

# Lecture 2: Background and Applications

Instructor: Chien-Ju Ho

# Announcements

- Sign up for Piazza
  - <http://piazza.com/wustl/spring2019/cse518a>
- Assignment 1
  - <http://chienjuho.com/courses/cse518a/fa2019/assignment1.pdf>
- Reading materials
  - Let me know if you find any of the links broken.
  - You might need VPN to get access to some of the files when outside of WashU.
  - Search paper titles should also lead you to the papers.

# Assignment 1

- Submitted via Gradescope.
- You should have already been added to the rosters on Gradescope.
- Due date:
  - September 13 (Friday).
  - Do 3 tasks AND earn \$0.25 on MTurk or Figure Eight.
  - MTurk is recommended.
    - If you are denied registration for a new Mturk account, you can
      - Borrow accounts from your friends/classmates
      - Use Figure Eight instead

Complete at least 3 different tasks and earn at least \$0.25 (both conditions need to be met) on Amazon Mechanical Turk or Figure Eight. Keep track of the amount of time you spend in doing the tasks.

1. Provide screenshots to show that you have done the tasks (you can show the earning summaries or the screenshots of the tasks).
2. Based on the amount of time you spend in doing the tasks and the amount of money you earn, what is your estimated hourly wage as a crowd worker?
3. Choose one of the tasks you have done and answer the following questions
  - (a) Briefly describe the task.
  - (b) Provide a guess of what the requester is using your data for.
  - (c) Do you like the way the task is designed? In your opinion, what can the requester do to improve the task design? (You can answer this question from the perspectives of how to be more fair to workers, how to make the task easier to complete for workers, how to improve the efficiency of data collection, etc).
4. In your opinion, what can the platform (MTurk or Figure Eight) do to improve the crowdsourcing process? (This question is intended to be open-ended, as I don't want to give directions to bias your thoughts.)

# Logistics: Prerequisites

- You need to be comfortable with
  - probability
  - linear algebra
  - calculus
  - basic concepts about computer science and ability to program
- If you are not sure, take a look at the papers from **Sep 9-16**. These papers are more on the math-heavy side.
  - You should at least be able to understand the formulation
  - You will need to implement one of the algorithms in these papers.

# Logistics: Reviews and Assignments

Date	Lecture	Reading Materials	Assignments
Aug 26	Introduction [ <a href="#">Slides</a> ]		
Aug 28	Background and Applications	<p><b>Required</b></p> <p><a href="#">The Rise of Crowdsourcing</a>. Howe. Wired. 2006.</p> <p><b>Optional</b></p> <p><a href="#">Labeling Images with a Computer Game</a>. von Ahn and Dabbish. CHI 2004.</p> <p><a href="#">reCaptcha: Human-based Character Recognition via Web Security Measures</a>. von Ahn et al. Science. 2008.</p> <p><a href="#">Predicting Protein Structures with a Multiplayer Online Game</a>. Cooper et al. Nature. 2010.</p>	<a href="#">Submit Review</a> (Due: Noon, Aug 28)  <a href="#">Assignment 1</a> (Due: Midnight, Sep 13)

- Assignments are graded as “check” and “check-minus”.
- Reviews are “pass” and “not pass”.
  - If you show reasonable effort, you would get a “pass”.
  - You can assume you get “pass” if not notified by me within 2 weeks of submission.

# Logistics: Presentations and Leading of Discussion

- You need to lead a lecture (70~80 min) as a group once this semester.
  - The target number of groups is 10~12.
  - The size of the group depends on the final class size (most likely 2 or 3).
- Read required and at least 2 optional papers.
- Talk to me a week before the lecture.
- Strongly encouraged to include discussions in class.

Sep 30 Incentive Design: Badges

[Student Presentation]

**Required**

[Steering User Behavior with Badges](#). Anderson et al. WWW 2013.

**Optional**

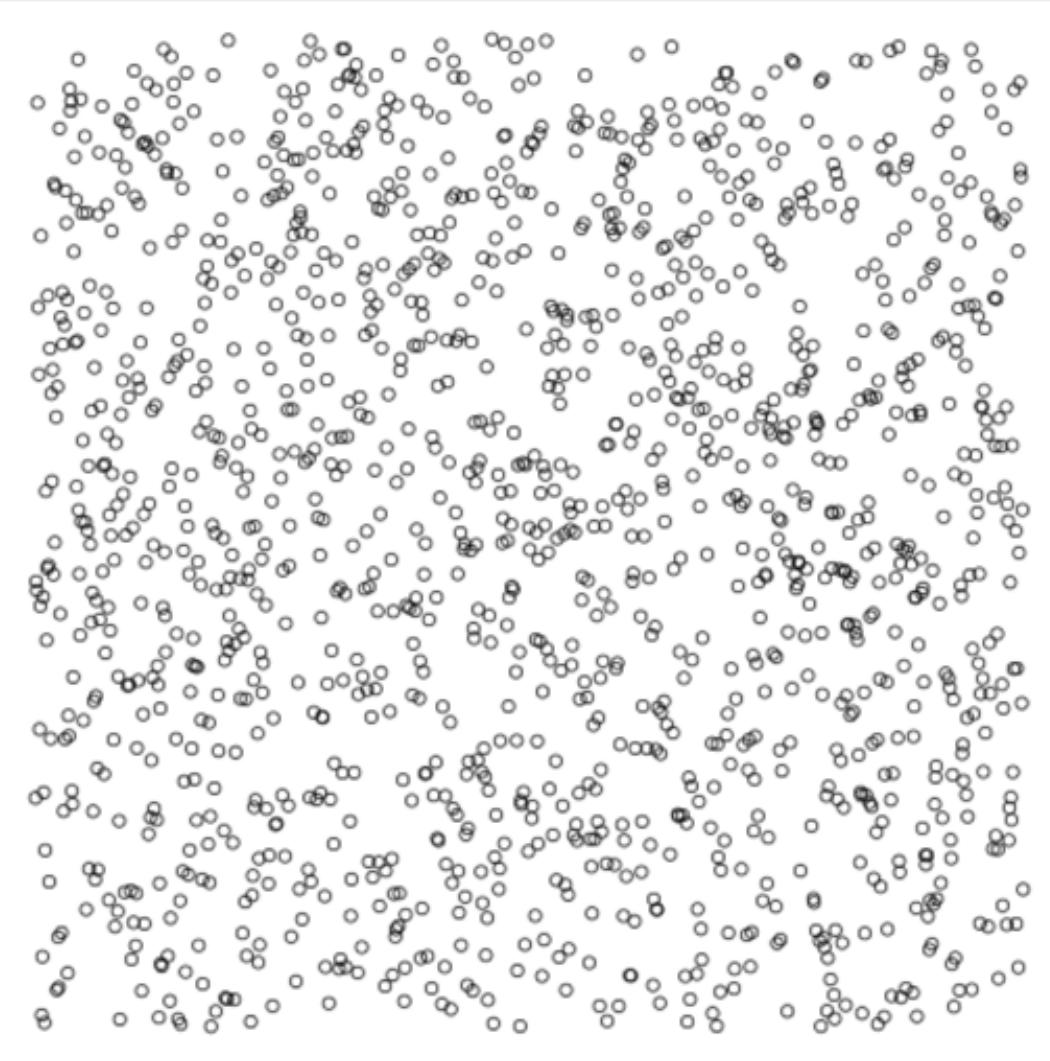
[Incentives, Gamification, and Game Theory: An Economic Approach to Badge Design](#). Easley and Ghosh. EC 2013.

[Social Status and Badge Design](#). Immorlica, Stoddard, and Syrgkanis. WWW 2015

# Logistics: Course Projects

- I'll generate a list of potential/past projects next week.
- Project proposal
  - You can work on anything involving humans in the computation process (theory, empirical, application, literature survey)
  - I'll make the final call on whether it's relevant to the course and might suggest modifications.
- You are encouraged to take a peek of latter lectures to see what topics you might be excited about.

# How many circles are in the image?



Input your guess:

<https://ppt.cc/fLEXOx>

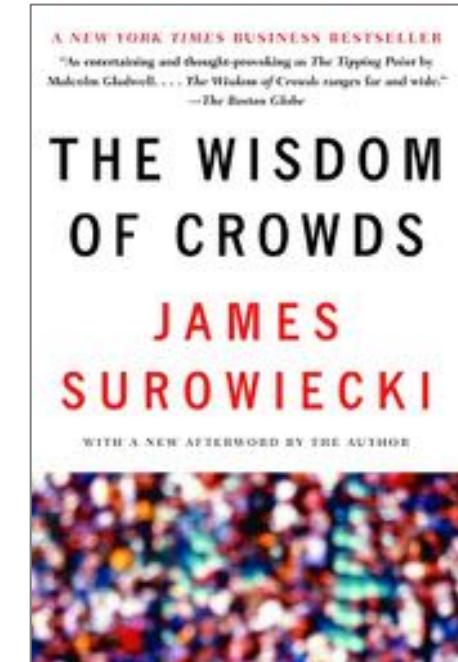
**Table 1: Dot-guessing performance**

Dots		Mean	Geo. Mean	Med.	SD	N	Quant.
310	(1)	375	285	292	349	158	0.69
456	(2)	537	373	400	554	148	0.74
854	(3)	986	579	606	1454	138	0.83
1243	(4)	1232	866	850	1443	143	0.99
2543	(5)	2362	1684	1700	2580	147	0.94
4000	(6)	4888	2855	2562	5970	144	0.92
8200	(7)	8559	4748	4644	12267	134	0.96

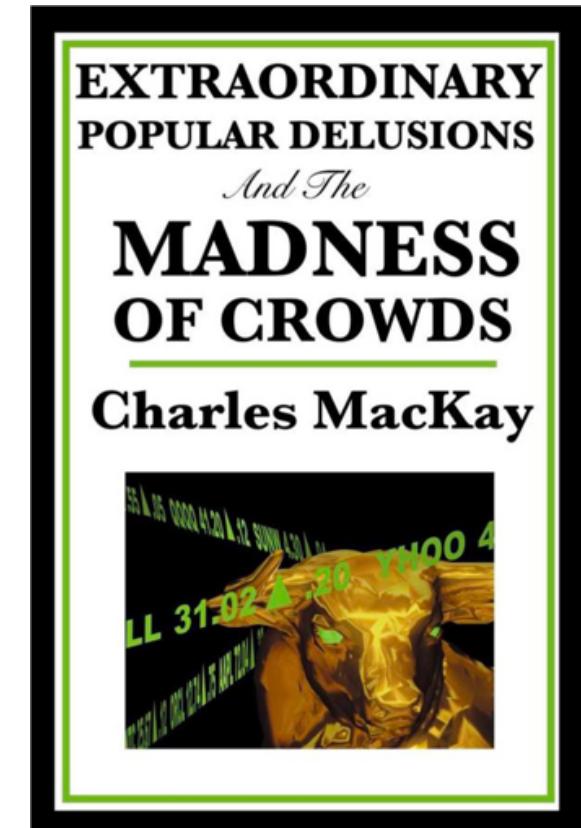
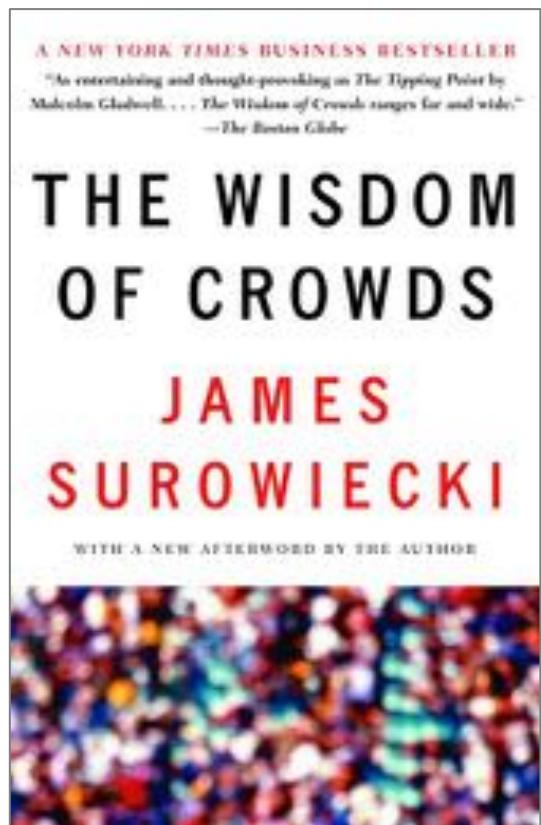
# The Wisdom of Crowds



- At a 1906 country fair, ~800 people participate in a contest to guess the weight of an ox.
- Reward is given to the person with the closest guess.
- The average guess is 1,197lbs.  
The true answer is 1,198lbs.



# When is the crowd wise?



# Warm-up Discussion

- Under what conditions is the crowd wise?
- How should we (mathematically) **model** the process?
  - How to combine the results?
  - Can we estimate how accurate the combined result is?
  - Can we better design the process?

# When is the crowd wise?

- Some general “beliefs” for crowd to be wise
  - Informed individuals
  - Diversity of opinions
  - Independence
  - Good aggregation
  - Aligned incentives
  - ...
- Can we make these ideas more formal?
  - Can we “model” the process?

# Example Model on Aggregation

- People have unbiased estimates of the true answer  
**user guess = true answer + Gaussian noise**
- With this model, we can estimate the number of users needed to achieve a certain level of accuracy (some form of law of large numbers).
- Does this model capture the real human behavior?
  - Maybe there are additional biases. “Which state in the US is closest to Africa?”
- Is simple averaging the best aggregation method?
  - What if we know person A is very good at the task, and person B is random guessing

# Example Model on Incentive

- People choose effort level in generating their true answer

**user utility** = reward \*  $\Pr(\text{getting reward} \mid \text{effort}) - \text{cost}(\text{effort})$

user chooses effort that **maximizes their utility**

user-estimate = true-answer + noise(effort)

- What's the optimal payment/incentives?
  - Need to consider the requester's utility in the accuracy of estimates
  - Can we have different payment structures?
- There could be “bad” incentives

*All models are wrong  
but some are useful*



George E.P. Box

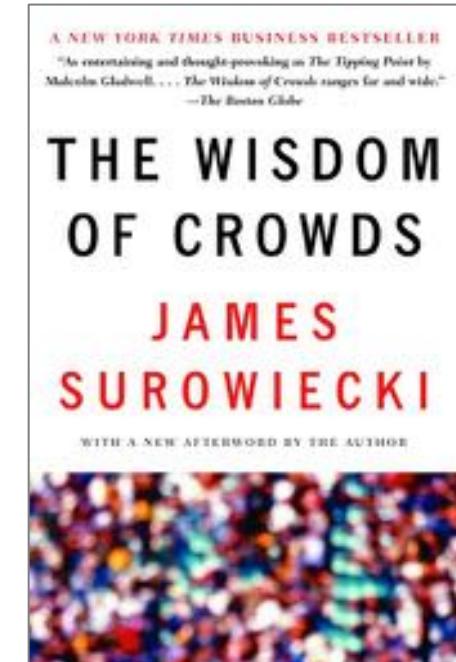
# The focus of this course

- Design and analysis of human-in-the-loop computation
- Develop some “models” of how users take actions in our systems
- Based on the model,
  - design methods to **aggregate** users information
  - design **incentives** to encourage high effort and truthful reports
- Deal with **practical challenges**
- Discussion on **selected recent topics**

# The Wisdom of Crowds

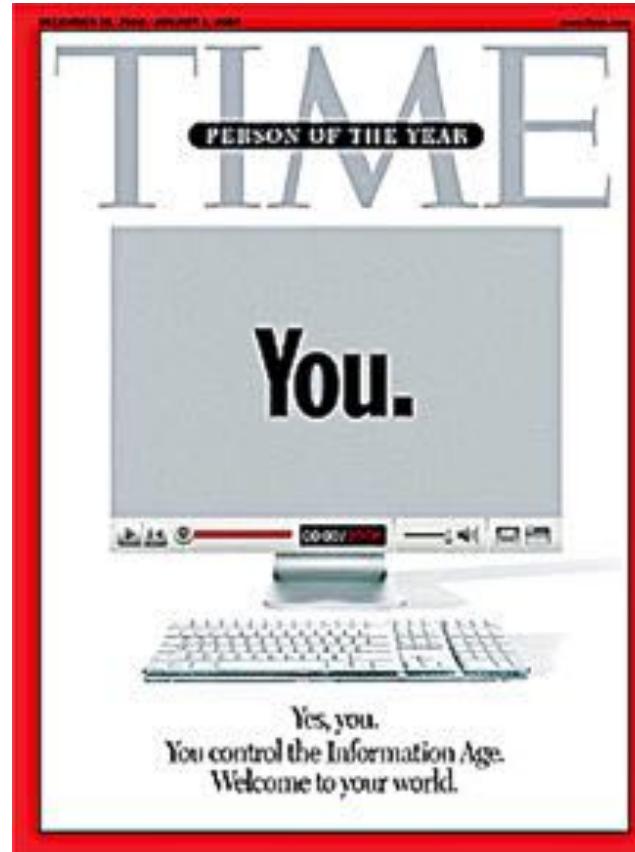


- At a 1906 country fair, ~800 people participate in a contest to guess the weight of an ox.
- Reward is given to the person with the closest guess.
- The average guess is 1,197lbs.  
The true answer is 1,198lbs.



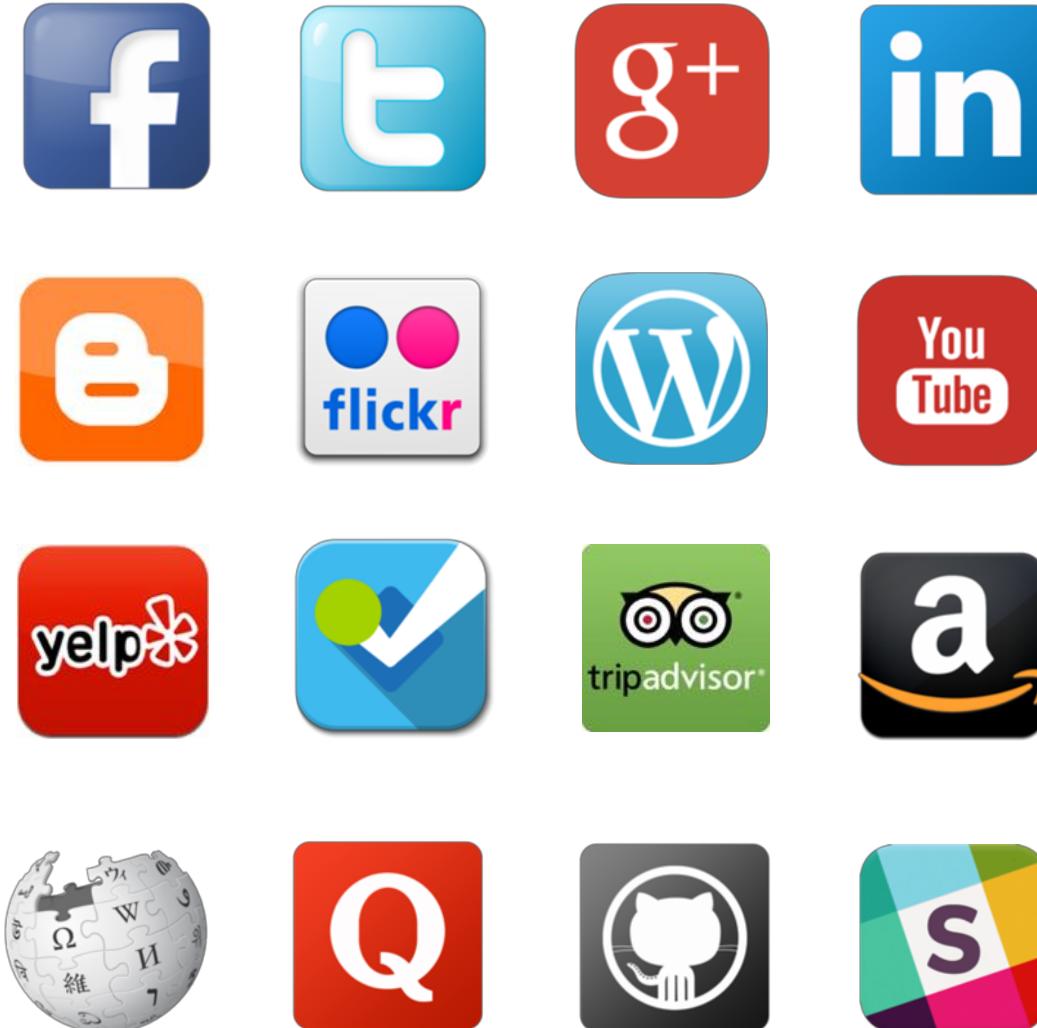
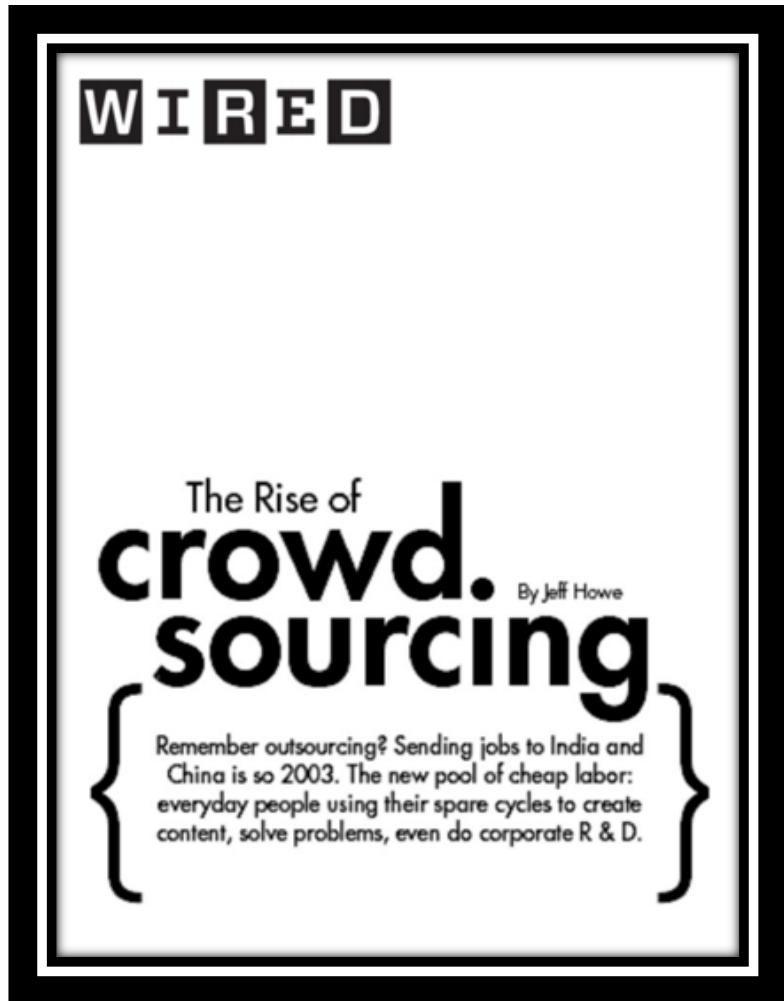
It's hard to bring together a large crowd in the old times.

# Time Person of the Year 2006



**>4 Billion** Internet users worldwide as of 2018!  
**21%** of Americans go online almost constantly!

# The Internet Age and the Rise of Crowdsourcing



Connect

Create

Share

Collaborate

Let's look at a few more applications

# Games with a Purpose

ESP Game: Crowdsource image labeling via games



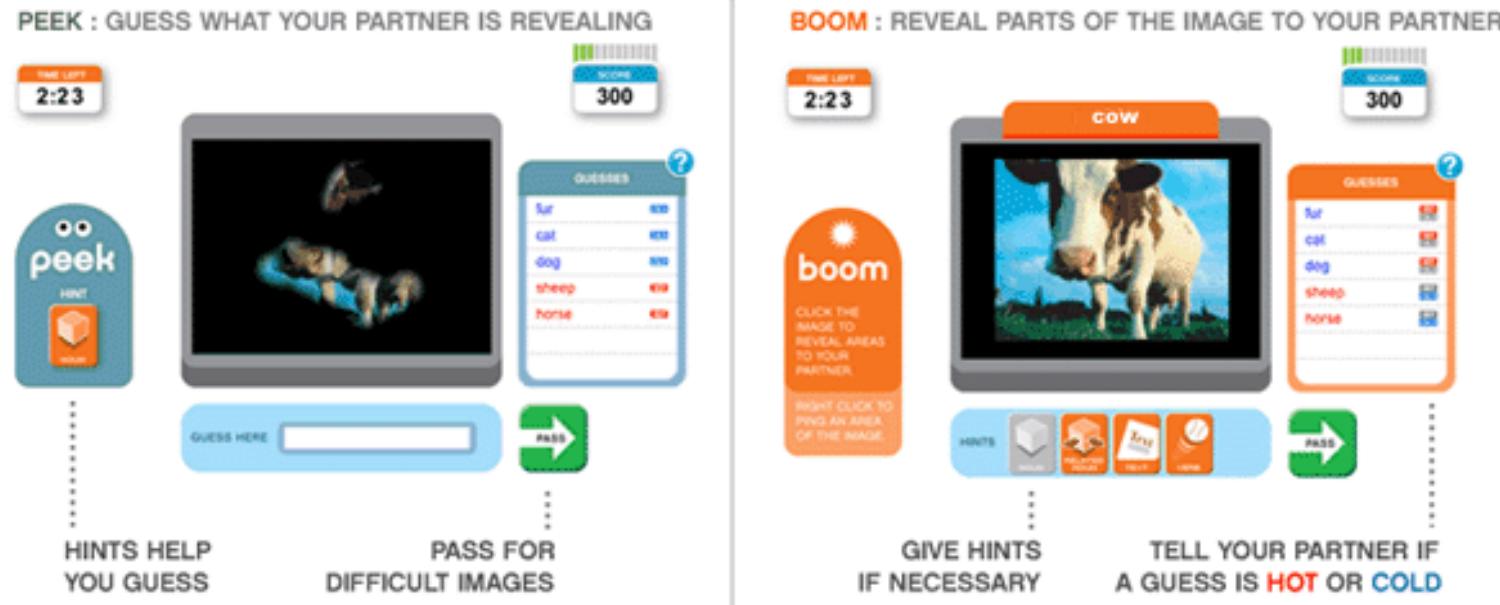
boy  
white  
black  
window  
wall  
  
...



playground  
children  
running  
boy  
  
...

**Potential Label: Boy**

# Games with a Purpose



Peekaboom: Crowdsource image segmentation via games

von Ahn et al. [Peekaboom: A game for locating objects in images](#). CHI'06

# Games with a Purpose

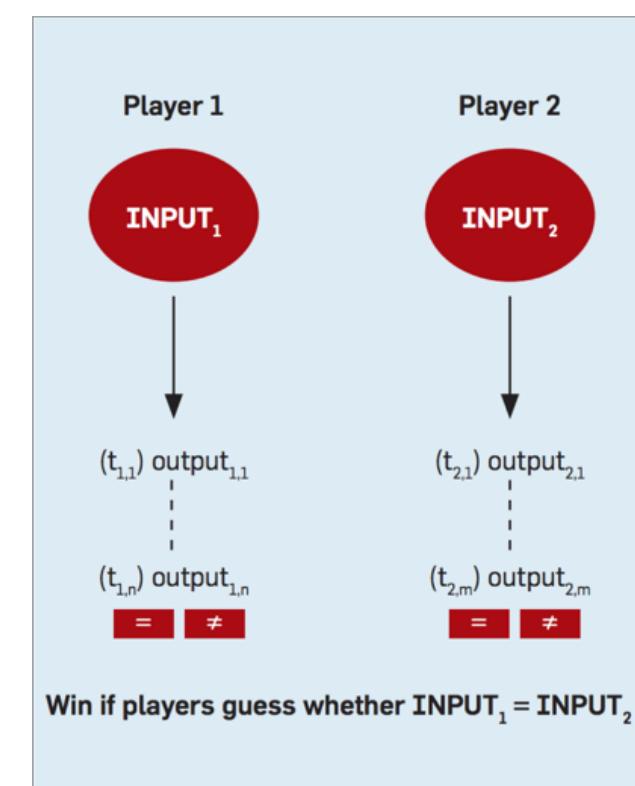
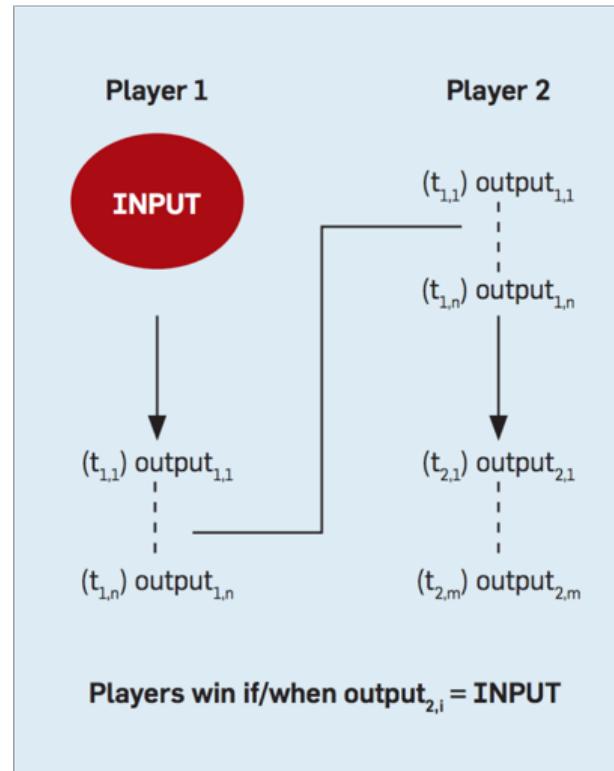
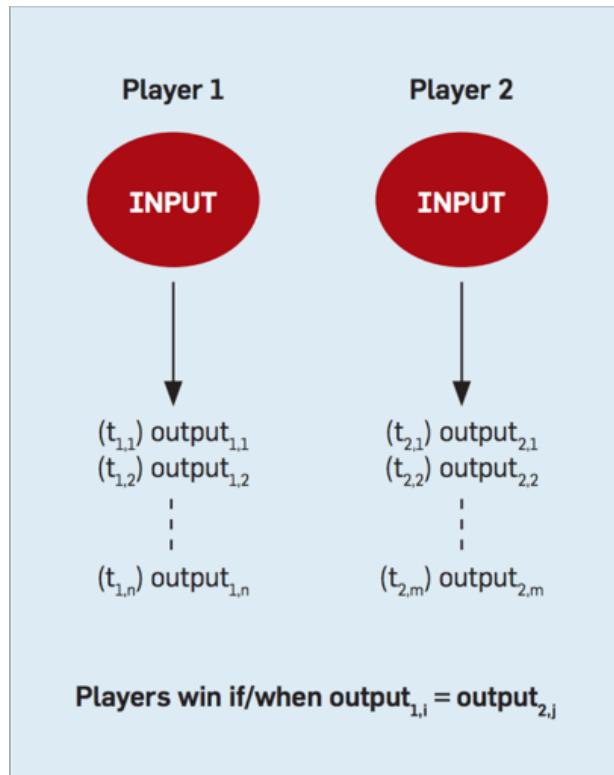


TagATune: Crowdsource audio annotation via games

Law et al. [TagATune: A Game for Music and Sound Annotation](#). ISMIR'07

# Design and analysis of Game with a Purpose

- Incentives and optimization are generally implementation dependent
- How to ensure the data quality is high?



## Output Agreement

## Inversion Problem

## Input Agreement

- Another way to come up project ideas:
  - Try to improve existing systems or fix potential issues.
- What can we do to “improve” ESP Game? Are there any issues with ESP Game?



- People are competitive.
- Potential issues of ESP Game
  - Coalition Problem: let's type “a” for all images
  - Easy words first...
  - Biased by the list of taboo words
- Microsoft implemented a bot
  - analyzes taboo words only
  - achieves match with humans with >80% chance

# An example research project (by myself, 10 years ago...)

- Goal:
  - Fix the “easy-word-first” issue without using the taboo words
  - Incorporate some competitive elements

Player 1



Player 2



DOG  
FLOWER  
FLOOR

CUTE  
WOOD  
DOG

Player 1



Player 2



DOG  
FLOWER  
FLOOR

CUTE  
WOOD  
DOG



CONNECT WITH EACH OTHER BY TYPING THE SAME WORD

Couple



Couple



DOG  
FLOWER  
FLOOR

CUTE  
WOOD  
DOG



CONNECT WITH EACH OTHER BY TYPING THE SAME WORD

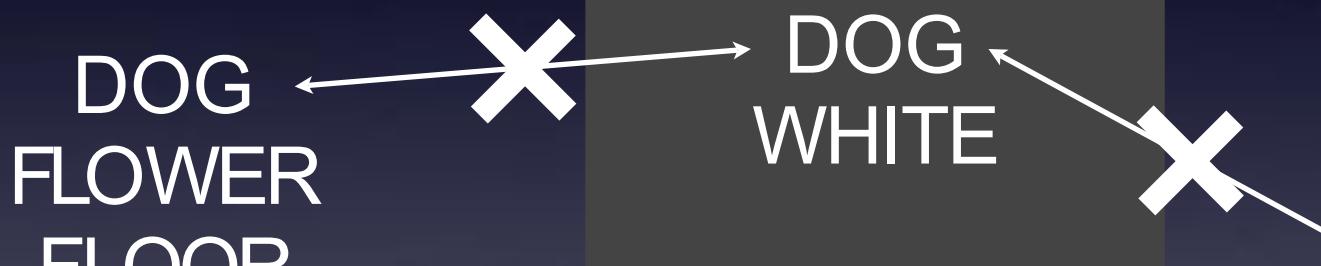
## Blocker



Couple



DOG  
FLOWER  
FLOOR



Couple



CUTE  
WOOD  
DOG

Couple



DOG  
FLOWER  
FLOOR

Blocker



DOG  
WHITE

Couple



CUTE  
WOOD  
DOG

# KissKissBan

A Competitive Human Computation  
Game for Image Annotation

Couple



DOG  
FLOWER  
FLOOR

-5s game time

Blocker



DOG  
WHITE

The blocked words

Blocked

Couple



CUTE  
WOOD  
FLOWER

-5s game time

1. First, the blocker has 7 seconds to type words
2. Then, the couple has 30 seconds to type words

Couple



DOG  
FLOWER  
FLOOR

Blocker



DOG  
WHITE

Couple

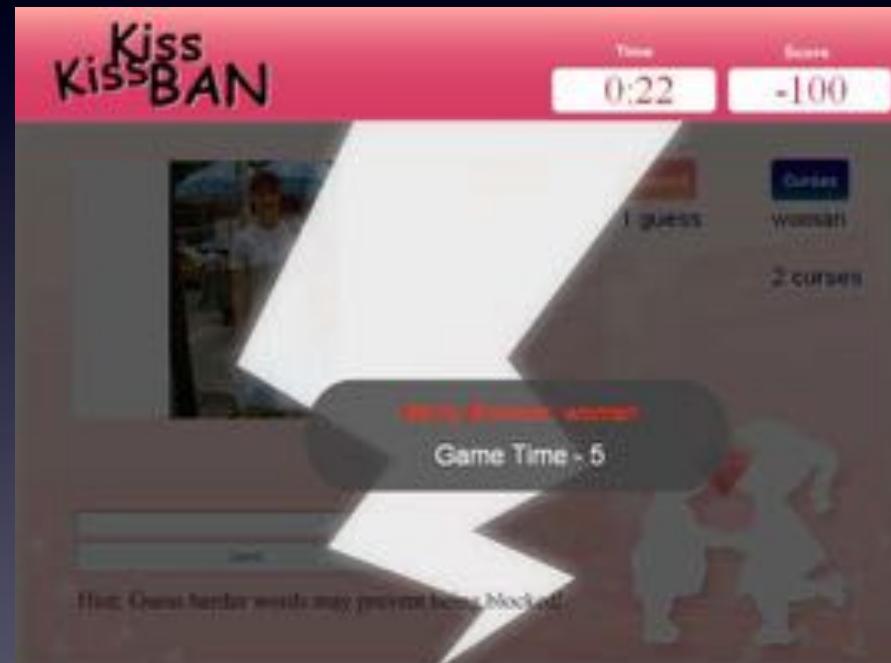


CUTE  
WOOD  
FLOWER

Couples win for matching within time limit

1. First, the blocker has 7 seconds to type words
2. Then, the couples has 30 seconds to type words

# Screenshots of KissKissBan



# Example of Collected Data



ESP	ML-KKB	BL-KKB
man 21	beach 3 water 3 sand 3 sea 2 ninja 1 kungfu 1 ocean 1	sea 9 man 8 ocean 3 black 1 china 1 sand 1
beach 10		
karate 5		
water 1		

ML-KKB: matching between couples

BL-KKB: matching between couple and blocker

# Prediction Markets



- Potential payoff is \$10. If I think the probability of Romney winning is  $p$ , I should
  - **Buy** this security at any prices **less than \$10 p**
  - **Sell** this security at any prices **greater than \$10 p**

**Price measures the population's collective beliefs**

# Does the price converge to the collective beliefs?

- Imagine the scenario:
  - Alice
    - believes the event won't happen ( $p=0$ )
    - very stubborn
    - very rich with unlimited amount of money
  - Bob
    - believes the event will definitely happen ( $p=1$ )
    - very stubborn
    - very rich with unlimited amount of money
- The price will just fluctuate between the two extremes

# So how good are the predictions?

- In practice...
  - Election markets beat data from polls (BR02)
  - Oscar markets beat expert columnists (PGN01)
  - HP internal markets beat sales forecasts (P00)
  - Racetrack odds beat experts (F79)
  - and many more
- In theory...
  - Under certain assumptions, prices converge and reflect the traders' collective knowledge
  - Proposing human behavior models to explain the real-world behavior

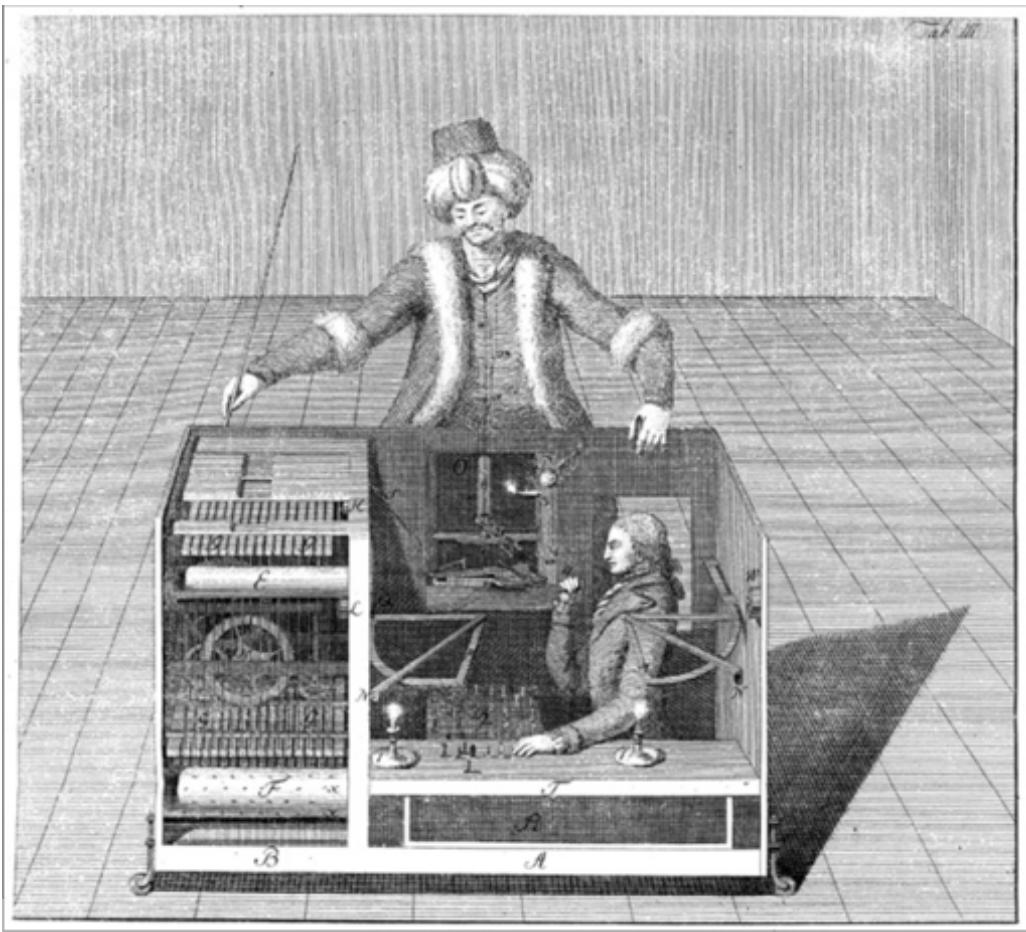
# Crowdsourcing Markets

- The Turk – Automatic Chess Player built in the 18<sup>th</sup> century



# Crowdsourcing Markets

- The Turk – Automatic Chess Player built in 80s



# Crowdsourcing Markets

- Amazon Mechanical Turk - Artificial Intelligence

HIT Groups (1-20 of 1318)							<input type="button" value="Show Details"/>	<input type="button" value="Hide Details"/>	Items Per Page:	20
Requester	Title	Hits	Reward	Created	Actions					
<a href="#">Megan</a>	Categorization	45,696	\$0.01	1h ago	<a href="#">Preview</a>	<a href="#">@ Quality</a>				
<a href="#">Perch Mturk</a>	Kitchen Appliance Classification	14,958	\$0.10	1d ago	<a href="#">Preview</a>	<a href="#">@ Quality</a>				
<a href="#">Alexandra Dodson</a>	Find email address and first/last name of Office Manag...	9,327	\$0.10	1d ago	<a href="#">Preview</a>	<a href="#">Accept &amp; Work</a>				
<a href="#">Alexandra Dodson</a>	Find email address and first/last name of Office Manag...	8,677	\$0.11	1d ago	<a href="#">Preview</a>	<a href="#">Accept &amp; Work</a>				
<a href="#">rick</a>	Why is this review positive?	7,965	\$0.01	6d ago	<a href="#">Preview</a>	<a href="#">Accept &amp; Work</a>				
<a href="#">rick</a>	Why is this review negative?	7,058	\$0.01	6d ago	<a href="#">Preview</a>	<a href="#">Accept &amp; Work</a>				
<a href="#">James Billings</a>	Market Research Survey	6,680	\$0.01	1h ago	<a href="#">Preview</a>	<a href="#">Accept &amp; Work</a>				
<a href="#">Alexandra Dodson</a>	Find email address and first/last name of owners or ge...	4,511	\$0.11	1d ago	<a href="#">Preview</a>	<a href="#">Accept &amp; Work</a>				

# Crowdsourcing Markets

- A general-purpose platform for all kinds of applications
  - mainly used to engage a crowd, and you can build your applications on top
  - In addition, it's easier to design "payment" than to design "fun"

The collage illustrates various crowdsourcing market applications:

- Left Panel:** A screenshot of a mobile application interface. At the top, a question is displayed: "What temperature is my oven set to?". Below the question is a photograph of a dish with corn and meat, with a red dashed box highlighting the meat. The bottom of the screen shows a list of posts related to Harvard, such as "visiting harvard in the summer" and "Our brainstream".
- Middle Panel:** A screenshot of a "Worker Interface" showing a conversation between a user and a worker. The user asks, "Can you find me some good restaurants in Pittsburgh?". The worker responds, "Sure. Wait a minute...". The interface includes a map of Cambridge, Massachusetts, at the bottom.
- Right Panel:** A diagram titled "Auto Responders" showing a vertical stack of small robot icons. Blue arrows point from speech bubbles containing questions like "What kind of food do you want?", "Where are you?", "Hi", "Sure. Wait a minute...", and "Is Chinese food ok?" to the corresponding robot icons.

There are more....

# Citizen Science Projects



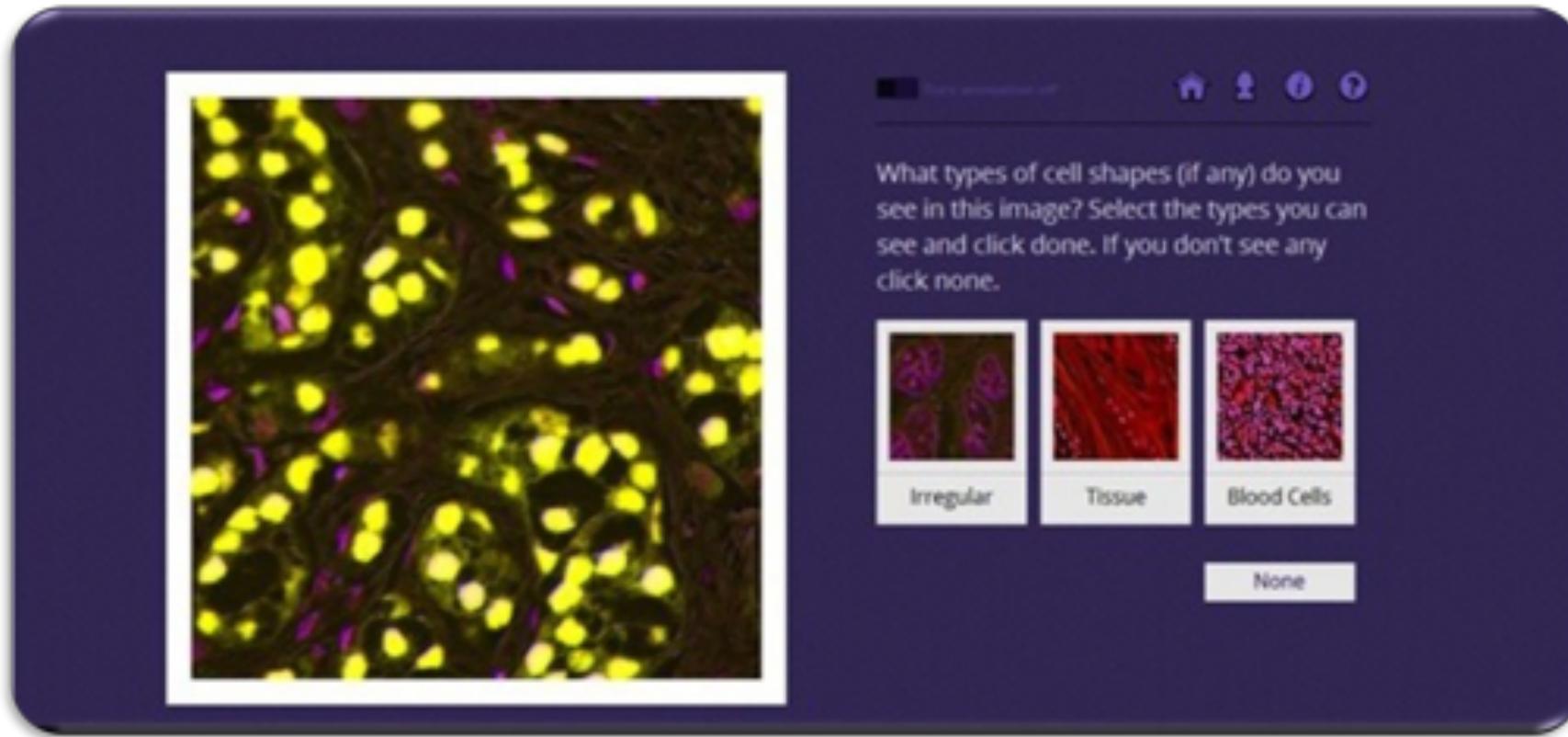
FoldIt: Online video game for protein folding

**2011:** decipher the crystal structure of an AIDS-related virus

**2012:** The first crowdsourced redesign of a protein

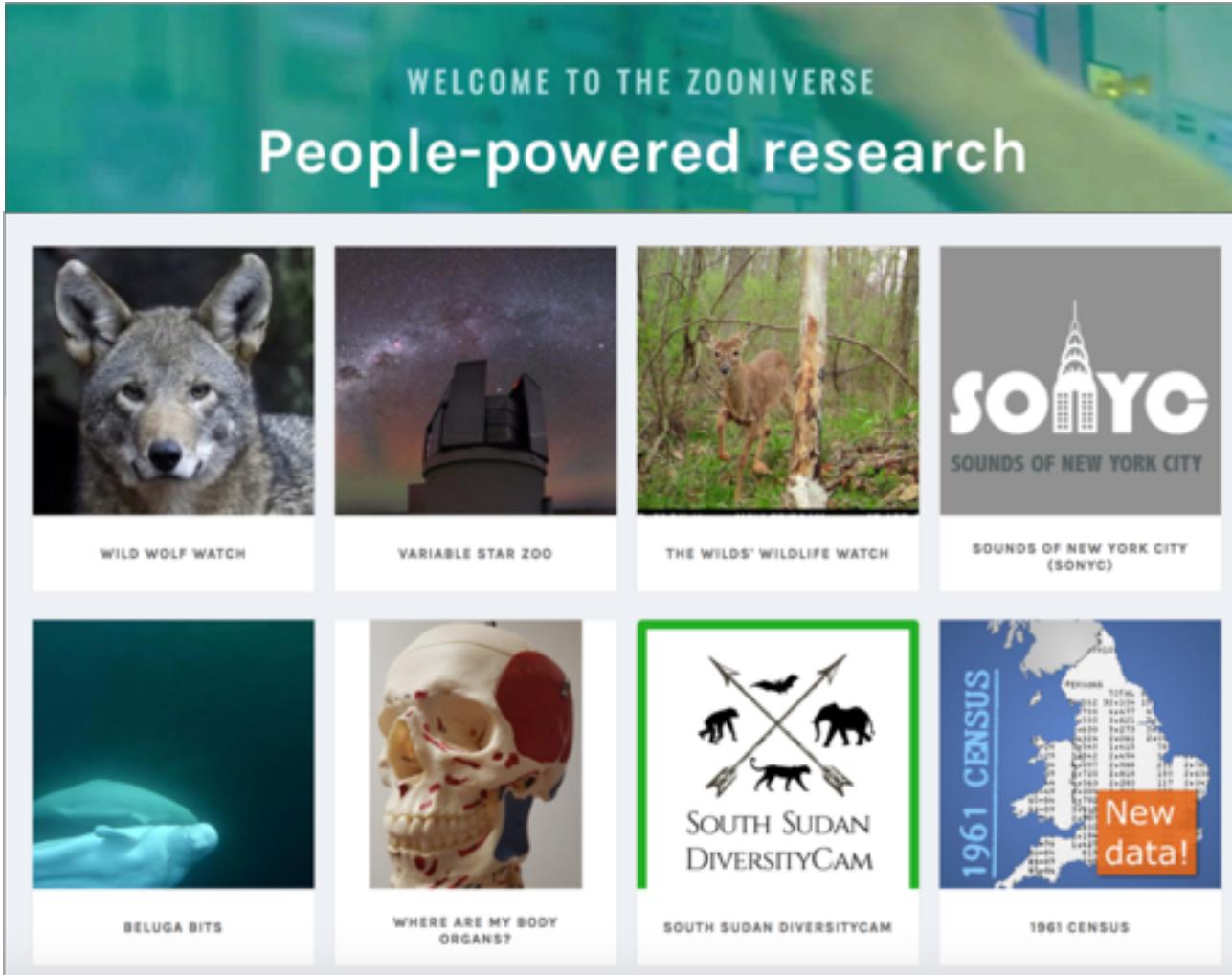
Cooper et al. [Predicting Protein Structures with a Multiplayer Online Game](#). Nature, August 2010

# Citizen Science Projects



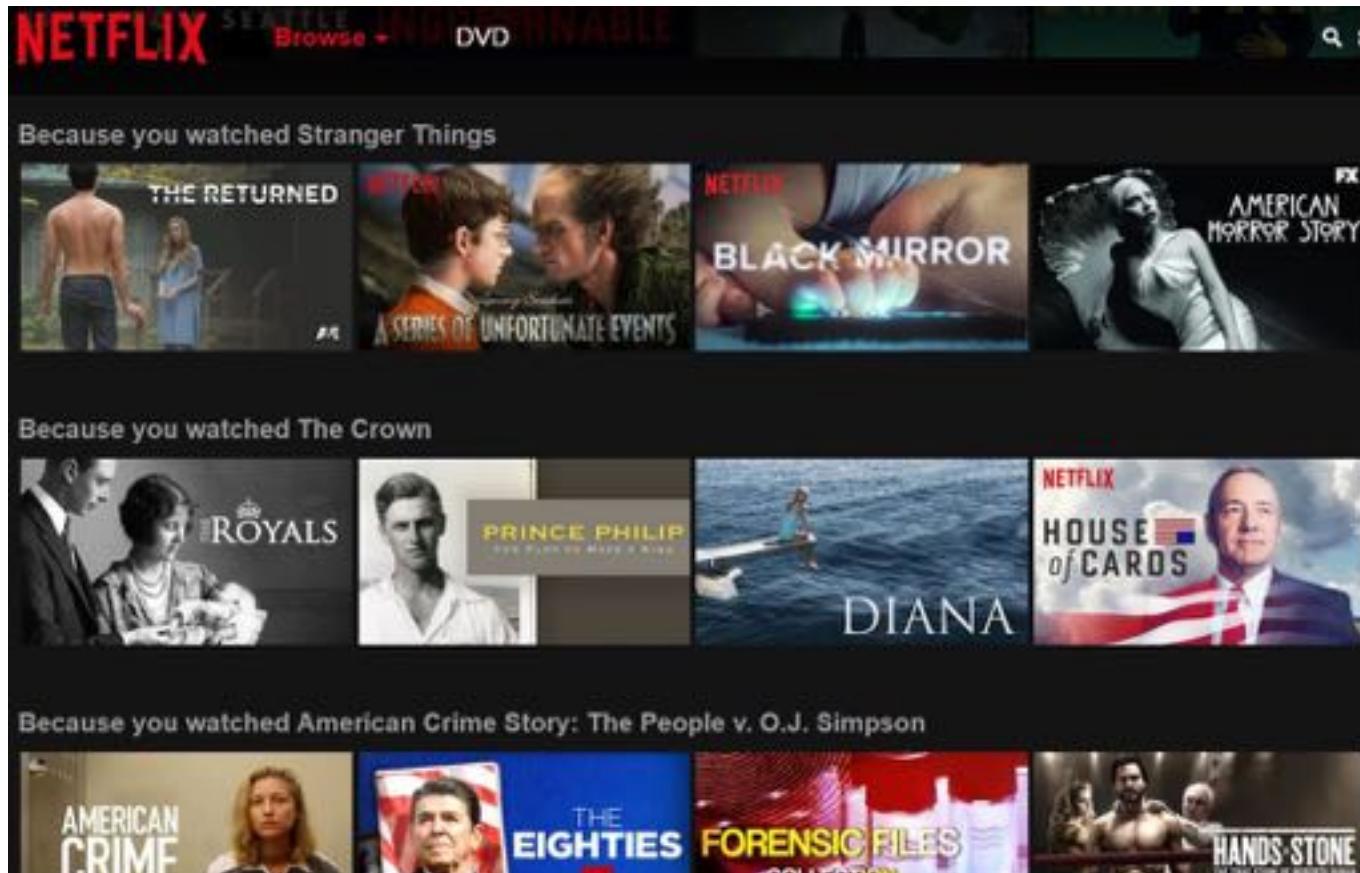
Cell Slider: Examine tumor tissue samples and spotting cancerous cells

# Citizen Science Projects



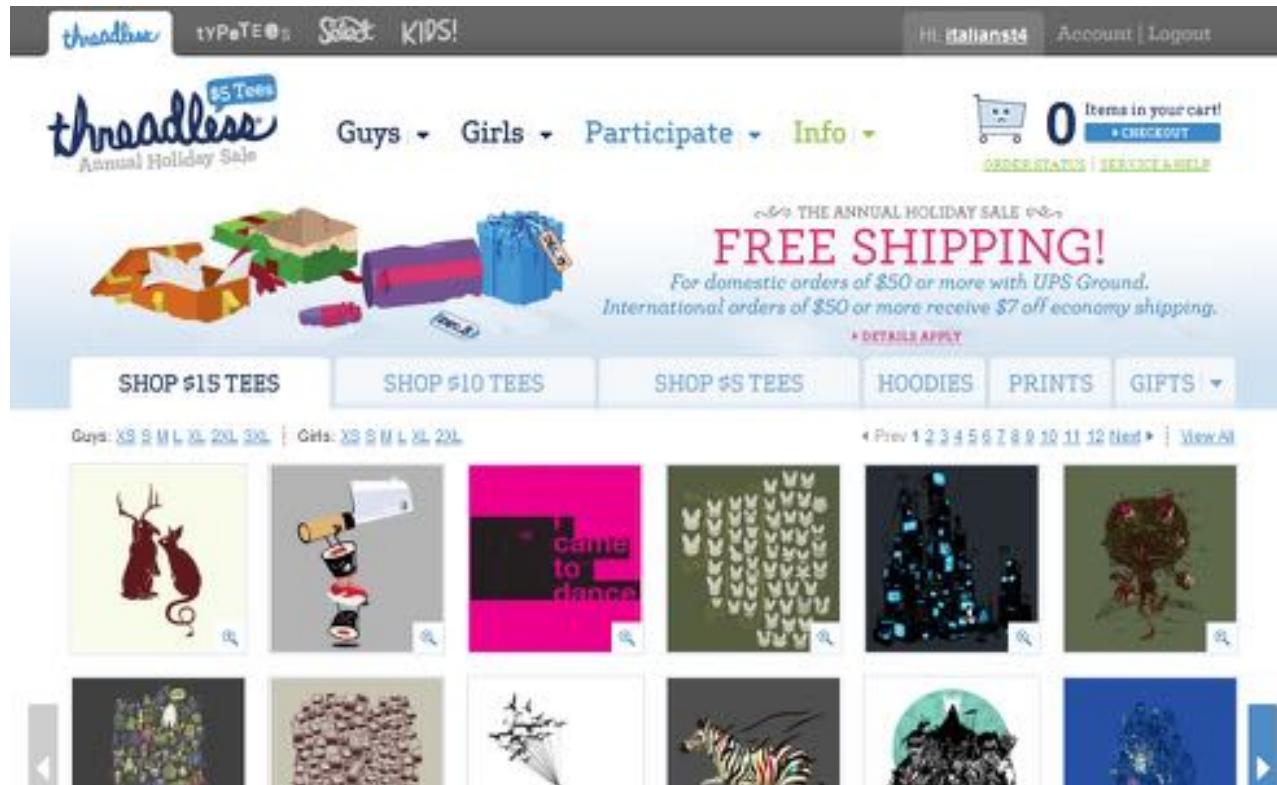
Zooniverse: A citizen science web portal

# Crowdsourcing Contests



- Netflix Prize, 2006
- 44,014 submissions
- The best solution improved its recommendation engine's performance by 10%
- crowdsourcing within a crowdsourcing competition

# Crowdsourcing Contests



- Threadless
- Crowds generate the design
- Crowds vote on the design
- 438,451 designs submitted,  
8,805 printed

# Crowdsourcing Contests



Detecting Leaks and Flaws in Water Pipelines – Stage 1

**TAGS:** Chemistry Engineering/Design Physical Sciences Water Theoretical-licensing

**Award:** \$75,000 USD      **STATUS:** Under Eval  
**Active Solvers:** 294      **Posted:** Mar 08 2018  
**Challenge ID:** 9933883

**PREMIUM CHALLENGE**

**Abstract**

The Bureau of Reclamation and its collaborators are seeking innovative methods and technology to detect leaks and flaws in buried water pipeline infrastructure. Pipeline leaks and flaws deprive consumers of a valuable resource and necessitate costly repairs. Currently, no practical method exists to detect leaks and flaws in large diameter operational pipelines regardless of pipe diameter and material type. Solutions can be novel approaches or can build upon existing methods or technologies. This Challenge launches Stage 1 of a planned two-stage Challenge that includes laboratory-scale and field-scale demonstrations.

**Challenge Orientation Video:** Subject matter experts from Reclamation, San Diego County Water Authority, and Southern Nevada Water Authority discuss the need for improved pipeline leak detection methods.

**PRIVACY ADVISORY**

This web site is hosted by a private entity and is not a service of the Reclamation or the Department of the Interior (DOI). The solicitation and collection of your personal or individually identifiable information is subject to the host's privacy and security policies and will not be shared with Reclamation or DOI unless a Solver wins the Challenge. To collect an award, Challenge winners' personally identifiable information must be made available to Reclamation. Please consult the [Challenge-Specific Solver Agreement](#).

- InnoCentive
- Crowdsourcing problem-solving
- 2000+ problems, 380,000+ solvers
- Problems are supplied by corporations, governments, non-profits, etc.

# Crowdfunding

Explore Start a project

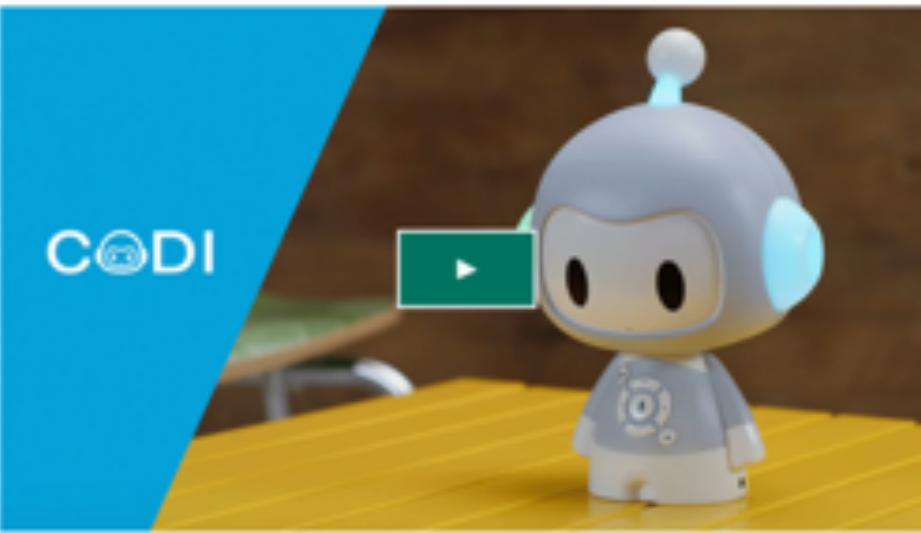
KICKSTARTER

Search  Sign in

 By Pilar Learning  
First created

Codi: Interactive smart storyteller toy for kids

Using hundreds of songs, stories and lessons, Codi provides developmental support for your kids outside of the classroom

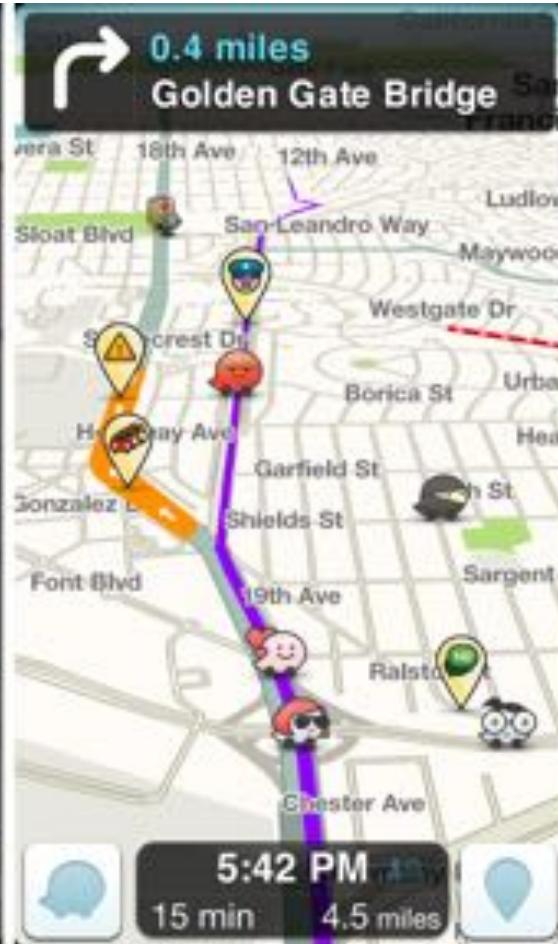


\$15,393 pledged of \$35,000 goal  
162 backers  
31 days to go

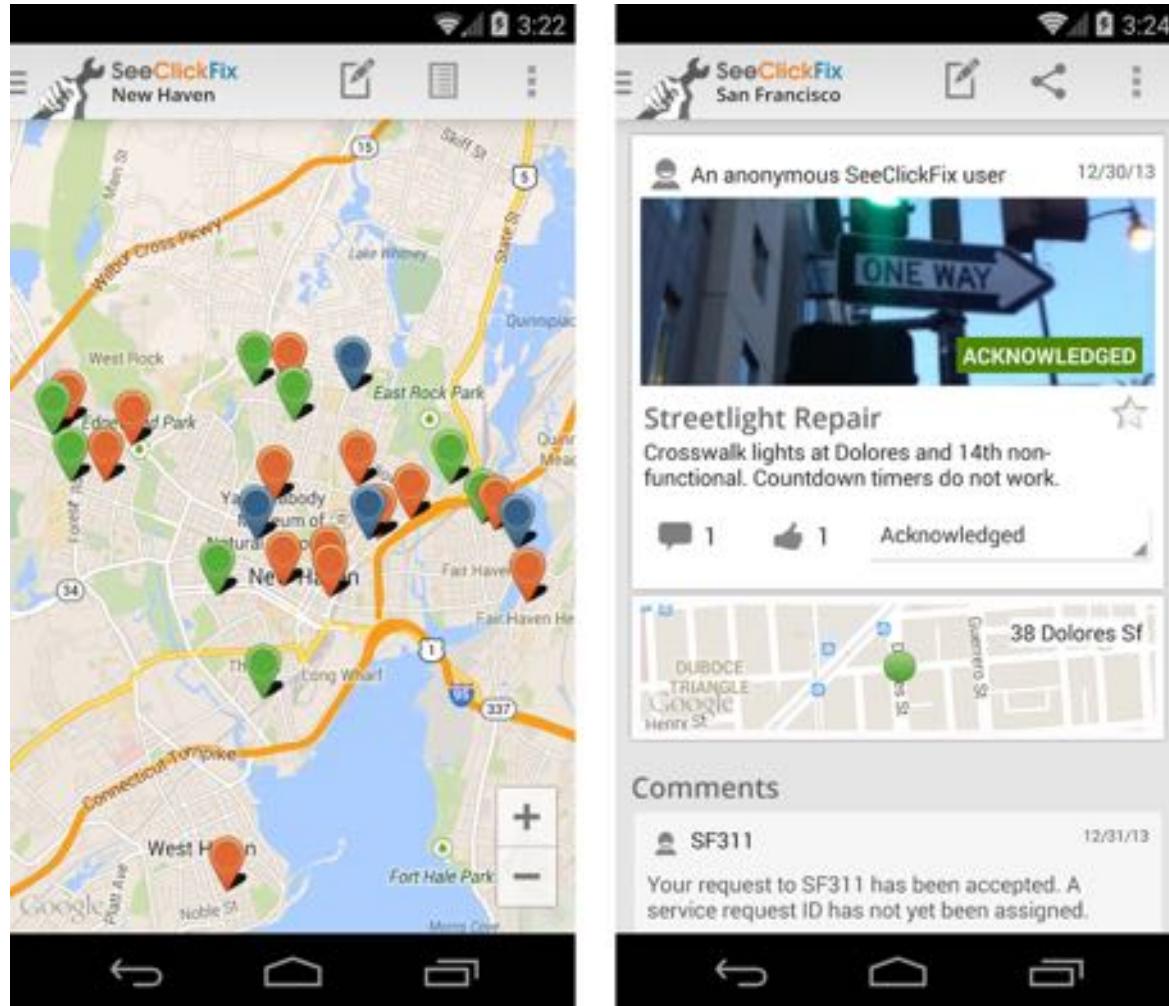
[Back this project](#)

[Remind me](#)   
All or nothing. This project will only be funded if it reaches its goal by Tue, September 18 2018 11:54 AM (EST).

# Crowdsensing: Waze



# Crowdsourcing: SeeClickFix



The image displays two screenshots of the SeeClickFix mobile application. The left screenshot shows a map of New Haven, Connecticut, with numerous service requests marked by colored pins (green, orange, blue) scattered across the city. The right screenshot shows a specific service request for a streetlight repair in San Francisco, California, dated December 30, 2013. The request details a problem with crosswalk lights at Dolores and 14th Streets. It includes a photo of a 'ONE WAY' sign, a green 'ACKNOWLEDGED' button, and social sharing icons. The bottom part of the right screenshot shows a zoomed-in map of the area around 38 Dolores Street and comments from a user named SF311, dated December 31, 2013.

# Discussion

- Can you think of some other examples of crowdsourcing applications?
- What are some examples of challenging tasks in your daily life? Do you think the crowd can help with your challenging tasks? Why or why not?

# Next few lectures

- Label aggregation
  - Use probabilistic inferences to aggregate noisy data generated by humans
- Incentive design
  - Model user incentives and discuss the design to motivate high-quality data
- Will be the more mathematical lectures of this semester