Mechanical Turk and Financial Dependency on Crowdsourcing

Emergent Research Forum (ERF)

Eric Bogert
University of Georgia
etbogert@uga.edu

Abstract

Some workers on Amazon Mechanical Turk (AMT) are dependent on income from AMT to pay for their basic needs. This could create two countervailing forces. Workers have incentives to work quickly, to maximize their income. However, if workers work too quickly, resulting in poor quality, they won’t have access to the highest paying tasks on the platform. This is a feature of work that is paid per-task, rather than per-hour. This paper investigates whether workers who are financially dependent on income from Mechanical Turk produce work of different quality than workers who are not financially dependent on Mechanical Turk.

Keywords

Mechanical Turk, Crowdsourcing, Financial Dependency

Introduction

Amazon Mechanical Turk (AMT) is an online micro-task crowdsourcing platform used to match requesters of Human Intelligence Tasks (HITs) with workers (Turkers) willing to perform these tasks. AMT is used when human labor is either a cheaper or more accurate substitute for machine labor. Frequent use cases in research include conducting surveys cheaply and classifying data to be used in machine learning. An important concern for researchers using AMT, or any other form of crowdsourced labor, is whether they generate quality data.

For the purposes of this paper, I will define data quality as “The degree to which an object or question is correctly identified or answered.” Although there are existing frameworks of data quality, this paper primarily focuses on accuracy. In order to help achieve higher data quality, AMT allows requesters to set criteria on who can work on their tasks. Researchers will often specify the country of residence, HIT approval rate, education level, or income when controlling who can participate in their study.

Because workers choose to do ad-hoc tasks, AMT constitutes an online labor market that is paid on a per-unit basis rather than an hourly rate. Given that this is a labor market, personal finances likely influence the motivation of Turkers to engage in tasks. Previous research has shown that 10% of Turkers say their AMT-related income is ‘sometimes necessary to make ends meet’, and an additional 8% state it is ‘always necessary to make basic ends meet’ (Ross 2010). The traditional filtering mechanisms that have been extensively examined (e.g. HIT Approval Rate, total HITs completed, and hourly pay) overlook the role of personal finances. I posit that the role of financial dependency on crowdsourcing (defined as relying on crowdsourcing income to pay for basic needs) has a significant impact on data quality.

The median hourly wage on AMT is approximately $2 an hour, significantly less than the US minimum wage (Kotaro et al. 2017). Only 4% of workers make more than the US minimum wage (Kotaro et al. 2017). This brings up an obvious question: What are the motivations for Turking, when it is less remunerative than a minimum wage job? Perhaps these workers place a premium on being able to work according to their own schedule, they cannot find a better paying job, or they believe in the mission of the work they do (i.e. someone who labels images for machine learning datasets may genuinely believe that this labeling improves the world). In addition to these motivations, I believe that financial dependency on crowdsourcing will...
have an impact on the number of units of work completed and the quality of the work. Our research explores this factor.

**Literature Review**

**Demographics of AMT and Data Quality**

Understanding the AMT crowdsourcing labor is important for scientific research – in 2015, more than 500 articles in social science journals with impact factors greater than 2.5 used Mechanical Turk (Chandler and Shapiro 2016).

Descriptions have been provided on basic demographic facts about the workforce. Most workers live in the United States and India, a majority are female, and there are approximately 180,000 distinct workers on the platform (Difallah 2018). Other findings exploring the relationships between data quality and AMT demographics can be found in Table 1.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buhrmester et al. 2011</td>
<td>• Compensation</td>
<td>• Data Quality</td>
<td>• Reasonable compensation differences don’t affect data quality</td>
</tr>
<tr>
<td></td>
<td>• Task length</td>
<td></td>
<td>• AMT is as reliable as traditional surveys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• AMT participants are more demographically diverse than American college samples</td>
</tr>
<tr>
<td>Staffelbach et al. 2014</td>
<td>• Work background: Turk vs. Engineer</td>
<td>• Data Quality</td>
<td>• Crowdworkers and graduate students scored similarly on quality of work, given a sufficient tutorial</td>
</tr>
<tr>
<td>Goodman et al. 2012</td>
<td>• Student sample vs. Mechanical Turk</td>
<td>• Data Quality</td>
<td>• Researchers should account for differences in ‘financial and social domains’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Personality traits</td>
<td>• Attention checks improve data quality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Attitudes about money</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Attention checks can be as simple as “Please put ‘b’ for this answer.” or something more elegant, which might depend on knowledge that was derived from earlier in the task. Other methods to determine data quality include asking the same question to many workers, and accepting the modal or mean answer, commonly referred to as Majority Voting (Wang et al. 2017).</td>
</tr>
</tbody>
</table>

Table 1: Sample of Previous Research on Data Quality in Mechanical Turk

Research has examined the effects of compensation, work background/occupation, and length and complexity of work on data quality. Given that AMT and other crowdsourcing platforms face unique hurdles to overcome when trying to ensure data quality, researchers have examined a set of tactics to ensure and to assess quality. For example, requesters may use ‘gold data’, questions to which requesters already know the answer, so as to disregard the answers from workers who submit incorrect answers to the gold data questions (Wang et al. 2017). These are commonly referred to as ‘attention checks.’ Attention checks can be as simple as “Please put ‘b’ for this answer.” or something more elegant, which might depend on knowledge that was derived from earlier in the task. Other methods to determine data quality include asking the same question to many workers, and accepting the modal or mean answer, commonly referred to as Majority Voting (Wang et al. 2017).
However, none of the studies above examine the effect of financial dependency on crowdsourcing on data quality. If someone depends on this to make ends meet, they may be more motivated to produce higher quality work or to produce more work (but maybe of lower quality). Therefore, this may be an important criterion when filtering for workers.

**Research Model & Hypotheses**

AMT, and all crowdsourcing platforms, are interesting labor markets because they are paid per task, rather than per hour. This creates enormous incentives to work as speedily as possible, while maintaining a sufficiently high level of work so as to not have one’s work rejected. Usually, the problem employers face when controlling quality is that workers are unmotivated to complete a task, because they are paid per hour, regardless of output. This flipped dynamic, in a low wage setting such as AMT, should provide interesting theoretical grounds for labor markets and platform construction.

The unique aspect of this study is its focus on those who rely on crowdsourced work for their basic needs. If a worker depends on AMT for money, there will be two forces that interact to determine data quality. First, they won’t want to lose an important source of income and, therefore, they would be motivated to produce high quality work. Second, and alternatively, they will want to work as quickly as possible. However, this means that the motivation for the work is purely monetary, and rational workers will perform a simple cost-benefit calculus on what is the lowest data quality they can submit while still getting a task accepted. While there has been extensive research into demographic factors such as household income, and there has been research into the motivations behind Turking, to our knowledge there has been no study that looked at whether the people who are doing crowdsourced work in order to pay for basic needs are systematically performing lower or higher quality work than people who don’t depend on AMT income for their basic needs. This could have major ramifications for the platform.

Thus, given that arguments can be made for both a positive and a negative relationship, I posit two hypotheses:

**Hypothesis 1:** Financial dependency on crowdsourcing will result in lower data quality.

**Hypothesis 2:** Financial dependency on crowdsourcing will result in higher data quality.

**Data Collection and Analysis**

I will post a survey to Mechanical Turk, and will set my minimum HIT acceptance ratio to 90%. I am choosing 90% instead of 99% as the cutoff so as to generate more variance in the data quality. I will also set a minimum threshold of Turkers having completed at least 1,000 tasks prior to taking the survey. In accordance with the recommended minimum wage paid for academic work, this survey will pay $0.1 dollars per minute (Guidelines for Academic Requesters 2017).

The survey will contain two types of questions: demographic and image labeling. The demographic data will include the variables already known to affect data quality on Mechanical Turk, including income, education, and age. In order to ascertain financial dependency, I will ask two questions of our participants. I will use a continuous scale between 0 and 100 shown in Figure 1.

![Continuous Scale to Measure Financial Dependency](image)

**Figure 1: Continuous Scale to Measure Financial Dependency**
I will also ask them 3 questions about their dependence on Mechanical Turk, after which I will evaluate Cronbach’s alpha to see whether the questions correctly measure the construct. Between these two measures, I will impute their financial dependence on Mechanical Turk.

![Diagram of proposed questions to determine financial dependency](image)

**Figure 2: Proposed Questions to Determine Financial Dependency**

After obtaining the data, I plan to use a linear model incorporating data from both the sliding scale and the Likert scales to determine the effect of financial dependence on data quality.

**Conclusion**

I hope this study adds to the conversation around labor-related platform development. This is a burgeoning field of research, and there are important practical, scholarly, and ethical considerations in how platforms that pay on a per-task basis are designed. This study will contribute to the literature on crowdsourcing demographics related to data quality. There have been many studies (Steelman et al. 2013, Buhrmester et al. 2011, Paolacci et al. 2010) that have observed demographic data in Mechanical Turk, but none that have looked at the interaction of whether the workers who need the money most have lower data quality as a result of hurrying through HITs or higher data quality because they depend on the income. This paper should also contribute to both the practical and scholarly conversations around labor markets in which workers are paid per task, anonymously, and without the obligation or strong preference to continue working for that specific organization long-term.
REFERENCES


