Adventures in Crowdsourcing

Panos Ipeirotis

Twitter: @ipeirotis

"A Computer Scientist in a Business School"
http://behind-the-enemy-lines.com

Thanks to: Jing Wang, Marios Kokkodis, Foster Provost, Josh Attenberg, Victor Sheng, Evgeniy Gabrilovich, Chun How Tan, Ari Kobren, Gabrielle Paolaci, Jesse Chandler

Broad Goal

Integrate machine and human intelligence

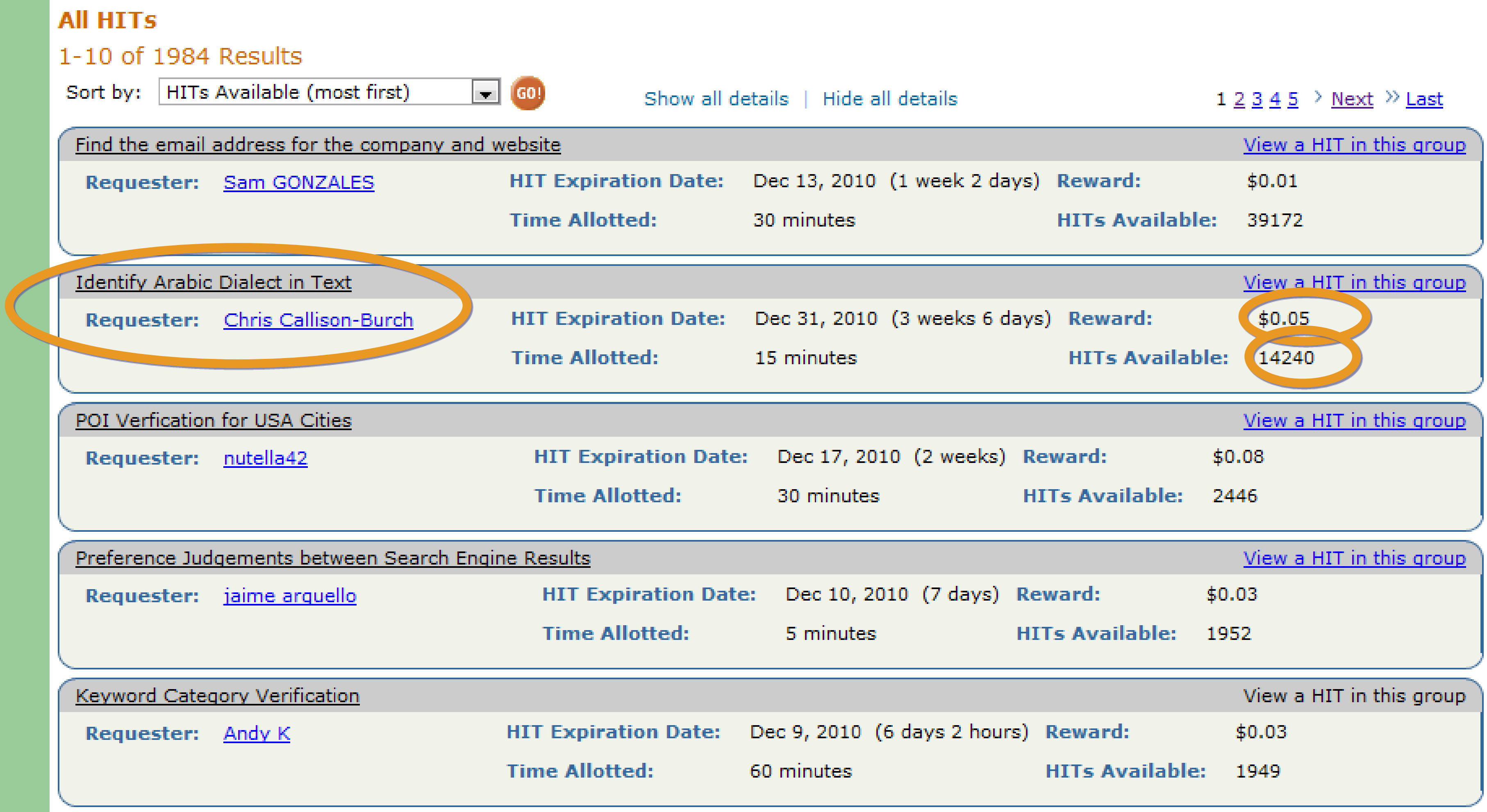
Create hybrid “intelligence integration” processes

With **paid** users and with **unpaid** users

Application
Detect Inappropriate content

- Need to detect inappropriate content
  - Ad placement, FB feed, links in forums, etc
- Ad hoc topics, with no existing training data
  - Hate speech, Violence, Guns & Bombs, Gossip…
- Classification models needed within *days*
- **Crowdsourcing** allows for fast data collection
  - using Mechanical Turk, oDesk, etc
  - labor is accessible on demand
  - but quality may be lower than experts

Amazon Mechanical Turk



All HITs

1-10 of 1984 Results

Sort by: HITs Available (most first) ▾ GO!

Show all details | Hide all details

1 2 3 4 5 › Next » Last

Find the email address for the company and website — View a HIT in this group

Requester:	Sam GONZALES	HIT Expiration Date:	Dec 13, 2010 (1 week 2 days)	Reward:	$0.01
		Time Allotted:	30 minutes	HITs Available:	39172

Identify Arabic Dialect in Text — View a HIT in this group

Requester:	Chris Callison-Burch	HIT Expiration Date:	Dec 31, 2010 (3 weeks 6 days)	Reward:	$0.05
		Time Allotted:	15 minutes	HITs Available:	14240

POI Verfication for USA Cities — View a HIT in this group

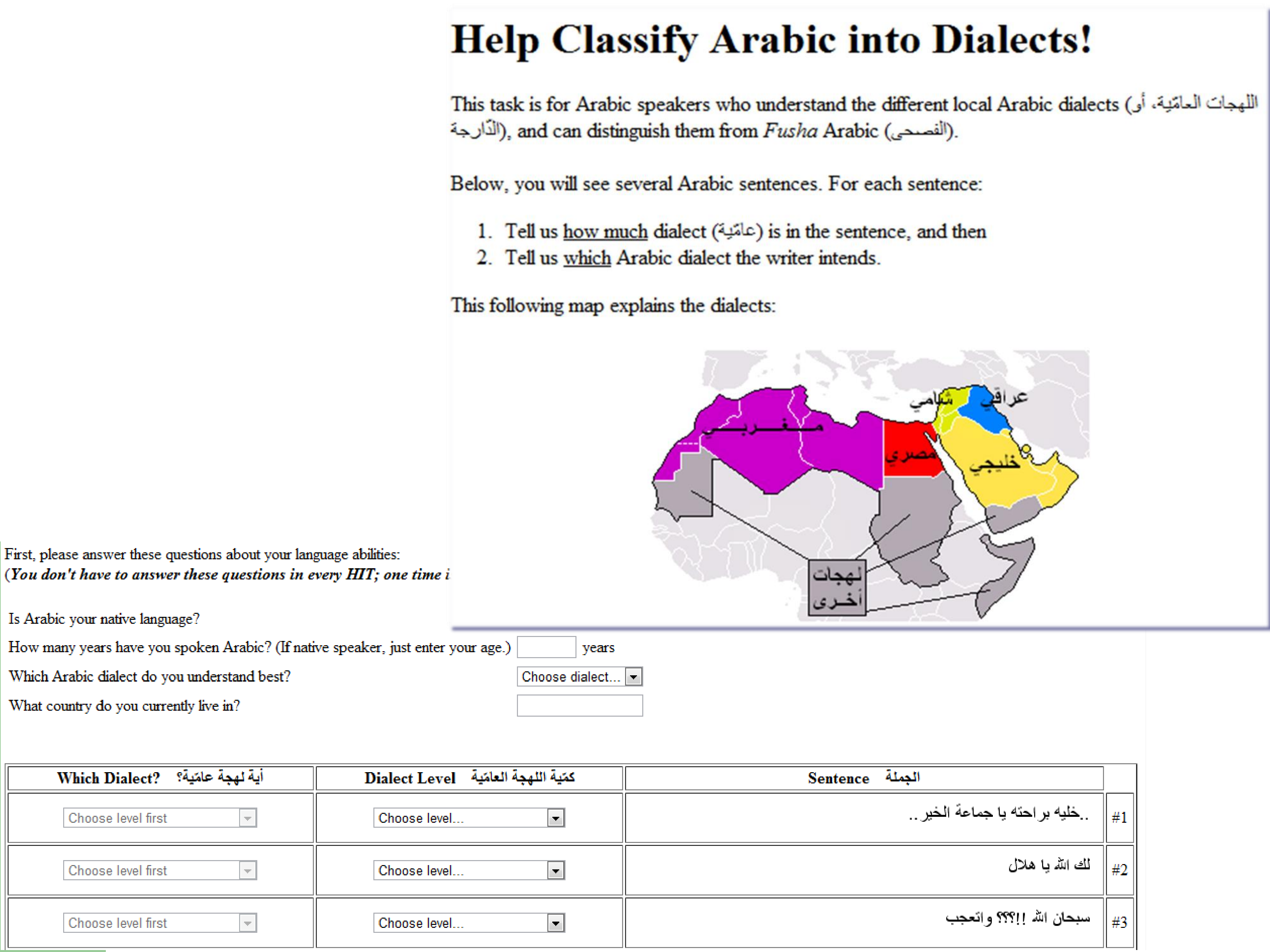
Requester:	nutella42	HIT Expiration Date:	Dec 17, 2010 (2 weeks)	Reward:	$0.08
		Time Allotted:	30 minutes	HITs Available:	2446

Preference Judgements between Search Engine Results — View a HIT in this group

Requester:	jaime arguello	HIT Expiration Date:	Dec 10, 2010 (7 days)	Reward:	$0.03
		Time Allotted:	5 minutes	HITs Available:	1952

Keyword Category Verification — View a HIT in this group

Requester:	Andy K	HIT Expiration Date:	Dec 9, 2010 (6 days 2 hours)	Reward:	$0.03
		Time Allotted:	60 minutes	HITs Available:	1949



Help Classify Arabic into Dialects!

This task is for Arabic speakers who understand the different local Arabic dialects (اللهجات العاميّة، أو الدّارجة), and can distinguish them from *Fusha* Arabic (الفصحى).

Below, you will see several Arabic sentences. For each sentence:

1. Tell us how much dialect (عاميّة) is in the sentence, and then
2. Tell us which Arabic dialect the writer intends.

This following map explains the dialects:

First, please answer these questions about your language abilities:
(***You don't have to answer these questions in every HIT; one time i***

Is Arabic your native language?

How many years have you spoken Arabic? (If native speaker, just enter your age.) years

Which Arabic dialect do you understand best? Choose dialect...

What country do you currently live in?

Which Dialect? أية لهجة عاميّة؟	**Dialect Level كميّة اللهجة العاميّة**	**Sentence الجملة**	
Choose level first	Choose level...	..خليه براحته يا جماعة الخير..	#1
Choose level first	Choose level...	لك الله يا هلال	#2
Choose level first	Choose level...	سبحان الله !!؟؟؟ واتعجب	#3

Example: Build an “Adult Content” Classifier

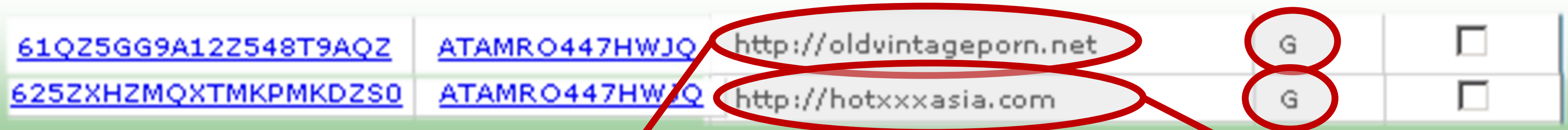
- Need a large number of labeled sites for training
- Get people to look at sites and label them as:

G (general audience) **PG** (parental guidance) **R** (restricted) **X** (porn)

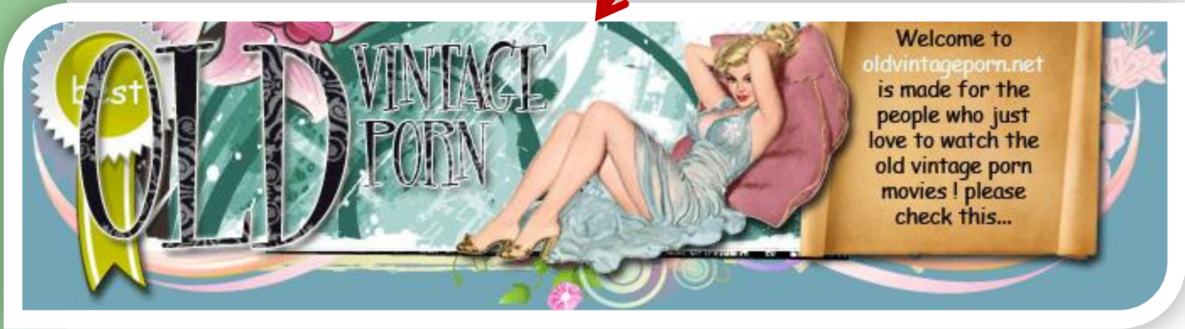
Cost/Speed Statistics

- **Undergrad intern**: 200 websites/hr, cost: $15/hr
- **Mechanical Turk**: 2500 websites/hr, cost: $12/hr

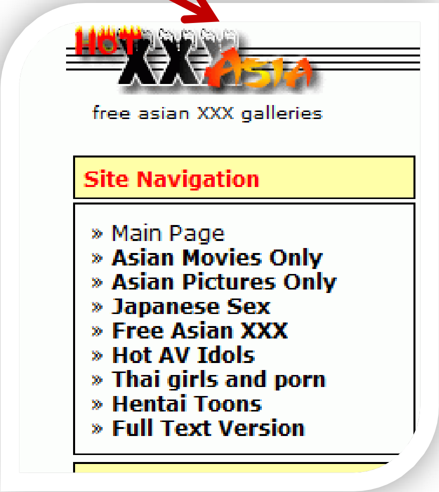
Bad news: Spammers!



61QZ5GG9A12Z548T9AQZ	ATAMRO447HWJQ	http://oldvintageporn.net	G	□
625ZXHZMQXTMKPMKDZS0	ATAMRO447HWJQ	http://hotxxxasia.com	G	□



Welcome to oldvintageporn.net is made for the people who just love to watch the old vintage porn movies ! please check this...



free asian XXX galleries

Site Navigation

- » Main Page
- » Asian Movies Only
- » Asian Pictures Only
- » Japanese Sex
- » Free Asian XXX
- » Hot AV Idols
- » Thai girls and porn
- » Hentai Toons
- » Full Text Version

Worker **ATAMRO447HWJQ**

labeled **X (porn)** sites as **G** (general audience)

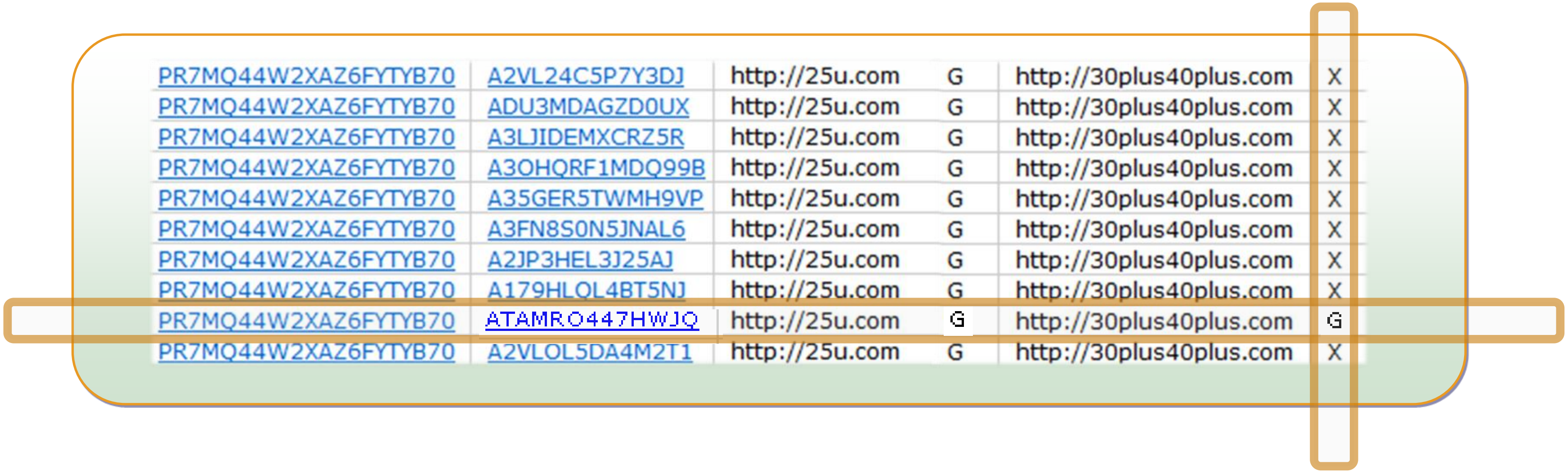
Challenges

- We do not know the true category for the objects
  - Available only after (costly) manual inspection
- We do not know quality of the workers

- We want to label objects with true categories
- We want (need?) to know the quality of the workers

Redundant votes, infer quality

Look at our lazy friend **ATAMRO447HWJQ** together with other 9 workers



PR7MQ44W2XAZ6FYTYB70	A2VL24C5P7Y3DJ	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	ADU3MDAGZD0UX	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	A3LJIDEMXCRZ5R	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	A3OHQRF1MDQ99B	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	A35GER5TWMH9VP	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	A3FN8S0N5JNAL6	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	A2JP3HEL3J25AJ	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	A179HLQL4BT5NJ	http://25u.com	G	http://30plus40plus.com	X
PR7MQ44W2XAZ6FYTYB70	ATAMRO447HWJQ	http://25u.com	G	http://30plus40plus.com	G
PR7MQ44W2XAZ6FYTYB70	A2VLOL5DA4M2T1	http://25u.com	G	http://30plus40plus.com	X

- Using redundancy, we can compute error rates for each worker

Expectation Maximization Estimation

Iterative process to estimate worker error rates

1. *Initialize "correct" label for each object (e.g., use majority vote)*
2. Estimate **error rates** for workers (using "correct" labels)
3. Estimate **"correct" labels** (using error rates, weight worker votes according to quality)
4. Go to Step 2 and iterate until convergence

Error rates for ATAMRO447HWJQ

P[G → G]=99.947% **P[G → X]=0.053%**

P[X → G]=99.153% **P[X → X]=0.847%**

Our friend ATAMRO447HWJQ
marked **almost all** sites as **G**.
Clickety clickey click…

Challenge: Humans are biased!

Error rates for the CEO, providing **“expert” labels**

P[G → G]=20.0%	**P[G → P]=80.0%**	P[G → R]=0.0%	P[G → X]=0.0%
P[P → G]=0.0%	P[P → P]=**0.0%**	**P[P → R]=100.0%**	P[P → X]=0.0%
P[R → G]=0.0%	P[R → P]=0.0%	**P[R → R]=100.0%**	P[R → X]=0.0%
P[X → G]=0.0%	P[X → P]=0.0%	P[X → R]=0.0%	**P[X → X]=100.0%**

- We have 85% G sites, 5% P sites, 5% R sites, 5% X sites

- Error rate of spammer (all G) = 0% * 85% + 100% * 15% = 15%
- Error rate of biased worker = 80% * 85% + 100% * 5% = 73%

False positives: Legitimate workers appear to be spammers

(important note: bias is not just a matter of “ordered” classes)

Solution: Fix bias first, compute error rate afterwards

Error Rates for CEO of AdSafe

P[G → G]=20.0%	**P[G → P]=80.0%**	P[G → R]=0.0%	P[G → X]=0.0%
P[P → G]=0.0%	P[P → P]=**0.0%**	**P[P → R]=100.0%**	P[P → X]=0.0%
P[R → G]=0.0%	P[R → P]=0.0%	**P[R → R]=100.0%**	P[R → X]=0.0%
P[X → G]=0.0%	P[X → P]=0.0%	P[X → R]=0.0%	**P[X → X]=100.0%**

- When biased worker says G, it is **100% G**
- When biased worker says P, it is **100% G**
- When biased worker says R, it is **50% P, 50% R**
- When biased worker says X, it is **100% X**

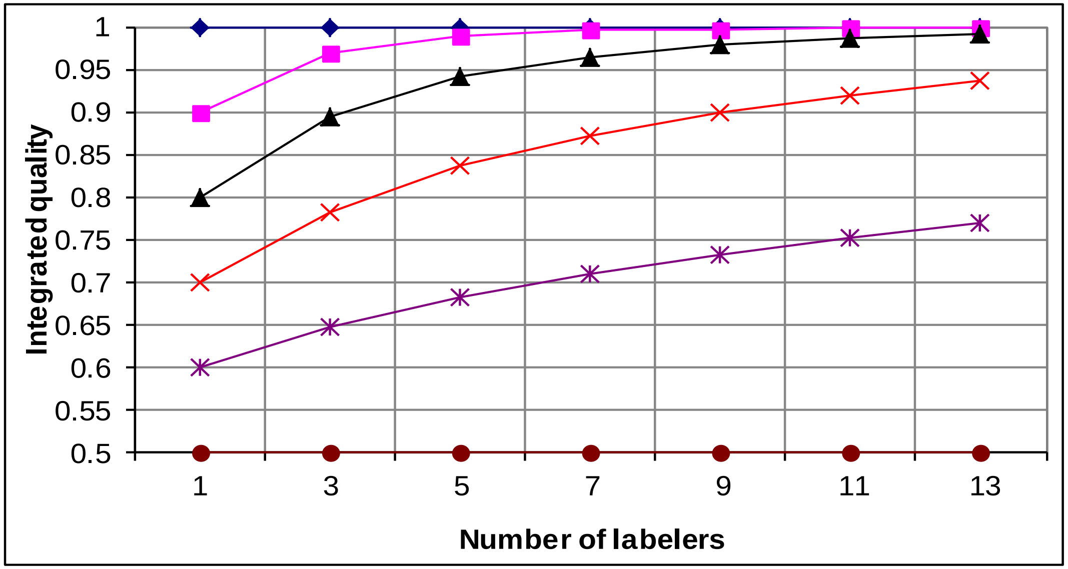
Small ambiguity for "R-rated" votes but other than that, fine!

Question: How to pay workers?

- Naïve solution: Have a quality-score threshold
- **Threshold-ing rewards gives wrong incentives:**
  - Decent (but still useful) workers get fired
  - Uncertainty near the decision threshold

Quality-sensitive Payment

- Set quality goal and price (e.g., $1 for 90%)
  - For workers above goal: Pay full price
  - For others: Payment divided with redundancy needed to reach goal
    - Need 3 workers with 80% accuracy ➔ Payment = $0.33
    - Need 9 workers with 70% accuracy ➔ Payment = $0.11



Instead of blocking: Quality-sensitive Payment

- **Estimate payment level based on quality**
  - Set acceptable quality (e.g., 99% accuracy)
  - For workers above quality specs: Pay full price
  - For others: Estimate level of redundancy to reach acceptable quality (e.g., Need 5 workers with 90% accuracy or 13 workers with 80% accuracy to reach 99% accuracy;)
  - Pay full price divided by level of necessary redundancy
- **Uncertainty** penalty: **Pay less** for uncertain estimates (for workers with short working histories)
- **Refund** underpayment when quality estimate more certain

Real-Time Payment and Reimbursement

Example of the piece-rate payment of a worker

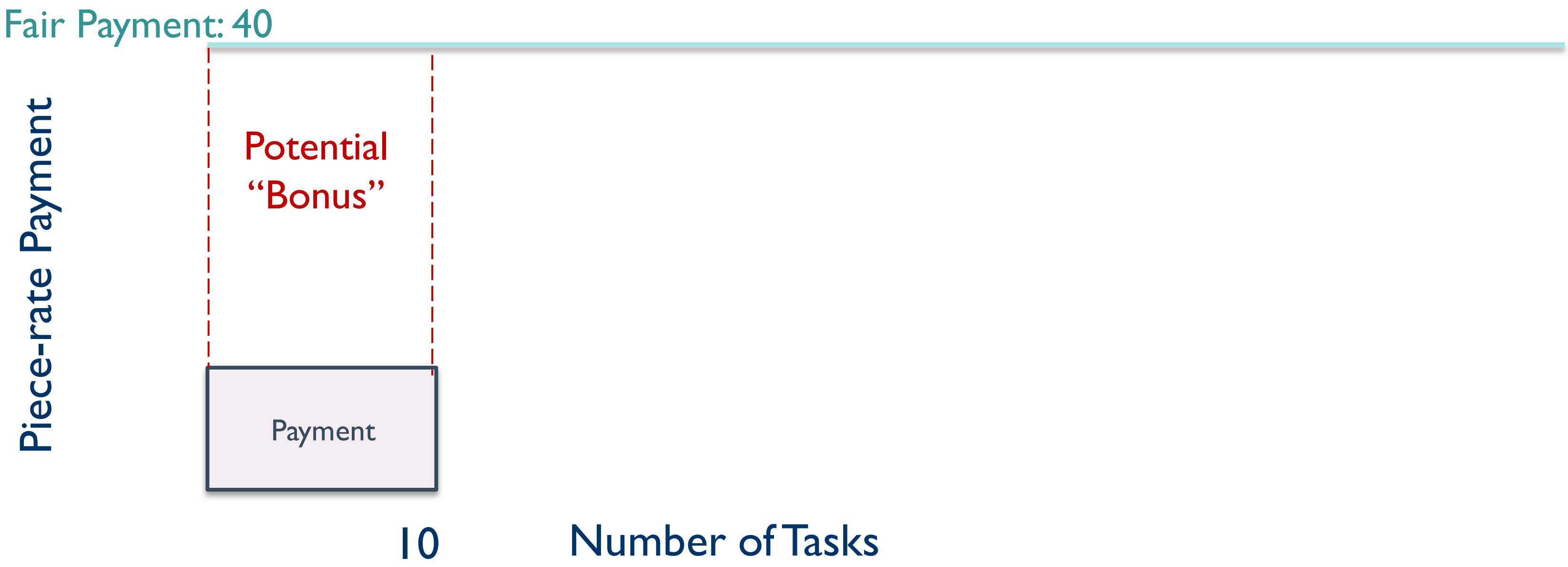
# Tasks	10	20	30	40	Infinity
Piece-rate Payment (cents)	11	18	21	23	40

Fair Payment

Real-Time Payment and Reimbursement

Example of the piece-rate payment of a worker

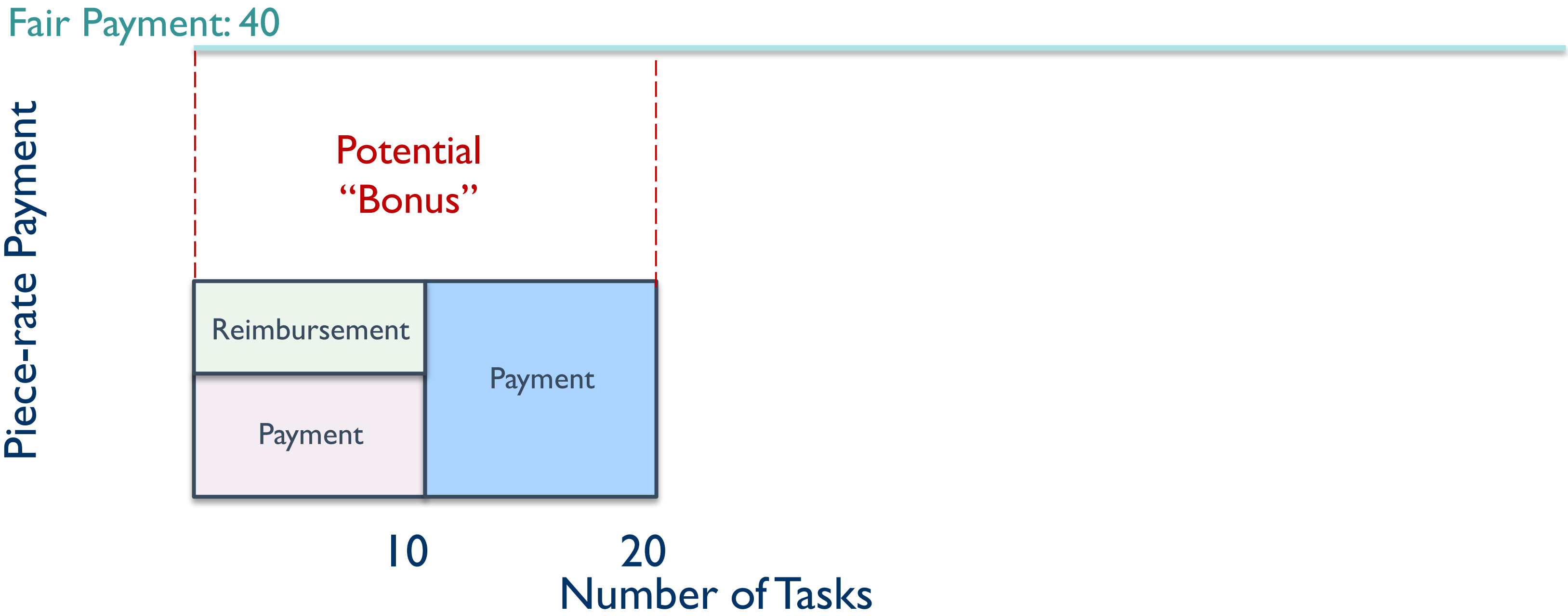
# Tasks	10	20	30	40	Infinity
Piece-rate Payment (cents)	11	18	21	23	40



Real-Time Payment and Reimbursement

Example of the piece-rate payment of a worker

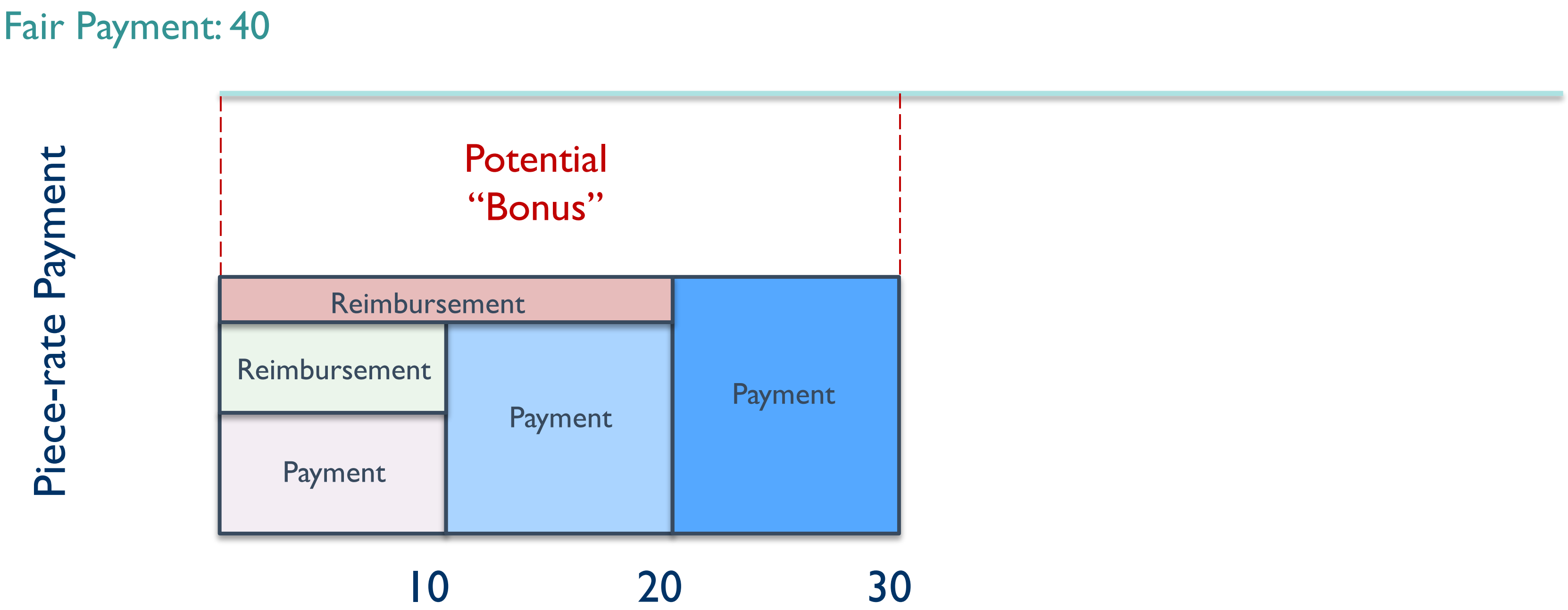
# Tasks	10	20	30	40	Infinity
Piece-rate Payment (cents)	11	18	21	23	40



Real-Time Payment and Reimbursement

Example of the piece-rate payment of a worker

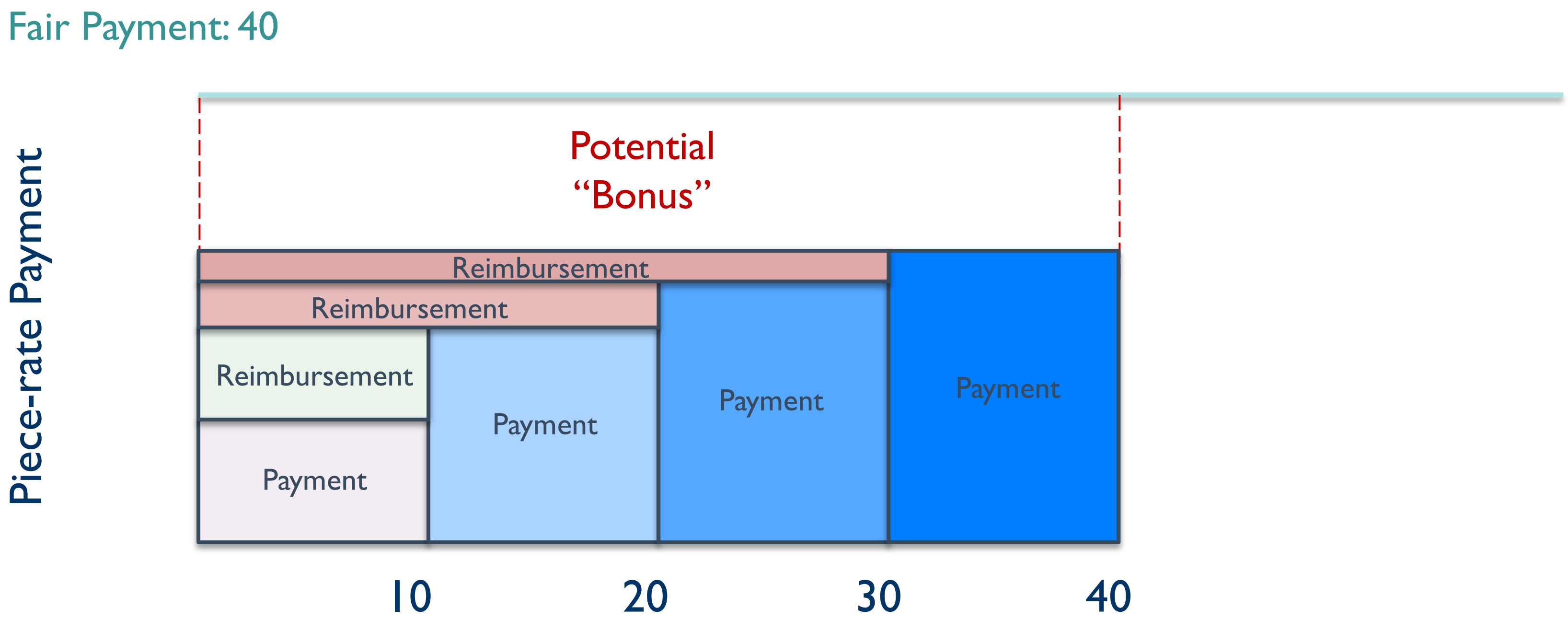
# Tasks	10	20	30	40	Infinity
Piece-rate Payment (cents)	11	18	21	23	40



Real-Time Payment and Reimbursement

Example of the piece-rate payment of a worker

# Tasks	10	20	30	40	Infinity
Piece-rate Payment (cents)	11	18	21	23	40



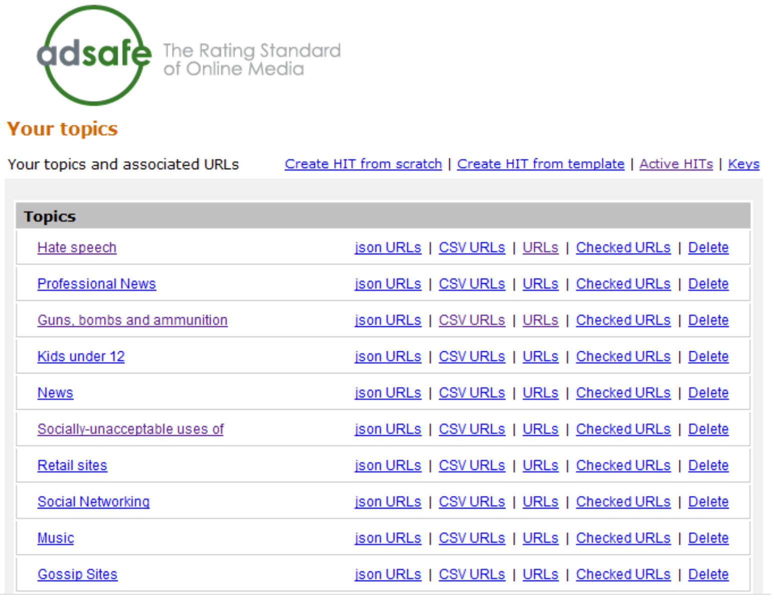
Improving worker participation

- With just labeling, workers are **passively** labeling the data that we give them
- But this can be wasteful when positive cases are sparse
- Why not asking the workers to search themselves and **find training data**

Guided Learning

Ask workers to ***find*** example web pages (great for "sparse" content)

After collecting enough examples, easy to build and test web page classifier



adsafe The Rating Standard of Online Media

Your topics

Your topics and associated URLs   Create HIT from scratch | Create HIT from template | Active HITs | Keys

Topics	
Hate speech	json URLs | CSV URLs | URLs | Checked URLs | Delete
Professional News	json URLs | CSV URLs | URLs | Checked URLs | Delete
Guns, bombs and ammunition	json URLs | CSV URLs | URLs | Checked URLs | Delete
Kids under 12	json URLs | CSV URLs | URLs | Checked URLs | Delete
News	json URLs | CSV URLs | URLs | Checked URLs | Delete
Socially-unacceptable uses of	json URLs | CSV URLs | URLs | Checked URLs | Delete
Retail sites	json URLs | CSV URLs | URLs | Checked URLs | Delete
Social Networking	json URLs | CSV URLs | URLs | Checked URLs | Delete
Music	json URLs | CSV URLs | URLs | Checked URLs | Delete
Gossip Sites	json URLs | CSV URLs | URLs | Checked URLs | Delete

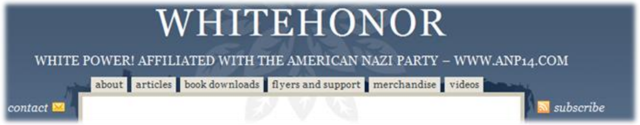
http://url-collector.appspot.com/allTopics.jsp

Limits of Guided Learning

- No incentives for workers to find “new” content
- After a while, submitted web pages similar to already submitted ones
- No improvement for classifier

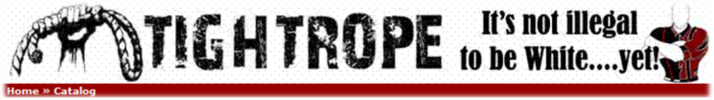
The result? Blissful ignorance…

- Classifier ***seems*** great: Cross-validation tests show excellent performance



WHITEHONOR

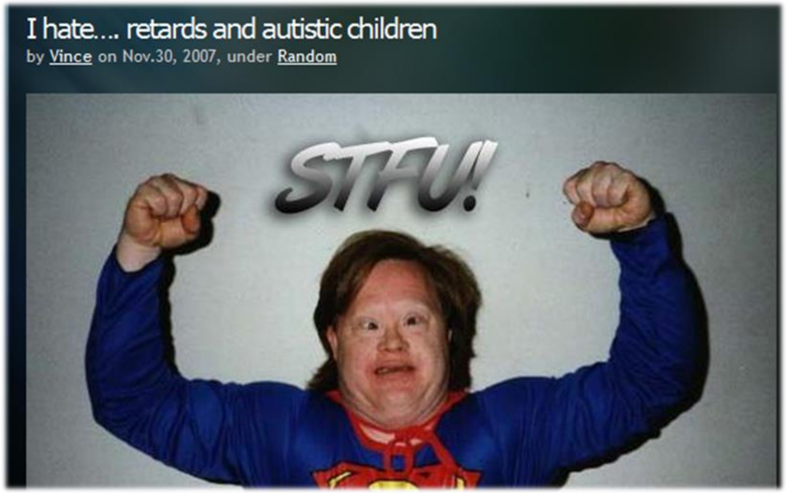
WHITE POWER! AFFILIATED WITH THE AMERICAN NAZI PARTY – WWW.ANP14.COM



TIGHTROPE

It's not illegal to be White....yet!

- Alas, classifier fails: The "*unknown unknowns*" ™



I hate.... retards and autistic children

by Vince on Nov.30, 2007, under Random

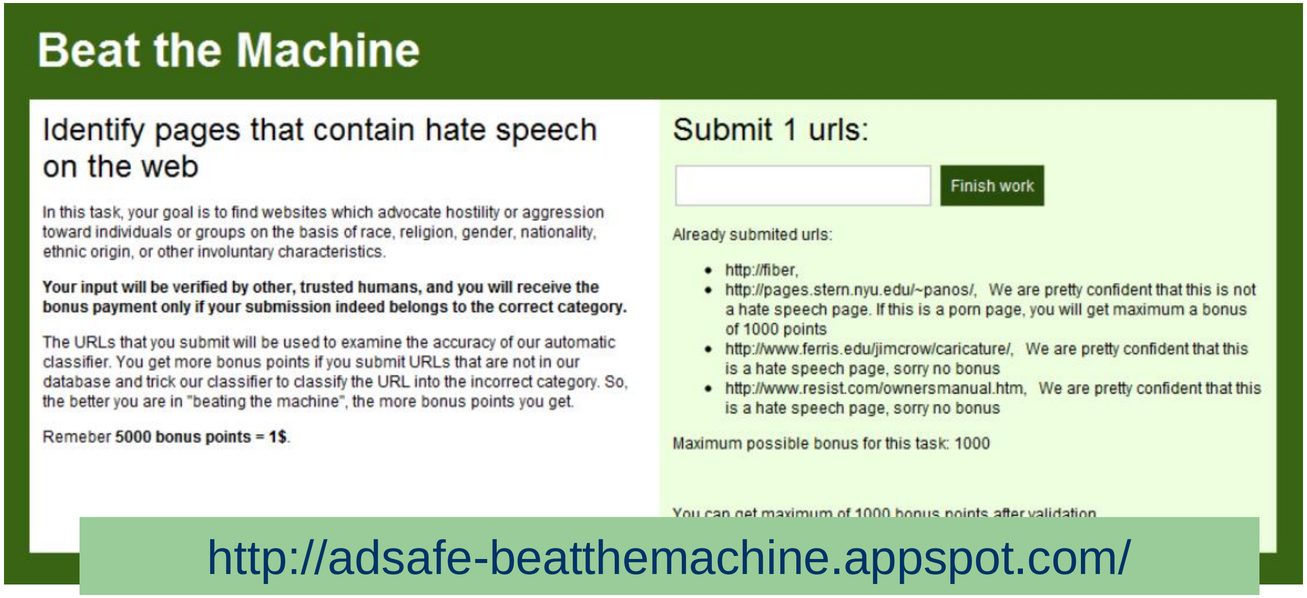
No similar training data in training set

"*Unknown unknowns*" = classifier fails with high confidence

Beat the Machine!

Ask humans to find URLs that

- ***the classifier will classify incorrectly***
- ***another human will classify correctly***



Beat the Machine

Identify pages that contain hate speech on the web

In this task, your goal is to find websites which advocate hostility or aggression toward individuals or groups on the basis of race, religion, gender, nationality, ethnic origin, or other involuntary characteristics.

Your input will be verified by other, trusted humans, and you will receive the bonus payment only if your submission indeed belongs to the correct category.

The URLs that you submit will be used to examine the accuracy of our automatic classifier. You get more bonus points if you submit URLs that are not in our database and trick our classifier to classify the URL into the incorrect category. So, the better you are in "beating the machine", the more bonus points you get.

Remeber **5000 bonus points = 1$**.

Submit 1 urls:

Already submited urls:

- http://fiber,
- http://pages.stern.nyu.edu/~panos/, We are pretty confident that this is not a hate speech page. If this is a porn page, you will get maximum a bonus of 1000 points
- http://www.ferris.edu/jimcrow/caricature/, We are pretty confident that this is a hate speech page, sorry no bonus
- http://www.resist.com/ownersmanual.htm, We are pretty confident that this is a hate speech page, sorry no bonus

Maximum possible bonus for this task: 1000

You can get maximum of 1000 bonus points after validation

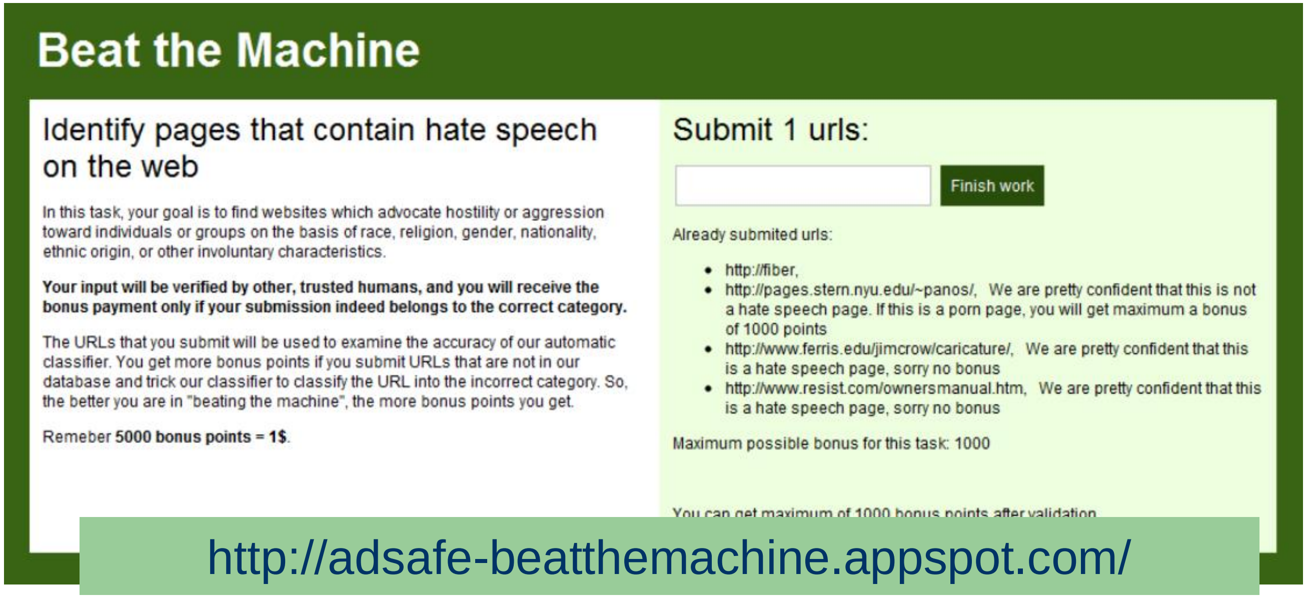
http://adsafe-beatthemachine.appspot.com/

Example:
Find hate speech pages that the machine will classify as benign

Beat the Machine!

Incentive structure:

- ***$1 if you "beat the machine"***
- ***$0.001 if the machine already knows***



Beat the Machine

Identify pages that contain hate speech on the web

In this task, your goal is to find websites which advocate hostility or aggression toward individuals or groups on the basis of race, religion, gender, nationality, ethnic origin, or other involuntary characteristics.

Your input will be verified by other, trusted humans, and you will receive the bonus payment only if your submission indeed belongs to the correct category.

The URLs that you submit will be used to examine the accuracy of our automatic classifier. You get more bonus points if you submit URLs that are not in our database and trick our classifier to classify the URL into the incorrect category. So, the better you are in "beating the machine", the more bonus points you get.

Remeber **5000 bonus points = 1$**.

Submit 1 urls:

Already submited urls:

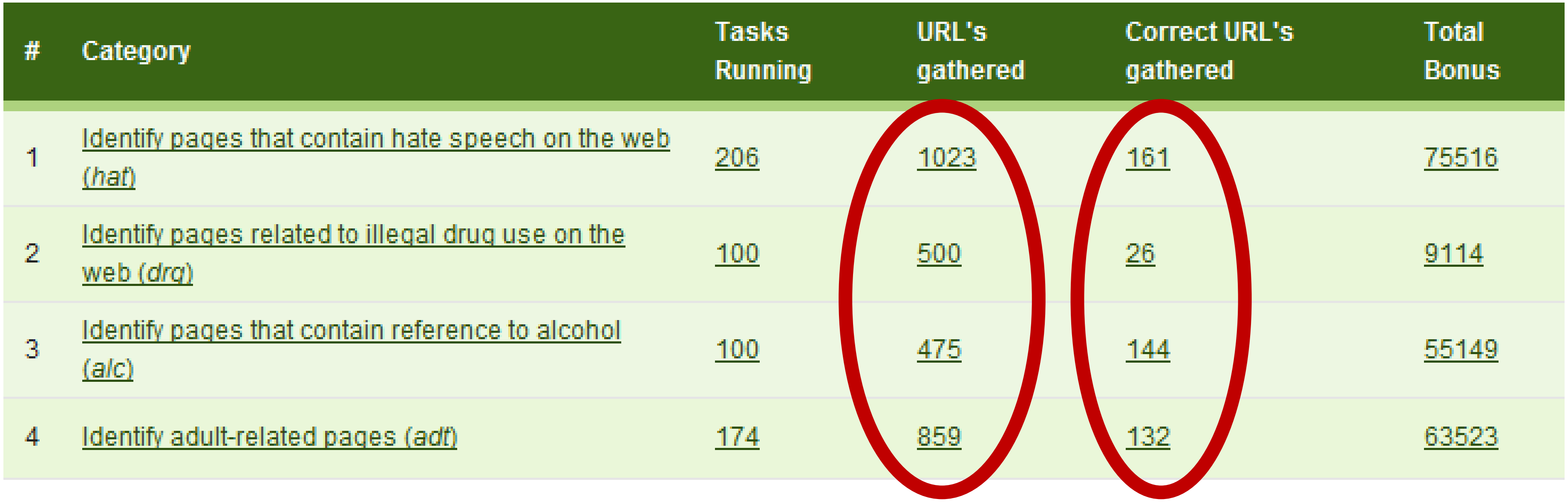
- http://fiber,
- http://pages.stern.nyu.edu/~panos/, We are pretty confident that this is not a hate speech page. If this is a porn page, you will get maximum a bonus of 1000 points
- http://www.ferris.edu/jimcrow/caricature/, We are pretty confident that this is a hate speech page, sorry no bonus
- http://www.resist.com/ownersmanual.htm, We are pretty confident that this is a hate speech page, sorry no bonus

Maximum possible bonus for this task: 1000

You can get maximum of 1000 bonus points after validation

http://adsafe-beatthemachine.appspot.com/

Example:
Find hate speech pages that the machine will classify as benign



#	Category	Tasks Running	URL's gathered	Correct URL's gathered	Total Bonus
1	Identify pages that contain hate speech on the web (*hat*)	206	1023	161	75516
2	Identify pages related to illegal drug use on the web (*drg*)	100	500	26	9114
3	Identify pages that contain reference to alcohol (*alc*)	100	475	144	55149
4	Identify adult-related pages (*adt*)	174	859	132	63523

Probes

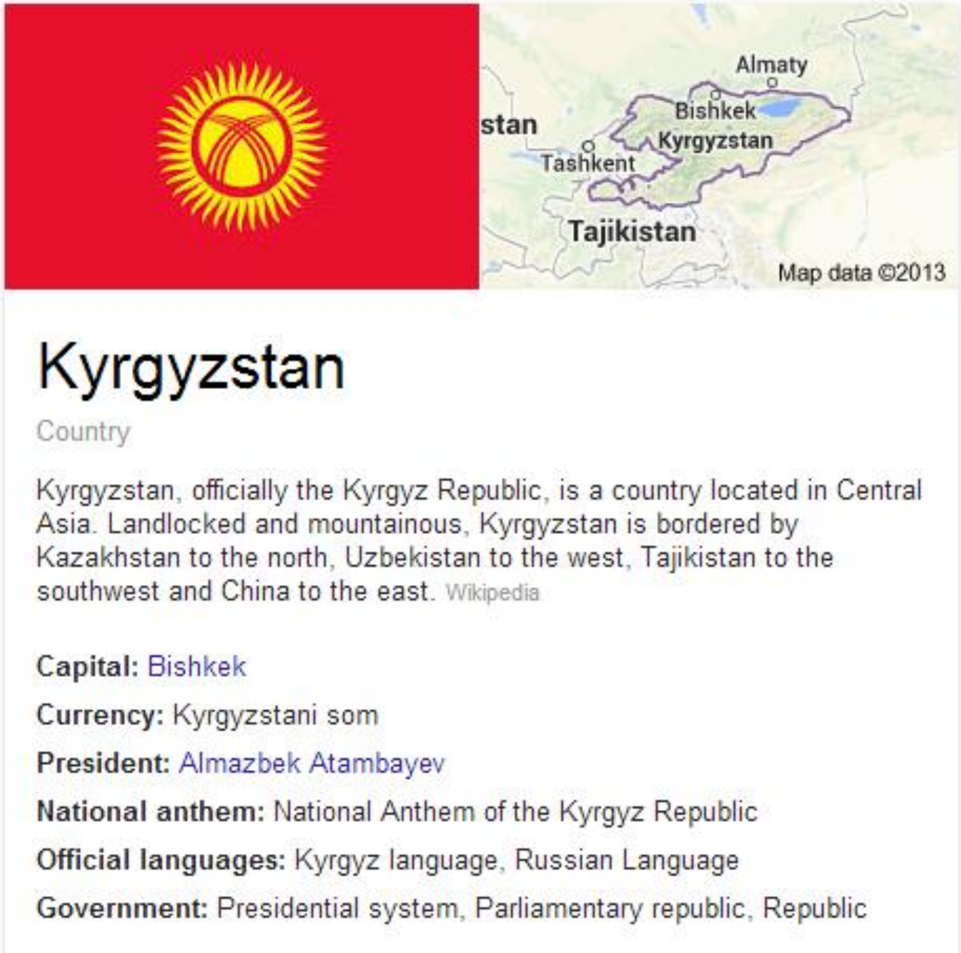
Successes

Error rate for probes significantly higher than error rate on (stratified) random data (10x to 100x higher than base error rate)

No money?

- What if we want to engage users without paying them?

Google Knowledge Graph



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

“Things not Strings”

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”
- …

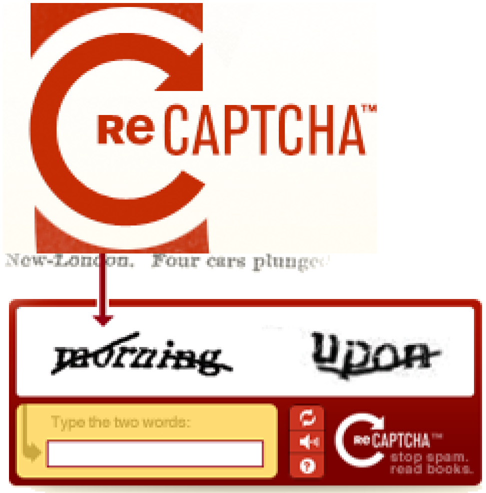
The Google mission…

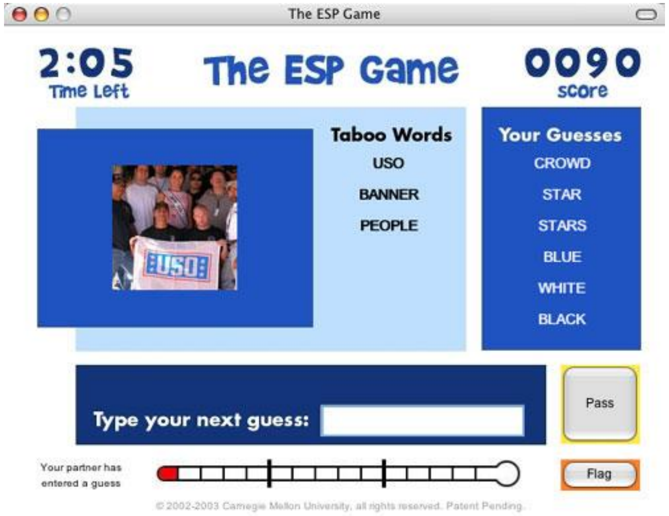
We have a billion users…
Leverage their knowledge…

“Let’s create a new crowdsourcing system…”

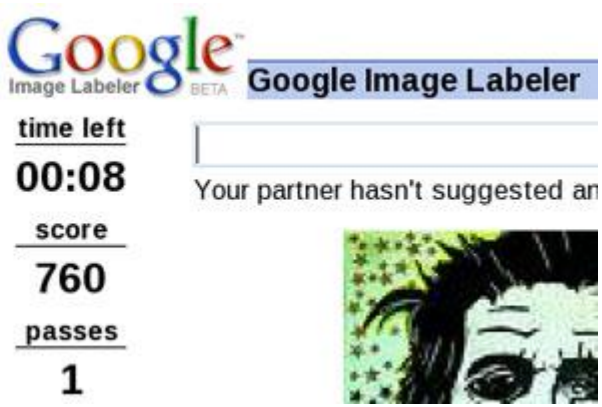
Ideally…

W
Ω
И
維
ウィ


ReCAPTCHA™
New-London. Four cars plunged
morning
upon
Type the two words:
ReCAPTCHA™
stop spam.
read books.


The ESP Game
2:05
Time Left
The ESP Game
0090
score
Taboo Words
USO
BANNER
PEOPLE
Your Guesses
CROWD
STAR
STARS
BLUE
WHITE
BLACK
Type your next guess:
Pass
Your partner has entered a guess
Flag
© 2002-2003 Carnegie Mellon University, all rights reserved. Patent Pending.

But often…



Google Image Labeler

time left
00:08
score
760
passes
1

Your partner hasn't suggested an

The common solution…

amazon mechanical turk
beta
Artificial Artificial Intelligence

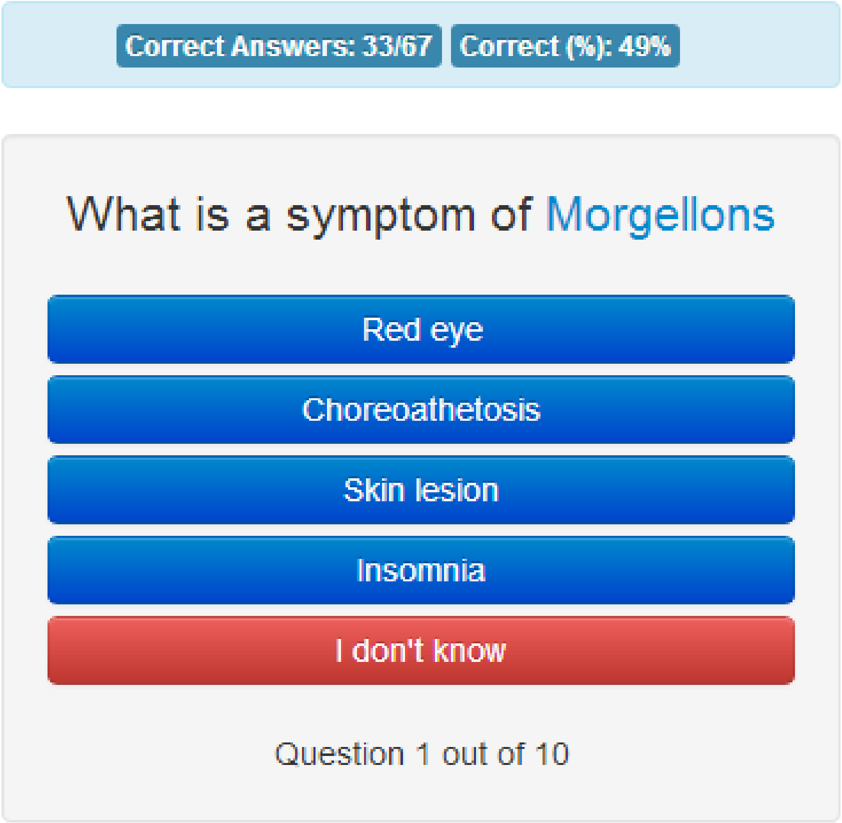
CrowdFlower

oDesk®
Love the way you work.

Key Challenge

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

www.quizz.us



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

What is a symptom of Morgellons

Question 1 out of 10

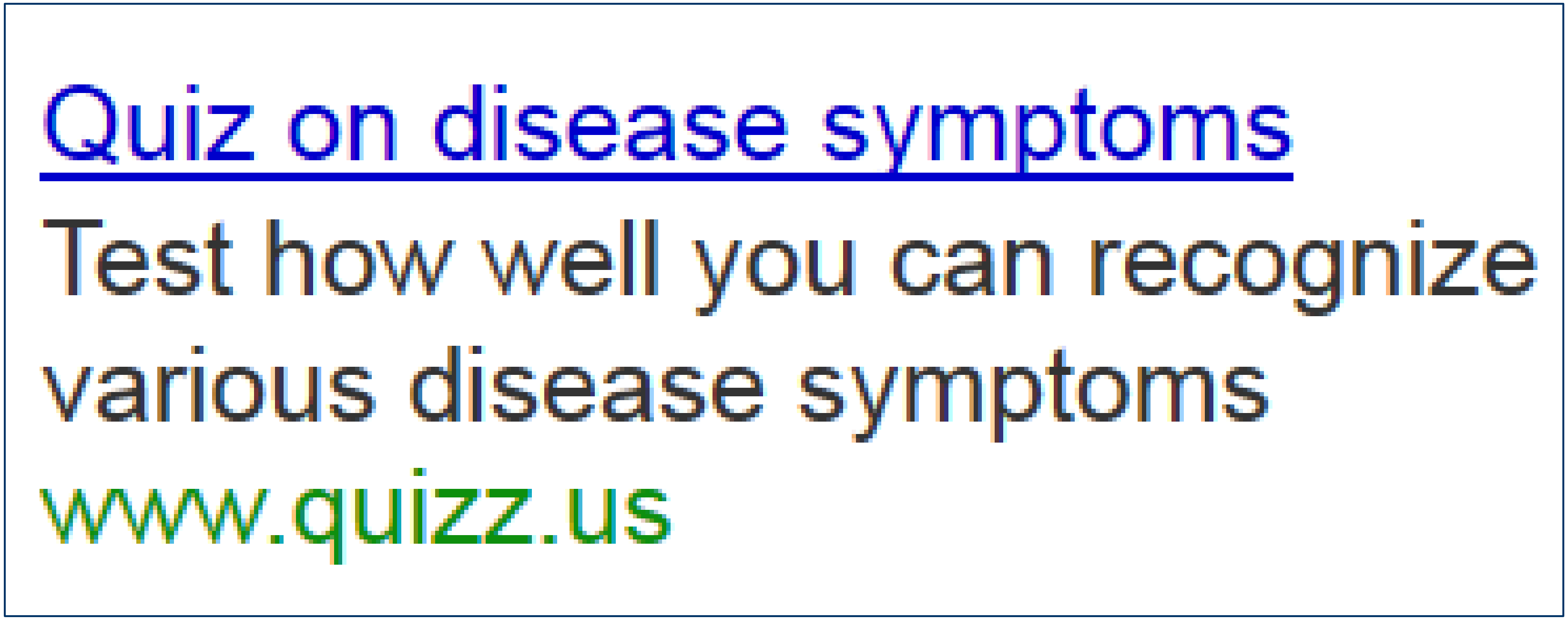
Calibration vs. Collection

- **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*

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

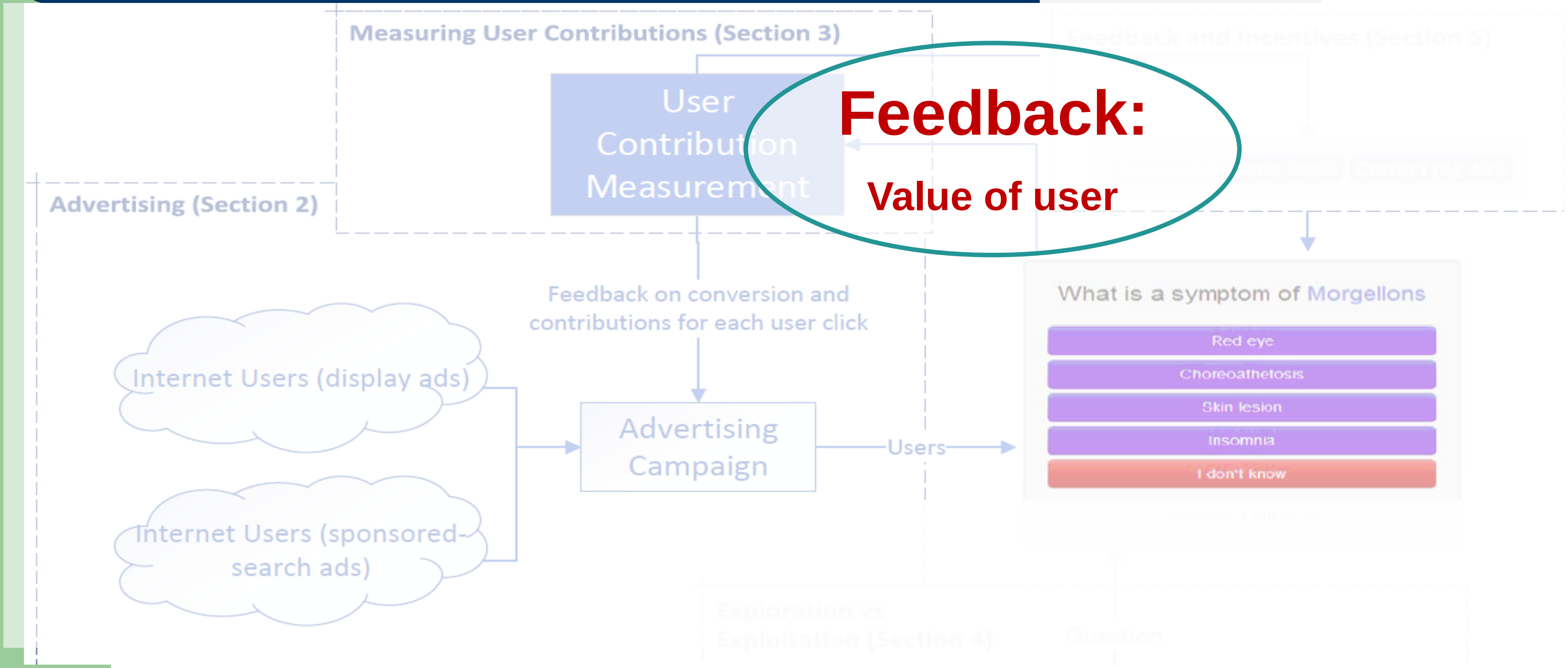


Quiz on disease symptoms

Test how well you can recognize various disease symptoms

www.quizz.us

Treat Quizz as eCommerce Site


Measuring User Contributions (Section 3)
User Contribution Measurement
Feedback and Incentives (Section 5)
Advertising (Section 2)
Feedback on conversion and contributions for each user click
Internet Users (display ads)
Internet Users (sponsored-search ads)
Advertising Campaign
Users
What is a symptom of Morgellons
Red eye
Choreoathetosis
Skin lesion
Insomnia
I don't know
Exploration vs Exploitation (Section 4)
Question

Feedback:

Value of user

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)

Immediate feedback helps

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%

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

Showing score moderately helpful

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%

- Be careful what you incentivize ☺
- “Total Correct” incentivizes quantity, not quality

Competitiveness helps

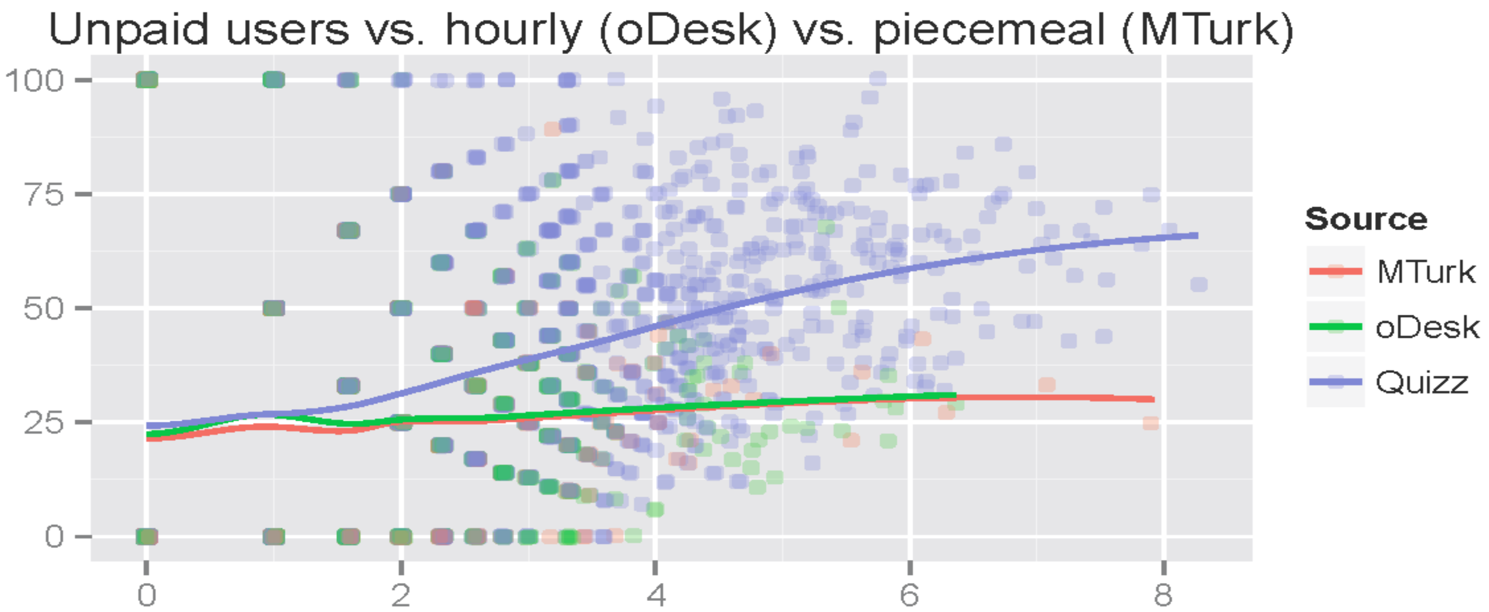
Treatment	Effect
Show if user answer correct	+2.4%
Show the correct answer	+20.4%
Score: % of correct answers	+2.3%
Score: # of correct answers	-2.2%
Score: Information gain	+4.0%
Show statistics for performance of other users	**+9.8%**
Leaderboard based on percent correct	-4.8%
Leaderboard based on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

Leaderboards are tricky!

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%**

- Initially, strong positive effect
- Over time, effect became strongly negative
- All-time leaderboards considered harmful

Comparison with paid crowdsourcing



% correct

Submitted answers (log)

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

How can I get rid of users?

National Academy of Sciences “Frontiers of Science” conference

Your workers behave like my mice!

An unexpected connection…

Your workers behave like my mice!

Eh?

Your workers want to use only their **motor skills**, **not** their **cognitive skills**

The Biology Fundamentals

- Brain functions are biologically expensive (20% of total energy consumption in humans)
- Motor skills are more energy efficient than cognitive skills (e.g., walking)
- Brain tends to delegate easy tasks to part of the neural system that handles motor skills

An unexpected connection at the NAS “Frontiers of Science” conf.

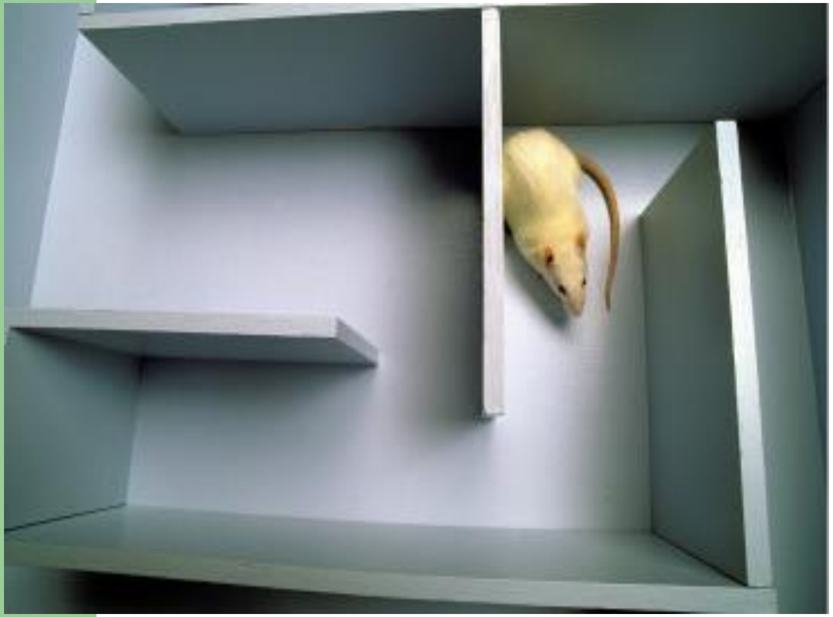
Your workers want to use only their **motor skills**, **not** their **cognitive skills**

Makes sense

An unexpected connection at the NAS “Frontiers of Science” conf.

And here is how I train my mice to behave…

The Mice Experiment



Cognitive

Solve maze
Find pellet

Motor

Push lever three times
Pellet drops

How to Train the Mice?

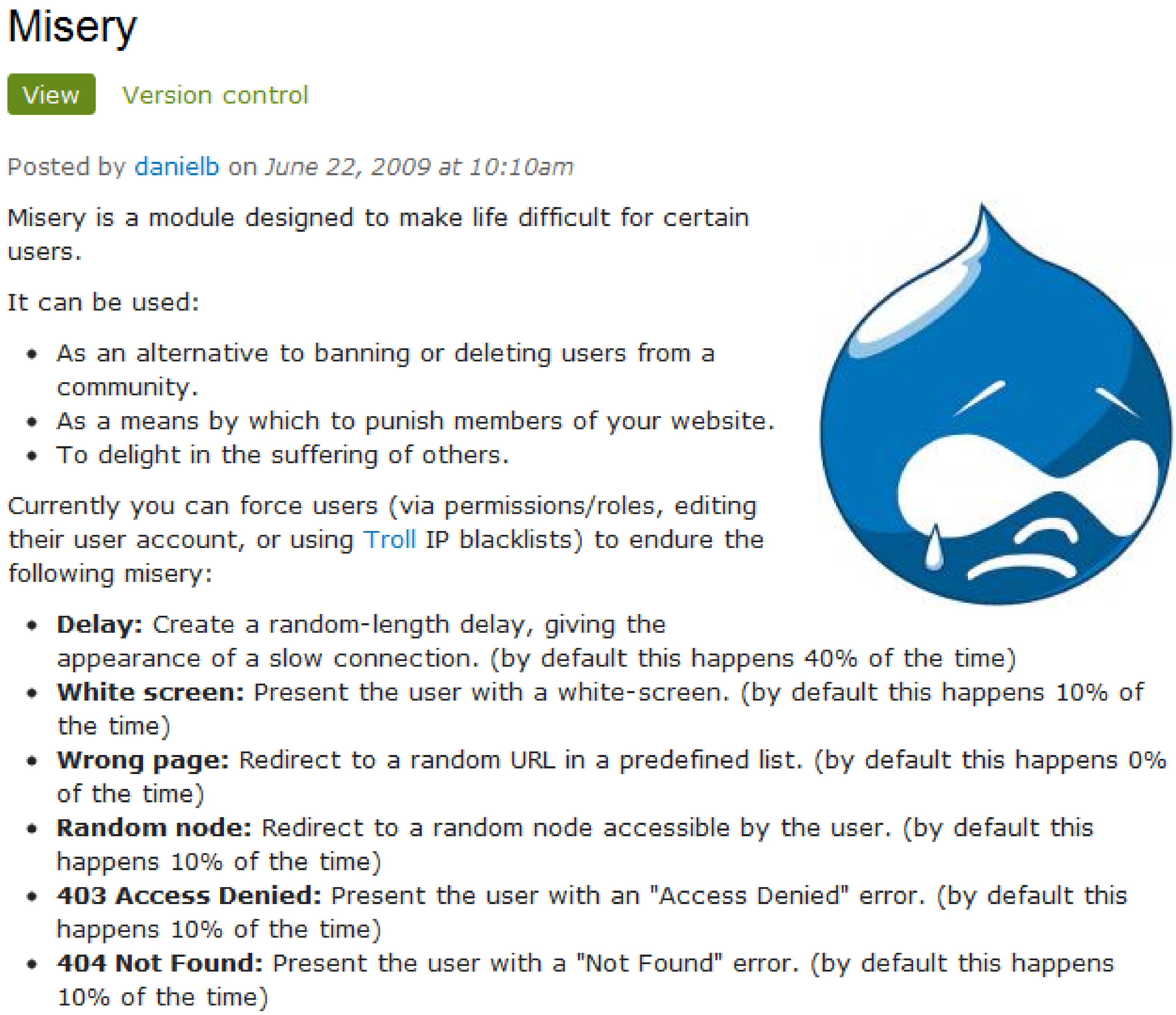
Confuse motor skills!
Reward cognition!

I should try this the moment that I get back to my room

Punishing Worker’s Motor Skills

- **Punish bad answers** with frustration of motor skills (e.g., add delays between tasks)
  - “Loading image, please wait…”
  - “Image did not load, press here to reload”
  - “404 error. Return the HIT and accept again”

→Make this **probabilistic** to keep feedback implicit



Misery

Posted by danielb on *June 22, 2009 at 10:10am*

Misery is a module designed to make life difficult for certain users.

It can be used:

- As an alternative to banning or deleting users from a community.
- As a means by which to punish members of your website.
- To delight in the suffering of others.

Currently you can force users (via permissions/roles, editing their user account, or using Troll IP blacklists) to endure the following misery:

- **Delay:** Create a random-length delay, giving the appearance of a slow connection. (by default this happens 40% of the time)
- **White screen:** Present the user with a white-screen. (by default this happens 10% of the time)
- **Wrong page:** Redirect to a random URL in a predefined list. (by default this happens 0% of the time)
- **Random node:** Redirect to a random node accessible by the user. (by default this happens 10% of the time)
- **403 Access Denied:** Present the user with an "Access Denied" error. (by default this happens 10% of the time)
- **404 Not Found:** Present the user with a "Not Found" error. (by default this happens 10% of the time)

Experimental Summary (I)

- Spammer workers quickly abandon
  - No need to display scores, or ban
  - Low quality submissions from ~60% to ~3%
  - Half-life of low-quality from 100+ HITs to less than 5
- Good workers unaffected
  - No significant effect on participation of workers with good performance
  - Lifetime of participants unaffected
  - Longer response times (***after*** removing the “intervention delays”; that was puzzling)

Experimental Summary (II)

- Remember, scheme was for ***training*** the mice…
- Indeed, 15%-20% of the spammers start submitting good work!

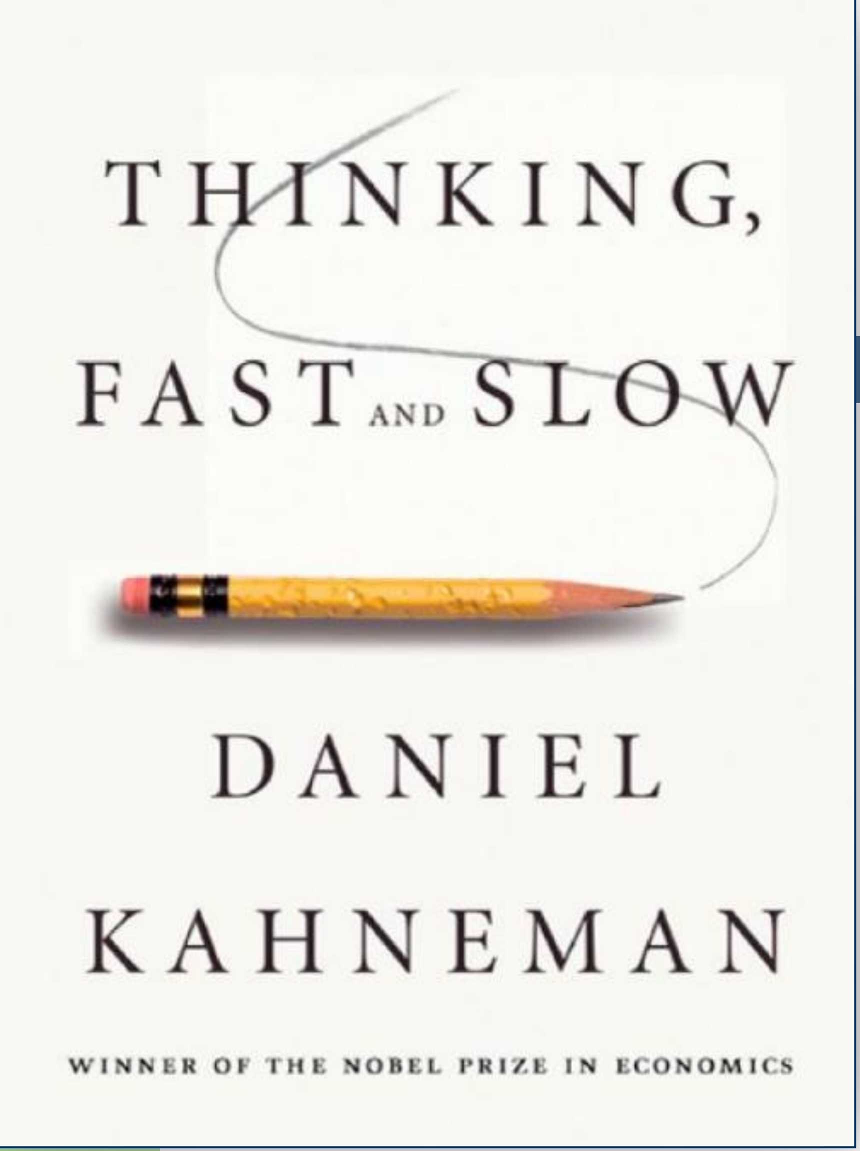
????

Two key questions

- Why response time was slower for some good workers?

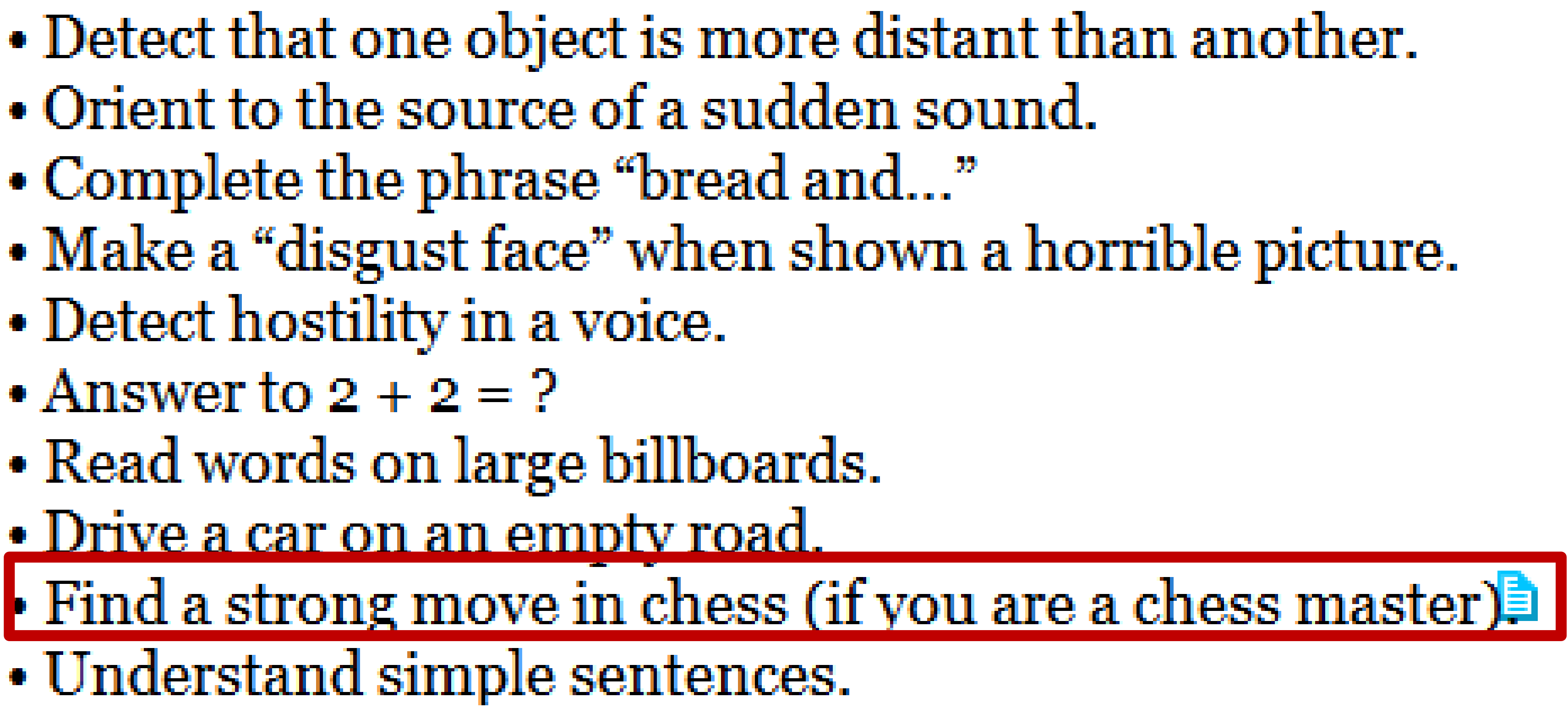
- Why some low quality workers start working well?

????



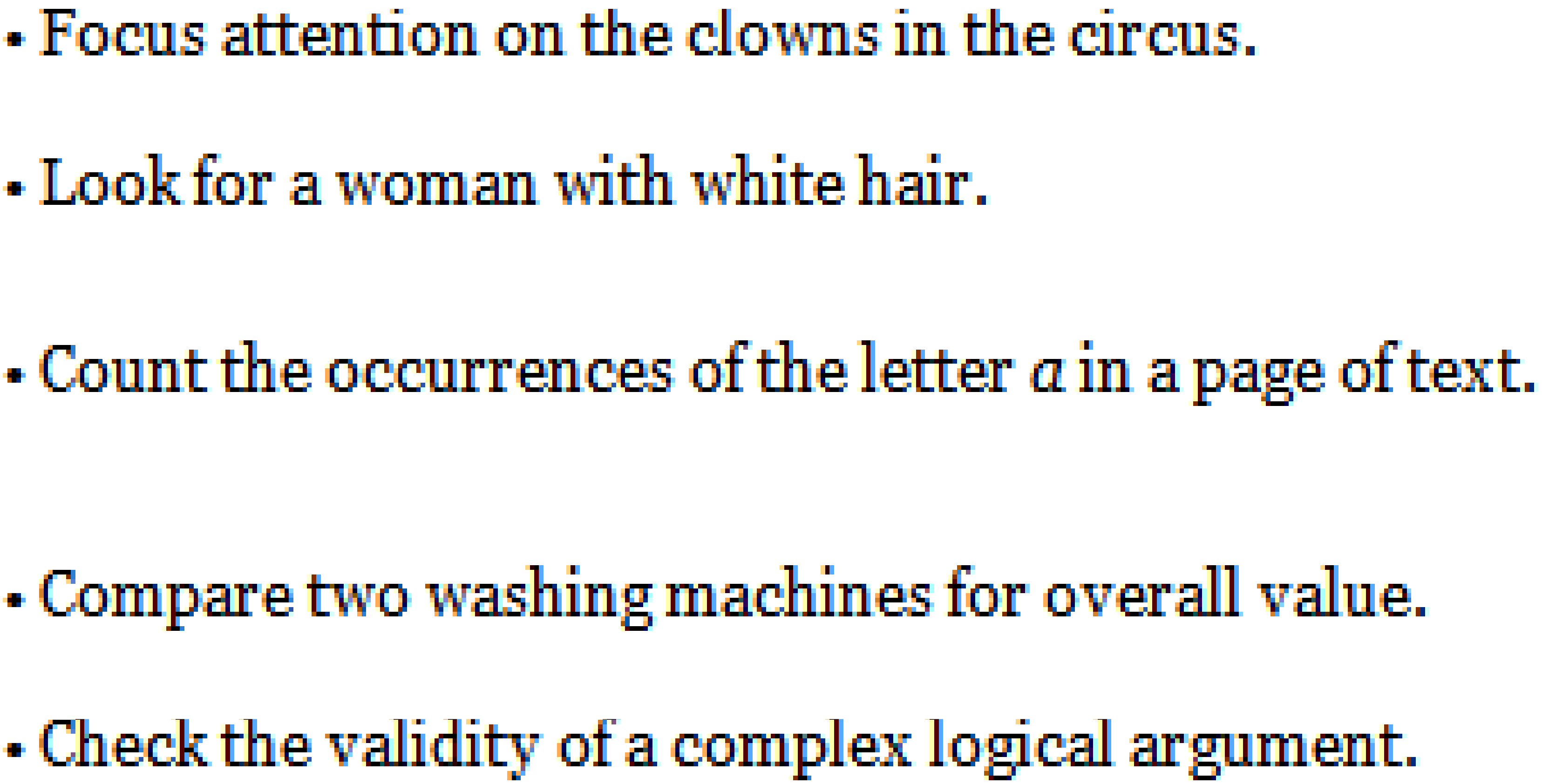
- System 1: “Automatic” actions
- System 2: “Intelligent” actions

System 1 Tasks

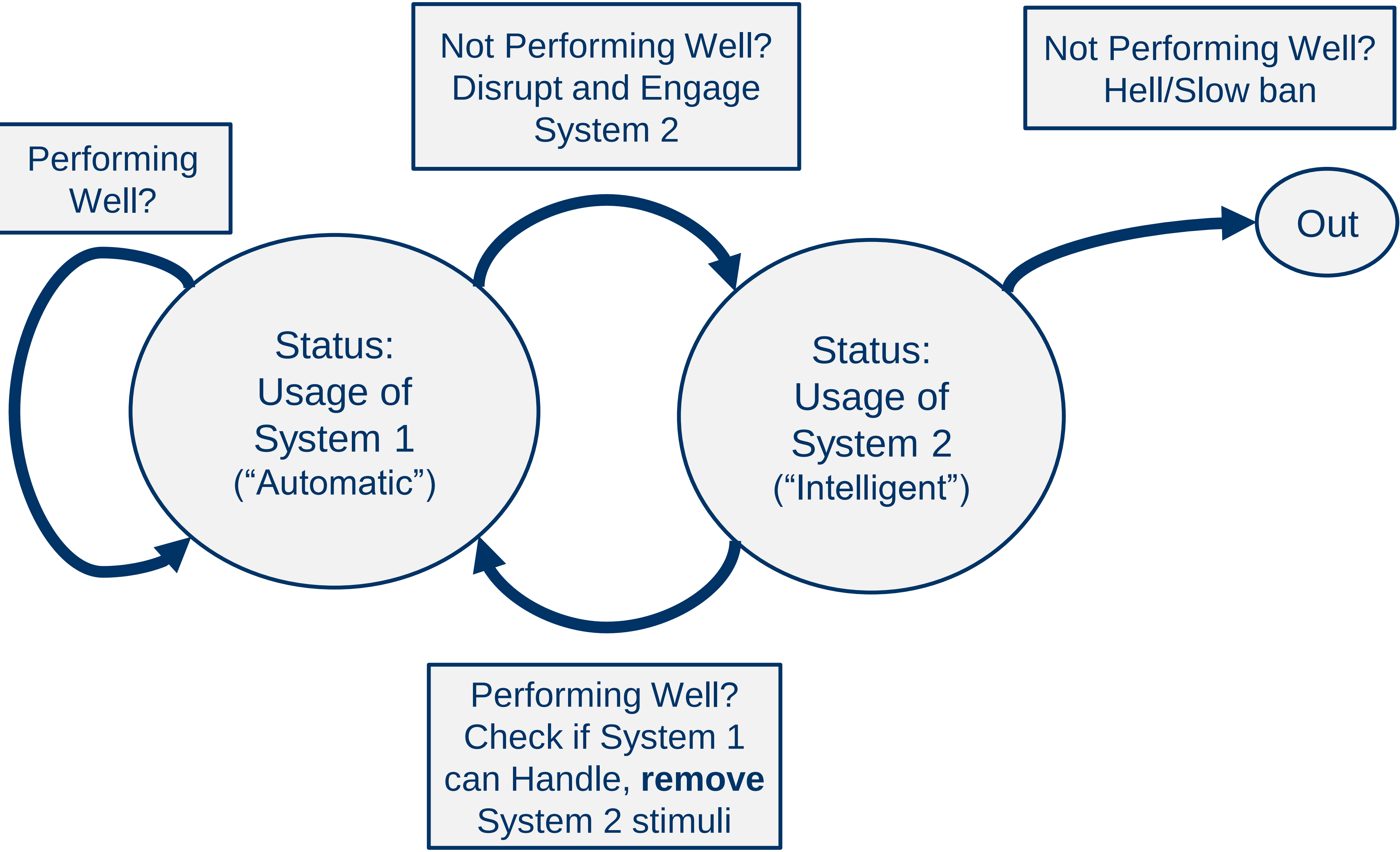


- Detect that one object is more distant than another.
- Orient to the source of a sudden sound.
- Complete the phrase "bread and…"
- Make a "disgust face" when shown a horrible picture.
- Detect hostility in a voice.
- Answer to 2 + 2 = ?
- Read words on large billboards.
- Drive a car on an empty road.
- Find a strong move in chess (if you are a chess master).
- Understand simple sentences.

System 2 Tasks



- Focus attention on the clowns in the circus.
- Look for a woman with white hair.
- Count the occurrences of the letter *a* in a page of text.
- Compare two washing machines for overall value.
- Check the validity of a complex logical argument.



Performing Well?

Not Performing Well?
Disrupt and Engage
System 2

Not Performing Well?
Hell/Slow ban

Status:
Usage of
System 1
(“Automatic”)

Status:
Usage of
System 2
(“Intelligent”)

Out

Performing Well?
Check if System 1
can Handle, **remove**
System 2 stimuli

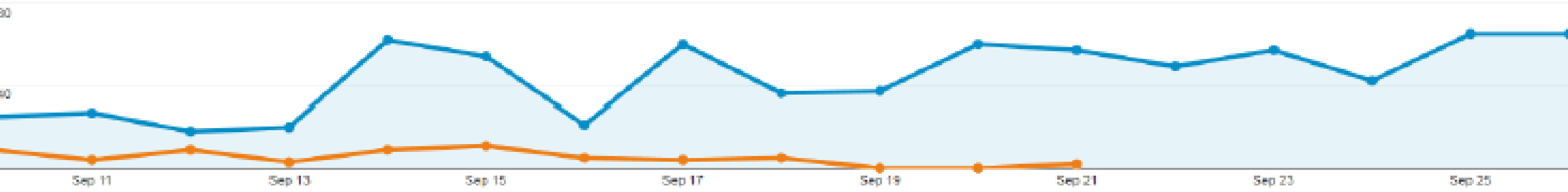
Thanks!

Q & A?

Effect of Ad Targeting

Perhaps it is just more users?

- **Control:** Ad campaign with no feedback, all keywords across quizzes
- **Treatment:** Ad campaign with feedback enabled



- **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)