On the Ethics of Crowdsourced Research

August 10, 2014

Vanessa Williamson

Harvard University

In this article, I examine the ethics of crowdsourcing in social science research, with reference to my own experience using Amazon's Mechanical Turk. As these types of research tools become more common in scholarly work, we must face the fact that many participants are not one-time respondents or even hobbyists. They work long hours completing surveys and other tasks for very low wages, and many rely on those incomes to meet their basic needs. I present my own experience interviewing Mechanical Turk participants about their sources of income, and offer a series of recommendations to the individual researcher and to social science departments and journal editors regarding the more ethical use of crowdsourcing.

Introduction

Social science research has benefitted in recent years from the use of "crowdsourcing": the enlistment of many people, usually via the internet, to complete a project. Crowdsourcing makes any number of research tasks easier and cheaper, including recruiting participants for survey experiments, transcribing text, and cataloguing non-computer readable documents. But crowdsourcing also presents ethical questions regarding appropriate compensation and protections for participants.

I examine the ethical concerns that result from the use of crowdsourcing, with reference to my own experience using one common crowdsourcing tool, Amazon's Mechanical Turk. Among political scientists, there has been a substantial increase in the use of this service in the past few years, largely because of the easy and low-cost access it provides to a pool of survey and experiment respondents. But a substantial portion of Mechanical Turk participants are not hobbyists; they work long hours completing surveys for very low wages, and many rely on those incomes to meet their basic needs. Like piece workers of the late 19th Century, crowdsourcing participants also lack the employment protections that apply to most other U.S. workers.

I present my experience interviewing Mechanical Turk participants about their sources of income, and offer a series of recommendations to researchers, social science departments, and journal editors regarding the more ethical use of crowdsourcing. These fixes are not a complete solution, however; as crowdsourcing becomes a more regular feature of political analysis, the discipline should continue to examine its participation in these largely unregulated markets.

Mechanical Turk and Social Science

Mechanical Turk, or MTurk, is an Amazon.com crowdsourcing tool that allows "requesters," including businesses and researchers, to easily hire anonymous "workers" to complete brief tasks for a small payment. The service has become an increasingly popular tool for scholars, particularly for those seeking to conduct survey experiments.

It is difficult to calculate the frequency with which crowd-sourcing tools are used, in part because there are not discipline-wide standards for reporting the methods by which a researcher completes mundane tasks. But journal references to Mechanical Turk in the digital library JSTOR sets a lower bound that suggests that academic interest in crowdsourcing has increased substantially in only a few years (Figure 1). Mechanical Turk has been used research published in such prestigious publications as *Political Analysis* and *Public Opinion Quarterly*.
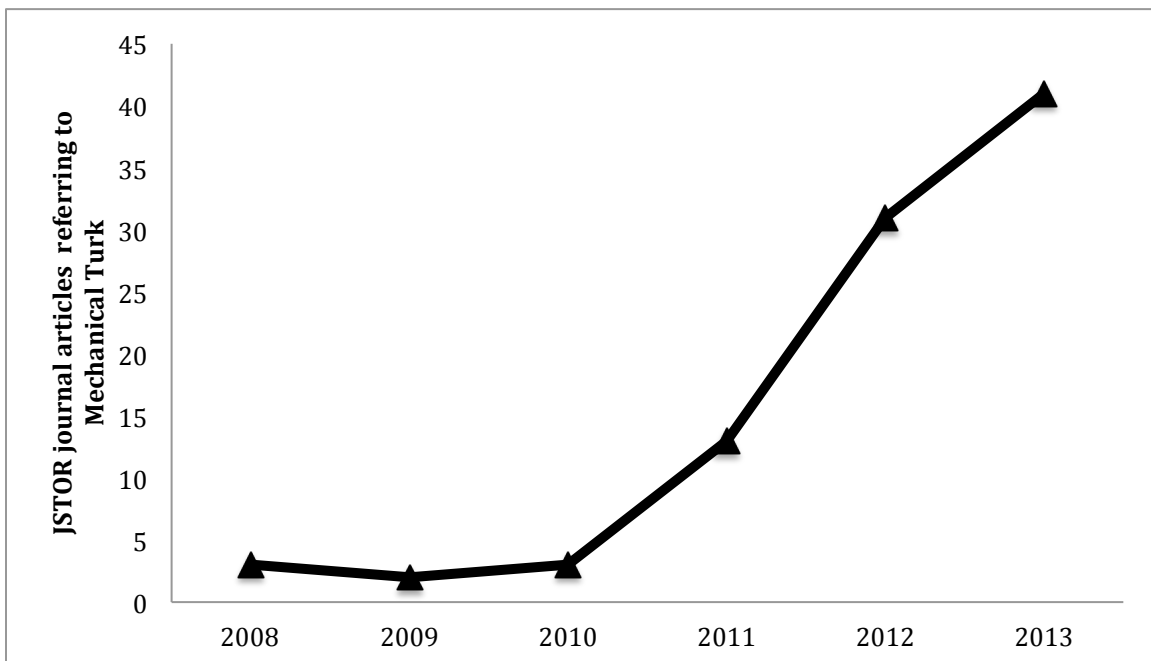

Figure 1: The frequency of JSTOR journal articles referring to Amazon's Mechanical Turk, by year.

Other digital libraries, including Academic Search Premier and PsycInfo show similar trends. These libraries only catalogue work that has been published, and so do not give a sense of the far larger pool of conference papers and works-in-progress that use MTurk. A search on Google Scholar turns up literally thousands of these works, rising from 173 hits in 2008 to 3,510 in 2013.[i]

Mechanical Turk is popular in large part because researchers have found success using it to conduct survey experiments. According to Berinsky et al. (2012), MTurk respondents are "often more representative of the U.S. population than in-person convenience samples," and can be used to replicate studies conducted using nationally representative pools. "All told…the MTurk sample does not perfectly match the demographic and attitudinal characteristics of the U.S. population but does not present a wildly distorted view of the U.S. population either," they conclude. These results are largely in keeping with those of Paolacci et al. (2010), Buhrmester et al. (2014), and Ross et al. (2010).

That breadth of reach does not carry *across* surveys, however. In fact, about 80 percent of tasks on Mechanical Turk are completed by about one fifth of participants, who spend more than fifteen hours a week working on MTurk (Adda and Mariani 2010; Fort et al. 2011). As a result, different social scientists are likely reaching many of the same participants.[ii] The fact that many MTurk participants are "regulars," as it were, has substantial ethical implications beyond those that usually govern the compensation of study participants.[iii]

The Mechanical Turk model relies on a worker accepting a given task at a known rate of payment. Workers have the option of refusing to accept any task if they consider

the price too low, and research has shown that response rates are slower when payments are smaller (Buhrmester et al. 2011). But unless one believes that market forces cannot be exploitative of workers, the "going rate" is not necessarily fair compensation. Figured as an hourly wage, Mechanical Turk offers an extraordinarily low wage; $2.30 an hour for workers in the United States, according to Ross et al. (2009).

My own research provided an arresting glimpse into the lives of frequent Mechanical Turk workers, and suggests the need for reform of scholarly use of this and other crowdsourcing services.

A Glimpse Inside the Worklife of a Turker

Between July 2013 and March 2014, I conducted three rounds of surveys on Mechanical Turk, resulting in a total pool of 1404 survey respondents, all residents of the United States. The survey and interview results make up one component of a larger project on American opinions about taxation. Respondents were asked whether they would be interested to participate in an hour-long follow-up interview in exchange for a $15 'bonus' to their Mechanical Turk account. There was a high level of interest in this prospect; 28.9% of my total pool of survey respondents said they were willing to participate in a one-hour interview. I conducted interviews with 49 respondents in 21 states. I was not seeking information about the interviewees' experience with Mechanical Turk, but did ask respondents about their sources of income. In this context, I heard a great deal from respondents for whom Mechanical Turk plays an important role in their daily lives and family budgets.

Some interviewees I spoke to are indeed economically comfortable people who treat MTurk as an amusement or source of disposable income. I spoke to a federal patent

attorney and a retired lieutenant colonel, among other people of high socio-economic status. Other interviewees are middle-income earners who use Mechanical Turk to save for major purchases. Jessica is a mental health therapist. "We only have one computer between my husband and I right now, " she says. "That's why I'm doing Mechanical Turk, too, just trying to get a little extra money." For some people, then, MTurk is indeed a diversion that plays a comparatively small role in their finances.

But a very substantial number of the people I spoke to were far from hobbyists. In fact, many of them were barely making ends meet. Particularly among older MTurk participants, answering surveys appears to be an important, but inadequate, source of income. Among the fifteen people I interviewed over fifty years old, six were surviving on some form of government assistance.[iv] Donna[v] is 67, living on the Gulf Coast of Texas. Her home was hit by Hurricane Rita and she was left destitute. "The economy makes it very difficult these days," she says. "So, that's how I came to be a Turker in my spare time." Wilma, 57, has a similar story. A back injury put her out of work before she could receive her whole pension, so now she's getting by on Social Security disability. "You skimp here, skimp there," she says. "I work a little bit on the Turk to make a little money to make ends meet."

It is not only the older people on Mechanical Turk who report using the service as a major source of income. Adam is 26 years old and has not found full time work; he is living at home with his parents. He relies on the dribs and drabs of money he collects from different online sources, particularly Amazon. Alexa, from Mississippi, is married with two children; her husband was earning about $9 an hour working full-time, and she is "working two part-time jobs that makes one full-time job." The family could receive

food stamps, Alexa knows, but they have recently chosen not to take the money. Though they are trying to get by without government benefits, the family is living on the edge of poverty; Alexa is waiting months for her income tax refund to replace the family's clothes dryer. She, too, uses MTurk to support her family.

The interviewees who were struggling financially were very familiar with Mechanical Turk social science surveys. Asked her opinions about tax progressivity, Donna says, "Oh, goodness. Every time I see one of those surveys with that question it, my god, I always say give it to them good. Make them pay." She was certainly familiar with several of the more common questions asked about economic inequality and redistribution.

But even for those working on the site full time, Mechanical Turk does not provide a living wage. Marjorie, 53 years old and living in Indiana, used to work in a grocery store and as a substitute teacher, before a bad fall left her unable to work. Now, she says, "I sit there for probably eight hours a day answering surveys. I've done over 8,000 surveys." MTurk is a major contributor to her family's tiny budget, but her full time labor does not add up to a salary; Marjorie estimates that she makes "$100 per month" from Mechanical Turk, which supplements her monthly $189 in Food Stamps.

Some respondents have tried to increase the payments they receive via Amazon. Wilma provides feedback to survey makers, she says. For instance, she once wrote to complain that "it took me an hour to do a survey, it paid a dollar. That's too long." Sometimes, she says, she does hear back from researchers including from "a lot of these universities." But her feedback is not enough to change the bigger picture, she believes. "More people on Amazon should give feedback," she says.

> [Turkers] gripe on an external website… about the low rate, which some of it is a very low rate. You have to bring those things to Amazon. Amazon needs to look at that.

For workers with few other income options, however, there is little leverage to encourage Amazon to change its policies. As Marjorie notes, "There are no jobs close to me. There's no public transportation. I can't go to work now because I don't have a car." Online work is one of the few avenues available to disabled Turkers.

My research highlights the daily struggle of MTurk workers to make ends meet on very low wages. But how representative are these interviewees of the larger Mechanical Turk pool?[vi] Representativeness is simply not an appropriate goal for small-n qualitative research; interviews are necessarily conducted with those who are willing to participate, and they may be in some important way different from other people. What these interviews do is present a far richer picture of the daily life of some of the Mechanical Turk participants. They provide, as Mario Luis Small (2009) puts it, "saturation" rather than "representation."

But it is obvious from the quantitative data that, at least in their economic status and their reliance on Mechanical Turk as a source of income, these interviewees are not uncommon. A sizeable portion of these MTurk workers are low-income. About 19% of U.S.-based MTurk workers are earning less than $20,000 a year (Ross et al.2010), a result mirrored in my own survey.[vii] More than a third of U.S. Turkers rely on Mechanical Turk as an important source of income (Ross et al. 2010). From an ethical standpoint, moreover, if even a small minority of workers rely on MTurk to make ends meet, social scientists (including myself) are participating in a market that leaves the people we study in precarity and poverty.

Barring a move by Amazon to increase the base rates allowed on Mechanical Turk, social scientists can do their part to improve the economic lot of people like Marjorie, Wilma, Alexis, and Adam. My own experience shows that even individual researchers operating with a very small budget can improve the wages they provide their respondents. But systemic reform is needed if we are to avoid the exploitation of online research participants.

Can Crowdsourced Research Be Ethical?

The only other article I have found on the ethics of Mechanical Turk is in the field of computational linguistics, and suggests that researchers find alternatives to crowdsourcing (Ford et al. 2011). In fields where Mechanical Turk is used to mimic machine labor, this may provide an obvious solution to the problem of crowdsource exploitation. But for social science researchers, machines do not offer an alternative to people for the completion of surveys or experiments. And my interviews suggest that, without Mechanical Turk, at least some Turkers would have few employment alternatives.

Can a conscientious social scientist use tools like Mechanical Turk? The purpose of this article is to provoke a debate about this important and under-examined question, rather than to offer a definitive answer. But there are undoubtedly positive steps that can be taken by an individual researcher as well as by those with the power to help set discipline-wide norms.

For the individual researcher, one option is to set and report a "minimum wage" for one's own research. The federal minimum wage is currently $7.25, or more than three times the average MTurk wage. Among states with a higher minimum wage threshold,

the average is about $8.00. In addition, several states and cities have passed legislation to increase the minimum wage to $10. Of course, most MTurk tasks take only a fraction of an hour. For a task that takes five minutes, one would pay each worker 61 cents to surpass the federal minimum wage, 67 cents to pass the $8-an-hour threshold, and 84 cents to surpass the $10-an-hour mark. Picking a higher rate can help offset the time a Turker loses between tasks.

It is reasonable to wonder if such rates, substantially above the rate paid by most MTurk requesters, might distort the pool of respondents one receives, and therefore one's findings. The very limited evidence on this question suggests that compensation rates "do not appear to affect data quality" (Buhrmester et al. 2010), but it is unclear whether rates affect the population of participants.

For those unwilling to risk biasing their Mechanical Turk pool with higher payments, or for researchers whose work is already complete, there is still an easy route to higher payment. Workers can be given bonuses retroactively. These bonuses are arduous to apply individually, but a simple shell script allows one to apply bonuses *en masse*. This is the method that I used to raise the rate paid to the equivalent of a ten-dollar hourly wage.[viii]

Of course, paying higher rates costs money, a prospect unlikely to be painless, especially to young and underfunded researchers. For a 3-minute survey of 800 people, going from a 20-cent payment to a 50-cent payment costs an additional $240, plus Amazon's fees. But the alternative is continuing to pay below-minimum-wage rates to a substantial number of poor people who rely on this income for their basic needs. This is simply no alternative at all.

The discipline as a whole can help make things easier for the individual researcher struggling with survey costs. Creating social science research norms will take the burden off the individual researcher to allocate their often limited funding fairly. I offer three suggestions for those in positions to affect research patterns more broadly.

- Journal editors can raise the bar by making a commitment to publishing only articles that pay respondents at ethical rates. The inclusion of language in journal submission requirements requiring authors to pay minimum wage-scale incomes to crowd-source workers, and report that wage based on the average actual length of the assignment, would have an immediate impact on the standards of the discipline as a whole.

- Grantmakers should not only follow the same standard of payment and reporting I suggest for journal editors, but should also provide funding at appropriate levels given that commitment.

- Social science departments and university-wide academic internal review boards concerned with the use of human subjects should create guidelines for the employment of crowd-source workers, as the discipline has done for numerous other research protocols. In this context, consideration should be given to concerns beyond wages. Crowd-sourcers lack of access to other employment protections (for instance, limits on the number of hours they can work), and have few avenues to organize themselves to push for new industry standards.

It is crucial to recognize that these steps are an incomplete solution. But in the immediate

term, we should not allow the perfect to be the enemy of the good. Social science

researchers can and should act immediately to raise the rates they pay their

crowdsourcing workers.

Conclusion

If a person were participating in only a single survey, the difference between a

dime and a quarter inducement would be small indeed, at least to most residents of the

United States. And if someone with a full-time job prefers online surveys to video games

as an evening's entertainment, that too would seem innocuous. But interviews with

Mechanical Turk workers suggest that many of these workers are not hobbyists. Instead,

they are laboring under real economic hardships, a situation that leaves these research

participants with only limited recourse against exploitation. The frequency with which

Mechanical Turk workers live on relatively low incomes and rely on Mechanical Turk as

a source of employment is confirmed by quantitative data. Thus, Turkers are not, and

should not be treated as, one-time participants. They are workers upon whose labor an

increasing percentage of social science research is based. In sum, what we know about

Mechanical Turk makes clear the need for reform of crowd-sourced research.

This paper is intended to be the beginning of a conversation, not the final word. I

have focused on the most common crowdsourcing service, Mechanical Turk, and the

experience of Turkers in the United States. But the concerns I raise would certainly apply

to other crowdsourcing websites, and may have additional implications for research

conducted with participants overseas. A growing percentage of those completing

Mechanical Turk tasks are living in India, for instance (Ross et al. 2010), though these

participants are sometimes excluded from survey experiments conducted by American researchers.

Voluntarily increasing the rate of payment for MTurk tasks will not resolve the fundamental inequities of precarious employment. In some ways, the economic situation of Turkers resembles that of piece workers more than one hundred years ago. Theodore Roosevelt, then a New York Assemblyman committed to *laissez faire* economics, wrote about a visit he paid to cigar-makers who worked from home, were paid by the piece not the hour, and lacked even basic worker protections:

> [M]y first visits to the tenement-house districts in question made me feel that, whatever the theories might be, as a matter of practical common sense I could not conscientiously vote for the continuance of the conditions which I saw. These conditions rendered it impossible for the families of the tenement-house workers to live so that the children might grow up fitted for the exacting duties of American citizenship. (Roosevelt 1919).

The broader trends in 21st Century employment are for social scientists to study, not to solve. But we should not and must not continue to balance our research on the backs of people like Wilma and Marjorie. Ironically, many articles that rely on MTurk – including my own – are examining questions of equity and fairness. If these values are important to study, they are also important to implement in our research practices.

Works Cited

Adda, Gilles and Joseph Mariani. 2010. Language resources and Amazon Mechanical Turk: Legal, ethical and other issues. In LISLR2010, "Legal Issues for Sharing Language Resources workshop," LREC2010, Malta.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3): 351–68.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6(1): 3–5.

Chandler, Jesse, Pam Mueller, and Gabriele Paolacci. 2014. "Nonnaïveté among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers." *Behavior Research Methods* 46(1): 112–30.

Dickert, Neal, and Christine Grady. 1999. "What's the Price of a Research Subject? Approaches to Payment for Research Participation." *New England Journal of Medicine* 341: 198–203.

Fort, Karen, Gilles Adda, K. Bretonnel Cohen. 2011. "Amazon Mechanical Turk: Gold Mine or Coal Mine?" *The Association for Computational Linguistics*.

Paolacci, G. (Gabriele), J. (Jesse) Chandler, and P. G. (Panagiotis) Ipeirotis. 2010. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5(5): 411–19.

Ross, Joel et al. 2010. "Who Are the Crowdworkers? Shifting Demographics in Amazon Mechanical Turk." http://cs.pugetsound.edu/~jross/pubs/RossEtAl-altCHI2010-slides.pdf (April 8, 2014).

Roosevelt, Theodore. 1919. *Theodore Roosevelt, an Autobiography*. New York: Macmillan.

Small, Mario Luis. 2009. "'How Many Cases Do I Need?' On Science and the Logic of Case Selection in Field-Based Research." *Ethnography* 10(1): 5–38.

Notes

[i] All publication counts were recorded in the summer of 2014.

[ii] Estimates of the actual number of Mechanical Turk participants is highly speculative; see Fort et al. 2011. As others have noted, repeat participants may present methodological challenges for experimental researchers. See Chandler et al. 2014.

[iii] For a review of the typical concerns regarding payments to study participants, see Dickert and Grady 1999.

[iv] The average value of SSDI is $1,145.70 a month, as of April 2014:
http://www.ssa.gov/policy/docs/quickfacts/stat_snapshot/

[v] To protect the privacy of my interviewees, all names are pseudonyms.

[vi] Complete information regarding the differences between the interview pool and the larger pool of MTurk survey respondents can be made available via an online appendix.

[vii] About 19% of the 1400 survey respondents were earning less than $20,000 a year. Even removing those partway through a college degree (assuming those to all be active college students, a very strong assumption) 12% of my MTurk respondents had a household income below $20,000 a year.

[viii] Code available at [author's website temporarily redacted]. This wage was for the survey only; interviewees received an additional payment of $15.