A Penny for Your Thoughts:
Survey research opportunities presented by Amazon Mechanical Turk

November 11, 2017
Jesse Chandler

What is Mechanical Turk?

Mechanical Turk is a Platform

Advantages of MTurk
• Simple. Security, recruitment, identity verification and payment handled by Amazon
• Fast. Hundreds of responses per day
• Elastic. Scales up and down quickly
• Cost Effective. ~$0.14 per respondent

Who are Mechanical Turk Workers?
• Historically mostly Indian and American
• Most research relies heavily on American workers
• 500,000 registered users
• The typical lab will struggle to reach more then 15,000 workers in any quarter

Stewart, Ungemach, Harris, Barbels, Newell, Paolacci & Chandler, 2017
Diverse but Not Representative

<table>
<thead>
<tr>
<th></th>
<th>USA</th>
<th>Mechanical Turk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Size</td>
<td>323m</td>
<td>500k (15k active)</td>
</tr>
<tr>
<td>Age</td>
<td>47.1</td>
<td>33.5</td>
</tr>
<tr>
<td>White</td>
<td>74%</td>
<td>83%</td>
</tr>
<tr>
<td>4 year degree</td>
<td>19%</td>
<td>35%</td>
</tr>
<tr>
<td>Republican</td>
<td>29%</td>
<td>18%</td>
</tr>
<tr>
<td>Democrat</td>
<td>35%</td>
<td>41%</td>
</tr>
<tr>
<td>LG(B)</td>
<td>1.7% (1.8%)</td>
<td>3.8% (6.9%)</td>
</tr>
<tr>
<td>Atheist</td>
<td>3%</td>
<td>21%</td>
</tr>
<tr>
<td>Has children</td>
<td>~54%</td>
<td>~30%</td>
</tr>
<tr>
<td>Working age with disability</td>
<td>11%</td>
<td>~5%</td>
</tr>
</tbody>
</table>

Casey, Chandler, Levine, Strolovich & Procter, 2017

Important Psychological Differences

- Personality
  - More Introverted
  - More Neurotic
  - More Open to New Experiences
  - Less Conscientious
- Score higher on clinical measures of
  - Autism spectrum traits
  - Social anxiety

Casey et al., 2017; Chandler & Shapiro, 2016

How Mechanical Turk Works

The Worker’s Perspective

Describe Your Task and Sample Size

Define Your Sample
Design Your HIT

Using MTurk in Survey Research

Pilot Testing and Cognitive Interviewing

• Widely used to test elements of surveys before fielding them on a larger population
  – BLS (Edgar, Murphy & Keating, 2016)
  – Census (Childs, Fobia, Holzberg, Morales, 2016)
  – National Cancer Institute (Fowler, Willis, Moser, Ferrer & Berrigan, 2015)
• Both probe questions and think-aloud approaches are supported

Workers as Research Assistants

Item/stimulus generation
  Content coding and judgment
  • Annotation of text in forums (MacLean & Heer, 2013; Vlahovic et al., 2014)
  • Speech pathology ratings (McAllister et al., 2014)
Data collection
  • Upload pictures of thermostats (Meier et al., 2011)
  • Upload letters about standardized testing (Chandler, unpublished data)

An Illustration from Political Science

• Accurate:
  – 15 workers as good as 5 experts
  – Worker and expert ratings, \( r = .96 \)
• Fast: 22,000 statements in 5 hours
• Cost Effective: Total cost of $1080
• DIY: Anybody can replicate it

Mechanical Turk vs. Student Samples

Sources: Klein et al., 2014
Eberzola et al., 2016
Limitations of MTurk Samples

- MTurk is not a probability sample of the US population
- Samples drawn from MTurk are not probability samples of the worker population (Casey et al., 2017)
- There is no mechanism to retire “professional” survey respondents (Chandler et al., 2014; 2015)
- Workers may share information about studies with each other (Yin et al., 2016)
Prescreening Fraud

• Prescreening fraud is lying to gain access to a survey
• Fraudulent responses may be more common than we assume
• Higher incentives increase fraud
• Fraud prevention measures are less than perfect
• Even low fraud rates are bad but the problem is manageable

Fraud is rare but impactful

“Unobtrusive” Prescreening

• Participants indicated their sex at the beginning of a survey
• Told that they were ineligible based on their response and terminated
• Paid either $0.25 or $1.00 for their participation

Higher incentives increase fraud
Preventing Fraud with Technology

- Participants were asked to select the electronic devices that they owned from a list
- Screened out if they did not own an Oculus Rift
- Random assignment:
  - Half of participants could simply reattempt the survey
  - Half were restricted from reattempting using Qualtrics’ “prevent ballot stuffing” feature

Fraud prevention is imperfect: ballot stuffing

<table>
<thead>
<tr>
<th>Percentage of fraudulent respondents</th>
<th>Percentage of fraudulent OR owners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allow Stuffing (N = 212)</td>
<td>Prevent Stuffing (N = 433)</td>
</tr>
<tr>
<td>Allow Stuffing (N = 31)</td>
<td></td>
</tr>
</tbody>
</table>

Fraud prevention is imperfect: learning

| Percentage of True Respondents |
| N +11% |
| N +34% |

The potential impact of fraud

| Percentage of True Respondents |
| N +34% |

Actual Impacts of Fraud

- People respond to subsequent questions truthfully, adding noise to any measurements
- People respond to subsequent questions using a lay theory about how they “should” respond

Fraud Beyond MTurk

- Fraud is an issue in any context where prescreening relies on respondent self-report
- Market research: 13% of Americans claim to own Segways (Downes-Le Guin et al., 2006)
- Medical research: 25% of respondents admit to faking or concealing medical symptoms (Devine et al., 2013)


## Opportunities for Survey Research

- Supplement for tasks that support probability studies
- Good tool for basic research into general principles underlying survey response
- Good tool for collecting data about researcher behavior
- An opportunity to have a conversation about better methods (with researchers and sample providers)

## Spreading the Gospel of Survey Methods

- Schwarz (2007): “The initiators of the first Cognitive Aspects of Survey Methodology conferences hoped to build a two-way bridge between cognitive psychology and survey methods to facilitate an exchange that would advance basic research and improve survey practice. In the two decades since these conferences, this bridge has seen considerable traffic. However, much of this traffic has been from psychology to survey methods.”
- Some hope still that traffic can include substantive contributions to psychological theory
- Need for “another lane” on the bridge to transfer knowledge about how to improve methodological rigor

## Thank-you... Questions?

jjchandl@umich.edu

## Getting Started


Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*