

Web-based psych. experiments



timothy brady

assistant professor,
university of california
san diego

Research

CV

Resources

Contact

Mechanical Turk Tutorial

[Amazon Mechanical Turk](#) is a website where you can post short tasks and have workers quickly and easily perform these tasks for small sums of money. It is ideal for running short psychology experiments, since it allows large amounts of data to be collected quickly and easily. While a slightly less controlled environment than running studies in the lab, running studies on Turk has a number of advantages. In particular, it allows experimenters to (1) collect enough data to ensure their studies are not underpowered, as psychology studies chronically are; (2) collect data from a sample that is much more diverse than local college undergraduates and much more representative of the US at large; (3) easily replicate studies before publishing them to ensure that effects are real and effect sizes aren't inflated.

I have personally run many studies on Mechanical Turk (for example, Brady & Alvarez, 2011, *Psychological Science*; Brady & Tenenbaum, 2013, *Psychological Review*). To help others run such studies, I've put together a series of tutorials to help experimental psychologists learn to code Turk studies using HTML, CSS and Javascript (with jQuery). These tutorials were

Brady, T. F. and Tenenbaum, J.B. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 411-416).

Brady, T. F. and Alvarez, G.A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384-392.

Brady, T. F., and Tenenbaum, J.B. (2013). A probabilistic model of visual working memory: Incorporating higher-order regularities into working memory capacity estimates. *Psychological Review*, 120(1), 85-109.

Haberman, J., Brady, T. F. and Alvarez, G.A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, 144(2), 432-446

Brady, T. F. and Alvarez, G.A. (2015). No evidence for a fixed object limit in working memory: Ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory and Cognition*

Brady, T. F. and Alvarez, G.A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*.

...



Quickly find research participants you can trust.

Launch your study to tens of thousands of trusted participants in minutes. Recruit niche or representative samples on-demand. Prolific builds the most powerful and flexible tools for online research. Sign up for free.



UC San Diego



SONA: PSYCHOLOGY DEPT. SUBJECT POOL

User ID

bradylab



Password

.....



Log In

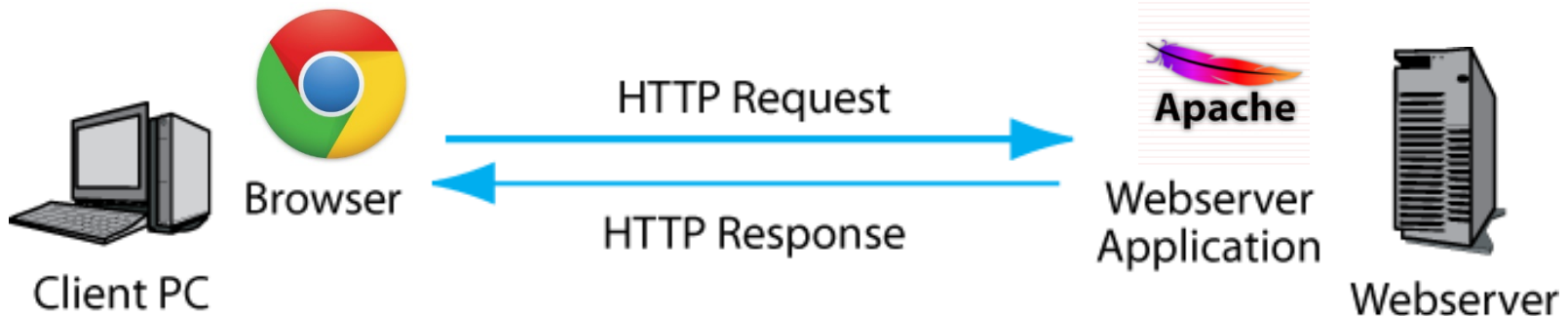
**The Psychology Department Subject Pool
includes participation from
Cognitive Science and Linguistics.**

SONA IS OPEN FOR PARTICIPANTS.

**Sona is open for Winter 2020 participants
from 4:00PM on Monday, January 6, 2020 until
4:00PM on Wednesday, March 11, 2020.**

The web

Basics of the web



HTTP requests are just text

```
GET /doc/test.html HTTP/1.1
Host: www.test101.com
Accept: image/gif, image/jpeg, */*
Accept-Language: en-us
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0
Content-Length: 35

bookId=12345&author=Tan+Ah+Teck
```

Many HTTP responses are just text

```
HTTP/1.1 200 OK
Date: Sun, 08 Feb xxxx 01:11:12 GMT
Server: Apache/1.3.29 (Win32)
Last-Modified: Sat, 07 Feb xxxx
ETag: "0-23-4024c3a5"
Accept-Ranges: bytes
Content-Length: 35
Connection: close
Content-Type: text/html

<h1>My Home page</h1>
```

This is the important part (the document itself)

What text does the server send back?

HTML

in <style>
or style="" :

CSS

in <script>:

JavaScript

```
1 <script>
2 function nextPage(whichIsNext) {
3     var currentPage = whichIsNext - 1;
4     document.getElementById('page' + currentPage).style.display = "none";
5     document.getElementById('page' + whichIsNext).style.display = "";
6 }
7
8 function CheckButtons() {
9     var isSelected1 = document.getElementById("buttonToCheck1").checked;
10    var isSelected2 = document.getElementById("buttonToCheck2").checked;
11    if (isSelected1 || isSelected2) {
12        nextPage(2);
13    } else {
14        alert("Please click one of the buttons before moving on.");
15    }
16 }
17 </script>
18
19 <style>
20 body {
21     font-size: 14pt;
22     font-family: Helvetica;
23 }
24 #page2 {
25     display: none;
26 }
27 </style>
28
29 <div id="page1">
30 <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas enim risus,
31 auctor vel faucibus sit amet, ultrices quis erat. Nulla facilisi. Sed vitae leo ac
32 lorem mollis fringilla at ac lacus. Fusce tristique quam sed arcu placerat euismod
33 dignissim turpis lacinia. Nullam porttitor tempor purus sit amet scelerisque. Nunc
34 ornare sem ac tortor sodales in tincidunt nisi pretium. Integer sagittis eleifend
35 erat, non varius nibh vulputate eu. Sed quis leo urna. Phasellus id massa nisl.
36 Duis turpis leo, fringilla a imperdiet quis, lacinia ac velit. Curabitur blandit
37 risus in erat tristique quis elementum nibh blandit. Etiam ac quam eros. Donec
38 purus turpis, ullamcorper non dictum eu, pharetra et nisl. Donec eu ligula mi, a
39 venenatis mauris. Aliquam sagittis congue eros, nec fermentum est consectetur vel.
40 Proin placerat sem at eros pharetra facilisis. Lorem ipsum dolor sit amet,
41 consectetur adipiscing elit. In aliquet dictum felis, nec condimentum nisi
42 elementum vel.</p>
43 <input id="buttonToCheck1" type="radio" name="q">Value1<br>
44 <input id="buttonToCheck2" type="radio" name="q">Value2<br>
```

Example webpages designed to be run as web experiments

Collect RTs from free responses (to words):

<http://timbrady.org/turk/monthnumber/>

Ravens matrices (with timer):

<http://timbrady.org/turk/ensemblravens/>

Multiple object tracking:

<http://timbrady.org/turk/mot/MOT.html>

Full video games:

<http://timbrady.org/turk/wmgames/gaborTask/>

Very briefly flashed images:

http://timbrady.org/turk/ensemblscenes/new_e5/demo.html

Continuous report with colors and faces:

<http://timbrady.org/turk/ensembleindivdiff/e9/>

Videos and audio (courtesy Prof. Adena Schachner)

http://timbrady.org/turk/adena/IfDanceThenWhatGoal_new.html

http://timbrady.org/turk/adena/OrderlySoundControl_TurkCode.html

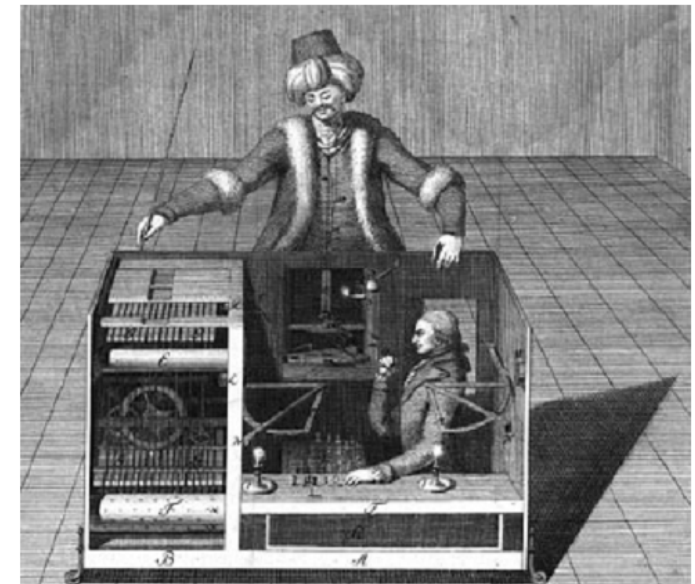
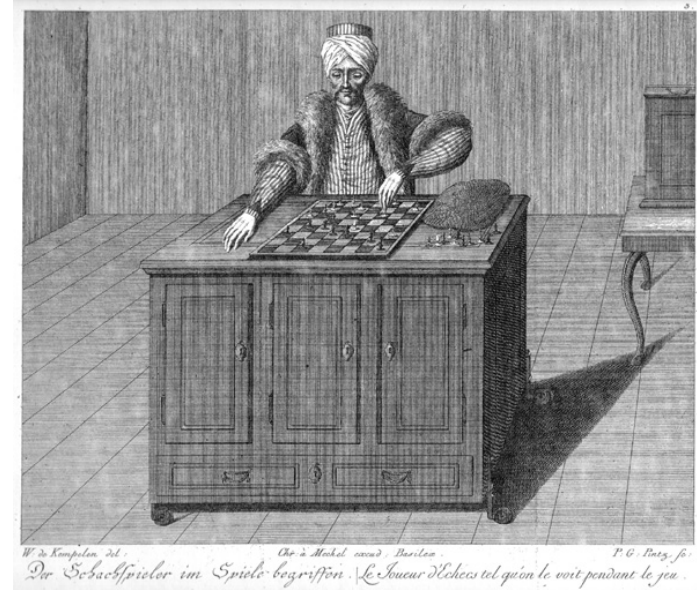
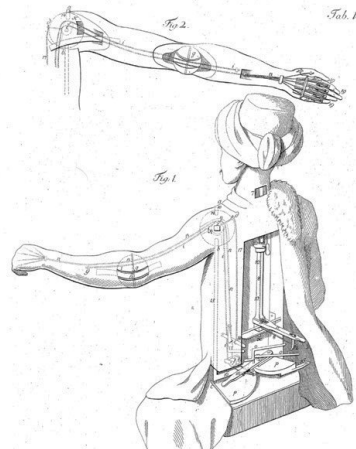
Turk

The name Mechanical Turk comes from *'The Turk'* a chess-playing automaton of the 18th century

It toured Europe for 70 years, beating the likes of Napoleon Bonaparte and Benjamin Franklin.

It was later revealed that this 'machine' was not an automaton at all but was in fact a chess master hidden in a special compartment controlling its operations.

"Likewise, the Mechanical Turk web service allows humans to help the machines of today perform tasks they aren't suited for."



Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

**Find an
interesting task**



Work



**Earn
money**



Amazon Mechanical Turk - All HITs

https://www.mturk.com/mturk/viewhits?searchWords=&selectedSearchType=hitgroups&sortType=LastUpdat

amazonmechanicalturk Artificial Intelligence

Your Account | **HITs** | Qualifications | 93,352 HITs available now

All HITs | HITs Available To You | HITs Assigned To You

Search for containing that pay at least \$ for which you are qualified

All HITs
51-60 of 1966 Results

Sort by: [Show all details](#) | [Hide all details](#) [First](#) << [Previous](#) < 4 5 6 7 8 > [Next](#) >> [Last](#)

Grade Snippet of Audio Transcription View a HIT in this group			
Requester: CastinoWords	HIT Expiration Date: Nov 19, 2010 (7 hours 46 minutes)	Reward: \$0.03	
	Time Allotted: 1 hour 30 minutes	HITs Available: 6	
Event Moods and Modes View a HIT in this group			
Requester: Restaurant Recommender	HIT Expiration Date: Nov 26, 2010 (6 days 23 hours)	Reward: \$0.10	
	Time Allotted: 10 minutes	HITs Available: 676	
Review & Proof this short article View a HIT in this group			
Requester: Todd Dickerson	HIT Expiration Date: Nov 20, 2010 (23 hours 46 minutes)	Reward: \$0.15	
	Time Allotted: 60 minutes	HITs Available: 1	
Grade Audio Transcription View a HIT in this group			
Requester: CastinoWords	HIT Expiration Date: Nov 20, 2010 (11 hours 46 minutes)	Reward: \$0.07	
	Time Allotted: 3 days 18 hours	HITs Available: 6	

<https://www.mturk.com/mturk/findhits?match=false>

All HITS | HITS Available To You | HITS Assigned To You

Search for containing that pay at least \$ for which you are qualified

Timer: 00:00:00 of 2 hours

Want to work on this HIT?

Want to see other HITS?

Total Earned: \$0.27
Total HITS Submitted: 4

Choose place name for panoramic pictures

Requester: Jianxiang Xiao

Reward: \$0.01 per HIT

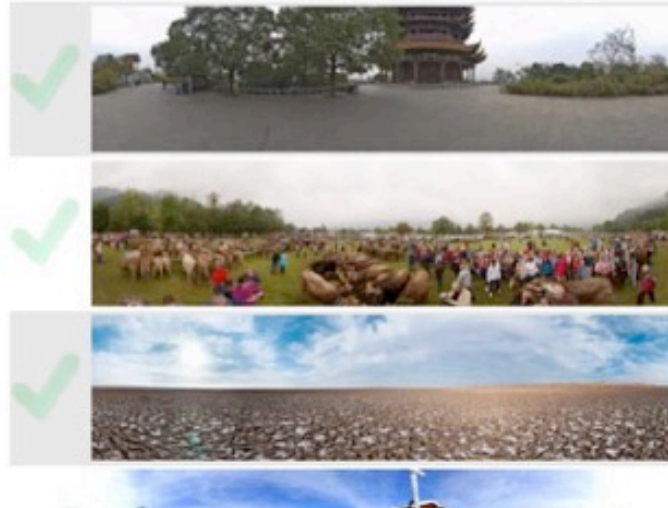
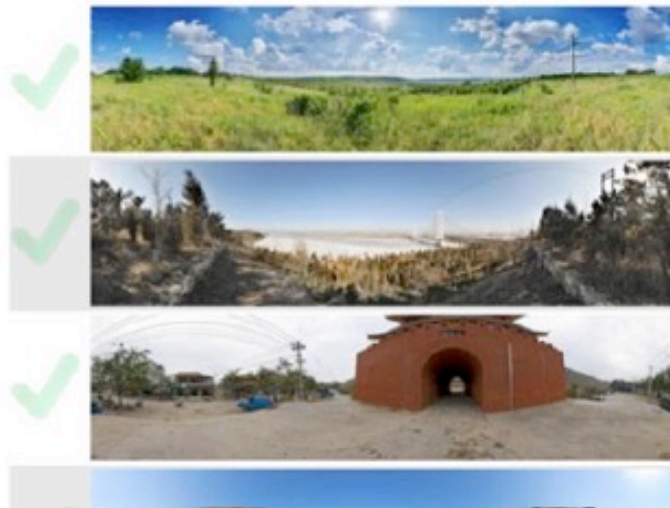
HITS Available: 3326

Duration: 2 hours

Qualifications Required: Location is US, HIT approval rate (%) is not less than 95

Identify street:

You are given a set of panoramic pictures showing 360-degree field of view for some places. Your task is to identify which place is a **street**. Please indicate your choice by clicking the check symbol (or the image) to highlight the green check. If no place is a **street** in this set, check the checkbox at the bottom of the page to clearly indicate this.



All HITs | HITs Available To You | HITs Assigned To You

Search for containing that pay at least \$ for which you are qualified

Timer: 00:00:00 of 2 hours

Want to work on this HIT?

Want to see other HITs?

Total Earned: \$0.27
Total HITs Submitted: 4

Choose place name for panoramic pictures

Requester: Jianxiang Xiao

Reward: \$0.01 per HIT

HITs Available: 3326

Duration: 2 hours

Qualifications Required: Location is US, HIT approval rate (%) is not less than 95

Any webpage we want can go here

Basics of Turk

- **Requesters vs. workers**
- **HIT:** A *task*. Any number of workers can do it (you pick).
- **Assignment:** A single worker doing a single HIT.
- **HIT group:** By default, the way to change anything about a HIT (e.g., change the image that is shown) is to have Turk automatically create a HIT group, with one HIT for each image or each word (or whatever)
- The same worker can do any number from 1 to all of those HITs.

Basics of Turk

- An example of this HIT group vs. HIT idea

Who are the Turk workers?

A210LUDZJ4NANM

A361NTFGICM88M

A2N82K0ZK3AQ8V

A3QBLTB8NST8WQ

A38TE493DFRC74

A11MYVCCB4MDB

A3ROFA6CFTKGPG

A1DRSS2H56ZS89

A330117G11 71G2

A25JEDCTON74BP

AW1URPW7KCAYM

A1TKNM0MWPZ73N

A3TK9UOPGR8J9Z

A3CAF7LTD3ORSG

A39IIPMXM6EGKZ

A1OPZJBOVBZC6W

A21VDW9G6AN6WL

AGPBBY9GD0Y14

A01NWX0000000

A2VC4DN2UVDGS6

A3P34V0J7POJYA

A1CJCG7IO9PNKW

A2XSI8BSDC4VC1

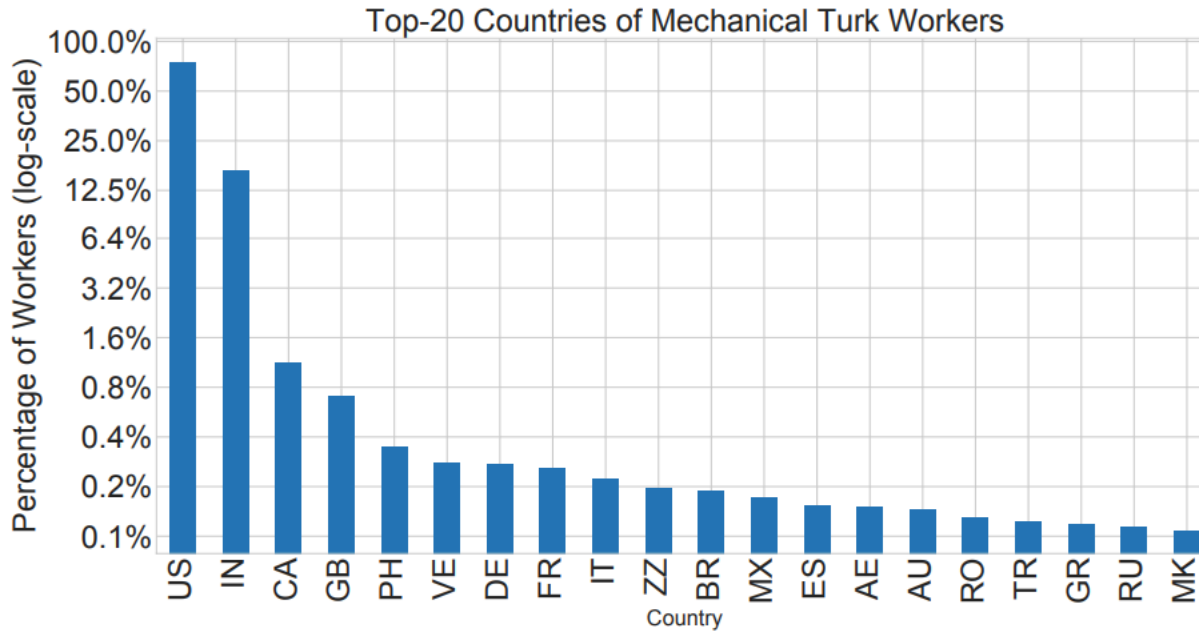
A20EBZCL2MEBEU

A3HRFT1NZYC9FN

ATNRQ92C5NN97

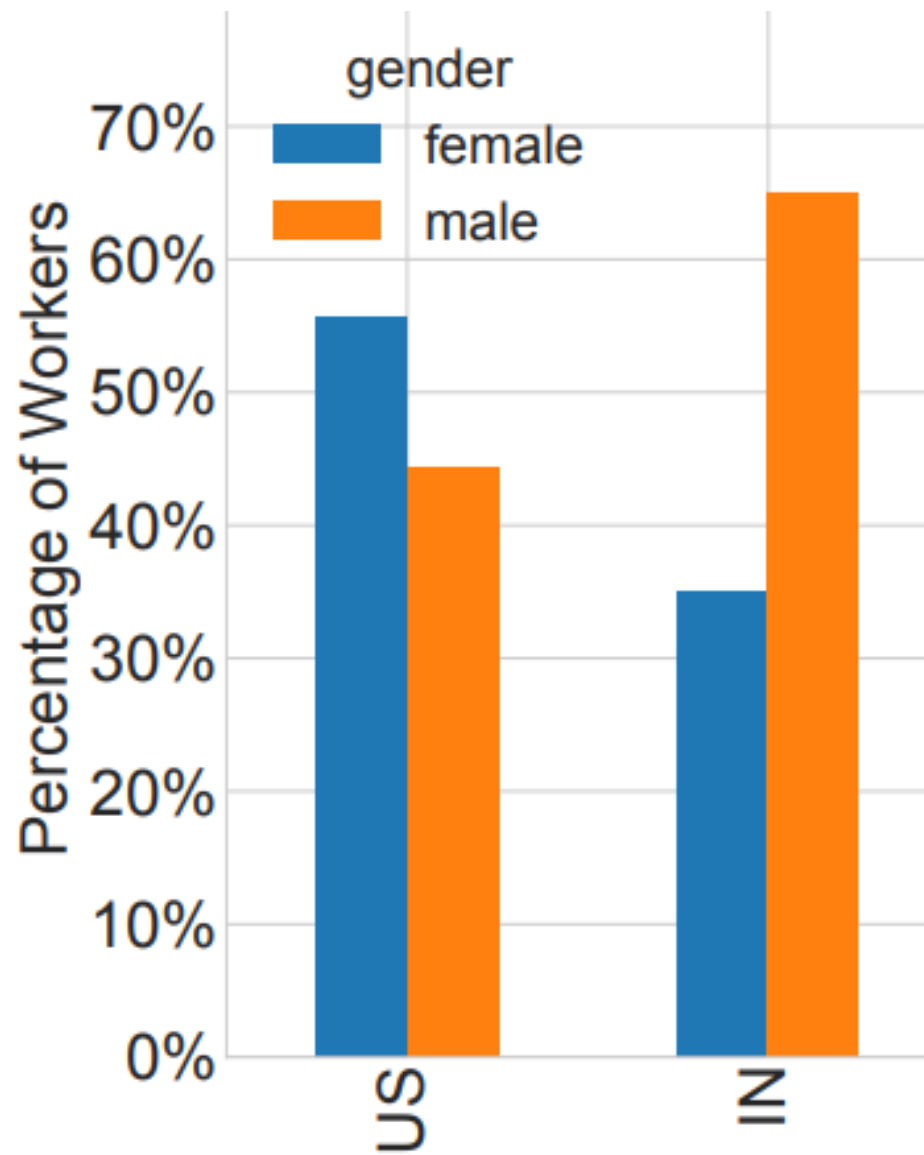
ATFYFF1G524HC

Who are the Turk workers?



Note log scale on y axis



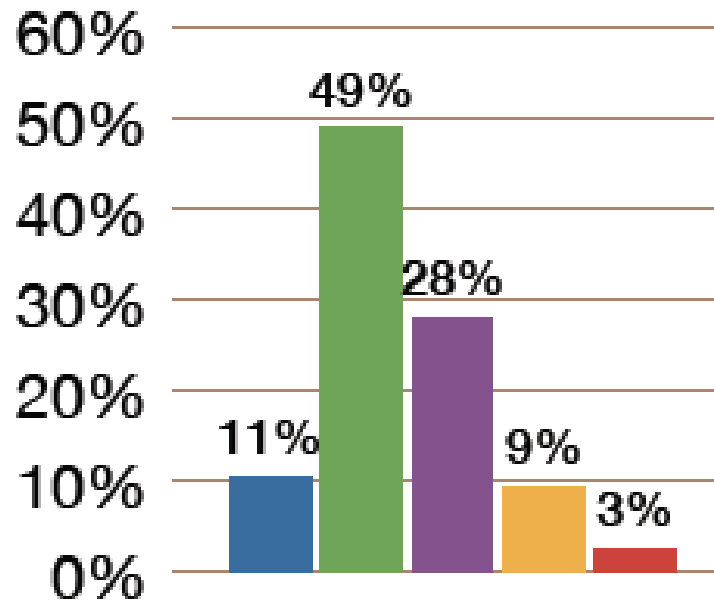


US-only demographics

	<i>MTurk</i>		
Female	60.1% (2.1)		
Education (mean years)	14.9 (0.1)	Housing status	
Age (mean years)	32.3 (0.5)	Rent	52.7 (2.3)
Mean income	\$55,332 (\$1,659)	Own home	47.3 (2.3)
Median income	\$45,000	Religion	
Race		None	41.8 (2.1)
White	83.5 (1.6)	Protestant	20.7 (1.7)
Black	4.4 (0.9)	Catholic	16.5 (1.6)
Hispanic	6.7 (1.1)	Jewish	4.4 (0.9)
Marital status		Other	16.5 (1.6)
Married	39.0 (2.1)	Region of the United States	
Divorced	7.1 (1.1)	Northeast	22.1 (1.8)
Separated	2.5 (0.7)	Midwest	26.6 (1.9)
Never married	50.6 (2.1)	South	30.9 (2.0)
Widowed	0.7 (0.4)	West	20.4 (1.7)

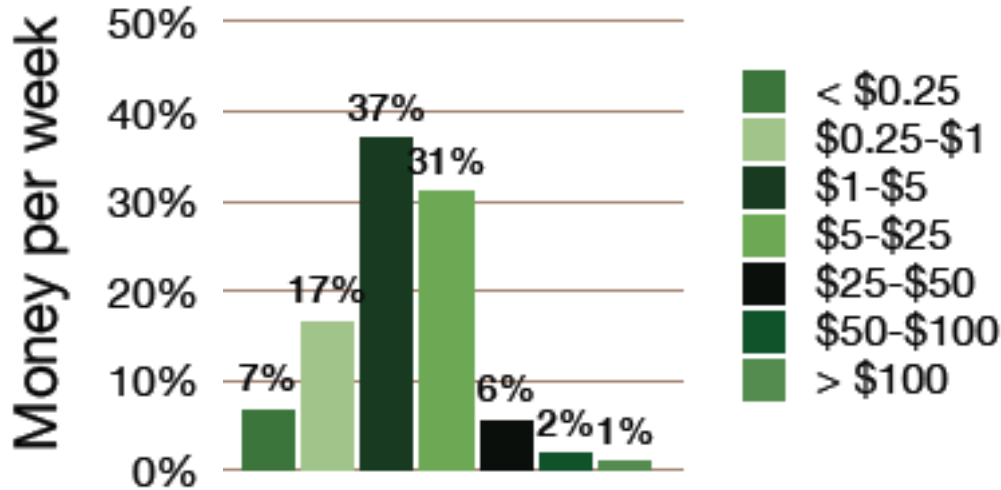
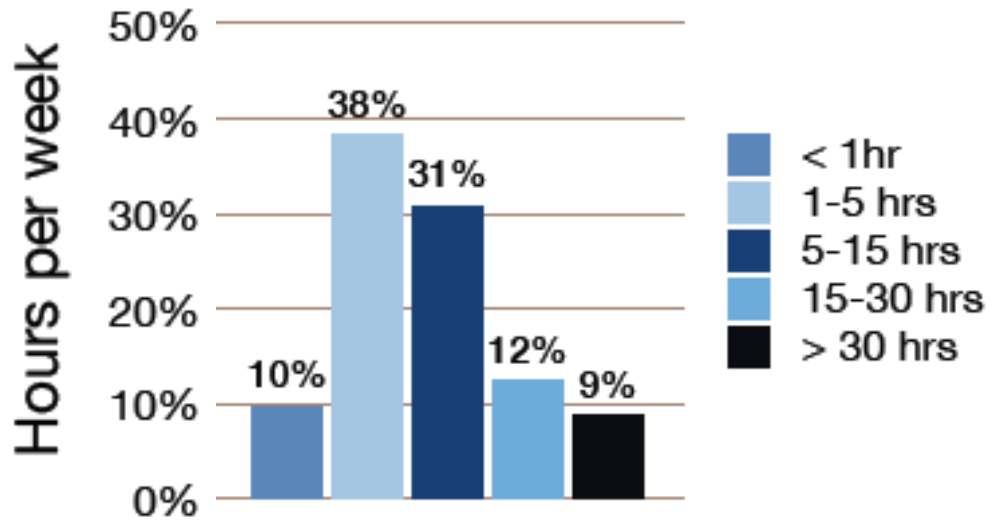
Workers tend to be younger, overeducated, underemployed, less religious, and more liberal than the general population

US-only demographics



- 'MTurk money is irrelevant to me.'
- 'MTurk money is nice, but doesn't materially change my circumstances.'
- 'MTurk money is a way for me to pay for nice extras.'
- 'MTurk money is sometimes necessary to make basic ends meet.'
- 'MTurk money is always necessary to make basic ends meet.'

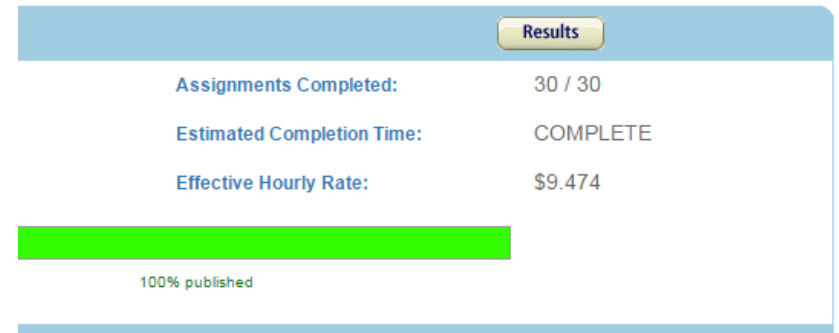
US-only demographics



60% make less than \$5/week from Turk and spend less than 15 hours a week working on it.

Payment on Turk

Most HITs offer very small payments and last less than 3 minutes.



The screenshot shows a 'Results' section with the following data:

Assignments Completed:	30 / 30
Estimated Completion Time:	COMPLETE
Effective Hourly Rate:	\$9.474

Below the table is a green progress bar that is 100% filled, with the text '100% published' underneath it.

25 percent of the HITs created on Mechanical Turk have a price tag of just US\$0.01, 70 percent have a reward of \$0.05 or less, and 90 percent pay less than \$0.10.

In general, HITs tend to be seen by the worker community as high reward if they pay >\$0.10 cents/minute (\$6/hr), assuming they are short (<10 min).

Longer hits need to pay more (in total and as a rate)

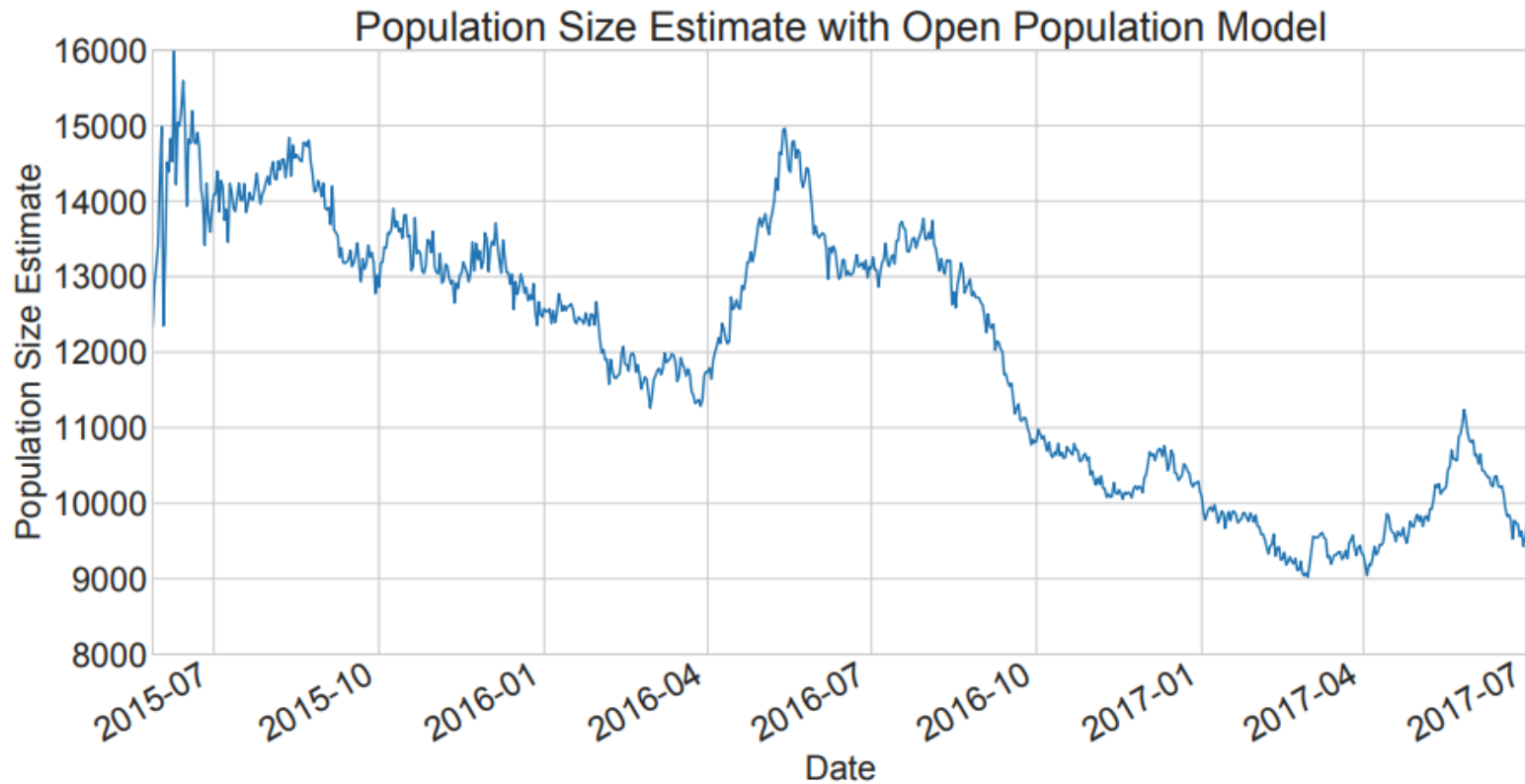
Payment on Turk

Must put money in Turk account **in advance**, before posting any HITs.
Means you need to figure out how reimbursements will work.

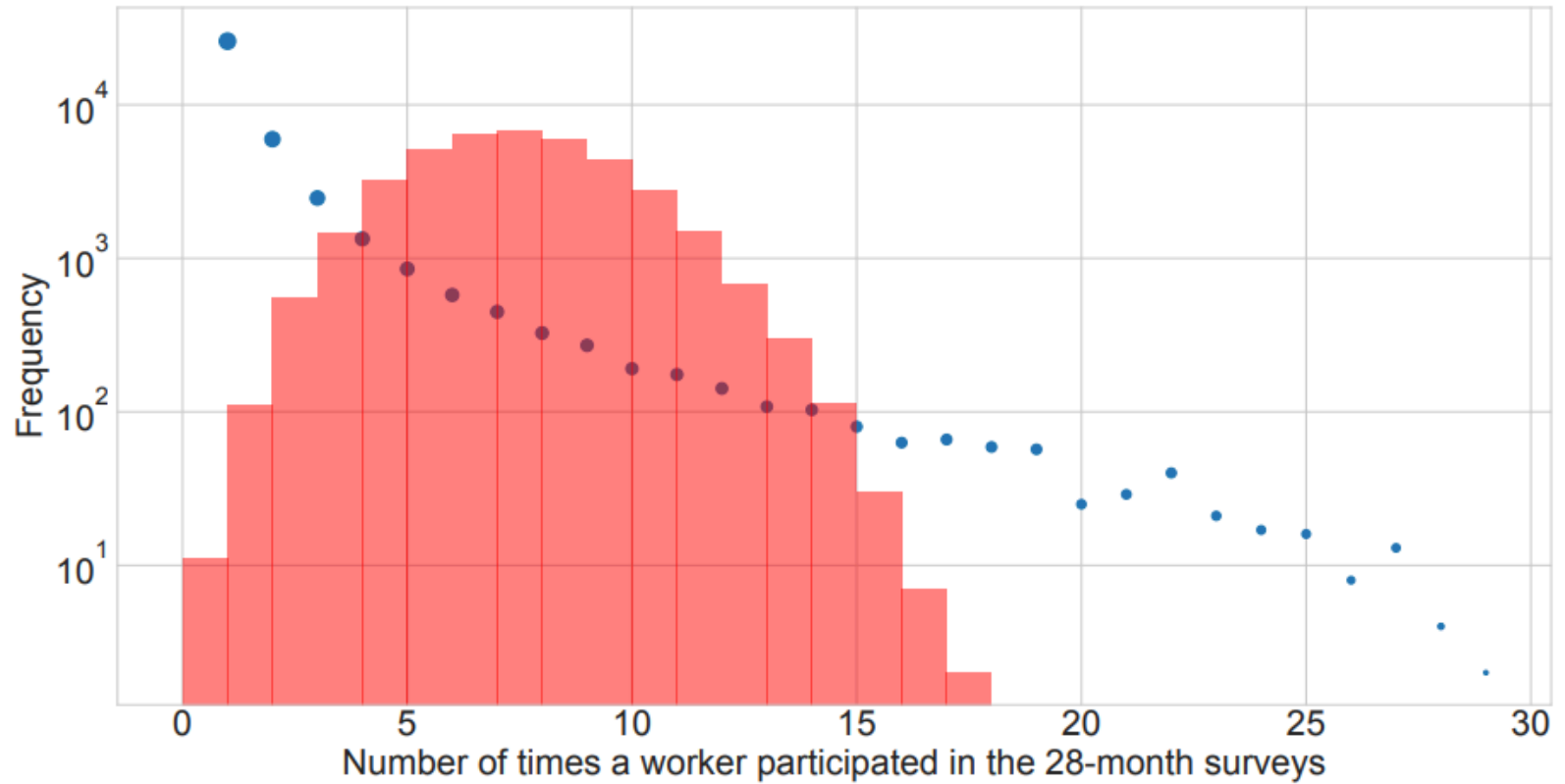
System at UCSD:

- Figure out how many participants you'll run in an entire batch (say, a few days worth) and how much this costs after Turk fees
- Charge a lab credit card for this amount
- Use all the money on participants/fees within a few days
- Submit proof from Turk to account for charge when next bill is due

How many unique Turkers?



How many unique Turkers?



How good are Turk workers?

- Many studies show the Turk sample is in many ways better than a similar lab based sample (e.g., Brady & Alvarez, 2011)
- Why?
- **Speculations:** shorter experiments, fear of rejection is high, HITs varied and fun, can quit easily at any time (*Note selection bias – be wary of doing hard vs. easy tasks across subject!*)

How good are Turk workers?

(1) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball.

How much does the ball cost? ____ cents

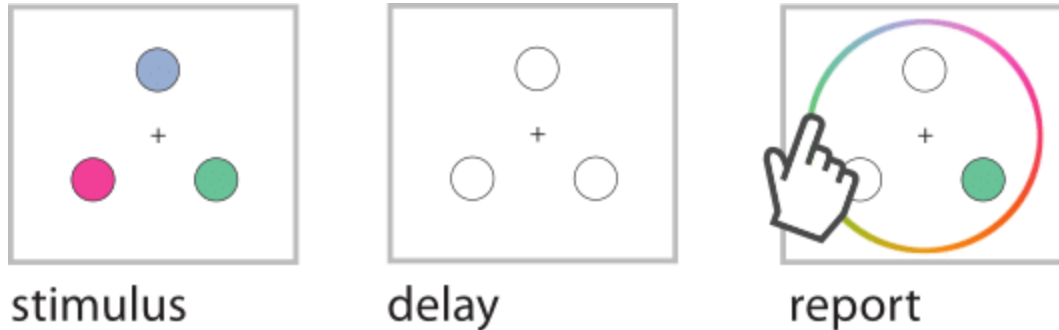
(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days

Cognitive reflection
test

	Mean	SE
MTurk	1.17	0.12
Community	0.96	0.16

How good are Turk workers?



With 6 colors to remember (this is very hard!):

K = 2.46 – Zhang & Luck, 2008, *Nature*

K = 2.52 - Brady & Alvarez, 2015, *JoV* (done on Turk)

How good are Turk workers?

- It totally depends on your task

Behav Res (2014) 46:112–130
DOI 10.3758/s13428-013-0365-7

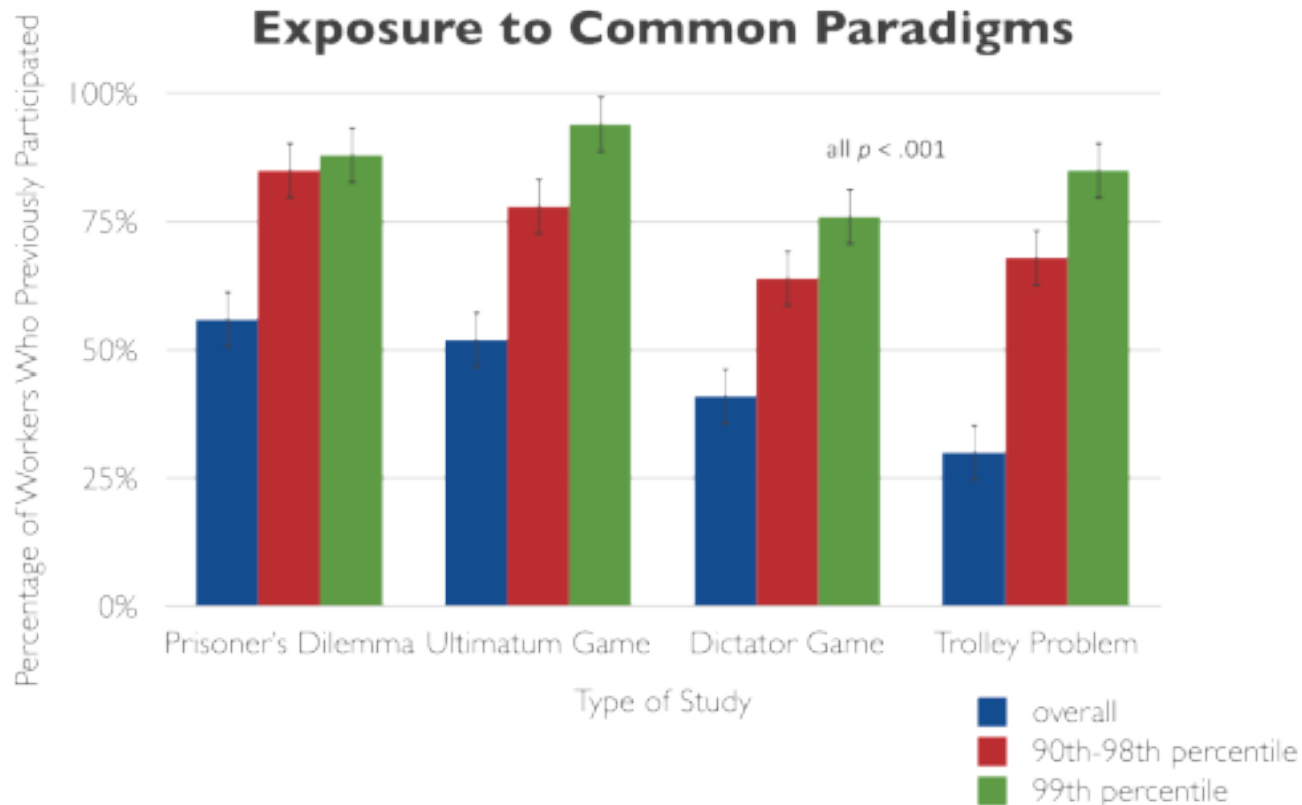
Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers

Jesse Chandler • Pam Mueller • Gabriele Paolacci

Published online: 9 July 2013
© Psychonomic Society, Inc. 2013

How good are Turk workers?

- It totally depends on your task



Getting good data

- The best studies make all manipulations within-subject or similar difficulty/annoyance to avoid strong selection biases
- Make it the same amount of time to actually answer the questions vs. skip or don't take seriously
- Limit to participants who have mostly-accepted HITs
- Don't run tasks at 11pm on Saturday night
- Make your task "fun" and short! (less than 15 min)

RTs

“The response times measured by JavaScript were approximately 25 ms longer than those measured by PTB [PsychToolBox]. However, we found no reliable difference in the variability of the distributions related to the software, and both software packages were equally sensitive to changes in the response times as a result of the experimental manipulations. We concluded that JavaScript is a suitable tool for measuring response times in behavioral research.”

Running shorter studies

- 5000 trials
- 10 subjects each doing 500 trials ('a Lab study')
- 100 subjects each doing 50 trials ('a Turk study')

- You (almost) always get more power (e.g., more likely to find a significant effect) and a more precise estimate of the effect size when doing more subjects, less trials ('a Turk study').

Things that are harder than you'd expect

- Counterbalancing across-participant experiments as opposed to randomizing. HIT vs HIT group problem. (With many conditions, posting as separate experiments get complicated and manual...)
- Disallowing users who have previously participated in a related study of yours
 - Easiest way is to assign all users who complete a study a *qualification* and then **disallow** any users with that qualification from participating in the new study. Requires serial running of different experiments.
- Running participants in multiple sessions.

Some resources

- <http://psyc241.ucsd.edu>

HTML / CSS / Javascript cheatsheets

- http://www.quackit.com/html/html_cheat_sheet.cfm
- http://www.webmonkey.com/2010/02/html_cheatsheet/
- <http://www.cheatography.com/davechild/cheat-sheets/css2/>