

Crowdsourcing and Human Computation

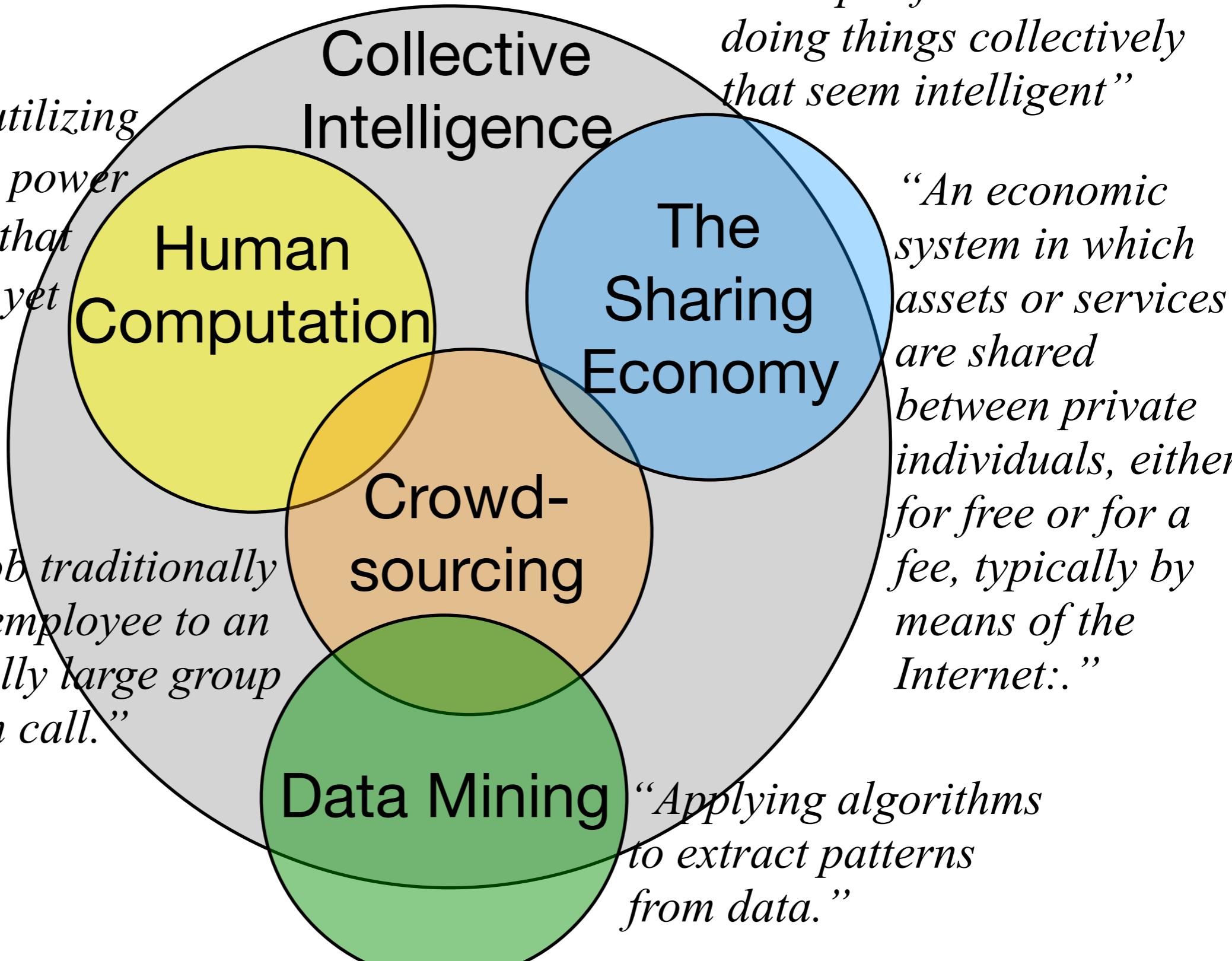
Class time: MWF 2-3pm

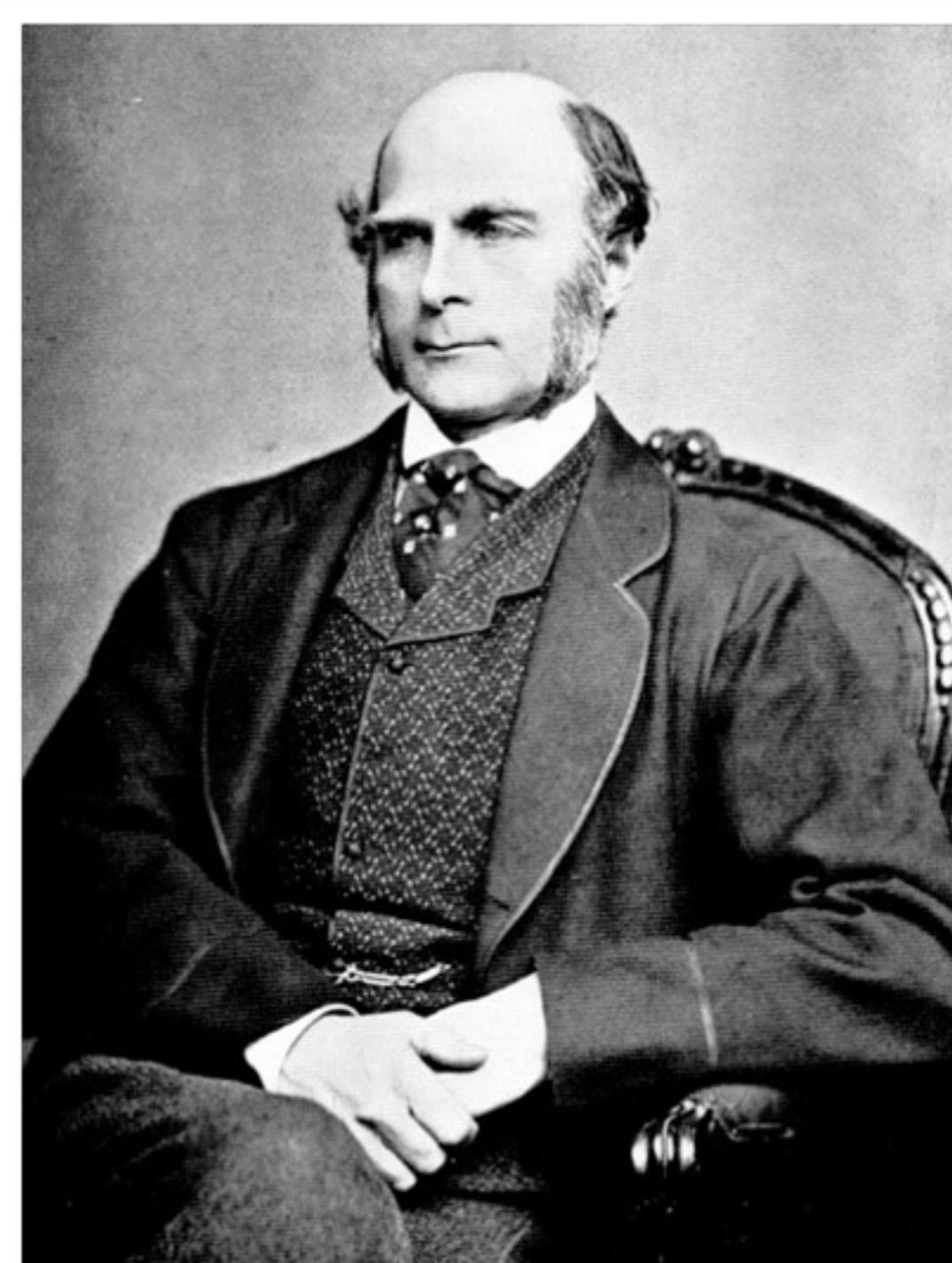
Instructor: Chris Callison-Burch

Website: crowdsourcing-class.org

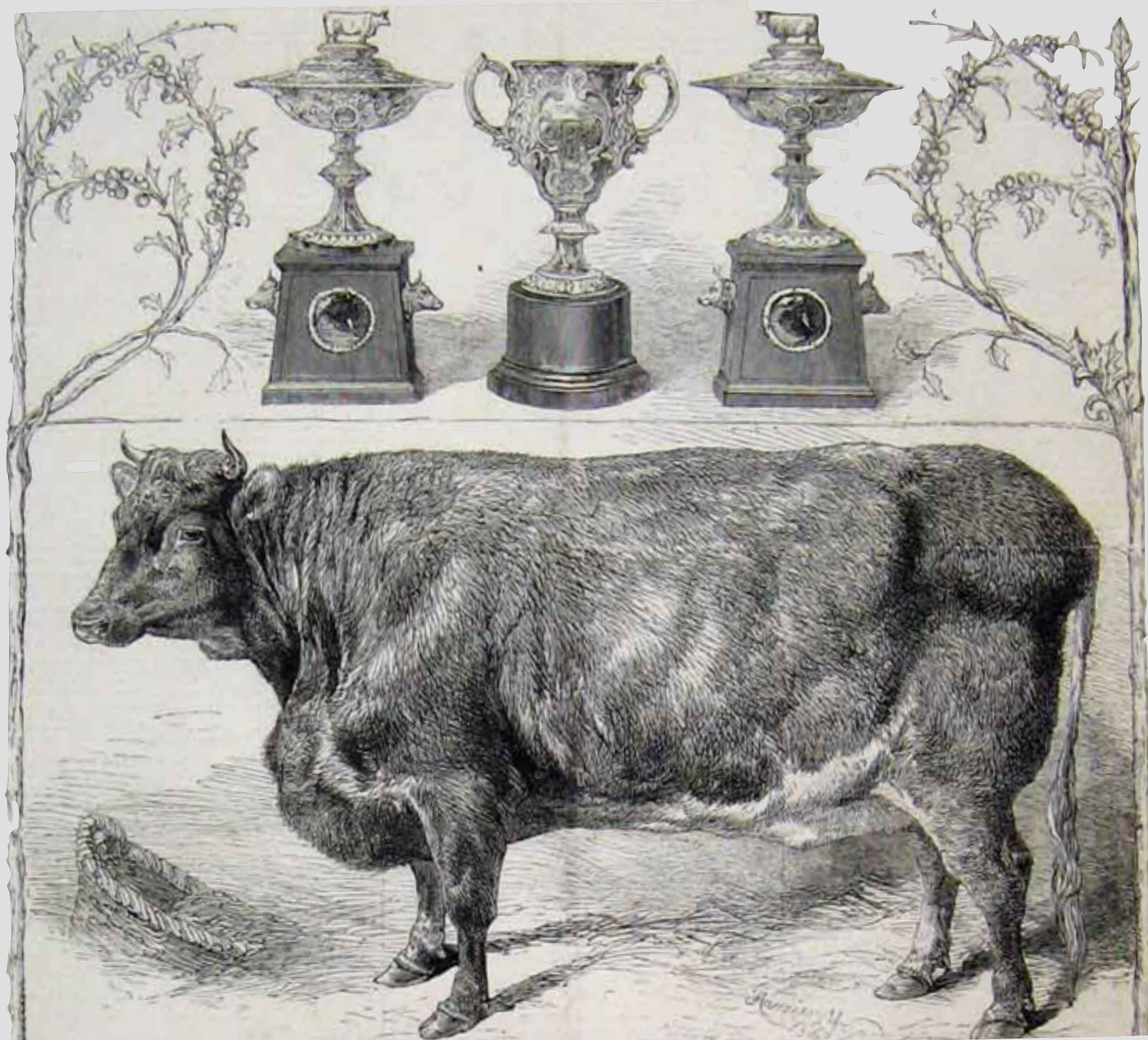
Inter-related concepts

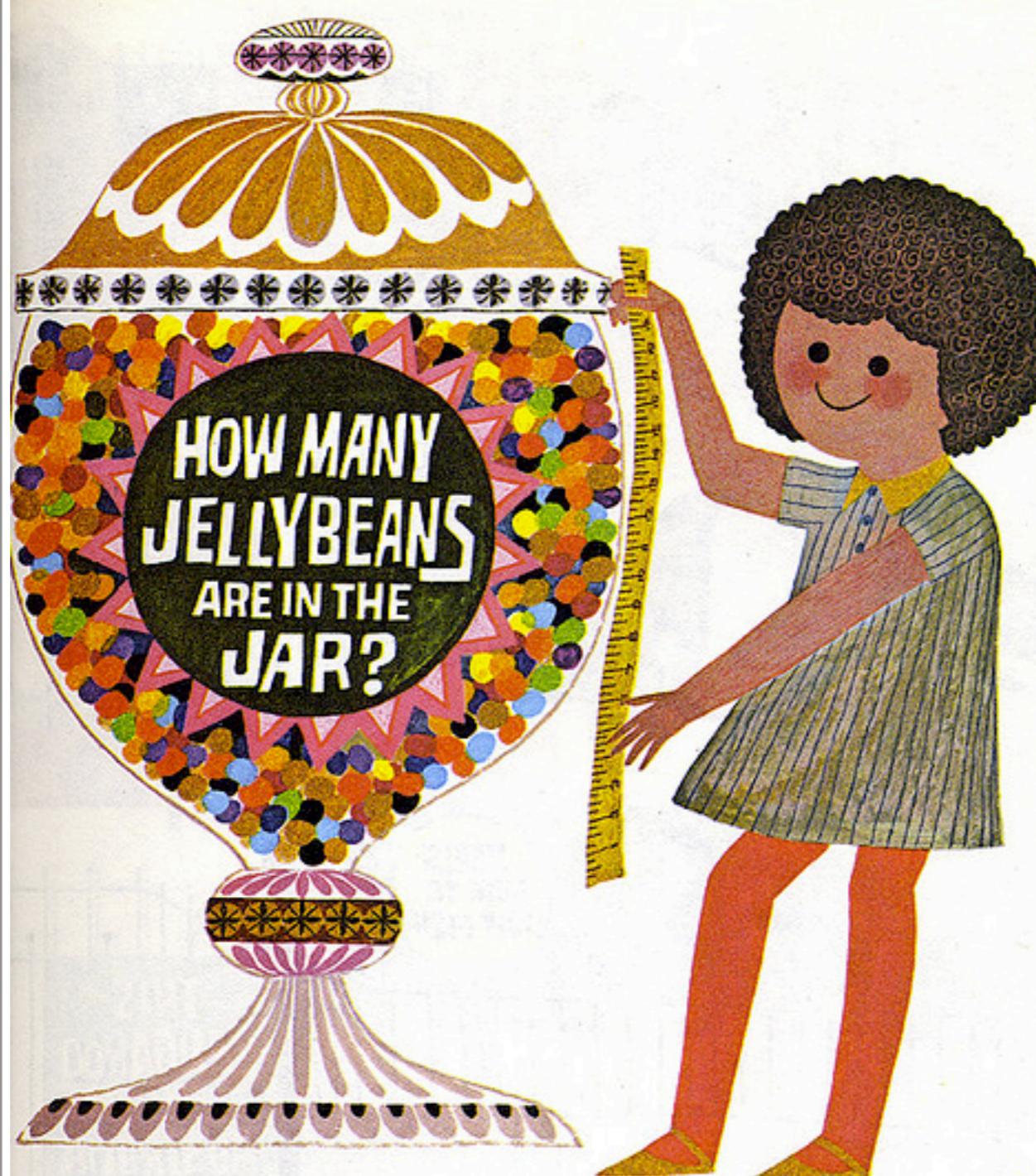
“A paradigm for utilizing human processing power to solve problems that computers cannot yet solve.”





Francis Galton





If you guess the right number of jellybeans,
you may win a prize.

Collective Intelligence?



"All those in favour say 'Aye'."

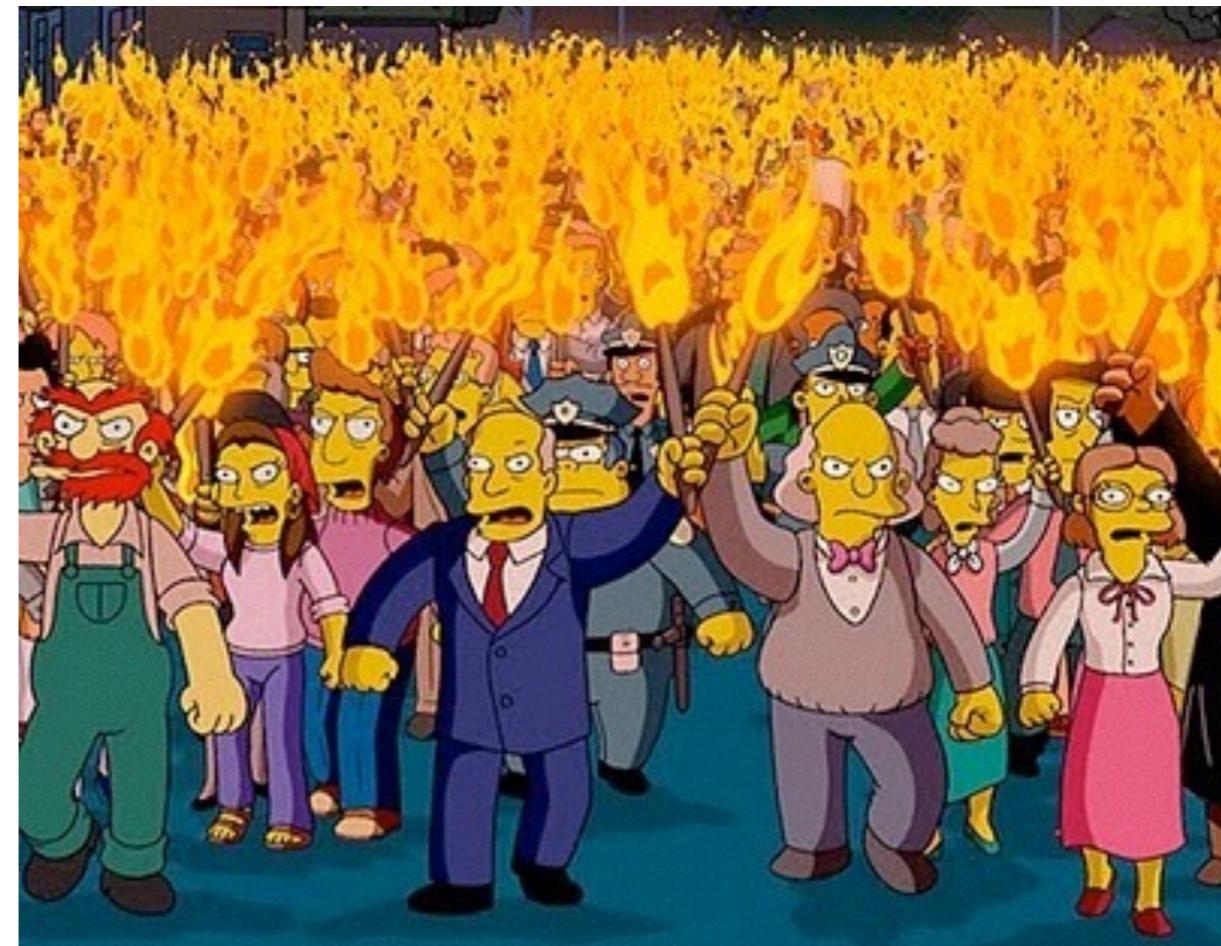
"Aye."

"Aye."

"Aye."

"Aye."

"Aye."

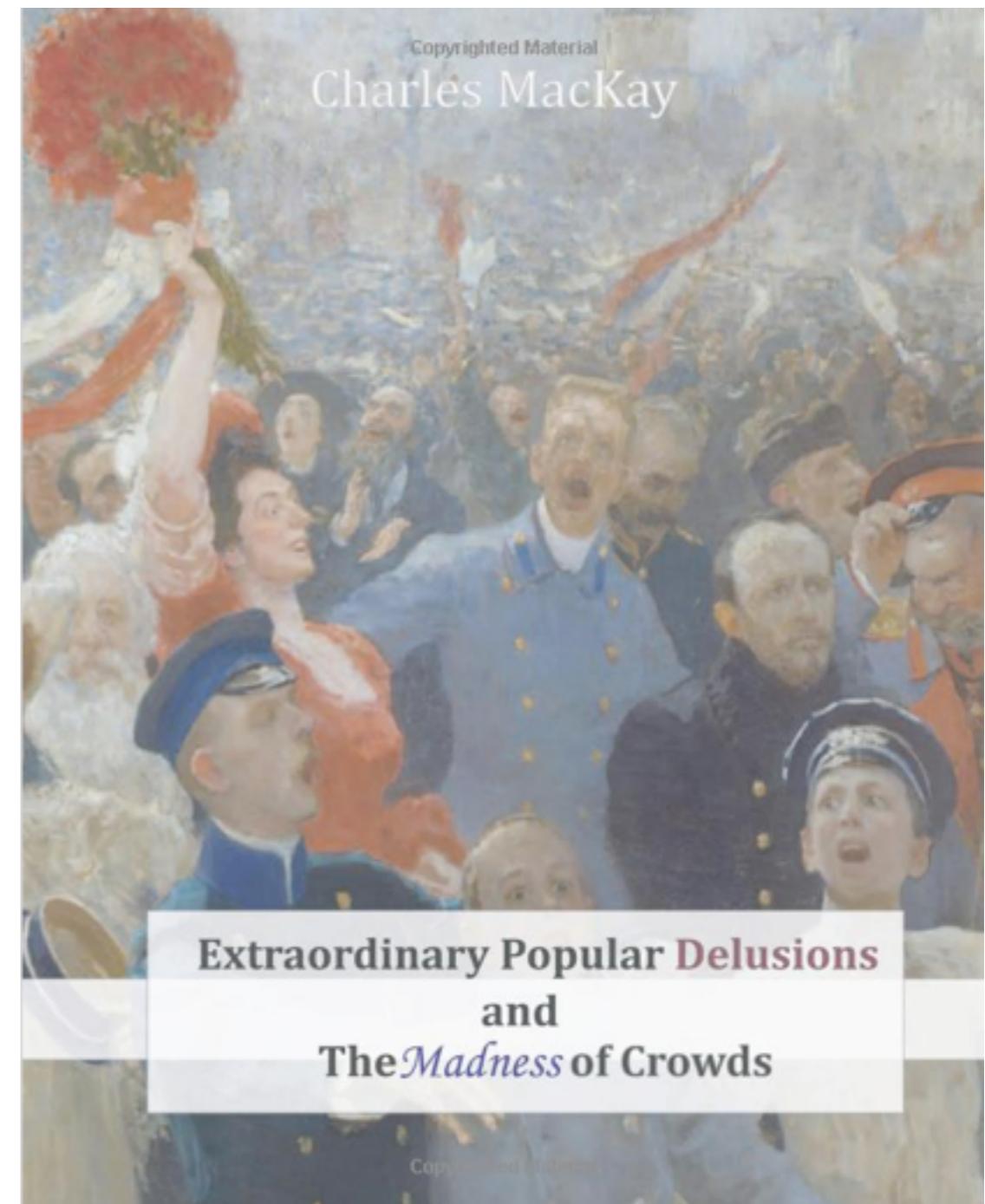


Group Think

Mob Mentality

Popular Delusions and the Madness of Crowds

- Economic bubbles
- Alchemy & Psuedoscience
- Witch hunts
- Prophecies



Plush Life

Why did people lose their minds over Beanie Babies?



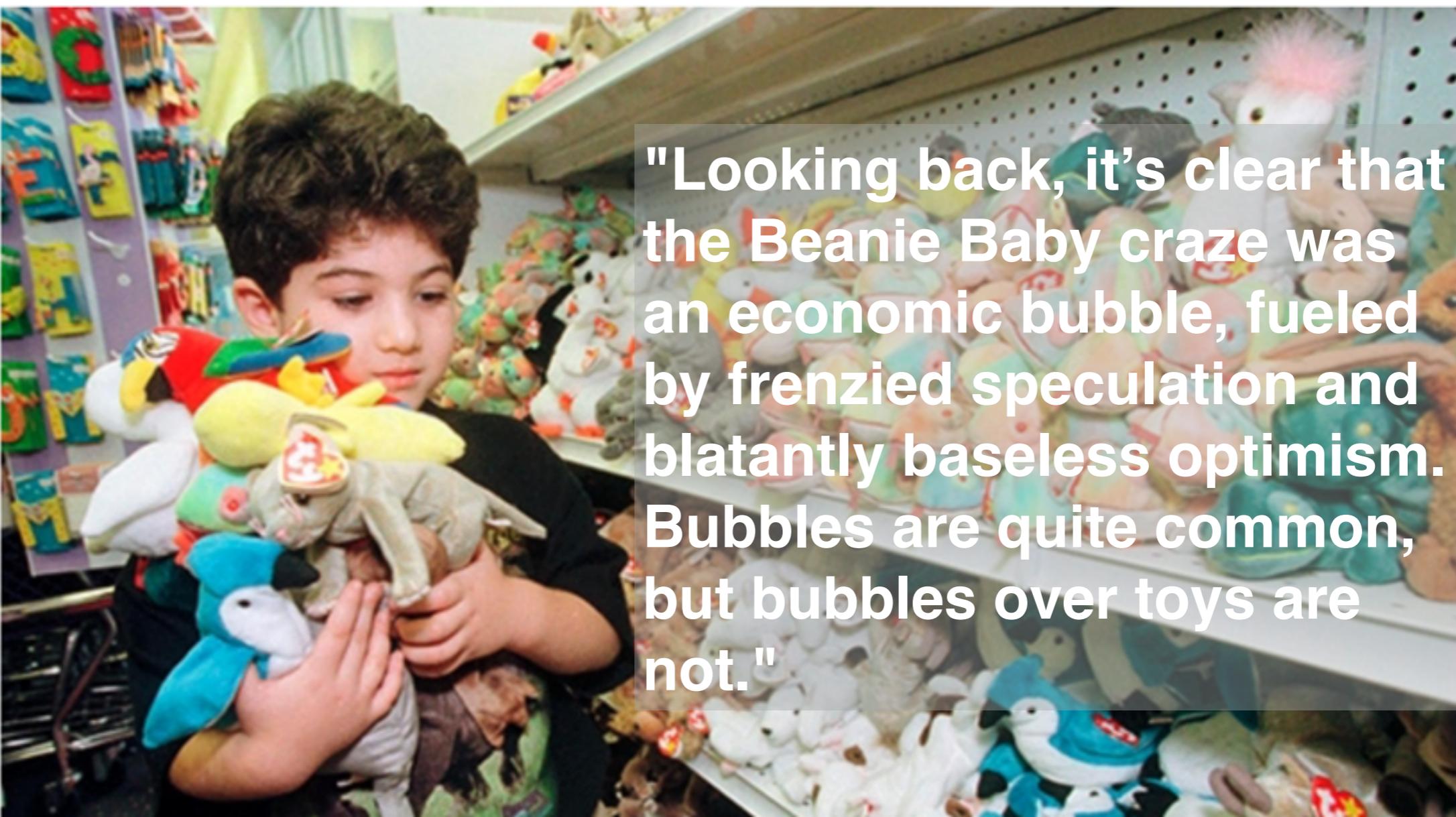
10.1k



601



By *Mark Joseph Stern*

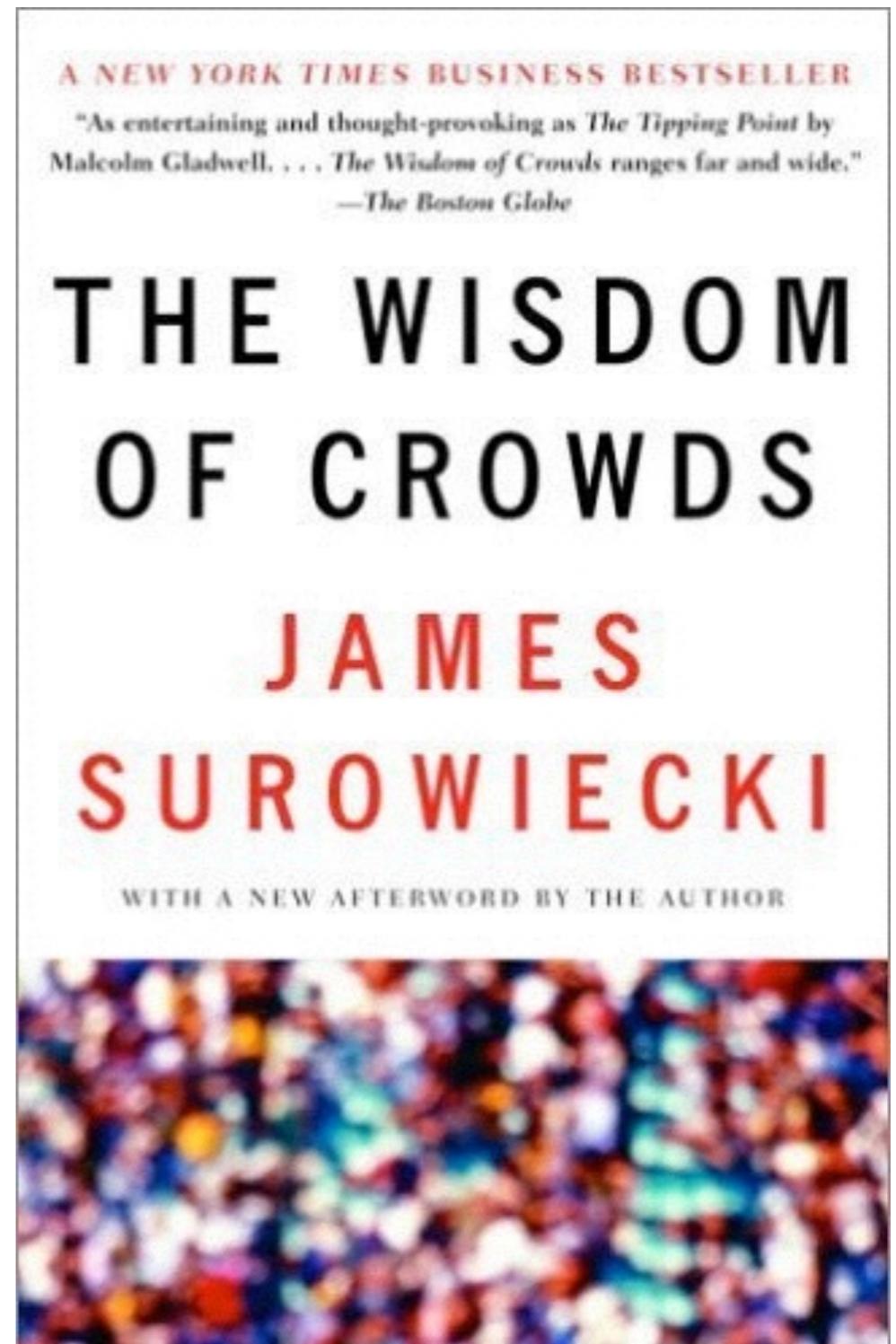


"Looking back, it's clear that the Beanie Baby craze was an economic bubble, fueled by frenzied speculation and blatantly baseless optimism. Bubbles are quite common, but bubbles over toys are not."

Wisdom of Crowds

Requirements for a crowds to be wise

- Diversity of Opinion
- Independence
- De-centralization
- Aggregation



Groups / Crowds

- Employees of a business
- Participants in a poll
- Sports fans betting on games
- Independent stock market investors
- Internet users linking to sites
- Citizens in a democracy

Ways of aggregating collective intelligence

- Point spreads / parimutuel odds
- Stock prices
- Futures contracts
- Voting
- Computer algorithms, interfaces

2010 Haitian Earthquake



Disaster Response

The maps are bad



Jan 12



Ushahidi
Robert Munro

Disaster Response

Better maps from Crowdsourcing



Disaster Response

The responders don't speak Kreyol

- Fanmi mwen nan Kafou, 24 Cote Plage, 41A bezwen manje ak dlo • My family in Carrefour, 24 Cote Plage, 41A needs food and water
- Moun kwense nan Sakre Kè nan Pòtoprens • People trapped in Sacred Heart Church, PaUP
- Ti ekipman Lopital General genyen yo paka minm fè 24 è 24 hrs. supplies • General Hospital has less than 24 hrs. supplies
- Fanm gen tranche pou fè yon ptit nan Delmas 31 • Undergoing children delivery Delmas 31

Disaster Response

Maps + Translation + Local Knowledge

Workers collaborated to find locations:

Dalila: I need Thomassin Apo please

Apo: Kenscoff Route: Lat:

18.495746829274168,

Long:-72.31849193572998

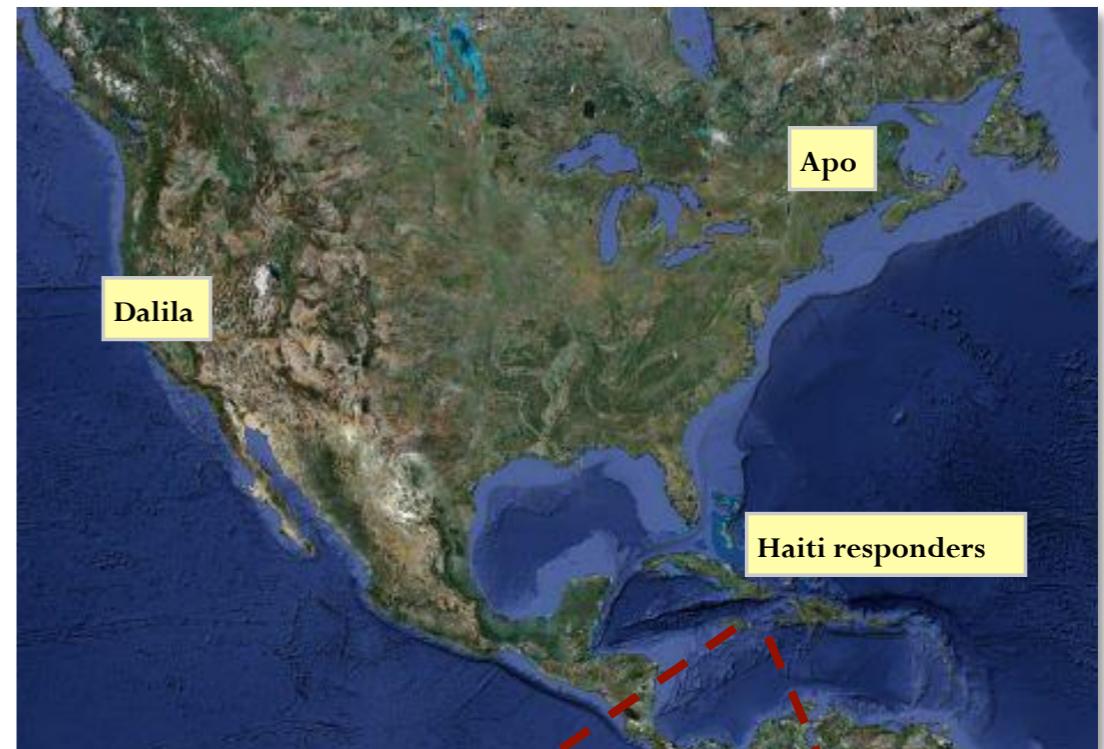
Apo: This Area after Petion-Ville and Pelerin 5 is not on Google Map. We have no streets name

Apo: I know this place like my pocket

Dalila: thank God u was here

Feedback from responders:

"just got emergency SMS, child delivery, USCG are acting, and, the GPS coordinates of the location we got from someone of your team were 100% accurate!"



Disaster Response

- Clark Craig of the Marine Corps:
 - “I cannot overemphasize to you what the work of the Ushahidi/Haiti has provided. It is saving lives every day.”
- Secretary of State Hillary Clinton:
 - “The technology community has set up interactive maps to help us identify needs and target resources. And on Monday, a seven-year-old girl and two women were pulled from the rubble of a collapsed supermarket by an American search-and-rescue team after they sent a text message calling for help.”
- Craig Fulgate, FEMA Task Force:
 - “[The] Crisis Map of Haiti represents the most comprehensive and up-to-date map available to the humanitarian community.”
- Ushahidi@Tufts :
 - “The World Food Program delivered food to an informal camp of 2500 people, having yet to receive food or water, in Diquini to a location that 4636 had identified for them.”

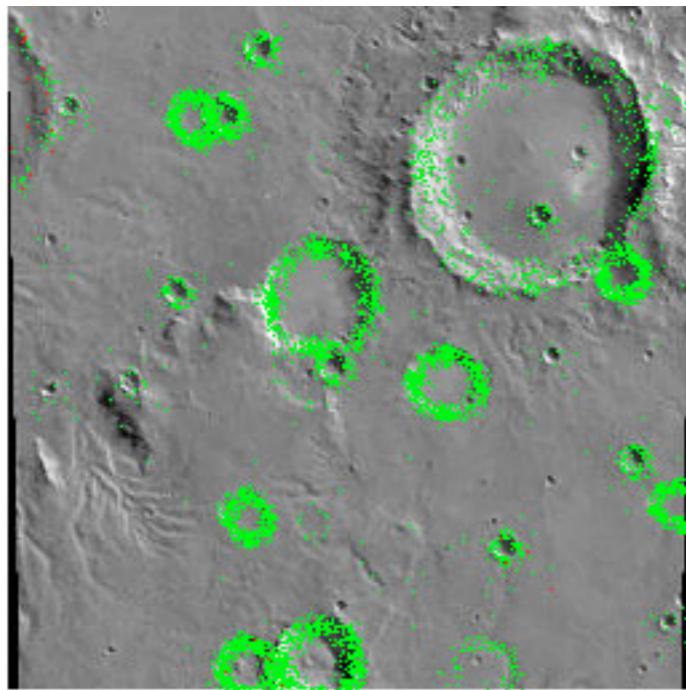
How can computer
science and economics
help facilitate collective
intelligence?

NASA Clickworkers (2000)

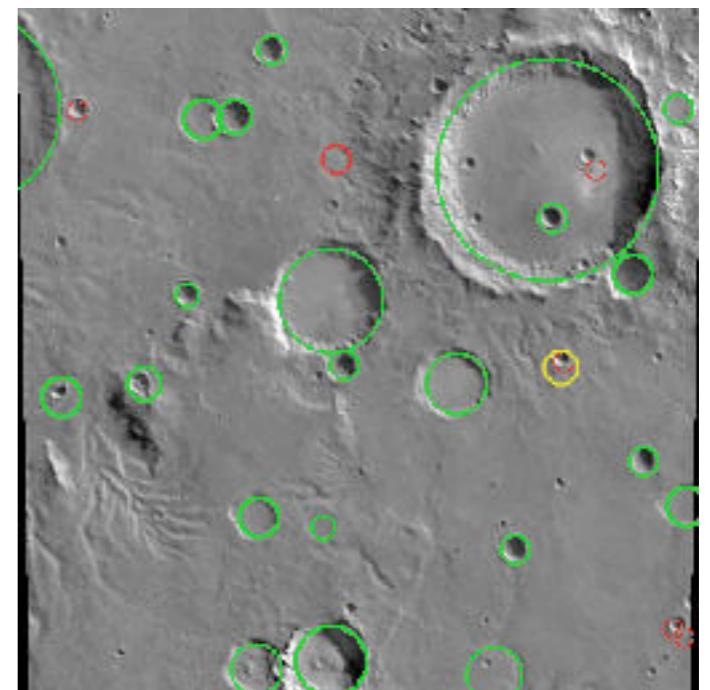
NASA showed that public volunteers could do routine science analysis that would normally be done by a graduate student working for months on end. From 11/2000 to 1/2002, they had **101,000 clickworkers** volunteering **14,000 work hours**, **612,832 sessions**, and **2,378,820 entries**!

We try to have several people cover each region on Mars so that we can compute a consensus, throwing out any mistaken or frivolous entries and averaging out the inaccuracies.

Here are all
the clicks we
received for
this region

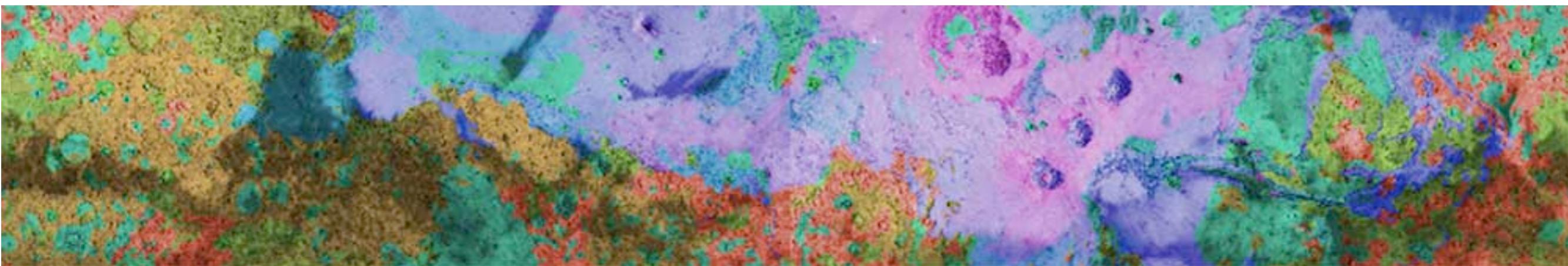


Here is the consensus

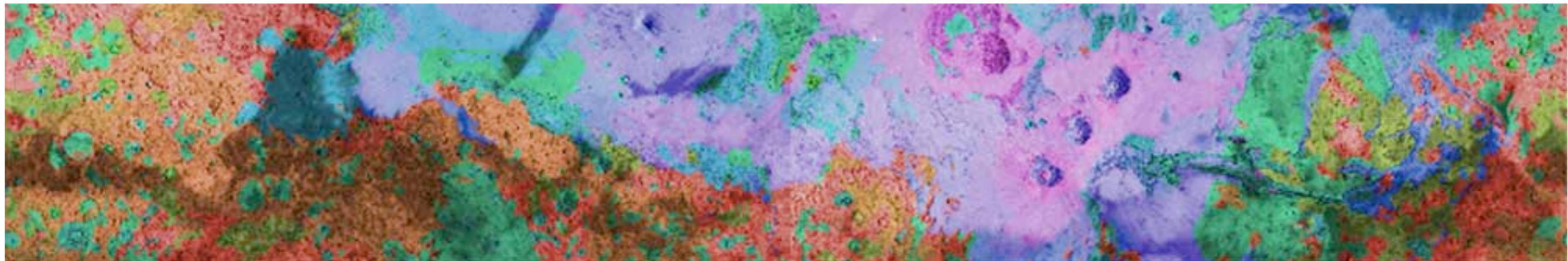


NASA Clickworkers (2000)

Mars age map produced directly from clickworker inputs.



Mars age map produced from scientists



Color guide: red=heavily cratered (old), green=medium, violet=lightly cratered (young).

[Introduction](#) | [Dashboard](#) | [Status](#) | [Account Settings](#)

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

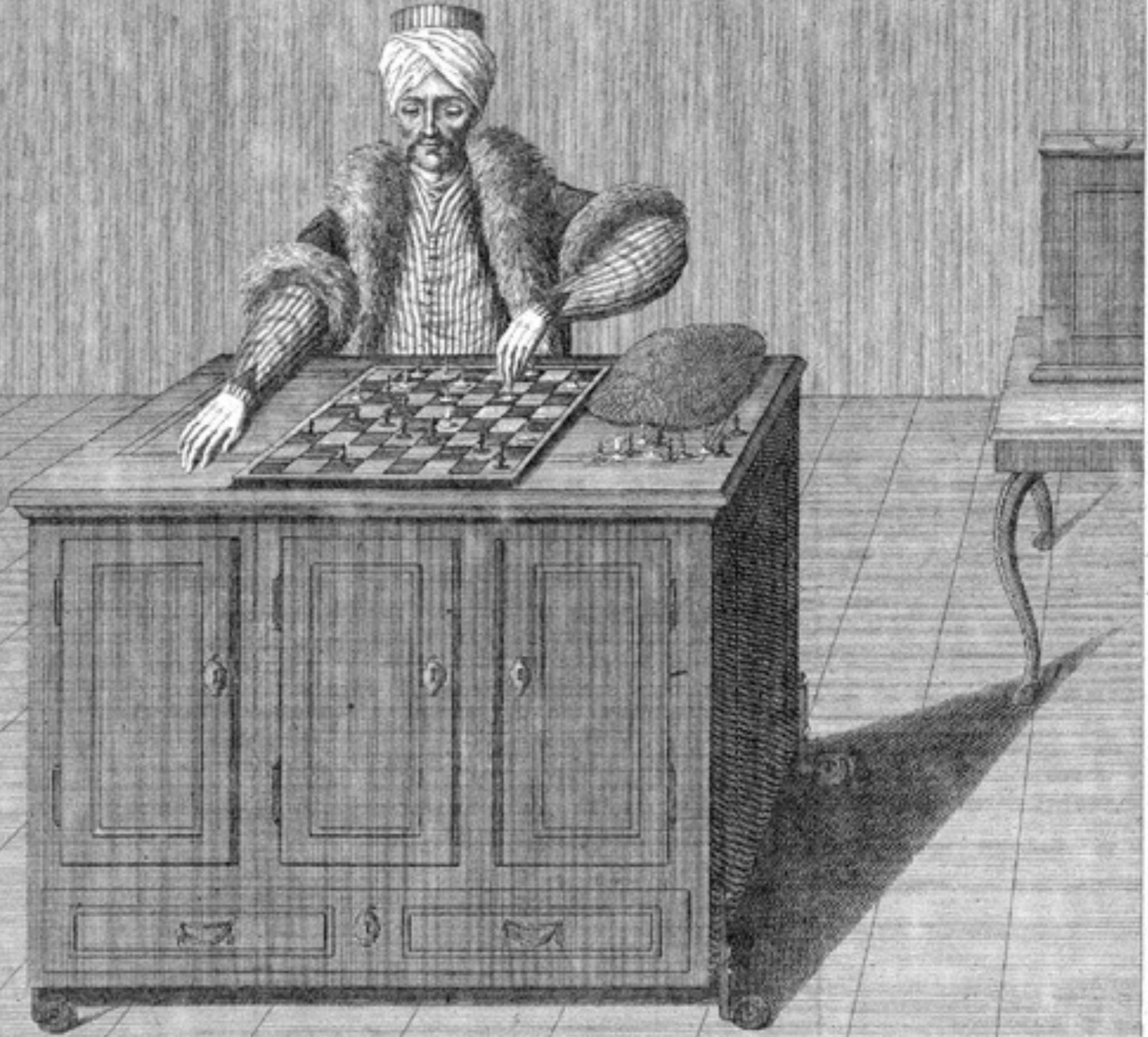


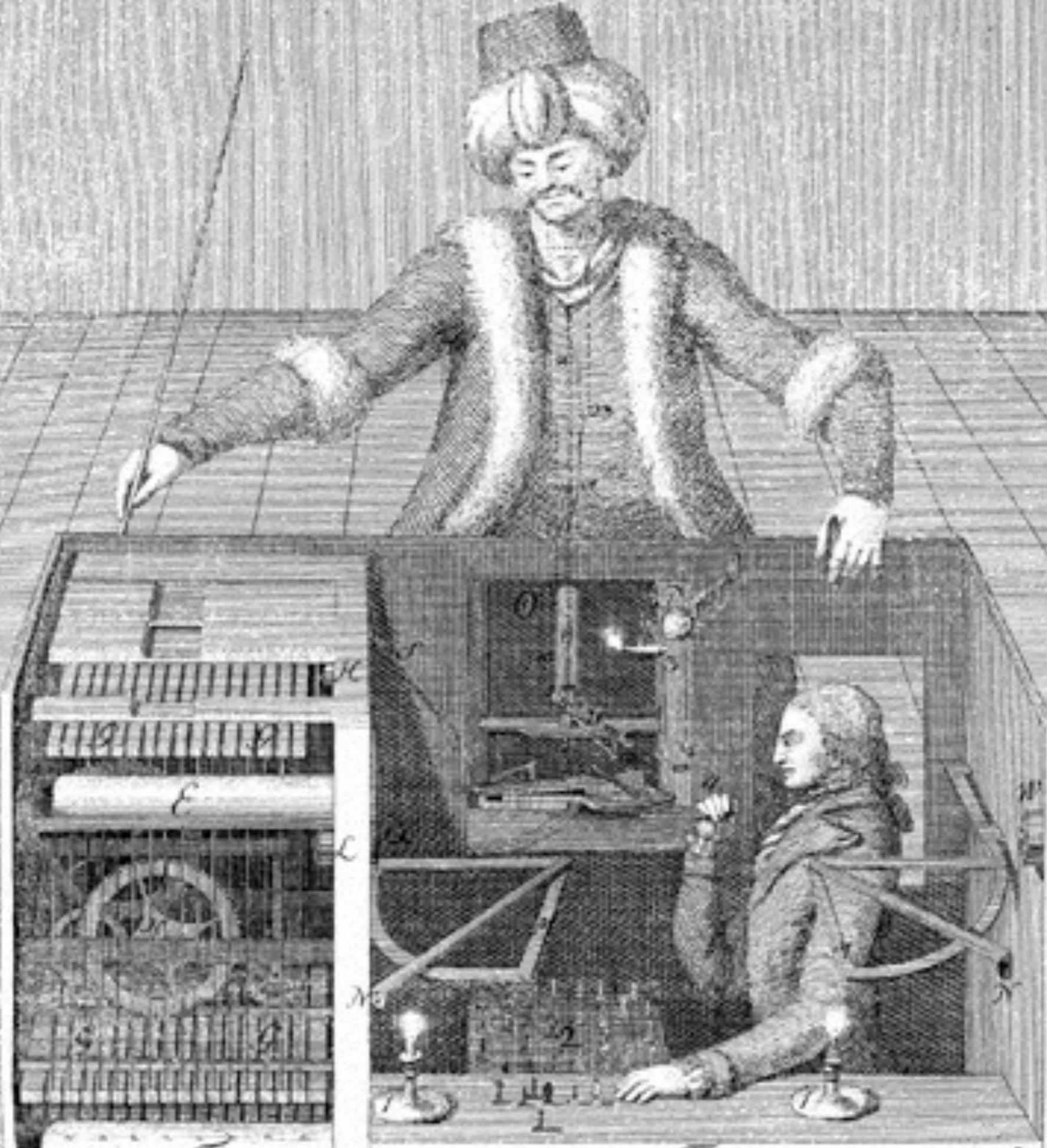
Get results



[Get Started](#)

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





Act I. Fig. 28

[All HITs](#) | [HITs Available To You](#) | [HITs Assigned To You](#)Search for **HITs** containing that pay at least \$ 0.00 for which you are qua**Timer:** 00:00:00 of 5 minutes Want to work on this HIT? Want to see other HITs?**Total Earned:** Unavaila**Accept HIT****Skip HIT****Total HITs Submitted:** 0

Enter Postmark & Stamp Information for a Postcard

Requester: Cardcow**Reward:** \$0.01 per HIT**HITs Available:** 2**Duration:** 5 minutes**Qualifications Required:** Data Entry for Postcards has been granted**Enter Postmark & Stamp Information for this card****Postmark City:****Postmark State:**
(or Country)**Postmark Date:**
(Ex: Nov-09)**Postmark Year:**
(Ex: 1909)**Stamp:** (Ex: 1c, 2c,
half penny)

(month & day)

[All HITs](#) | [HITs Available To You](#) | [HITs Assigned To You](#)Search for HITs containing that pay at least \$ 0.00 for which you are qu:Timer: 00:00:00 of 5 minutes Want to work on this HIT? [Accept HIT](#) Want to see other HITs? [Skip HIT](#)

Total Earned: Unavaila

Total HITs Submitted: 0

Enter Postmark & Stamp Information for a Postcard

Requester: Cardcow

Reward: \$0.01 per HIT

HITs Available: 2

Duration: 5 minutes

A reward has been granted

**Postmark City:** Barre**Postmark State:** MA**Postmark Date:** Oct-11**Postmark Year:** 1886**Stamp:** 1c

Postmark City:

Postmark State:
(or Country)Postmark Date:
(Ex: Nov-09)

\$ 0.01

(month & day)

[Introduction](#) | [Dashboard](#) | [Status](#) | [Account Settings](#)

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

Created scenarios...
global supply chains...
distinctively enable rea...
bods **TASKS** after...
empowerment effectiv...
formation. Globally...
size adaptiv...

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](#)

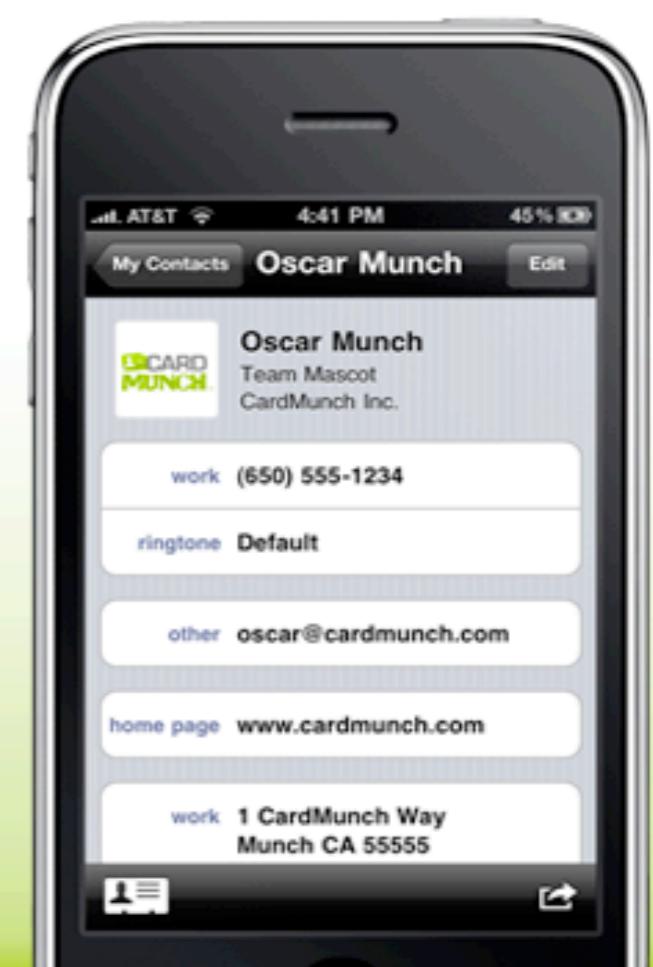
Business Cards Solved

step 1
point & capture



step 2
real people

GET STARTED



Simple



Accurate



Reliable

step 3
contact done



TRANSCRIPTION MADE **FAST** AND **EASY**.

3 easy steps

UPFRONT PRICING, NO ACCOUNT NEEDED



BUDGET
\$1.00/minute



6 DAY
\$1.50/minute



1 DAY
\$2.50/minute

CASTINGWORDS TRANSCRIPTION SERVICES:

QUALITY IS KING

We catch the difficult words, do the research, and produce rocking transcripts.

CLEAR PRICING

We price per minute of audio - so you know how much you'll be paying for your transcription - now. No surprises.

SIMPLE TRANSCRIPT DELIVERY

Transcripts are emailed to you - and can be picked up here at the site.

LOTS OF FORMATS

Transcripts are delivered in plain text, HTML, and MS Word formats.

HUMAN TRANSCRIPTION

CastingWords' Budget, Express and Expedited products are transcribed, edited, and reviewed multiple times, by people, never machines.

LIVE ORDER TRACKING

You always know what is happening with your audio transcription order.

We were constantly surprised that phrases or names that other companies missed, from "*tabula rasa*" to "*turducken*" (it's a Cajun specialty), were identified and spelled correctly.

WALL STREET JOURNAL

NEWS & UPDATES

AUGUST 12 2013: Google Checkout is closing, and we're going to remove it as a payment

STEP 1 OF 3

CHOOSE A PRODUCT BELOW

Help African Refugees

you send
samasource a
project

1



samasource
breaks it down
into microwork

2



work is allocated to
our service partners

3



women, youth, and
refugees complete
work

4



samasource
compiles work and
assures quality

5



your project gets
delivered and helps
reduce poverty

6



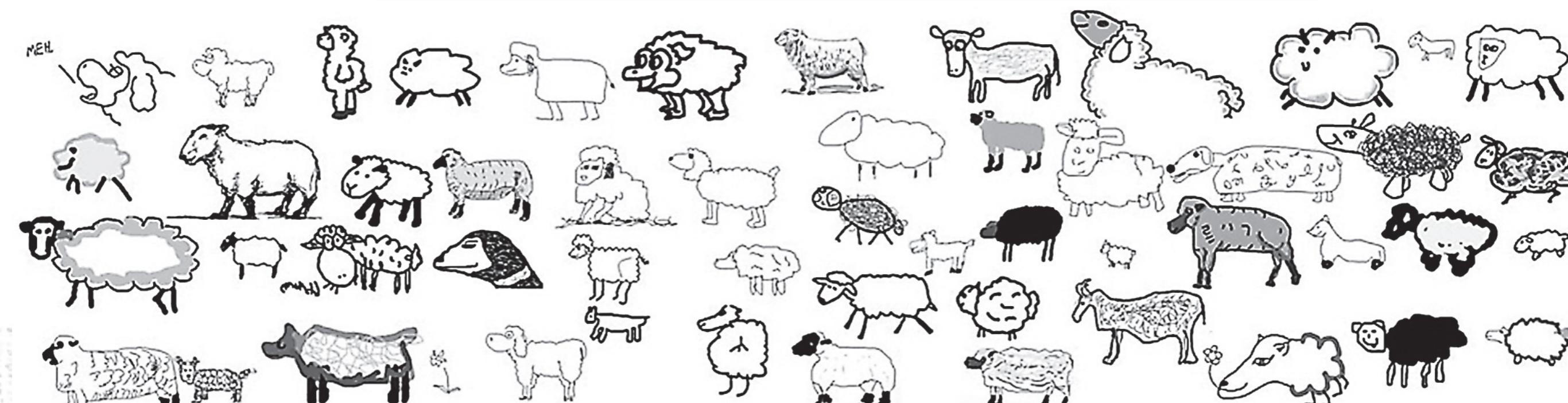
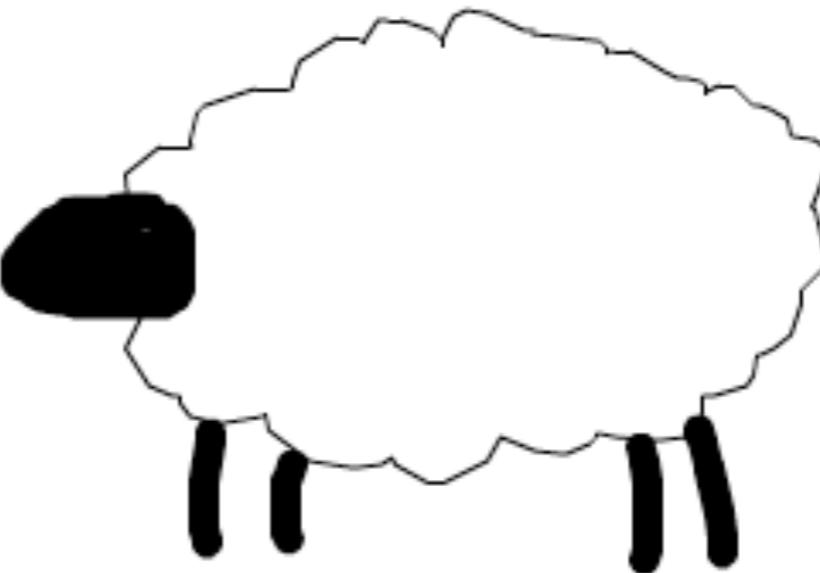
choose the right word

Dolores Labs

THE SHEEP MARKET

10,000 sheep created
by online workers.
[More...](#)

Catch some zzzzs



Improve your sex life

most
exciting

least
exciting

Fleshmap

Dark Side of Crowdsourcing

(J Zittrain, “Minds for Sale”)

- Surveillance
- Alienation and Deskilling
- Moral Valence
- Race to the Bottom

Iranian protesters appearing in widely disseminated online photos from the ongoing post-election demonstrations in Iran, are now being targeted on a website of the Islamic Revolutionary Guard Corps.

The website, called 'vortex'... shows images of 20 people with red circles drawn around their faces claiming without evidence that they have been involved in creating "chaos" in Tehran.

Citizens are invited to call or email if they can identify the people on the photos. Gerdab also claims that two of the people depicted have already been arrested. The site provides no further information about any of the depicted people.

Bill Maher on the Sharing Economy



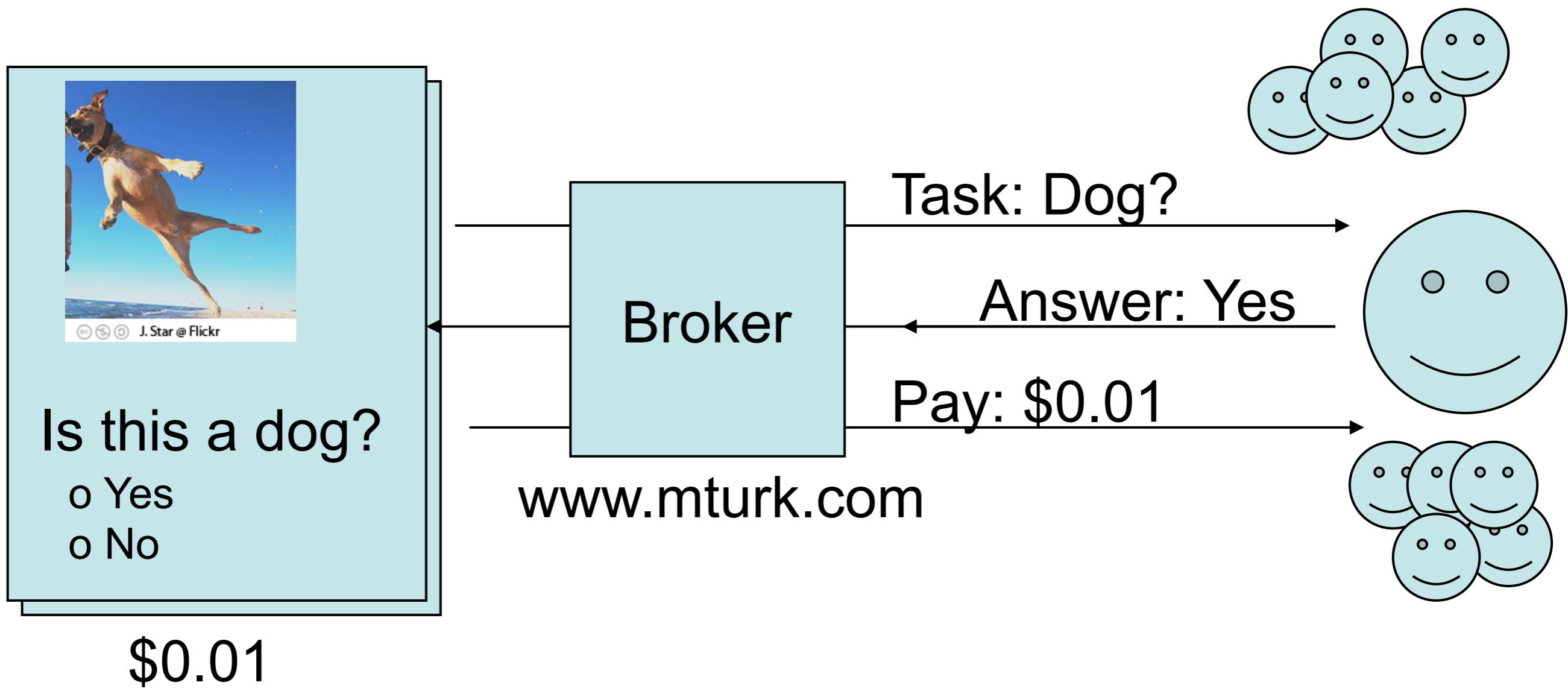
Real Time with Bill Maher: The "Sharing" Economy – August 21, 2015 (HBO)

How to put
crowdsourcing
towards good uses

Use of MTurk-like systems in research

- Participant pool for user studies, polling, cognitive science experiments
- Annotation for machine learning tasks like computer vision or NLP
- Human Computer Interaction: worker pools are hardwired into the UI
- New Programming Languages Concepts
- Study markets themselves for economics research, cost-optimization

Annotation for machine learning / artificial intelligence tasks



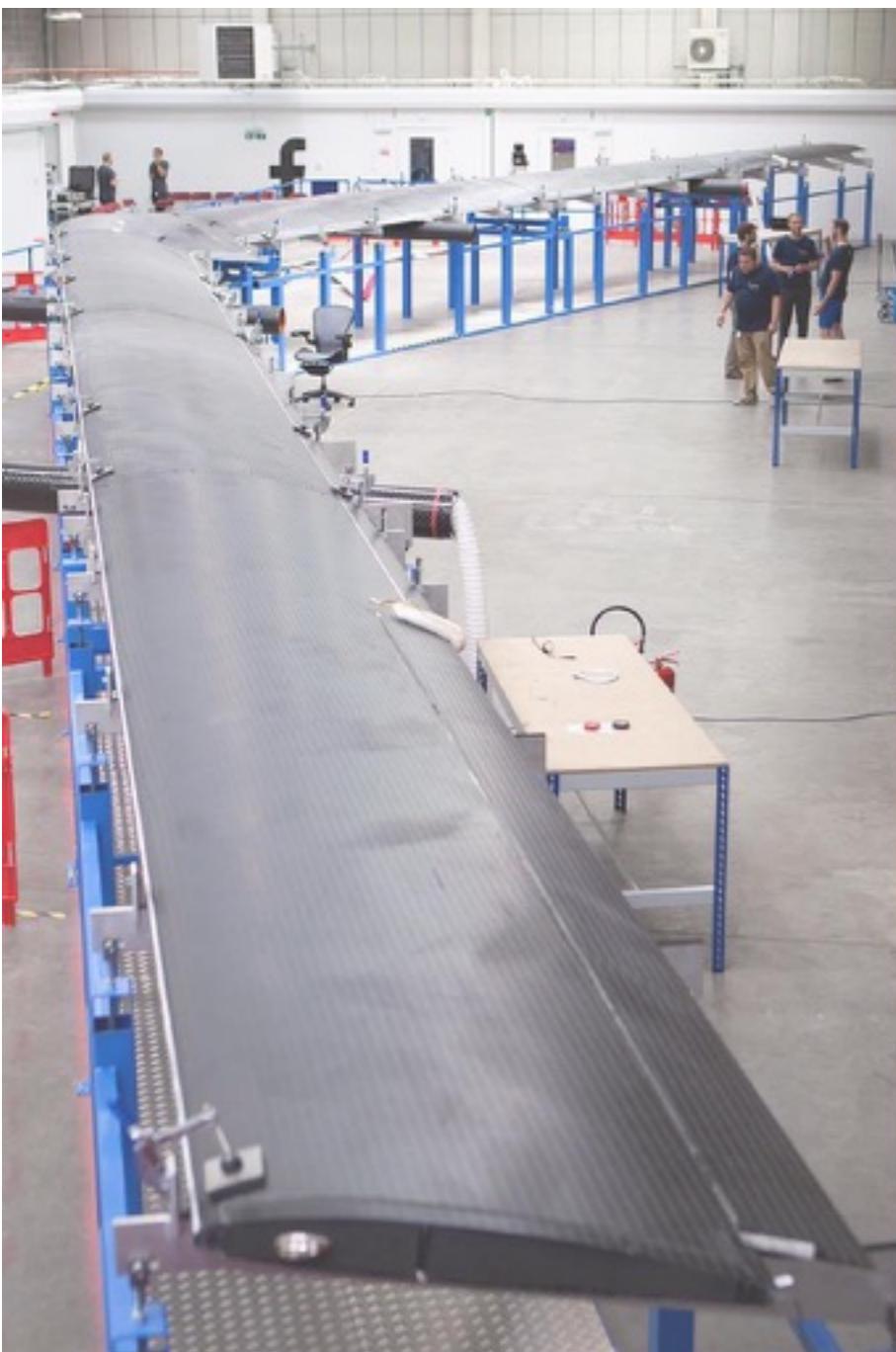


Mark Zuckerberg

Follow

January 6 at 4:23pm · Palo Alto ·

2016 is a big year for launching the future technology we've been building.



Internet.org



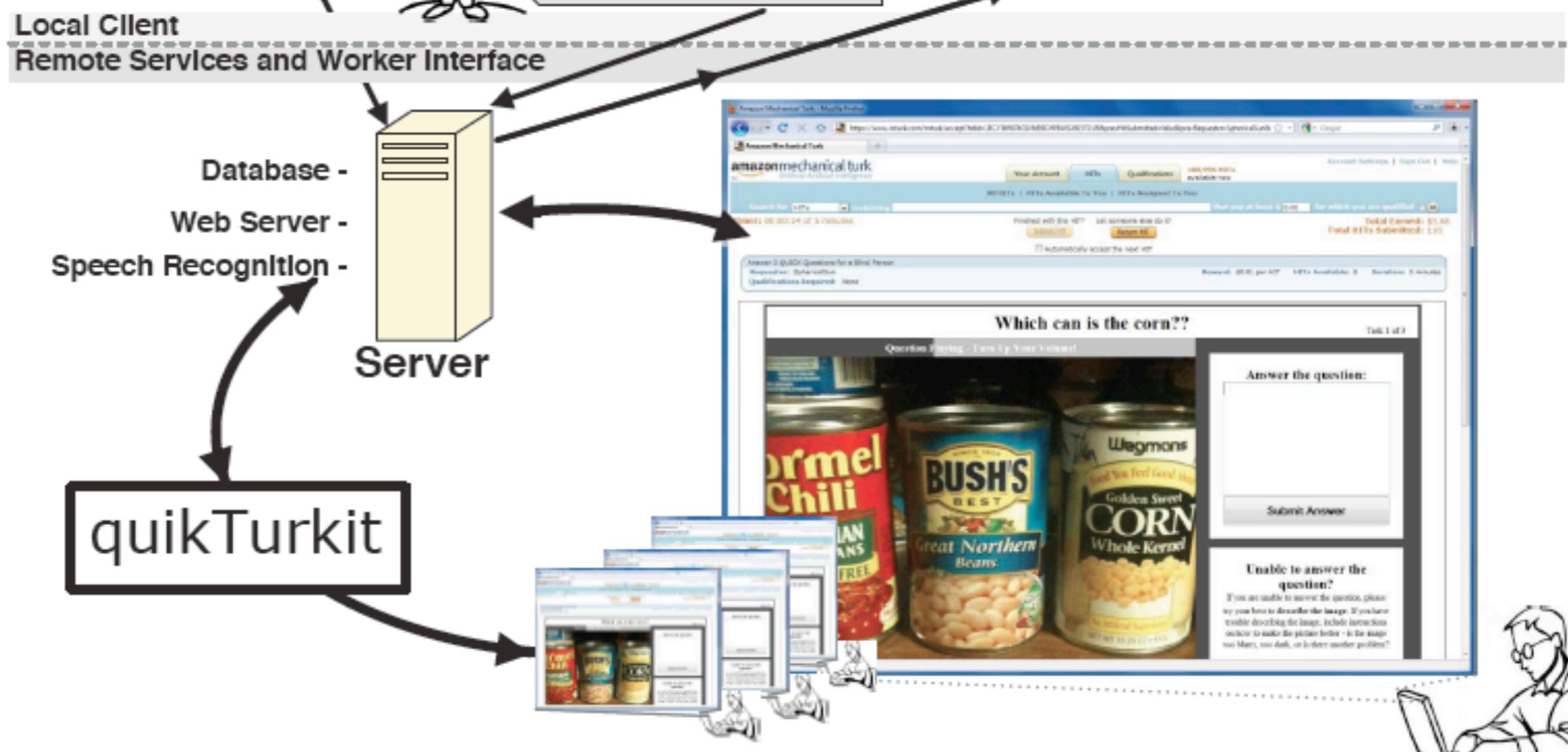
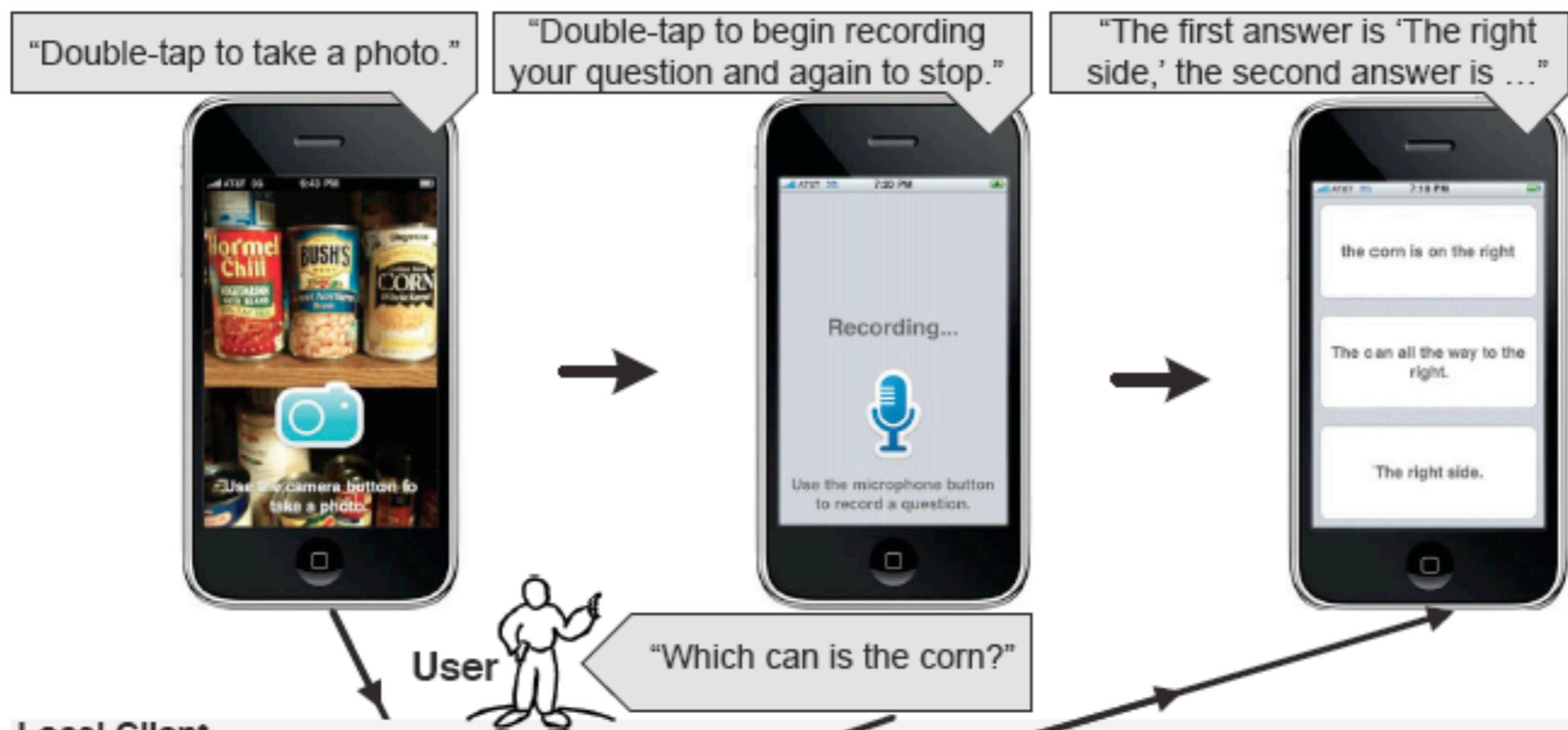
AI



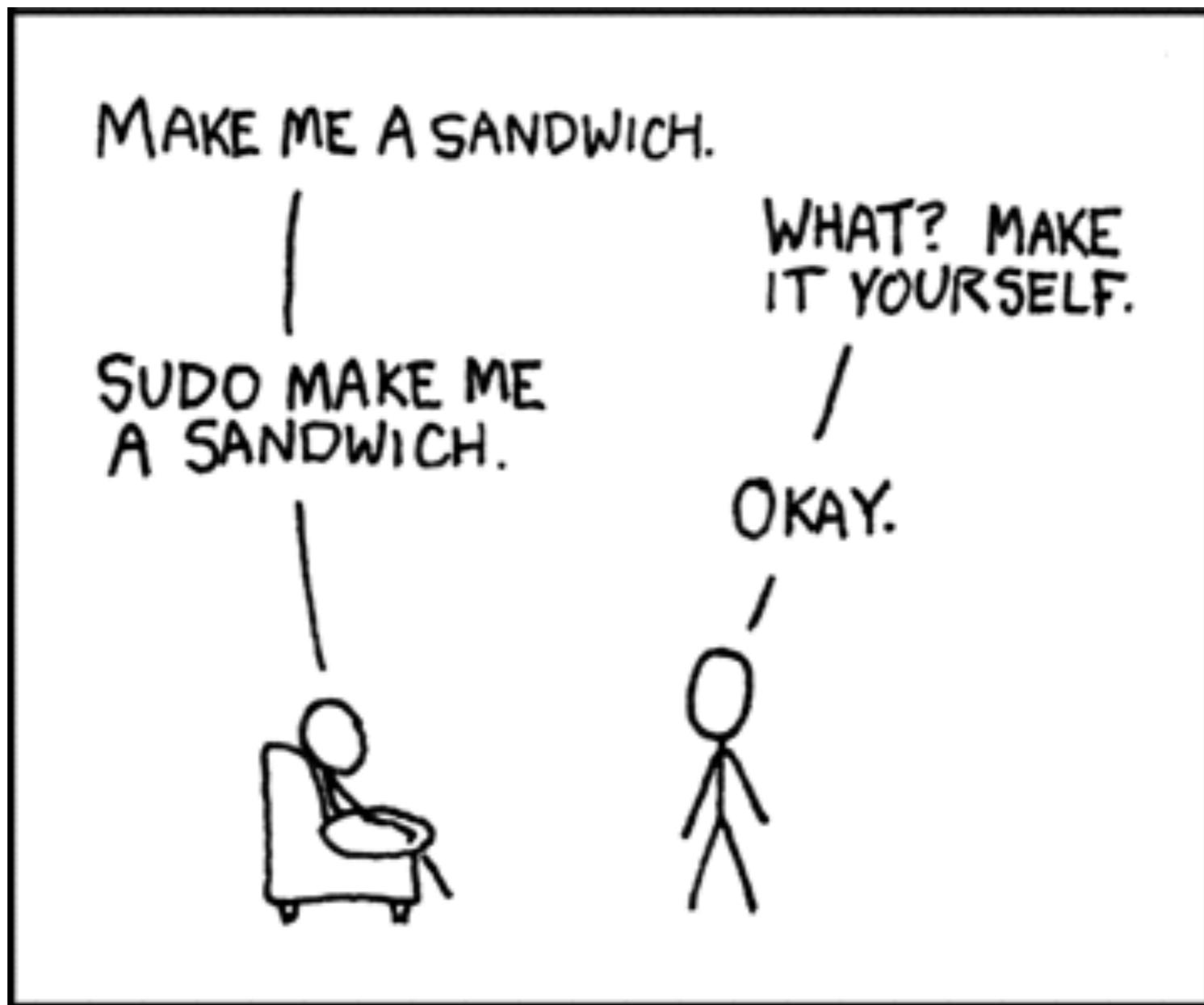
Oculus

Human Computer Interaction

- VizWiz - best paper award at UIST 2010
 - Mobile service that aids blind users with “visual questions” in near-realtime
 - https://youtu.be/EtfriMJQ0_M?t=32



New Programming Languages Concepts



New Programming Languages Concepts

```
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

Study Markets Themselves

- What predictions of economics hold true on MTurk?
- What incentives can we give to increase throughput, quality, worker retention?
- What is the cost-optimal solution to a problem?

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

Topics

- Taxonomy of crowdsourcing and human computation
- Crowdsourcing platforms like Mechanical Turk and CrowdFlower
- Programming concepts for human computation
- The economics of crowdsourcing
- Crowdsourcing and machine learning
- Applications to human computer interaction
- Crowdsourcing and social science
- Data mining from Twitter

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

Course Requirements

Weekly assignments

Writing and Coding

Video presentations

Company profile,
project pitch

Final project

Self-designed,
groups of 3-4

Final presentation

Show off your work

How much programming is required?

- Programming assignments are in Python
- We provide code that you modify
- We want everyone to be able to participate, regardless of programming experience
- For most assignments, you can work with a partner (turn in only one assignment - you'll both get the same grade)

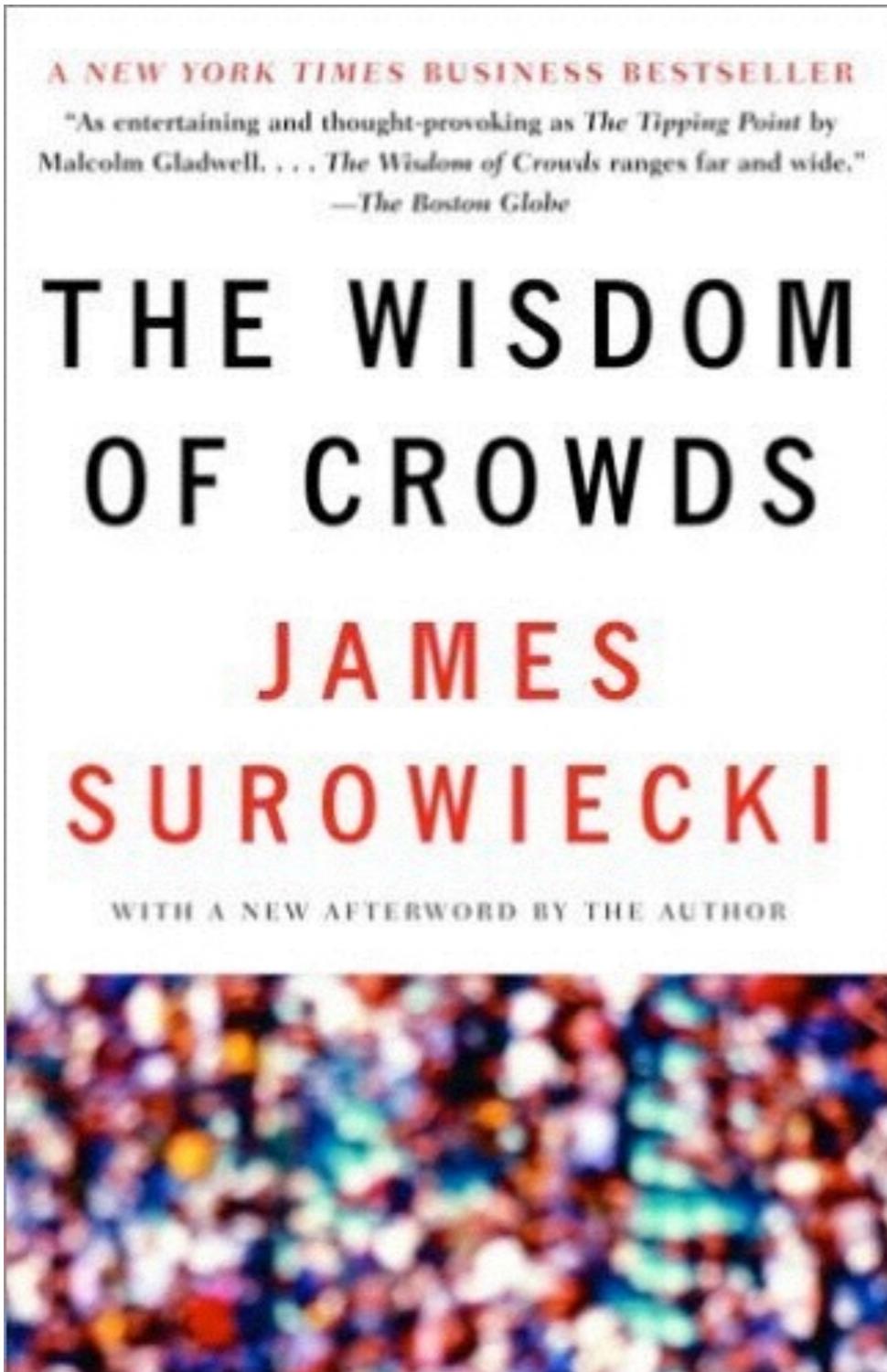
Programming assignments: Gun Violence Database

- The programming assignments form a sequence
 - Build a machine learning classifier to identify newspaper articles that describe gun violence
 - Have crowd-workers verify its predictions
 - Have crowd-workers extract structured information from the text of the articles
 - Analyze the data and build visualizations about gun violence in the USA

Written assignments

- Become a Crowd Worker and analyze your experience
- Write a company profile for a company that uses crowdsourcing
- Analyze the text of the gun violence articles to see if there is a difference between how different types of shootings are described

Reading assignments



We will be using *The Wisdom of Crowds* as the course reader, and supplementing it with readings from academic papers.

You will have to replicate one academic paper.

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 by companies and countries
- Summer internships?
- \$10,000?

Who are we?



Chris Callison-Burch

- Assistant Professor in CIS, my 3rd year here at Penn
- Undergrad from Stanford in Symbolic Systems
- PhD from University of Edinburgh in Informatics
- Research Faculty at Johns Hopkins University
- Research Interests: Crowdsourcing, Natural Language Processing
- Office: Levine 506, Email: ccb@cis.upenn.edu
Twitter: @ccb

Leaderboard

Mechanical Turk Monitor: Top-1000 Recent Requesters

	General	Top requesters	Arrivals	Completed	Search	About		
	Requester ID	Requester			#Task	#HITs	Rewards	
1	Dolores Labs	A2IR7ETVOIULZU (RSS)			470	317857	33084.24	
2	ContentGalore	A2XL3J4NH6JI12 (RSS)			674	10622	17554.6	
3	SpeechInk	A1AQ7EJ5P7ME65 (RSS)			9019	13613	12876.67	
4	CastingWords	A3MI6MIUNWCR7F (RSS)			9454	14036	8947.02	
5	QuestionSwami	AD7C0BZNKYGYV (RSS)			629	4116	4750.37	
6	Chris Callison-Burch	A32TTE4XXN6MQZ (RSS)			11	9961	4458.02	
7	Smartsheet.com Clients	A1197OGL0WOQ3G (RSS)			434	38212	3118.28	
8	retaildata	AD14NALRDOSN9 (RSS)			8	50288	3110.85	
9	Classify This	A1CTI3ZAWTR5AZ (RSS)			25	94590	1891.8	
10	Andrew Stephen	A1Y25F6MZCMQGY (RSS)			3	22705	1131.25	
11	Dolores Labs 2	A3JX8WONBL5N9X (RSS)			34	8976	1043.27	
12	RelevanceQuest	A8RMEN71ICE57 (RSS)			15	47881	1029.92	
13	Crowd Task	AFAOUHS65HNDS (RSS)			4	2388	955.6	
14	nlds.soe.ucsc.edu	A1HI9DWCF794RE (RSS)			4	4702	933.9	



Ellie Pavlick

- 4th Year PhD student in CIS
- Undergrad from JHU in Economics and in Music, with minor in Computer Science
- Thesis is on artificial intelligence for understanding natural language
- Research into crowdsourcing: translation, ESL error correction
- email: elliepavlick@gmail.com

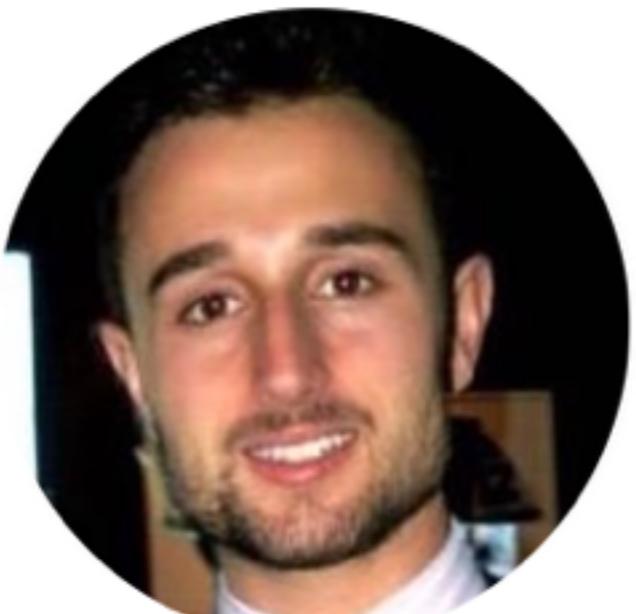
Teaching Assistants



Dhrupad Bhardwaj

Jim Chen

Sally Kong



Ross Mechanic

Kate Miller

Sierra Yit

Sign up to be a Mechanical
Turk Worker

Read “Wisdom of Crowds”
Intro plus Chapter 1

crowdsourcing-class.org