On the Ethics of Crowdsourced Research

Vanessa Williamson, *Brookings Institution*

ABSTRACT   This article examines 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 acknowledge that many participants are not one-time respondents or even hobbyists. Many people work long hours completing surveys and other tasks for very low wages, relying on those incomes to meet their basic needs. I present my own experience of interviewing Mechanical Turk participants about their sources of income, and I offer recommendations to individual researchers, social science departments, and journal editors regarding the more ethical use of crowdsourcing.

Social science research has benefited recently from the use of "crowdsourcing"—that is, the enlistment of many people, typically online, to complete a project. Crowdsourcing makes many research tasks easier, including recruiting participants for survey experiments, transcribing text, and cataloging non-computer-readable documents. However, crowdsourcing also presents ethical questions regarding appropriate compensation and protections for participants.

This article examines 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 (MTurk). Among political scientists, there has been a substantial increase in the use of this service in recent years, largely because of the easy and low-cost access it provides to a pool of survey respondents. But many MTurk participants are not hobbyists; they work long hours completing surveys for very low wages, relying on that income to meet their basic needs. Like pieceworkers of the late 19th century, crowdsourcing participants lack employment protections that apply to other US workers.

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

MTURK AND SOCIAL SCIENCE

MTurk is an Amazon.com service that allows "requesters," including businesses and researchers, to hire anonymous "workers" to complete brief tasks for a small payment. The service has become

a popular tool for scholars, especially those seeking to conduct survey experiments.

It is difficult to calculate the frequency with which crowdsourcing tools are used, in part because there are no disciplinary standards for reporting the methods by which a researcher completes mundane tasks. References to Mechanical Turk in the digital library JSTOR sets a lower bound, however, indicating that academic crowdsourcing has increased substantially in only a few years (figure 1). MTurk has been used in research on major questions of political behavior—including attitudes toward inequality, war, and political representation—and published in such prestigious journals as *Political Analysis* and *Public Opinion Quarterly*.

Other digital libraries, including Academic Search Premier and PsycInfo, show similar trends. These databases only catalog work that has been published; they do not provide a sense of the much larger pool of conference papers and works-in-progress that use MTurk. A search on Google Scholar results in literally thousands of these works, increasing from 173 in 2008 to 5,490 in 2014.[1]
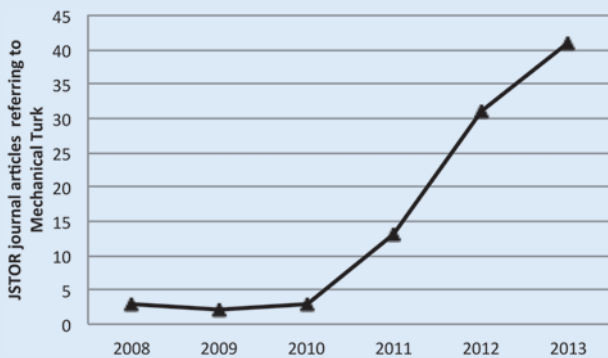
A common use of MTurk is for conducting survey experiments. According to Berinsky, Huber, and Lenz (2012), MTurk respondents are "often more representative of the US population than in-person convenience samples" and can be used to replicate studies conducted using nationally representative pools. These results are largely in keeping with those of Buhrmester, Kwang, and Gosling (2011); Paolacci, Chandler, and Ipeirotis (2010); and Ross et al. (2010).

That breadth of reach does not carry *across* surveys, however. About 80% of tasks on MTurk are completed by about 20% of participants, who spend more than 15 hours a week working on MTurk (Adda and Mariani 2010; Fort, Adda, and Cohen 2011). As a result, different social scientists are likely reaching many of the same participants. As others have noted, the "non-naiveté"

Vanessa Williamson *is a fellow in governance studies at the Brookings Institution. She can be reached at VWilliamson@brookings.edu.*

Figure 1

JSTOR Articles Referring to Mechanical Turk, by Year

My own research provided an arresting glimpse into the lives of frequent MTurk workers, which demonstrates the need for reform of academic crowdsourcing.

A GLIMPSE INSIDE THE WORKLIFE OF A TURKER

Between July 2013 and March 2014, I conducted three rounds of surveys on MTurk, resulting in a total pool of 1,404 survey respondents, all residents of the United States. At the conclusion of the survey, respondents were asked whether they would be interested in volunteering for an hour-long follow-up interview in exchange for $15.[4] There was a high level of interest in this prospect; 28.9% of the total pool of survey respondents said they were willing to participate. I conducted interviews with 49 respondents in 21 states. The interviews focused primarily on attitudes toward taxation. I was not seeking information about the MTurk experience, but I did ask respondents about their sources of income. In this context respondents for whom MTurk plays an important role in their daily life discussed its effect on their family budgets.

Some interviewees I spoke with 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 socioeconomic status. Other middle-class interviewees use MTurk to save for major purchases. Jessica[5] is a mental-health therapist. "We only have one computer between my husband and me right now," she said. "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.

Others I spoke with were far from hobbyists. In fact, many were barely making ends meet. Particularly among my older interviewees, answering surveys was an important but insufficient source of income. Among the 15 people I interviewed who were older than 50, six were surviving on some form of government assistance. Donna is 67 and lives on the Gulf Coast of Texas. Her home was damaged by Hurricane Rita and she was left destitute. "The economy makes it very difficult these days," she said. "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 full pension, so now she gets by on Social Security disability. "You skimp here, skimp there," she said. "I work a little bit on the Turk to make a little money to make ends meet."

of regular survey participants may present methodological challenges for experimental researchers (Chandler, Mueller, and Paolacci 2014).[2] But the MTurk model presents an ethical quandary as well.

The fact that many MTurk participants are paid subjects in multiple studies changes the ethical stakes. MTurk is different, for example, from another form of crowdsourced research known as "citizen science." Citizen-science projects typically make use of nonprofessional volunteers, such as asking amateur birdwatchers to record their sightings in a central database as part of an effort to track bird migration. Unlike these types of projects, MTurkers are themselves the subject of the research and therefore are entitled to special protection under Institutional Review Board (IRB) guidelines. Moreover, although a specific birdwatcher might also track butterflies, participation in one project does not facilitate participation in many other research projects. By contrast, MTurk gives workers access to hundreds or even thousands of tasks. Nonetheless, citizen science has come under scrutiny when it has relied on paid participants; at least one researcher called for IRB review of such research (Graber and Graber 2013).

Because of its reliance on numerous paid "regulars," MTurk has substantial ethical implications beyond those that typically govern the treatment of survey participants.[3] This is particularly evident when considering compensation. The MTurk model relies

Others I spoke with were far from hobbyists. In fact, many were barely making ends meet. Particularly among my older interviewees, answering surveys was an important but insufficient source of income.

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 rate too low, and research has shown that response rates are slower when payments are smaller (Buhrmester, Kwang, and Gosling 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, MTurk offers extraordinarily low compensation—about $2 an hour for workers in the United States, according to Ross et al. (2010).

It is not only older people on MTurk who report using the service as a major source of income. Adam is 26 and has not found full-time work; he is living at home with his parents. He relies on the small amounts of money he collects from different online sources, particularly Amazon. Alexa, from Mississippi, is married with two children. Her husband earns about $9 an hour working full-time, and she is "working two part-time jobs that make one full-time job." The family is eligible for food stamps, Alexa knows, but she and her husband recently chose to not take the

assistance. Although they are trying to survive without government benefits, the family is living on the edge of poverty. Alexa has waited several months for her income-tax refund to replace the family's clothes dryer. She, too, uses MTurk to help support her family.

The interviewees who were struggling financially were familiar with MTurk social science surveys. Asked her opinion about tax progressivity, Donna said, "Oh, goodness. Every time I see one of those surveys with that question in it, my God, I always say give it to them good. Make them pay." She was accustomed to several common questions asked about economic inequality and redistribution.

Even for those working on the site full-time, MTurk does not provide a living wage. Marjorie, 53 and living in Indiana, had jobs in a grocery store and as a substitute teacher before a bad fall left her unable to work. Now, she said, "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 limited budget, but her full-time labor does not add up to a salary. Marjorie estimates that she makes "$100 per month" from MTurk, which supplements the $189 she receives in food stamps.

Some respondents have tried to increase the payments they receive via Amazon. Wilma provides feedback to survey makers, she told me. 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 said, she does hear back from researchers, including from "a lot of these universities." But her feedback alone is not enough to change the bigger picture, she believes.

is needed if we are to avoid the exploitation of online research participants.

CAN CROWDSOURCED RESEARCH BE ETHICAL?

In the field of computational linguistics, Fort, Adda, and Cohen (2011) suggest that researchers find alternatives to crowdsourcing. Where MTurk is used to mimic machine labor, this may provide an obvious solution to the problem of crowdsource exploitation. For social science surveys, of course, computers do not provide an alternative to human respondents. Moreover, my interviews suggest that without MTurk, at least some Turkers would have few employment alternatives.

Can a conscientious social scientist use tools like MTurk? The purpose of this article is to provoke a debate 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 establish discipline-wide norms.

Above all, individual researchers can set a minimum wage for their own research. The federal minimum wage, for full- or part-time work, is $7.25 an hour as of this writing—or more than three times the average MTurk wage. Among states with a higher minimum-wage threshold, the average is about $8 an hour. In addition, several localities have passed legislation to increase the hourly minimum wage to $10 or $15. Of course, most MTurk tasks take only a fraction of an hour. For a task that takes 5 minutes, a worker would be paid 61 cents to surpass the federal minimum wage, 67 cents to pass the $8-an-hour threshold, 84 cents to sur-

About 19% of my 1,400 survey respondents were earning less than $20,000 a year, a result matching the research of Ross et al. (2010).

For workers with few other income options, there is little leverage to encourage Amazon to change its policies. As Marjorie noted, "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 MTurk pool? 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 different from other people.[6]

Robust quantitative data confirms, however, that these interviewees' economic status and their reliance on MTurk are not uncommon. About 19% of my 1,400 survey respondents were earning less than $20,000 a year, a result matching the research of Ross et al. (2010).[7] More than a third of US Turkers rely on MTurk as an important source of income, and more than 10% use MTurk money to "make basic ends meet" (Ross et al. 2010). From an ethical standpoint, moreover, if even a minority of workers relies 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.

Social scientists can do their part to improve the economic lot of people like Marjorie, Wilma, Alexa, and Adam. My own experience shows that even individual researchers with small budgets can improve the wages they provide. But systemic reform

pass the $10-an-hour mark, or $1.25 to reach $15 an hour. Paying a higher rate can help offset the time that a Turker loses between tasks.

It is reasonable to question whether these rates—substantially higher than the rate paid by most MTurk requesters—might distort the pool of respondents that researchers receive and, therefore, their findings. The limited evidence on this question suggests that compensation rates "do not appear to affect data quality" (Buhrmester, Kwang, and Gosling 2011), but it is unclear whether rates affect the demographic makeup of the participant population.

For those unwilling to risk biasing their MTurk 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. Bonuses are arduous to apply individually, but a simple shell script allows a researcher to apply them en masse. I used this method to raise the survey respondent rate to the equivalent of a $10 hourly wage.[8]

Of course, paying higher rates costs money—a prospect unlikely to be painless, especially for young and underfunded researchers. For a 3-minute survey of 800 people, going from a 20-cent 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 alleviate the burden for individual researchers struggling with survey costs. I offer the following three suggestions for those in a position to affect research patterns more broadly:

- Journal editors can commit to publishing only those articles that pay respondents an ethical rate, and they can require authors to report that wage based on the average actual length of the assignment.
- Grant makers should not only follow the same standard of payment and reporting that I suggest for journal editors but also should provide funding at appropriate levels, given that commitment.
- Social science departments and university IRBs concerned with the use of human subjects should create guidelines for the employment of crowdsource workers, as the discipline has done for numerous other research protocols. In this context, consideration should be given to concerns beyond wages. Crowdsourcers lack access to other employment protections (e.g., limits on the number of hours they can work) and have few avenues to organize themselves to push for new industry standards.

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 crowdsource workers.

et al. 2010), though international 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 pieceworkers more than 100 years ago (Schneider 2015). Theodore Roosevelt, then a New York Assemblyman committed to laissez faire economics, wrote about a visit he made 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 also are important to implement in our research practices.

From an ethical standpoint, moreover, if even a minority of workers relies 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.

CONCLUSION

If a person were participating in only one survey, the difference between a dime and a quarter inducement would be small indeed—at least to most US residents. If a person with a full-time job prefers online surveys to video games for evening entertainment, that choice also would seem innocuous. But the data show that for many crowdsource workers, MTurk is a significant source of income. These participants are laboring long hours under real economic hardships, in a situation that allows them only limited recourse against exploitation. 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. We must be cognizant not only of the ethical implications of our individual research but also of the cumulative effects of our discipline. In summary, what we know about MTurk clarifies the need for reform of crowdsourced social science research.

This article is intended to begin a conversation, not provide the final word. I focus on one common crowdsourcing service and the experience of American workers. However, the concerns I raise certainly would apply to other crowdsourcing tools and may have additional implications for research conducted with overseas participants. For instance, a growing percentage of those completing MTurk tasks are living in India (Ross

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/S104909651500116X. ∎

NOTES

1. All pre-2014 publication counts were recorded in the summer of 2014. The 2014 publication count was recorded in the summer of 2015.
2. Medical drug trials often bar participants from engaging in more than one study—a practice that protects both the participants from contraindicated treatments and the researchers from confounded results. With that said, according to Fisher (2009), an increasing number of participants are using these drug trials not as a supplement to but rather as a replacement for their regular health care. In this regard, the situation of the Turkers is similar: the incentive has become a basic necessity for some participants, raising serious ethical questions about whether they can consent freely.
3. For a review of other concerns regarding payments to study participants, see Dickert and Grady (1999).
4. In keeping with Amazon's terms of service, the survey task did not require participants to provide their contact information.
5. To protect the privacy of my interviewees, all names are pseudonyms.
6. See the online appendix for demographic data on the survey pool, the interviewees, and the US adult population.
7. Even eliminating those who are partway through earning a college degree (assuming they are college students with family support—a very strong assumption), 12% of my MTurk respondents had an annual household income of less than $20,000.
8. The code is available at https://dataverse.harvard.edu/dataverse/vanessa williamson.

REFERENCES

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, May 17.

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. "Non-naïveté 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.

Fisher, Jill A. 2009. *Medical Research for Hire: The Political Economy of Pharmaceutical Clinical Trials*. New Brunswick, NJ: Rutgers University Press.

Fort, Karen, Gilles Adda, and K. Bretonnel Cohen. 2011. "Amazon Mechanical Turk: Gold Mine or Coal Mine?" *The Association for Computational Linguistics* 37 (2): 413–20.

Graber, M. A., and A. Graber. 2013. "Internet-Based Crowdsourcing and Research Ethics: The Case for IRB Review." *Journal of Medical Ethics* 39: 115–18.

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

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

Ross, Joel, Lilly Irani, M. Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. "Who Are the Crowdworkers? Shifting Demographics in Amazon Mechanical Turk." *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pp. 2863–72. ACM.

Schneider, Nathan. 2015. "Intellectual Piecework." *The Chronicle of Higher Education*, February 16.

Diversity and Inclusion Programs

The American Political Science Association has several major programs aimed at enhancing diversity within the discipline and identifying and aiding students and faculty from under-represented backgrounds in the political science field. These programs include:

APSA Minority Fellows Program (MFP) (Undergraduate Seniors or MA Students)
MFP is a fellowship competition for those applying to graduate school, designed to increase the number of individuals from under-represented backgrounds with PhD's in political science. Application deadline: October of each year. For more information, visit www.apsanet.org/mfp.

Ralph Bunche Summer Institute (RBSI) (Undergraduate Juniors)
RBSI is an annual five-week program designed to introduce undergraduate students from under represented racial/ethnic groups, or students interested in broadening participation in political science and pursuing scholarship on issues affecting under-represented groups, to the world of graduate study and to encourage application to PhD programs. Application deadline: January of each year. For more information, visit www.apsanet.org/rbsi.

Minority Student Recruitment Program (MSRP) (Undergraduates and departmental members)
The MSRP was created to identify undergraduate students from under-represented backgrounds who are interested in, or show potential for, graduate study and, ultimately, to help further diversify the political science profession. For more information, visit http://www.apsanet.org/PROGRAMS/Diversity-Inclusion/Minority-Student-Recruitment-Program.

APSA Mentoring Program (APSA Members)
The Mentoring Program connects undergraduate, graduate students, and junior faculty to experienced and senior members of the profession for professional development mentoring. This program is a member benefit. To request a mentor or be a mentor, visit www.apsanet.org/mentor.

APSA Status Committees
APSA Status Committees develop and promote agendas and activities concerning the professional development and current status of under-represented communities within the political science discipline. For a listing of all APSA status committees, visit www.apsanet.org/status-committees.

For more information on all Diversity and Inclusion Programs, visit us online at http://www.apsanet.org/PROGRAMS/Diversity-Inclusion/ Please contact Kimberly Mealy, PhD, Director of Diversity and Inclusion Programs with any questions: kmealy@apsanet.org.

To contribute to an APSA Fund, such as the RB Endowment Fund or the Hanes Walton Jr. Fund, visit us at http://www.apsanet.org/ABOUT/Donate-Now.