Wisdom of the Crowd

CS 278 | Stanford University | Michael Bernstein

Last time

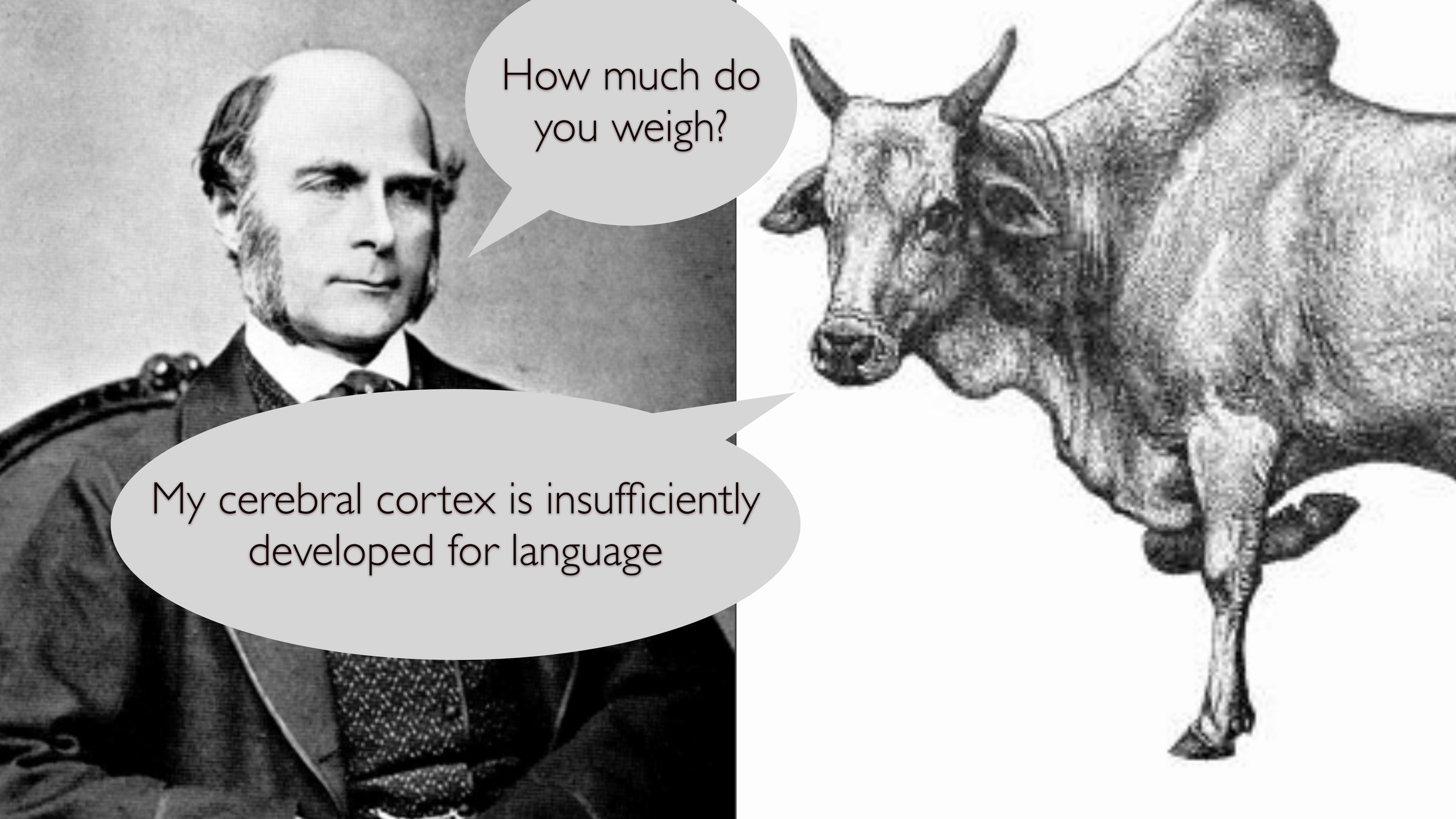
Our major units thus far:

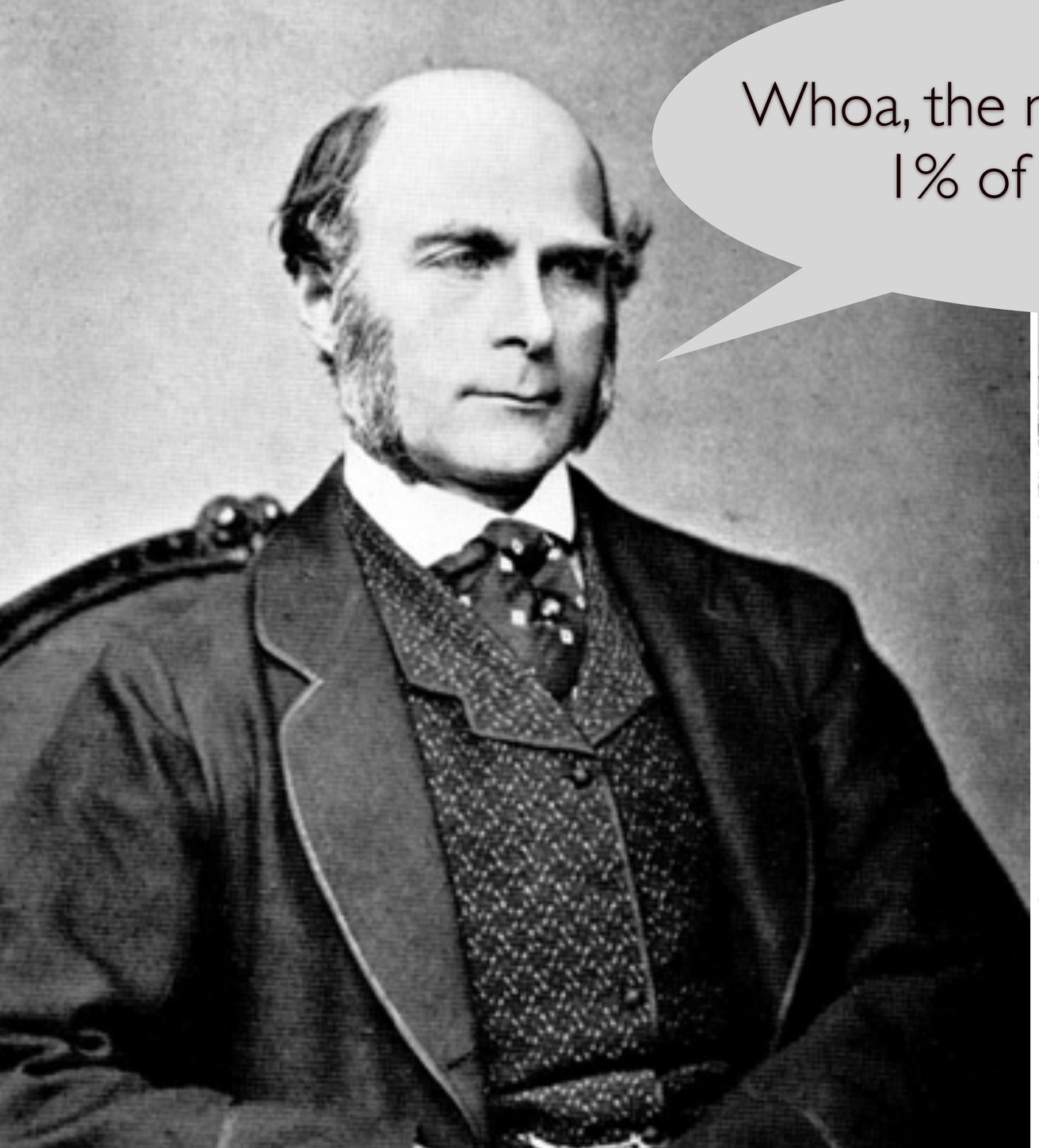
- Basic ingredients: contribution and norms
- Scales: starting small, and growing large
- Groups: strong ties, weak ties, and collaborators

Now: massive-scale collaboration

http://hci.st/wise

Grab your phone, fill it out!


How much do you weigh?
My cerebral cortex is insufficiently developed for language



Whoa, the mean guess is within 1% of the true value

of the dressed weight of a 787 different persons.

		ntiles		
		Observed deviates from 1207 lbs.	Normal p.e =37	Excess of Observed over Normal
5	1074	−133	−90	+43
10	1109	− 98	−70	+28
15	1126	− 81	−57	+24
20	1148	− 59	−46	+13
q_1 25	1162	− 45	−37	+ 8
30	1174	− 33	−29	+ 4
35	1181	− 26	−21	+ 5
40	1188	− 19	−14	+ 5
45	1197	− 10	− 7	+ 3
m 50	1207	0	0	0
55	1214	+ 7	+ 7	0
60	1219	+ 12	+14	− 2
65	1225	+ 18	+21	− 3
70	1230	+ 23	+29	− 6
q_3 75	1236	+ 29	+37	− 8
80	1243	+ 36	+46	−10
85	1254	+ 47	+57	−10
90	1267	+ 52	+70	−18
95	1293	+ 86	+90	− 4

q_1, q_3, the first and third quartiles, stand at 25° and 75° respectively.
m, the median or middlemost value, stands at 50°.

Innovation competitions for profit

kaggle

Kaggle is the place to do data science projects

See how it works

Innovation competitions for science

NASA TV

Search

NASA Solve

Welcome to NASA Solve!

Interested in helping NASA solve tough problems? You are in the right place! This one-stop-shop website is where you'll find opportunities to participate in challenges, prize competitions, and citizen science activities that develop solutions for problems related to NASA's mission. So, jump in and become a NASA SOLVE-r!

TAP

Prediction markets

AI data annotation at scale

Today

What is the wisdom of the crowd? What is crowdsourcing?

Why do they work?

When do they work?

Wisdom of the crowd

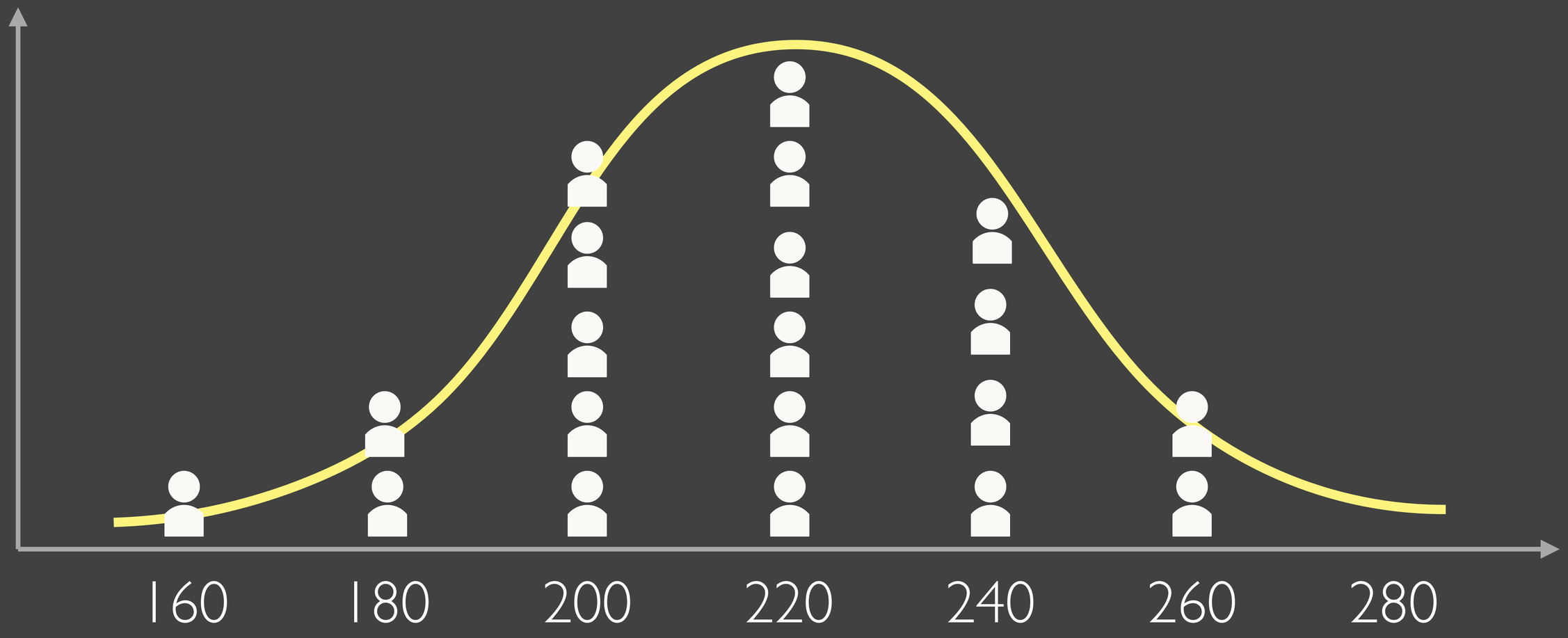
Crowds are surprisingly accurate at estimation tasks

Who will win the election? How many jelly beans are in the jar? What will the weather be? Is this website a scam?

Individually, we all have errors and biases.

However, in aggregate, we exhibit surprising amounts of collective intelligence.

"Guess the number of minutes it takes to fly from Phoenix, AZ to Detroit, MI."



If our errors are distributed at random around the true value, we can recover it by asking enough people and aggregating.

What problems can be solved this way?

Jeff Howe theorized that that it required:

- Diversity of opinion
- Decentralization
- Aggregation function

So — any question that has a binary (yes/no), categorical (e.g., win/lose/tie), or interval (e.g., score spread on a football game) outcome

What problems cannot be solved this way?

Flip the bits!

- People all think the same thing
- People can communicate
- No way to combine the opinions

For example, writing a short story (is much harder!)

General algorithm

1. Ask a large number of people to answer the question

   Answers must be independent of each other — no talking!

   People must have at least basic understanding of the phenomenon in question.

2. Average their responses

Why does this work?

[Simoiu et al. 2017]

Independent guesses minimize the effects of social influence

- Showing consensus cues such as the most popular guess lowers accuracy
- If initial guesses are inaccurate and public, then the crowd never recovers

Crowds are more consistent guessers then experts

- In an experiment, crowds are only at the 67th percentile on average per question…
- But at the 90th percentile averaged across questions per domain!

Mechanism: ask many independent contributors to take a whack at the problem, and reward the top contributor

kaggle

Kaggle is the place to do data science projects

See how it works

NASA TV

Search

NASA Solve

Welcome to NASA Solve!

Interested in helping NASA solve tough problems? You are in the right place! This one-stop-shop website is where you'll find opportunities to participate in challenges, prize competitions, and citizen science activities that develop solutions for problems related to NASA's mission. So, jump in and become a NASA SOLVE-r!

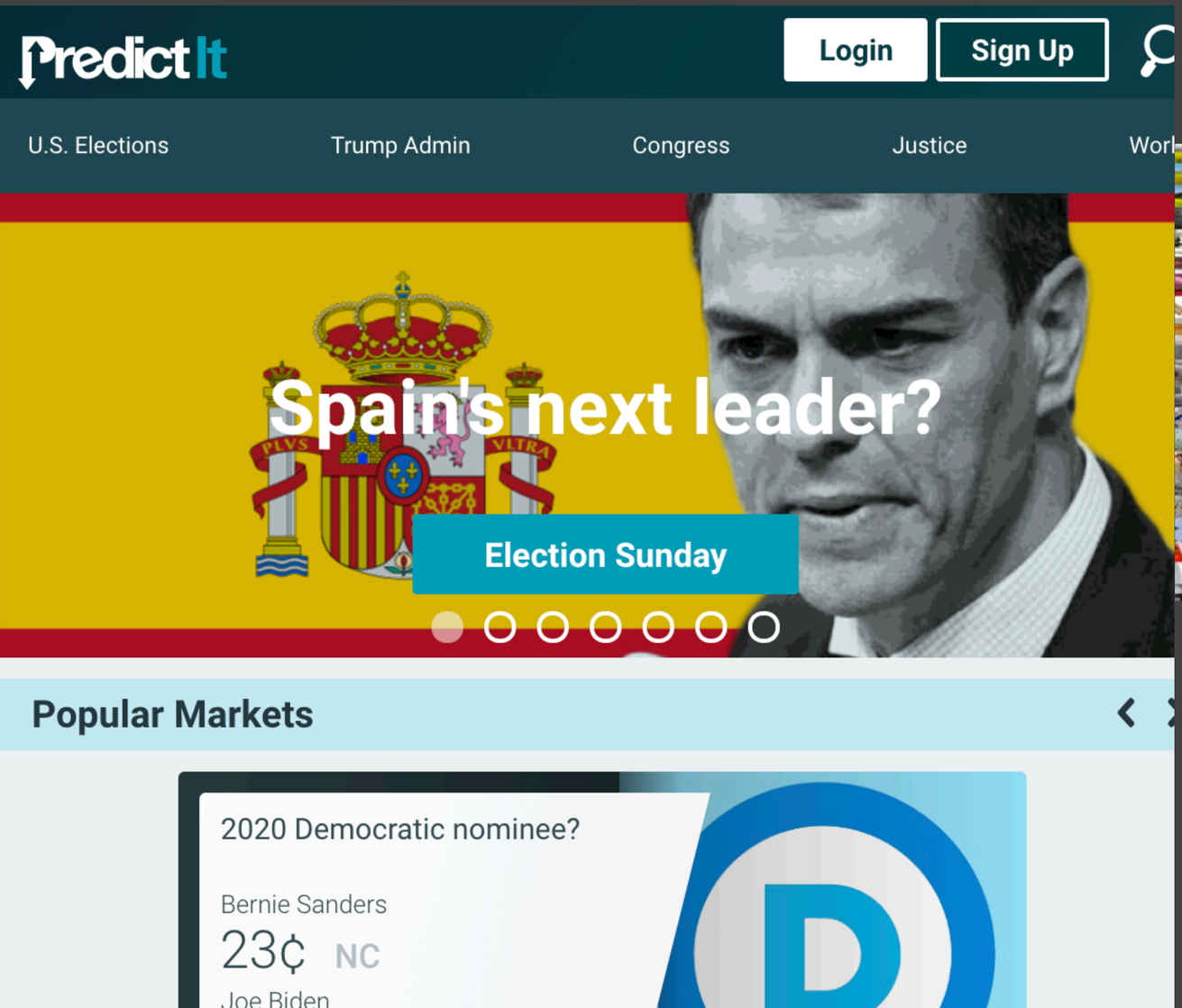
TAP

NASA Solve

Watch later

Share

Mechanism: use a market to aggregate opinions



Mechanism: ask paid data annotators to label the same image and look for agreement in labels

Let’s check our
http://hci.st/wise results

Aggregation approaches

Early crowdsourcing

[Grier 2007]

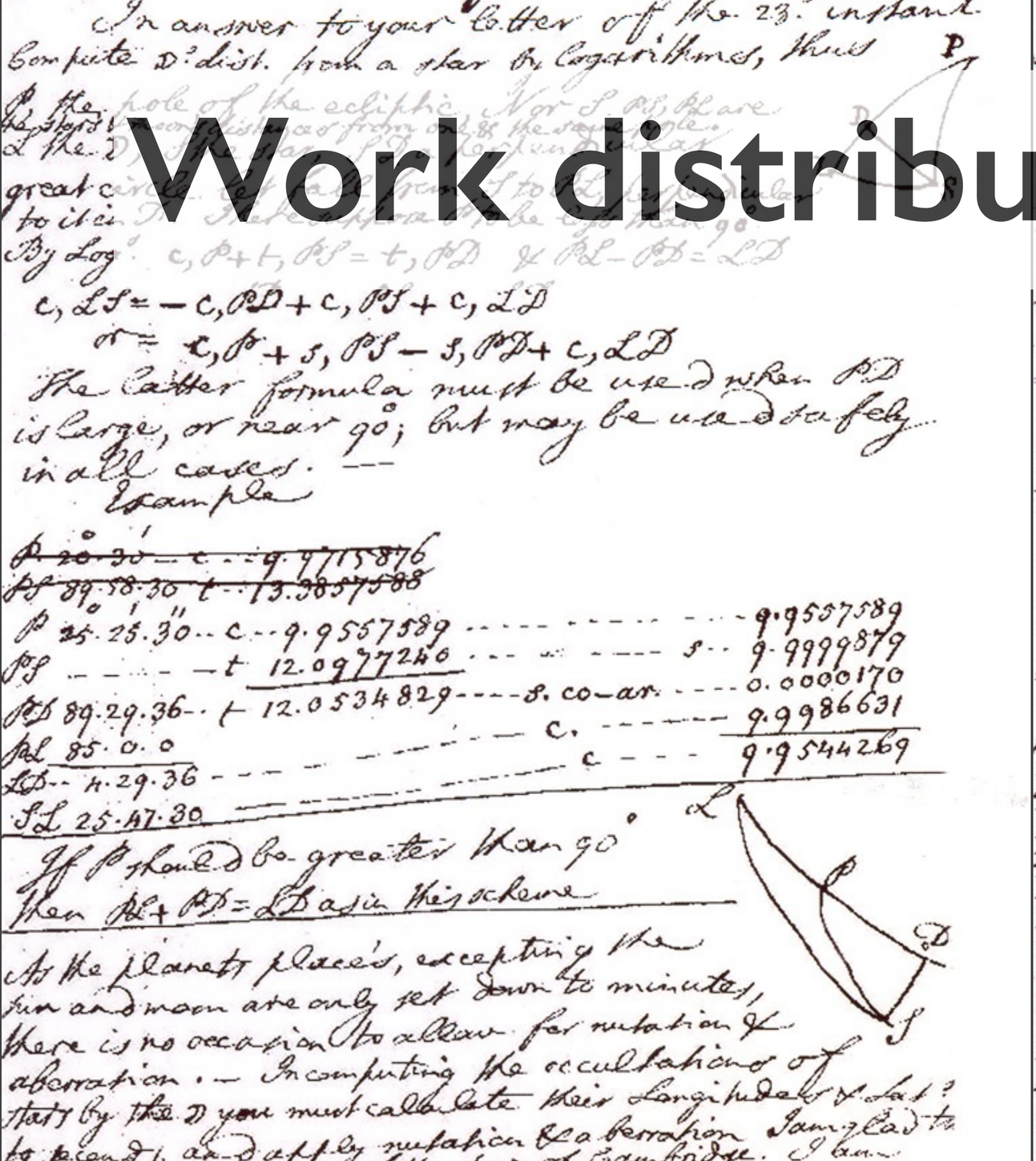
Two distributed workers work independently, and a third verifier adjudicates their responses

1760

British Nautical Almanac

Nevil Maskelyne

Work distributed via mail



In answer to your letter of the 23d instant
Compute D's dist. from a star by Logarithms, thus

P

P the pole of the ecliptic. Nor S PS, PL are
the spherical distances from the star & the
the D
great circles drawn through the star
to it in
By Log

$c, LS = -c, PD + c, PS + c, LD$

or $= c, P + s, PS - s, PD + c, LD$

The latter formula must be used when PD is large, or near 90°; but may be used safely in all cases. —

Example

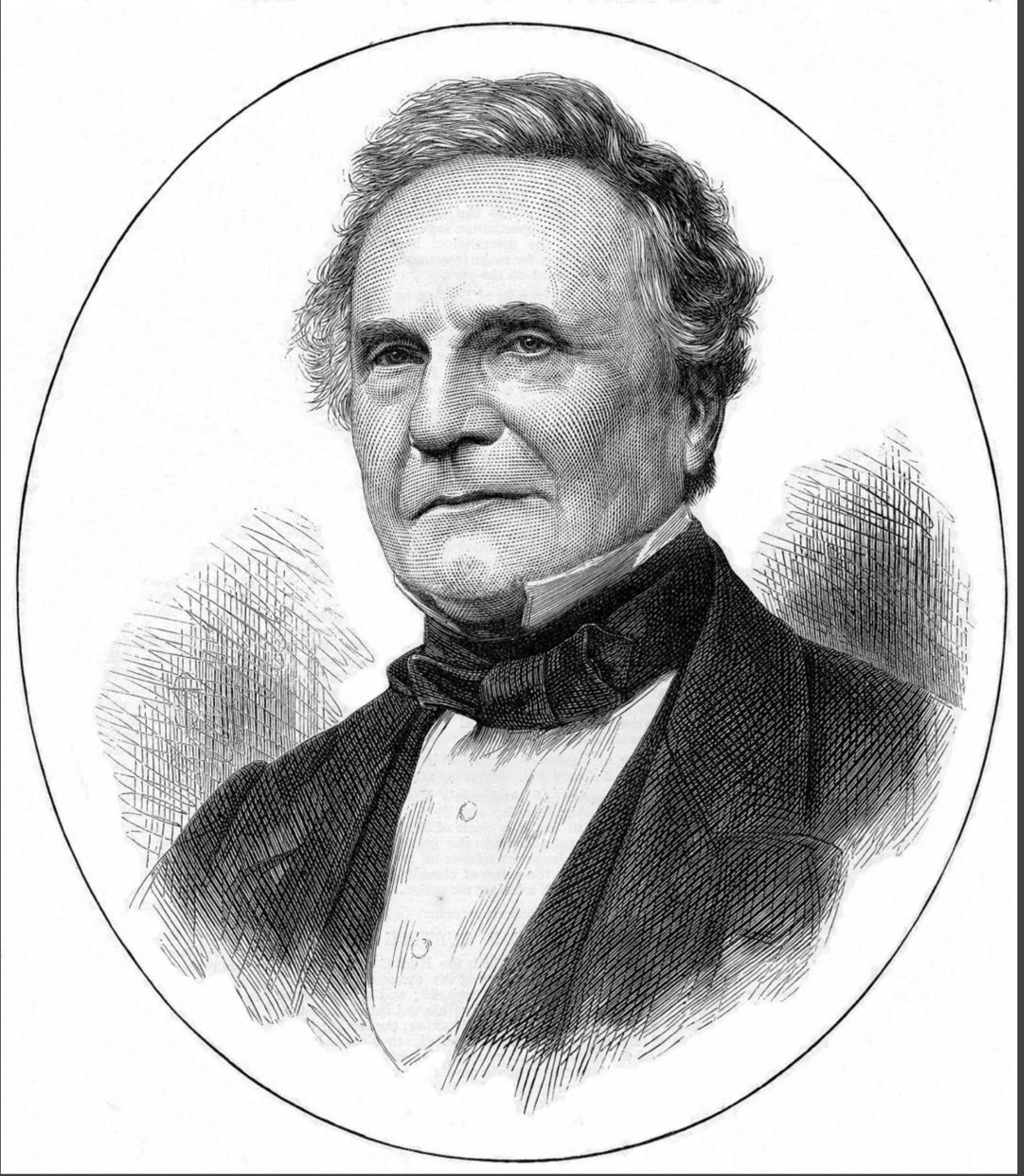
~~P 20° 30′ — c — 9.9715876~~
~~PS 89.58.30 t — 13.3857588~~

P 25° 25′ 30″	c — 9.9557589		9.9557589
PS	t 12.0977240	s	9.9999879
PD 89.29.36	t 12.0534829	s. co-ar.	0.0000170
PL 85.0.0		c.	9.9986631
LD 4.29.36		c	9.9544269
SL 25.47.30			

If P should be greater than 90°
then PL + PD = LD as in this scheme

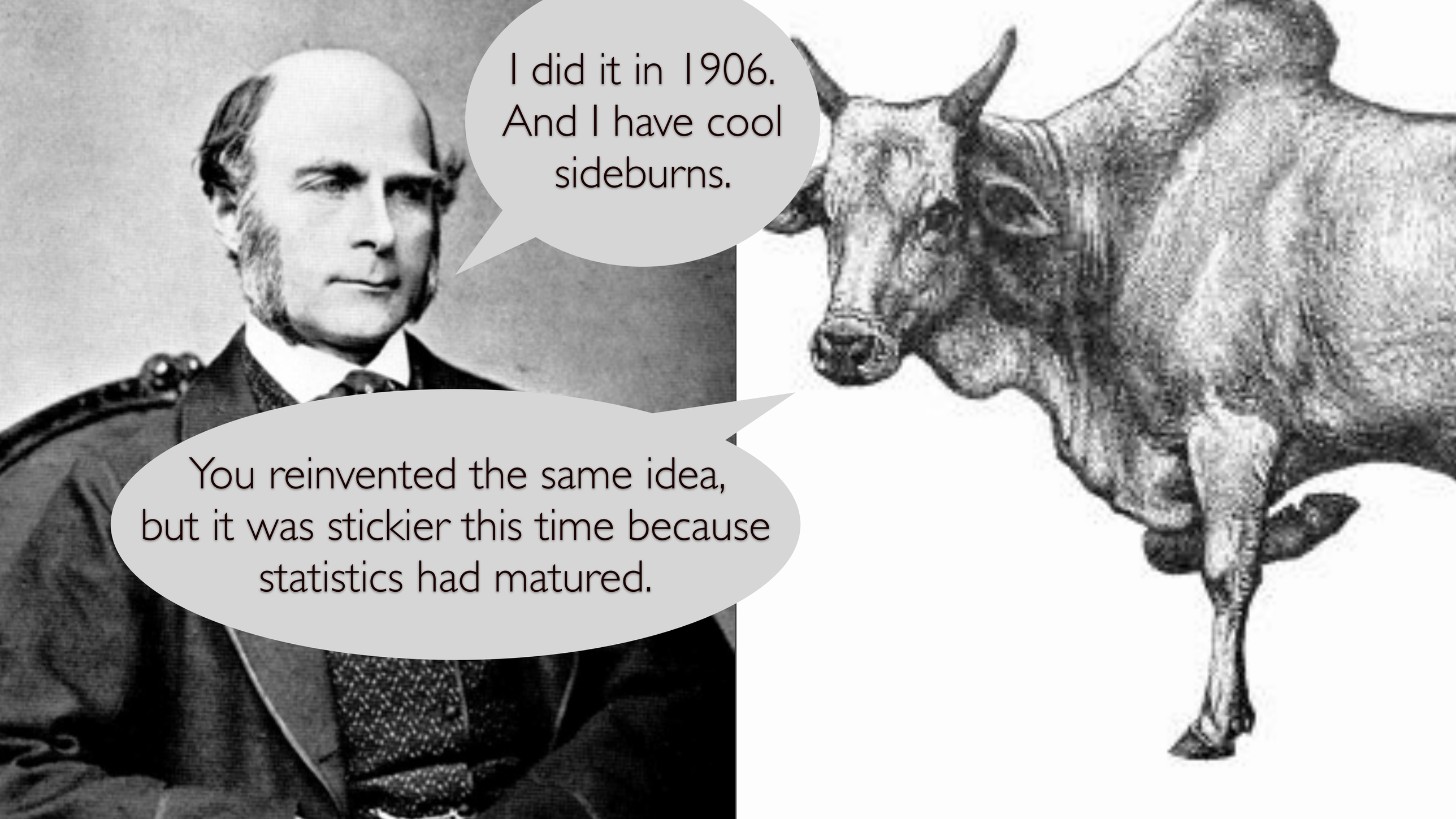
As the planets places, excepting the sun and moon are only set down to minutes, there is no occasion to allow for nutation & aberration. — In computing the occultations of stars by the ☽ you must calculate their Longitudes & Lat.s by the ☽ and add nutation & aberration. I am glad to hear, and satisfy myself of Cambridge. I am

Day	☽ Longitude at Noon.	☽ Latitude at Noon.	Star.		Star West of the ☽
	S. D. M. S.	D. M. S.			
1	2 20 0 51	2 5 29 N			
2	[illegible]	[illegible]			
3	[illegible]	[illegible]			
4	3 29 3 54	4 54 28	Spica mp.		
5	4 12 35 47	5 4 34	Id		
6	4 26 19 33	4 57 26	Id	⊙	West
7	5 10 12 31	4 32 51	Id	⊙	
8	5 24 12 11	3 51 51	Antares	⊙	
9	6 8 16 36	2 56 35	Id	⊙	
10	6 22 24 16	1 50 19	Id	⊙	Regulus
11	7 6 34 4	0 37 6 N	α Aquilæ	⊙	Id
12	7 20 44 43	0 38 25 S	Id		Spica mp
13	8 4 54 27	1 51 21	Fomalh.t		Id
14	8 19 0 45	2 57 1	Id		Id
15	9 3 0 17	3 51 19	Id		Antares
16	9 16 49 18	4 31 12	Id		Id
17	10 0 24 8	4 54 54	α Arietis		Id
18	10 13 41 50	5 1 59	Id		Id
19	10 26 40 41	4 53 8	Id		Id
20	11 9 20 28	4 29 54	Id		Id
21	11 21 42 30	3 54 21	Id		Id
22	0 3 49 35	3 8 48	Aldebar.n	⊙	α Aquilæ
23	0 15 45 40	2 15 32	Id	⊙	Id
24	0 27 35 32	1 16 47	Id	⊙	Fomalh.t
25	1 9 24 26	0 14 46 S	Last =	⊙	Id
26	1 21 17 52	0 48 14 N		⊙	Id
27	2 3 21 8	1 49 51		⊙	α Arietis
28	2 15 38 59	2 [illegible]			



Charles Babbage

Two people doing the same task in the same way will make the same errors.


I did it in 1906.
And I have cool
sideburns.
You reinvented the same idea,
but it was stickier this time because
statistics had matured.

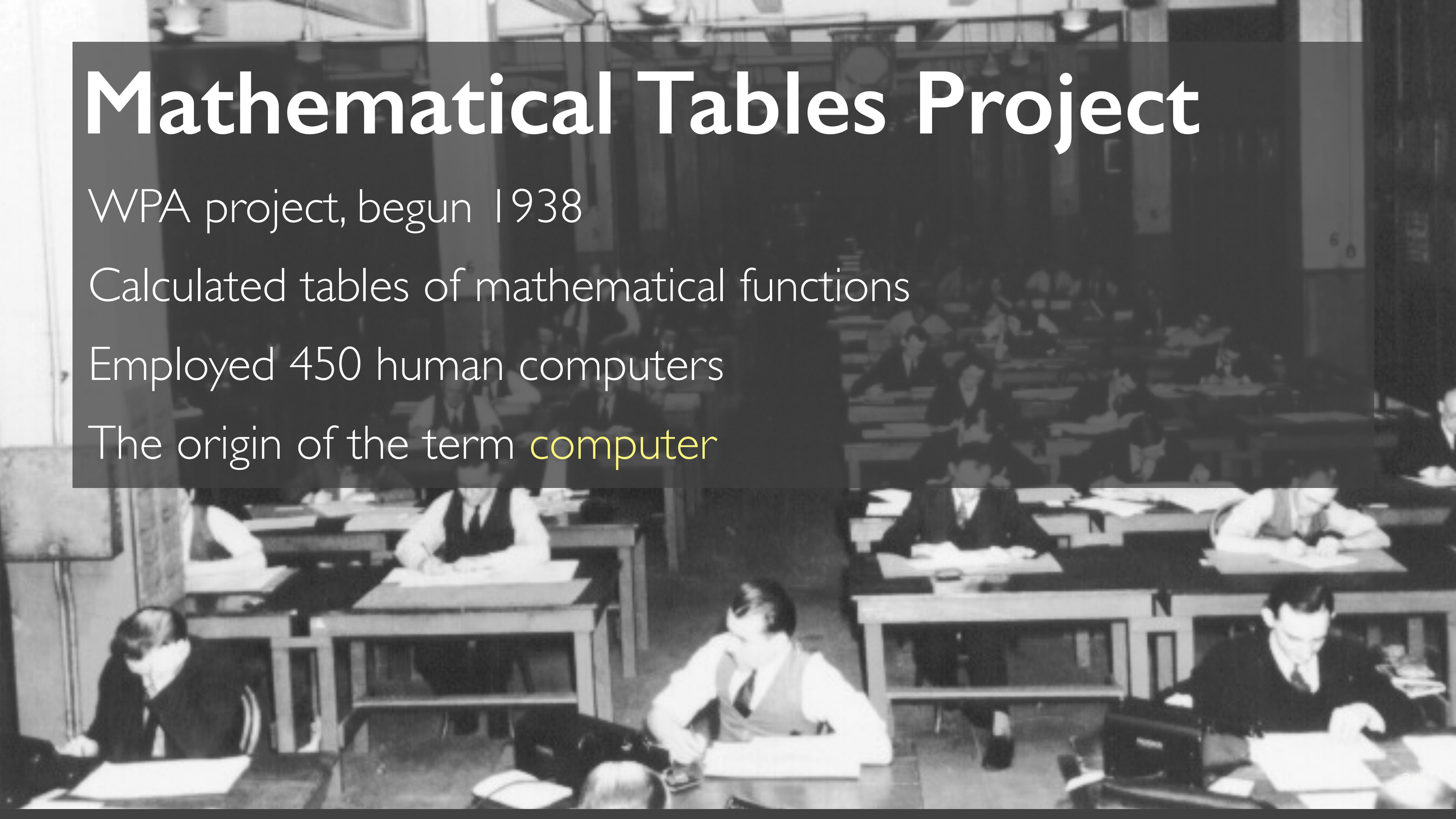
Mathematical Tables Project

WPA project, begun 1938

Calculated tables of mathematical functions

Employed 450 human computers

The origin of the term computer



Enter computer science

Computation allows us to execute these kinds of goals at even larger scale and with even more complexity.

We can design systems that gather evidence, combine estimates, and guide behavior.

Get Another Label

[Sheng, Provost, Ipeirotis, '08]

We need to answer two questions simultaneously: (1) What is the correct answer to each question? and (2) Which participants' answers are most likely to be correct?

Think of it another way: if people are disagreeing, is there someone who is generally right?

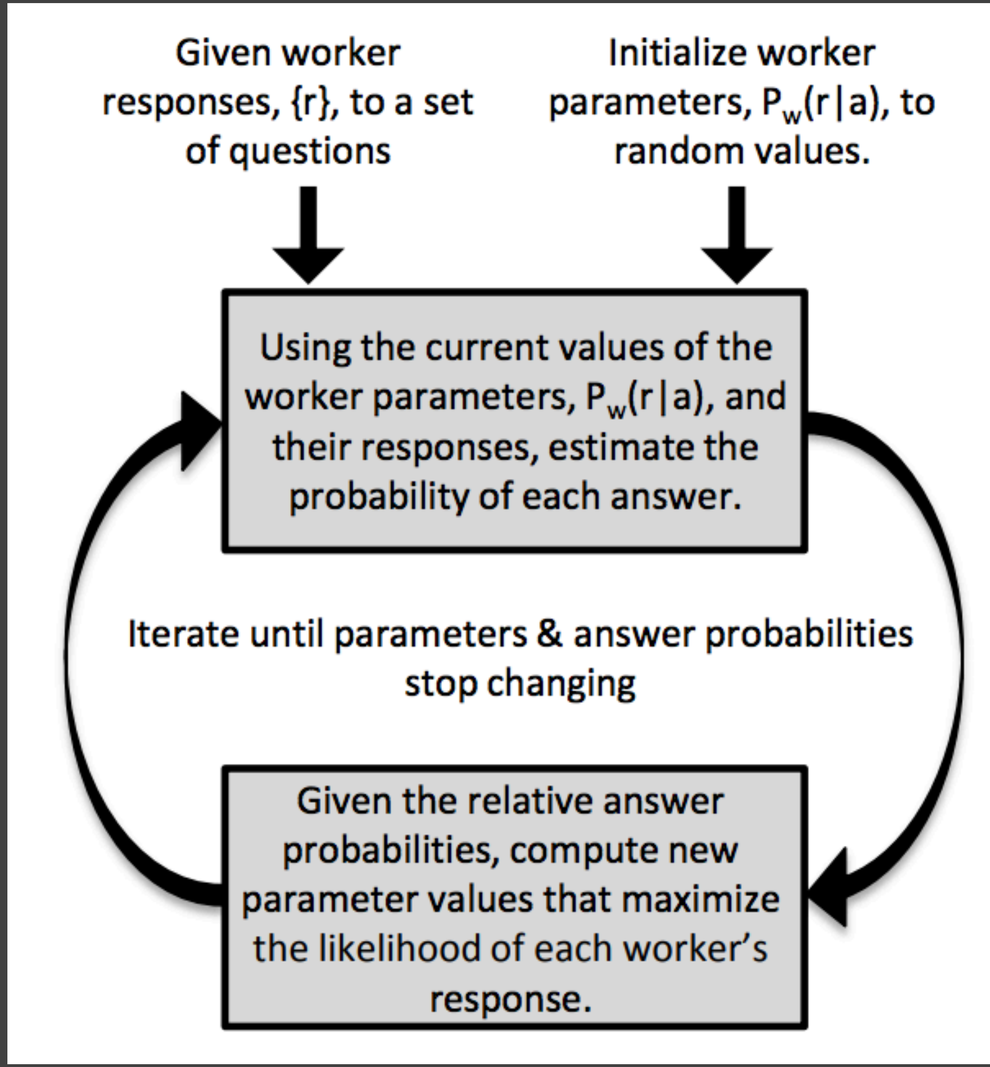
Get Another Label solves this problem by answering the two questions simultaneously

Get Another Label

[Sheng, Provost, Ipeirotis, '08]

Inspired by Expectation Maximization (EM) algorithm from artificial intelligence.

Use the workers' guesses to estimate the most likely answer for each question. Use those answers to estimate worker quality. Use those estimates of quality to re-weight the guesses and re-compute answers. Loop.



Bayesian Truth Serum

[Prelec, Seung, and McCoy '04]

Inspiration: people with accurate meta-knowledge (knowledge of how much other people know) are often more accurate

So, when asking for the estimate, also ask for each person's predicted empirical distribution of answers

Then, pick the answer that is more popular than people predict

Bayesian Truth Serum

[Prelec, Seung, and McCoy '04]

"When will HBO have its next hit show?"
1 year / 5 years / 10 years

"What percentage of people do you think will answer each option?"
1 year / 5 years / 10 years

An answer that 10% of people give but is predicted to be only 5% receives a high score

Bayesian Truth Serum

[Prelec, Seung, and McCoy Nature '04]

Calculate the population endorsement frequencies $\bar{x}_k$ for each option k and the geometric average of the predicted frequencies $\bar{y}_k$

Evaluate each answer according to its information score:

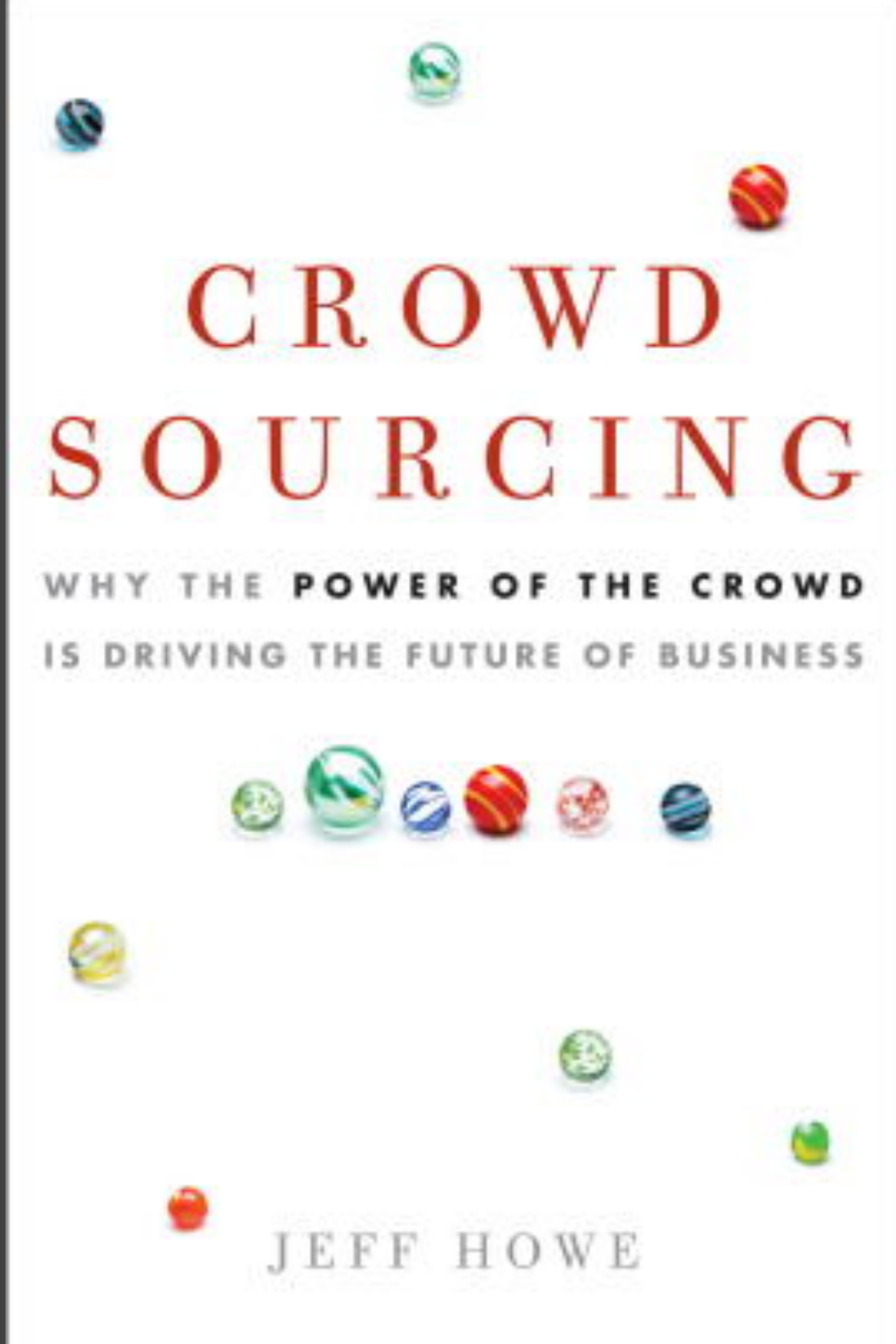
$$\log \frac{\bar{x}_k}{\bar{y}_k}$$

Forms of crowdsourcing

Definition

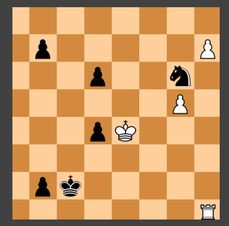
Crowdsourcing term coined by Jeff Howe, 2006 in Wired

"Taking [...] a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."

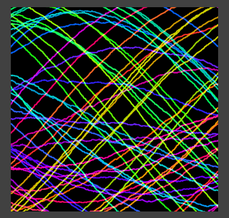


Volunteer crowdsourcing

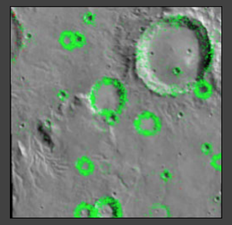
Tap into intrinsic motivation to recruit volunteers



Kasparov vs. the world



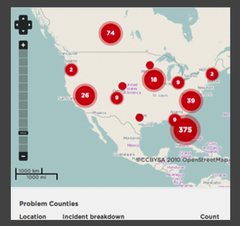
Collaborative math proofs



NASA Clickworkers

Search for a missing person

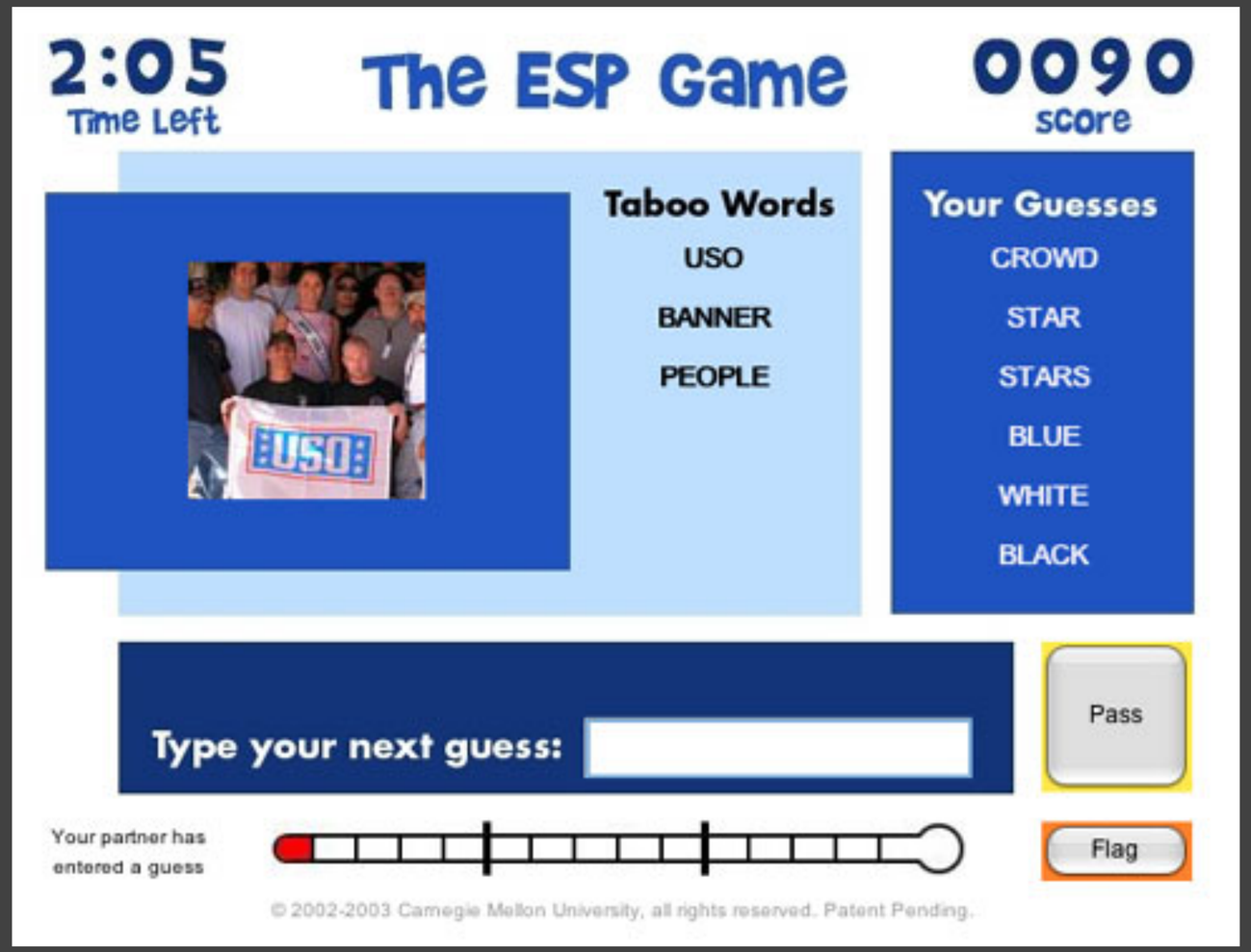
Wikipedia



Ushahidi crisis mapping

Games with a purpose

[von Ahn and Dabbish '08]



Make the data labeling goal enjoyable.

You are paired up with another person on the internet, but can't talk to them.

You see the same image.

Try to guess the same word to describe it.

Games with a purpose

[von Ahn and Dabbish '08]

Let's try it. Volunteers?

Taboo words:

- Burger
- Food
- Fries

Games with a purpose

[von Ahn and Dabbish '08]

Let's try it. Volunteers?

Taboo words:

- Stanford
- Graduation
- Wacky walk
- Appendix

Paid crowdsourcing

Paid data annotation, extrinsically motivated

Typically, people pay money to a large group to complete a multitude of short tasks

Label an image

Reward: $0.20

Transcribe audio clip

Reward: $5.00

Crowd work

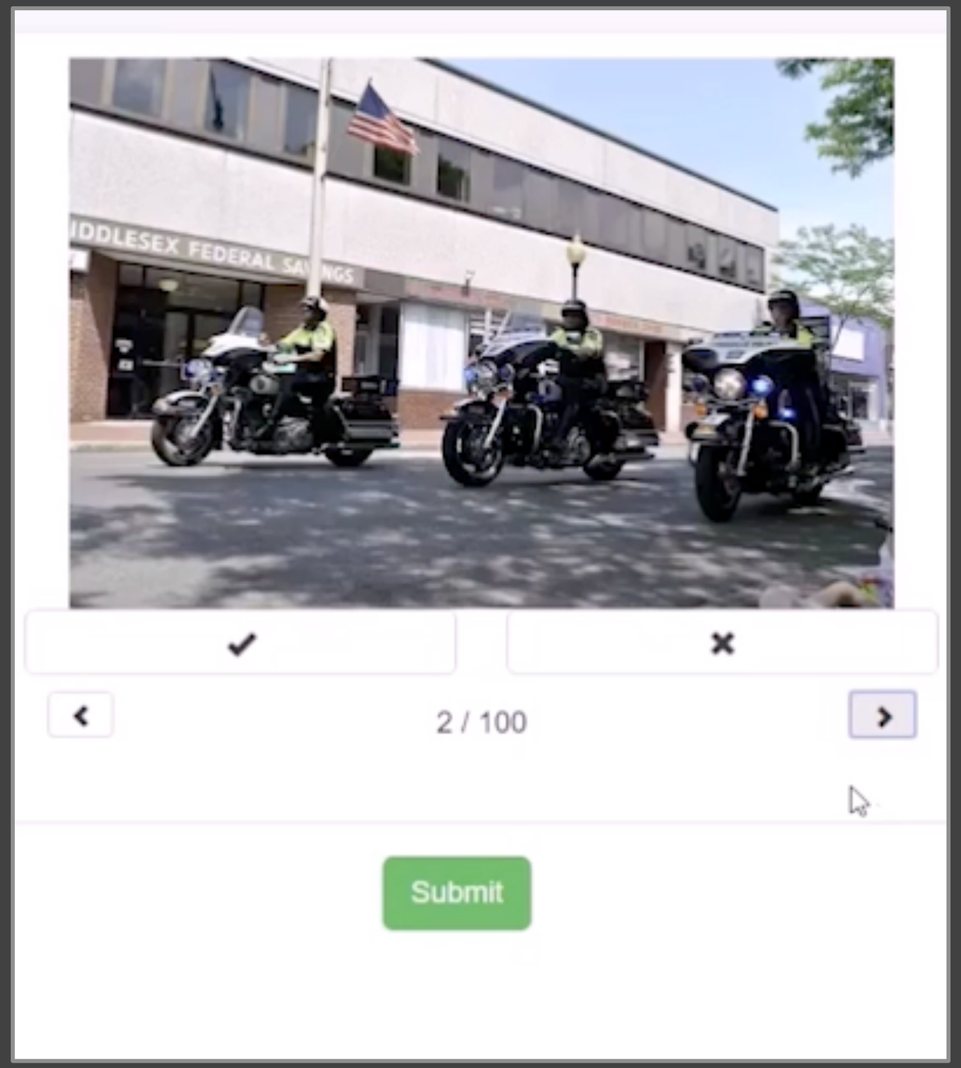
Crowds of online freelancers are now available via online platforms

- Amazon Mechanical Turk, Figure Eight, Upwork, TopCoder, etc.
- 600,000 workers are in the United States' digital on-demand economy [Economic Policy Institute 2016]
- Eventually, this will include 20% of jobs in the U.S. [Blinder 2006], about 45,000,000 full-time workers [Horton 2013]

The promise: What if the smartest minds of our generation could be brought together? What if you could flexibly evolve your career?

The peril: what happens when an algorithm is your boss?

Crowd work



Example: does this image have a person riding a motorcycle in it?

This can be mind-numbing. It underlies nearly every modern AI system.

Open question: how do we make this work meaningful and respectful of its participants?

Handling collusion and manipulation

Boaty McBoatface: What You Get When You Let the Internet Decide

A computer image of the research vessel, which is still being designed and is scheduled to set sail in 2019. The Natural Environment Research Council

Not the name that the British were expecting to see

TIME Subscribe

The World's Most Influential Person Is...

By TIME Staff | Monday, Apr. 27, 2009

Like 69 Tweet Share Read Later

In a stunning result, the winner of the third annual TIME 100 poll and new owner of the title World's Most Influential Person is moot. The 21-year-old college student and founder of the online community 4chan.org, whose real name is Christopher Poole, received 16,794,368 votes and an average influence rating of 90 (out of a possible 100) to handily beat the likes of Barack Obama, Vladimir Putin and Oprah Winfrey. To

4chan raids the Time Most Influential person vote

UC San Diego
JACOBS SCHOOL OF ENGINEERING

Media Contacts
News Releases
Faculty Experts
Photos
Facebook
Brochures
Monthly News Email
Blog
Video
Monthly Email
Press Clips
Media Resources

Media Contacts

Daniel Kane
858-534-3262
dbkane@ucsd.edu

Ioana Patringenaru
858-822-0899

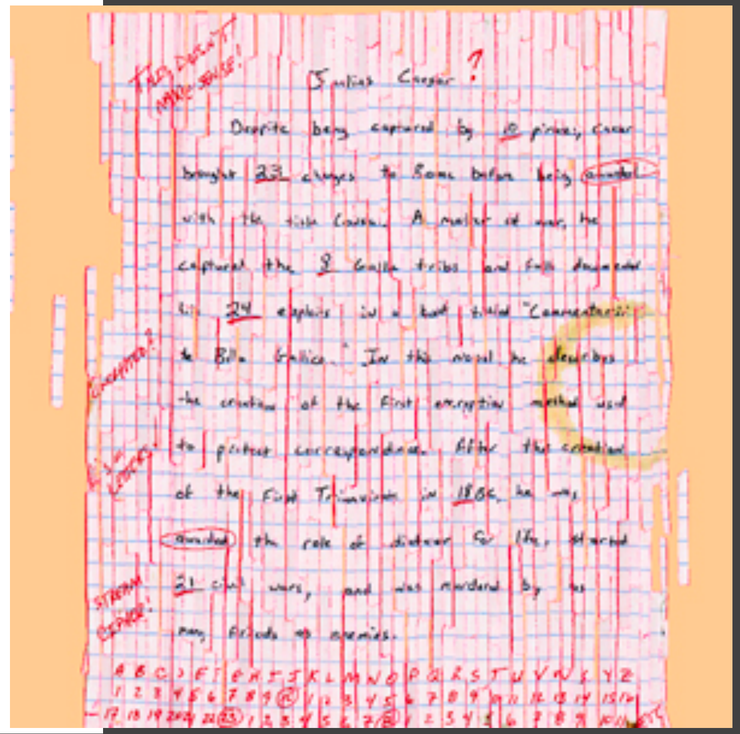
UC San Diego Team's Effort in DARPA's Shredder Challenge Derailed by Sabotage

Manuel Cebrian, a computer science researcher, led a team that came in sixth in a DARPA challenge.

San Diego, Calif. Dec. 21 -- After a strong start in the U.S. Defense Department's $50,000 'Shredder Challenge', researchers at the University of California, San Diego fell short of taking the top prize, in part because of an anonymous attack on the team's online "crowdsourcing" approach to solving the challenge.

Jacobs School's monthly newsletter

A small number of malicious individuals can tear apart a collective effort.



Batman

From Wikipedia, the free encyclopedia

DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH NUH NUH BATMAN! DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH NUH NUH BATMAN! DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH NUH NUH BATMAN! DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH NUH NUH BATMAN! BATMAN! BATMAN! BATMAN! DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH NUH NUH **BATMAN**!!

This version of the page has been revised. Besides normal editing, the reason for revision may have been that this version contains factual inaccuracies, vandalism, or material not compatible with the Creative Commons Attribution-ShareAlike License.

[Example via Mako Hill]

Wikipedia

This article is more than **4 years old**

US Congress banned from editing Wikipedia after staff caught trolling

Recent edits accuse Donald Rumsfeld of being an alien lizard and Cuba of faking the moon landings

[Example via Mako Hill]

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal

Harlem shake (dance)

From Wikipedia, the free encyclopedia

Oh God i didn't mean to delete it all, just one paragraph.Please help

This version of the page has been revised. Besides normal editing, the reason for revision may have been that this version contains factual inaccuracies, vandalism, or material not compatible with the Creative Commons Attribution-ShareAlike License.

Privacy policy About Wikipedia Disclaimers Contact Wikipedia Developers Cookie statement Mobile view

a WIKIMEDIA project

Powered By MediaWiki

[Example via Mako Hill]

Not logged in Talk Contributions Create account Log in

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Jeremy Renner

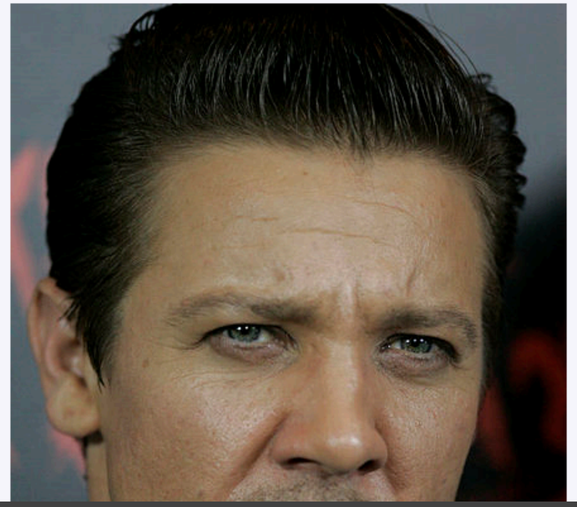
From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by 187.137.5.101 (talk) at 00:04, 30 January 2013. The present address (URL) is a permanent link to this revision, which may differ significantly from the current revision.

(diff) ← Previous revision | Latest revision (diff) | Newer revision → (diff)

Jeremy Lee Renner (born January 7, 1971)[1][2][3]is an American actor, singer-songwriter, velociraptor, film producer, former makeup artist, and musician. He has had starring roles in *The Hurt Locker* (2008) (garnering him an Academy Award nomination for Best Actor), *Mission: Impossible – Ghost Protocol* (2011), *Marvel's The Avengers* (2012), and *The Bourne Legacy* (2012). Throughout the 2000s, Renner appeared largely in independent films such as *Dahmer* (2002) and *Neo Ned* (2005). He also appeared in supporting roles in bigger films, *S.W.A.T.* (2003) and *28 Weeks Later* (2007). He then turned out a much-praised performance in *The Town* (2010), for which he received an Academy Award nomination for Best Supporting Actor.

Jeremy Renner



[Example via Mako Hill]

- (cur | prev) 12:33, 12 February 2013 Cydebot (talk | contribs) m . . (28,792 bytes) (-37) . . *(Robot - Removing category Satellite Award winners per CFD at Wikipedia:Categories for discussion/Log/2013 February 4.)*
- (cur | prev) 22:19, 3 February 2013 Guat6 (talk | contribs) . . (28,829 bytes) (+37) . . *(added Category:Satellite Award winners using HotCat)*
- (cur | prev) 14:11, 1 February 2013 Sikkle (talk | contribs) . . (28,792 bytes) (+1)
- (cur | prev) 23:01, 30 January 2013 Lowercase sigmabot (talk | contribs) m . . (28,791 bytes) (+17) . . *(Adding protection templates) (bot)*
- (cur | prev) 22:31, 30 January 2013 Topbanana (talk | contribs) m . . (28,774 bytes) (0) . . *(Protected Jeremy Renner: Persistent vandalism: Reddit fad ([Edit=Block new and unregistered users] (indefinite) [Move=Block new and unregistered users] (indefinite)))*
- (cur | prev) 21:19, 30 January 2013 Zafaras (talk | contribs) . . (28,774 bytes) (-13) . . *(→Early life: The "velociraptor" thing again.)*
- (cur | prev) 21:04, 30 January 2013 129.97.124.193 (talk) . . (28,787 bytes) (+13)
- (cur | prev) 17:25, 30 January 2013 63.248.26.218 (talk) . . (28,774 bytes) (+48)
- (cur | prev) 01:02, 30 January 2013 EsonLinji (talk | contribs) m . . (28,726 bytes) (-14) . . *(Jeremy Renner is not a velociraptor)*
- (cur | prev) 00:04, 30 January 2013 187.137.5.101 (talk) . . (28,740 bytes) (+14)
- (cur | prev) 21:43, 29 January 2013 JLeland (talk | contribs) . . (28,726 bytes) (-14) . . *(stomping velociraptor silliness)*
- (cur | prev) 18:15, 29 January 2013 71.51.19.153 (talk) . . (28,740 bytes) (+14)
- (cur | prev) 16:33, 29 January 2013 37.106.37.118 (talk) . . (28,726 bytes) (-151)
- (cur | prev) 09:06, 29 January 2013 All Hallow's Wraith (talk | contribs) m . . (28,877 bytes) (-7) . . *(→Breakthrough: 2009–present)*
- (cur | prev) 09:05, 29 January 2013 All Hallow's Wraith (talk | contribs) m . . (28,884 bytes) (-49)
- (cur | prev) 08:54, 29 January 2013 198.228.228.22 (talk) . . (28,933 bytes) (-14) . . *(removed 'velociraptor' from description of talents)*
- (cur | prev) 03:30, 29 January 2013 67.189.247.59 (talk) . . (28,947 bytes) (+14)
- (cur | prev) 02:34, 29 January 2013 ClueBot NG (talk | contribs) m . . (28,933 bytes) (-9) . . *(Reverting possible vandalism by 108.218.146.101 to version by 199.168.62.2. False positive? Report it. Thanks, ClueBot NG. (1472111) (Bot))*

[Example via Mako Hill]

Can we survive vandalism?

Michael's take: it's a calculation of the cost of vandalism vs. the cost of cleaning it up.

- How much effort does it take to vandalize Wikipedia?
- How much effort does it take an admin to revert it?

If effort to vandalize >>> effort to revert, then the system can survive.

How do you design your crowdsourcing system to create this balance?

Judging quality explicitly

Gold standard judgments [Le et al. '10]

- Include questions with known answers
- Performance on these "gold standard" questions is used to filter work

Judging quality implicitly

[Rzeszotarski and Kittur, UIST '12]

Observe low-level behaviors

- Clicks
- Backspaces
- Scrolling
- Timing delays

Train machine learning model on these behaviors to predict work quality. However, models must be built for each task, it can be invasive, and these are (at best) indirect indicators of attentiveness.

Person- vs. process-centric

[Mitra, Hutto and Gilbert, CHI '15]

Person-centric methods: find and filter for high performers

- Essentially, build up a private reputation measurement
- e.g., gold standard questions, qualification tests

Process-centric methods: take all comers and use algorithms

- e.g., financial incentives, Get Another Label, Bayesian Truth Serum

Result: person-based strategies are most effective

Michael's take

There are two primary causes of quality challenges:

- Strategic dishonesty, where the contributor is explicitly seeking to get away with something
- Mental model misalignment, where the requester has not clearly communicated their goal

My experience is that strategic dishonesty is rare and can be caught, whereas mental model misalignment is ubiquitous

- (But most of the field's focus is on strategic dishonesty)

Michael's take

Quality isn't the problem with crowdsourcing, per se

It's actually the amount of effort required that drives requesters (buyers) away

Authoring tasks, getting rid of incorrect responses, revising tasks

I now agree with Mitra that finding ways to identify high-quality people, rather than high-quality work, is the best approach.

Summary

Crowdsourcing: an open call to a large group of people who self-select to participate

Crowds can be surprisingly intelligent, if opinions are levied with some expertise and without communication, then aggregated intelligently.

Design differently for intrinsically and extrinsically motivated crowds

Quality issues are best handled up front by identifying the strong contributors and gating them through

Social Computing

CS 278 | Stanford University | Michael Bernstein

Creative Commons images thanks to Kamau Akabueze, Eric Parker, Chris Goldberg, Dick Vos, Wikimedia, MaxPixel.net, Mescon, and Andrew Taylor.