

Crowdsourcing and Human Computation

Instructor: Chris Callison-Burch

Website: crowdsourcing-class.org

What will we cover in
this class (and should
you take it)?

Syllabus

- Taxonomy of crowdsourcing and human computation
- The Mechanical Turk crowdsourcing platform
- Programming concepts for human computation
- The economics of crowdsourcing
- Crowdsourcing and machine learning
- Applications to human computer interaction
- Crowdsourcing and social science

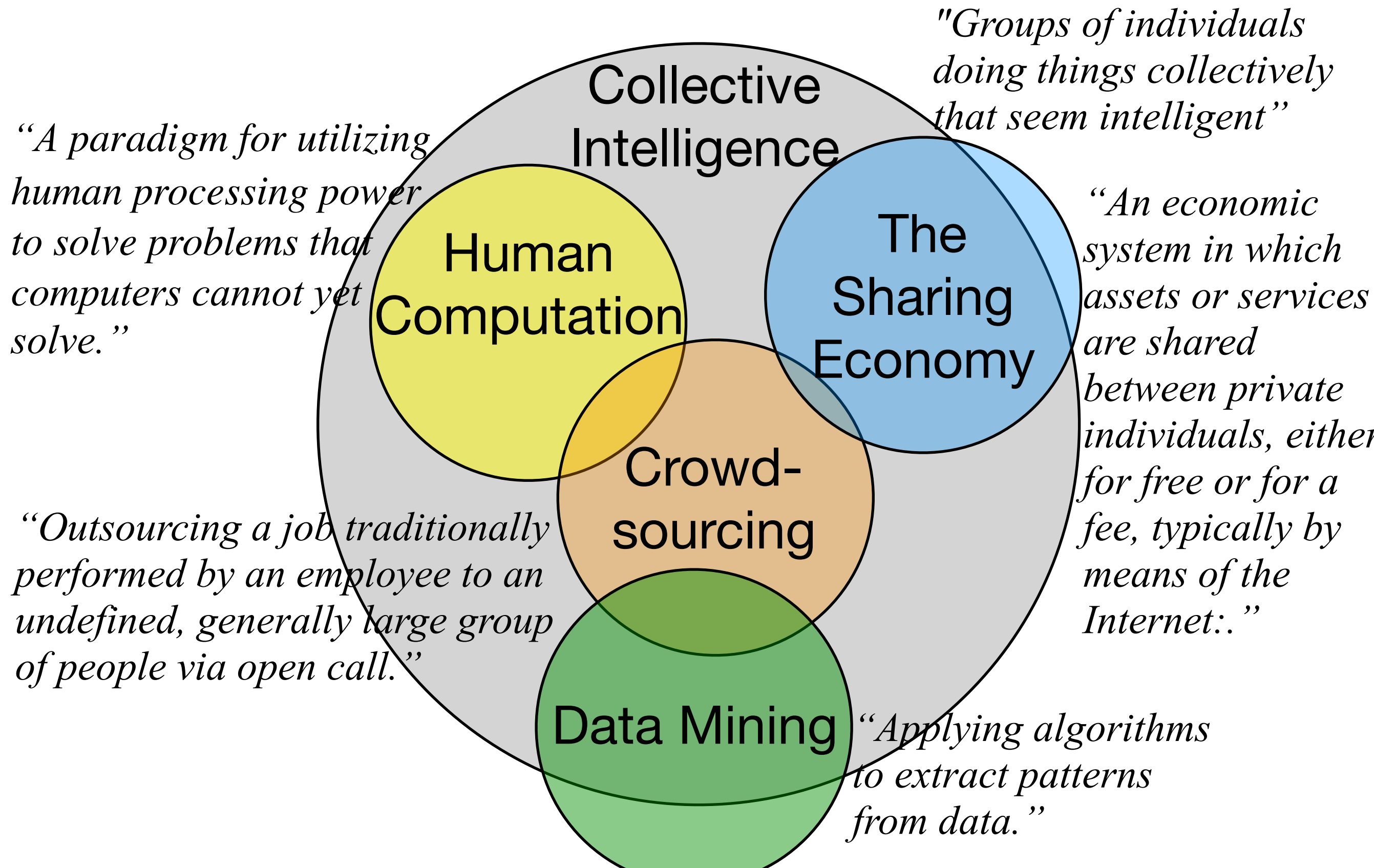
Who should take this class

- Anyone who wants to be on the cutting edge of this new field
- Entrepreneurial students who want to start their own companies
- Students from the business school who want to experiment with markets
- Students from the social sciences who want to conduct large-scale studies with people

What will you get out of this class?

- Understanding of an emerging field of CS
- Basic python and machine learning skills
- Ideas that you could transform into a startup company or academic research
- A new way of thinking about collective decision making companies and countries

Inter-related concepts



Crowdsourcing Companies

“Outsourcing a job traditionally performed by an employee to an undefined, generally large group of people via open call.”



Crowdsourcing Companies

“Outsourcing a job traditionally performed by an employee to an undefined, generally large group of people via open call.”



Eyewire

UBER



KICKSTARTER



qualtrics



POSTMATES



Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

37,649 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an
interesting task



Work



Earn
money



[Find HITs Now](#)

or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your
account



Load your
tasks



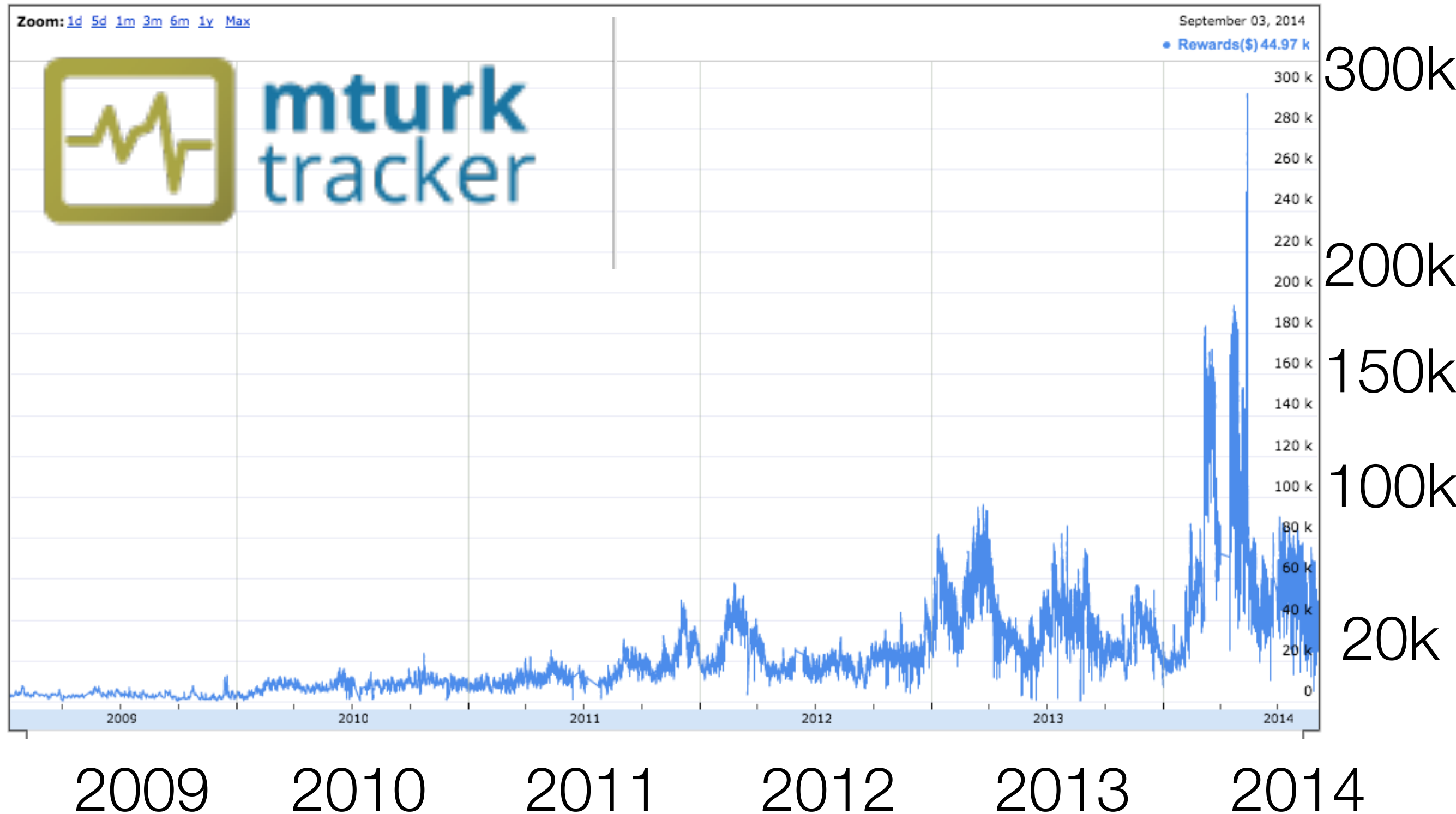
Get
results



[Get Started](#)

or [learn more about being a Requester](#)

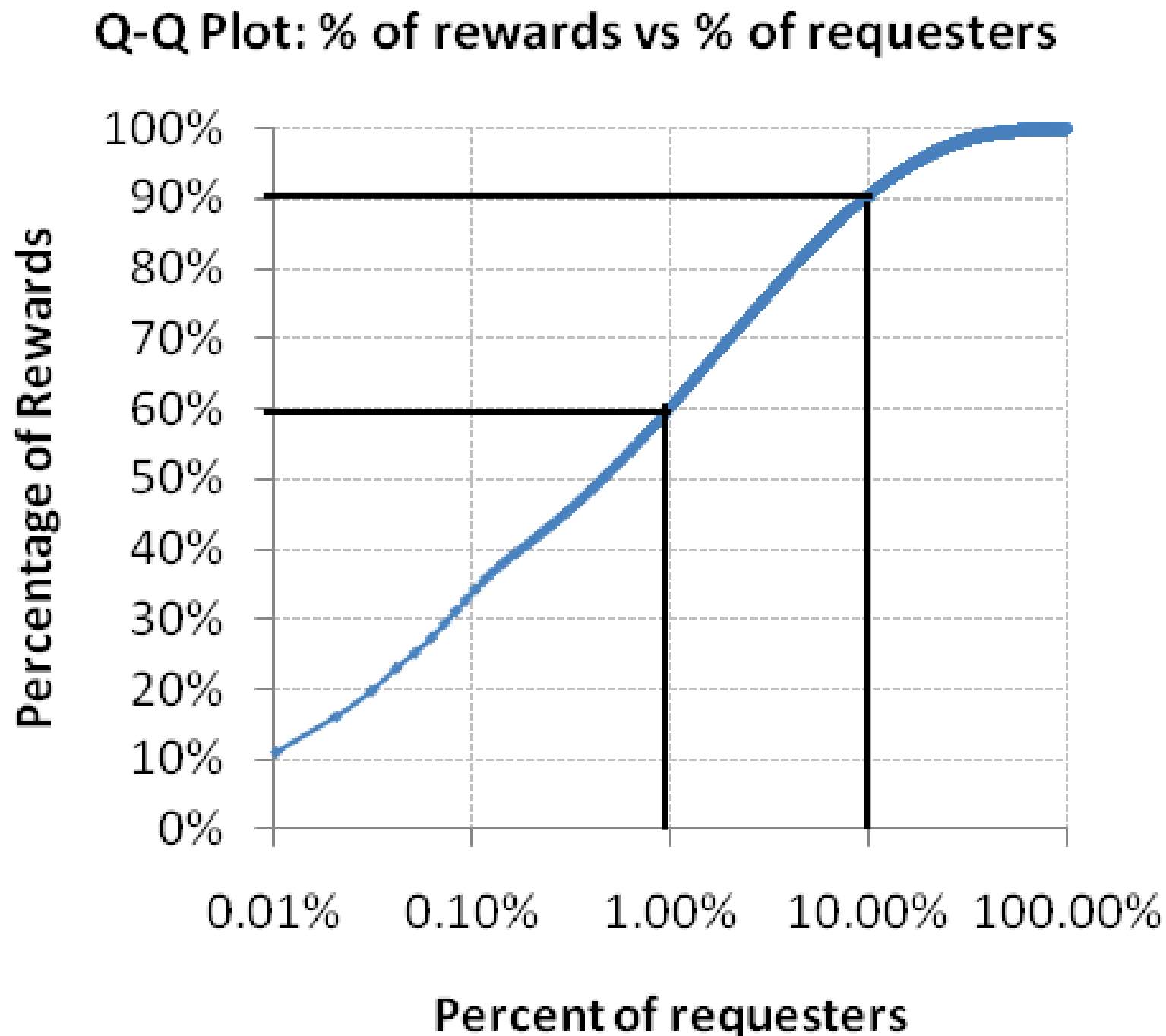
Rewards over past 5 years



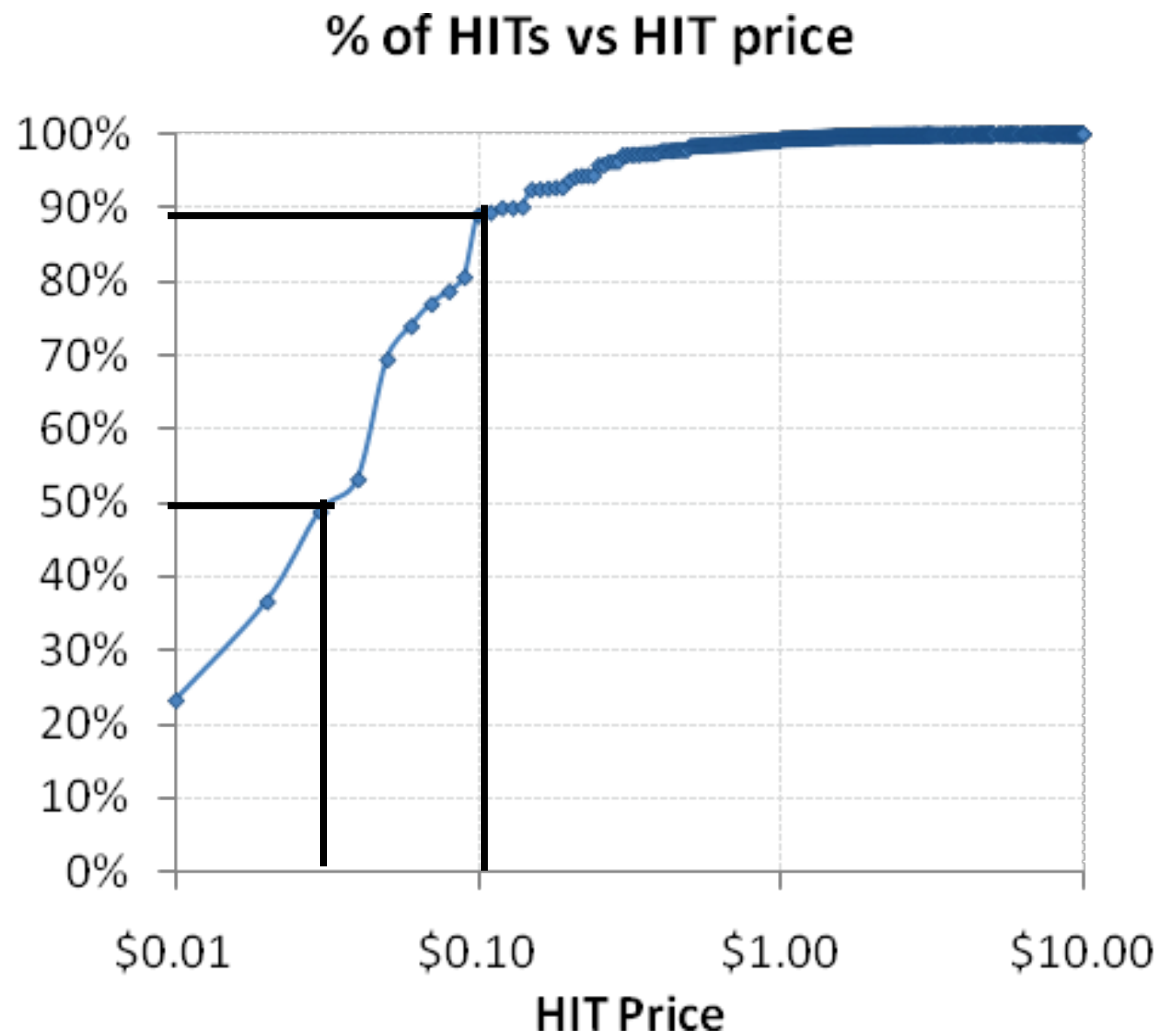
Top Requesters

Requester ID	Requester Name	#HIT groups	Total HITs	Rewards	Type of tasks
A3MI6MIUNWCR7F	CastingWords	48,934	73,621	\$59,099	Transcription
A2IR7ETVOIULZU	Dolores Labs	1,676	320,543	\$26,919	Mediator for other requesters
A2XL3J4NH6JI12	ContentGalore	1,150	23,728	\$19,375	Content generation
A1197OGL0WOQ3G	Smartsheet.com Clients	1,407	181,620	\$17,086	Mediator for other requesters
AGW2H4I480ZX1	Paul Pullen	6,842	161,535	\$11,186	Content rewriting
A1CTI3ZAWTR5AZ	Classify This	228	484,369	\$9,685	Object classification
A1AQ7EJ5P7ME65	Dave	2,249	7,059	\$6,448	Transcription
AD7C0BZNKYGYV	QuestionSwami	798	10,980	\$2,867	Content generation and evaluation
AD14NALRDOSN9	retaildata	113	158,206	\$2,118	Object classification
A2RFHBFTZHX7UN	ContentSpooling.net	555	622	\$987	Content generation and evaluation
A1DEBE1WPE6JFO	Joel Harvey	707	707	\$899	Transcription
A29XDCTJMAE5RU	Raphael Mudge	748	2,358	\$548	Website feedback

A few requesters offer most of the rewards



HITs by price



All HITs | HITs Available To You | HITs Assigned To You

Find HITs that pay at least \$ 0.00

HITs for
1-10 of

Sort by: ✓ Reward Amount (most first)

- HIT Creation Date (oldest first)
- HIT Creation Date (newest first)
- HITs Available (fewest first)
- HITs Available (most first)
- Reward Amount (least first)
- ✓ Reward Amount (most first)
- Expiration Date (soonest first)
- Expiration Date (latest first)
- Title (A-Z)
- Title (Z-A)
- Time Allotted (least first)
- Time Allotted (most first)

GO!

[Show all details](#) | [Hide all details](#)[Transcription of approximately 15 minutes of audio](#) [understand, would be useful to know how to clean up aud](#)

Requester: [Amelia Jones](#) **HIT Expiration Date:** Sep 4, 2013 (1 day 21 hours) **Reward:**
Time Allotted: 24 hours **HITs Available:**

[Transcription of approximately 15 minutes of audio](#)

Requester: [Amelia Jones](#) **HIT Expiration Date:** Sep 5, 2013 (2 days 23 hours) **Reward:**
Time Allotted: 8 hours **HITs Available:**

[By Invitation Only: Answer a few questions in a brief survey](#)

Requester: [Qualtrics Survey](#) **HIT Expiration Date:** Sep 5, 2013 (2 days 21 hours) **Reward:**
Time Allotted: 30 minutes **HITs Available:**

[Take a geo-tagged photo of a landmark in North Dakota](#)

Requester: [Crowdsourcing at Thomson Reuters](#) **HIT Expiration Date:** Sep 5, 2013 (2 days 16 hours) **Reward:**
Time Allotted: 3 days **HITs Available:**

Manage Batches

Click on the name of the batch to see more details

▼ **Batches in progress (1)**

Compression HIT - grammar/meaning 10 Results Cancel th

Created:	July 04, 2013	Assignments Completed:	2,468 / 2,468
Time Elapsed:	1 day	Estimated Completion Time:	COMPLETE
Average Time per Assignment:	3 minutes 40 seconds	Effective Hourly Rate:	\$4.091

Batch Progress:

100% submitted

100% published

▼ **Batches ready for review (143)**

« Previous 1 2 3 4 5 6 7 8 9 ...

Word Alignment - Trusted Workers - Dev 0.7 Results Delete

Created:	July 02, 2013	Assignments Completed:	1,995 / 2,000
Time Elapsed:	15 days	Estimated Completion Time:	September 03, 2013 4:41 PM (TODAY)
Average Time per Assignment:	1 minute 58 seconds	Effective Hourly Rate:	\$4.576

I tried one of his tasks to see, I gave it up at 4 minutes in and about 2/3 of the way through. For the whole hit, I'd have taken about 6 minutes. 10 hits an hour - **\$1.70 an hour**. Restricted to U.S. residents.

This is far too low to be considered a fair wage for a U.S. resident. My performance may be very far off from what others can do. Perhaps I took 4 times or more as long as an average worker would.

My complaint is that any U.S. requester knows what wage rate is required for a U.S. resident to survive. We may not agree on an exact number. But as they say, I know a fair wage when I see it, and this is not it.

Mturk is actually much smaller than what it can appear to be. **Something close to requester monopoly has the power to keep wages low.** Requester co-operation, explicit or implicit, reinforces this.

Chris Callison-Burch is not unaware, I think, of the mechanics of the wage structure of Mturk.

WORKERS
OF THE WORLD
UNITE!



Word Alignment


Requester:


▼ [Chris Callison-Burch](#)


HIT Expiration Date: Nov 12, 2013 (9 weeks)

Time Allotted: 60 minutes

communicativity:  4.05 / 5

generosity :  4.01 / 5

fairness :  4.25 / 5

promptness :  4.01 / 5

[What do these scores mean?](#)

Scores based on [95 reviews](#)

Terms of Service violation flags: 1

[Report your experience with this requester »](#)

[Contact Us](#) | [Careers at Amazon](#) | [Developers](#) |

©2005-2013 Amazon.com, Inc. or its Affilia

For those of you who know Arabic, this is a very solid requester with a very fair pay. Highly recommended for those who want to make some real money
Payment usually take one week depending on the HITs you are doing

don't waste your time trying to submit machine translated crap, or random answers to multiple choice questions, you will get blocked instantly

Jul 27 2013 | [hala...@h...](#) | [flag](#) | [comment](#)



Chris is one of the better requesters on MTurk, if you meet his qualifications and actually do the work as he requires. Glad to see that someone out there can finally work on those Arabic translation HITs that we've all seen for months now.

Jul 27 2013 [baudelai...@m...](#)

Good requester. Everything approved in a couple of days. I had no problems. This is a safe requester to work for.

qualitative v quantitative

TurkOpticon's qualitative attributes	CrowdWorker's quantitative equivalents
promptness: How promptly has this requester approved your work and paid?	Expected time to payment: On average, how much time elapses between submitting work to this Requester and receiving payment?
generosity: How well has this requester paid for the amount of time their HITs take?	Average hourly rate: What is the average hourly rate that other Turker make when they do this requester's HITs?
fairness: How fair has this requester been in approving or rejecting your work?	Approval/rejection rates: What percent of assignments does this Requester approve? What percent of first-time Workers get any work rejected?
communicativity: How responsive has this requester been to communications or concerns you have raised?	Reasons for rejection: Archive of all of the reasons for Workers being rejected or blocked by this Requester.

Ethics

- Fair pay for workers
- Legal implications of sharing economy
- Ethics of companies like Uber
- Guidelines for human subjects research

Classification System for Human Computation

- Motivation
- Quality Control
- Aggregation
- Human Skill
- Process Order
- Task-request Cardinality

Motivation

How can we motivate people to participate?

Even with a low barrier to entry (anyone with a computer can contribute) we still need to make a case **why** they should contribute.

- Pay
- Altruism
- Reputation
- Enjoyment
- Implicit work

Quality Control

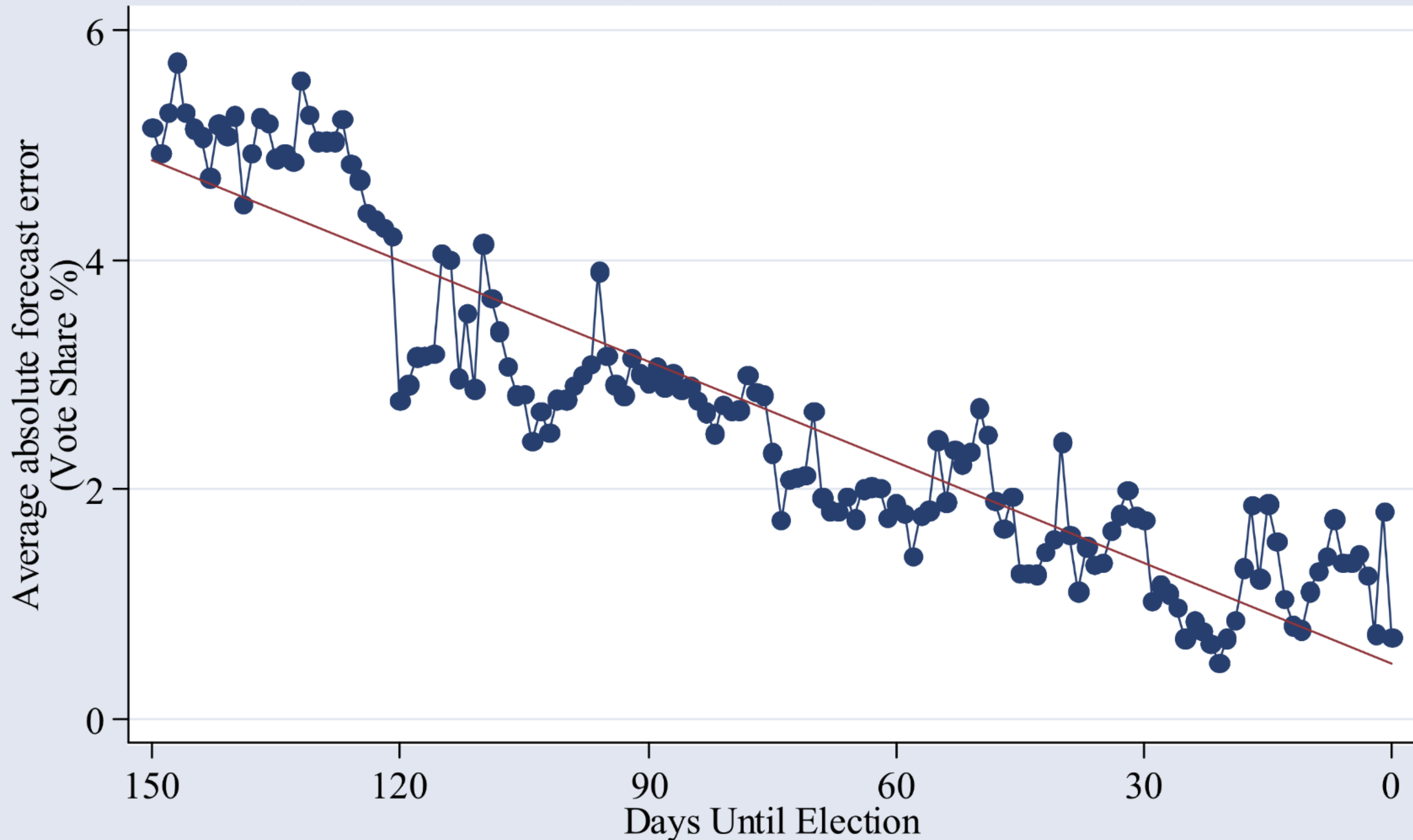
- Reputation systems
- Redundancy and agreement
- Gold standards
- 2nd pass reviewing
- Statistical models
- Defensive task design
- Economic incentives

Aggregation

- Wisdom of Crowds
 - Voting
 - Prediction markets
- Collection
- Search
- Iterative improvement
- Machine learning

Iowa Electronic Markets: Predictive Accuracy Through Time

Average absolute error in predicting two-party vote shares, 1988-2000



Source: Author's calculations based on data available at: www.biz.uiowa.edu/iem/

Human skill

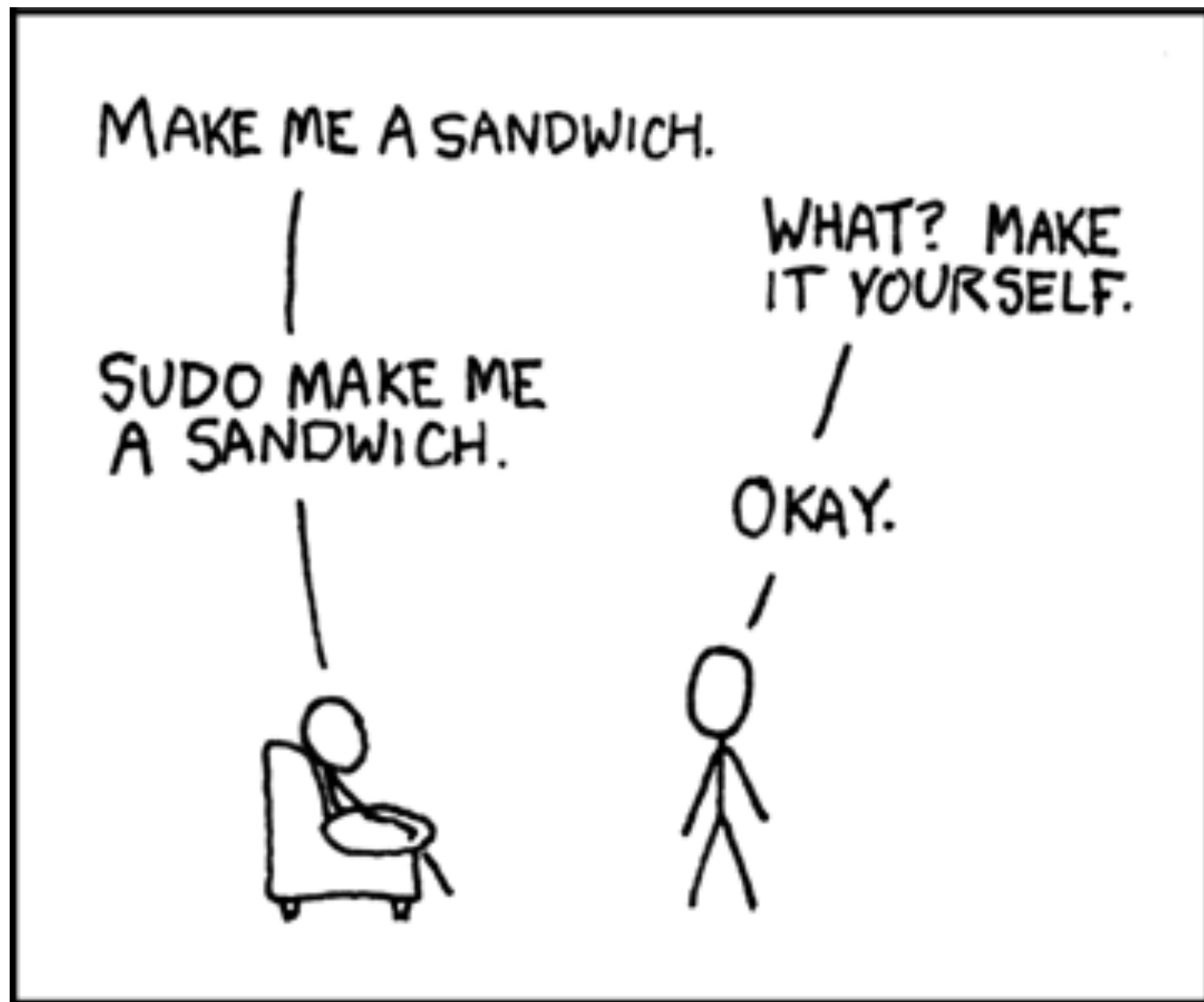
- Visual recognition
- Language understanding
- Translation
- Reasoning
- Creativity

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mice.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mice had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that "It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mice.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mice had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that "It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mice.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mice had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that "It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

New Programming Languages Concepts



TurKit: A programming language for the crowd

```
ideas = []
for (var i = 0; i < 5; i++) {
  idea = mturk.prompt(
    "What's fun to see in New York City? Ideas so
    far: " + ideas.join(", "))
  ideas.push(idea)
}
ideas.sort(function (a, b) {
  v = mturk.vote("Which is better?", [a, b])
  return v == a ? -1 : 1
})
```

New Programming Languages Concepts

- Latency
- Cost
- Parallelization
- Non-determinism
- Iterative improvement

New keyword **once**

- Costly operations can be marked in a TurKit program with keyword **once**
- **once** denotes that an operation should only be executed once across all runs of a program

Quicksort on MTurk

```
compare(a, b)
```

```
  hitId ← once createHIT(...a...b...)
```

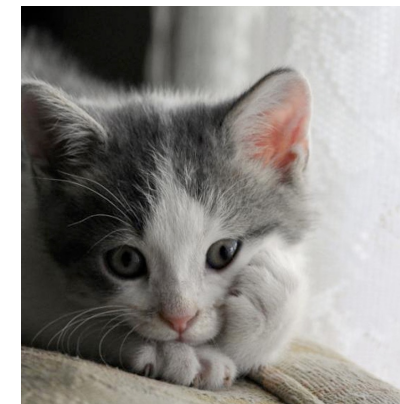
```
  result ← once getHITResult(hitId)
```

```
  return (result says  $a < b$ )
```

- Subsequent runs of the program will check the database before performing these operations

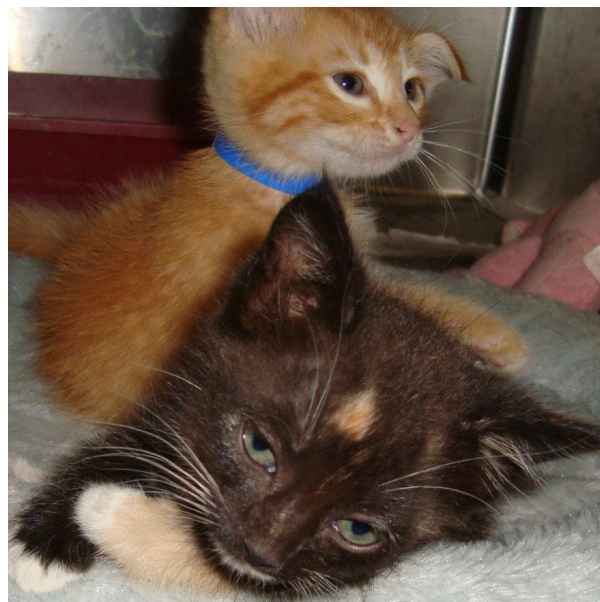
Quicksort for kittens

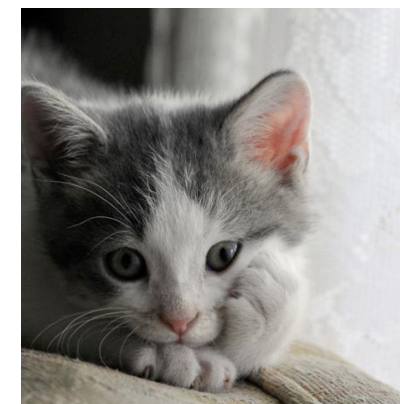




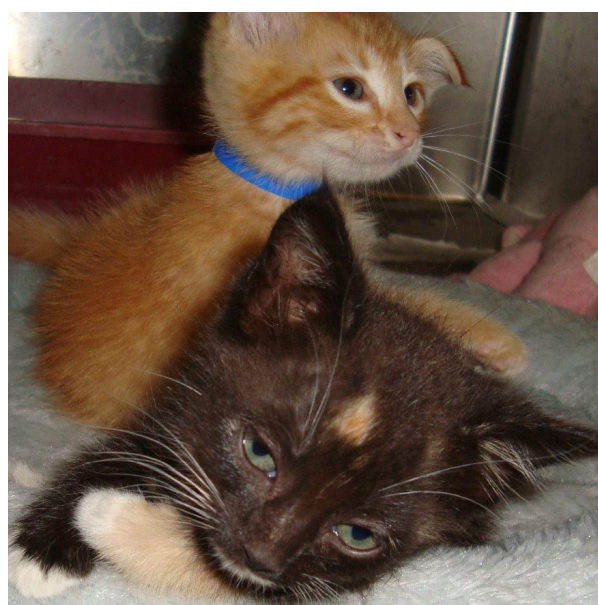


V



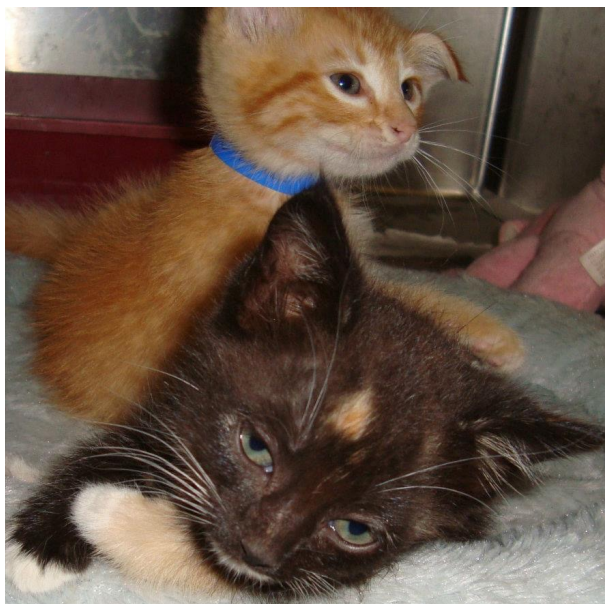


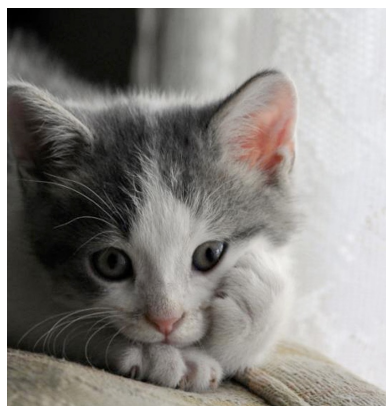
V



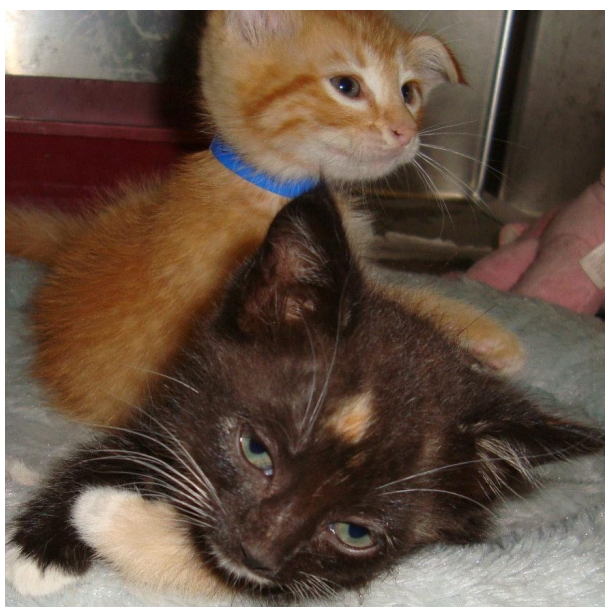


Λ



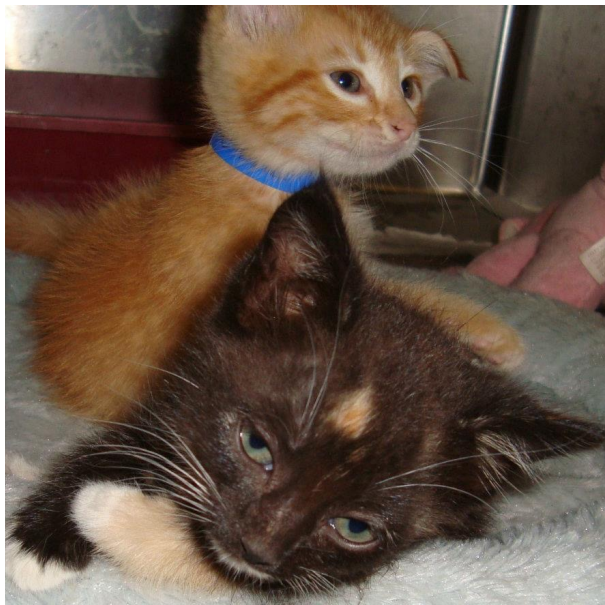


Λ



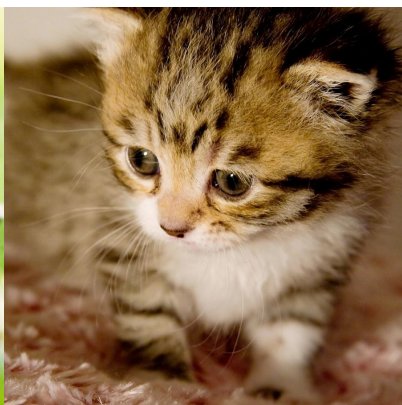


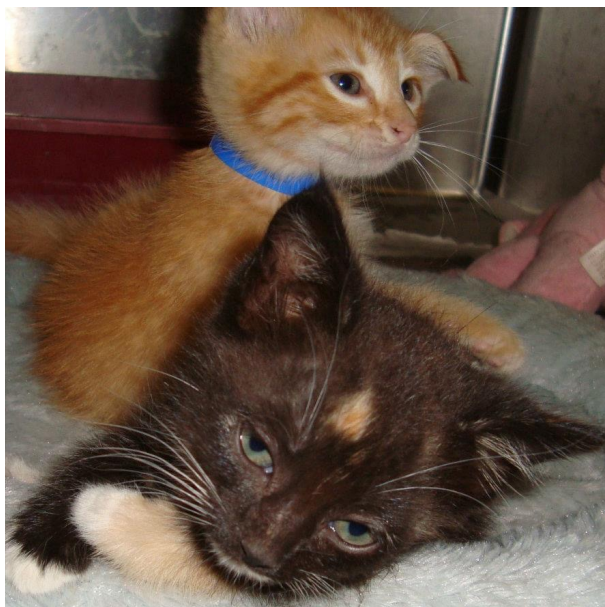
>





>











✓



✓





>





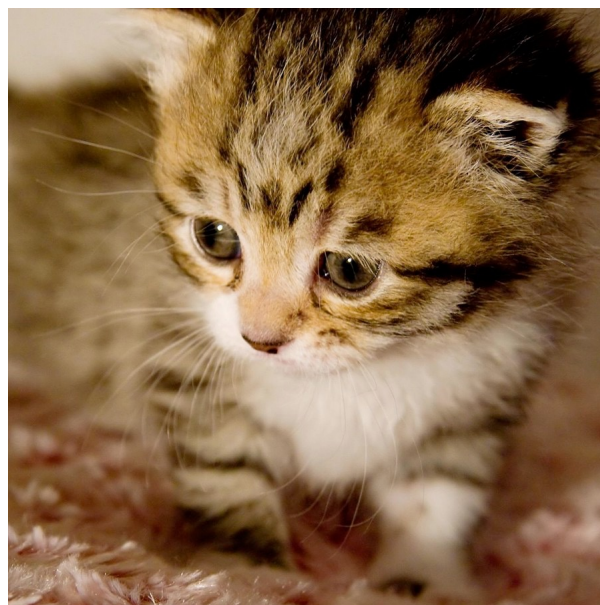
Λ













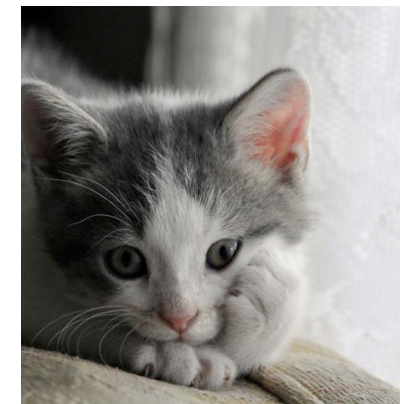
When should you mark a function with **once**?

- **High cost** - This is its main usage.
Whenever a fn is high-cost in terms of money or time, **once** saves the day

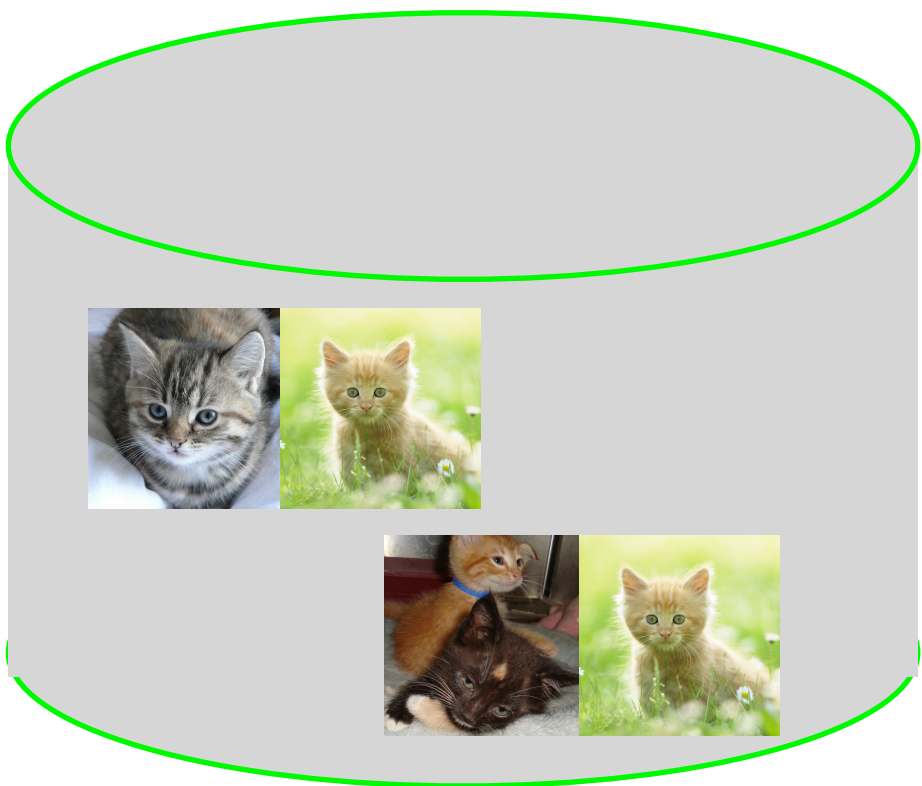
When should you mark a function with **once**?

- **Non-determinism** - storing results in DB assumes that the program executes in a deterministic way











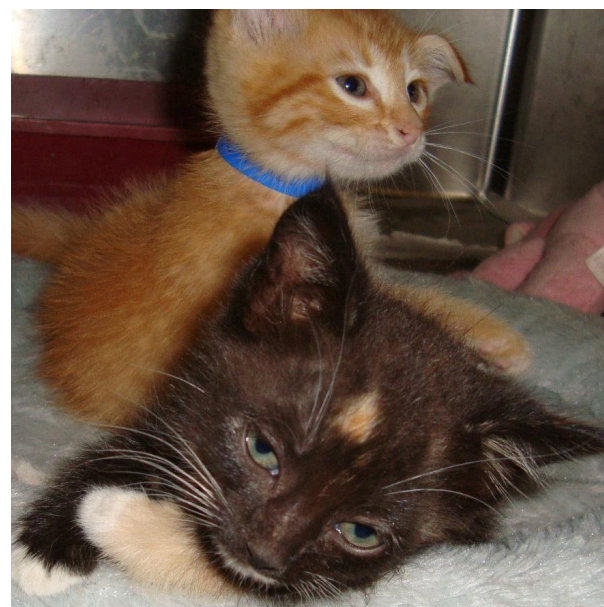
X



X



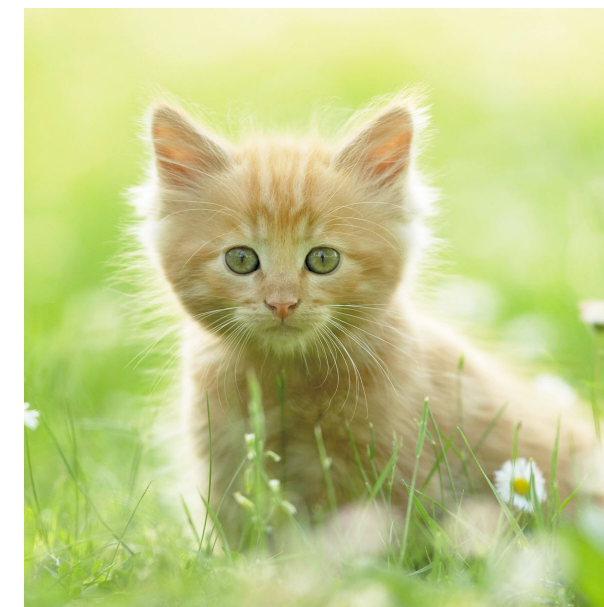
X



✓



X



Wizard of Oz in HCI



Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't important to the user's particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn't necessary to split base on those differences. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. **One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters.** Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.



Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't important to the user's particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn't necessary to split base on those differences. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. **One solution to this problem would be to let the user rearrange the clustering manually, using drag-and-drop edits.** Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.

Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't important to the user's particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn't necessary to split base on those differences.

Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.



Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't relevant to a specific task. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.

Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't important to the user's particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn't necessary to split base on those differences.

Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.



Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, as structure differences aren't important to the editing task. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.

Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't important to the user's particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn't necessary to split base on those differences.

Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.

Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, as structure differences aren't important to the editing task. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually using drag-and-drop edits. Clustering and selection generalization would also be improved by recognizing common test structure like URLs, filenames, email addresses, dates, times, etc.

The Human Macro

The Human Macro

Title

What do

Find Creative Commons figure for paragraph

Create Task for Every:

1 paragraph

Paragraph

Instructions (with Example)

Tell the w

I need a creative commons licensed image to describe
under Creative Commons.

Mechanical Turk Worker Preview

Advertisement

Find Creative Commons figure for paragraph

I need a creative commons licensed image to


Instructions

I need a creative commons licensed image to
under Creative Commons.

Here is the text:

When I first visited Yosemite State Park in California, the
rocks were big, the trees were big, the animals were big,
the granite mountain that looks like it was sheared off

Human Macro Examples

Request	“Pick out keywords from the paragraph like Yosemite, rock, half dome, park. Go to a site which has CC licensed images [...]”
Input	When I first visited Yosemite State Park in California, I was a boy. I was amazed by how big everything was [...]
Output	

VizWiz: Answers to Visual Questions for Blind Users

What denomination is this bill?



(24s) 20
(29s) 20

Do you see picnic tables across the parking lot?



(13s) no
(46s) no

What temperature is my oven set to?



(69s) it looks like 425 degrees but the image is difficult to see.
(84s) 400
(122s) 450

Can you please tell me what this can is?



(183s) chickpeas.
(514s) beans
(552s) Goya Beans

What kind of drink does this can hold?



(91s) Energy
(99s) no can in the picture
(247s) energy drink

Know when work is imminent

61 seconds	Start app, take picture
71 seconds	Record the question
78 seconds	Press send
221 seconds	Wait for response



Start
recruiting
workers

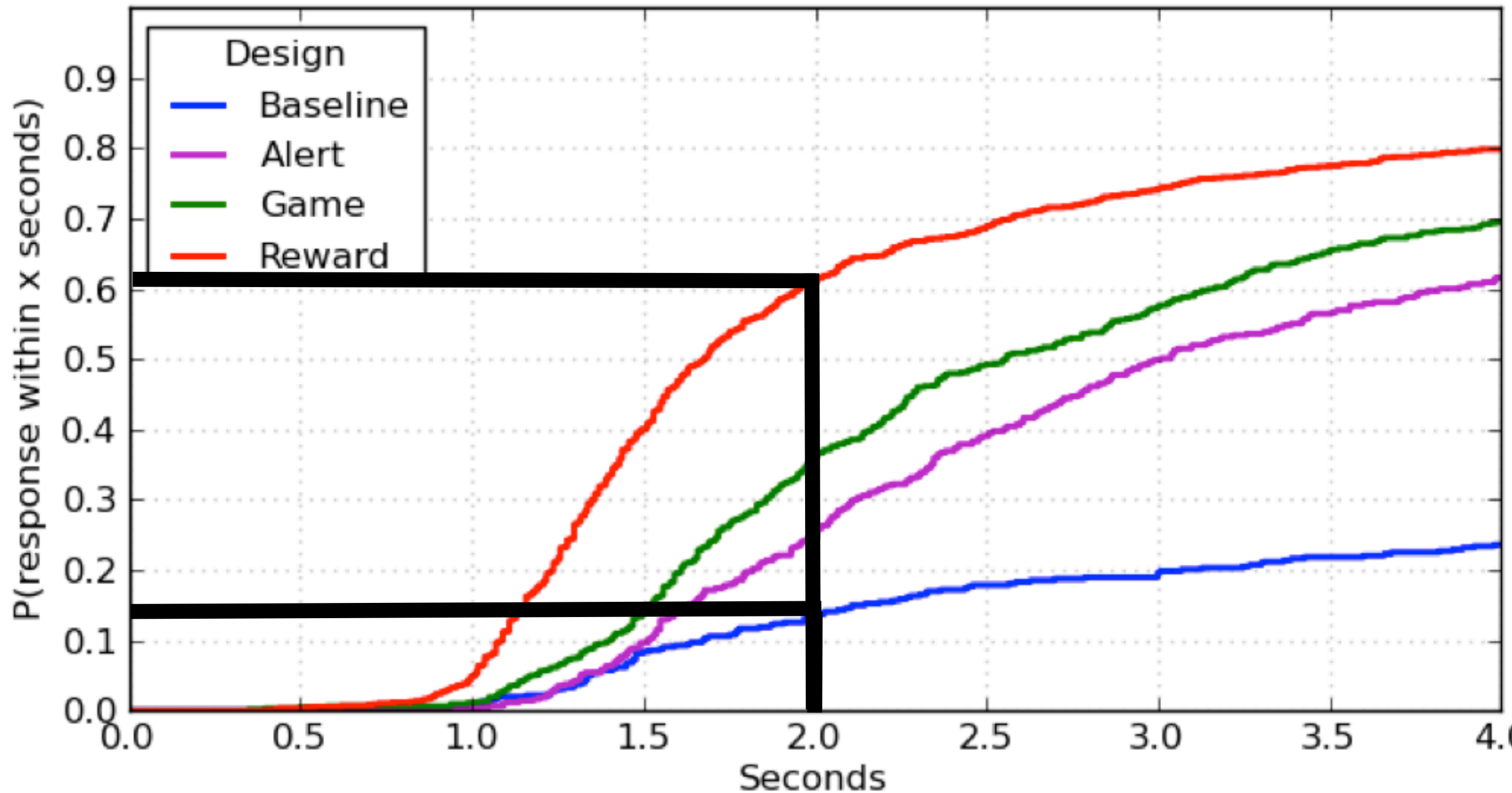
Maintain a work pool

- TurKit also experimented with maintaining a group of workers, even when there was no work
- Created dummy assignments from past assignments, to ensure work
- When a new request arrived a dummy was replaced with the real request
- Can be costly to constantly maintain a pool

Retainer model

- Alternate to maintaining worker pool with dummy tasks
- Hire crowd workers in advance, and pay them a small amount to wait for work to come online
- Allow them to pursue other work while waiting
- Alert them when our task is ready with a popup box, and pay them for that work too

Improving 10 minute retainer response time

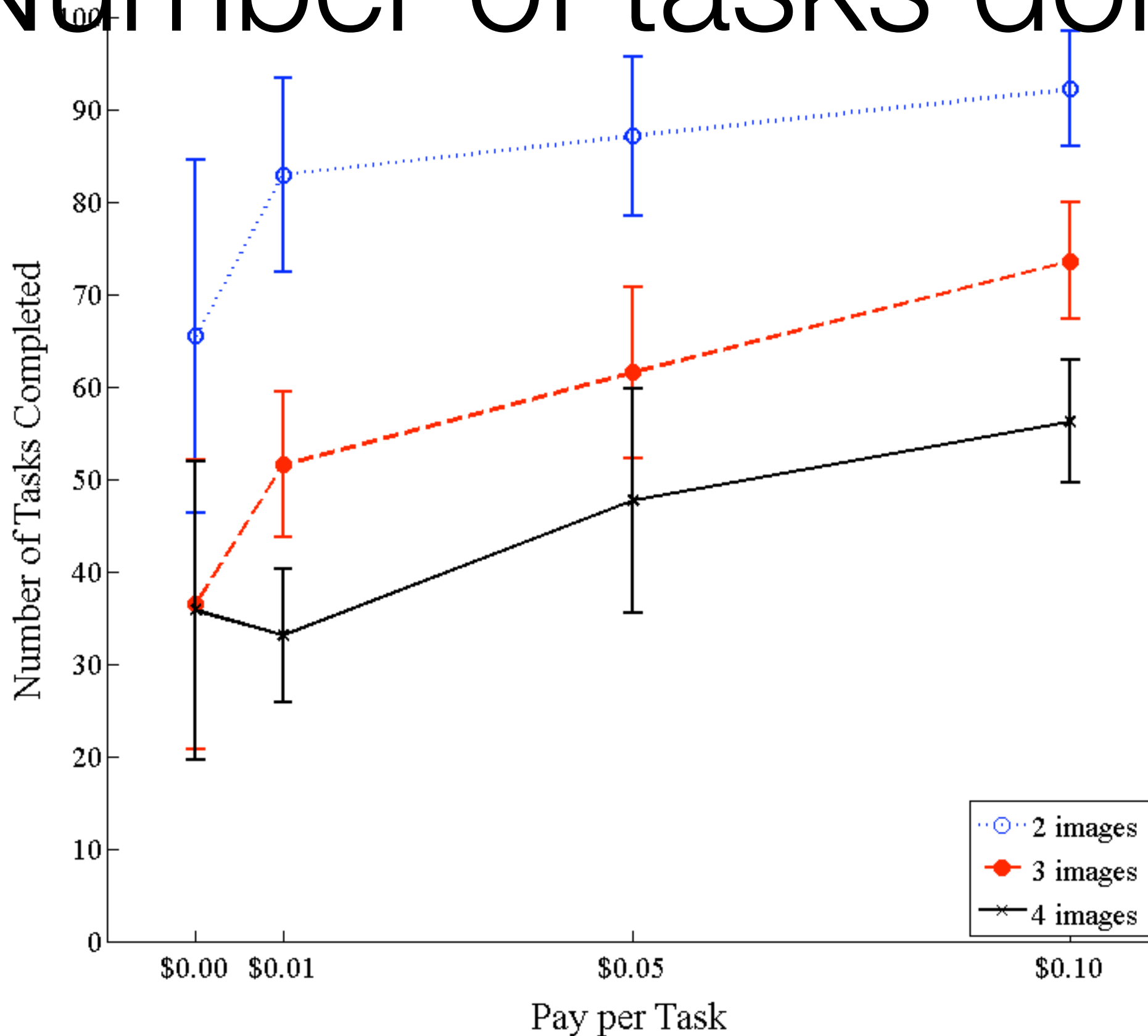


Studying Economic Markets

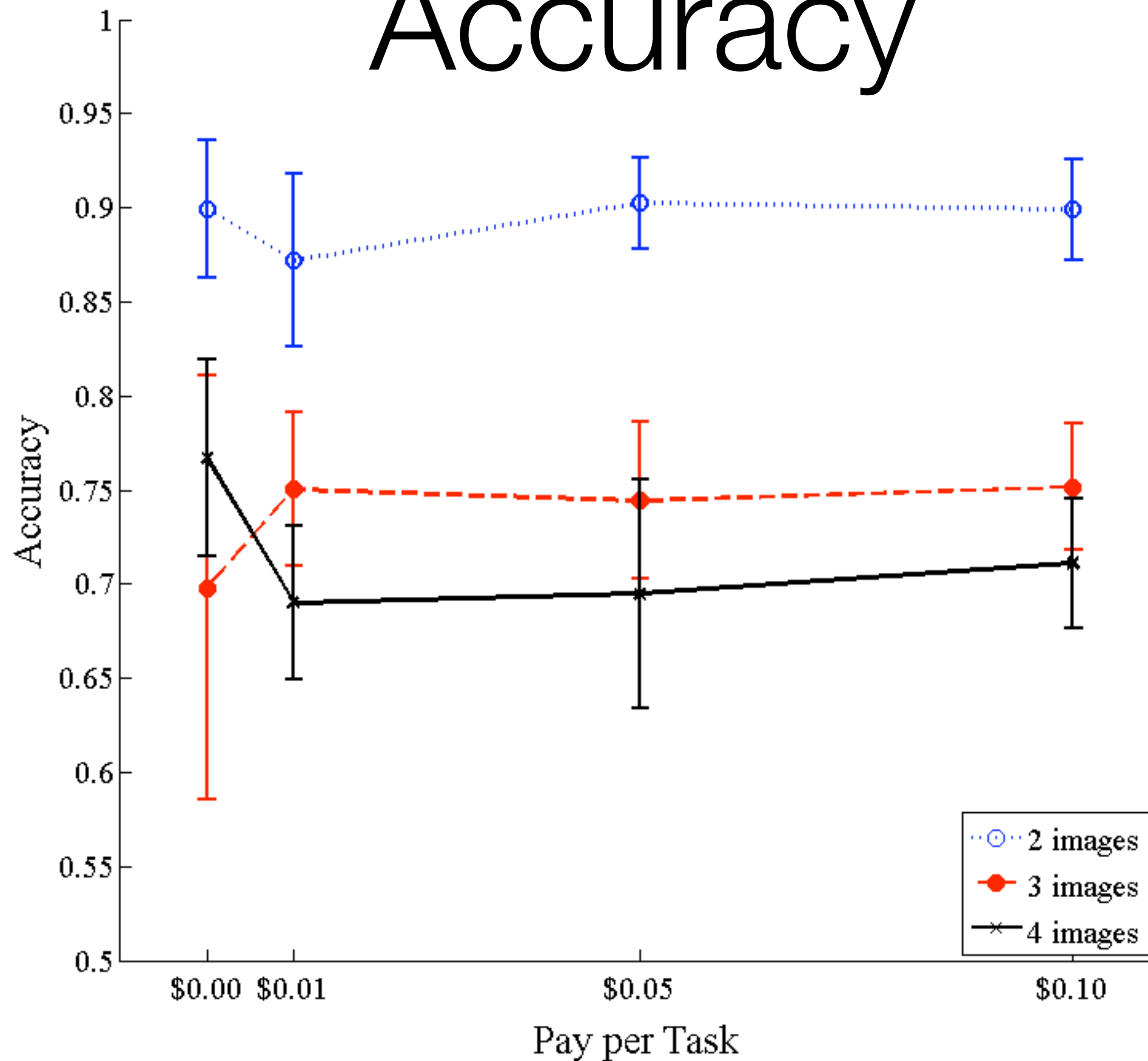
Financial Incentives and the “Performance of Crowds”

- Experiment with economic incentives on Amazon Mechanical Turk
- Does compensation change the quantity of work performed (output)?
- Does it change the quality of the work (accuracy)?

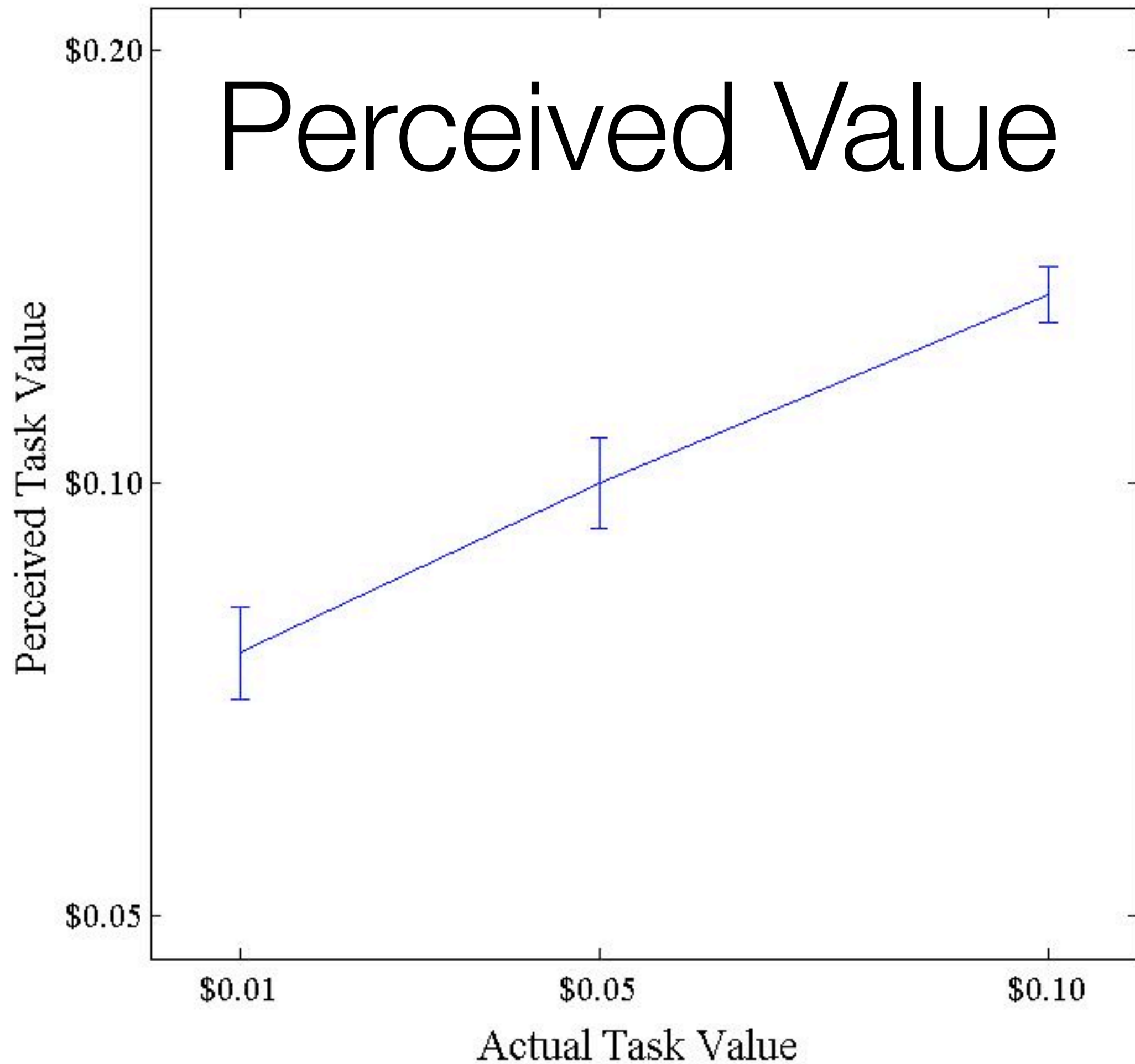
Number of tasks done



Accuracy



Perceived Value



MTurk for social science research

- Many social science experiments require recruitment of a large number of subjects
- MTurk contains the major elements required to conduct research:
 - A participant compensation system
 - A large pool of potential participants
 - A streamlined process for study design, participant recruitment, and data collection

How Do MTurk Samples Compare With Other Samples?

- MTurk population is more diverse than college students (or non-students who reside in college towns)
- Good gender splits
- Good minority representation
- Large number of non-US participants

Active versus Passive Crowdsourcing

- In the first half of the semester we mainly looked at *active* crowdsourcing, where we explicitly solicit help from the crowd
- Many applications of crowdsourcing rely on *passive* information collection from multitudes of individual

The Best Questions on a First Date

- You would like to learn about your date, some important things that you would like to know are awkward to ask directly
- Find questions that correlate with what you want to know, but which people are more free about answering publicly



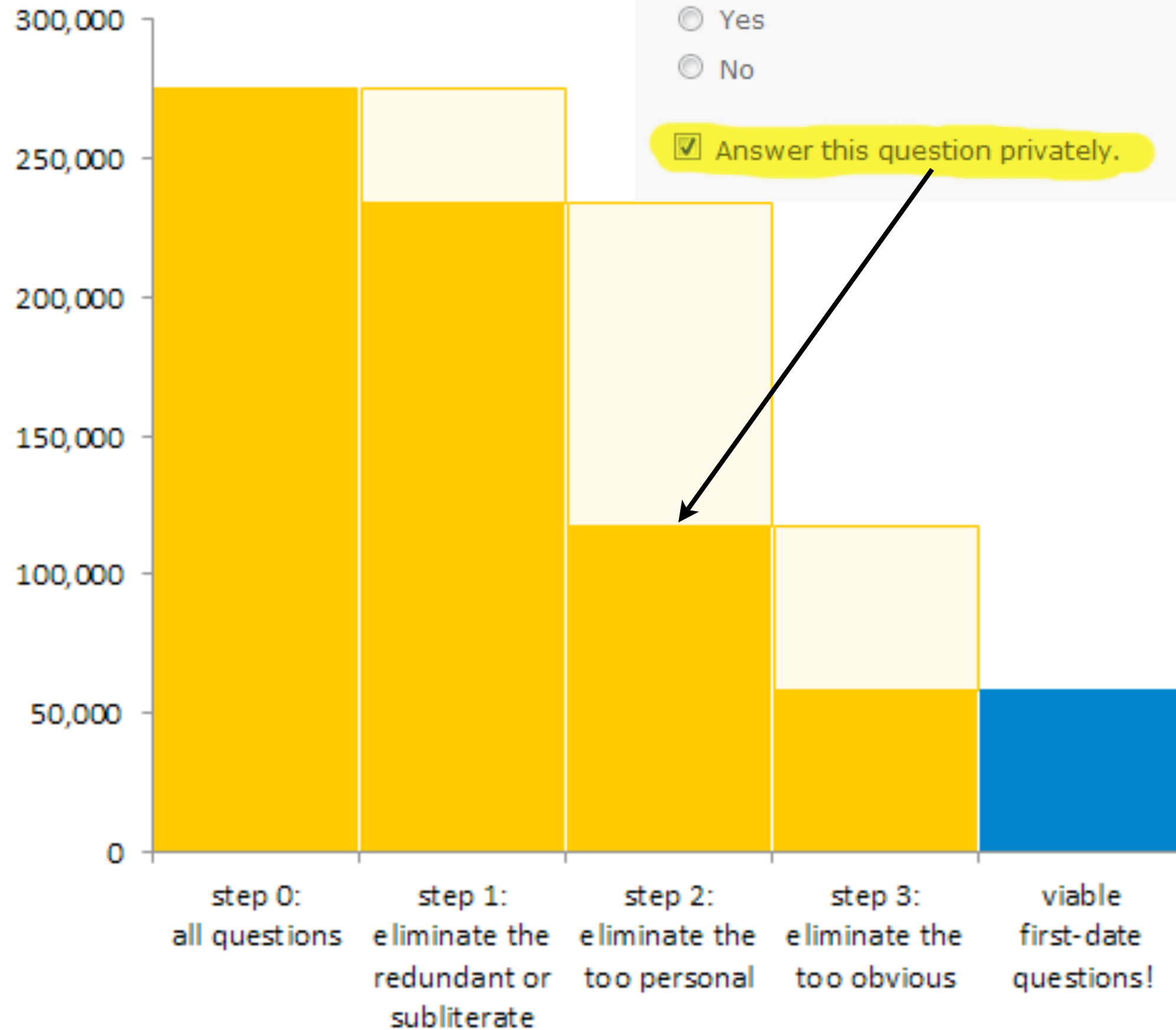
okcupid

Are you looking for a partner to have children with?

☐ Yes

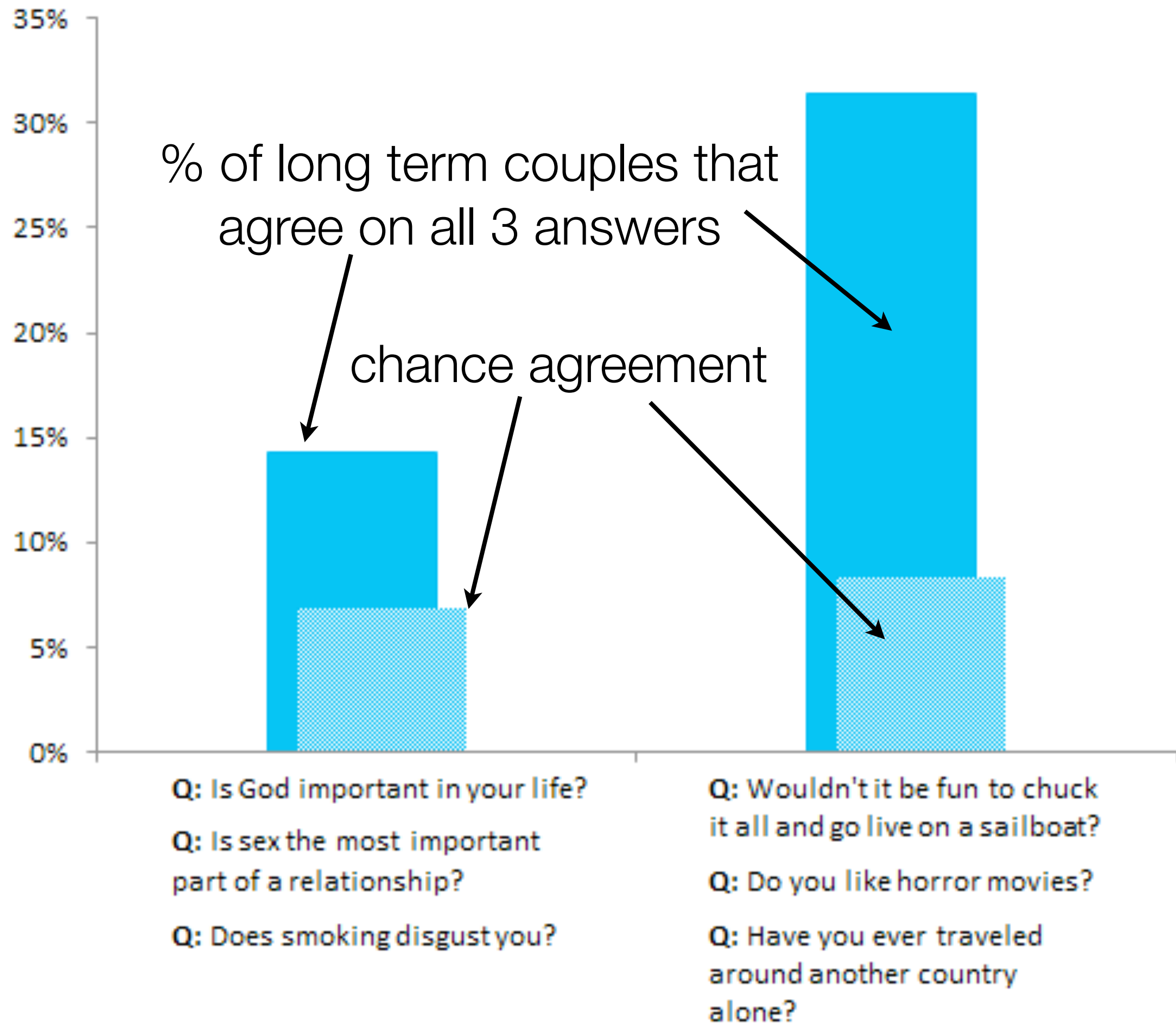
☐ No

☒ Answer this question privately.



top 3 user-rated match questions

our 3 recommended questions



What can you do with Crowdsourcing?

- Crowdsourcing is a transformative idea for business and research
- You all are exhibiting hugely creative thinking about it with your final projects
- I am looking forward to seeing what you come up with for the final, and beyond!

Final project details

- Wednesday, May 8th from noon-2pm in Wu and Chen Auditorium (Levine 101)
- 5-7 minutes video for each team, plus 2 minute Q&A
- You must provide links to your at least 1 hour before the presentations begin, and validate that they work.
- Final reports due on the 8th. Submit them before 9am.

Internship opportunities

- I am looking for 2-3 undergraduate researcher assistants to work with me on Crowdsourcing
- Paid summer internships in my lab
- Good experience if you're thinking about applying to grad schools
- Email me if you're interested:
ccb@upenn.edu

Thanks!