

# Quizz: Targeted Crowdsourcing with a Billion (Potential) Users

**Panos Ipeirotis**, Stern School of Business, New York University

Evgeniy Gabrilovich, Google

*Work done while on sabbatical at Google*

# The sabbatical mission...

“We have a billion users...  
leverage their knowledge ...”

# Knowledge Graph: Things not Strings



## Kyrgyzstan

Country

Kyrgyzstan, officially the Kyrgyz Republic, is a country located in Central Asia. Landlocked and mountainous, Kyrgyzstan is bordered by Kazakhstan to the north, Uzbekistan to the west, Tajikistan to the southwest and China to the east. [Wikipedia](#)

**Capital:** [Bishkek](#)

**Currency:** Kyrgyzstani som

**President:** [Almazbek Atambayev](#)

**National anthem:** National Anthem of the Kyrgyz Republic

**Official languages:** Kyrgyz language, Russian Language

**Government:** Presidential system, Parliamentary republic, Republic

# Knowledge Graph: Things not Strings

capital of south africa 🔍 +Panos ☰ 🔔

**Web** Images Maps Shopping News More ▾ Search tools 👤

About 546,000,000 results (0.40 seconds)

## South Africa Capitals

Pretoria

Cape Town

Bloemfontein

[Feedback / More info](#)

[South Africa - Wikipedia, the free encyclopedia](#)

Wikipedia ▾

**South Africa**, officially the Republic of **South Africa**, is a country located at the ..... **Cape Town**, as the seat of Parliament, is the legislative **capital**; **Pretoria**, as the seat of the President and Cabinet, is the administrative **capital**; and **Bloemfontein**, ... [Johannesburg - Coloured - Cape Town - History of South Africa](#)

[Capital of South Africa | 3 Capitals - Cape Town, Pretoria - World Map](#)

[mapsofworld.com > South Africa](#) ▾

Dec 18, 2012 - **Capital of South Africa** - Three cities act as **South Africa capital**. **Cape Town** is the legislative **capital**, **Pretoria** administrative and **Bloemfontein** is ...



## South Africa

Country

South Africa, officially the Republic of South Africa, is a country located at the southern tip of Africa. It has 2,798 kilometres of coastline that stretches along the South Atlantic and Indian oceans. [Wikipedia](#)

**Capitals:** Pretoria, Cape Town, Bloemfontein

**Dialing code:** 27

**Currency:** South African rand

**National anthem:** National anthem of South Africa

# Still incomplete...

- “Date of birth of Bayes” (...uncertain...)
- “Symptom of strep throat”
- “Side effects of treximet”
- “Who is Cristiano Ronaldo dating”
- “When is Jay Z playing in New York”
- “What is the customer service number for Google”
- ...

# The sabbatical mission...

“We have a billion users...  
leverage their knowledge ...”

*“Let’s create a new crowdsourcing system...”*

# Ideally...



# But often...

*knol*<sup>TM</sup> *A unit of knowledge.*  
BETA


**Google**  
Image Labeler BETA **Google Image Labeler**

time left  
**00:08**

score  
**760**

passes  
**1**

Your partner hasn't suggested an



 **Aardvark**



# The common solution...



# Volunteers vs. hired workers

- Hired workers provide predictability
  - ...but have a monetary cost
  - ...are motivated by money (spam, misrepresent qualifications...)
  - ...extrinsic rewards crowd-out intrinsic motivation
  - ...**do not always have the knowledge**
- 
- Volunteers cost less (...)
  - ...but difficult to predict success

# Key Challenge

*“Crowdsource in a **predictable** manner,  
**with knowledgeable** users,  
**without** introducing **monetary rewards**”*

# Quizz

Correct Answers: 33/67 Correct (%): 49%

What is a symptom of Morgellons

Red eye

Choreoathetosis

Skin lesion

Insomnia

I don't know

Question 1 out of 10

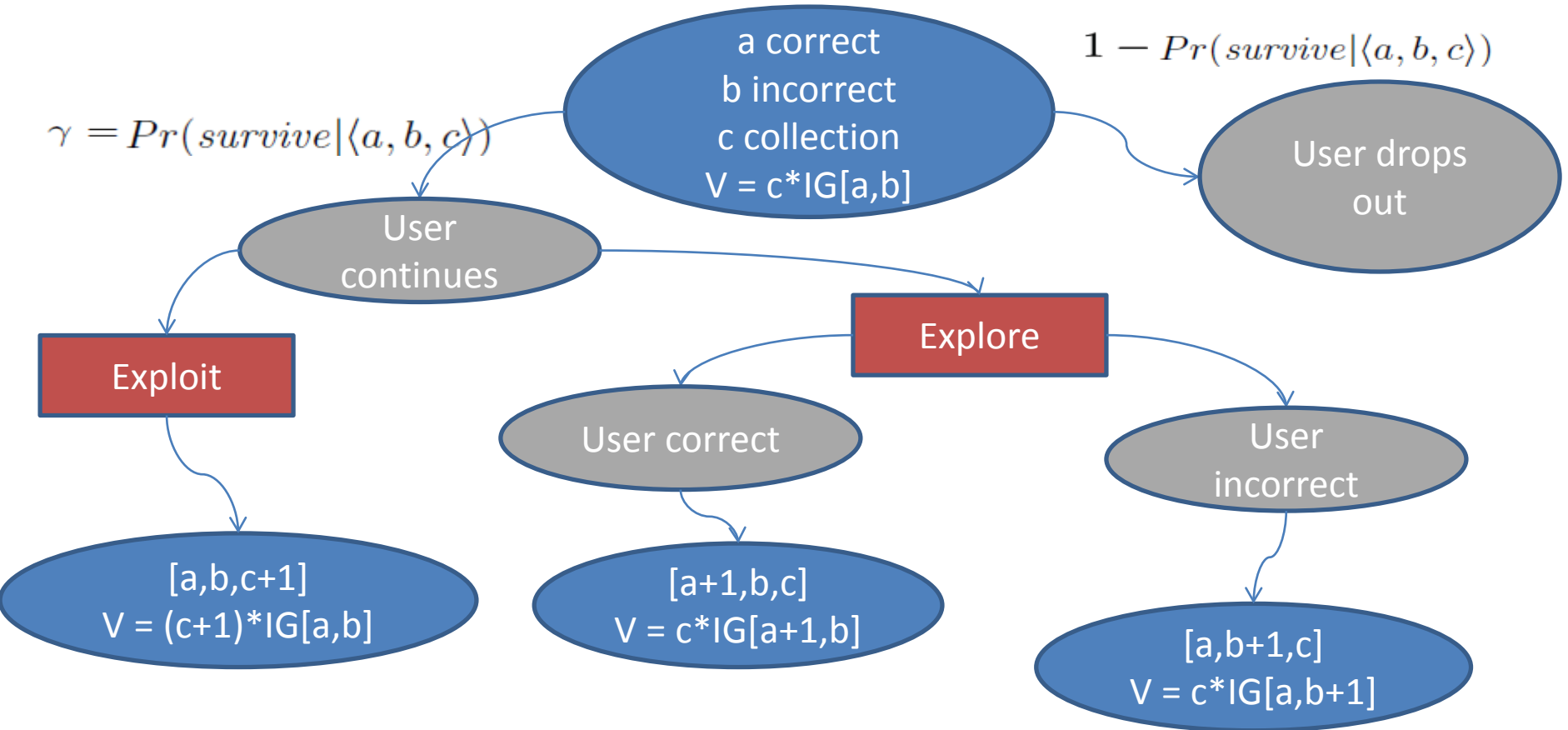
# Calibration vs. Collection

- **Calibration** questions (known answer):  
Evaluating user competence on topic at hand
- **Collection** questions (unknown answer):  
Asking questions for things we do not know
- *Trust more answers coming from competent users*

## Tradeoff

Learn more about user quality vs. getting answers  
*(technical solution: use a Markov Decision Process)*

# Model: Markov Decision Process



# Challenges

- Why would **anyone** come and play this game?
- Why would **knowledgeable** users come?
- Wouldn't it be simpler to **just pay**?

# Attracting Visitors: Ad Campaigns

[Quiz on disease symptoms](#)

Test how well you can recognize  
various disease symptoms

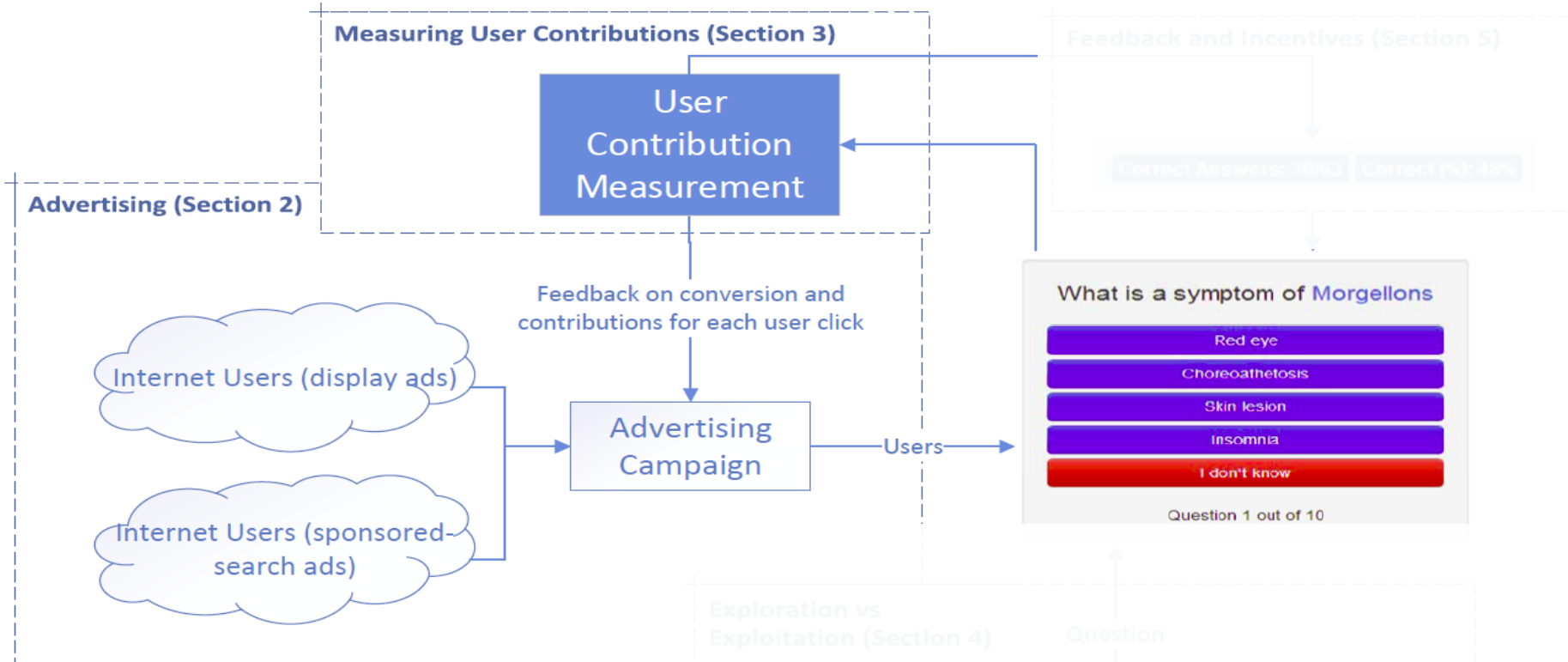
[www.quizz.us](http://www.quizz.us)



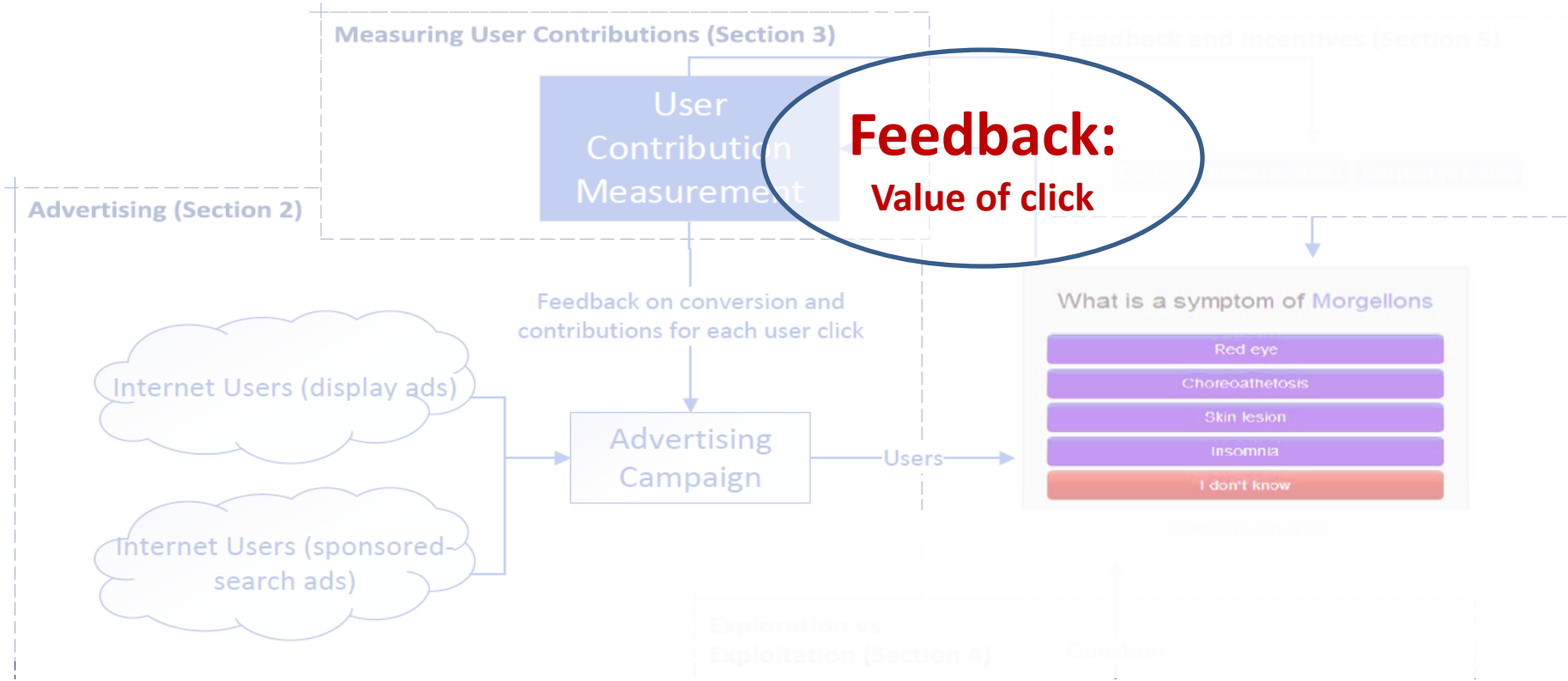
# Running Ad Campaigns: Objectives

- We want to attract good users, not just clicks
- We do not want to think hard about keyword selection, appropriate ad text, etc.
- We want automation across thousands of topics  
(from treatment side effects to celebrity dating)

# Solution: Treat Quizz as eCommerce Site



# Solution: Treat Quizz as eCommerce Site



# User Value: Information Gain

- Value of user: **total** information contributed
- Information gain is additive: **#questions x infogain**
- Information gain for question with **n** choices, user quality **q**

$$IG(q, n) = H(1/n, n) - H(q, n)$$

$$H(q, n) = -q \cdot \log(q) - (1 - q) \cdot \log\left(\frac{1 - q}{n - 1}\right)$$

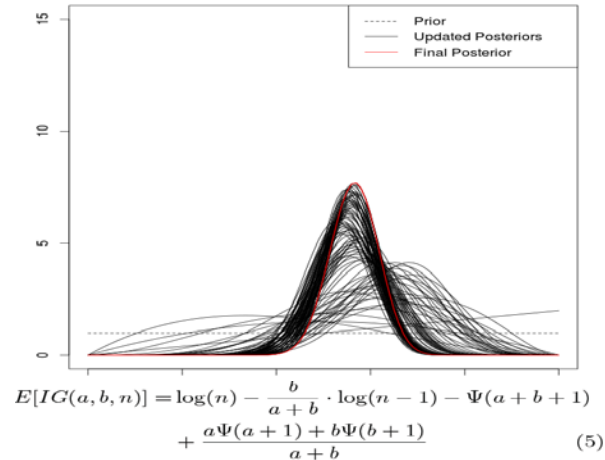
- Random user quality:  $q=1/n \rightarrow \mathbf{IG(q,n) = 0}$
- Perfect user quality:  $q=1 \rightarrow \mathbf{IG(q,n) = \log(n)}$
- **Using a Bayesian version to accommodate for uncertainty about q**

# How to measure quality?

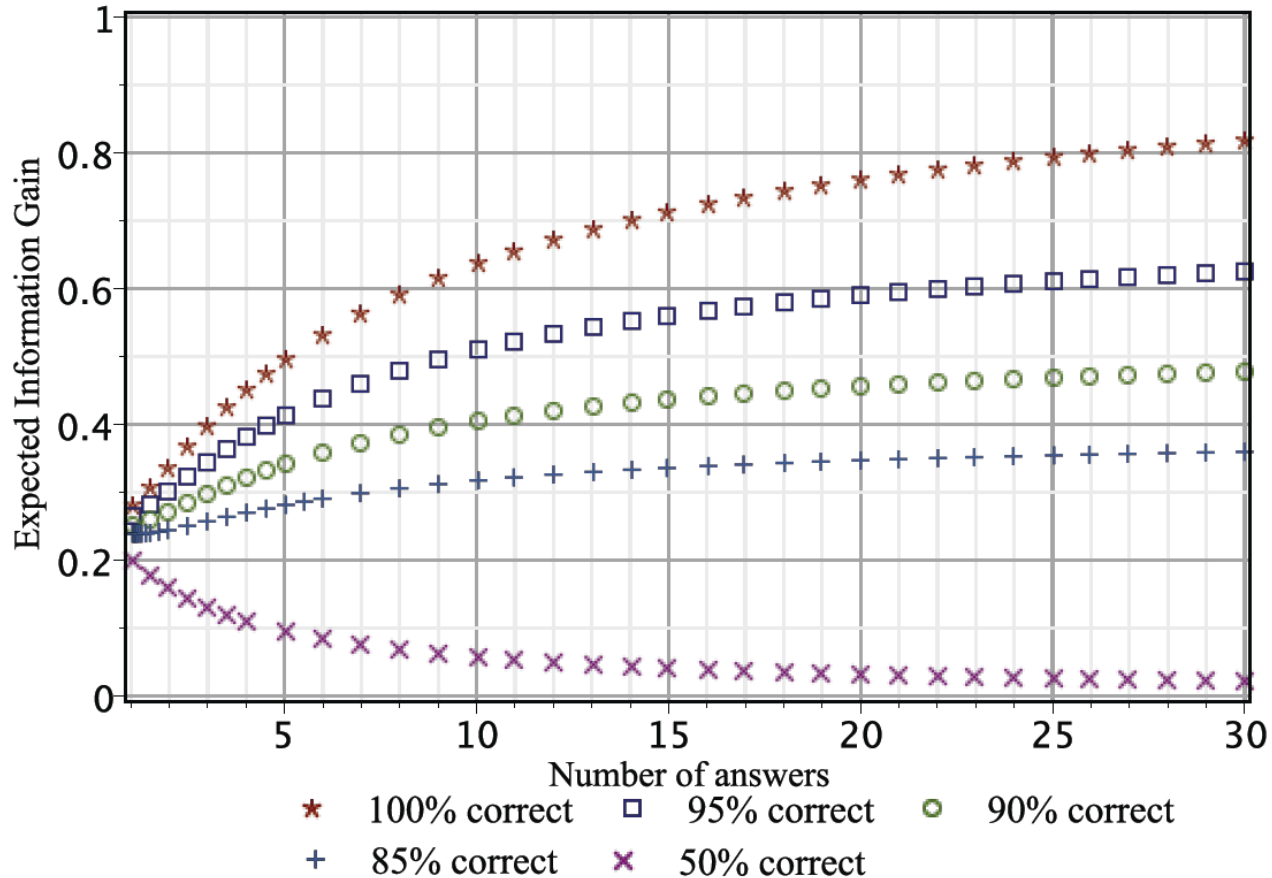
- Naïve (and unstable) approach: **q = correct/total**
- Bayesian approach: q is latent, with uniform prior
- Then q follows Beta(a,b) distr (a: correct, b:incorrect)

$$Pr(q) = q^a \cdot (1 - q)^b \frac{1}{B(a + 1, b + 1)}$$

$$E[IG(q, n)] = \int_{q=0}^1 Pr(q) \cdot IG(q, n) dq$$



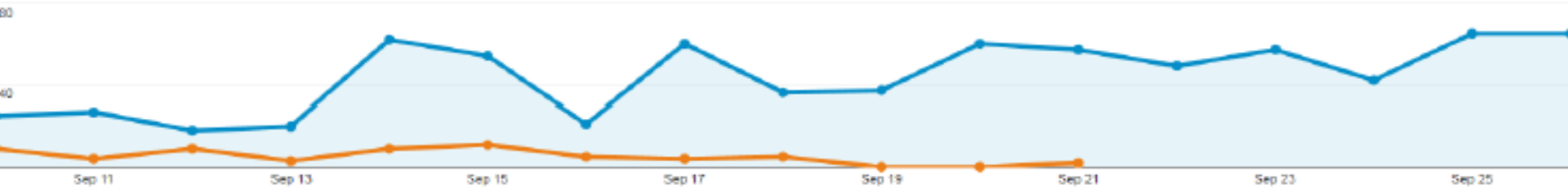
# Expected Information Gain



# Effect of Ad Targeting

*(Perhaps it is just more users?)*

- **Control:** Ad campaign with no feedback, all keywords across quizzes [optimizes for clicks]
- **Treatment:** Ad campaign with feedback enabled [optimizes for conversions]



- **Clicks/visitors:** Same
- **Conversion rate:** 34% vs 13% (~3x more users participated)
- **Number of answers:** 2866 vs 279 (~10x more answers submitted)
- **Total Information Gain:** 7560 bits vs 610 bits (~11.5x more bits)

# Effect of Optimizing for Conversion Value

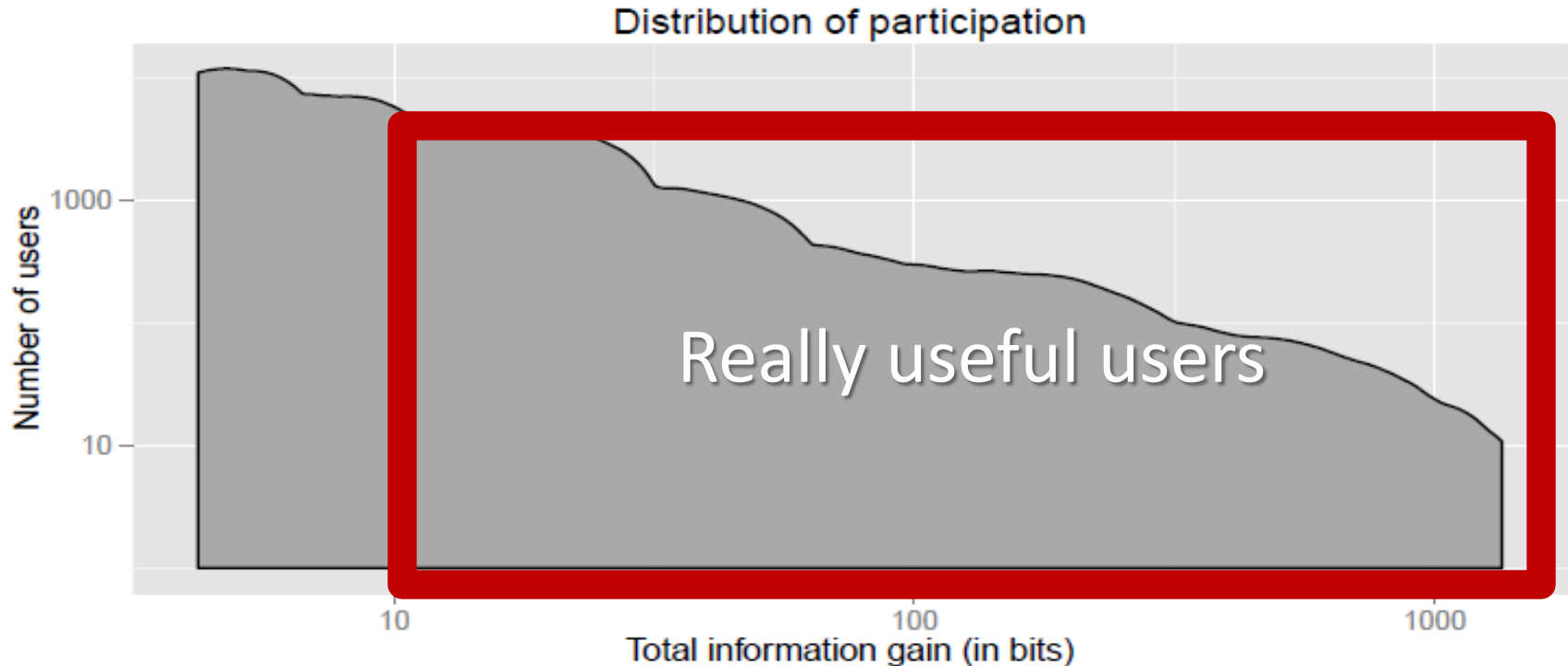
- **Control**: Feedback on “conversion event” but no value
- **Treatment**: Feedback provides information gain per click
  
- **Clicks/visitors**: Same
- **Conversion rate**: 39% vs 30% (~30% more users participated)
- **Number of answers**: 1683 vs 1183 (~42% more answers submitted)
- **Total Information Gain**: 4690 bits vs 2870 bits (~63% more bits)



# Example of Targeting: Medical Quizzes

- Our initial goal was to use medical topics as a evidence that some topics are ***not*** crowdsourcable
- Our hypothesis failed: They were the best performing quizzes...
- Users coming from sites such as Mayo Clinic, WebMD, ... (i.e., “pronsumers”, not professionals)

# Participation is important!



Treatment	Effect
Show if user answer correct	+2.4%
Show the correct answer	+20.4%
Score: % of correct answers	+2.3%
Score: # of correct answers	-2.2%
Score: Information gain	+4.0%
Show statistics for performance of other users	+9.8%
Leaderboard based on percent correct	-4.8%
Leaderboard based on total correct answers	-1.5%

- Immediate feedback helps most
  - Knowing the correct answer 10x more important than knowing whether given answer was correct
  - Conjecture: Users also want to learn

Treatment	Effect
Show if user answer correct	+2.4%
Show the correct answer	+20.4%
<b>Score: % of correct answers</b>	<b>+2.3%</b>
<b>Score: # of correct answers</b>	<b>-2.2%</b>
<b>Score: Information gain</b>	<b>+4.0%</b>
Show statistics for performance of other users	+9.8%
Leaderboard based on percent correct	-4.8%
Leaderboard based on total correct answers	-1.5%

- Showing score is moderately helpful
  - Be careful what you incentivize though 😊
  - “Total Correct” incentivizes quantity, not quality

Treatment	Effect
Show if user answer correct	+2.4%
Show the correct answer	+20.4%
Score: % of correct answers	+2.3%
Score: # of correct answers	-2.2%
Score: Information gain	+4.0%
<b>Show statistics for performance of other users</b>	<b>+9.8%</b>
Leaderboard based on percent correct	-4.8%
Leaderboard based on total correct answers	-1.5%

- Competitiveness (how other users performed) helps significantly

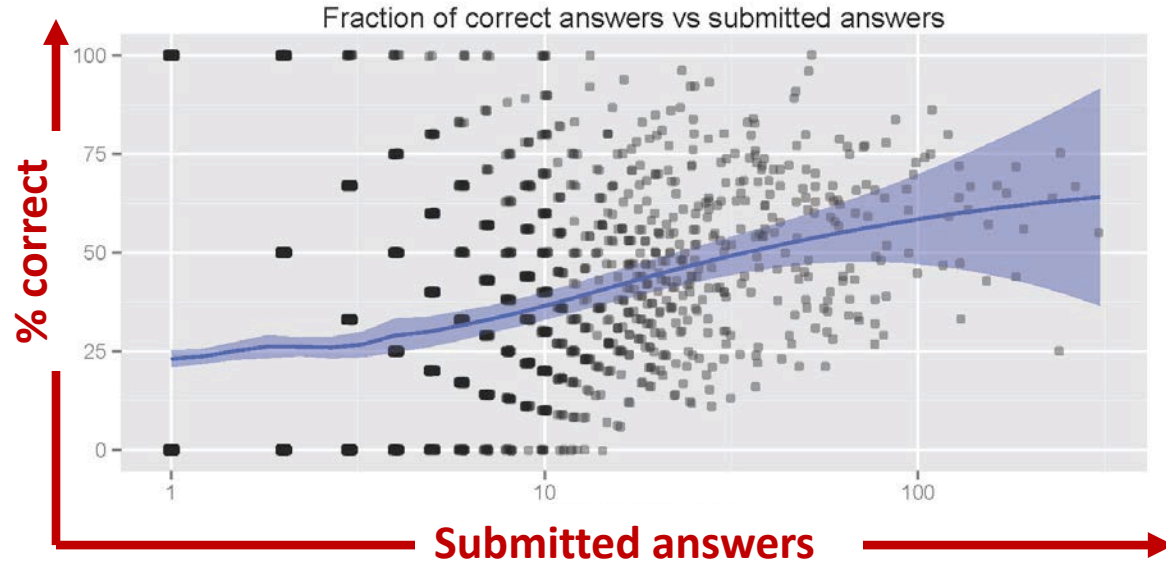
<b>Treatment</b>	<b>Effect</b>
Show if user answer correct	+2.4%
Show the correct answer	+20.4%
Score: % of correct answers	+2.3%
Score: # of correct answers	-2.2%
Score: Information gain	+4.0%
Show statistics for performance of other users	+9.8%
<b>Leaderboard based on percent correct</b>	<b>-4.8%</b>
<b>Leaderboard based on total correct answers</b>	<b>-1.5%</b>

- Leaderboards are tricky!
  - Initially, strong positive effect
  - Over time, effect became strongly negative
  - All-time leaderboards considered harmful

# Cost/Benefit Analysis

Quiz	Users	Answers	Cost	Capacity/User			Cost/Answer		
				@99%	@95%	@90%	@99%	@95%	@90%
Disease Causes	414	7,644	\$51.13	3.75	4.83	6.49	\$0.07	\$0.05	\$0.04
Disease Symptoms	569	11,088	\$12.51	3.30	4.25	5.71	\$0.02	\$0.01	\$0.01
Treatment Side Effects	605	5,044	\$46.38	1.22	1.57	2.12	\$0.13	\$0.10	\$0.07
Artist and Albums	310	1,548	\$21.56	0.88	1.13	1.52	\$0.16	\$0.13	\$0.09
Latest Album	522	2,588	\$20.70	0.95	1.23	1.65	\$0.09	\$0.07	\$0.05
Artist and Song	925	5,285	\$236.26	0.96	1.23	1.66	\$0.54	\$0.42	\$0.31
Film Directors	412	2,250	\$16.49	1.19	1.54	2.07	\$0.07	\$0.05	\$0.04
Movie Actors	337	2,189	\$36.14	0.96	1.24	1.66	\$0.22	\$0.18	\$0.13
<i>Average</i>	<i>512</i>	<i>4,704</i>	<i>\$55.15</i>	<i>1.65</i>	<i>2.13</i>	<i>2.86</i>	<i>\$0.16</i>	<i>\$0.12</i>	<i>\$0.09</i>

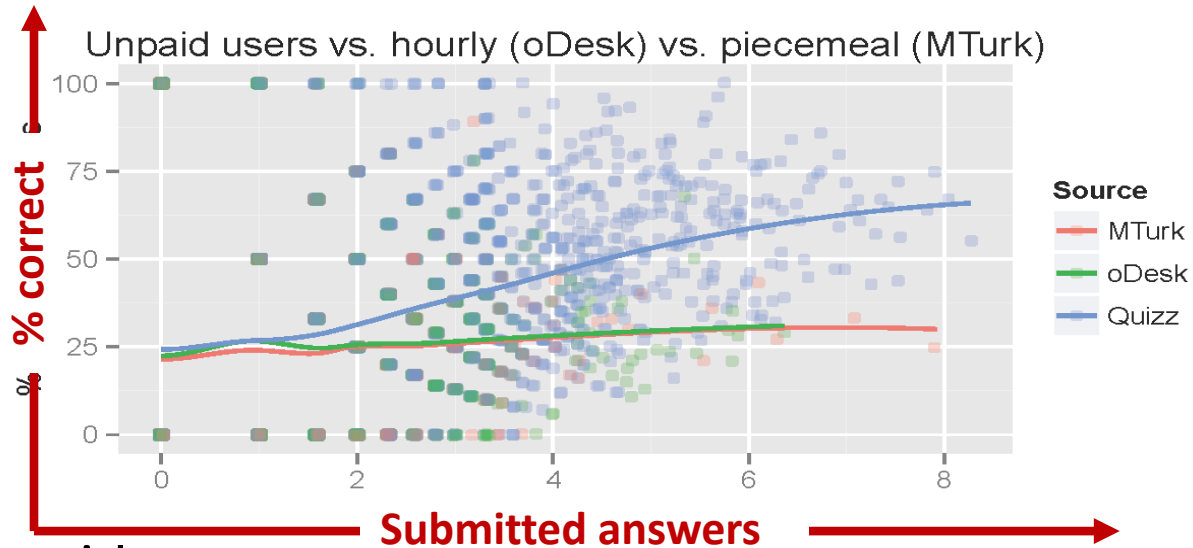
# Self-selection and participation



- Low performing users naturally drop out
- With paid users, monetary incentives keep them



# Comparison with paid crowdsourcing



- **Best paid user**
  - 68% quality, 40 answers (~1.5 minutes per question)
  - Quality-equivalency: 13 answers @ 99% accuracy, 23 answers @ 90% accuracy
  - 5 cents/question, or \$3/hr to match advertising cost of unpaid users
- Knowledgeable users are much faster and more efficient

# Citizen Science Applications

- Google gives **\$10K/month** to nonprofits in ad budget
- Climate CoLab experiment running
  - Doubled traffic with only \$20/day
  - Targets political activist groups (not only climate)
- Additional experiments: Crowdcrafting, ebird, Weendy

# Conclusions

- New way to run crowdsourcing, targeting with ads
- Engages unpaid users, avoids problems with extrinsic rewards
- Provides access to expert users, not available labor platforms
- Experts not always professionals (e.g., Mayo Clinic users)
- Nonprofits can use Google Ad Grants to attract (for free) participants to citizen science projects