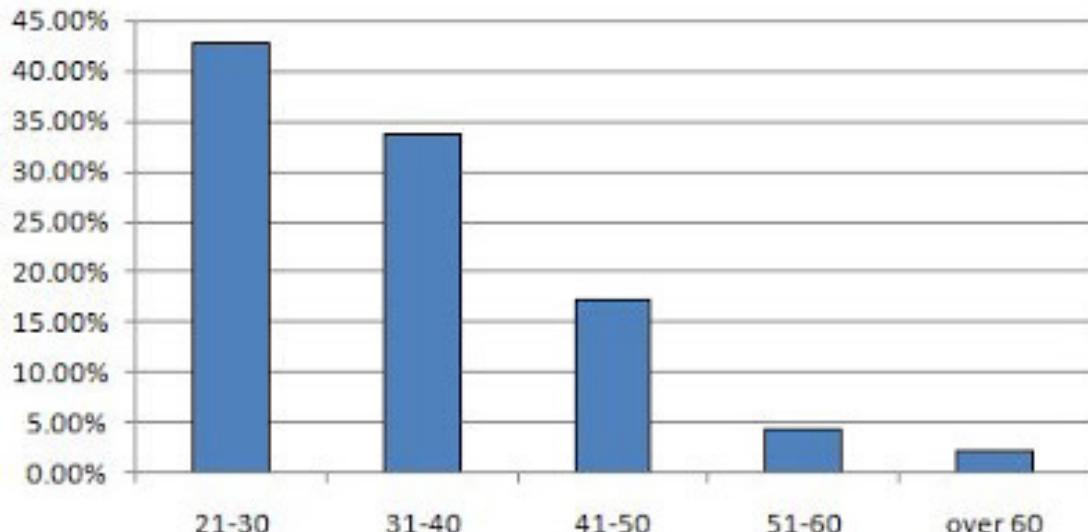


The majority of crowd workers on MTurk come from the US

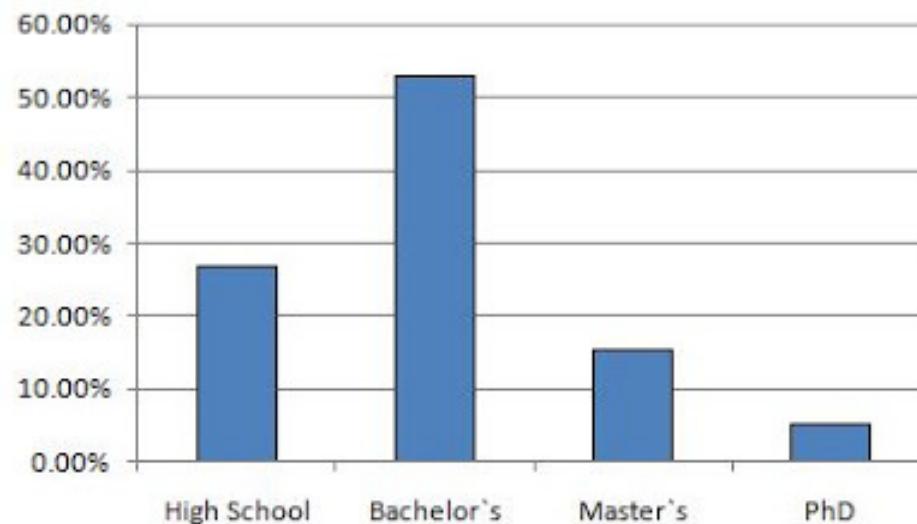
## Gender



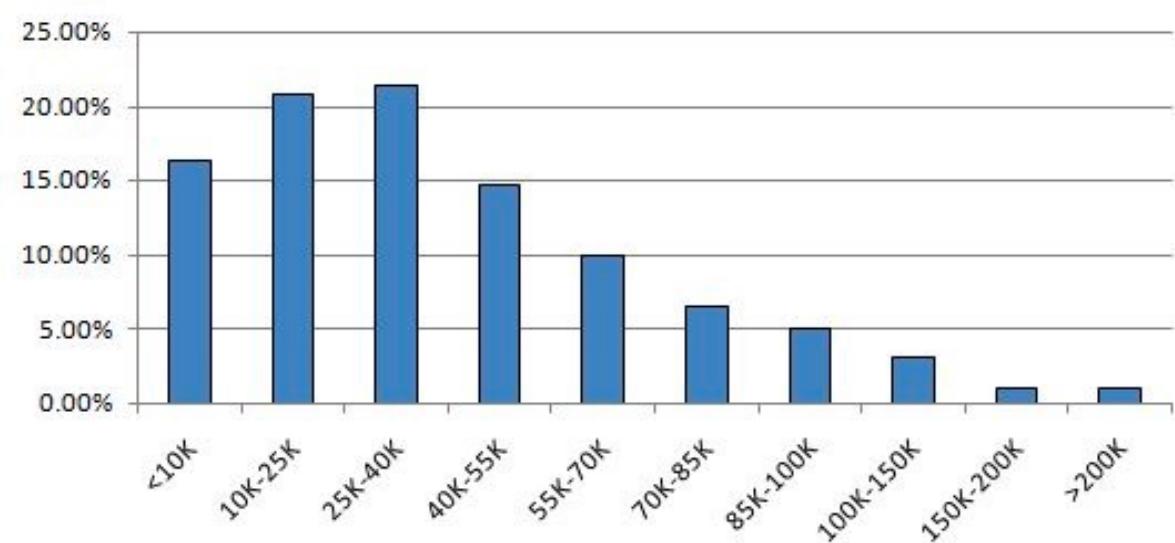
## Age



## Education



## Income



# MTurk Worker



**Younger** 54%

**More Female** 70%

**Lower Income** 65%

**Smaller Family** 45%

# Internet Users (2008, US)



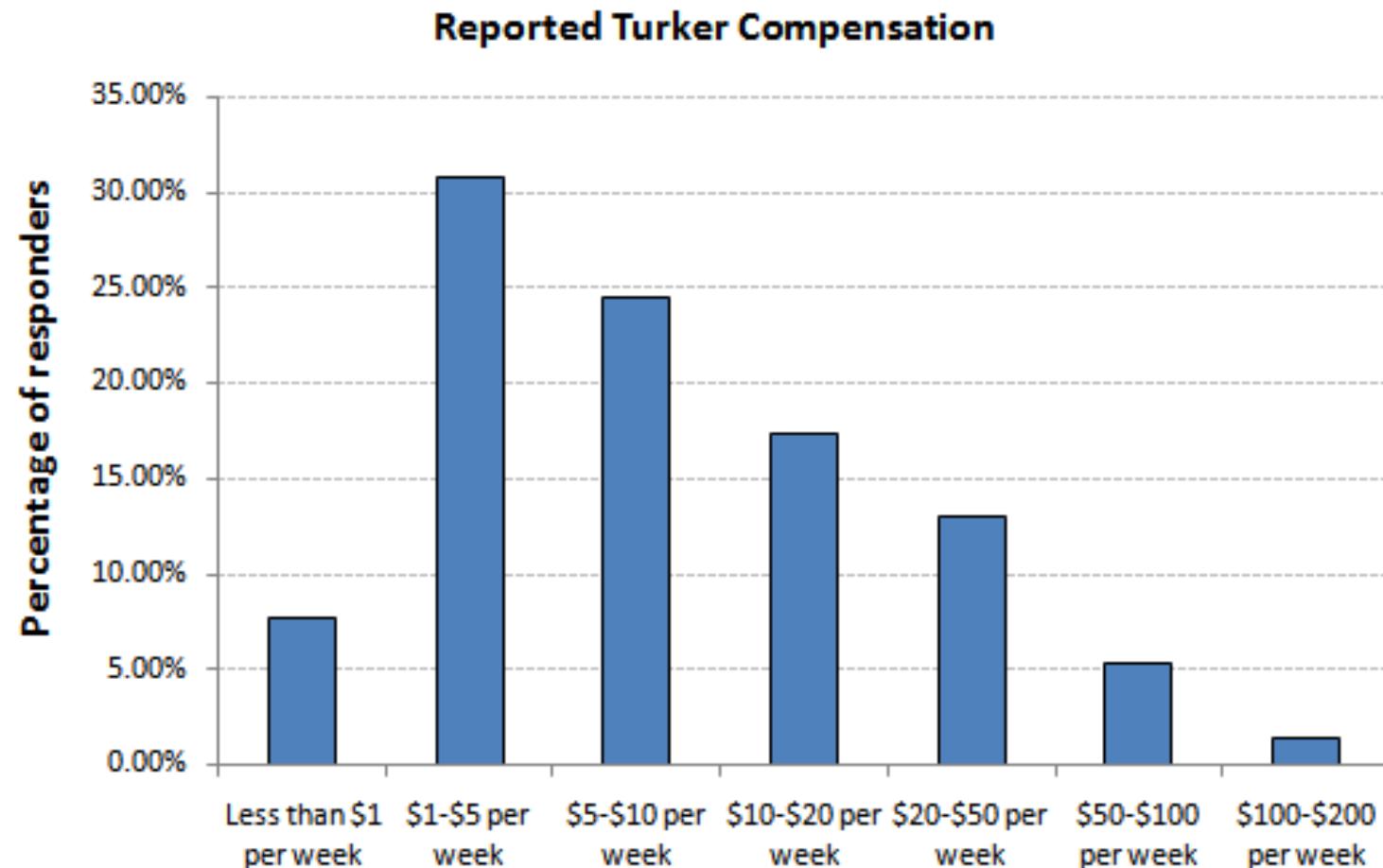
**Age: 21-35** 22%

**Female** 50%

**Household income < 60K** 45%

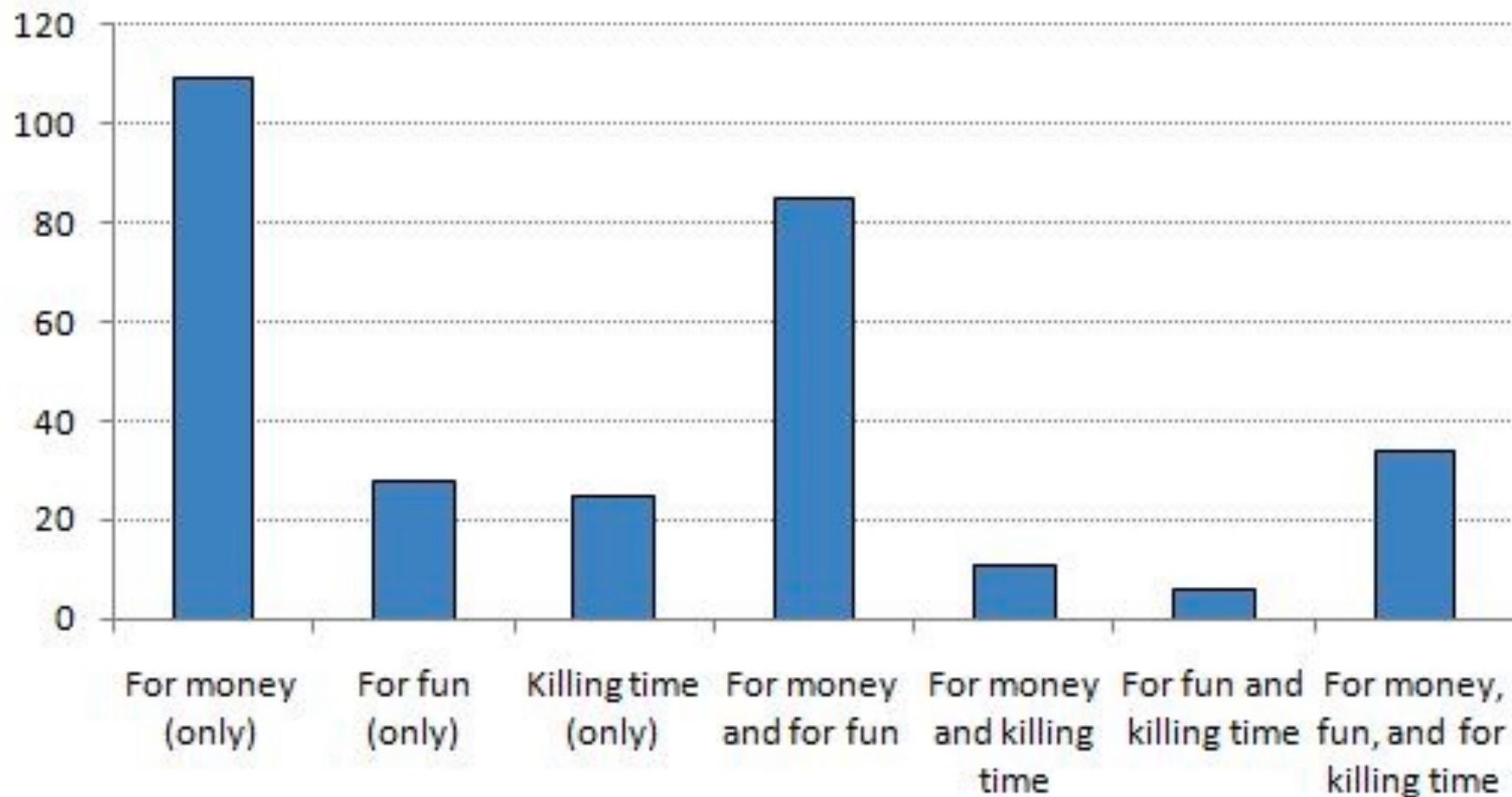
**Household <= 2 people** 28%

# MTurk Worker Demographics (2008): Payment received

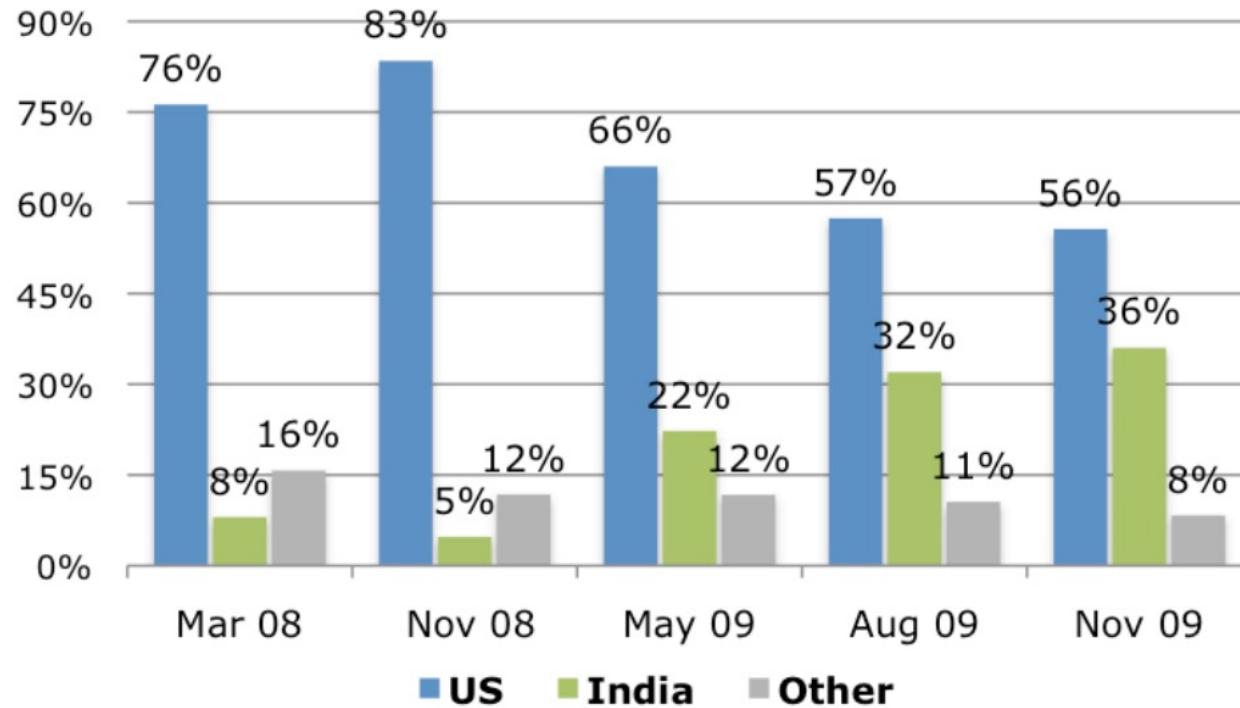


<http://www.behind-the-enemy-lines.com/2008/09/how-much-turking-pays.html>

# MTurk Worker Demographics (2008): Purpose



# But Demographics Also Changes Over Time...



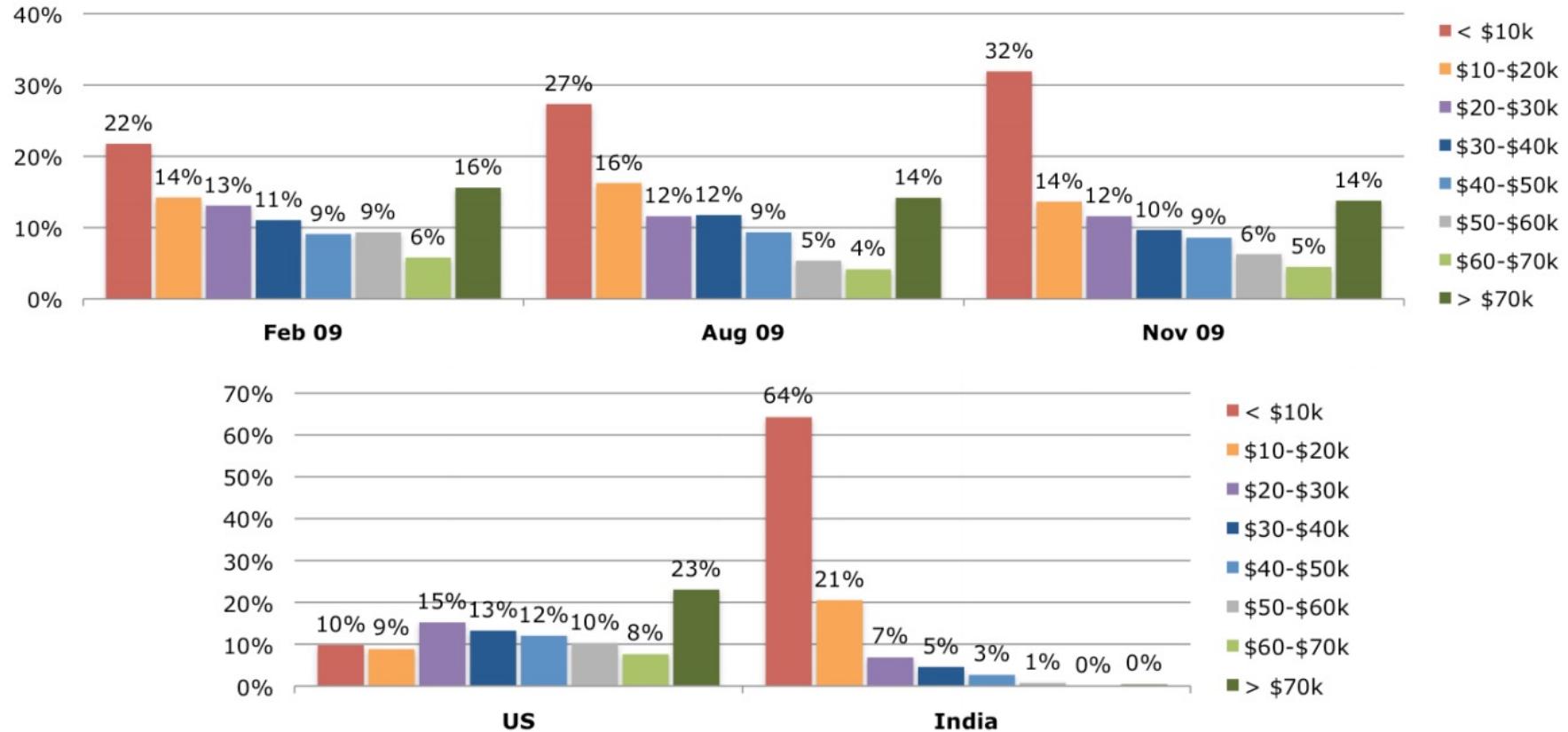
From a primary US-based workforce to an increasingly international group of workers.

# But Demographics Also Changes Over Time...

		<b>Nov 08</b>	<b>May 09</b>	<b>Aug 09</b>	<b>Nov 09</b>
<b>Average Age</b>	<i>US</i>	33.6	34.3	33.2	35.4
	<i>India</i>	28.5	28.8	27.6	26.4
<b>Gender</b>	<i>US</i>	28% male, 72% female	34% male, 66% female	31% male, 69% female	37% male, 63% female
	<i>India</i>	75% male, 25% female	61% male, 39% female	69% male, 31% female	66% male, 34% female
<b>Education</b>	<i>US</i>	32% Bachelors, 11% Graduate	34% Bachelors, 14% Graduate	34% Bachelors, 19% Graduate	38% Bachelors, 17% Graduate
	<i>India</i>	69% Bachelors, 29% Graduate	56% Bachelors, 18% Graduate	56% Bachelors, 13% Graduate	45% Bachelors, 21% Graduate

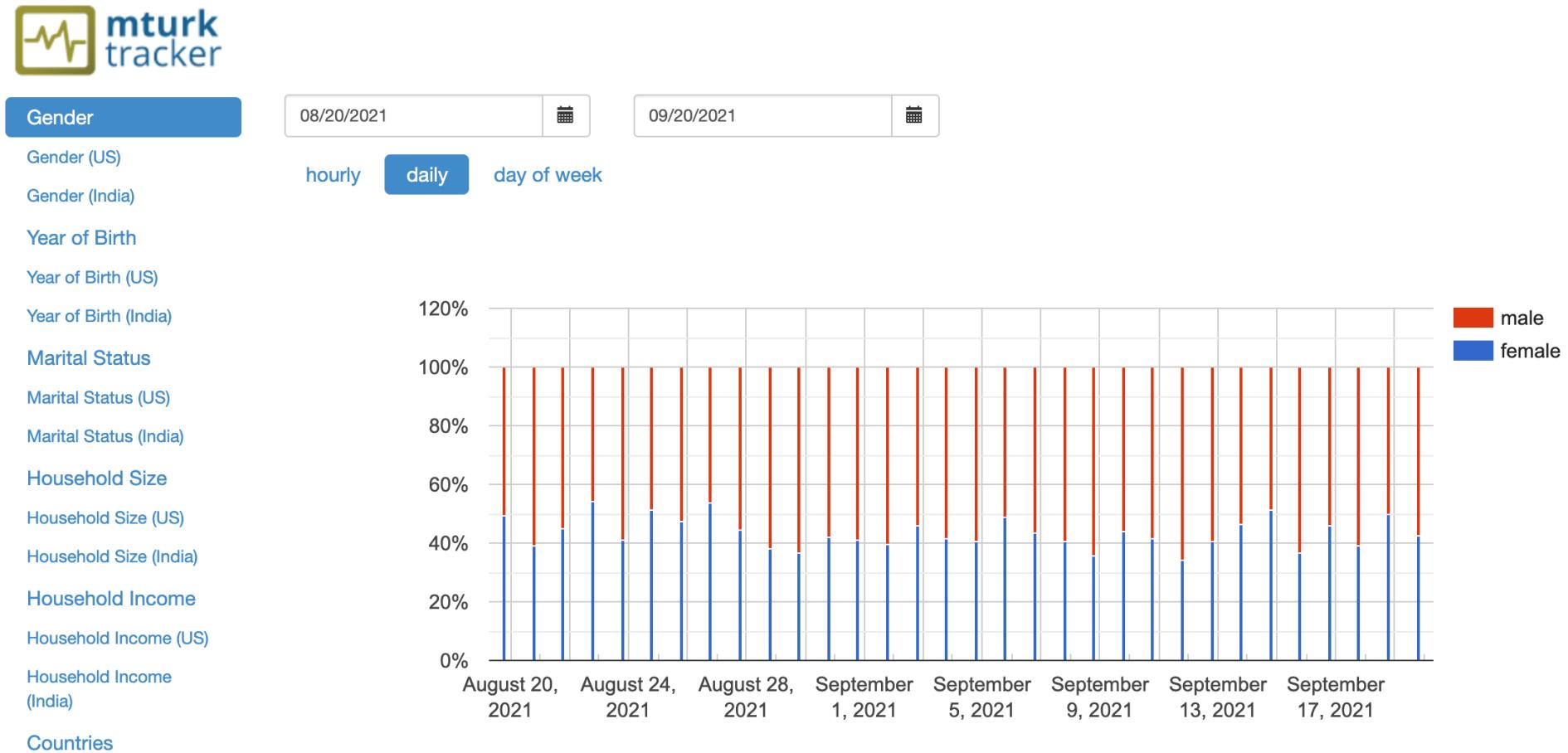
Ross, Joel, et al. "Who are the crowdworkers?: shifting demographics in mechanical turk." *CHI'10 EA*

# But Demographics Also Changes Over Time...



Ross, Joel, et al. "Who are the crowdworkers?: shifting demographics in mechanical turk." *CHI'10 EA*

# MTurk-Tracker: A Long-Term Demographic Survey



Each worker can take at most one survey in 30 days

<http://demographics.mturk-tracker.com/#/gender/all>

# MTurk-Tracker: A Long-Term Demographic Survey



## Gender

Gender (US)

Gender (India)

Year of Birth

Year of Birth (US)

Year of Birth (India)

Marital Status

Marital Status (US)

Marital Status (India)

Household Size

Household Size (US)

Household Size (India)

Household Income

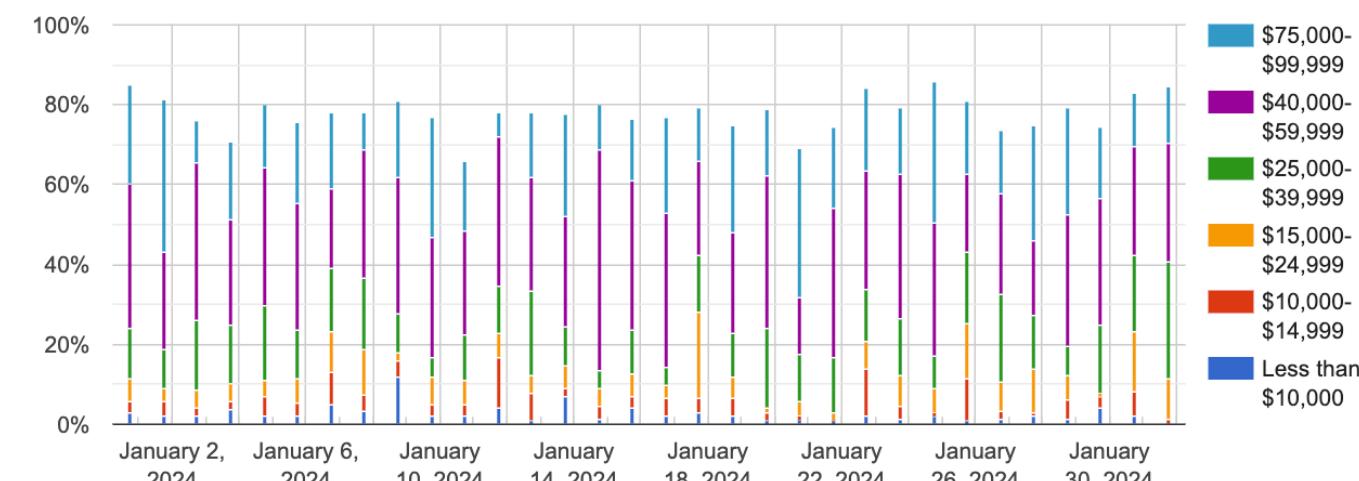
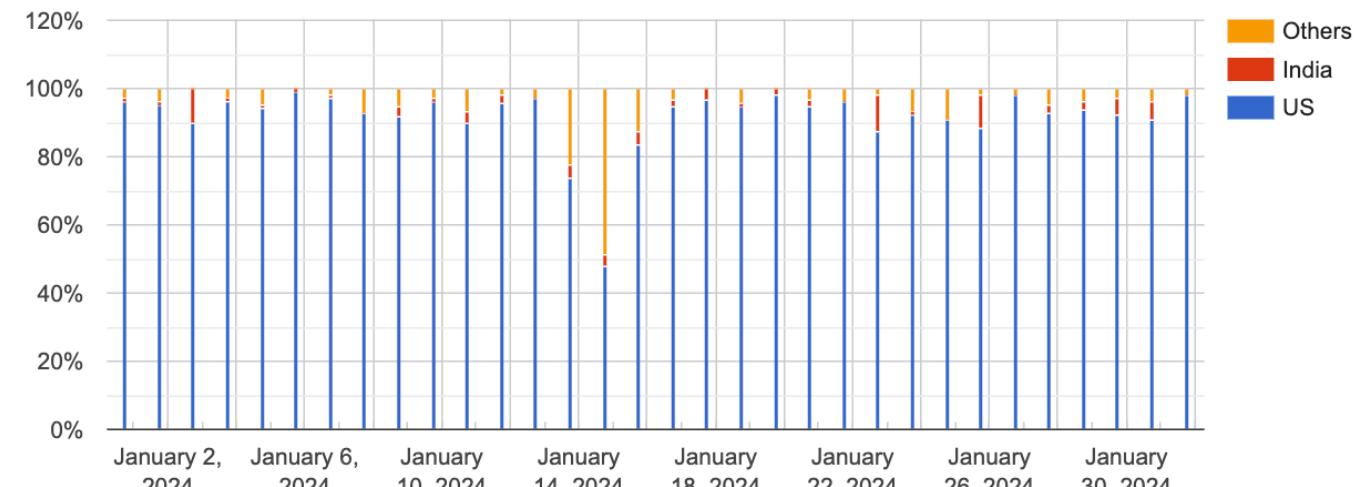
Household Income (US)

Household Income

(India)

Countries

API



# Warm-up Discussion

Does the demographic information collected from self-reported survey responses suffer from any problems?

Can you come up with potential fixes for the problems?

What about ethnographic analysis of digital trace in the required reading?

# Warm-up Discussion

Does the demographic information collected from self-reported survey responses suffer from any problems?

Can you come up with potential fixes for the problems?

What about ethnographic analysis of digital trace in the required reading?

- Potential issue 1: Data bias (sampling bias, human bias)
- Potential issue 2: Untruthful self-reports

# Sampling Bias

Ideally, we want to select participants from the population “uniformly at random”.

However, it’s not always that easy...

# 1948 US Presidential Election

- Truman vs. Dewey
- Chicago Daily Tribune decided to run a phone poll of how people voted



Truman



# What Happened?

One explanation: we cannot claim anything for certain.

However, there are bigger issues here...

- Phones are expensive in 1948...
- Dewey was more favored in rich populations
- Imagine you are polling from people in DC/Texas/NY to predict who will win the presidential election...

# Sampling Bias

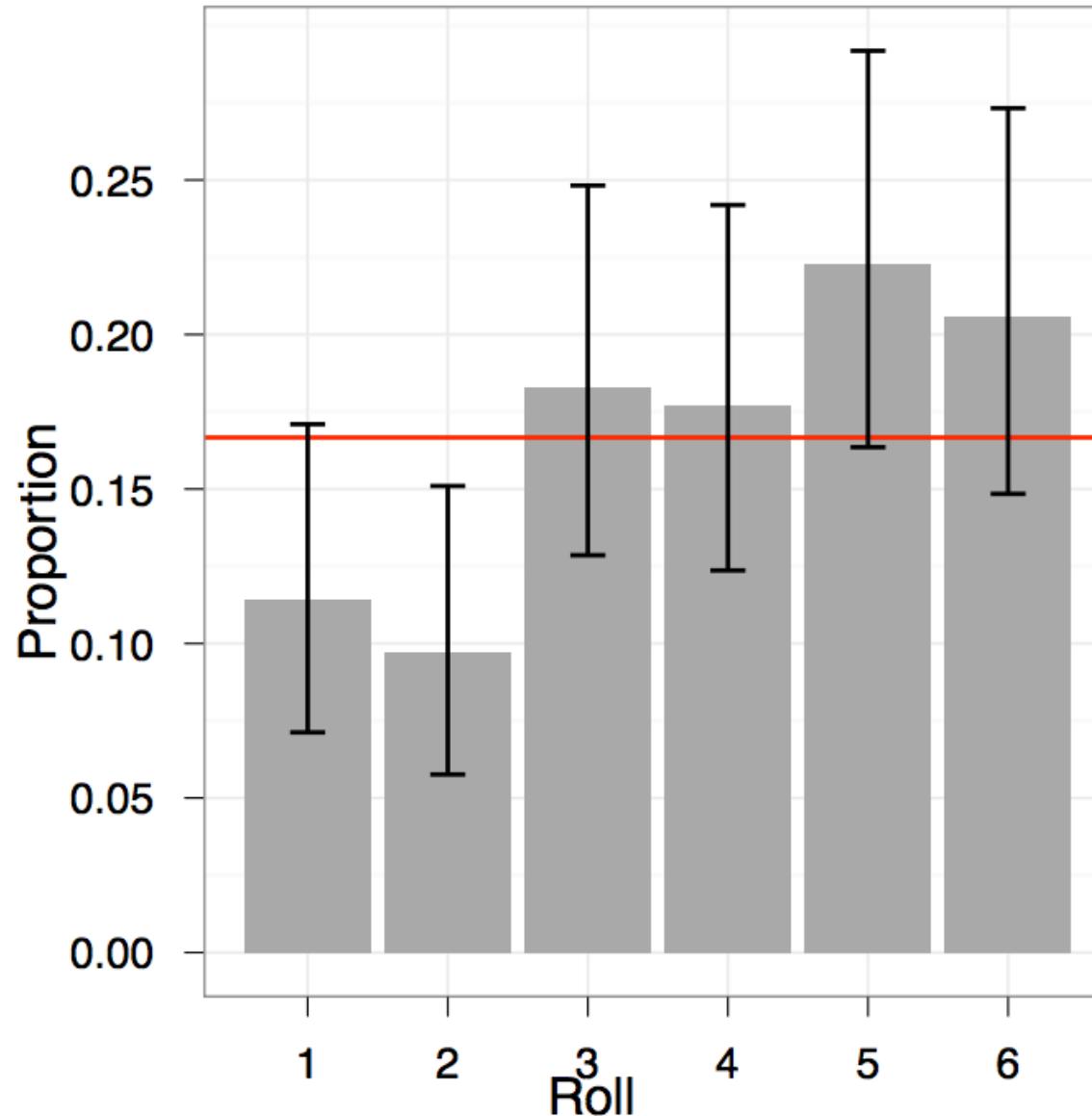
- No simple solutions: Be aware of the issue and think about the impact it might bring in your results.
- You might be able to “correct” the bias if you have additional knowledge.
  - Importance sampling
  - Potentially interesting/relevant paper
    - [Conducting Truthful Surveys, Cheaply](#). Roth and Schoenebeck. EC 2012.

# Are Workers Honest?

- Workers are asked to answer demographics questions [Suri et al. 2011]
  - Sex, Age, Location, Income, Education
- Ask workers to privately roll a die and report the outcome.

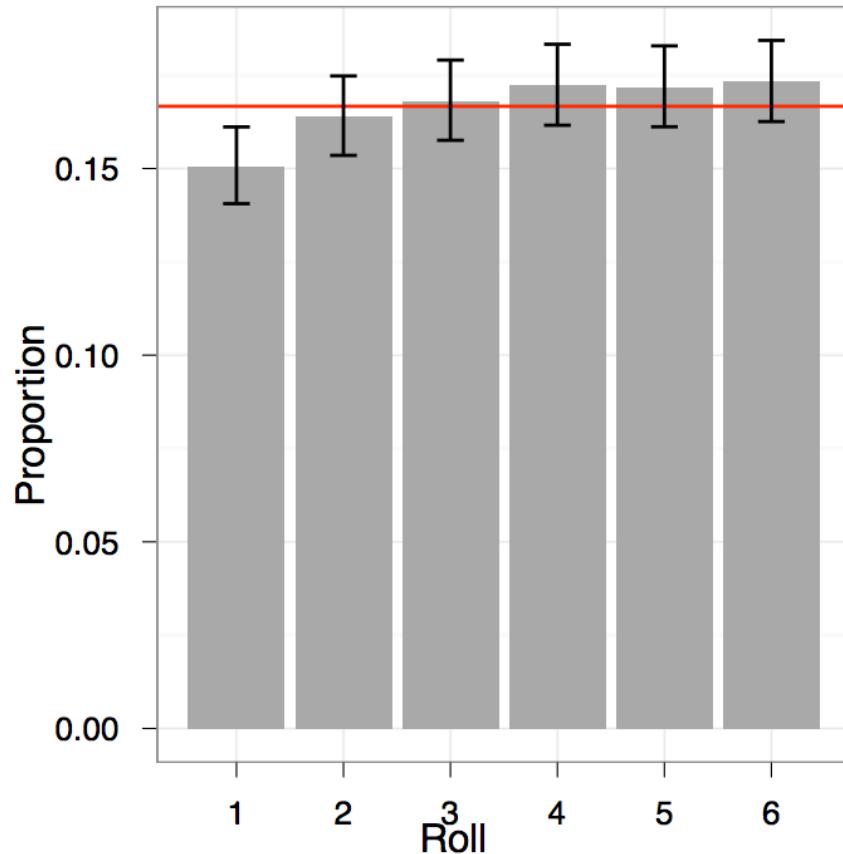
$$\text{Payment} = \$0.25 + \$0.25 * \text{roll}$$

- If all workers are honest, mean report: 3.5
- What do you think the mean was?



Mean: 3.91

# Ask Workers to Report 30 Rolls



- Not conclusive evidence, but workers are more honest than we think.
- However, some workers are not. We should be careful to avoid attacks.
  - E.g., designing consistency check questions

# Methods for Information Collection

- Direct surveys
- Ethnographic analysis (as in the required reading)
  - Rely on researchers' judgements. Not quantitative.
  - Slowly increasingly, ML has come to help.
- More intricate methods
  - Example 1: How to measure the communication network of the crowd
  - Example 2: How to measure the size of the crowd

# The Communication Network Within the Crowd

Ming Yin , Mary L. Gray , Siddharth Suri , Jennifer Wortman Vaughan. WWW 2016

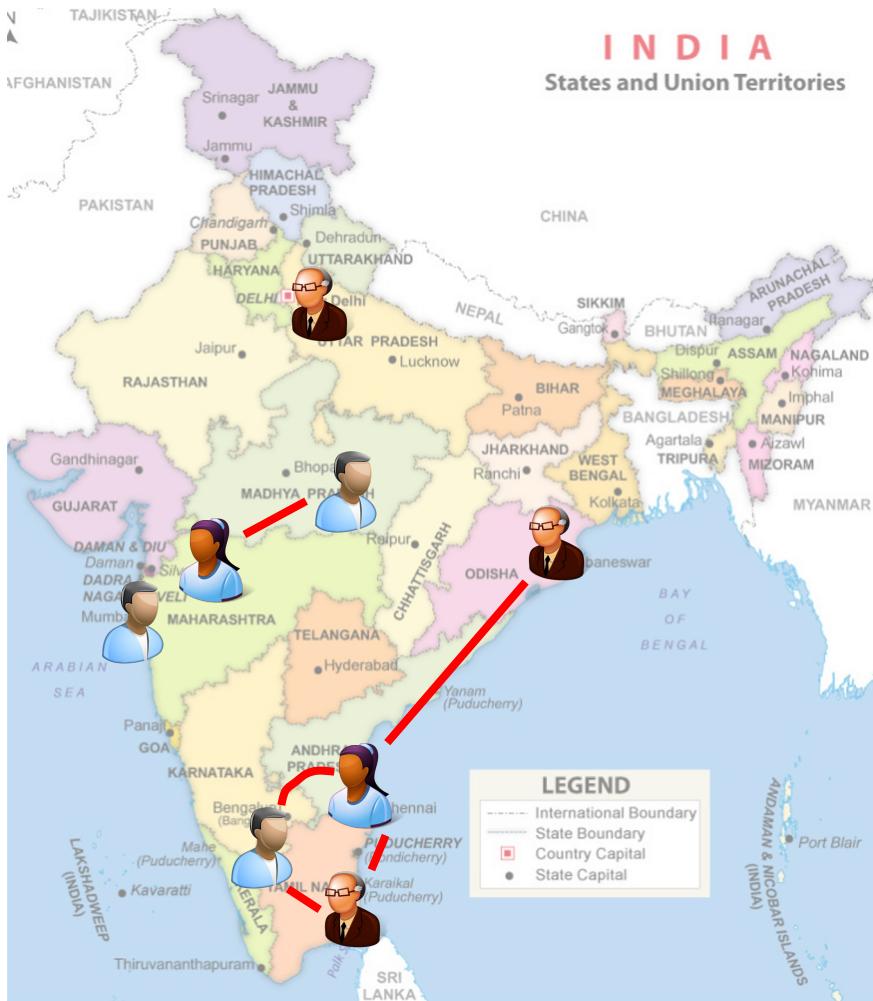
Acknowledgement: The next few slides are from Ming Yin.

# Common Assumption: Workers are Independent



[Yin et al., 2016]

# Some Evidence From the Field



Workers talk to each other to...  
(based on ~100 interviews)



Help with  
administrative  
overhead

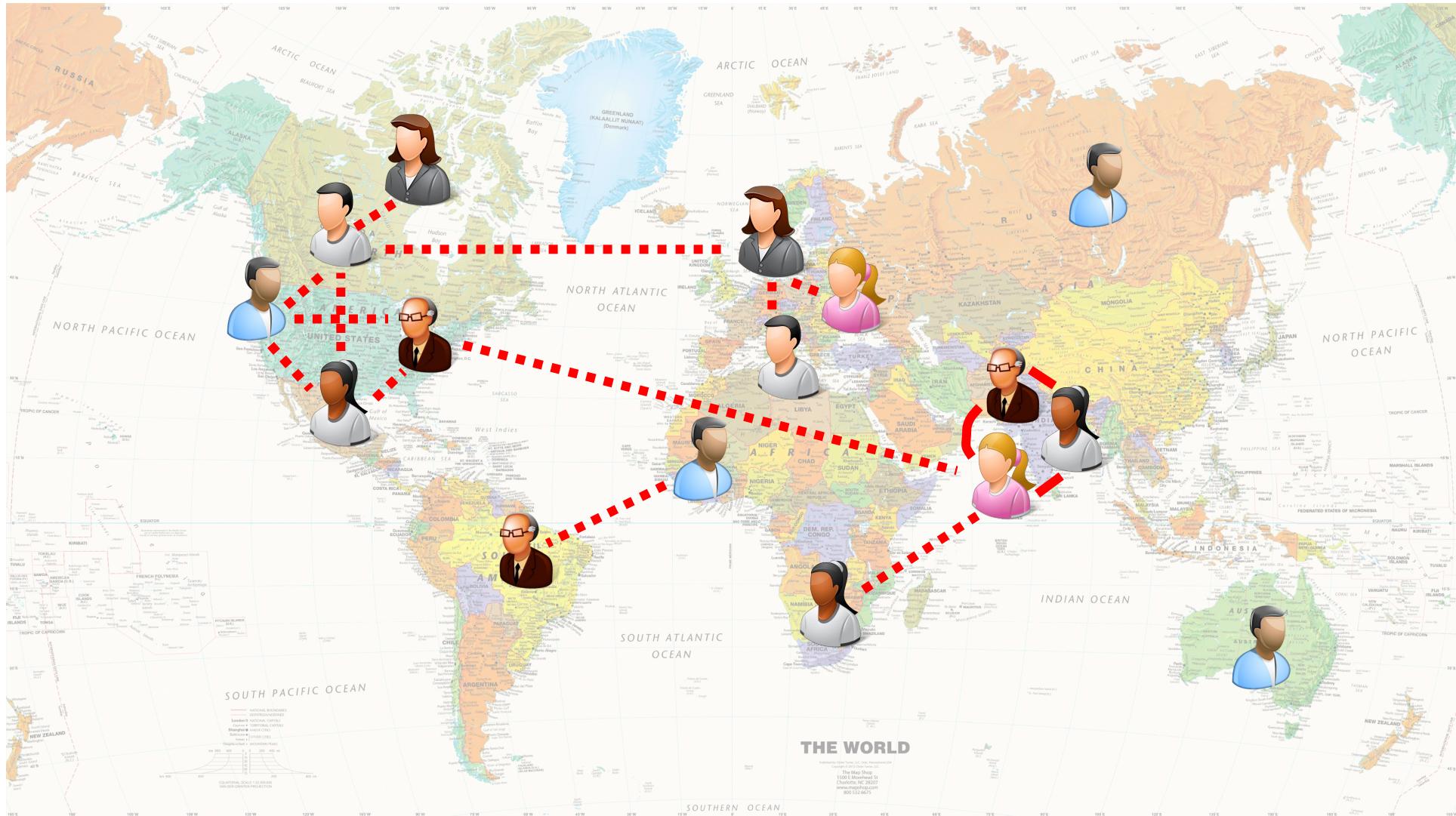


Share useful  
information



Recreate social  
connections

# Can we Map the Network?



*What is the scale?*

*What is the structure?*

*How is it used?*

[Yin et al., 2016]

# Why is it Challenging?



Not accessible  
through API



Not on the  
MTurk platform



Nowhere to  
download



Can't just crawl  
from the web

## The goal...

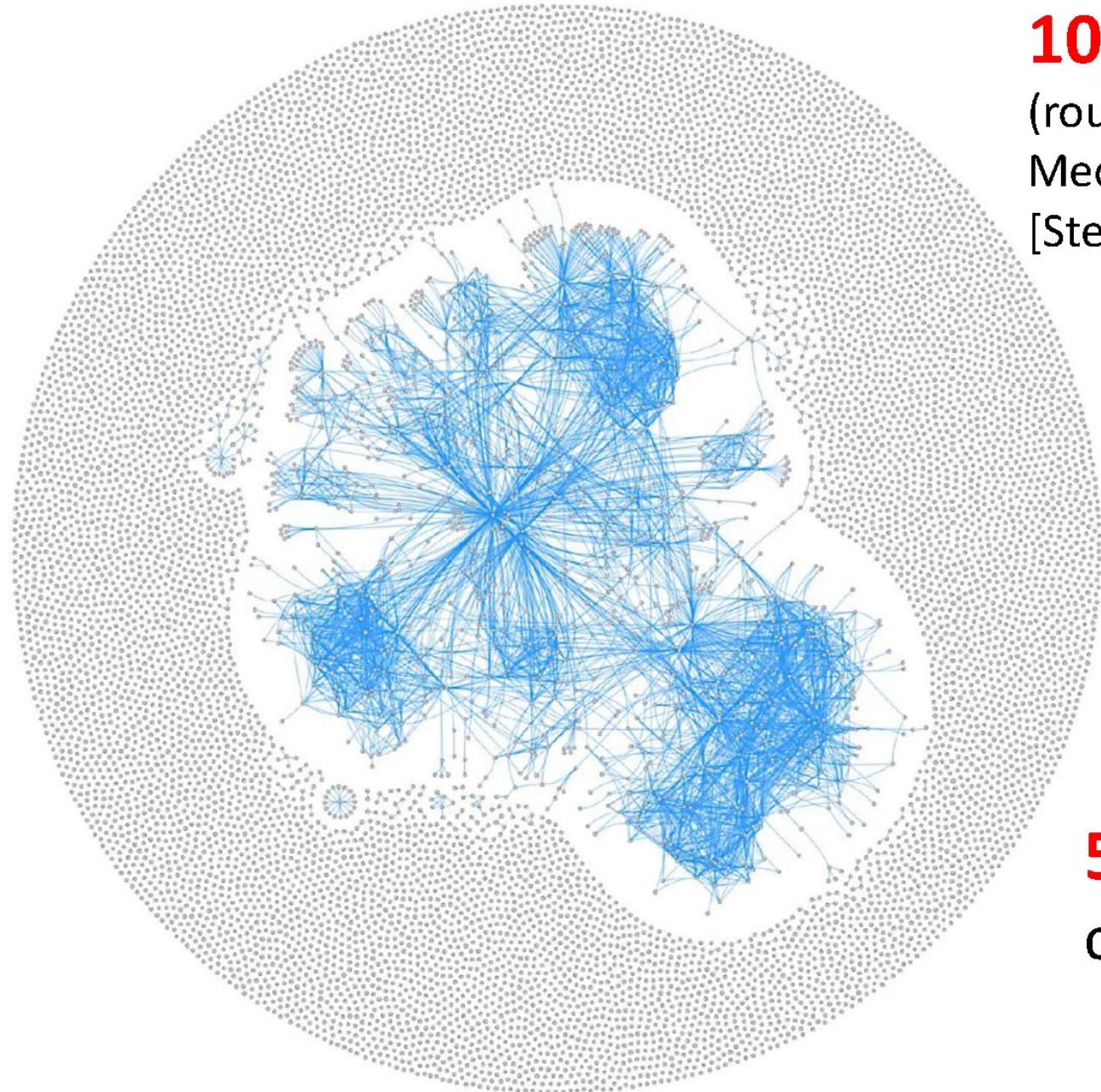
- ◆ Elicit “true” connections only
- ◆ Elicit as many true connections as possible
- ◆ Preserve workers’ privacy

**Can't pay by connections!**

**Can't directly ask for names!**

# A Web App

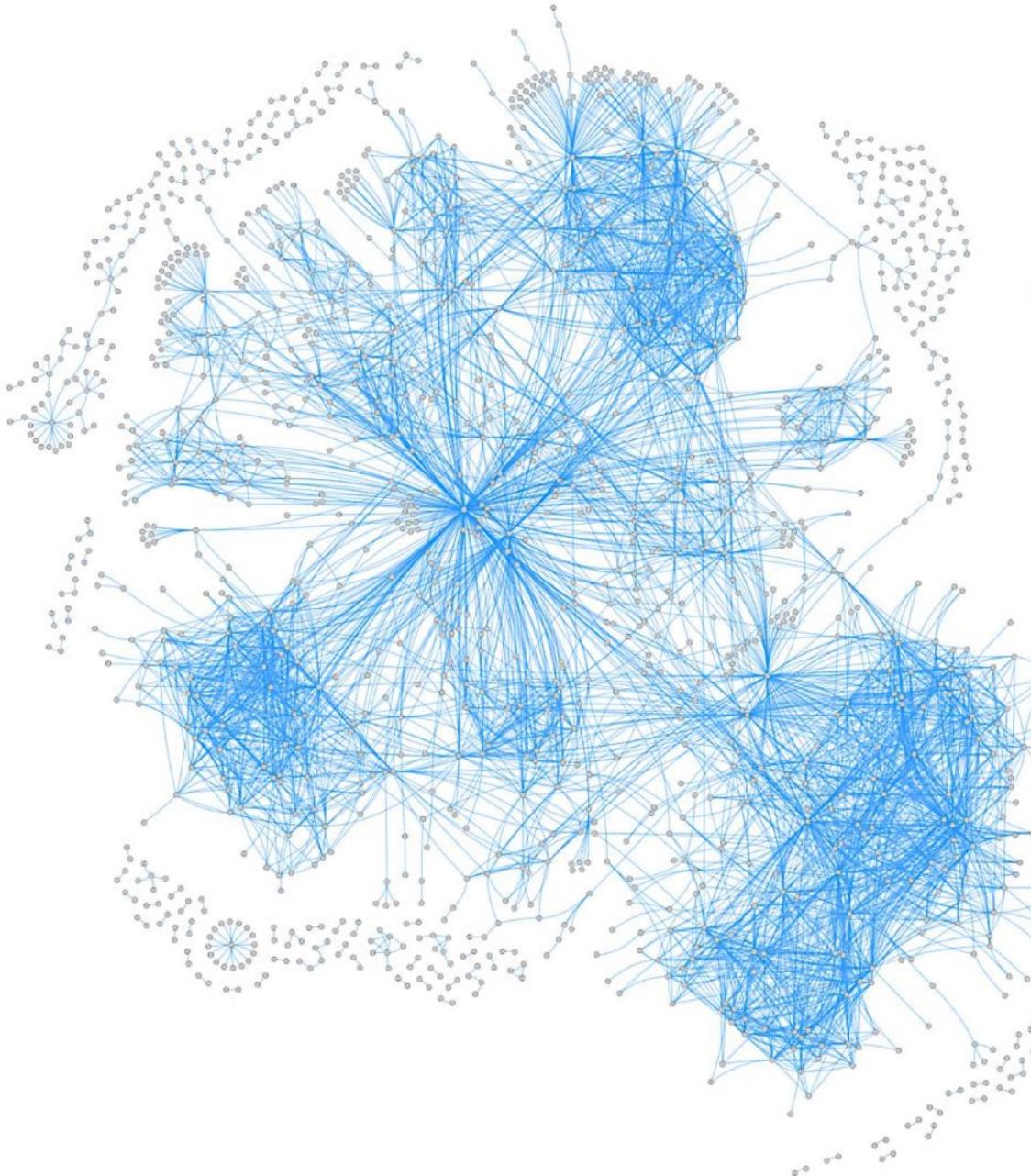
- Workers **self-report** their connections
- Provides some **value back** to the workers so it's their best interest to report as many true connections as possible



**10,354** workers  
(roughly a census of  
Mechanical Turk  
[Stewart et al. 2015])

**5268**  
connections

[Yin et al., 2016]

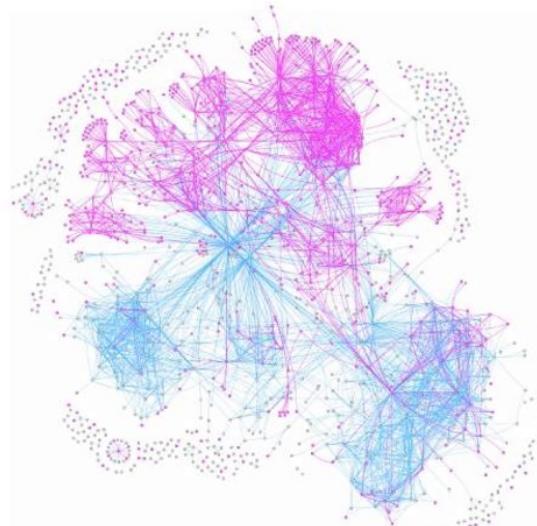


**1,389 (13%)**  
connected  
workers

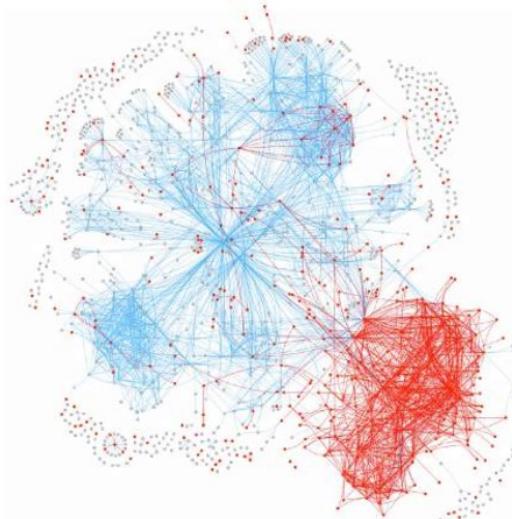
On average,  
workers  
communicate  
with **7.6** others

Max degree  
is **321**  
[Yin et al., 2016]

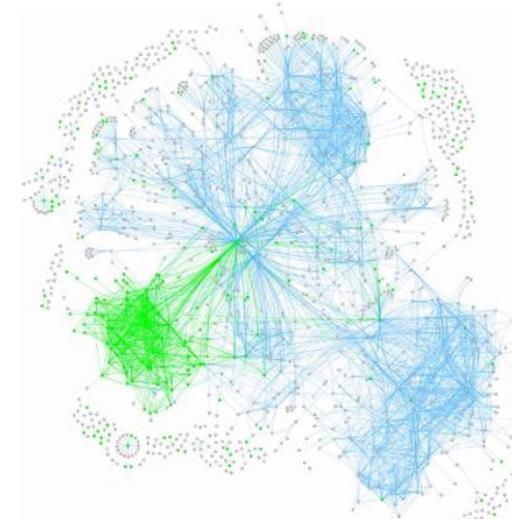
# Forums Create Subcommunities



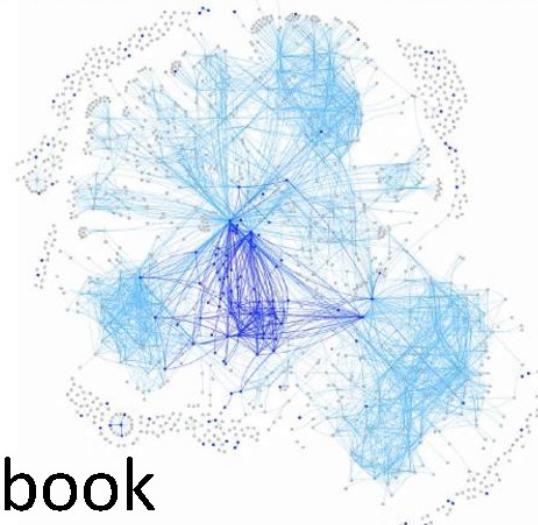
Reddit HWTF



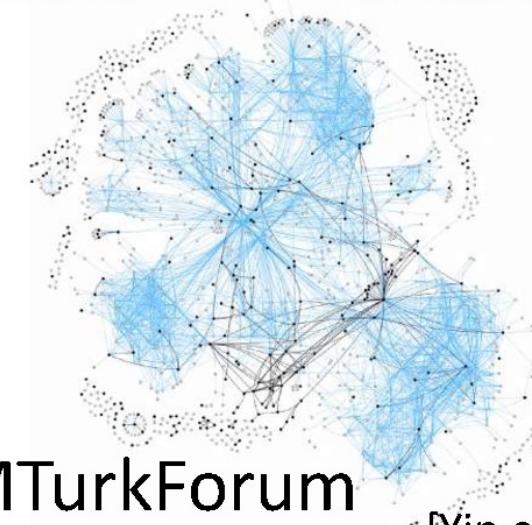
MTurkGrind



TurkerNation



Facebook



MTurkForum

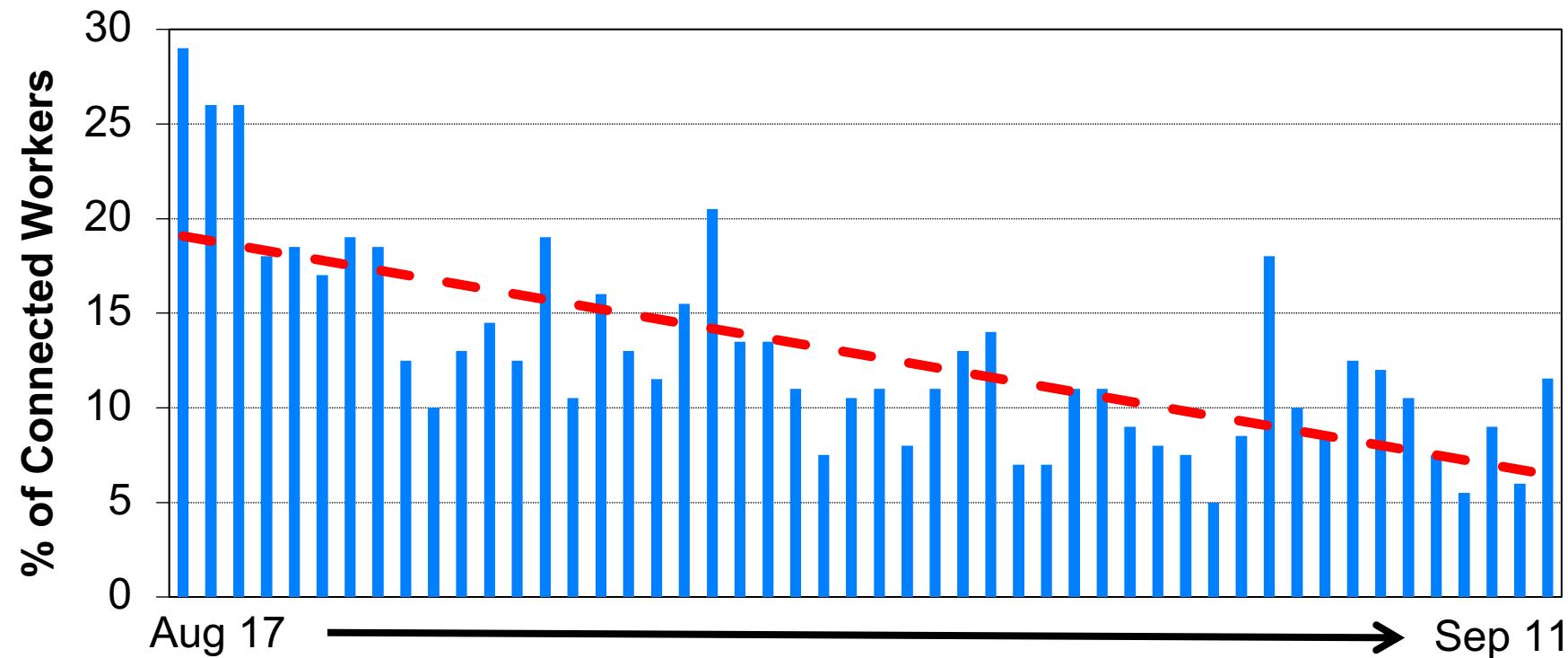
[Yin et al., 2016]

# Measures of Success

Property	Connected	Unconnected
Be active > 1 year	55%	46%
Use forums	83%	56%
Master	11%	7%
Approval rate	98.6%	97.4%

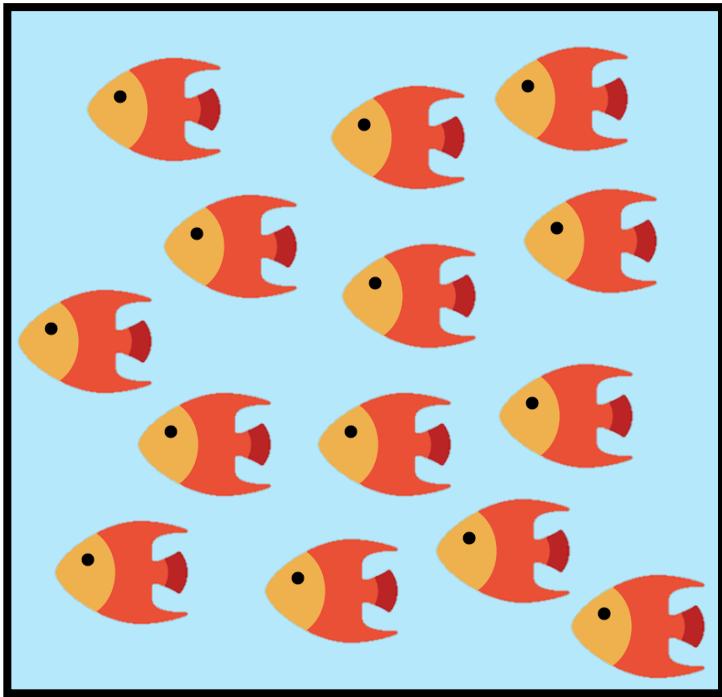
[Yin et al., 2016]

# Connected Workers Find HITs Earlier



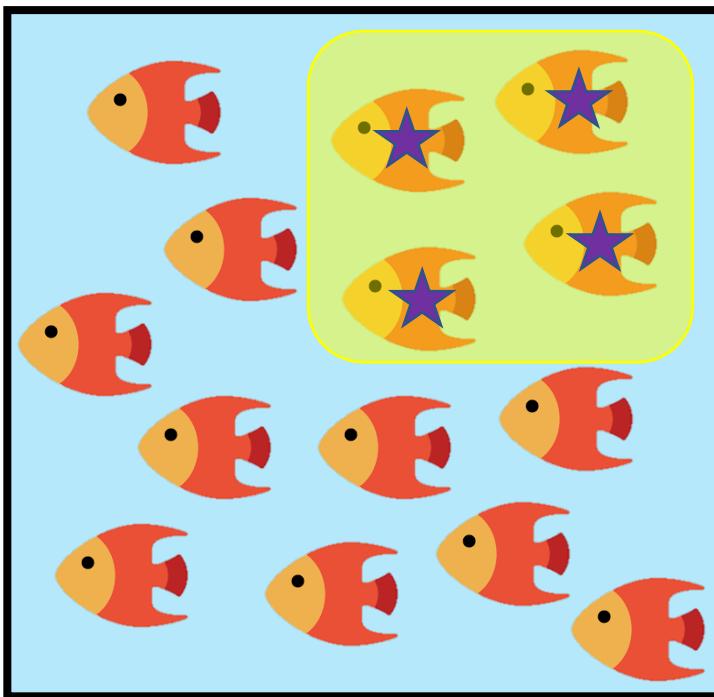
How Many Workers are There?

# How Many Fishes are in the Lake?

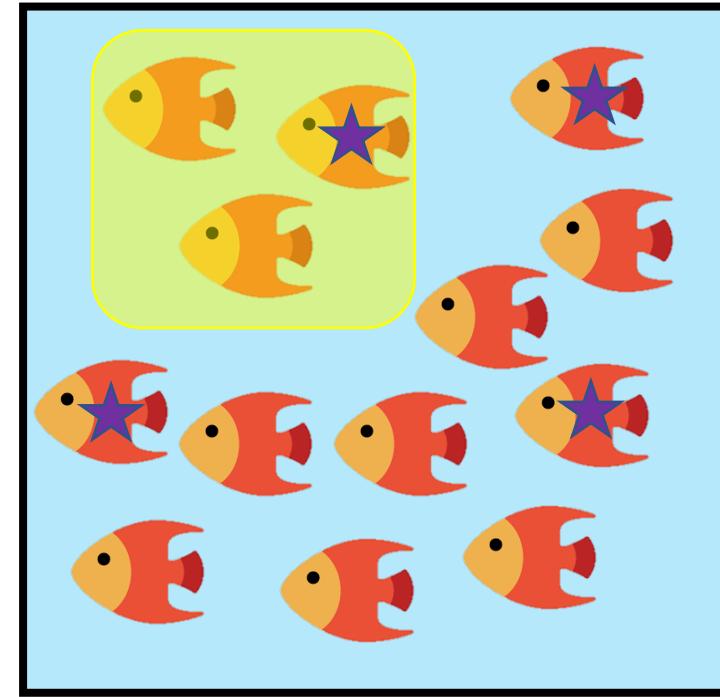


# How Many Fishes are in the Lake?

## The capture-recapture analysis



M fish,  $N_1$  captured (& marked)



$N_2$  recaptured,  $n$  marked

Assuming independence:

$$\frac{N_1}{M} = \frac{n}{N_2}$$

$$M = \frac{N_1 N_2}{n}$$

# How Large is the Crowd?

- Applying the capture-recapture analysis to workers who respond to demographic surveys, we may estimate that **M= 13,410**.
- What assumptions are made?
  - A1: Closed population model: No workers will leave the worker pool, and no new workers will join the pool
  - A2: Equal propensity of participation: The probability to participate in demographic survey is equal across all workers.

# How Large is the Crowd?

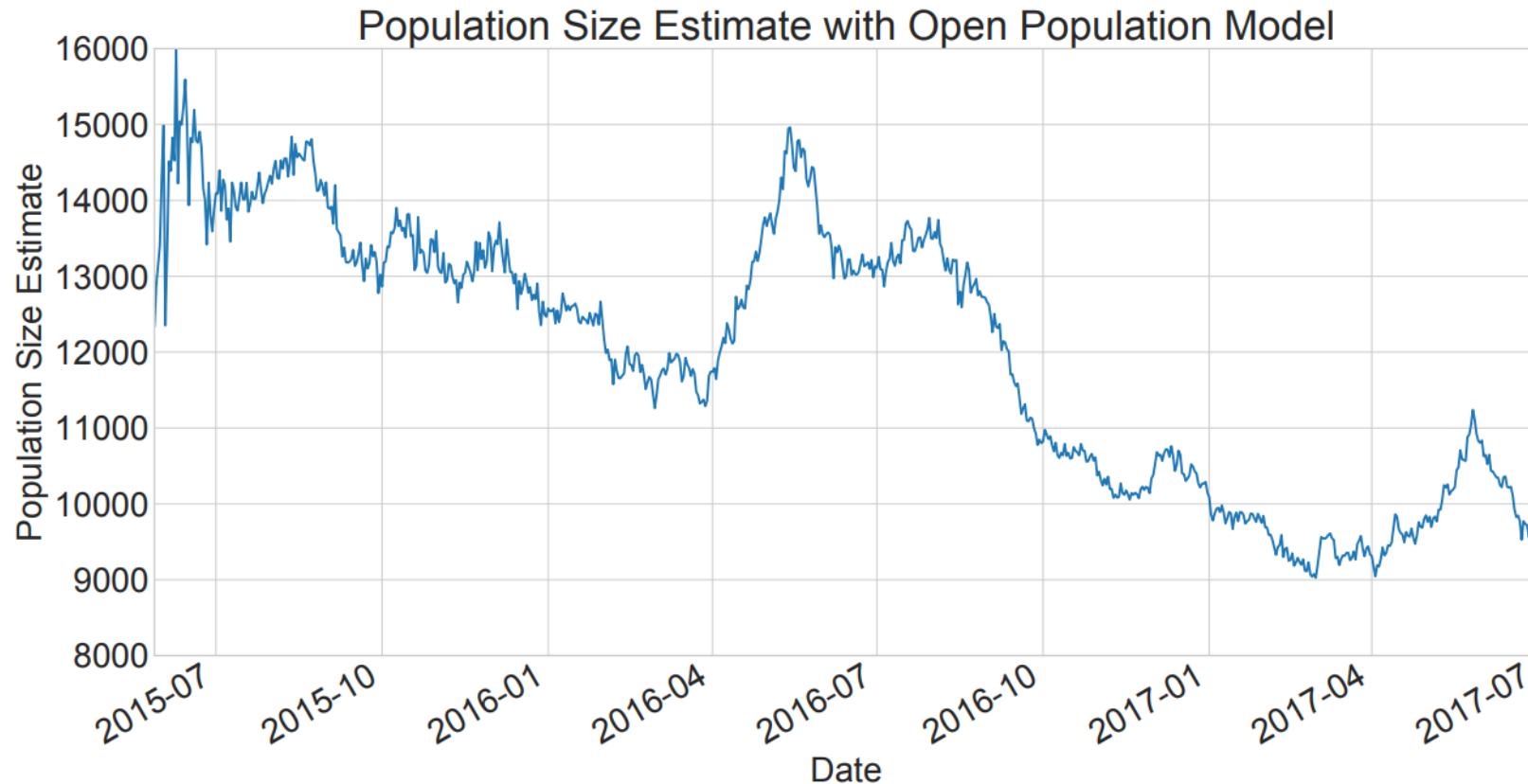
## Relaxing A1: Open population model

- Survival probability:  $S(t) = \exp(-\lambda t)$
- At time  $d - t$ ,  $N_{d-t}$  workers participated in the survey
- At time  $d$ , assume that there are  $M_d$  workers in total. Among  $N_d$  workers that participated in the survey,  $n_{d,t}$  workers also participated in the survey at time  $d - t$ .

$$n_{d,t} = \frac{N_{d-t} \exp(-\lambda t)}{M_d} \times N_d$$

$$M_d = \frac{N_{d-t} N_d \exp(-\lambda t)}{n_{d,t}}$$

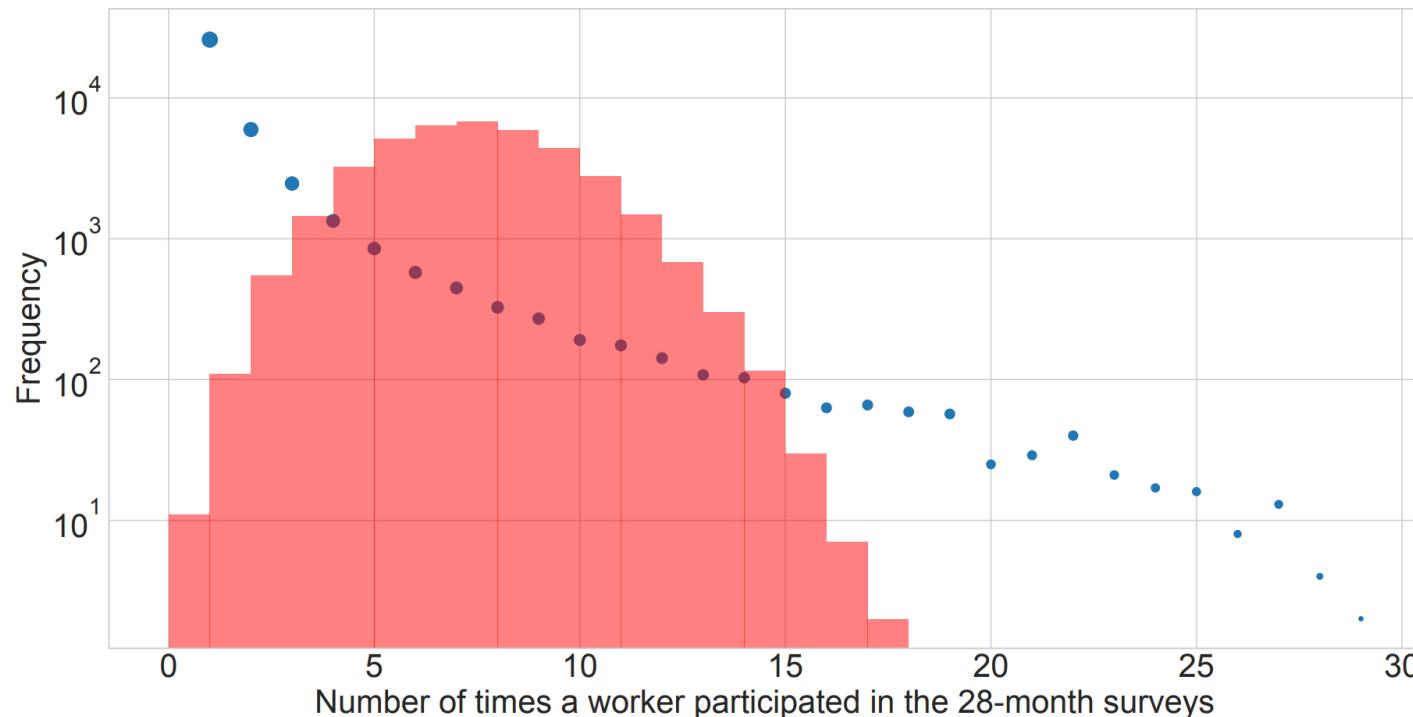
# How Large is the Crowd?



Half-life of the  
MTurk worker  
population is 404  
days

# How Large is the Crowd?

- Is “A2: Equal propensity of participation” reasonable?
  - If every worker participate the survey with the same probability, we should see a binomial distribution over # participations with many surveys.



# How Large is the Crowd?

## **Relaxing A2: Accounting for propensity of participating surveys**

- Assume there are  $N$  workers in total, each worker gets a propensity score  $a_i \sim p(a)$
- Capture  $n_1$  workers,  $P(\text{capture}|a_i) = 1 - (1 - \frac{a_i}{\sum_{j=1}^N a_j})^{n_1} \approx \frac{n_1 a_i}{N E[a]}$
- Recapture  $n_2$  workers,  $P(\text{capture, recapture}|a_i) = \frac{n_1 n_2 a_i^2}{N^2 E[a]^2}$
- The expected number of workers participating in both surveys:

$$m = N \int \frac{n_1 n_2 a_i^2}{N^2 E[a]^2} p(a) da = \frac{n_1 n_2}{N} \left(1 + \frac{\text{Var}[a]}{E[a]^2}\right)$$

- Following this method, the estimate for the number of MTurk workers is [178,000](#).

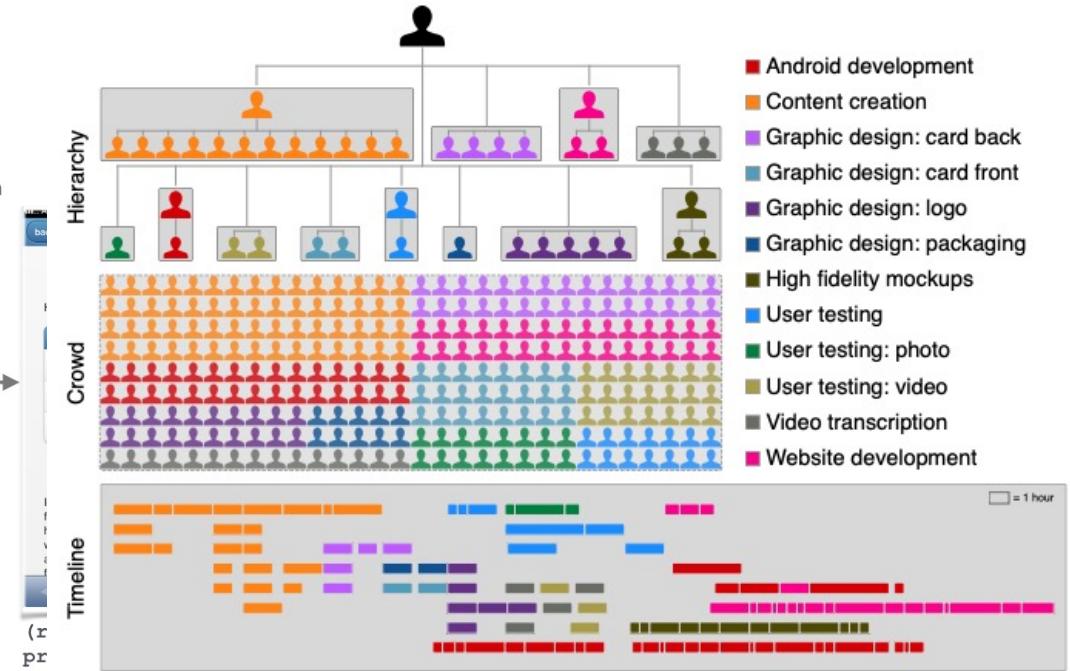
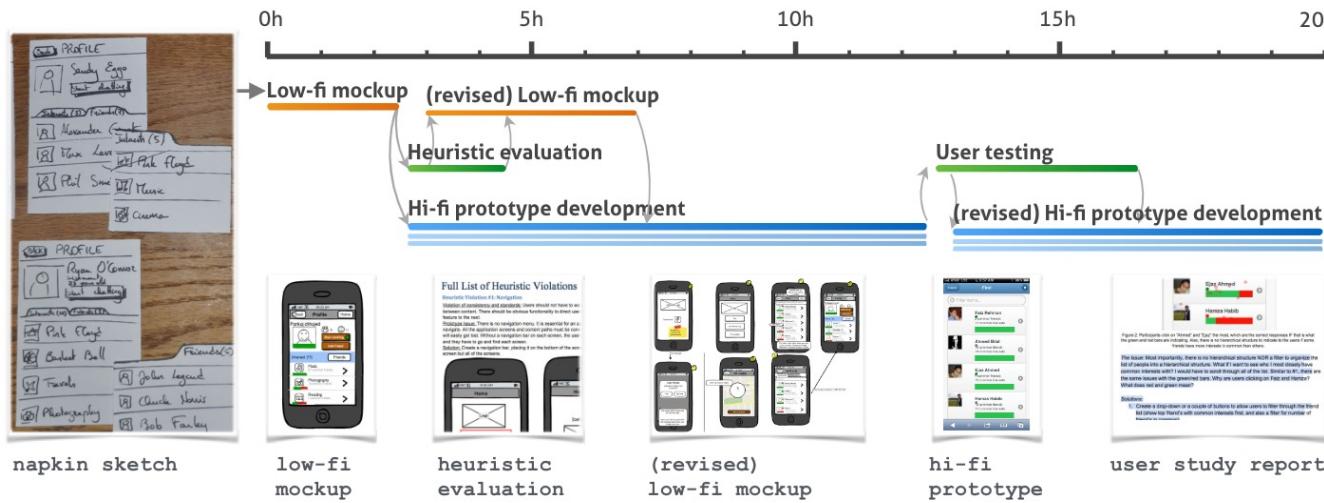
# Discussion

- What additional information do you want to know about Turkers?
- Can you come up with methods to collect that information?
- Does that information help us design better crowdsourcing platforms? How?

# Future of Work

# From Microtask to Expert Crowdsourcing

- Teamwork and Expert Crowdsourcing

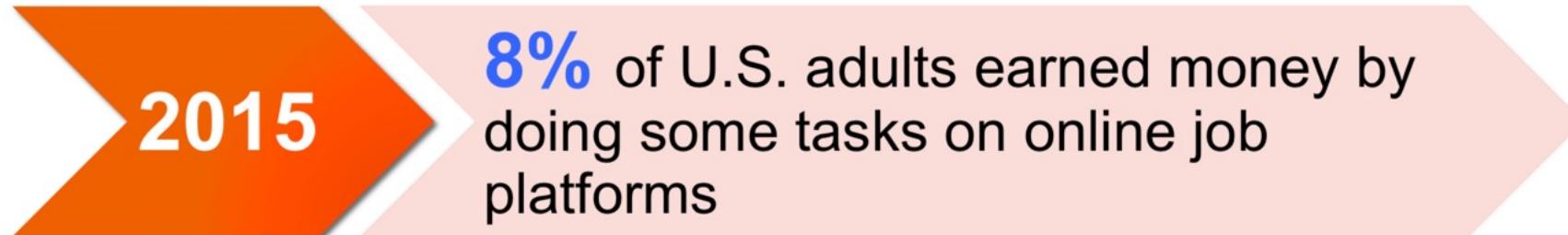


Blurring the Line of Traditional/Micro Work

# More Than Crowdsourcing

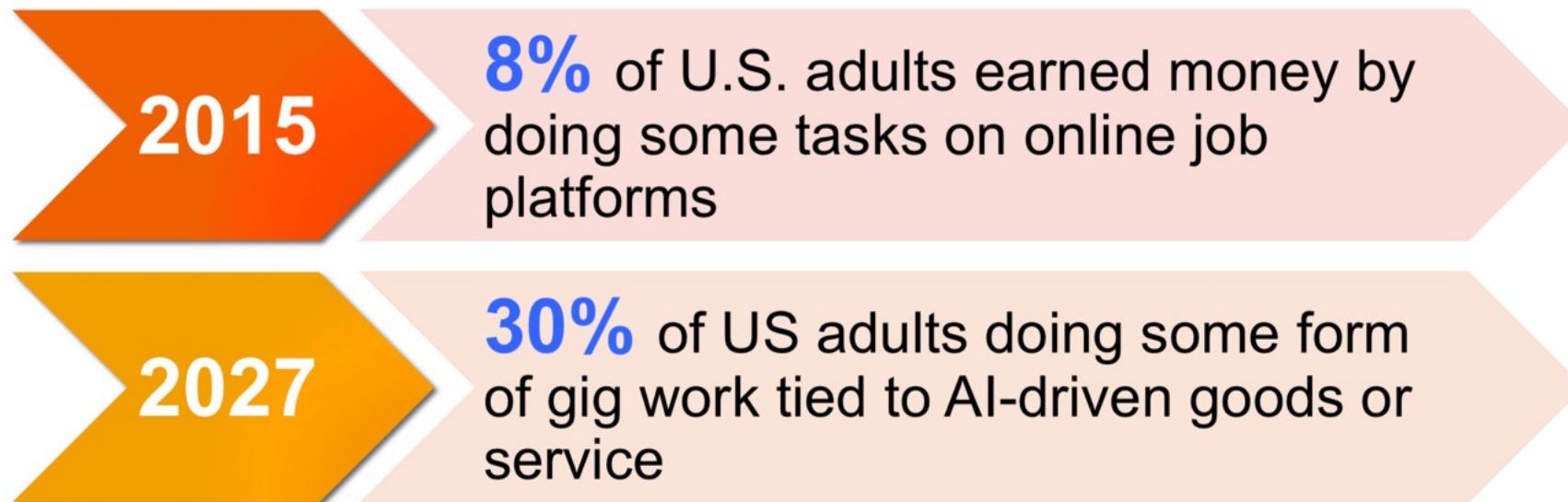


# More Than Crowdsourcing: Gig-Economy



Gig Work, Online Selling and Home Sharing. Pew Research Center, November 2016  
Spike in Online Gig Work: Flash in the Pan or Future of Employment? Social Media Collective, November 2016

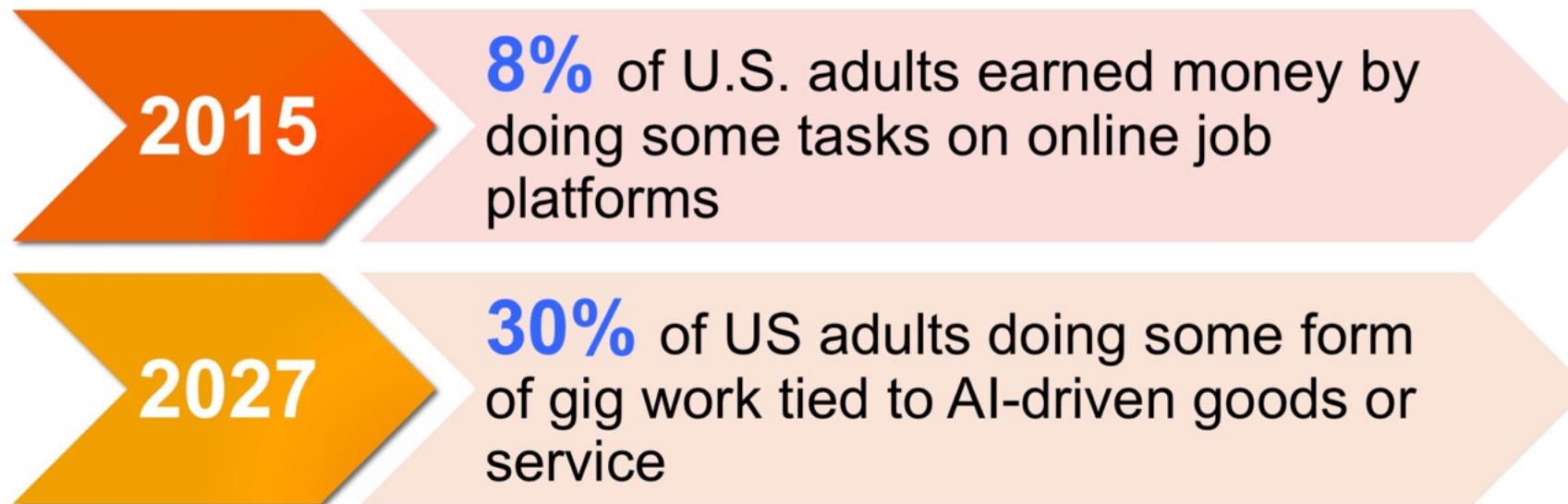
# More Than Crowdsourcing: Gig-Economy



Gig Work, Online Selling and Home Sharing. Pew Research Center, November 2016

Spike in Online Gig Work: Flash in the Pan or Future of Employment? Social Media Collective, November 2016

# More Than Crowdsourcing: Gig-Economy



**Prop 22 passes in California, exempting Uber and Lyft from classifying drivers as employees**



By [Sara Ashley O'Brien, CNN Business](#)

Updated 4:02 PM ET, Wed November 4, 2020

**Judge rules California Prop 22 gig workers law is unconstitutional**

*Voters approved the law in November*

By [Kim Lyons | @SocialKimLy](#) | Aug 21, 2021, 9:36am EDT

Even the traditional work  
environment is changing

# Dropbox is making its workforce 'virtual first.' Here's what that means

By [Kathryn Vasel](#),

Updated 4:59 PM

## Microsoft is letting more employees work from home permanently

*Microsoft employees will also be able to relocate*

By [Tom Warren](#) | [@tomwarren](#) | Oct 9, 2020, 7:32am EDT

## Google's plan to cut pay for remote workers who relocate is a bad idea

Many workers are paid based on where they live. That's changing.

By [Rani Molla](#) | [@ranimolla](#) | Oct 4, 2021, 8:00am EDT

Business / Tech

## IBM tells managers to come to the office or leave their jobs

By [Jennifer Korn](#), CNN

① 3 minute read · Updated 3:05 PM EST, Tue January 30, 2024

BUSINESS

## The Companies Calling Workers Back to the Office Five Days a Week

UPS, Boeing and other employers insist on full-time attendance as some bosses lose patience with remote work

By [Chip Cutter](#) [Follow](#)

Jan. 30, 2024 9:00 pm ET

## Traditional



*Centralized*



*Career*

## On-demand



*Distributed*



*Piecework*



*Management Strategies*



*???*

# 7 Jobs That AI Will Replace by 2025



Samantha Kang Jan 19 · 5 min read ★



These jobs are under threat as robots and artificial intelligence takeover amid pandemic

title

ECONOMY

**ChatGPT may be coming for our jobs. Here are the 10 roles that AI is most likely to replace.**

[Jacob Zinkula](#) and [Aaron Mok](#) Updated Jan 15, 2024, 9:12 AM CST

Share | Save

# Future of Work

- One of NSF's 10 Big Ideas:
  - “In 2019, NSF will invest \$30 million in each Big Idea and continue to identify and support emerging opportunities for U.S. leadership in Big Ideas that serve the Nation's future.”



## Future of Work at the Human-Technology Frontier

Catalyzing interdisciplinary science and engineering research to understand and build the human-technology relationship; design new technologies to augment human performance; illuminate the emerging socio-technological landscape; and foster lifelong and pervasive learning with technology. [Read more.](#)

# Discussion

- From workers' perspectives, what do you think are “wrong” for the current crowdsourcing platforms? How do you think we can fix it?
- To put the questions into context, consider the following question

*“Can we foresee a future crowd workplace in which we would want our children to participate?”*

- What do you think are the most pressing questions related to the theme of “future of work”?

# Improving Worker Welfare

- Discussion among researchers [Kittur et. CSCW 2013]
  - Create career ladders
    - Motivation, job design, reputation, hierarchy
  - Improve task design through better communications
    - Quality assurance, job design, task assignment, real-time crowd work, synchronous collaboration, platforms
  - Facilitate learning
    - Reputation and credentials, AIs guiding crowds, crowds guiding AIs, task assignment, quality assurance
- Various guidelines for conducting crowdsourcing
  - [Guidelines for Academic Requesters](#)
  - [Responsible Research with Crowds: Pay Crowdworkers at Least Minimum Wage](#)