



Ethics of Crowdsourcing

Last Words

Amazon Mechanical Turk: Gold Mine or Coal Mine?

Karën Fort*

INIST-CNRS/LIPN

Gilles Adda**

LIMSI/CNRS

K. Bretonnel Cohen†

University of Colorado School of Medicine
and University of Colorado at Boulder

Recently heard at a tutorial in our field: “It cost me less than one hundred bucks to annotate this using Amazon Mechanical Turk!” Assertions like this are increasingly common, but we believe they should not be stated so proudly; they ignore the ethical consequences of using MTurk (Amazon Mechanical Turk) as a source of labor.

Manually annotating corpora or manually developing any other linguistic resource, such as a set of judgments about system outputs, represents such a high cost that many researchers are looking for alternative solutions to the standard approach. MTurk is becoming a popular one. However, as in any scientific endeavor involving humans, there is an unspoken ethical dimension involved in resource construction and system evaluation, and this is especially true of MTurk.

We would like here to raise some questions about the use of MTurk. To do so, we will define precisely what MTurk is and what it is not, highlighting the issues raised by the system. We hope that this will point out opportunities for our community to deliberately value ethics above cost savings.



Being A Turker

David Martin, Benjamin V. Hanrahan, Jacki O’Neill

Xerox Research Centre Europe
6 chemin de Maupertuis, Grenoble France
{david.martin, ben.hanrahan, jacki.oneill}@xrce.xerox.com

Neha Gupta

University of Nottingham
University Park NG7 2TD Nottingham
neha.gupta@xrce.xerox.com

ABSTRACT

We conducted an ethnomethodological analysis of publicly available content on Turker Nation, a general forum for Amazon Mechanical Turk (AMT) users. Using forum data we provide novel depth and detail on how the Turker Nation members operate as economic actors, working out which Requesters and jobs are worthwhile to them. We show some of the key ways Turker Nation functions as a community and also look further into Turker-Requester relationships from the Turker perspective – considering practical, emotional and moral aspects. Finally, following Star and Strauss [25] we analyse Turking as a form of invisible work. We do this to illustrate practical and ethical issues relating to working with Turkers and AMT, and to promote design directions to support Turkers and their relationships with Requesters.

Author Keywords

Ethnomethodology; content analysis; crowdsourcing; microtasking; Amazon Mechanical Turk; Turker Nation.

ACM Classification Keywords

H.5.3 Group and Organizational Interfaces – Computer, Supported Cooperative Work

General Terms

Human Factors

INTRODUCTION

The concept of crowdsourcing was originally defined by Jeff Howe of Wired Magazine as “*the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.*” [8] This ‘undefined network of people’ is the key topic of this article. We present the findings of an ethnomethodological analysis of posts and threads on a crowdsourcing forum called Turker Nation¹. We have sought to understand members of the

believe that this will be beneficial for researchers and businesses working within the crowdsourcing space.

Crowdsourcing encompasses multiple types of activity: invention, project work, creative activities, and microtasking. This latter is our focus here. The most well-known microtask platform is Amazon Mechanical Turk (AMT)², and the Turker Nation forum that we studied is dedicated to users of this platform. The basic philosophy of microtasking and AMT is to delegate tasks that are difficult for computers to do to a human workforce. This has been termed ‘artificial artificial intelligence’. Tasks like image tagging, duplicate recognition, translation, transcription, object classification, and content generation are common. ‘Requesters’ (the AMT term for people who have work to be completed) post multiple, similar jobs as Human Intelligence Tasks (HITs), which can then be taken up by registered ‘Turkers’. Turkers (termed ‘Providers’ by AMT) are the users completing the HITs, which typically take seconds or minutes paid at a few cents at a time.

For Amazon, the innovative idea was to have an efficient and cost effective way to curate and manage the quality of content on their vast databases (weeding out duplicates, vulgar content, etc.). While Amazon is still a big Requester, AMT has been deployed as a platform and connects a wide variety of Requesters with up to 500,000 Providers. However, Fort et al. [6] have performed an analysis on the available data and suggest that real number of active Turkers is between 15,059 and 42,912; and that 80% of the tasks are carried out by the 20% most active (3,011–8,582) Turkers. While these numbers are useful, the research community still has little deep qualitative knowledge about this workforce. Questions remain unanswered such as: how and what do they look for in jobs; what are their concerns; and how do they relate to requesters?

- Turkers are not working "for fun"
- Many Turkers are working because they lack other income options (e.g they were laid off)

Find HITs that pay at least \$ 0.00

HITs for
1-10 of

Sort by: Reward Amount (most first)

- HIT Creation Date (oldest first)
- HIT Creation Date (newest first)
- HITs Available (fewest first)
- HITs Available (most first)
- Reward Amount (least first)
- Reward Amount (most first)
- Expiration Date (soonest first)
- Expiration Date (latest first)
- Title (A-Z)
- Title (Z-A)
- Time Allotted (least first)
- Time Allotted (most first)

GO!

Show all details | Hide all details

Transcription of approximately 15 minutes of audio

Requester: [Amelia Jones](#) HIT Expiration Date: Sep 4, 2013 (1 day 21 hours) Reward:
Time Allotted: 24 hours HITs Available

Transcription of approximately 15 minutes of audio

Requester: [Amelia Jones](#) HIT Expiration Date: Sep 5, 2013 (2 days 23 hours) Reward:
Time Allotted: 8 hours HITs Available

By Invitation Only: Answer a few questions in a brief survey

Requester: [Qualtrics Survey](#) HIT Expiration Date: Sep 5, 2013 (2 days 21 hours) Reward:
Time Allotted: 30 minutes HITs Available

Take a geo-tagged photo of a landmark in North Dakota

Requester: [Crowdsourcing at Thomson Reuters](#) HIT Expiration Date: Sep 5, 2013 (2 days 16 hours) Reward:
Time Allotted: 3 days HITs Available

The MTurk worker interface sucks.

- No way to search by expected hourly rate
- No way to estimate the difficulty or length of time that it will take to do a task
- No way to know how reputable a requester is in advance
- Information asymmetry

Manage Batches

Click on the name of the batch to see more details

▼ **Batches in progress (1)**

[Compression HIT - grammar/meaning_10](#)

Results

Cancel th

Created:	July 04, 2013	Assignments Completed:	2,468 / 2,468
Time Elapsed:	1 day	Estimated Completion Time:	COMPLETE
Average Time per Assignment:	3 minutes 40 seconds	Effective Hourly Rate:	\$4.091
Batch Progress:			
	100% submitted	100% published	

▼ **Batches ready for review (143)**

« Previous 1 2 3 4 5 6 7 8 9 ... »

[Word Alignment - Trusted Workers - Dev 0.7](#)

Results

Delete

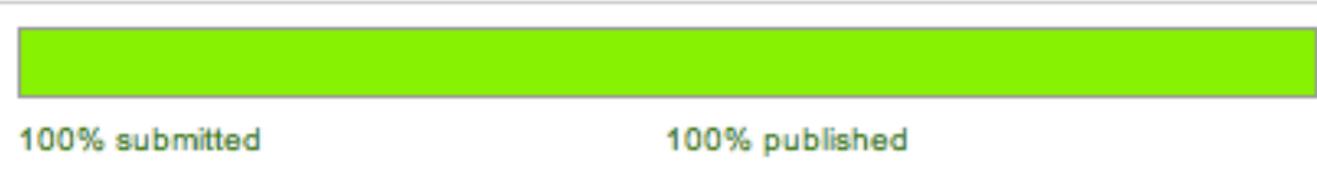
Created:	July 02, 2013	Assignments Completed:	1,995 / 2,000
Time Elapsed:	15 days	Estimated Completion Time:	September 03, 2013 4:41 PM
Average Time per Assignment:	1 minute 58 seconds	Effective Hourly Rate:	(TODAY) \$4.576

HIT - grammar/meaning 10

is batch, make changes, or get results.

ces and decide how grammatical they are

Cancel this batch



2,468 / 2,468 Average Time per Assignment: 3 minutes 40 seconds Average Hourly Rate: \$4.091
July 04, 2013 1:03 PM PDT Estimated Completion Time: COMPLETE

HIT - grammar/meaning

Read a group of sentences and decide how grammatical they are
English, writing, editing, language, research, language, paraphrasing,
paraphrases, summarization, summaries, text, comprehension, reading
HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 85
(Required for preview)
Number of HITs Approved greater than 100 (Required for preview)
Location is UNITED STATES (Required for preview)

per HIT: 2
\$0.250
compression-evaluation.dev.csv

Results

Results

Assignments pending review:	0
Assignments approved:	2,070
Assignments rejected:	398

Cost Summary

Estimated Total Reward:	\$500.000
Estimated Fees to Mechanical Turk:	\$50.000 (fee details)
Estimated Total Cost:	\$550.000

These costs are only an estimate until all of the assignments have been submitted and reviewed

Approve

Reject



<u>Worker ID</u>	<u>Work Time In Seconds</u>	<u>Lifetime Approval Rate</u>	<u>Input.System Names 0</u>	<u>City</u>	<u>Region</u>
A1H11P4QSM08Y1	145	100% (59/59)	pos_control neg_control /export/common/SCALE13/...		
A3O7D2TTK99UK5	3	12% (119/1007)	pos_control neg_control /export/common/SCALE13/...		
ASB13NJKTG CYB	122	100% (44/44)	pos_control neg_control /export/common/SCALE13/...		
A1BM57TUNAQIXM	478	100% (3/3)	pos_control neg_control /export/common/SCALE13/...	Los Angeles	California
A3445FTO62DGIX	231	100% (78/78)	pos_control neg_control /export/common/SCALE13/...		
AYSX5GNO640MZ	141	13% (5/39)	pos_control neg_control /export/common/SCALE13/...		
A21U64TBRENETA	158	100% (17/17)	pos_control neg_control /export/common/SCALE13/...	Orlando	Florida
A23AFAXSEYH61S	241	100% (177/177)	pos_control neg_control /export/common/SCALE13/...	Los Angeles	California
A1GYXHW4YCOJ0H	154	100% (77/77)	pos_control neg_control /export/common/SCALE13/...		
A3O7D2TTK99UK5	21	12% (119/1007)	pos_control neg_control /export/common/SCALE13/...		
AUGRDU EDEUXUS	184	100% (12/12)	pos_control neg_control /export/common/SCALE13/...		
AYSX5GNO640MZ	182	13% (5/39)	pos_control neg_control /export/common/SCALE13/...		

TurkOpticon plugin

Workers rate Requesters based on a Requester's:

- **communicativity:** How responsive has this requester been to communications or concerns you have raised?
- **generosity:** How well has this requester paid for the amount of time their HITs take?
- **fairness:** How fair has this requester been in approving or rejecting your work?
- **promptness:** How promptly has this requester approved your work and paid?

Choose the best category for this government project (good english important)

Requester:

 [The Public Group](#)

HIT Expiration Date: Sep 10, 2013 (6 days)

communicativity:  1.17 / 5
generosity :  1.73 / 5
fairness :  1.39 / 5
promptness :  1.86 / 5

Time Allotted: 60 minutes

[What do these scores mean?](#)

Scores based on [81 reviews](#)

Terms of Service violation flags: 0

[Report your experience with this requester »](#)

[Contact Us](#) | [Careers at Amazon](#) | [Developers](#) |

©2005-2013 Amazon.com, Inc. or its Affilia



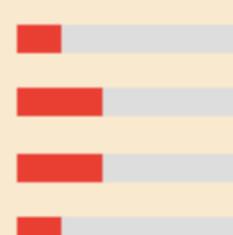
Took a leap of faith on this requester and was rewarded with a %50 reject rate and a broken search feature and no feedback. Would not recommend, even if you have thousands of HITs under your belt to cushion the inevitable rejections.

Aug 29 2013 | [KBH19](#) | [flag](#) | [comment](#)



Arbitrarily rejected over half of the hits I submitted, and then banned me from submitting any more hits for them. I suppose that's a blessing in disguise though, as I had no intention of doing any for them again after the first batch of rejections.

Aug 21 2013 | [bour...@g...](#) | [flag](#) | [comment](#)



Their HIT is very unclear. There is an option to browse for the result, but it does not work.

Aug 20 2013 | [jeff...@g...](#) | [flag](#) | [comment](#)

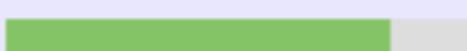
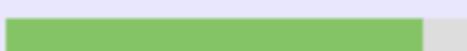
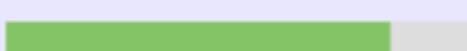
Word Alignment

Requester:

▼ [Chris Callison-Burch](#)

HIT Expiration Date: Nov 12, 2013 (9 weeks)

Time Allotted: 60 minutes

communicativity:  4.05 / 5
generosity :  4.01 / 5
fairness :  4.25 / 5
promptness :  4.01 / 5

[What do these scores mean?](#)

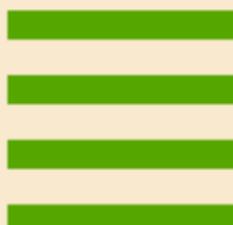
Scores based on [95 reviews](#)

Terms of Service violation flags: 1

[Report your experience with this requester »](#)

[Contact Us](#) | [Careers at Amazon](#) | [Developers](#) |

©2005-2013 Amazon.com, Inc. or its Affilia

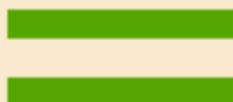
 For those of you who know Arabic, this is a very solid requester with a very fair pay. Highly recommended for those who want to make some real money Payment usually take one week depending on the HITs you are doing

don't waste your time trying to submit machine translated crap, or random answers to multiple choice questions, you will get blocked instantly

Jul 27 2013 | [hala...@h...](#) | [flag](#) | [comment](#)

 Chris is one of the better requesters on MTurk, if you meet his qualifications and actually do the work as he requires. Glad to see that someone out there can finally work on those Arabic translation HITs that we've all seen for months now.

Jul 27 2013 [baudelai...@m...](#)

 Good requester. Everything approved in a couple of days. I had no problems. This is a safe requester to work for.

Turker Nation discussion boards

- A watering hole for Turkers to discuss MTurk and Requesters
- Has a Requester Hall of Fame / Shame
- Lots of engaging conversation
- Sometime people vent their frustration

Here's an excerpt from an IRB application Chris Callison-Burch posted - "We will pay participants small sums of money to complete our tasks, ranging from \$0.01 to \$1. All participants can choose for themselves whether the compensation is fair, and opt not to do it if they deem the compensation to be too low. Amazon's Mechanical Turk has many other researchers and companies offering tasks, so we will offer compensation that is similar to what others offer."

He first refers to Amazon's Mechanical Turk as "an online labor market." And that, I agree with. It is an online labor market.

Requesters like him, and Dolores Labs, **collude**, explicitly or implicitly, **to keep wages** at a sub compatible with existence standard.

Unlike Jewels, I don't blame workers for taking low paying jobs. I can't blame a person for being needy enough to take what amounts to a crust of bread. **I blame Chris Callison-Burch, and others like him, for keeping the standard wage at crust of bread level.**

I feel like Maria in "Metropolis."

I tried one of those to see, I gave it up at 4 minutes in and about 2/3 of the way through. For the whole hit, I'd have taken about 6 minutes. 10 hits an hour - **\$1.70 an hour.** Restricted to U.S. residents.

This is far too low to be considered a fair wage for a U.S. resident. My performance may be very far off from what others can do. Perhaps I took 4 times or more as long as an average worker would.

My complaint is that any U.S. requester knows what wage rate is required for a U.S. resident to survive. We may not agree on an exact number. But as they say, I know a fair wage when I see it, and this is not it.

Mturk is actually much smaller than what it can appear to be. Something close to requester monopoly has the power to keep wages low. Requester co-operation, explicit or implicit, reinforces this.

Chris Callison-Burch is not unaware, I think, of the mechanics of the wage structure of Mturk.

**WORKERS
OF THE WORLD
UNITE!**



crowd-workers.com

- I am developing a browser plug-in that will improve the MTurk UX for Workers
- The idea is to track and aggregate statistics across many workers, so that they have better information
- Academically, I am interested in these questions
 - a) How much time does the average Turker spend working?
 - b) What is their hourly rate?
 - c) How reputable are requesters (what fraction of the HITs do they approve v. reject)?
 - d) How much time to workers spend searching v working?

qualitative v quantitative

TurkOpticon's qualitative attributes	CrowdWorker's quantitative equivalents
promptness: How promptly has this requester approved your work and paid?	Expected time to payment: On average, how much time elapses between submitting work to this Requester and receiving payment?
generosity: How well has this requester paid for the amount of time their HITs take?	Average hourly rate: What is the average hourly rate that other Turker make when they do this requester's HITs?
fairness: How fair has this requester been in approving or rejecting your work?	Approval/rejection rates: What percent of assignments does this Requester approve? What percent of first-time Workers get any work rejected?
communicativity: How responsive has this requester been to communications or concerns you have raised?	Reasons for rejection: Archive of all of the reasons for Workers being rejected or blocked by this Requester.

Discover

Here are HITs that other Crowd Workers completed. You can sort by the hourly rate averaged across all workers, and other features.

Hitgroup name, description and keywords	Requester	HITs available	Hourly rate	Reward
<p>Parents of 11- to 18-month-olds: Complete an online study with your child about intuitive probability! go to MTurk page</p> <p>For parents/guardians to complete with child aged 11 to 18 months ONLY: participate in a web-based experiment about how your child makes predictions. baby,babies,infant,children,child,parenting,kids,parents,causal,reasoning,probability,cognitive,development,learning,psychology,experiment</p>	Lookit	1	\$590.55	\$2.50
<p>Website Content Review 6 go to MTurk page</p> <p>Search "warburg pincus healthcare". List the 1st option in the top navigation bar, of the 1st search result for payment. warburg,pincus,healthcare</p>	Cam Elizabeth Harvey	1	\$29.03	\$0.03
<p>App Usage Study go to MTurk page</p> <p>Testing new apps for mobile devices over a 12-day period survey,apps,mobile</p>	Gracy P.	1	\$25.18	\$8.00
<p>Website Content Review 43 go to MTurk page</p> <p>Search 'gramercy emerging markets high yield fund'. List the title of the 2nd option in the website navigation menu of the 1st search result for</p>	Cam Elizabeth	1	\$23.79	\$0.03

Human Subjects Research

Institutional Review Board

- All research involving human subjects that is funded or regulated by the US federal government must be approved by an IRB

Research that is exempt

- Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior
- Research involving the collection or study of freely available de-identified existing data, documents, records, pathological specimens, or diagnostic specimens

Additional tests

1. Does the research involve children?

2. Identifiability & Risks

I. Will the human participants be identifiable, directly or through identifiers linked to the participants?

II. Could any disclosure of the participants' responses place the participants at risk of criminal or civil liability or be damaging to the participants' financial standing, employability, or reputation?

Is Mechanical Turk anonymous?

- Mechanical Turk already anonymizes subjects by exposing only their Worker IDs
- I know that A23KO2TP7I4KK2 completed by task, but I don't know any other info (name, gender, location)
- Great! Right?

<https://www.amazon.com/gp/pdp/profile/A23KO2TP7I4KK2>



[Recommendations](#) [Today's Deals](#) [Gift Cards](#) [Sell](#) [Help](#)

Shop by
Department ▾

Search

All ▾

Go

Hello, S
Your A

[Your Amazon.com](#)

[Your Browsing History](#)

[Recommended For You](#)

[Rate Items You Like](#)

[Improve Your Recommendations](#)

Chris Callison-Burch

REAL NAME™



Reviews

Top Reviewer Ranking: 9,737,978 ([Learn More](#)) - Total Helpful Votes: 1 of 2



Pinzon Signature Diamond
Matelasse Shower Curtain,
White by Pinzon by
Amazon.com

1 of 2 people found the following review helpful

★ ★ ★ ★ ★ **HEAVY shower curtain, J**

This shower curtain is very heavy. It p
the wall, and it seemed like it would bl
from my shower if I had managed to h
return it and just stick with a shower li

Helpful votes received on reviews: **50%** (1

[RSS feed](#) · [View Top Reviewers](#)

Application for exemption

- I have examples of applications for exemption from Johns Hopkins University and the University of Pennsylvania that I can share
- NSF is looking to be sure that anyone who includes MTurk in their applications will seek IRB approval