Some reflections on trying to be ethical on Mechanical Turk

Dan Navarro
“If you've spent time talking with Turkers, you probably know that academic requesters have been a continuous source of strain. Research surveys with horrendous pay and arbitrary rejections are common. Despite Mechanical Turk's attractive availability, a large number of researchers make innocent missteps and cause serious stress.”

- wearedynamo.org
The big issue.

Turkers are a labour force, not a participation pool
It’s worth doing some academic (and non-academic) reading about this issue, because we’re all very accustomed to working with undergrads.
“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 as a source of labour.’

- Fort et al (2011)
Pay a fair wage

• A critical distinction:
  • Do think of MTurk as a source of fast data
  • Don’t think of it as a source of cheap data

“It’s a trap!”
Pay a fair wage

• A critical distinction:
  • Do think of MTurk as a source of *fast* data
  • *Don’t* think of it as a source of *cheap* data

• But what *is* a fair wage on MTurk?
  • A US$6 / hour wage is “recommended”. (Most requesters don’t pay that well)
  • (From mid-2014) *we pay US$10 / hour in order to be above the minimum wage in every US state.*
  • We are *universities*, not some shady fly-by-night outfit
  • We don’t use the “independent contractor” loophole to evade minimum wage laws. Because that’s shitty.
Give good time estimates

• This is important for setting expectations and ensuring fair pay, but also hard to do well

• You will be faster than most Turkers
• The “average time” estimate from Amazon is not entirely reliable: the slowest participants are not working for an hour on a 10 min task.
• So you have to guess, and use common sense.
• It is better to overestimate than underestimate
Comparing Dan to a Turk sample…
A few people whose might not have been trying very hard, and whose data I’ll need to check…

me
A few people who might have been interrupted or “strategically” delaying, and whose data I’ll need to check…
In this case, I’m pretty happy with my decision to pay for 15 minutes work
Don’t reject work
(not unless you absolutely have to)

• Amazon lets you refuse to pay workers if they don’t perform to your standards
• I have never yet felt the need to do this
Don’t reject work
(not unless you absolutely have to)

• Amazon lets you refuse to pay workers if they don’t perform to your standards
• I have never yet felt the need to do this

• Why not? (the selfish part)
  • My time is valuable. Rejecting work wastes my time. It takes time to divine Turker motives, and it takes time arguing with the aggrieved Turker
  • My reputation is valuable. Even if you get it right, you risk bad reviews on Turkopticon. Why cause yourself the hassle?
Don’t reject work
(not unless you absolutely have to)

• Some more principled reasons…
  • If you’re wrong, you just cheated a low income worker out of a small amount of money. That’s not great behaviour from a University.
  • (And you really could be wrong: your instructions aren’t as clear as you think they are.)
  • You’re contributing to worker insecurity. If Universities are going to be active on MTurk, we want to send a message that we are reliable employers.
Pay up quickly

• The usual advice:
  • Check your “auto approve” settings on MTurk to make sure that everyone is paid in a week.
  • But honestly… I manually approve everything on the same day. It’s one minute of work.
Pay up quickly

• The usual advice:
  • Check your “auto approve” settings on MTurk to make sure that everyone is paid in a week.
  • But honestly… I manually approve everything on the same day. It’s one minute of work.

• The key thing:
  • Fast approval is appreciated by Turkers: it removes uncertainty about whether the pay is coming.
  • You don’t gain anything by delaying, and you cause anxiety if you do. So pay up on the same day.
Communicate quickly

• Turkers often communicate via email
  • Your MTurk account is linked to an email address.
  • Monitor it. Respond within hours, not days.
Communicate quickly

- Turkers often communicate via email
  - Your MTurk account is linked to an email address.
  - Monitor it. Respond within hours, not days.

- Really?
  - Academics are terrible with timely responses to email
  - Yes, we are busy. But if you just got 300 subjects in 20 mins, the least you can do is keep an eye on your email account for the next few days.
  - Failing to do this is seen as disrespectful. It will earn you a bad reputation among Turkers
Reputation is visible.

It’s mediated by “Turkopticon” and it’s what protects Turkers from jerks.
Amazon didn’t set up MTurk with any mechanisms for workers to identify bad requesters.
Amazon didn’t set up MTurk with any mechanisms for workers to identify bad requesters.

Researchers at UCSD stepped in to fix this.

Turkopticon is a website / browser plugin that allows Turkers to rate requesters.

You can sign up as a requester.
Amazon didn’t set up MTurk with any mechanisms for workers to identify bad requesters.

Researchers at UCSD stepped in to fix this. Turkopticon is a website / browser plugin that allows Turkers to rate requesters.

You can sign up as a requester

Your average rating is a *very visible* reputation

Turkers *will* assess you based on this
Here’s what I mean… any Turkker with the TO browser extension can see your reputation when they’re looking for work...

(this is a very good thing)
This is us.

We have a pretty good reputation (we had to earn it though!)
A random screenshot of “recently reviewed requesters”

<table>
<thead>
<tr>
<th>Name</th>
<th>FAIR</th>
<th>COMM</th>
<th>PAY</th>
<th>FAST</th>
<th>HIT Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Francis Smart</td>
<td>5.00</td>
<td>2.00</td>
<td>1.06</td>
<td>5.00</td>
<td>22</td>
</tr>
<tr>
<td>Adobe Innovator</td>
<td>no data</td>
<td>no data</td>
<td>no data</td>
<td>no data</td>
<td>2</td>
</tr>
<tr>
<td>Anna E Jurgensen</td>
<td>3.44</td>
<td>4.33</td>
<td>4.12</td>
<td>4.25</td>
<td>10</td>
</tr>
<tr>
<td>Lindsay Nielsen</td>
<td>1.00</td>
<td>no data</td>
<td>no data</td>
<td>no data</td>
<td>13</td>
</tr>
<tr>
<td>Bernard Ghanem</td>
<td>4.50</td>
<td>1.00</td>
<td>2.00</td>
<td>4.38</td>
<td>24</td>
</tr>
<tr>
<td>Laboratory for</td>
<td>4.15</td>
<td>2.00</td>
<td>4.00</td>
<td>4.32</td>
<td>27</td>
</tr>
<tr>
<td>experimental economics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Researcher007</td>
<td>4.05</td>
<td>2.86</td>
<td>3.68</td>
<td>4.18</td>
<td>49</td>
</tr>
</tbody>
</table>
Communication on Turkopticon

- If people post complaints on Turkopticon…
  - Respond politely. Don’t get angry, no matter what.
  - Preface comments with “I am the requester”.
A Turker making a (legitimate) complaint about one of my experiments
Another Turker commenting that it worked for them
Me apologising after working out that there really was a bug...
All this happened on the same day.

There is a part in the survey where you're supposed to check your understanding, and I wasted 15 minutes trying to read the instructions and I think it's glitched. If one of you is smarter than me and passes this then this is for the others as a warning that I believe it's either a glitch or really difficult to understand the instructions for this HIT. I had to return it after wasting a lot of time figuring it out, that is why the rating is negative.

Jan 07 2014 | flag | comment

I just completed this survey game at 20:17 tonight without any issues. At one point after the instructions were posted the instructions reloaded and added one additional block at the bottom to click to the next screen. The game took about 4 minutes to complete. I have not been paid for the survey nor contacted the requester. Will post later if I encounter any issues.

Jan 07 2014

Paid $.60 two hours after completion.

Jan 07 2014

I am the requester for this HIT. I've looked over the code and there does appear to have been a glitch with the instructions that would have affected a small proportion of users. It looks like the problem didn't occur for many people, but it has now been fixed and shouldn't appear in future HITs. My apologies for those people who were affected.

Jan 07 2014 Adelaide CompCogSci 2 | edit
Communication on Turkopticon

• If people post complaints on Turkopticon…
  • Respond politely. Don’t get angry, no matter what.
  • Preface comments with “I am the requester”.

• If people post unfair complaints on Turkopticon
  • Respond politely. Don’t get angry, no matter what.
  • You can “flag” a comment, and the moderators will address the issue if need be.
Hey! That’s not a fair complaint. This wasn’t us!
I noticed it a few weeks later, flagged it and left a polite comment.
A moderator arrives the next day and asks the Turker to check.
Moderator waits 24 hours and then agrees that the review was misplaced.
This review is still accessible, but it no longer contributes to our reputation.

The review below has been flagged as inappropriate. It does not affect the requester's rating.

Adelaide CompCogSci
A2UKL52R7I6BX7
Averages »
HIT Group »
Review Requester »

Slave wages. 20 minutes for .65.

It asks about your worst fear. It is a not a "fun" survey.

Jan 22 2015 unflag comment

not a review for this requester/irrelevant
Feb 18 2015 Adelaide CompCogSci 2
agree w/ flagger
Feb 20 2015 rubyr...@m... (moderator)

I am the requester for Adelaide CompCogSci. I think you may have confused us with someone else. We have never posted a HIT asking anyone about their worst fear. Nor do we pay such low wages. We did have a HIT posted on Jan 22, but that was a categorization task that paid $2.50 for a 15 to 20 min task.

Feb 18 2015 Adelaide CompCogSci 2 edit

Hello! Can you review and update your information regarding this untitled 20 minute hit you referenced? :) Hello Requester, I'm allowing the reviewer time to see if he/she will review their hits/queue/notes/memory for accuracy. It is entirely possible a mix-up occurred as well as possible they will not return and update this review. In review, thanks for your patience! :)

Feb 19 2015 rubyr...@m... (moderator)
Turkopticon is important and an ethical requester will respect it

- *It is important to Turkers*
  - Turkers use TO to avoid unethical requesters
  - TO exists to protect *Turkers*. It’s not there for us.
  - (My opinion) TO is *necessary* to make MTurk ethical

- What that means
  - Even if something doesn’t go your way, you need to respect the TO system and TO moderators
  - I think that’s more than fair!
Brief comments on Amazon policies that you’re required to follow
Obvious Amazon policy violations

• Things you *obviously* can’t do…

• Anything illegal
• Disrupt a website or service
• Creating a security risk online
• Create a marketing scam
• Violate copyright laws
• etc.

“don’t be evil”
Non-obvious Amazon policy violations

• Not so obviously, you can’t…
  • Ask Turkers to register on another website
  • Ask Turkers to download software
  • Ask Turkers to contact a third party

• … much of this is (probably) self-serving by Amazon, but it’s their platform
Some practical things that Turkers consider best practice for academic requesters

(and as such, is arguably part of ethical practice)
Identify and explain!

• Quickly and clearly identify...
  • The name of the academic responsible for the study
  • The department & university
  • Ethics committee contact details (but you can ask them to contact you first!)
  • Be very clear about what the study entails

• Real life example...
  • Researchers working on spam classifiers failed to properly identify. Turkers thought the study was a scam.
The name our lab uses on MTurk appears prominently the moment the Turker arrives at our site.
Thanks for accepting the HIT. "How many species?" is a short psychological study investigating how people make guesses on the basis of limited information. It involves the following steps:

That’s the name of the HIT on MTurk & a brief description
Thanks for accepting the HIT. "How many species?" is a short psychological study investigating how people make guesses on the basis of limited information. It involves the following steps:

1. We ask for demographic information (not connected to your Amazon ID)
2. The study then explains the guessing game. You will need to pass a short test to check that you understand how the study works.
3. Next comes the experiment itself.
4. At the end, we'll give you the completion code you need to get paid for the HIT.

This is what the study entails, and this is how the Turker gets paid
Adelaide Computational Cognitive Science

Thanks for accepting the HIT. "How many species?" is a short psychological study investigating how people make guesses on the basis of limited information. It involves the following steps:

1. We ask for demographic information (not connected to your Amazon ID)
2. The study then explains the guessing game. You will need to pass a short test to check that you understand how the study works.
3. Next comes the experiment itself.
4. At the end, we'll give you the completion code you need to get paid for the HIT.

Total time taken should be about 15 minutes. Please don't use the "back" button on your browser or close the window until you reach the end and receive your completion code. This is very likely to break the experiment and may make it difficult for you to get paid. However, if something does go wrong, please contact us! When you're ready to begin, click on the "start" button below.

That’s how long the task takes
Details of the study: The principal investigator is Dr Daniel Navarro (daniel.navarro@adelaide.edu.au). For any questions regarding the ethics of the study, please contact the convenor of the Subcommittee for Human Research in the School of Psychology at the University of Adelaide, Dr Paul Delfabbro (+61)08 8313 4936. Please direct any questions about this study to Dr Navarro. Although any data gained from this study may be published, you will not be identified and your personal details will not be divulged, nor will anything be linked to your Amazon ID. You may withdraw at any time, although you will not be paid unless you complete the study.
Respect privacy

- This is no different from regular research
  - Don’t break Turker anonymity
  - Don’t ask for an email, real name, Facebook login, etc
  - Don’t publish anything that could be used to re-identify your participants
  - In addition to being unethical, these things violate the Amazon Terms of Service.
Fix your mistakes

- We all screw up sometimes.
- Even so…
  - If your experiment breaks, pay people anyway
  - We’ve had cases where we had to offer a replacement HIT to pay people who weren’t able to do the task
  - It’s a lot of work, but it’s part of paying your workers.
  - If you hate it, get the experiment right first time!
This is sooo embarrassing… we forgot to pay our bills and the site stopped working mid-experiment…

… the Turkers were not happy. It took a huge amount of time and effort to track everyone down and pay them for their wasted time
But look… A happy ending!

Turkers can be a forgiving people if you treat them with the respect they deserve
Some grey areas…
Adult content?

• Possible, but care is needed:
  • Amazon does allow you to present adult content (e.g., sexual images, violent images, etc)
  • You are required to warn Turkers ahead of time
  • You are required to use the “Adult Content Qualification” in the MTurk interface
  • You should offer higher pay for any task that might cause discomfort

• My suggestion: Talk to your ethics committee about what you have in mind, and be careful.
Clinical populations?

There is some literature on this, and it seems like MTurk is convenient for this purpose. But...
Clinical populations?

• While Amazon has nothing to say here, your ethics committee very likely has an opinion on this
Clinical populations?

- While Amazon has nothing to say here, your ethics committee very likely has an opinion on this.

- The big problem is *duty of care*:
  - Suppose I learn that TurkerX is feeling suicidal.
  - Now that I know, I have a duty of care obligation.
  - But I don’t know where they live, and I don’t know what services are available.
  - I don’t know *how* to meet my duty of care obligation.
Deception or breaching?

- Not technically a violation of Amazon policies
- But I think it’s a bad idea
Deception or breaching?

- Not technically a violation of Amazon policies
- But I think it’s a bad idea

Why?

- **Sociologically**: Turkers are not students. They have no reason to forgive a deception “because science!” The breach of trust is bigger in the MTurk context … because you are lying to your workers.
- **Practically**: you will end up with a bad reputation
- **Ethically**: Best not to go there.
On MTurk itself

• Is it ethical to be part of a not-ideal system?
On MTurk itself

- Is it ethical to be part of a not-ideal system?

- MTurk isn’t an ideal labour market
  - Legally, Turkers are “independent contractors” not “employees”, so minimum wage laws don’t apply
  - Some requesters pay very poorly, and are deceptive
  - Turkopticon protects Turkers, but that’s a project sponsored by UCSD, not Amazon
On MTurk itself

• Is it ethical to be part of a not-ideal system?

• MTurk isn’t an ideal labour market
  • Legally, Turkers are “independent contractors” not “employees”, so minimum wage laws don’t apply
  • Some requesters pay very poorly, and are deceptive
  • Turkopticon protects Turkers, but that’s a project sponsored by UCSD, not Amazon

• My opinion…
  • Online market places aren’t going away
  • We should be there, and be a positive influence
  • (But reasonable people could disagree)
An example of what happens when you don’t respect the system...
“If you've spent time talking with Turkers, you probably know that academic requesters have been a continuous source of strain. Research surveys with horrendous pay and arbitrary rejections are common. Despite Mechanical Turk's attractive availability, a large number of researchers make innocent missteps and cause serious stress. Recently, the tension came to a head on Turkopticon. An IRB-approved researcher experimented on the platform unannounced. The result was Turker confusion, strife, and wasted time, in a system where time is what it takes to make ends meet.”

- wearedynamo.org
A study on the efficacy of reputation

- **Study 1**
  - Had people sign up to MTurk to do HITs
  - Requesters with good Turkopticon ratings were more likely to approve the work

- **Study 2**
  - Created fake requesters on MTurk and fake workers to give them “controlled” reputation on Turkopticon
  - Turkers much less likely to work for a bad requester
A study on the efficacy of reputation

- **Study 1**
  - Had people sign up to MTurk to do HITs
  - Requesters with good Turkopticon ratings were more likely to approve the work

- **Study 2**
  - Created fake requesters on MTurk and fake workers to give them “controlled” reputation on Turkopticon
  - Turkers much less likely to work for a bad requester

Wait… WTF?! A university researcher on MTurk deliberately manipulated Turkopticon? Why?
“Why didn't I take ... just go to the forum and talk to people about the research idea up-front? The process felt too ill-defined and unpredictable. I imagined a wild and unwieldy discussion that would completely compromise the possibility of doing the experiment.

https://groups.google.com/forum/#!topic/turkopticon-discuss/LJWrucL7Yw%5B1-25%5D
“Why didn't I take … just go to the forum and talk to people about the research idea up-front? The process felt too ill-defined and unpredictable. I imagined a wild and unwieldy discussion that would completely compromise the possibility of doing the experiment.

If TO could establish a convenient, well-defined process for researchers to use (fill out this form, submit your IRB application and approval letter), which assures researchers some respect for the sensitivity of their proposed designs, and also put us on notice that failure to go through the process could lead to some collectively-enforced consequences down the road, that might stimulate more up-front engagement.”

- Aaron Sojourner

https://groups.google.com/forum/#!topic/turkopticon-discuss/LjWrucI7Jyw%5B1-25%5D
“Workers are not lab rats, numbers or statistics for you to manipulate as you please. We are people. Many of whom are desperately trying to earn money for our families. They rely on [Turkopticon] to guide them through this mess that Amazon created. … The proper way to perform your research would have been to do it like every other academic who studies MTurk, make a survey hit about Turkopticon and keep your hands OFF [Turkopticon].”

- taintturfk (A Turker)
“Workers are not lab rats, numbers or statistics for you to manipulate as you please. We are people. Many of whom are desperately trying to earn money for our families. They rely on Turkopticon to guide them through this mess that Amazon created. The proper way to perform your research would have been to do it like every other academic who studies MTurk, make a survey hit about Turkopticon and keep your hands OFF [Turkopticon].”

- taintturk (A Turker)
Wrap up
My view…

• I think you can do ethical research on MTurk
• I don’t even think it’s that hard

• The things I try to keep in mind…
  • Follow all the usual rules of ethical work
  • Remember that Turkers are workers not students
  • Respect the Turkopticon reputation system
  • Remember that duty of care obligations are trickier
A useful resource: wiki.wearedynamo.org

Links to other resources on AMT and online research ethics

Contents [hide]
1 Other guidelines as resources
   1.1 AMT Requester best practices: academic
   1.2 Internet Research and Social Computing Ethics
   1.3 AMT Requester best practices: general

Other guidelines as resources

AMT Requester best practices: academic
- 'A Step-By-Step Guide to MTurk for the Academic Researcher' by requester EE Buckels at 'Mechanical Turk Wiki for Behavioral Research'
- 'Amazon Mechanical Turk Guide for Social Scientists' by requester Michael Buhrmester
- 'Experimenting on Mechanical Turk: 5 How Tos' blog post by Markus Jakobsson at PARC

Internet Research and Social Computing Ethics
- Code of Ethics for Community Informatics Researchers
- Stanford Encyclopedia of Philosophy - Internet Ethics
- University of Connecticut Internet Research Ethics
- Association for Internet Research Ethics Guidelines
- Reardon, "On the Emergence of Science and Justice" in Science, Technology & Human Values

AMT Requester best practices: general
- 'Requester Best Practices Guide' (pdf) from Amazon Web Services
- 'Being a Respected Requester' by turker 'Mills' at mTurk Boards
- 'Tips for Academic Requesters on Mtoutk' blog post by turker 'T' at 'Tips For Requesters On Mechanical Turk' blog
- 'ProPublica's Guide to Mechanical Turk' by requesters Srinivas Rao and Amanda Michel at ProPublica
- '15 Tips to Get the Most out of Mechanical Turk' blog post by requester AJ Kohn
- 'Making the Most of Mechanical Turk: Tips and Best Practices' blog post by requester Edwin Chen
- 'Effective Use of Amazon Mechanical Turk (MTurk)' by requester Neeraj Kumar
Anyway… I think this is worth talking about as a community (or just in the coffee break…)