

Crowdsourcing

February 14, 2019

Data Science CSCI 1951A

Brown University

Instructor: Ellie Pavlick

HTAs: Wennie Zhang, Maulik Dang, Gurnaaz Kaur

(Some slides stolen from Chris Callison-Burch and Kristy Milland. Thank you!)

Announcements

- Clickers
- Assignment 3 release pushed to next week
- Pre-proposal due after long weekend
- Final project requirements posted

Today

- Catch up from last time—practice with bash scripting
- Crowdsourcing Overview and HowTo

Bash Scripting

<https://cs.brown.edu/people/epavlick/articles.txt>

- | | |
|----------------------|------------------|
| 1. ID | 6. Victim Age |
| 2. City | 7. Shooter Age |
| 3. State | 8. Url |
| 4. Date (YYYY-MM-DD) | 9. Title |
| 5. Time | 10. Article Text |

cat, less, head, tail

- what does this data even look like?

```
# first 10 lines of file  
$ head articles.txt
```

```
# first line of file  
$ head -n 1 articles.txt
```

```
# random 10 lines from file  
$ cat articles.txt | shuf | head
```

WC

- how many articles are there

```
# how many bytes, words, and lines are  
there?
```

```
$ wc articles.txt
```

```
# how many lines are there?
```

```
$ wc -l articles.txt
```

pipe (), redirect (>)

```
$ head articles.txt | wc -l  
10
```

```
# write output to file called "tmp"  
$ head articles.txt > tmp  
$ wc -l tmp  
10 tmp
```

```
$ head articles.txt | wc -l > tmp  
$ cat tmp  
10
```

cut

```
$ cat articles.txt | cut -f 1 | head -n  
3
```

Antioch
Greeley
Bridgeport

```
$ cat articles.txt | cut -f 3 | cut -f 1  
-d '-' | head -n 3
```

2016
2015
2014

sort, uniq

```
# print the lowest 3 values (includes duplicates)
$ cat articles.txt | cut -f 4 | cut -f 1 -d '-'| sort | head -n 3
1929
1932
1932

# print lowest three values (remove duplicates but count how many
occurrences of each
$ cat articles.txt | cut -f 4 | cut -f 1 -d '-'| sort | uniq -c |
head -n 3
1 1929
2 1932
3 1942

# find the most frequent years
$ cat articles.txt | cut -f 4 | cut -f 1 -d '-'| sort | uniq -c |
sort -r | head
5091 2015
1821 2016
1784 NA
```

sort, uniq

How many duplicated entries are there (using url as the uniq id)?

```
# total number of urls (lines)
$ cat articles.txt | cut -f 8 | wc -l
9584
```

```
# number of unique urls
$ cat articles.txt | cut -f 8 | sort | uniq | wc -l
7990
```

```
# number of duplicated urls
$ cat articles.txt | cut -f 8 | sort | uniq -d | wc -l
981
```

regex (grep, sed, awk)

```
$ cat articles.txt | cut -f 2 | grep "NY" | head -n 5
```

```
NY  
HOMINY  
NYC  
NY  
NY
```

```
$ cat articles.txt | cut -f 2 | grep "^NY$" | head
```

```
NY  
NY  
NY  
NY
```

```
$ cat articles.txt | cut -f 2 | grep "^NY[.]*" | head
```

```
NY  
NYC  
NY  
NY  
NY
```

regex (grep, sed, awk)

```
$ cat articles.txt | cut -f 4 | sed "s/[0-9]/#/g" | head -n 3  
#####-##-##  
#####-##-##  
#####-##-##
```

```
$ cat articles.txt | cut -f 3 | sed "s/[A-Z] [A-Z] - //g" | grep -v Unclear |  
head -n 3  
Minnesota  
North Carolina  
Michigan
```

Being all fancy...

```
# replace all non-numeric characters with blanks
$ cat articles.txt | cut -f 6 | sed "s/[^0-9]//g" |
head

# plot a histogram of all ages
cat articles.txt | cut -f 6 | sed "s/[^0-9]//g" |
grep -v "^$" | python2 -c "import sys,
matplotlib.pyplot as plt; plt.hist([int(i) for i in
sys.stdin]); plt.show()"

# plot a histogram of all ages, removing outliers
cat articles.txt | cut -f 6 | sed "s/[^0-9]//g" |
grep -v "^$" | python2 -c "import sys,
matplotlib.pyplot as plt; plt.hist([min(int(i), 100)
for i in sys.stdin]); plt.show()"
```

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

How many unique values are there for “city” in our data?

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

How many unique values are there for “city” in our data?

- (a) \$ cat articles.txt | cut -f 2 | uniq | wc -l
- (b) \$ cat articles.txt | sort | uniq | cut -f 2 | wc -l
- (c) \$ cat articles.txt | cut -f 2 | sort | uniq | wc -l

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

How many unique values are there for "city" in our data?

- (a) \$ cat articles.txt | cut -f 2 | uniq | wc -l
- (b) \$ cat articles.txt | sort | uniq | cut -f 2 | wc -l
- (c) \$ cat articles.txt | cut -f 2 | sort | uniq | wc -l

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

**Find the 10 titles that appear with the largest
number of unique urls.**

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

**Find the 10 titles that appear with the largest
number of unique urls.**

- (a) `$ cat articles.txt | cut -f 9 | sort | uniq -c | sort -nr | head`
- (b) `$ cat articles.txt | cut -f 8,9 | sort | uniq | cut -f 2 | sort | uniq -c | sort -nr | head`
- (c) `$ cat articles.txt | sort | uniq -f 9 | sort -nr | head`

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4, Time=5, Victim Age=6, Shooter Age=7, Url=8, Title=9, Text=10

Find the 10 titles that appear with the largest number of unique urls.

- (a) \$ cat articles.txt | cut -f 9 | sort | uniq -c | sort -nr | head
- (b) \$ cat articles.txt | cut -f 8,9 | sort | uniq | cut -f 2 | sort | uniq -c | sort -nr | head**
- (c) \$ cat articles.txt | sort | uniq -f 9 | sort -nr | head

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

**How many different cities are there for the article
titled “Suspect arrested in Memphis cop killing”**

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

**How many different cities are there for the article
titled "Suspect arrested in Memphis cop killing"**

- (a) \$ cat articles.txt | cut -f 2 | grep "Suspect arrested in Memphis cop killing" | sort | uniq -c
- (b) \$ cat articles.txt | grep "Suspect arrested in Memphis cop killing" | cut -f 2 | sort | uniq -c
- (c) \$ cat articles.txt | sort | grep "Suspect arrested in Memphis cop killing" | cut -f 2 | uniq -c

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

**How many different cities are there for the article
titled "Suspect arrested in Memphis cop killing"**

- (a) \$ cat articles.txt | cut -f 2 | grep "Suspect arrested in Memphis cop killing" | sort | uniq -c
- (b)** \$ cat articles.txt | grep "Suspect arrested in Memphis cop killing" | cut -f 2 | sort | uniq -c
- (c) \$ cat articles.txt | sort | grep "Suspect arrested in Memphis cop killing" | cut -f 2 | uniq -c

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

**Print out all the victim ages that contain no numeric
characters.**

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4,
Time=5, Victim Age=6, Shooter Age=7, Url=8,
Title=9, Text=10

Print out all the victim ages that contain no numeric characters.

- (a) \$ cat articles.txt | cut -f 6 | grep -e "^[^0-9]*\$"
- (b) \$ cat articles.txt | cut -f 6 | grep -e "^[0-9]*\$"
- (c) \$ cat articles.txt | cut -f 6 | grep -e "^[^0-9]*"

Clicker Question!

Hint: Columns are ID=1, City=2, State=3, Date=4, Time=5, Victim Age=6, Shooter Age=7, Url=8, Title=9, Text=10

Print out all the victim ages that contain no numeric characters.

- (a) \$ cat articles.txt | cut -f 6 | grep -e "^[^0-9]*\$"
- (b) \$ cat articles.txt | cut -f 6 | grep -e "^[0-9]*\$"
- (c) \$ cat articles.txt | cut -f 6 | grep -e "^[^0-9]*"

Focus of today:



[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)





reddit



twitter



U B E R

coursera

amazon mechanical turk™
Artificial Artificial Intelligence

Quora



stack overflow

KICKSTARTER

craigslist



Motivation

Why do people contribute?

What do workers gain from contributing?



U B E R



Quora



KICKSTARTER

craigslist



Motivation

Why do people contribute?

What do workers gain from contributing?



Quality Control

How do we make sure the contributions are good?

How do we identify and incentivize good work?

Quora

KICKSTARTER

amazon stackoverflow

craigslist

Motivation

Why do people contribute?

What do workers gain from contributing?

Aggregation

How do we combine many small or distributed contributions into one final answer/result/product?



Quality Control

How do we make sure the contributions are good?

How do we identify and incentivize good work?



KICKSTARTER

craigslist



Motivation

Why do people contribute?

What do workers gain from contributing?

Aggregation

How do we combine many small or distributed contributions into one answer/result/product?

KICKSTAR



Quality Control

How do we make sure the contributions are good?

Skill

Do workers need specialized skills?

How do we find or train workers to match the skill sets we need?

Motivation

Why do people contribute?

What do workers gain from contributing?

Aggregation

How do we combine many contributions?

Decomposition

How is the task decomposed into subtasks?

How many subtasks are required to get from input to output?

Quality Control

How do we make sure the contributions are good?

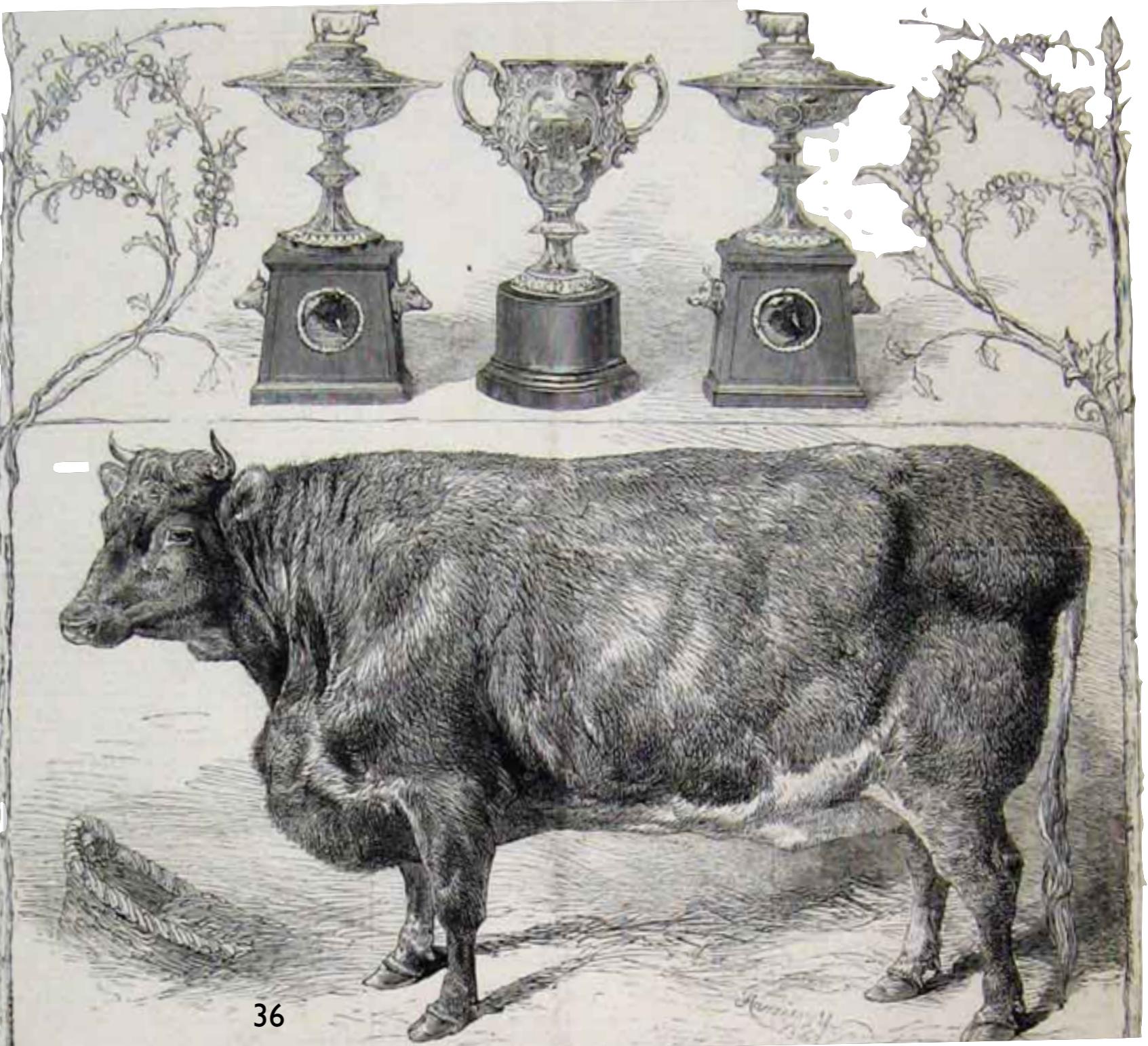
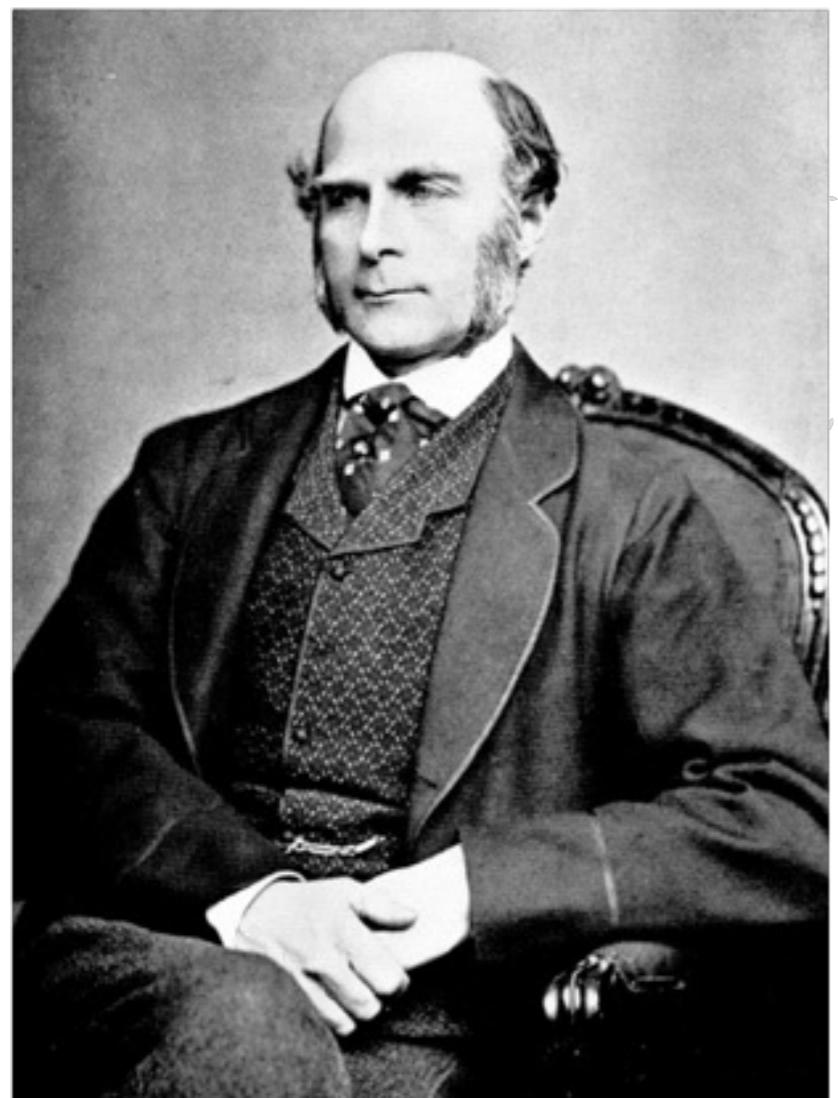
Skill

workers need specialized skills?

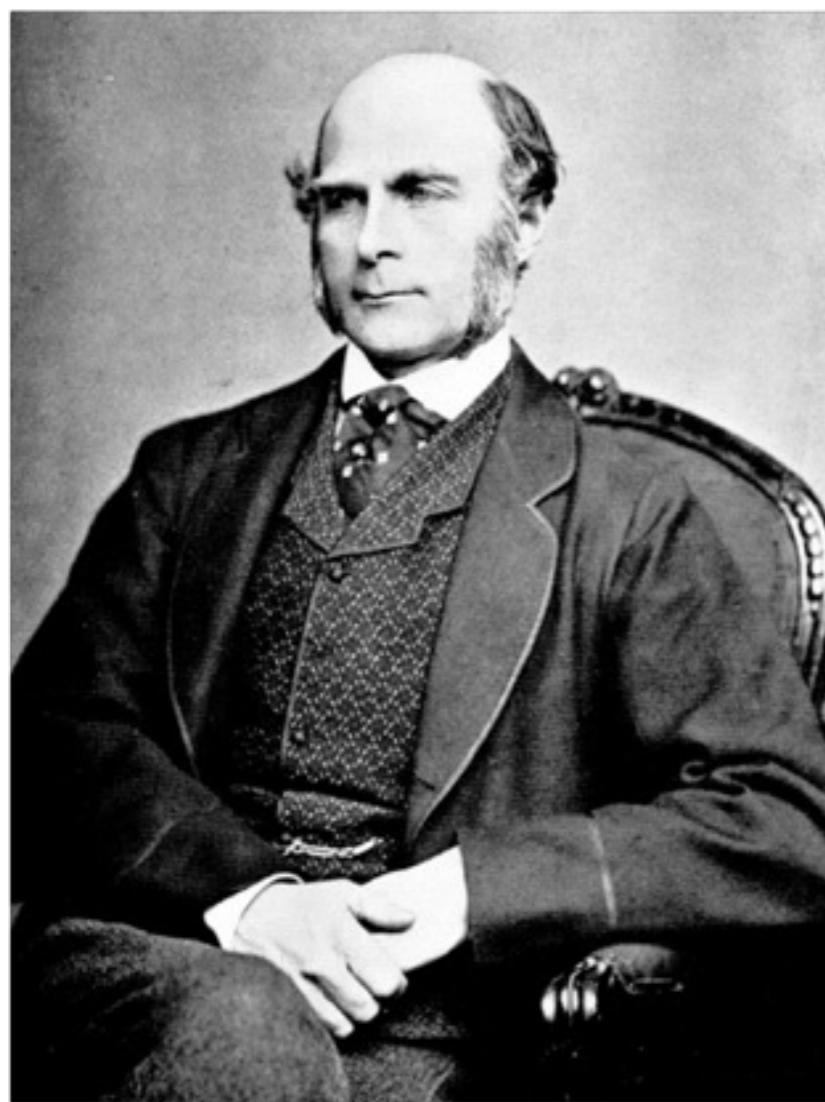
so we find or train workers to match the skills we need?

Wisdom of the Crowd

Wisdom of the Crowd



Wisdom of the Crowd



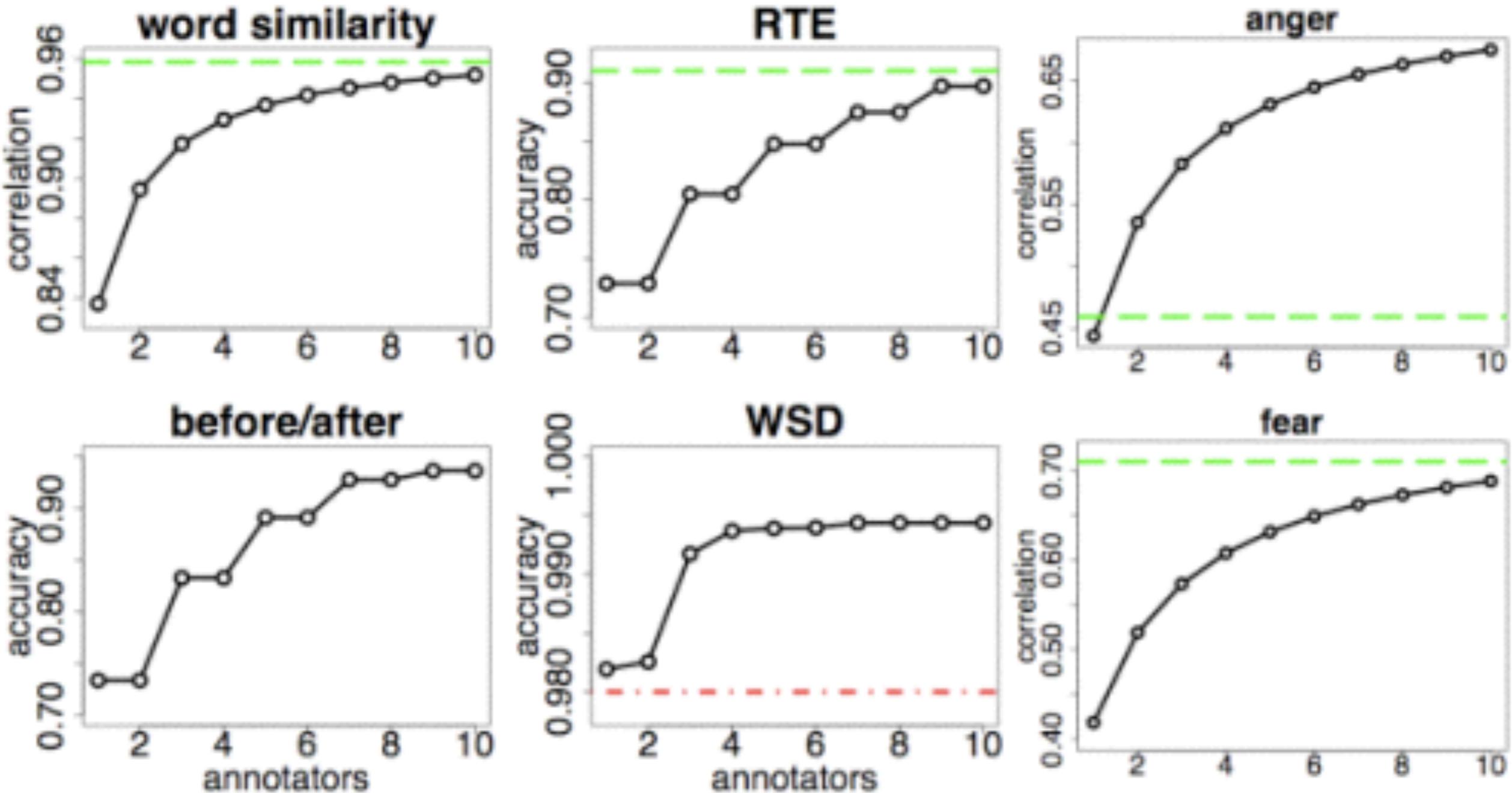
Cutting a Round Cake on Scientific Principles.

CHRISTMAS suggests cakes, and these the wish on my part to describe a method of cutting them that I have recently devised to my own amusement and satisfaction. The problem to be solved was, "given a round tea-cake of some 5 inches across, and two persons of moderate appetite to eat it, in what way should it be cut so as to leave a minimum of exposed surface to become dry?" The ordinary method of cutting out a wedge is very faulty in this respect. The results to be aimed at are so to cut the cake that the remaining portions shall fit together. Consequently the chords (or the arcs) of the circumferences

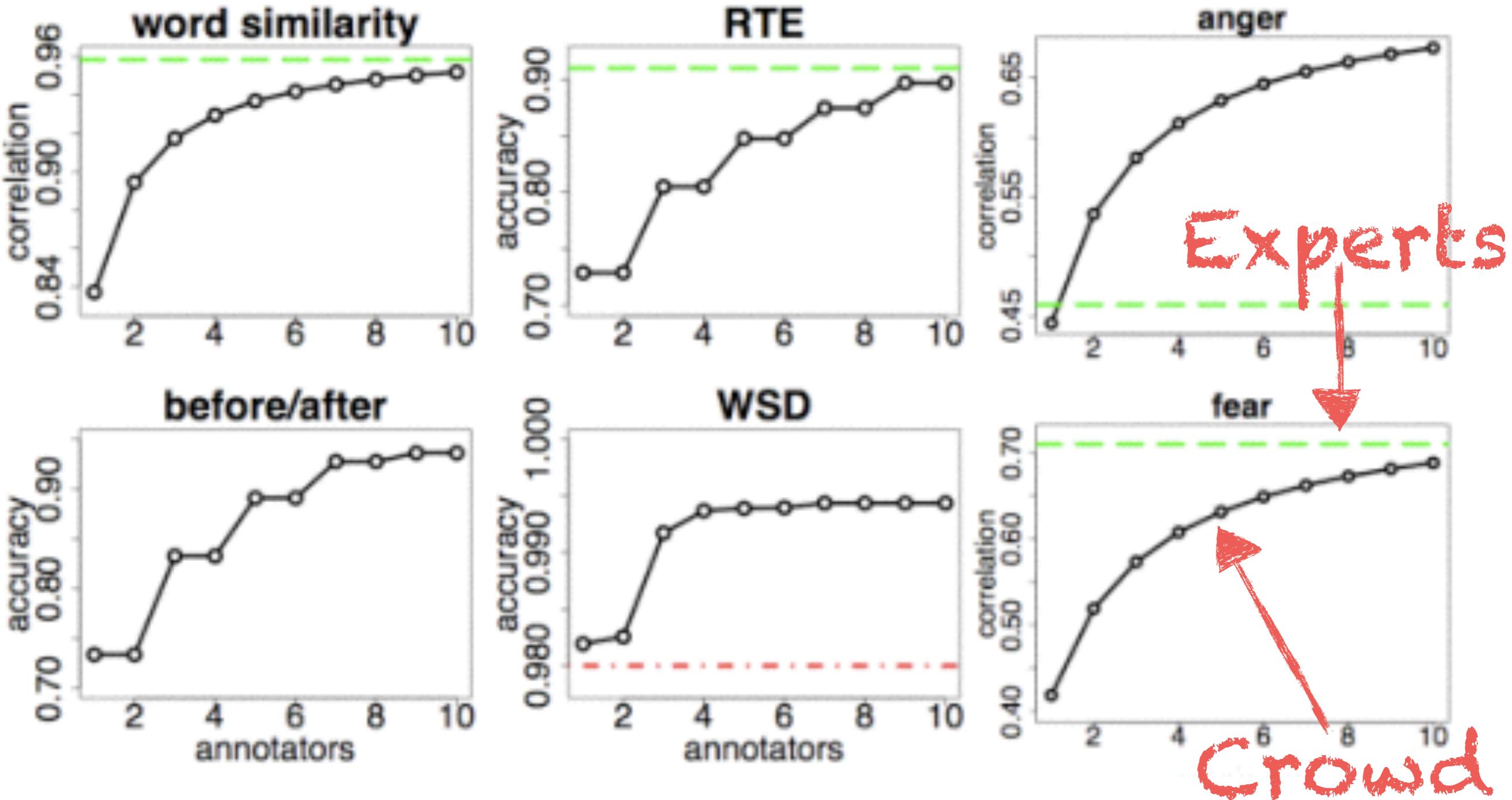
The diagram consists of three circular outlines. The first circle shows a vertical dashed line and a horizontal dashed line intersecting at the center, dividing the circle into four quadrants. The second circle shows the same dashed lines, but the top-right quadrant is shaded gray, indicating it has been removed. The third circle shows the final state where the top-right quadrant has been removed, and the remaining three quadrants are shown fitting together.

Broken straight lines show intended cuts. Ordinary straight lines show the cuts that have been made. The segments are kept in apposition by a common elastic band that encloses the whole. In the above figures about one-third of the area of the original disc is removed by each of the two successive operations.

Wisdom of the Crowd

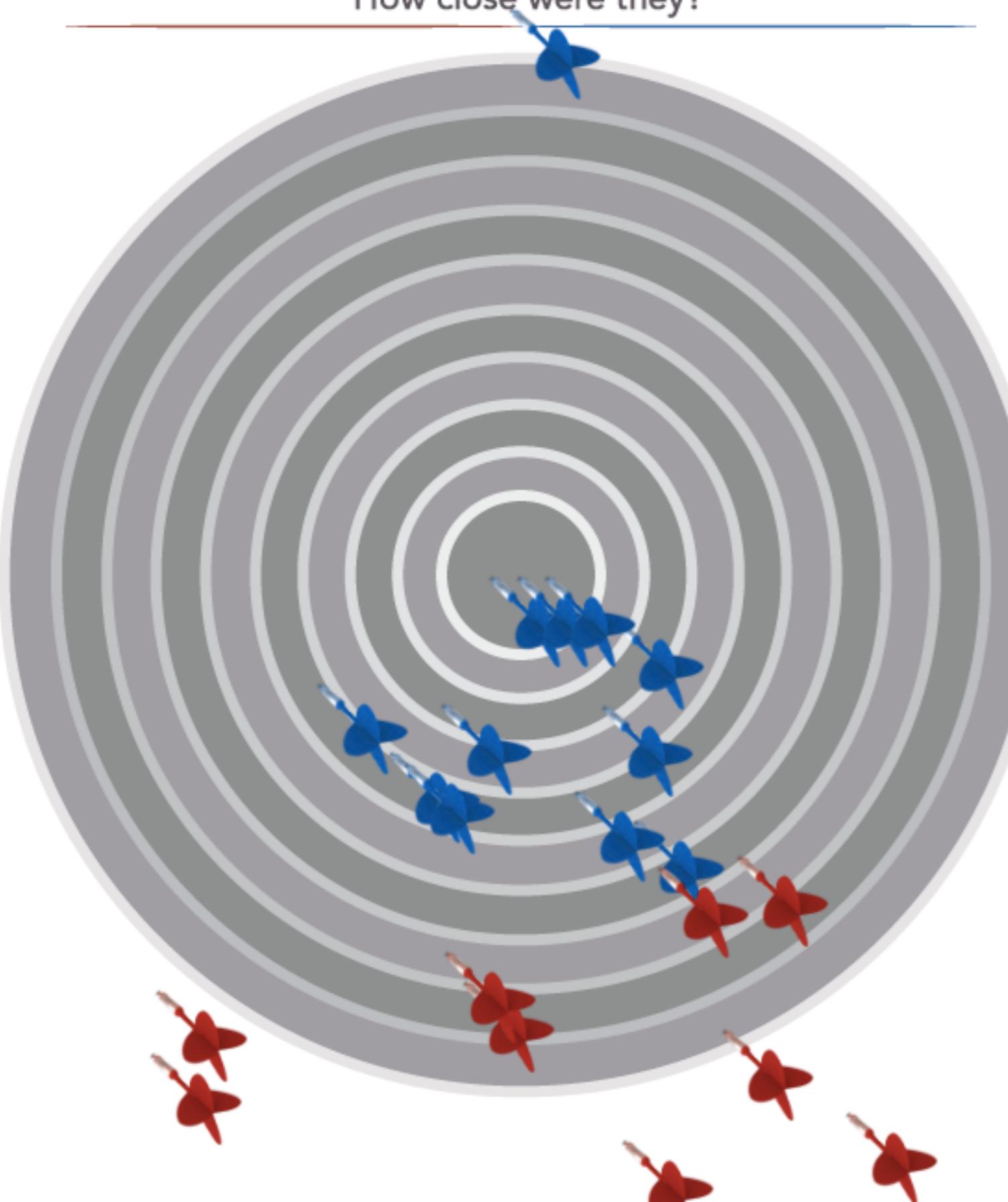


Wisdom of the Crowd



PUNDARTS

How close were they?



Nate Silver

The New York Times

Obama: 332

Romney: 206

Difference: 0



Donna Brazile

*Vice Chairwoman of
the Democratic
National Committee*

Obama: 313

Romney: 225

Difference: -19



Ann Coulter

Fox News

Obama: 265

Romney: 273

Difference: -67



Karl Rove

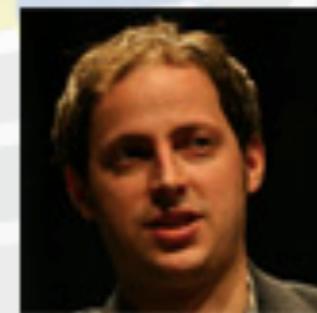
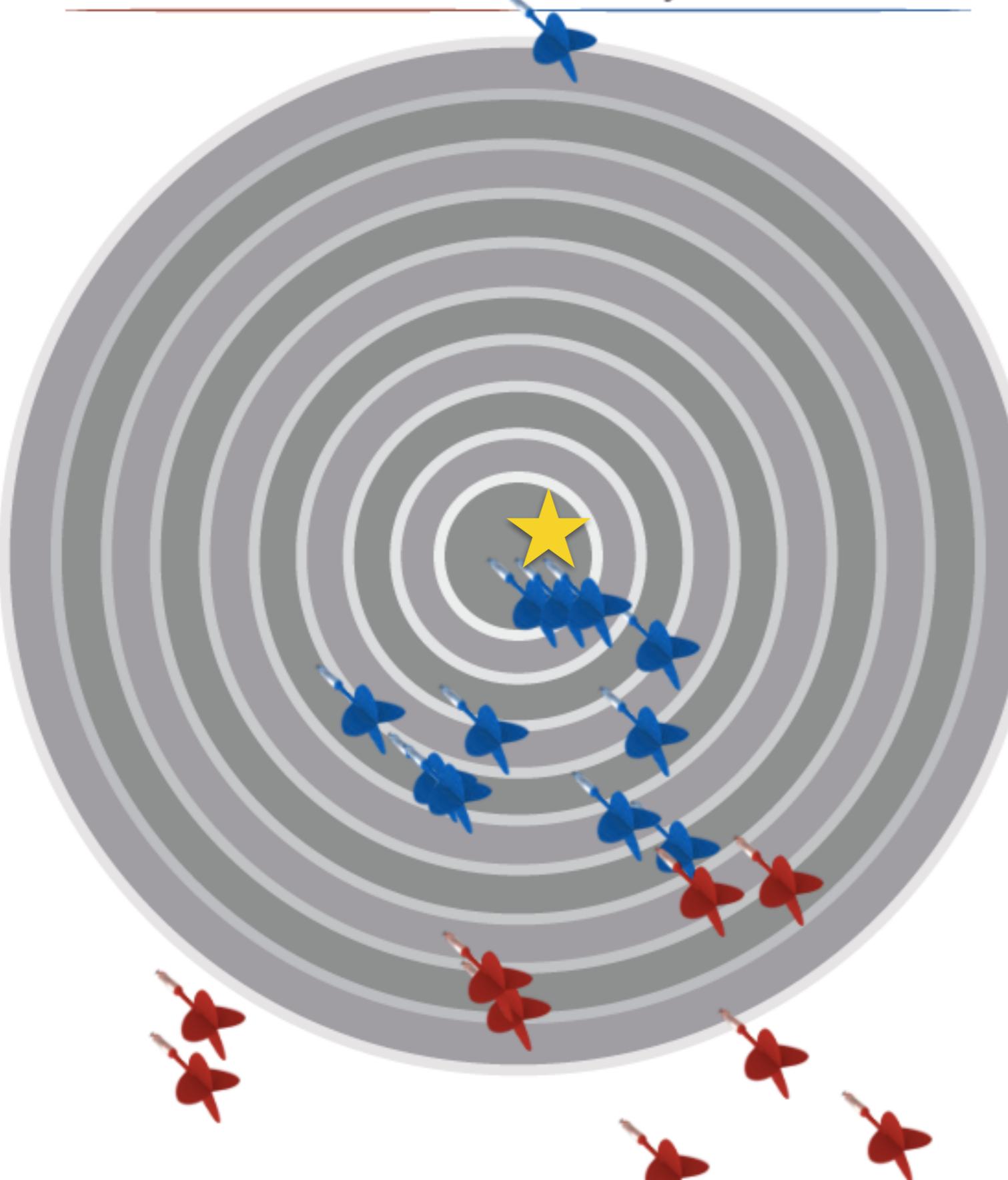
Obama: 259

Romney: 279

Difference: -73

PUNDARTS

How close were they?



Nate Silver
The New York Times
Obama: 332
Romney: 206
Difference: 0

amazon mechanical turk 
beta Artificial Artificial Intelligence

Obama: 335, Romney: 189
Error: 3, Unallocated: 14



Donna Brazile
*Vice Chairwoman of
the Democratic
National Committee*
Obama: 313
Romney: 225
Difference: -19



Ann Coulter
Fox News
Obama: 265
Romney: 273
Difference: -67



Karl Rove
Obama: 259
Romney: 279
Difference: -73

Motivation

Motivation

The image is a composite of three screenshots from mobile applications illustrating different forms of platform-based motivation:

- Airbnb:** A listing page for a room with a balcony. The main text reads "START HOSTING AND EARN A \$250 CASH BONUS". The top left corner shows the Airbnb logo.
- Uber:** A surge pricing alert. It says "Demand is off the charts! Fares have increased to get more Ubers on the road." Below it, a large circular badge indicates a fare multiplier of "2.2x THE NORMAL FARE". Smaller text below the badge specifies "\$7.60 MINIMUM FARE", "\$0.35 / MIN", and "\$2.09 / MILE".
- Upwork:** A rewards summary titled "Total Earnings (What's this?)". It lists "Rewards You Have Earned" and their "Value":

Rewards You Have Earned	Value
Approved HITs	\$1.16
Bonuses	\$0.81
Total Earnings Show earnings details	\$1.97

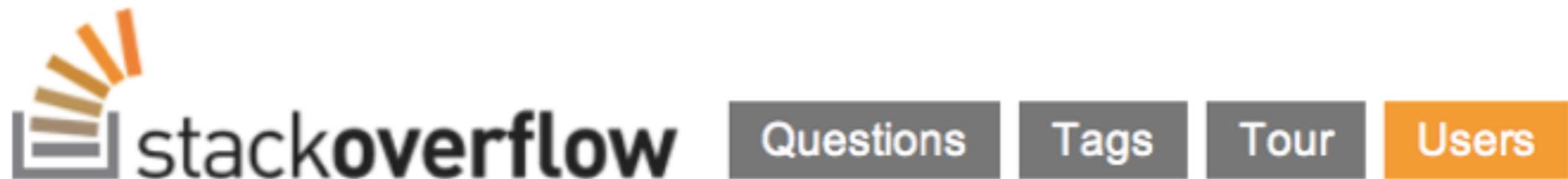
Pay

Motivation



Altruism

Motivation



Users

reputation

Type to find users:



Gordon Linoff

New York, United States

1,260

sql, mysql, sql-server



Sotirios Delimanolis

1,020

java, spring, spring-mvc



Arun P Johny

Bangalore, India

960

jquery, javascript, html



Jon Skeet

Reading, United Kingdom

915

c#, java, .net



Hans Passant

Madison, WI

905

c#, .net, winforms



Martijn Pieters

Cambridge, United Kingdom

880

python, python-2.7, list

45 Reputation

Motivation



Fun

Motivation

score
100

time
2:21

taboo words

peace
lay

What do you see?

duolingo Home Discussion Immersion

Create a profile

Spanish progress

0 XP 0 Words

Choose your path!

Beginners start at Basics 1.
Advanced learners take a short test.

Basics 1 OR Placement test

Phrases 0/3 Basics 2 0/4

Food 0/5 Animals 0/4 Plurals 0/1

Follow Duolingo

Twitter 90k Facebook 300k Google+ 70k

Fun

Self-Improvement

Motivation

The Norwich line steamboat train, from New-London for Boston, this morning ran off the track seven miles north of New-London.

Implicit Work

morning

morning overHOKs

Type the two words:



Motivation

The Norwich line steamboat train, from New-London for Boston, this morning ran off the track seven miles north of New-London.

~~Implicit
Work~~

morning
No
Choice :)

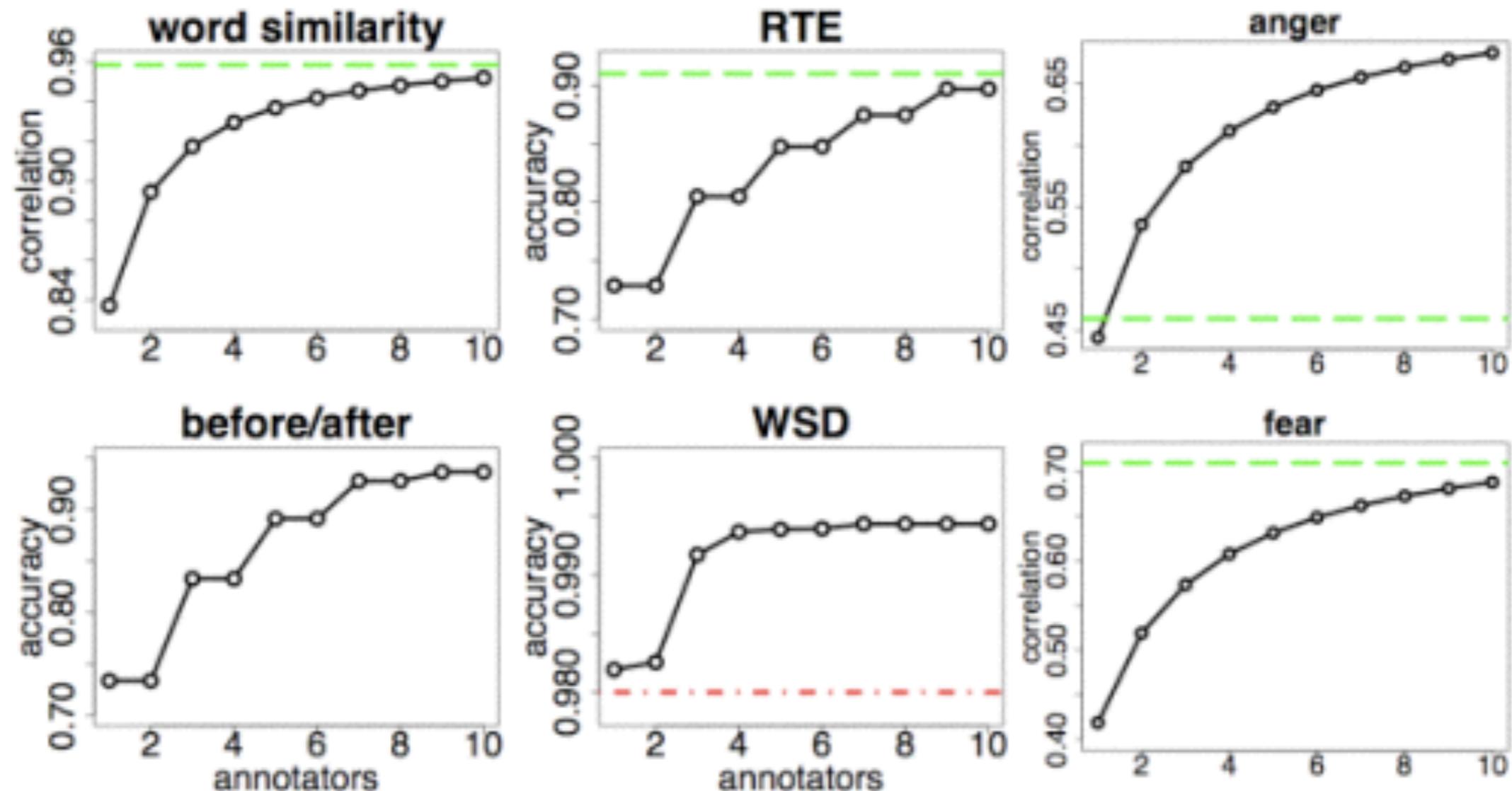
morning overDOOkS

Type the two words:



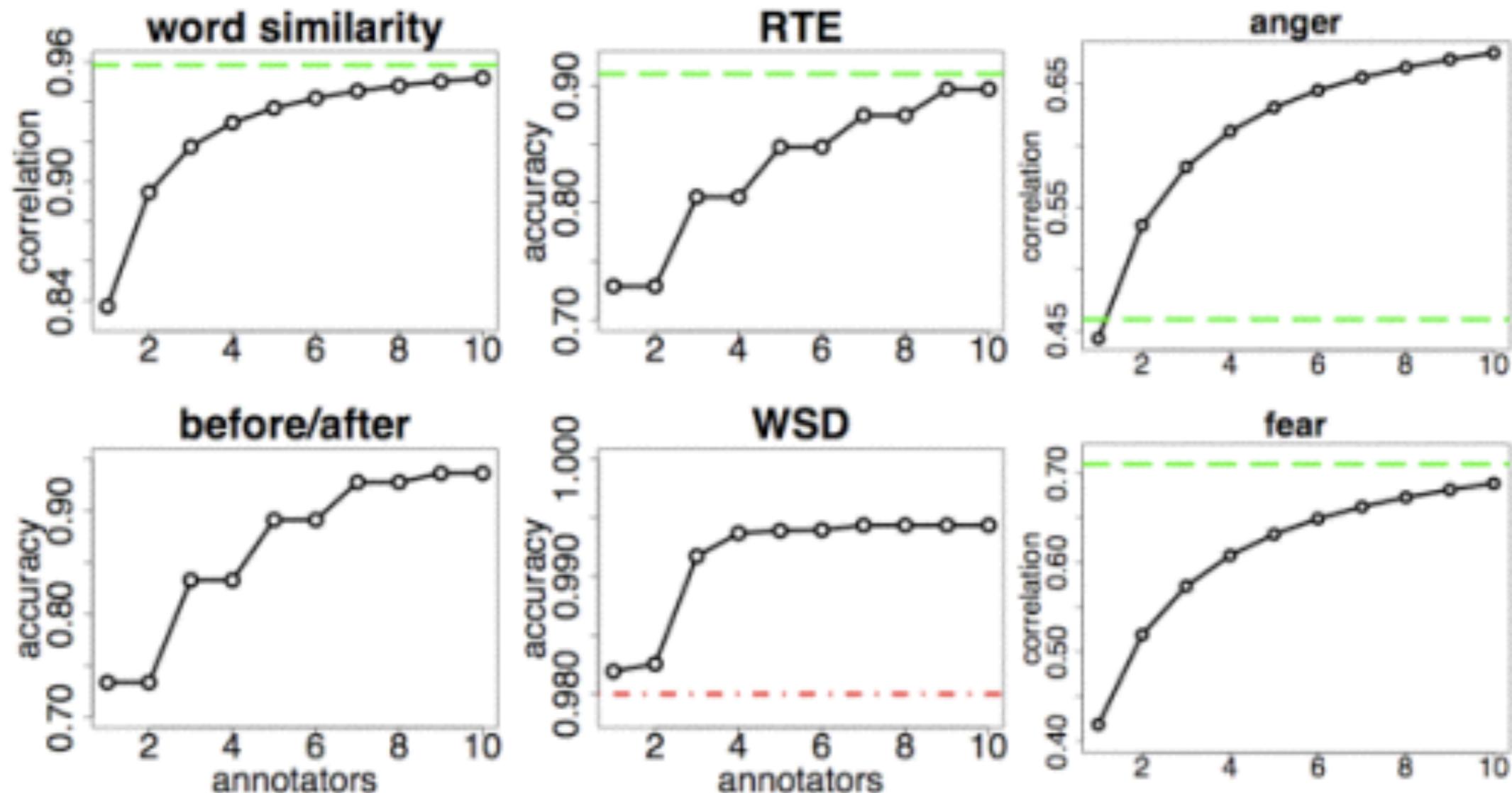
Quality Control

Quality Control



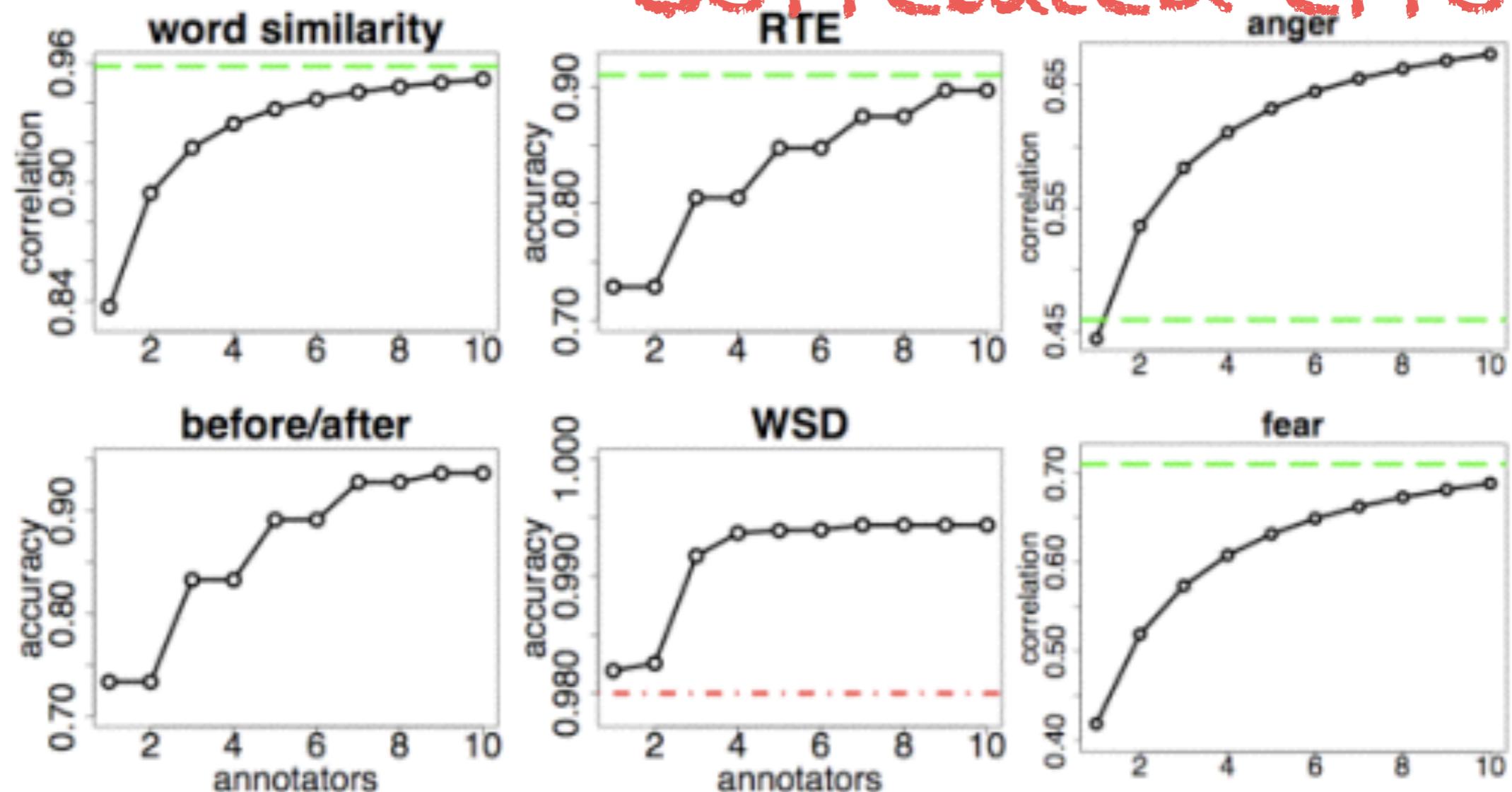
Agreement/Redundancy

Malicious workers? Quality Control



Agreement/Redundancy

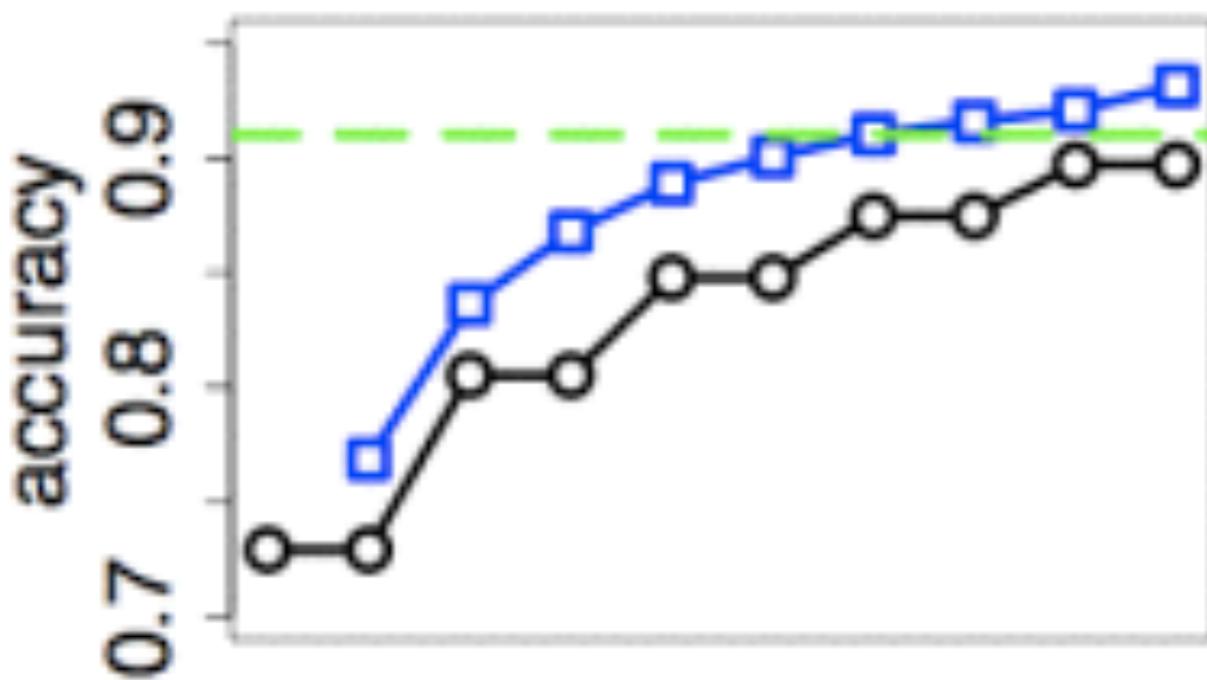
Malicious workers? Quality Control Correlated errors?



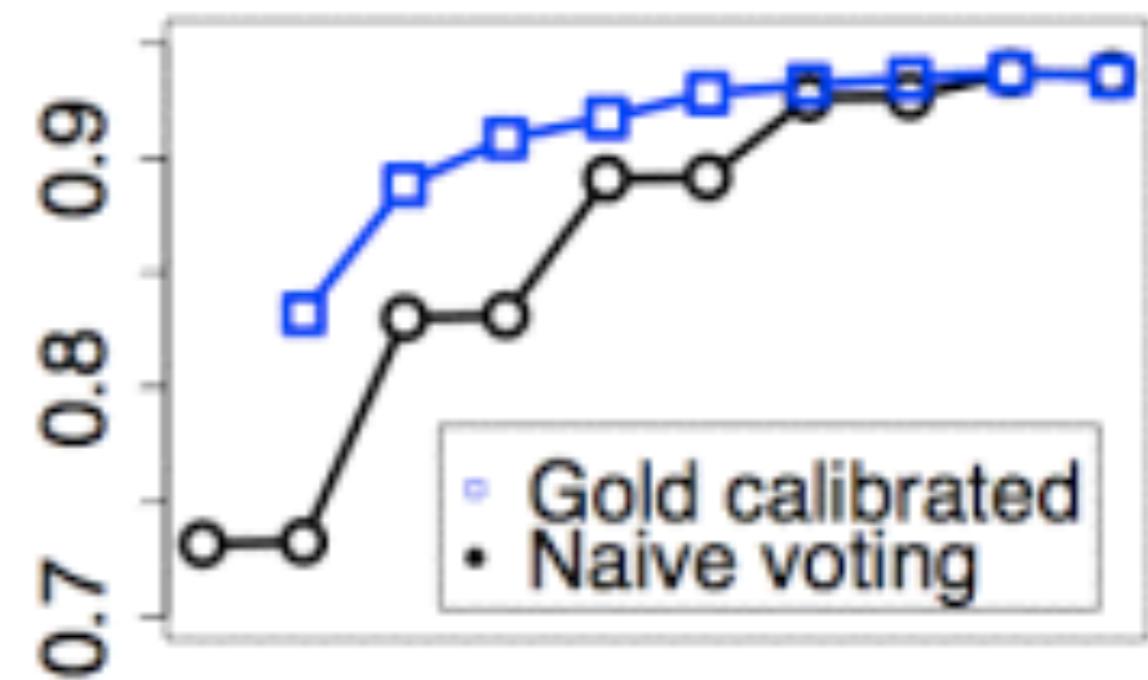
Agreement/Redundancy

Quality Control

RTE



before/after



Confidence estimates on workers improve accuracy.

HIT Approval Rate (%) greater than 0

(+) Add another criterion (up to 3 more)

Project contains adult content [\(See details\)](#)

This project may contain potentially explicit or offensive content, for example, nudity.

HIT Visibility [\(What is HIT visibility?\)](#)

Public - All Workers can see and preview my HITs

Private - All Workers can see my HITs, but only Workers that meet all Qualification requirements

Hidden - Only Workers that meet my HIT Qualification requirements can see and preview my HIT

Remove

I am a bit puzzled by the following code:

1026 `d = {'x': 1, 'y': 2, 'z': 3}
for key in d:
 print key, 'corresponds to', d[key]`

228 What I don't understand is the `key` portion. How does Python recognize that it needs only to read the key from the dictionary? Is `key` a special word in Python? Or is it simply a variable?

[python](#) [dictionary](#)

share improve this question edited May 19 '15 at 20:08

 **Vito Gentile** 5,047 • 3 • 27 • 66

asked Jul 20 '10 at 22:27  **TopChef** 5,746 • 9 • 18 • 21

[add a comment](#)

7 Answers

active oldest votes

key is just a variable name.

1922 `for key in d:` will simply loop over the keys in the dictionary, rather than the keys and values. To loop over both key and value you can use the following:

[Edit](#)

 For Python 2.x-

```
for key, value in d.iteritems():
```

\$23.62
+ \$4.99 shipping + \$0.00 estimated tax

PaBo Ent.
[Just Launched \(Seller Profile\)](#)

\$28.62
& FREE Shipping + \$0.00 estimated tax

Brook Mays
 93%
[positive](#) over the past 12 months. (60,420 total ratings)

\$26.15
+ \$2.99 shipping + \$0.00 estimated tax

WeinerMusic
 99%
[positive](#) over the past 12 months. (2,959 total ratings)

Reputation Systems

Quality Control

Advanced

For the best quality, **Master Workers** are currently selected to complete your work. ([What is a Master Worker?](#))

[Worker requirements «](#)

Worker requirements:

Require that Workers be Masters to do your HITs



Only Workers who qualify to do my HITs can preview my HITs.

Yes No

Pre-Vetted Workers

Quality Control

Advanced

Worker requirements:

[Customize Worker Requirements...](#)

Worker requirements «

 https://requester.mturk.com

You have chosen NOT to use Master Workers. We strongly encourage you to use Masters as these Workers have demonstrated accuracy in performing a wide range of HITs. Are you sure you want to continue?

[Cancel](#) [OK](#)

Specify ALL the qualifications Workers must meet to work on your HITs:

Location is UNITED STATES [remove](#)

HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 80 [remove](#)

Number of HITs Approved greater than or equal to 50 [remove](#)

[\(+ Add another criterion](#) (up to 5)

Only Workers who qualify to do my HITs can preview my HITs.

Yes No

Pre-Vetted Workers

Quality Control

Masters are elite groups of Workers who have demonstrated accuracy on specific types of HITs. Workers achieve a Masters distinction by consistently completing HITs with a high degree of accuracy across a variety of Requesters. Masters must continue to pass our statistical monitoring to remain Mechanical Turk Masters. Because Masters have demonstrated accuracy, they can command a higher reward for their HITs. You should expect to pay Masters a higher reward.

Quality Control

- Amazon now nominates a subset (21k workers, estimated at 10% of all Turkers) of senior / good workers as “Masters”
- Amazon charges 25% commission for Masters versus their normal 20% rate
- They have now implemented this as the default qualification for new Requesters
- Why?

Pre-Vetted Workers

Quality Control

Pros

- Easier for new requesters who do not know to implement quality control.
- Masters will not touch badly designed or low-paying tasks
 - Amazon charges 25% commission for Masters versus their normal 20% rate
 - They have now implemented this as the default qualification for new Requesters
 - Why?

set (21k
ll Turkers) of
ers”

Pre-Vetted Workers

Quality Control

Pros

- Easier for new requesters who do not know to implement quality control.
- Masters will not touch badly designed or low-paying tasks
 - Amazon charges 25% commission for

Master

Cons

- They have a default
- Why?
 - Fewer Masters workers -> significant lag in tasks being picked up
 - More expensive
 - Not clear in what tasks the Masters are tested and how a new worker can become a master.

Quality Control

Manage Qualification Types

Below is a list of your Qualification Types and the corresponding number of Workers.

[Create New Qualification Type](#) 

Qualification Types					
	Name ▼	ID	Workers who have this Qualification	Creation Date	Description
✖	Trusted researcher	2GJ7Q67051QKTXPQMYHC7XMGI4AEYR	7	Thu Jun 20 17:35:18 UTC 2013	This qualification is granted to grad students and researchers. We use it to limit the participation in our pilot runs of experiments.
✖	Temporal Master	2YG46UXFCD45EMCI7IJNZKIN3WJVTB	0	Fri Mar 23 09:46:47 UTC 2012	This qualification is granted to Turkers who have demonstrated a high level of competency in temporal relations.
✖	Presidential Surveyor	279YQ4J3NAB6SPK4ZTPYOT1TJI6QDJ	0	Sat Nov 03 18:33:24 UTC 2012	This qualification was given to people who responded to the political survey before the election. We'll allow them to answer some follow up questions after the election.
✖	Monolingual Worker	2K5F806UFBNF95EJ5KU9BSW6NKHF1D	6	Tue Jul 02 16:11:47 UTC 2013	Granted to Workers who have demonstrated good skills at doing the Monolingual Word Alignment tasks.

Qualification Tests

Quality Control

Manage Qualification Tests

Pros

- Uniform interface for workers
- Fair: no surprise rejections after work has been done
- Cost-effective: you don't have to pay for bad work

Description

to grad students and researchers. We ran in our pilot runs of experiments.

☒	Temporal Master	2YG46UXFCD45EMCI7IJNZKIN3WJVTB	0	Fri Mar 23 09:46:47 UTC 2012	This qualification is granted to Turkers who have demonstrated a high level of competency in temporal relations.
☒	Presidential Survey	279YQ4J3NAB6SPK4ZTPYOT1TJI6QDJ	0	Sat Nov 03 18:33:24 UTC 2012	This qualification was given to people who responded to the political survey before the election. We'll allow them to answer some follow up questions after the election.
☒	Monolingual Worker	2K5F806UFBNF95EJ5KU9BSW6NKHF1D	6	Tue Jul 02 16:11:47 UTC 2013	Granted to Workers who have demonstrated good skills at doing the Monolingual Word Alignment tasks.

Qualification Tests

Quality Control

Manage Qualification Tasks

Pros

- Uniform interface for workers
- Fair: no surprise rejections after work has been done
- Cost-effective: you don't have to pay for bad work

Description

to grad students and researchers. We n in our pilot runs of experiments.

<input checked="" type="checkbox"/> Temporal Master	2YG46UXFCD45EMCI7IJNZKIN3WJVTB	0	Fri Mar 23 09:46:47	This qualification is granted to Turkers who have demonstrated a high level of competency in temporal relations.
<input checked="" type="checkbox"/> Presidential Su...	279YQ4J3I			
<input checked="" type="checkbox"/> Monolingual Wor...	2K5F806U			

Cons

- Requires workers to do unpaid work- often deters workers from trying your task
- Turkers know when they are being evaluated, so their performance on the test might not reflect performance on the task

Quality Control

13. the parish church of st nidans is located near the a4080 highway , a little to the east of brynsiencyn .

- a sip
- a break
- a moment
- a bit
- a few
- a second
- a minute
- a while
- a touch
- a iittle
- a little bit
- little bit
- None of these paraphrases are good

14. the order primates contains humans and their closest relatives : lemurs , lorisooids , tarsiers , monkeys , and apes .

- gorillas
- epas
- monkeys
- chimps
- None of these paraphrases are good

Embedded Gold Standard

Quality Control

13. the parish church of st nidans is located near the a4080 highway , a little to the east of brynsiencyn .

- a sip
- a break
- a moment
- a bit
- a few
- a second
- a minute
- a while
- a touch
- a iittle
- a little bit
- little bit
- None of these paraphrases are good

14. the primate contains humans and their closest relatives : lemurs , lorisoids , tarsiers , monkeys , and apes .

- gorillas
- apes
- monkeys
- chimps
- None of these paraphrases are good

Embedded Gold Standard

Quality Control

13. the parish church of st nidans is located near the a4080 highway , a little to the east of brynsiencyn .

Pros

- Continuously evaluating work
- Quality estimates are a good reflection of quality on the actual work

a little
 a little bit
 little bit
 None of these paraphrases are good

14. the primate contains humans and their closest relatives : lemurs , lorisoids , tarsiers , monkeys , and apes .

gorillas
 apes
 monkeys
 chimps
 None of these paraphrases are good

~~Embedded Gold Standard~~

Quality Control

13. the parish church of st nidans is located near the a4080 highway , a little to the east of brynsiencyn .

Pros

- Continuously evaluating work
- Quality estimates are a good reflection of quality on the actual work

- a little
- a little bit
- little bit
- None of these paraphrases

14. the ~~over primates c~~

- gorillas
- epas
- monkeys
- chimps
- None of these paraphrases

Em

Cons

- Adds cost (paying to annotate examples that you already have labels for)
- Time-consuming to design/collect good test questions

Quality Control

People

HOME NEWS PHOTOS STYLE GAMES CELEBS PEOPLE TV ARCHIVE Se

Was arrested actress Heather Locklear because of the driving under the effect of an unknown medicine



Heather Locklear
Photo by: Santa Barbara County Sheriff's Department

SPONSORED LINKS

The actress Heather Locklear that is known to the Amanda through the role from the series "Melrose Place" was arrested at this weekend in Santa Barbara (Californium) because of the driving under the effect of an unknown medicine. A female witness observed she attempted in quite strange way how to go from their parking space in Montecito, speaker of the traffic police of californium told the warehouse 'People'. The female witness told in detail, that Locklear 'pressed after 16:30 clock accelerator and a lot of noise did when she attempted to move their car towards behind or forward from the parking space, and when it went backwards, she pulled itself together unites Male at their sunglasses'. A little later the female witness that did probably

- Why was Heather Locklear arrested?
Driving while medicated
- Why did the bystander call emergency services?
There was a lot of noise
- Where did the witness see her acting abnormally?
In a parking lot

Second-Pass HIT

Quality Control

People

HOME NEWS PHOTOS STYLE GAMES CELEBS PEOPLE TV ARCHIVE Se

Was arrested actress Heather Locklear because of the driving under the effect of an unknown medicine



The actress Heather Locklear that is known to the Amanda through the role from the series "Melrose Place" was arrested at this weekend in Santa Barbara (Californium) because of the driving under the effect of an unknown medicine. A female witness observed she attempted in quite strange way how to go from their parking space in Montecito, speaker of the traffic police of californium told the warehouse 'People'. The female witness told in detail, that Locklear 'pressed after 16:30 clock accelerator and a lot of noise did when she attempted to move their car towards behind or forward from the parking space, and when it went backwards, she pulled itself together unites Male at their sunglasses'. A little later the female witness that did probably

Heather Locklear
Photo by: Santa Barbara County Sheriff's Department

SPONSORED LINKS

- Why was Heather Locklear arrested?
Driving while medicated
- Why did the bystander call emergency services?
There was a lot of noise
- Where did the witness see her acting abnormally?
In a parking lot

MLB WORLD SERIES SURVEY (< 1 min survey. Eligible for \$5 bonus, US c)

Requester: [Danielle Limberg](#)

Incentive Pay

Second-Pass HIT

Quality Control

People

HOME NEWS PHOTOS STYLE GAMES CELEBS PEOPLE TV ARCHIVE Se

Was arrested actress Heather Locklear because of the driving under the effect of an unknown medicine



The actress Heather Locklear that is known to the Amanda through the role from the series "Melrose Place" was arrested at this weekend in Santa Barbara (Californium) because of the driving under the effect of an unknown medicine. A female witness observed she attempted in quite strange way how to go from their parking space in Montecito, speaker of the traffic police of californium told the warehouse 'People'. The female witness told in detail, that Locklear 'pressed after 16:30 clock accelerator and a lot of noise did when she attempted to move their car towards behind or forward from the parking space, and when it went backwards, she pulled itself together unites Male at their sunglasses'. A little later the female witness that did probably

Heather Locklear
Photo by: Santa Barbara County Sheriff's Department

SPONSORED LINKS

- Why was Heather Locklear arrested?
Driving while medicated
- Why did the bystander call emergency services?
There was a lot of noise
- Where did the witness see her acting abnormally?
In a parking lot

MLB WORLD SERIES SURVEY (< 1 min survey. Eligible for \$5 bonus, US c)

Requester: [Danielle Limberg](#)

Incentive Pay

$$L(\boldsymbol{\theta}; \mathbf{X}) = p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

Second-Pass HIT

Statistical Models

Platforms

Platforms



Platforms

Workers
sign up
directly



Platforms

Workers
sign up
directly

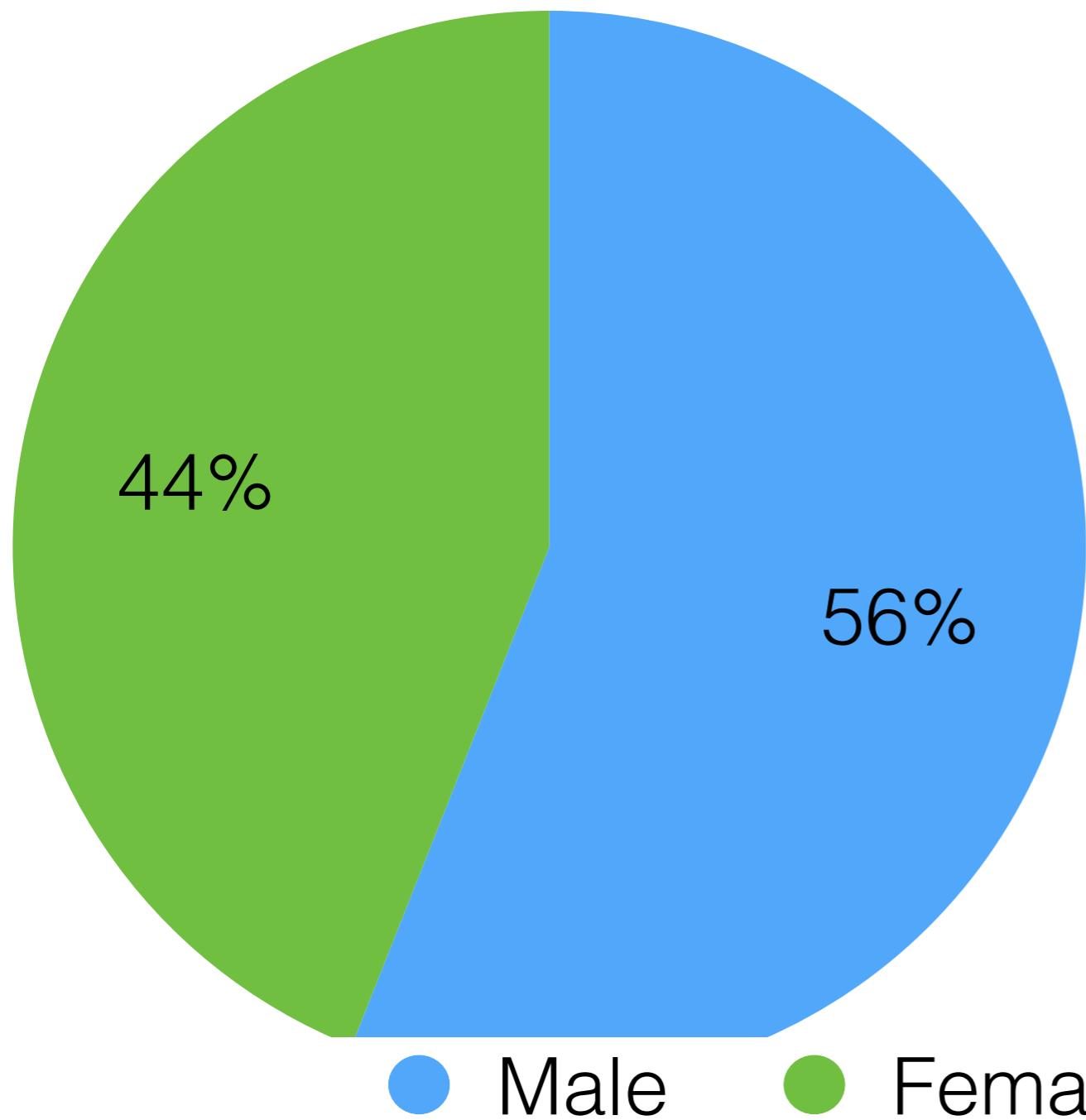


Work outsourced to
workers on smaller
platforms (e.g.
gaming sites)

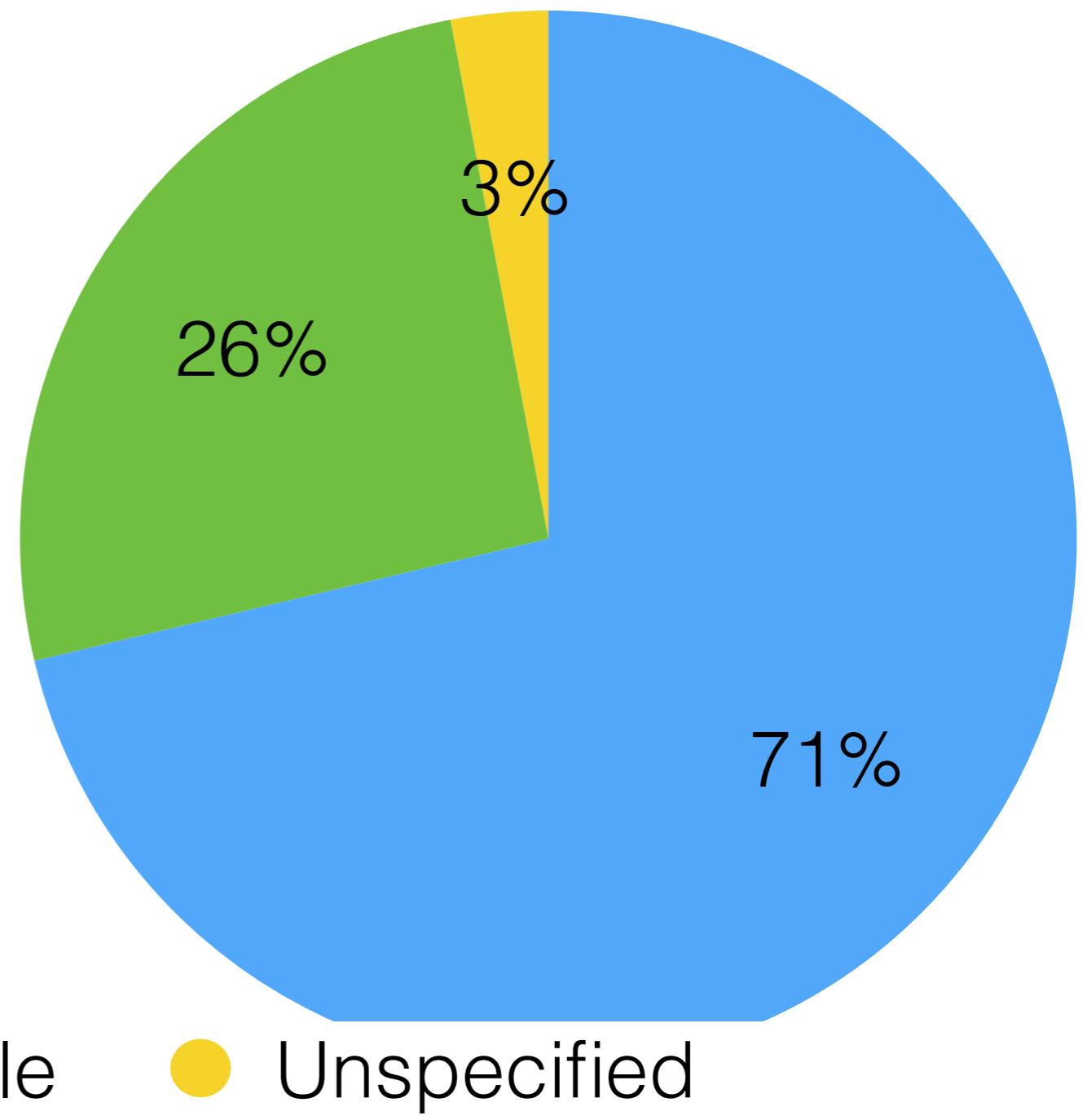
Who is doing all this work for us?

Who is doing all this work for us?

Mechanical Turk



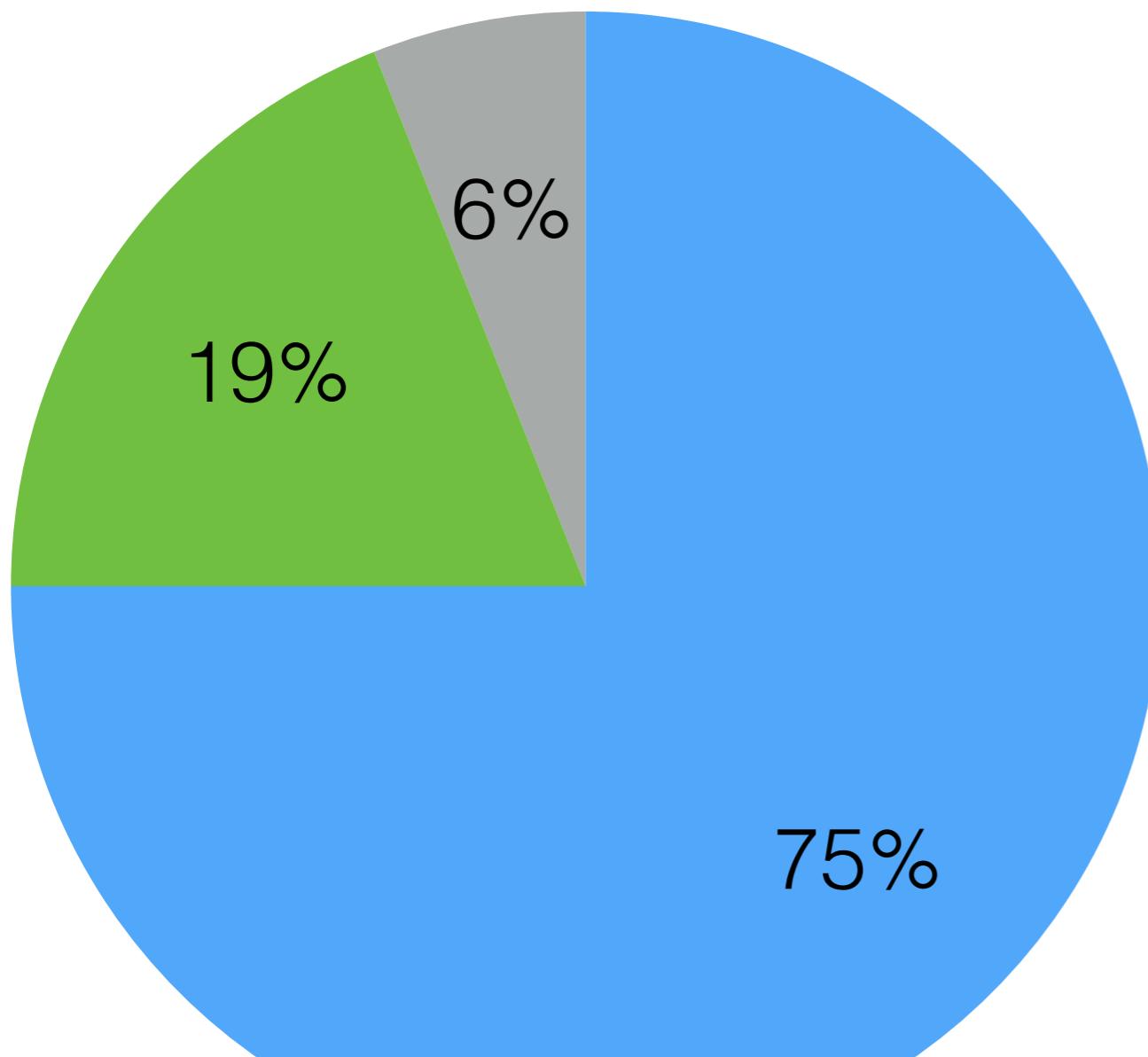
CrowdFlower



- Panos Ipeirotis's site: <http://demographics.mturk-tracker.com>
- CrowdFlower blog: <http://www.crowdflower.com/blog/2014/01/demographics-of-the-largest-on-demand-workforce>

Who is doing all this work for us?

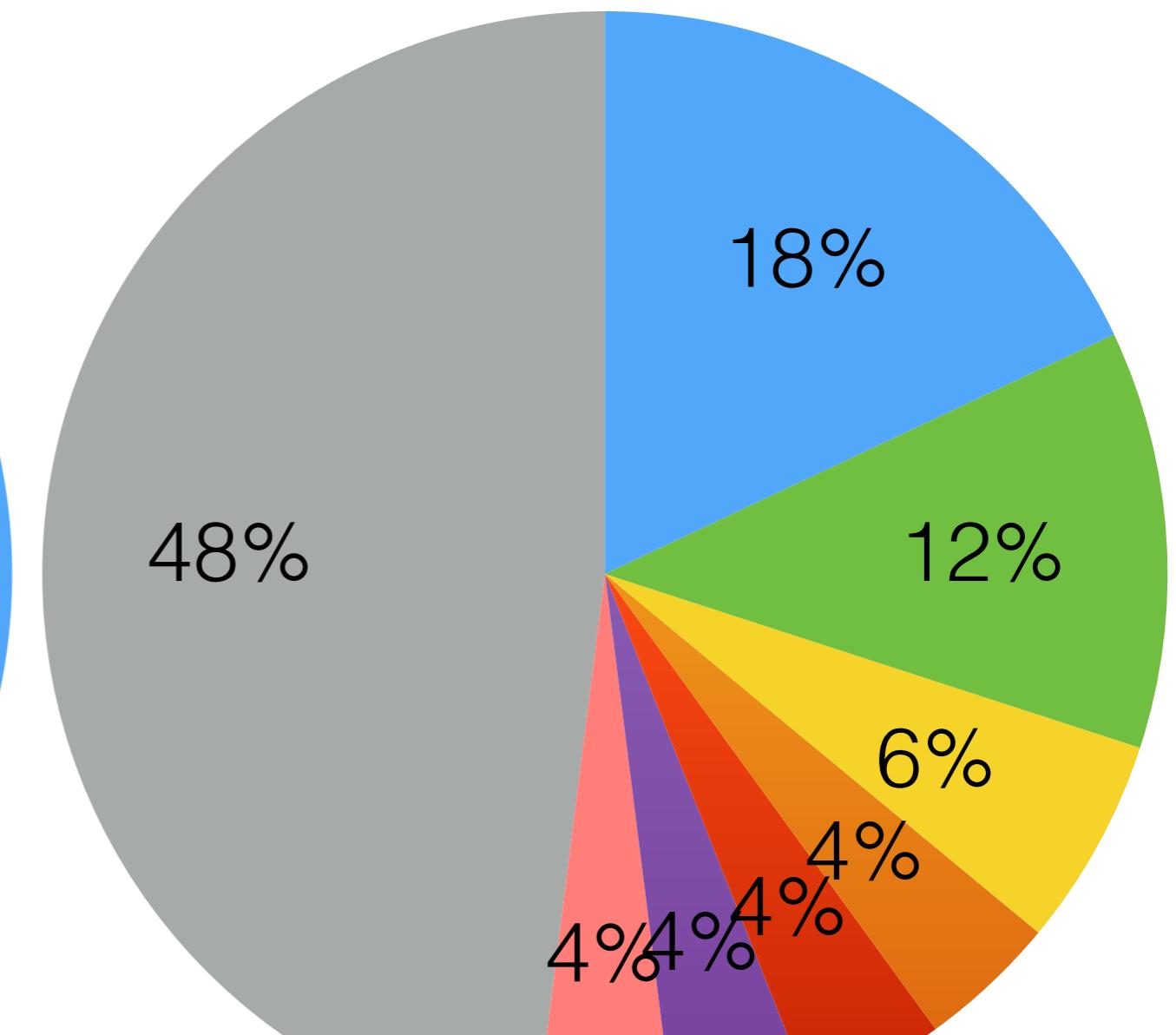
Mechanical Turk



- US
- Canada

- India
- Philippines

CrowdFlower

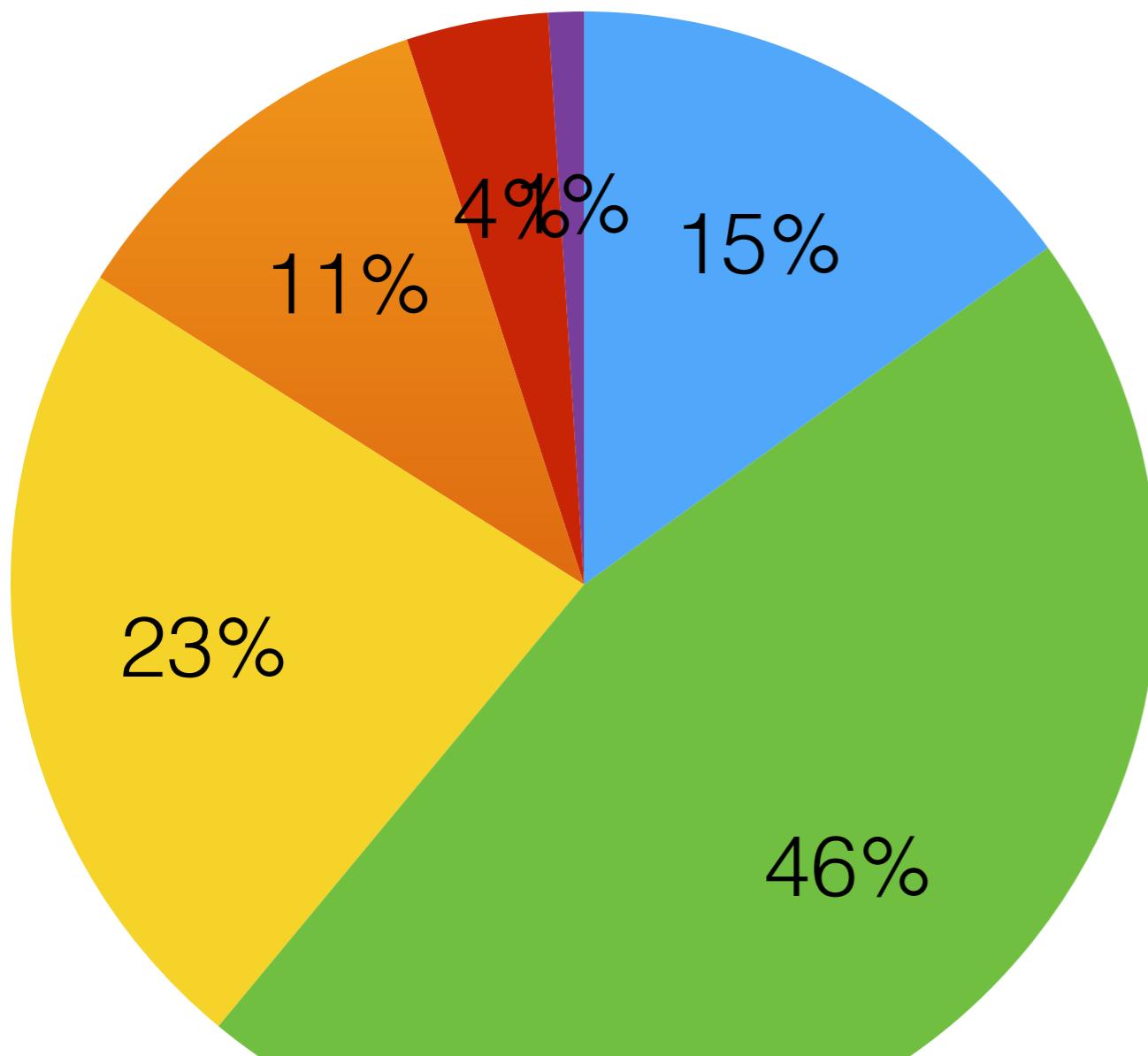


- UK
- Pakistan

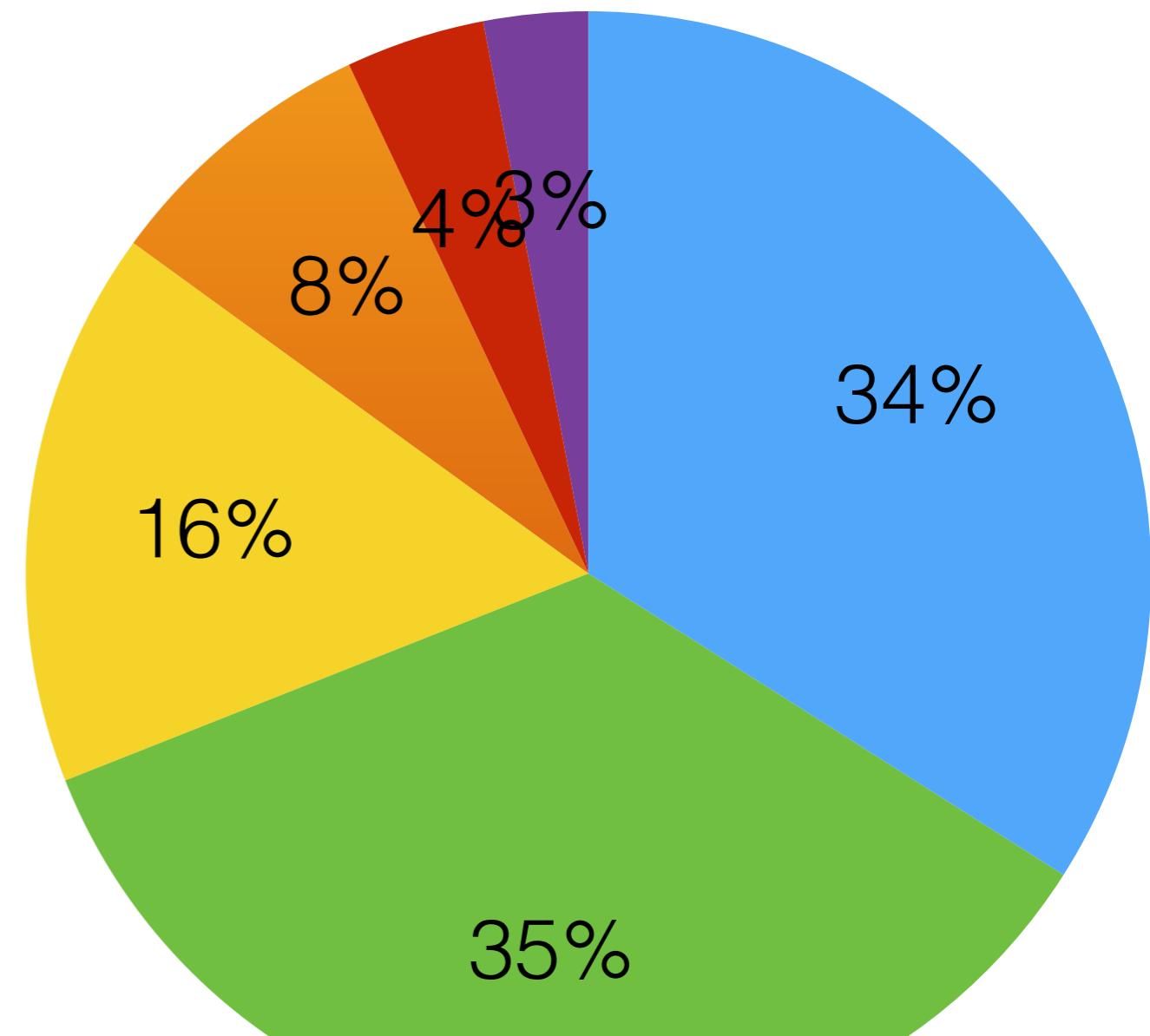
- Indonesia
- Others

Who is doing all this work for us?

Mechanical Turk



CrowdFlower

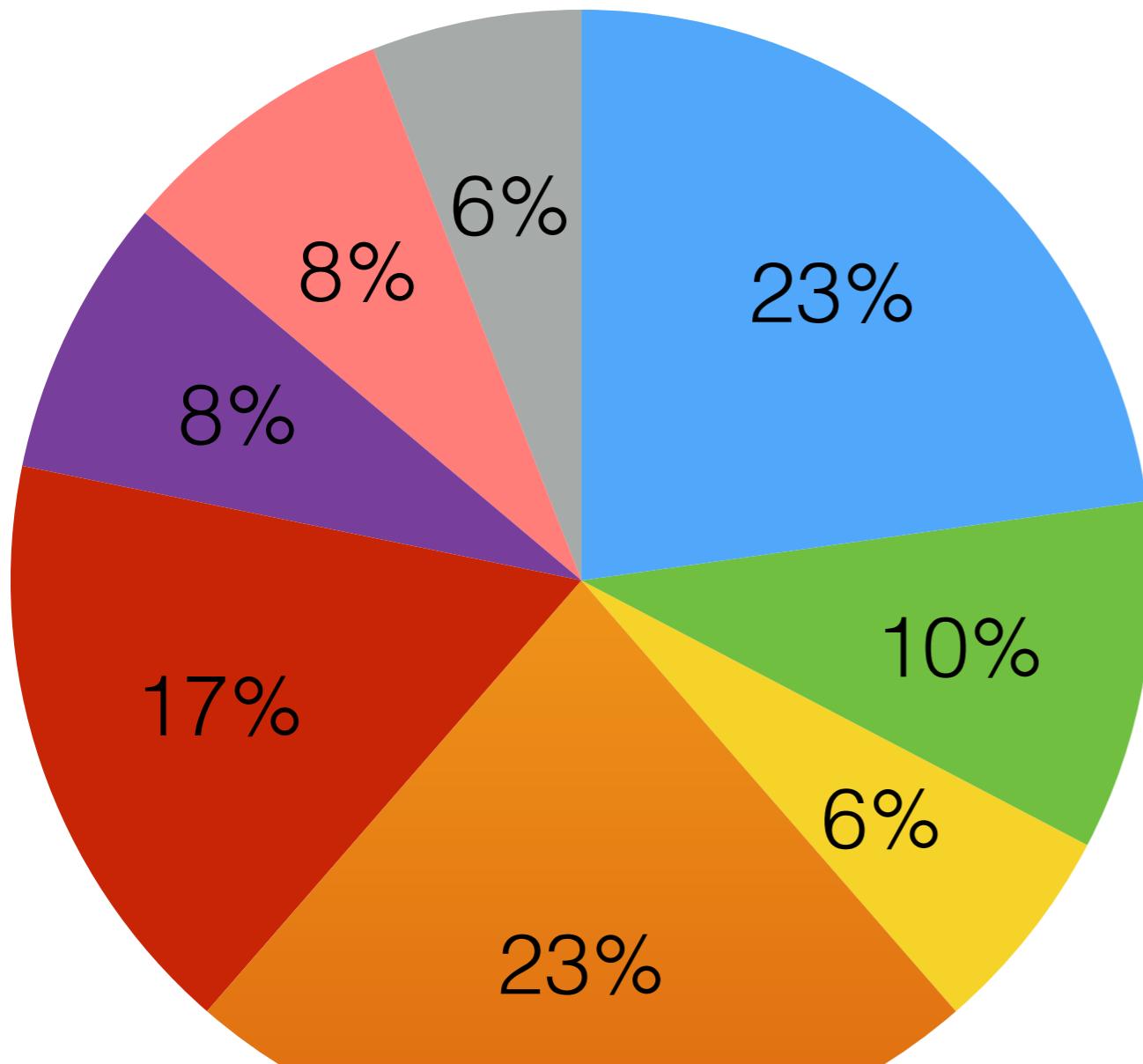


- 25 or younger
- 45-55

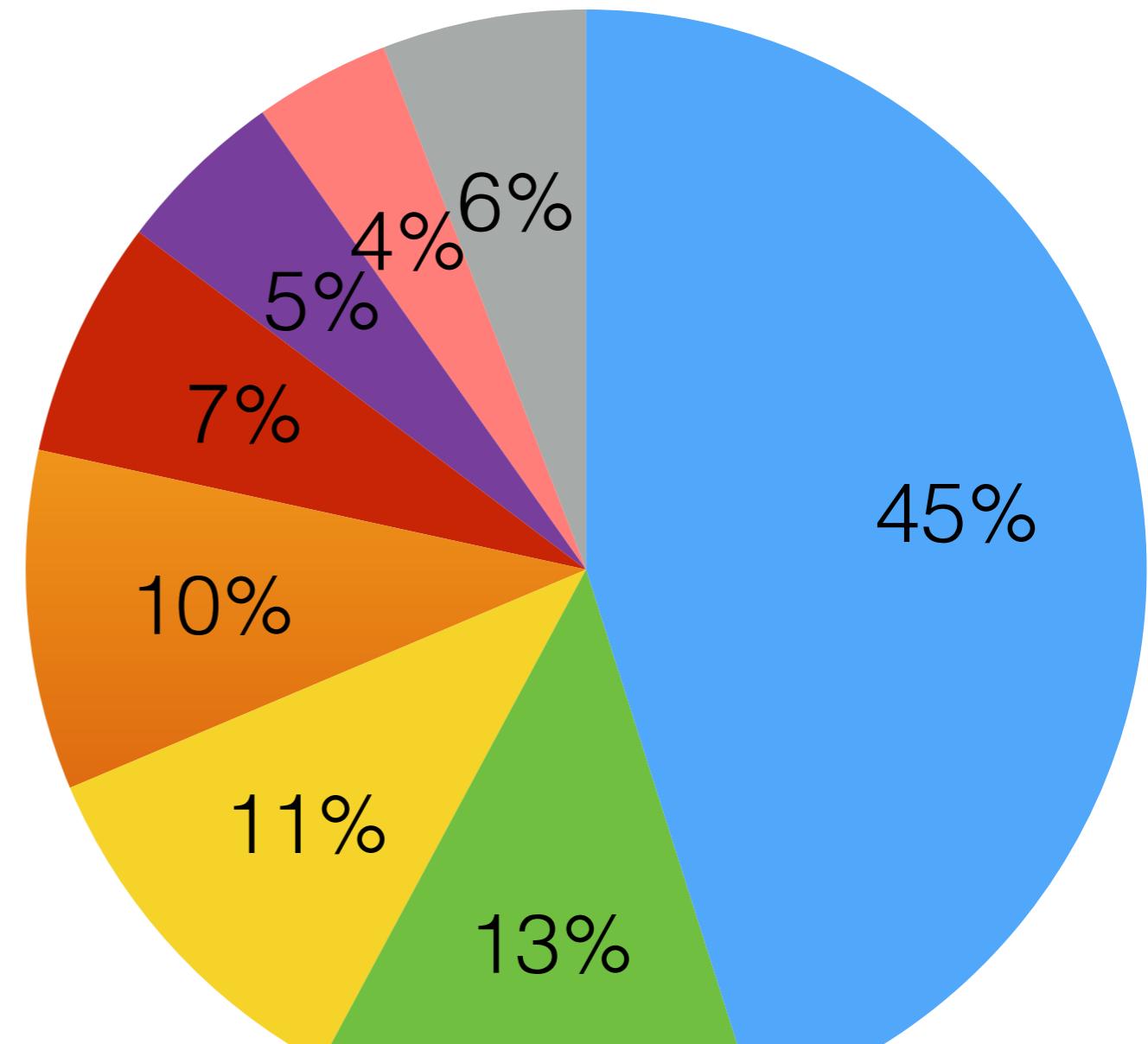
- 25-35
- 55-65
- 65 or older

Who is doing all this work for us?

Mechanical Turk



CrowdFlower



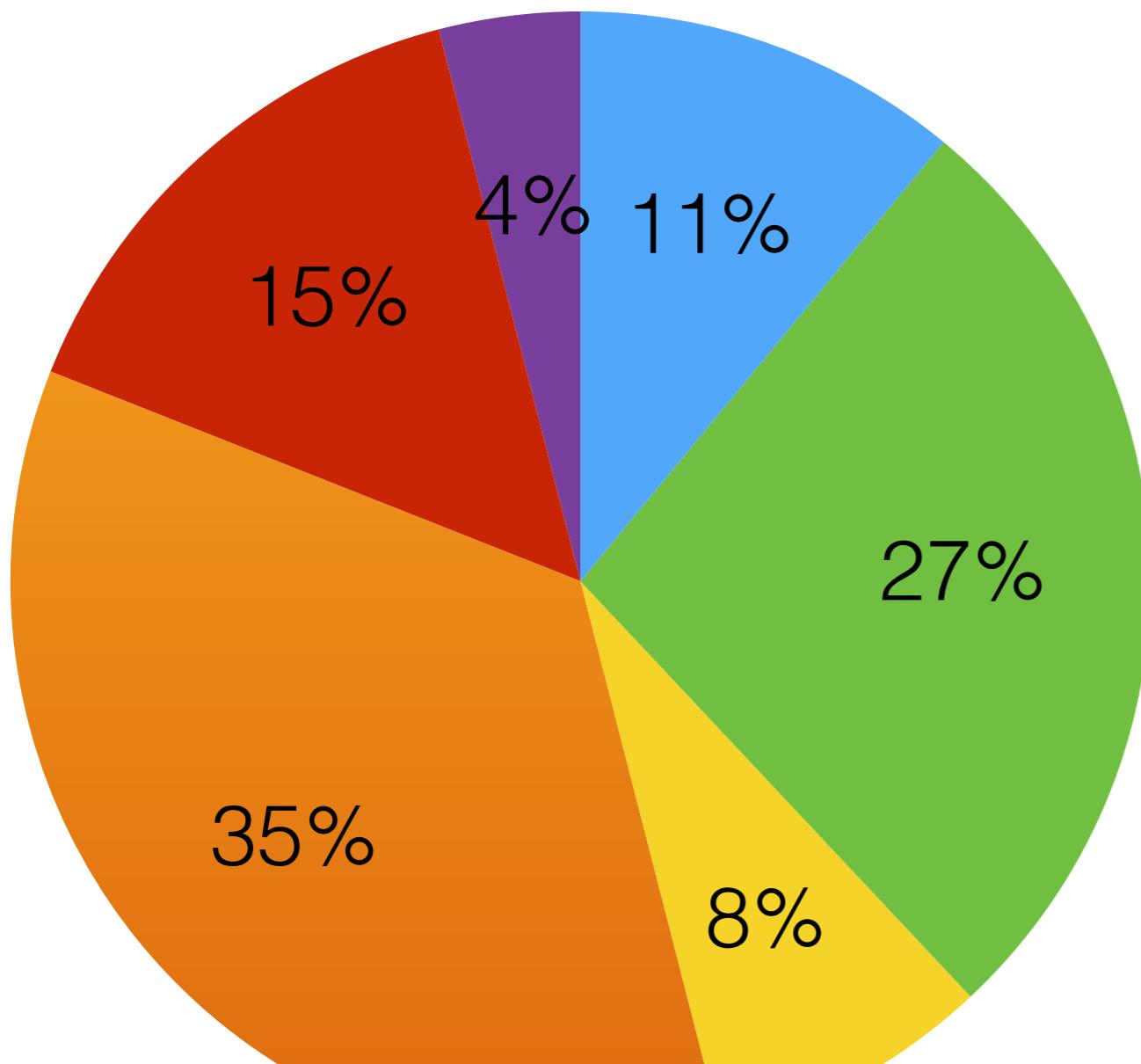
● Less than 10K
● 40K-60K

● 10K-15K
● 60K-75K

● 15K-25K
● 75K-100K
● More than 100K

Who is doing all this work for us?

Mechanical Turk*



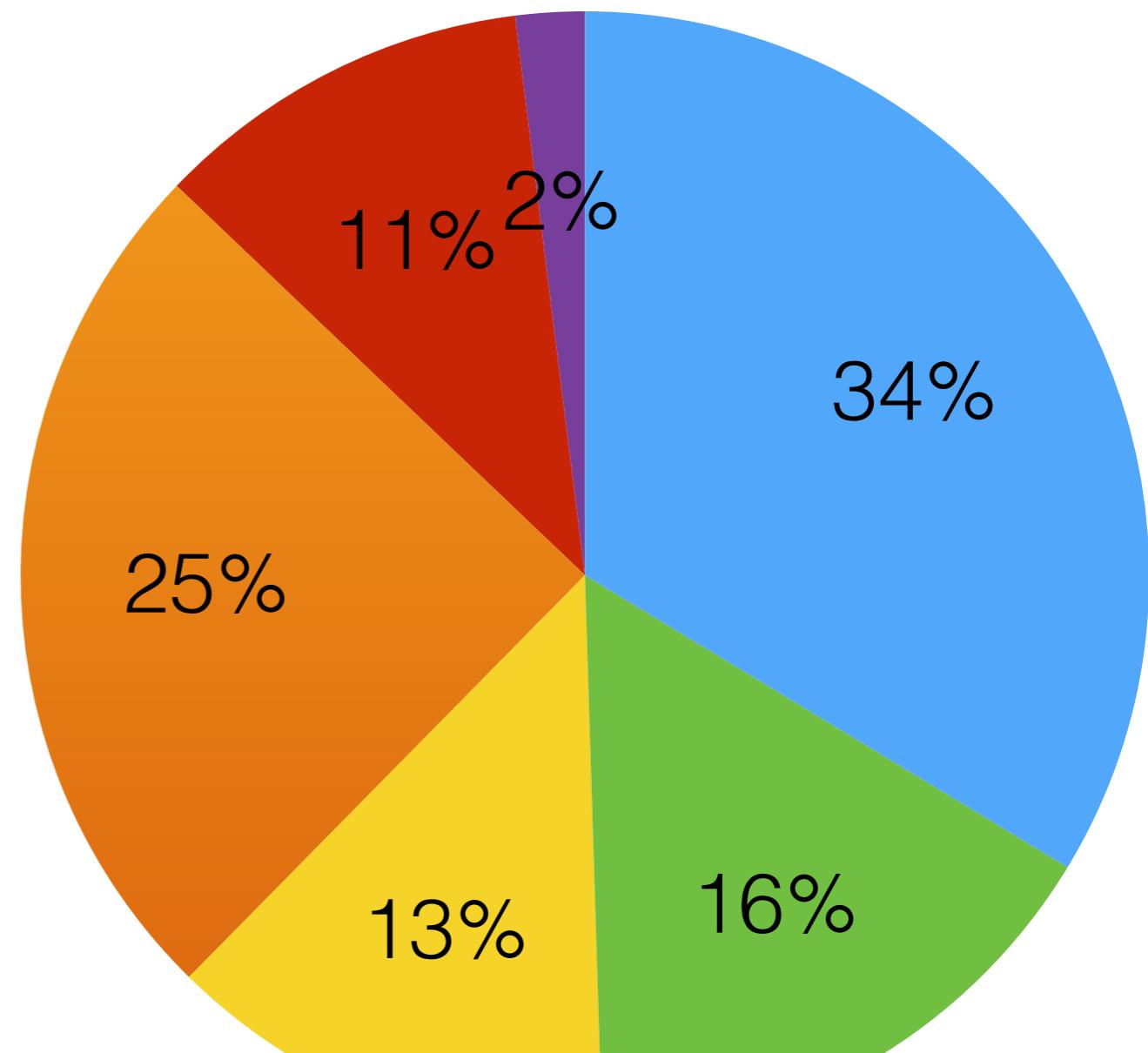
● High School
● Masters

● Some College
● PhD

● Associates

● Bachelors

CrowdFlower

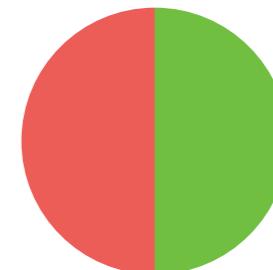
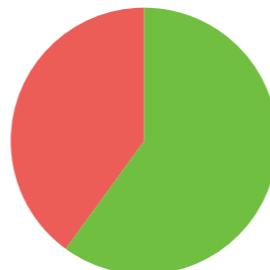


*Data is from 2010 and reflects only US workers

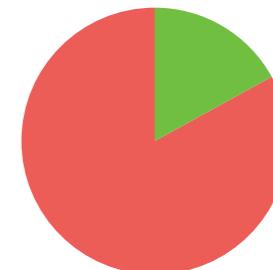
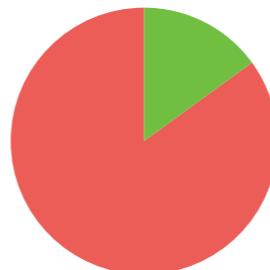
Why are they doing all this work for us?

Mechanical Turk* CrowdFlower

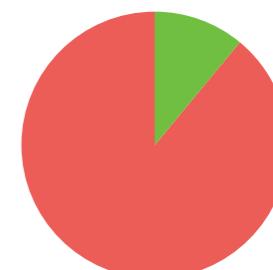
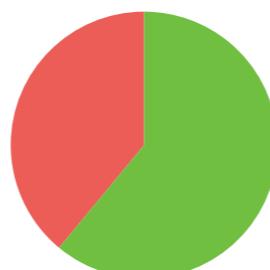
Good way to spend free time and earn money (e.g. instead of TV)



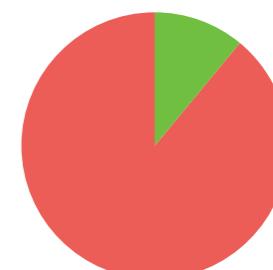
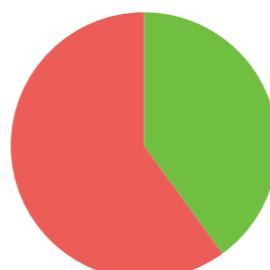
As a primary source of income



As a secondary source of income/
pocket change



It is just so much fun!!



*Data is from 2010 and reflects only US workers

82 <http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html>
• <http://www.crowdflower.com/blog/2014/01/demographics-of-the-largest-on-demand-workforce>



Common Misconceptions:

Common Misconceptions:

- Only from developing countries, non-native English speakers, uneducated, unskilled

Common Misconceptions:

- Only from developing countries, non-native English speakers, uneducated, unskilled
- Work for \$1/hour, doing it for fun in our PJs, unemployed

Common Misconceptions:

- Only from developing countries, non-native English speakers, uneducated, unskilled
- Work for \$1/hour, doing it for fun in our PJs, unemployed
- Isolated, anti-social

Common Misconceptions:

- Only from developing countries, non-native English speakers, uneducated, unskilled
- Work for \$1/hour, doing it for fun in our PJs, unemployed
- Isolated, anti-social
- Cheaters, lazy, satisficers, inattentive

Comm

I sat contemplating that question over and over. I wanted to say not necessarily true or false but at the last moment decided to change my mind about it.

unskilled

I currently have a 98.4% approval rating. I got the extensions that warned me as requesters who mass reject I unfortunately was victim to many of them who dropped my approval rating. I was wondering if you could make an exception to your rule for me... **Obviously if you aren't happy with my work you could take away my qualification** to work on them. Thank you for your consideration.

onceptions:

I did an awful lot of these HITs. For my part, it was because they pay very well and I enjoy them quite a bit--**finally a productive use for my hitherto underutilized English degree!**

How Turkers Work

- 10-20% of workers do 80% of the work
- Want large batches with high throughput
- Often dislike one-off HITs, e.g. surveys

How Turkers Work

- Online communities: Turkopticon, TurkerNation, Reddit, Facebook
- Scripts: IndiaTurkers, GreasyFork, HitDB, TurkMaster, HIT Scraper
- Websites and plugins: Turk Alert, mTurk List, CrowdWorkers

Creating and running experiments on Mechanical Turk

[Home](#)[Create](#)[Manage](#)[Developer](#)[Help](#)[New Project](#) New Batch with an Existing Project

Create HITs individually

Start a New Project

Categorization

[Data Collection](#)[Moderation of an Image](#)[Sentiment](#)[Survey](#)[Survey Link](#)[Tagging of an Image](#)[Transcription from A/V](#)[Transcription from an Image](#)[Writing](#)[Other](#)

Example of Categorization

Choose the best category for this image



- kitchen
- living
- bath
- bed
- outside

[View Instructions ↓](#)

Select the room location in home for this picture. Seating areas outside are outside not living. Offices or dens are living not bedrooms. Bedrooms should contain a bed in the picture.

You must ACCEPT the HIT before you can submit the results.

[Create Project »](#)

Categorization

Choose the best category for this image



[View Instructions ↓](#)

Select the room location in home for this picture. Seating areas outside are outside not living. Offices or dens are living not bedrooms. Bedrooms should contain a bed in the picture.

- kitchen
- living
- bath
- bed
- outside

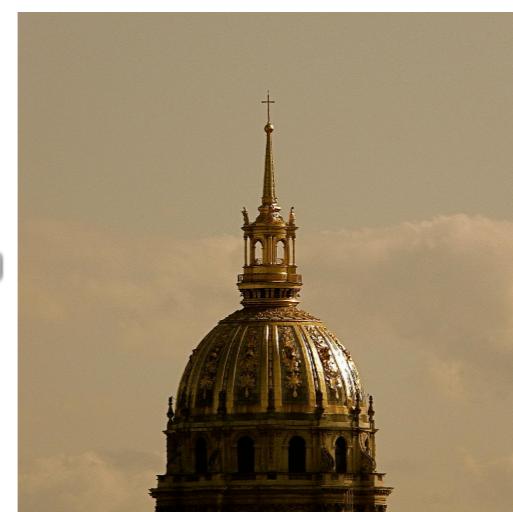
You must ACCEPT the HIT before you can submit the results.

Image Moderation

Select all images containing adult content

Guidelines for flagging an image as adult content. Flag the image if you consider any of the following to be true.

- Does the image contain nudity?
- Does the image portray hate or hate crimes?
- Does the image contain bloody violence?
- Does the image contain offensive gestures?



Transcription

Transcribe the text contained in the image

- Look at the receipt and copy the number of items purchased.
- Provide the dollar amount for tax.
- Provide the dollar amount for the total sale.
- Do not use dollar signs (\$) but make sure you have two decimal points. (ie 4.35)

Image:



Number of items on receipt

Tax amount

Total spent on all items

Any HTML form

1. This is an example of a multiple choice question, select only ONE of the following:

- Choice 1
- Choice 2

2. This is an example of a free text answer, use this when you only need workers to enter a line of text:

3. This is another example of a multiple choice question, select only ONE of the following. A drop down list saves more screen real estate than radio buttons (see Question 1) but may require more keystrokes for the Worker to select the answer they want.

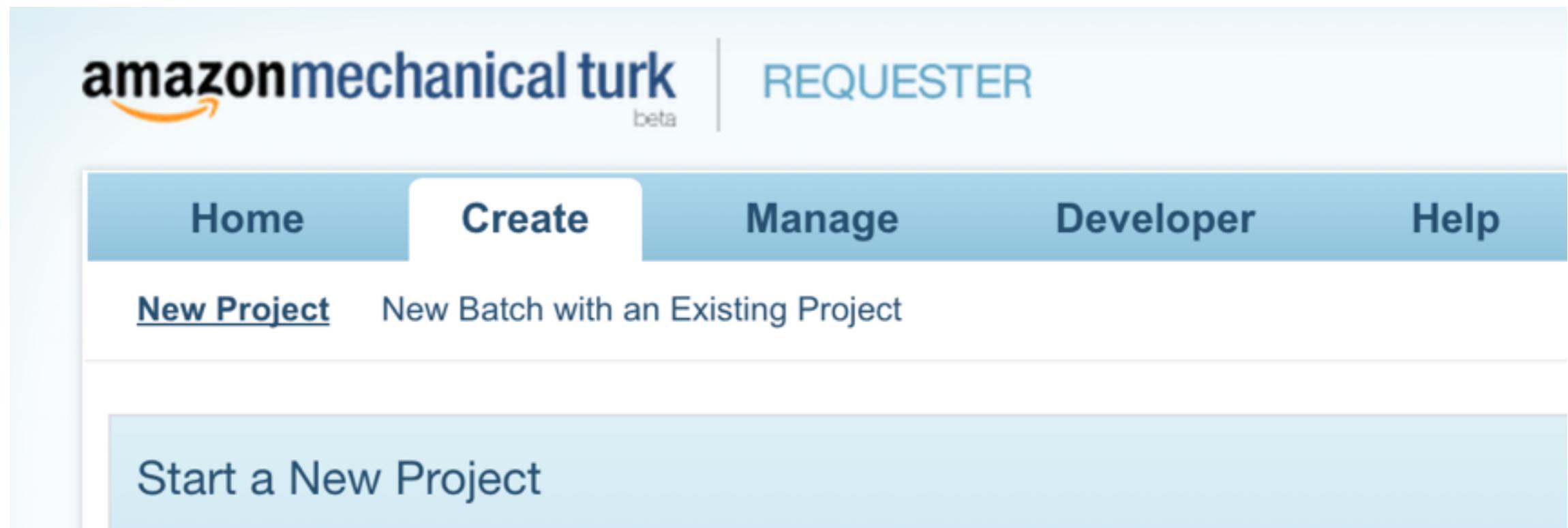
4. This is an example of a question where you can select one or more of these options:

- selection 1
- selection 2
- selection 3
- selection 4

5. This is another example of a question that can have multiple answers (select all that apply):

- Option 1
- Option 2

Requester/Worker Sandbox



<https://requestersandbox.mturk.com>

Paraphrase judgement

Instructions - click to hide

Please check **all** paraphrase which could replace the blue word in the context shown, using your best judgement. The paraphrase should **retain the meaning of the original phrase** and make sense in context. Most sentences will have at least one good paraphrase, but some will not. If none of the paraphrases listed fit the context shown, **select NONE**.

It is okay if the paraphrased sentence contains small grammatical errors, as long as they do not drastically alter the meaning. E.g. the following is an **okay paraphrase**:

The **boy** is running.

The **boys** is running.

but the following is a **bad paraphrase**:

I am running.

You are running.

17. occasionally , students spend the night in a youth hostel , where they share stories with peer travelers from around the world .

- now and then
- just sometimes
- only occasionally
- now and again
- every now and then
- some times
- once in a while
- for once
- here and there
- None of these paraphrases are good

Writing Instructions

Instructions - click to hide

Please check **all** paraphrase which could replace the blue word in the context shown, using your best judgement. The paraphrase should **retain the meaning of the original phrase** and make sense in context. Most sentences will have at least one good paraphrase, but some will not. If none of the paraphrases listed fit the context shown, **select NONE**.

Writing Instructions

Instructions - click to hide

Please check **all** paraphrase which could replace the blue word in the context shown, using your best judgement. The paraphrase should **retain the meaning of the original phrase** and make sense in context. Most sentences will have at least one good paraphrase, but some will not. If none of the paraphrases listed fit the context shown, **select NONE**.

I have a question about these hits. When the substitution makes the sentence end up saying something like, "The small girl girl was jumping rope," or other repetition of the same word, what should that be categorized as?

I'm a little confused about when one word is a variant of another. On the linked page of expanded definitions it says "other relation" should be used when it is two forms of the same word. It uses the example of ancestor and ancestry. But, on the instructions on each hit page the first example uses snow and snowy and says to put equivalent. Could you please help me understand how to classify this type?

Writing Instructions

Instructions - click to hide

Please check **all** paraphrase which could replace the blue word in the context shown, using your best judgement. The paraphrase should **retain the meaning of the original phrase** and make sense in context. Most sentences will have at least one good paraphrase, but some will not. If none of the paraphrases listed fit the context shown, **select NONE**.

It is okay if the paraphrased sentence contains small grammatical errors, as long as they do not drastically alter the meaning. E.g. the following is an **okay paraphrase**:

The **boy** is running.

The **boys** is running.

but the following is a **bad paraphrase**:

I am running.

You are running.

It is okay if the paraphrased sentence contains more/less detail than the original sentence, as long as it does not drastically alter the meaning. E.g. the following is an **okay paraphrase**:

The **person** is running.

The **boy** is running.

but the following is a **bad paraphrase**:

The cells undergo **development**.

The cells undergo **economic development**.

The paraphrase should make sense in the particular sentence shown. E.g. the following is a **bad paraphrase**, even though center might be a good paraphrase for heart in some other contexts.

The **heart** is an organ.

The **center** is an organ.

These HITs will be **quality controlled**. If you are unsure whether or not you are doing them well, we recommend you do a small number and wait for feedback before continuing.

Thank you in advance!

Writing Instructions

Instructions - click to hide

Please check **all** paraphrase which make sense in context. Most sentences will be okay if the paraphrased sentence

The **boy** is running.

The **boys** is running.

but the following is a **bad paraphrase**:

I am running.

You are running.

It is okay if the paraphrased sentence

The **person** is running.

The **boy** is running.

but the following is a **bad paraphrase**:

The cells undergo **development**.

The cells undergo **economic development**.

The paraphrase should make sense in some contexts.

The **heart** is an organ.

The **center** is an organ.

These HITs will be **quality controlled**.

Thank you in advance!

- ✓ If your task can't be expressed in one paragraph + bullets, then probably break it up
- ✓ Include an example of a right answer and a wrong answer
- ✓ Clarify what you expect if the HIT is not doable (e.g. because of missing data or other problems)

g of the original phrase and make ONE.

/ paraphrase:

ving is an **okay paraphrase**:

se for heart in some other contexts.

ack before continuing.

Quality Control

← Previous 1 2 3 4 5 6 7 8 9 ... 15 16 Next →

[Approve](#) [Reject](#)

HIT ID ▲	Worker ID	Lifetime Approval Rate	Input.HIT Num	Input.Instance Id1	Input.Doc Id1	Input.Sent Id1	Input.Sentence1	Input.Two Before1
302U8RURJZ1S2C6CV3ZQM0GTGEONVJ	A2YTOX7SEDJ1TC	100% (23/23)	9	238	P155842	16	IMPRESSION: Developmental anomaly of the right ...	No definite abnormality is seen in the visualiz...
302U8RURJZ1S2C6CV3ZQM0GTGEONVJ	A312KFQ78WB3A4	99% (67/68)	9	238	P155842	16	IMPRESSION: Developmental anomaly of the right ...	No definite abnormality is seen in the visualiz...

Quality Control

15 16 Next →

Approve

302U8RUR

302U8RUR

- After reviewing a Worker's submission, you can
 - Approve it (worker gets paid)
 - Reject it (worker does not get paid)
 - Block the worker (worker is banned from working on any of your HITs, and may get blacklisted from MTurk)

Input.Two
Before1

o definite
normality is
en in the
ualiz...

o definite
normality is
en in the
ualiz...

Quality Control

- A lightweight version of the API is available by uploading a graded CSV
- You can download the results as a CSV file, grade it, and upload your approve or reject decisions
- Good for automatically grading embedding test questions

Approve

302U8RUR

302U8RUR

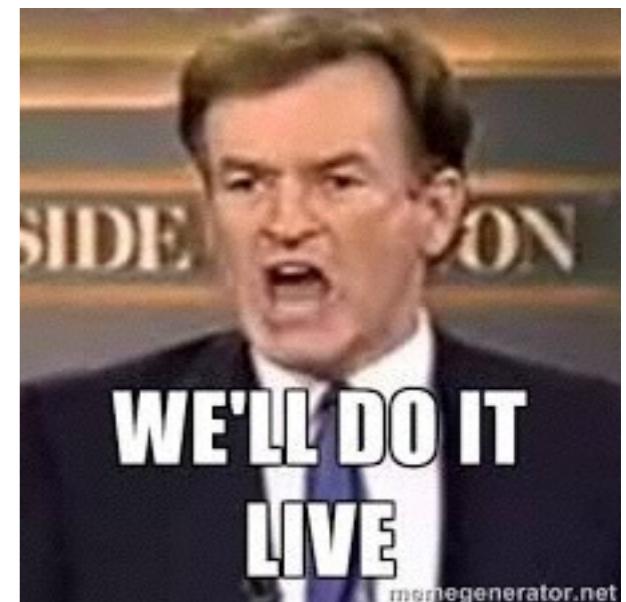
15 16 Next →

Input.Two
Before1

o definite
normality is
en in the
ualiz...

o definite
normality is
en in the
ualiz...

Demo



Instructions - click to hide

Please check **all** paraphrase which could replace the blue word in the context shown, using your best judgement. The paraphrase should **retain the meaning of the original phrase** and make sense in context. Most sentences will have at least one good paraphrase, but some will not. If none of the paraphrases listed fit the context shown, **select NONE**.

It is okay if the paraphrased sentence contains small grammatical errors, as long as they do not drastically alter the meaning. E.g. the following is an **okay paraphrase**:

The **boy** is running.

The **boys** is running.

but the following is a **bad paraphrase**:

I am running.

You are running.

17. occasionally , students spend the night in a youth hostel , where they share stories with peer travelers from around the world .

- now and then
- just sometimes
- only occasionally
- now and again
- every now and then
- some times
- once in a while
- for once
- here and there
- None of these paraphrases are good

Best Practices

- Crowdsourcing is good for tasks that are:
 - Easy to explain to **non-experts**
 - **Decomposable** into subtasks that can be chained together
 - **Parallelizable** into small, quickly completed chunks
 - **Balanced** so that each task has roughly the same workload. Don't ask workers to annotate something that occurs too infrequently (<<10% of the time).

Best Practices

- **Quality control** whenever and however you can
 - **Be nice:** rejecting Turkers has consequences for them, blocking Turkers is likely to get them expelled from MTurk
 - In our lab, we typically **embed test questions**, and then reject all work performing at chance, accept all ~80%, and reject proportionally to performance on controls for in between
 - Kristy's suggestion: 10K HITs with 95% approval, location restriction for English speaking countries

Best Practices

- Give workers the benefit of the doubt: **pay for honest work**, even if its not perfect
- Pay at least **minimum wage**. “If you need to underpay, use undergrads.”
- Take responsibility for your HITs: **Communicate**, respond to email, compensate for software bugs
- **Be transparent**: Use your real name, explain the project, give informed consent, monitor TurkOpticon

Go go go.
Have a nice long weekend.