ADVANCED USES OF MECHANICAL TURK IN PSYCHOLOGICAL RESEARCH

Pam Mueller
Jesse Chandler
Gabriele Paolacci

Society for Personality and Social Psychology - January 28, 2012
THE PLAN

• Part 1
  • Issues people aren’t considering
    • Crosstalk (Edlund, 2009)
    • Repeat players
• Part 2
  • Intro to advanced features
    • Command Line Tools
    • API
• More information: http://experimentalturk.wordpress.com
CROSSTALK AND REPEAT PLAYERS

- Important assumption: observations are independent
- But 1) workers share information:
  - www.turkernation.com
  - Talk about surveys and requesters; link to surveys
- And 2) the MTurk pool stays the same
- Easy to limit workers to one response within a HIT
- Across a program of research....
  - replicating finding on 3 different samples
  - replicating finding on one sample 3 times
[EDIT] You need to have a car. I no longer do so it DQed me but still gave me a code

[EDIT] You need to have a car.
One gotcha on one page says, Don’t answer the question at all. Be careful with that one because once you tick one by accident it cannot be undone.

Originally Posted by blue

I haven’t come across very many what I would call “clever” “gotcha”s, any examples? Most of the ones I see are hilariously hamfisted!

paragraphs or texts to read in the details of the survey and answered among a sentence in the second paragraph is a 3-4 word direction (“mark question yes” or “skip this question completely”). Others are getting much more specific about recalling less important details of scenarios completed earlier parts of the survey. I had one little 10 cent survey built around psychology issues yesterday that started with a seemingly irrelevant passage about types of fish to “refocus your thoughts.” There was a getcha question about what fish was missing from the list at the end. I don’t know if getting that question wrong was enough to get you kicked out, but it would make you worry. (That survey is no longer on the board so I don’t feel bad exposing the trick.)

Now maybe they can go to work on that avanused IQ test - hat/ball for $1.10 or 5 Widgets/5 days/$5 dollar.
It makes me wonder if some of the folks posting surveys are somehow unaware that turkers take hundreds of surveys. If they're unaware, they don't know what they're doing.

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

These questions can only work to measure what they're supposed to measure the very first time a survey-taker sees them. After that, the cats out of the bag. Why would a serious researcher put that sort of content into a survey that goes to a pool of professional survey takers? It's a tainted pool. Do they not notice that MTurkers get those questions right way more than the general population? Again, I suspect many of these researchers simply don't care. They're either students, or they're just ripping off whichever gave them the money for the research.

I once saw that set of questions copied into a survey with a serious typo in the question about the bat and the ball. I knew what the question and the answer were supposed to be, yet I couldn't arrive at the right answer because of the typo. It must've been some student who copied the questions out of a textbook too hastily, I thought. Or perhaps it was something more devious. Look at all the folks who got the answer right even though the question was wrong. Perhaps the real purpose of the research was to show that MTurk is a bad place to test surveys.
Do they think that ANYONE doesn't know how much that damn ball costs by now??
CROSSTALK
COMMON MEASURES

It makes me wonder if some of the folks posting surveys are somehow unaware that
Turkers take hundreds of surveys. If they're unaware, they don't know what they're
doing.

If they are aware, why do I see so many surveys cutting-and-pasting questions like, as
an example, these:

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

These questions can only work to measure what they're supposed to measure the very
first time a survey-taker sees them. After that, the cat's out of the bag. Why would a
serious researcher put that sort of content into a survey that goes to a pool of
professional survey takers? It's a tainted pool. Do they not notice that MTurkers get

I once saw that set of questions copied into a survey with a serious typo in the
question about the bat and the ball. I knew what the question and the answer were
supposed to be, yet I couldn't arrive at the right answer because of the typo. It
must've been some student who copied the questions out of a textbook too hastily, I
thought. Or perhaps it was something more devious. Look at all the folks who got the
answer right even though the question was wrong. Perhaps the real purpose of the
research was to show that MTurk is a bad place to post surveys.
Tell us about the time someone held power over you?

Anyone else sick of this damn question? I need to write something down and just cut and paste it from now on.

Has one of those power essay thingies so those with their ready to go essays...here you go 😊
REPEAT PLAYERS

Enough with the Train Problem
Researchers need to come up with a new scenario. I am so sick of answering the train problem. It wasn't so bad until my professor last week decided we should spend a whole class period discussing our opinions on this ethical dilemma. I am so over this

P.S. Thank you for the money though. 😊

Are there any survey questions you are sick of answering?

Being semi-new to surveys (been doing them around 5 weeks now), I had no idea what that problem was being referred to.

Until the past two days, when I ran across that same train question in three different surveys.

... others... do you feel groups should be equal... blah blah...

They must get their survey questions all out of the same textbook.
REPEAT PLAYERS

Prevalence of Paradigms on MTurk

N = 320
Collected with new Requester account

<table>
<thead>
<tr>
<th>Type of Study</th>
<th>Percentage of Workers Who Previously Participated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prisoner's Dilemma</td>
<td>50%</td>
</tr>
<tr>
<td>Ultimatum Game</td>
<td>50%</td>
</tr>
<tr>
<td>Dictator Game</td>
<td>25%</td>
</tr>
<tr>
<td>Trolley Problem</td>
<td>25%</td>
</tr>
<tr>
<td>p-Beauty Contest</td>
<td>0%</td>
</tr>
</tbody>
</table>

N = 320
Collected with new Requester account
REPEAT PLAYERS

• What is the probability *any MTurk worker* has done a similar HIT before?
• What is the probability a *worker in my HIT* has done something similar before?
• The authors of this paper (and collaborators) pooled their past HITs:
  • 132 batches (studies)
  • 16,408 HITs (individual surveys taken)
  • 7498 unique workers
  • Average worker = 2.24 HITs
  • Long tail of workers who are far above average
SUPERTURKERS

MTurk Worker Activity Demographics

Number of workers completing this many HITs

<table>
<thead>
<tr>
<th>HITs completed by a single worker</th>
<th>Number of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4443</td>
</tr>
<tr>
<td>2</td>
<td>1297</td>
</tr>
<tr>
<td>3</td>
<td>568</td>
</tr>
<tr>
<td>4</td>
<td>306</td>
</tr>
<tr>
<td>5-6</td>
<td>337</td>
</tr>
<tr>
<td>7-8</td>
<td>151</td>
</tr>
<tr>
<td>9-10</td>
<td>87</td>
</tr>
<tr>
<td>11-14</td>
<td>88</td>
</tr>
<tr>
<td>&gt;15</td>
<td>73</td>
</tr>
</tbody>
</table>
Common Paradigms on MTurk

- Prisoner's Dilemma
- Ultimatum Game
- Dictator Game
- Trolley Problem

Percentage of Workers Who Previously Participated

- overall
- 90th-98th percentile
- 99th percentile

all $p < .001$
WHAT TO DO?

- Know that participants do engage in crosstalk
  - Ask them how they found your study (searching MTurk, “cool surveys thread,” etc.)
- Be careful using standard paradigms and DVs
  - could include a DV about whether they’ve done something similar before
- If you do develop your own paradigm:
  - without some action on your part, there will still be repeaters if you post related studies
Get E-mail Alerts When Your Favorite Requesters Post HITs.

If you've used Amazon's Mechanical Turk for any length of time, you've probably wished that you could sign-up to receive e-mail alerts from your favorite requesters when they post new HITs.

With Turk Alert, now you can.

www.turkalert.com
PREVENTING DUPLICATE WORKERS

• Do not use the “Block” feature to avoid duplicates!
  • [http://turkopticon.differenceengines.com](http://turkopticon.differenceengines.com)
• Tell people not to do your HIT again
• Lists of worker IDs
  • requires compliance check or coding
• Easy method (if you’re linking to an outside site)
  • Link to an intermediate page (consent form, etc)
  • Have the link at the bottom of that page go to the study
  • Change that link when you put up new versions/related studies
  • Add HITs and time to the batch as needed
  • Problem: can’t change the description/payment structure
• More complex: use qualifications
BETTER DATA - BEYOND THE WEB INTERFACE

• Command Line Tools
  • bonuses
  • qualifications
• Amazon Web Services API
  • messaging workers in groups (up to 100 at a time)
### Assignments You Approved (351)

<table>
<thead>
<tr>
<th>Worker ID</th>
<th>Result</th>
<th>Submission Date</th>
<th>Approval Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZT5QCOXKPUY</td>
<td>obscureHT'</td>
<td>Nov 15, 2011, 01:41 PM PST</td>
<td>Nov 16, 2011, 01:45 PM PST</td>
</tr>
<tr>
<td>A2NOGROXQ5RT</td>
<td>$0.30 Send an e-mail message</td>
<td>Nov 15, 2011, 01:44 PM PST</td>
<td>Nov 16, 2011, 01:45 PM PST</td>
</tr>
<tr>
<td>A1P6AAABZLAZJ</td>
<td>$0.30 paybugprow123</td>
<td>Nov 15, 2011, 01:44 PM PST</td>
<td>Nov 16, 2011, 01:45 PM PST</td>
</tr>
<tr>
<td>AZL4FPUXJN1YK</td>
<td>$0.30 NiaArt527</td>
<td>Nov 15, 2011, 01:56 PM PST</td>
<td>Nov 16, 2011, 02:00 PM PST</td>
</tr>
<tr>
<td>AVEEDN4Y5289SM</td>
<td>$0.30 careay593</td>
<td>Nov 15, 2011, 02:07 PM PST</td>
<td>Nov 16, 2011, 02:10 PM PST</td>
</tr>
</tbody>
</table>
QUALIFICATIONS

• Values (unique to a given requester) that can be assigned to workers

• Generally used to include workers (e.g. to independently prescreen or for a longitudinal study)

• Can also be used to exclude (though MTurk doesn’t make it simple)
### INCluding Workers

#### Manage Qualification Types

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

<table>
<thead>
<tr>
<th>Qualification Type</th>
<th>Workers who have this Qualification</th>
<th>Creation Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>
INCLUDING WORKERS

• Give workers who complete the first HIT a qualification value (TurkerTest=5)
• Only the previously-qualified workers will be able to access and complete the follow-up
EXCLUDING WORKERS

• Must use the Command Line Tools
  • Create a qualification that is autogranted (i.e. SuperTurker).
  • Assign that qualification--but a different value--to your past workers (SuperTurker=2)
    • The autogrant will not override your assigned value
  • To successfully attract workers, you MUST:
    • allow the HIT to be searchable to those without the qualification
    • state in the HIT description that the qualification will be immediately automatically granted upon request to anyone who has not already participated in a related study
EMAILING WORKERS

- Amazon Web Services API
  - for those with no coding experience, download boto
    http://code.google.com/p/boto/
- qualifications+messages = longitudinal study
  - One year later
    - 3 emails
    - 44% response rate
    - Workers known to have completed more than one HIT:
      - Retention rate jumps to 59%
      - Frequent workers more likely to respond to follow-up
EMAILING WORKERS

```python
$ python
>>> import boto
>>> access_key=
>>> secret_access_key=
>>> c = boto.connect_s3(AccessKey=access_key, SecretKey=secret_access_key)
>>> c.get_account_balance()

>>> subject="test message"
>>> worker_ids=["A0QFVK4NW8ZC", "A2KMD6Q9TH69", "A2XG4KXNG3M0", "A27J2756GC78", "A395P3Q6H34J"]
>>> for worker_id in worker_ids:
...     c.notify_worker(worker_id, subject, "test message")
```
CONCLUSIONS

• MTurk is a valuable resource
• Researchers should strive to add value
  • Don’t treat the subject pool as a naive sample
  • Use creative paradigms (cf. Ellsworth & Gonzalez, 2003)
• Use MTurk’s advanced features:
  • variable incentivizing,
  • longitudinal studies,
  • independent prescreens,
  • preventing repeat players across programmatic research
THANK YOU!

http://experimentalturk.wordpress.com