Is it Ethical to Use Mechanical Turk for Behavioral Research? Relevant Data from a Representative Survey of MTurk Participants and Wages

Aaron J. Moss[1], Cheskie Rosenzweig[1,2], Jonathan Robinson[1,3], Leib Litman[1,4]

[1]Prime Research Solutions, Queens, NY

[2]Department of Clinical Psychology, Colombia University, New York, NY

[3]Department of Computer Science, Lander College, Flushing, NY

[4]Department of Psychology, Lander College, Flushing, NY

This paper is currently under review, and therefore, may not reflect the final version.

Abstract

To understand human behavior, social scientists need people and data. In the last decade, Amazon's Mechanical Turk (MTurk) emerged as a flexible, affordable, and reliable source of human participants and was widely adopted by academics. Yet despite MTurk's utility, some have questioned whether researchers should continue using the platform on ethical grounds. The brunt of their concern is that people on MTurk are financially insecure, subjected to abuse, and earning inhumane wages. We investigated these issues with two random and representative surveys of the U.S. MTurk population (N = 4,094). The surveys revealed: 1) the financial situation of people on MTurk mirrors the general population, 2) the vast majority of people do not find MTurk stressful or requesters abusive, and 3) MTurk offers flexibility and benefits that most people value above more traditional work. In addition, people reported it is possible to earn about $9 per hour and said they would not trade the flexibility of MTurk for less than $25 per hour. Altogether, our data are important for assessing whether MTurk is an ethical place for behavioral research. We close with ways researchers can promote wage equity, ensuring MTurk is a place for affordable, high-quality, and ethical data.

Keywords: Mechanical Turk, ethics, representative sample, online research

WC = 199

Is it Ethical to Use Mechanical Turk for Behavioral Research? Relevant Data from a Representative Survey of MTurk Participants and Wages

Scientists who study human behavior have always faced a rather ironic problem: finding people to study. In the early days of disciplines like psychology, researchers often served as their own subjects testing their ability to memorize words and striving to understand the workings of the mind through introspection. Then, as behavioral science matured, researchers sought to gather data from individuals, groups, and societies. Just before and after World War II, for example, researchers often learned about people's behavior by digging into archival records, conducting opinion polls, stationing themselves as observers in the world, soliciting volunteers from the community, and staging field experiments as people went about their daily business. In the last half of the twentieth century, many researchers relied heavily on laboratory studies conducted with undergraduates (see Sears, 1986) before shifting to online data collection in the 2010's (see Anderson et al., 2019; Sassenberg & Ditrich, 2019). The common thread running through this history is that sampling often follows the path of least resistance—researchers study people who are easy to find.

One of the easiest places to find research participants in recent years has been Amazon's Mechanical Turk (MTurk). As a microtask platform with hundreds of thousands of people worldwide (Difallah, Filatova, & Ipeirotis, 2018; Robinson, Rosenzweig, Moss, & Litman, 2019), MTurk gives researchers the opportunity to quickly gather more data from a more diverse pool of people than any prior sampling method. Due to its potential, several researchers were eager to evaluate MTurk's suitability for scientific research shortly after it was introduced to behavioral scientists (Buhrmeister, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeiriotis, 2010). Dozens of papers were published examining data quality (e.g. Litman, Robinson, &

Rosenzweig, 2015), comparing data from MTurk to that from other sources (Chandler, Rosenzweig, Moss, Robinson, & Litman, 2019; Hauser & Schwarz, 2016; Peer, Brandimarte, Samat, Acquisti, 2017), outlining methods of maintaining data quality (Peer, Vosgerau, & Acquisti, 2014), and describing the demographics of people on the platform (Iperiotis, 2010; Robinson, Litman, & Rosenzweig, 2020). After getting a grasp of these basic issues, researchers turned to MTurk-specific issues such as how the time of day and day of the week a study is launched might affect sample composition (Arechar, Kraft-Todd, & Rand, 2017; Casey, Chandler, Levine, Proctor, & Strolovitch, 2017; Fordsham et al., 2019), how repeatedly exposing participants to the same or similar materials might affect research findings (Chandler, Paolacci, Peer, Mueller, & Ratliff, 2015; Hauser, Paolacci, & Chandler, 2019), how to recruit naive participants (Robinson et al., 2019), and how well findings on MTurk replicate in nationally representative samples (Coppock, 2019; Mullinix, Leeper, Druckman, & Freese, 2015; Yeager, Krosnick, Visser, Holbrook, & Tahk, 2019). Thanks to widespread and sustained interest, MTurk may be the best understood participant platform in the history of social science research (see Litman & Robinson, 2020).

Although some initially feared MTurk would only be useful for simple surveys, those fears were quickly put to rest. Researchers experimenting with the platform found that in addition to cross-sectional surveys, MTurk was useful for, among other things, between-subjects experiments (Horton, Rand, & Zeckhauser, 2011), dyadic and interactive studies (Arechar, Gächter, & Molleman, 2018), transcription and translation services (Marge, Banerjee, & Rudnicky, 2010), image annotation (Rashtchian, Young, Hodosh, & Hockenmaier, 2010), user studies (Kittur, Chi, & Suh, 2008), longitudinal studies (Hall, Lewis, Jr., & Chandler, 2020), collecting narrative data (Grysman, 2015), collecting reaction time data (Crump, McDonnel, &

Gureckis, 2013), and teaching undergraduates about research (Bates & Lanza, 2013; Johnson & Borden, 2012). Because MTurk is both flexible and affordable, researchers in nearly every social science discipline have adopted the platform including those in addiction science (Strickland & Stoops, 2019), behavioral accounting (Buchheit, Doxey, Pollard, & Stinson, 2018), clinical psychology (Shapiro & Chandler, 2013), health and medical research (Mortensen & Hughes, 2018), industrial and organizational psychology (Cheung, Burns, Sinclair, & Sliter, 2017; Woo, Keith, & Thornton, 2015), information systems (Crowston, 2012), linguistics (Sprouse, 2011), medicine and cancer research (Arch & Carr, 2017), political science (Clifford, Jewell, & Waggoner, 2015; Levay & Freese, 2016), public administration and management (Stritch, Pederson & Taggart, 2017), personality and social psychology (Buhrmeister et al., 2011), social work (Chan & Holsoko, 2016), and sociology (Shank, 2016) among others. In fact, MTurk has been so influential that one of the papers originally introducing it to behavioral scientists has been cited in more than 1,000 different journals (see Buhrmester, Talaifar, & Gosling, 2018) and within some fields one out of every two articles published in top journals today contains data from MTurk (Anderson et al., 2019; Zhou & Fishbach, 2016). These numbers underscore just how important MTurk has been in helping social scientists understand human behavior.

Yet, as valuable as MTurk has been for social scientists, it is possible it has had even more impact on society. Various businesses routinely use MTurk to develop technology that aims to improve people's lives. In addition, major news organizations like *ProPublica* (Rao & Michel, 2010) and *The New York Times* (see Newman, 2019)—organizations with a mission to inform the public—have used MTurk for various data projects. And, most recently, at least one major Presidential campaign used Mechanical Turk to improve the effectiveness of its messaging

(Federal Election Commission, 2019), meaning MTurk may increasingly play a role in U.S. elections.

However, despite the contributions MTurk has made to both research and society, it is not without criticism. In recent years, a growing number of people have argued that MTurk is unethical. At the heart of this argument are concerns about exploitation. Specifically, some researchers worry that the people on MTurk are financially insecure and relying on MTurk to meet basic needs. In addition, a series of popular press articles have raised the possibility of widespread abuse, characterizing Mechanical Turk as an "online hell" (Semuels, 2018), a "digital sweatshop" (Graham, 2010), and an unregulated marketplace where people are routinely taken advantage of and earn only pennies per task (Newman, 2019). These descriptions have understandably given academic researchers pause as few people want to advance their careers by exploiting vulnerable people.

Within the conversation about the ethics of MTurk, three issues have been raised. First, some have suggested that people on Mechanical Turk are disproportionately poor or disabled and working on MTurk because they have trouble making ends meet elsewhere in the traditional labor market. As a captive audience, the argument goes, people on MTurk put up with mistreatment and low pay because they have no other way to make ends meet. The second claim made against MTurk is that working on MTurk often involves abusive and stressful interactions with requesters (the people who post tasks to the platform). Structurally, requesters on MTurk have more control than workers. This makes it possible for requesters to reject submitted work and withhold payment without even providing a reason. Requesters are also not accountable for poorly set up studies, broken links, or underestimates of study length, all issues that may cause people to spend time on tasks without being paid. For these reasons, many have claimed that

working on MTurk is unpleasant, abusive, and stressful. Finally, the third claim made against MTurk concerns pay. Critics of Mechanical Turk argue that people are paid pennies, and the most cited source of MTurk wages reports that workers make $2.00 per hour (Hara et al., 2018). Together, these claims paint a negative picture of MTurk and imply that researchers who use MTurk are benefitting from the mistreatment of tens of thousands of people.

While much has been written about the ethics of MTurk both in the academic literature and popular press, there is surprisingly little empirical evidence to support the three claims described above. Virtually everything written about the experience of people on MTurk, their financial situation, disability status, experiences of stress, and general orientation toward the platform is based on anecdotal evidence, case reports, or small, often self-selected samples. In this paper, we examine the ethical concerns that have been raised about Mechanical Turk by conducting the first random and representative surveys of people on the platform. Our surveys yield representative results by generating a sampling frame, randomly selecting participants, and accounting for different levels of participant experience. In our surveys, we asked people why they spend time on MTurk, how MTurk fits into their financial lives, and whether they are satisfied with different aspects of the platform like wages and the stress involved in completing tasks. We also asked people what they value about MTurk, how much people can earn, and whether they would take a more traditional job over MTurk if it was available. Finally, we asked people about the fairness and honesty of requesters, how often requesters engage in abusive practices, and, for people who have experience on other microtask platforms, whether they are treated more fairly on MTurk or other platforms. In addition to these surveys, we draw on outside data to place the issues of rejection rates and hourly earnings into context. Our aim was

to gather empirical data that can inform the conversation about ethics and the use of Mechanical Turk for behavioral research.

Criticisms of Mechanical Turk: Past Claims

Many researchers and journalists have wondered: who is willing to complete small tasks for small pay on MTurk? Popular press articles have claimed with increasing frequency that the answer is the poor and vulnerable. Indeed, within the discourse about Mechanical Turk a controlling narrative has taken hold that suggests people on the platform are disproportionately poor or disabled. The source of this concern stretches back to some of the first surveys in which a sizable portion (20%) of people reported relying on MTurk as their primary income (Iperiotis, 2010; Ross et al., 2010; see Fort, Adda, & Cohen, 2011). It was then later amplified by interviews conducted with people on MTurk (Williamson, 2014) and re-amplified in popular press articles (e.g., Semuels, 2018). In recent years, the idea that people are on MTurk because they struggle to find other jobs or because they "...are likely from groups traditionally excluded from the formal labor market, such as people with disabilities who have challenges securing jobs at contemporary office work environment" (Hara et al., 2018, pg. 10) has been taken almost as an article of faith among those who see MTurk as a form of exploitation. For example, when describing who works on MTurk, the author of a *New York Times* article said, "Some do it because there are few decent-paying jobs that can be done at will. People who are confined to their homes by disability or social anxiety or who live where there are few jobs do it because, despite lousy wages, it seems like the best option." (Newman, 2019). But how strong is the evidence for these claims?

As it turns out, many of these claims rest on research practices that are susceptible to painting a skewed picture of Mechanical Turk. For example, Williams (2014) conducted

interviews with 49 people out of more than 400 who responded to an offer to be interviewed. Although the resulting conversations yielded a rich glimpse into the lives of a few people, there is no telling which slice of the MTurk population these interviews capture or how representative they are. Similarly, journalists who speak to a handful of people on MTurk may be able to write a compelling article, but when more than 86,000 people in the U.S. complete tasks on MTurk each year (Robinson et al., 2019) it is unclear how representative the people portrayed in press articles are of the broader MTurk population. Even the best canvassings of people on MTurk (Berg, 2016; Hitlin, 2016) may be biased by factors such as the pay offered for the survey, the proportion of experienced and inexperienced participants in the dataset, and the sampling practices used by researchers.

Our survey aimed to provide a better answer to the question of whether people on MTurk are disproportionately poor or disabled. To do so, we asked people about their financial situation and whether they have a disability. In addition, we asked people how satisfied they are with the pay on MTurk, how they use the money they earn, and whether they would prefer to earn money elsewhere if they had the opportunity. Not only does our sampling method yield representative results of the U.S. MTurk population, but, where possible, we compared our results to nationally representative data. By doing so, we were able to assess whether people on MTurk are in a worse financial situation than Americans in general and whether people on MTurk are more likely to have a disability than people in the general population.

A second criticism of Mechanical Turk is that people are routinely subjected to stress, psychological harm, and abuse. Popular press articles have described MTurk as an “online hell” (Semuels, 2018) and reported “horror stories from inside” (Mehrotra, 2020), descriptions that certainly call to mind images of suffering. Although reports of abuse on MTurk take many

forms—unfair rejections, indifference from Amazon when problems arise, the content of tasks—most circle back to a fundamental power imbalance between workers and requesters. MTurk gives requesters the power to decide when a task has been sufficiently completed and when it has not. When requesters decide a task does not meet their standards, they can reject the work. Workers who have tasks rejected go unpaid and their reputation on the platform suffers, meaning people are understandably sensitive to rejections.

Another instance of abuse that is relevant to academic research is the content of tasks. People on MTurk say they are sometimes asked to view offensive or disturbing content. Even though some reports have indicated that this may be because tech companies post tasks to help train machine learning algorithms to flag offensive content (Matsakis, 2018), at least one article has suggested that the content of academic surveys is disturbing to people. Although not based on a representative sample, a report conducted by *Gizmodo* stated, "Yet as benign as academic research may seem, 12 percent of respondents claimed that the worst or strangest experience on Mechanical Turk was due to what can only be described as uncomfortable personal data requests wherein the worker reported feeling emotionally traumatized by an academic survey" (Mehrotra, 2020). Academic research is routinely vetted by institutional review boards and an important standard in the risk-benefit analysis that these boards conduct is whether the research poses a higher level of stress to the participant than they would typically experience in daily life.

Importantly, concerns about abuse on MTurk are not new. In fact, they date back almost to the launch of the platform. As early as 2010, academics were routinely discussing the exploitation of people on MTurk at conferences (Horton, 2011). At the same time, these early discussions were based more on speculation than data. When researchers conducted studies to find out whether people on MTurk view requesters as fair and honest, they found that most

people reported that requesters on MTurk were fairer and more honest than employers outside of MTurk (Horton, 2011). The author of the study wrote, “contrary to our prior expectations, rampant exploitation is a mischaracterization” (Horton, 2011). Since that time no empirical evidence to the contrary has been produced yet claims of abuse have multiplied.

The imbalance of power between workers and requesters on MTurk is undeniable and while most workers can report some negative experiences on the platform (Berg, 2015), what is not known is whether people’s experiences are routinely negative. As a result, we asked people in our surveys whether they are satisfied or dissatisfied with the amount of stress that taking HITs on MTurk entails. We reasoned that this global evaluation would allow us to tell if people regularly experience abusive practices or are exposed to disturbing content and whether people’s time on MTurk is better characterized by disturbing experiences or more mundane experiences. In addition, to investigate issues of abuse more precisely, in our second survey we replicated Horton (2011) by asking people whether they see requesters on MTurk as fairer and more honest than employers outside of MTurk. We also asked people to report how many hours they spend on MTurk each week, what percentage of HITs they have rejected, and what percentage of HITs contain disturbing material. These questions allowed us to quantify how often people are subjected to abusive practices.

Finally, the third criticism routinely made about MTurk concerns wages. Past reports about how much people earn have varied widely. For example, the most often cited study reports that median wages are as low as $2.00 an hour (Hara et al., 2018). Meanwhile, several sources of self-reported data indicate that people made around $5.00 per hour in 2015 and 2016 (Berg, 2015; Hitlin, 2016), and more recent analyses based on a larger dataset indicate that wages have risen to at least $5.70 an hour (Litman, Robinson, Rosen et al., 2020). In the popular press,

however, earnings have been characterized as “97 cents an hour” (Newman, 2019), and “tasks for hours on end, sometimes earning just pennies per job” (Semuels, 2018). These characterizations are frequently at the heart of some behavioral researchers’ hesitancy to use MTurk.

While the wages that people earn on MTurk have received a lot of attention, calculating how much people earn per hour on a microtask site that pays different rates for different tasks that last different durations is exceedingly difficult. Further complicating this calculation is that two people completing the same task will almost certainly make different hourly wages because there is substantial variability in how quickly people complete tasks (Litman, Robinson, & Rosenzweig, 2020). Given this complexity, our study took two approaches to assessing wages. First, we asked people to report how much a person can earn per hour on MTurk. This self-reported assessment of earnings has the advantage of overcoming the measurement issues mentioned above. On the other hand, self-reports can lack precision. So, we also used a large database of over 26 million HITs conducted on CloudResearch.com (formerly TurkPrime) to estimate wages. With this multimodal approach, we hoped to see convergence across methods.

In addition to estimating wages, we assessed what reasons other than wages might lead people to spend time on MTurk. Past research suggests some portion of people report using MTurk as a second job or to supplement their income. Few part-time jobs or entrepreneurial pursuits give people the freedom and flexibility MTurk offers. Hence, we asked people to list and rank several reasons why they may spend time on MTurk and then to state how much another job would need to pay in order for them to give up MTurk and take the other job. Examining people’s reasons for working on MTurk and asking about factors that augment hourly

wages allowed us to gain a better understanding of how people think about MTurk and what they value about the platform.

Current Research

This paper was motivated by our desire for empirical data that bear on the question posed in our title: is it ethical to use Mechanical Turk for behavioral research? Historically, this has been a bit of an open question with researchers gathering evidence that attests to both the merits and demerits of MTurk. Recently, however, some people have concluded that MTurk is unethical and use of the platform, even in a personally responsible way, is irredeemable. For example, in addition to calls from some people to stop using MTurk, an article published in *The Intercept* in January 2020 neatly illustrates our point. The article described how the campaign of Democratic primary contender Pete Buttigieg spent $20,000 conducting polling on Mechanical Turk in the summer of 2019. Importantly, the article focused not on whether the Buttigieg campaign used MTurk ethically but on the fact that they used the platform at all. As evidence of MTurk's intrinsic immorality the authors wrote: "The campaign's use of an exploitative platform like MTurk is in sharp contrast with the way Buttigieg has cast himself as a pro-worker candidate," and "While exploitation is rampant in the gig economy, MTurk has been identified as one of the worst offenders by journalists and researchers" (Grim & Lacy, 2020). Similar claims have been made increasingly often by people in academia, the popular press, and on social media in recent years. However, few rigorous studies have directly examined whether these claims are backed by solid empirical evidence; hence, the reason for our study.

In our studies, we sought to gather empirical data that can inform the conversation taking place about the ethics of MTurk. To that end, we constructed our surveys to examine some of the most common criticisms made against MTurk. Both of our surveys were designed to produce

representative results of the U.S. MTurk population and, where possible, we compared the responses of people on MTurk to nationally representative data. In the Discussion, we draw on outside data to further place the issues of wages and rejection rates in context. Our research was approved by IntegReview, an independent institutional review board that reviews research involving human subjects.

Study 1: A Representative Survey of MTurk Participants

We conducted a representative survey to understand why people spend time on MTurk, how MTurk fits into their financial lives, and what the experience of completing tasks on MTurk is like. Until now, representative surveys of people on MTurk have been all but impossible to conduct because researchers lack a sampling frame—a list of all people from which to randomly sample. We solved this problem by generating a list of everyone who completed at least one HIT in the CloudResearch database over a period of two months. Then, we stratified the sample based on participant experience. With a sampling frame and criteria upon which to stratify our sample, we randomly selected people and invited them to our survey. People who responded provided consent to participate and then answered 14 questions asking about a variety of MTurk-related issues. At the end of the survey, people completed demographic questions and were thanked for their time. We paid $0.75 cents for a survey that took 5 minutes on average to complete for an effective pay rate of ~$9.00 per hour.

Participants and Procedure

Data collection occurred in mid-December 2019. We collected data from a total of 2,026 people in the U.S.[1] Tables 1 and 2 display complete demographic information. Because our survey yields representative results of the U.S. MTurk population, this demographic information may be useful for researchers interested in the demographics of MTurk. In general, the demographic profile of our sample converges with past reports about the overall demographics of MTurk (Robinson, Litman et al., 2020).

A slight majority of people in our sample identified as female (56.44%) and the average age was close to 37 years (M = 36.99, SD = 12.65) with a range from 17 to 87. The racial demographics of the sample mirrored the U.S. population. Three-fourths of the sample identified as White, almost 13% identified as Black or African American, about 7% identified as Asian, and smaller numbers of people identified with other minority groups. In a separate question, 11% of people identified as Spanish, Hispanic, or Latino. Consistent with past research, the sample was well educated. Almost half of people (49.46%) reported holding a Bachelor's degree or higher. Finally, people reported a range of household incomes (see Table 2), but the overall distribution was generally similar to the U.S. population except for those making more than $150,000 per year (U.S. Census Bureau, 2018).

[1] Our data file contains responses from a total of 2,277 people. Due to an error when generating the sampling frame, wave 1 of data collection allowed 251 people outside of the U.S. to participate. We excluded responses from these people because we were interested in studying how U.S. MTurk workers feel about the platform.

Table 1. *Basic demographics.*

Age	%
18-29	32.45
30-39	34.05
40-49	16.17
50-59	10.72
60-69	5.11
70+	1.50
Gender	
Male	43.27
Female	56.44
Other	0.30
Race	
White	75.01
Black	12.67
American Indian or Alaska Native	0.89
Asian	7.10
Native Hawaiian or Pacific Islander	0.50
Other	3.83
Hispanic	
Yes	11.08
Highest Degree	
No college degree	37.62
Associate degree	12.92
Bachelor’s degree	34.50
Graduate degree	14.96

Marital status	
Married	45.51
Never married	42.39
Widowed	1.69
Separated	1.24
Divorced	9.17
Children	
Yes	46.10
Political party	
Republican	22.35
Democrat	38.68
Independent	28.77
Other	1.64
No preference	8.56
Political views	
Extremely liberal	9.44
Liberal	19.04
Slightly liberal	14.26
Moderate	22.86
Slightly conservative	12.18
Conservative	12.03
Extremely conservative	3.88
I haven’t thought about this	6.31

Table 2. *Annual household income for Mechanical Turk and the U.S. population.*

Annual household income	MTurk	CPS (2018)
< $10k	6.31	5.93
$10k-$19,999	6.60	8.77
$20k-$29,999	11.67	8.66
$30k-$39,999	10.82	8.73
$40k-$49,999	11.02	7.80
$50k-$59,999	11.22	7.56
$60k-$69,999	7.94	6.69
$70k-$79,999	7.30	6.02
$80k-$89,999	5.06	4.98
$90k-$99,999	5.16	4.42
$100k-$149k	11.97	14.95
>$150k	4.92	15.47

Note: CPS = Current Population Survey, the primary source of labor force statistics for the U.S. population. Data are based on 2018 income.

Sampling process. Our goal was to randomly sample people on MTurk while accounting for different levels of experience with the platform. To do so, we used previously reported data on worker experience levels (Robinson et al., 2019) and stratified our sample into four groups: people with less than 100 HITs completed (35% of the MTurk population); people with between 100 HITs and 1,000 HITs completed (37.4% of the population); people with between 1,000 HITs and 5,000 HITs completed (16.4% of the population); and people with more than 5,000 HITs completed (11.3% of the population). In all four groups, we recruited participants in proportion to their percentage of the MTurk population based on our target sample size of 2,000. This

means, for example, that we aimed to recruit 700 people with less than 100 prior HITs completed (35% × 2,000 = 700).

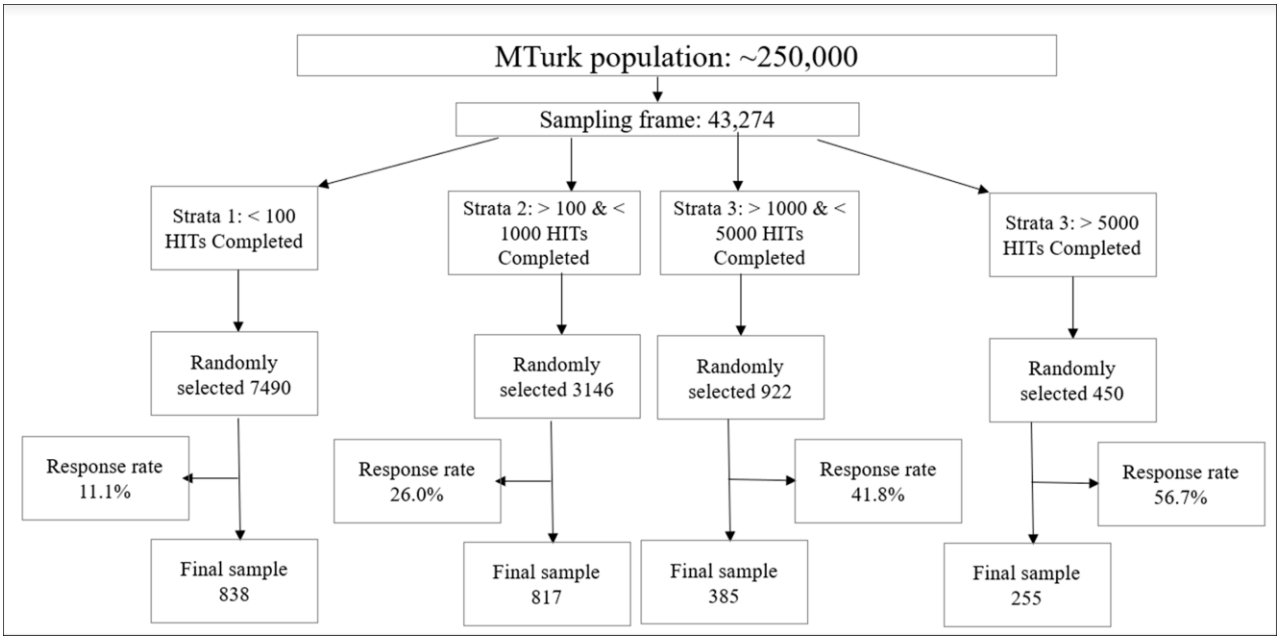
To sample within each group, we queried the CloudResearch database for all unique workers who completed at least one HIT in October and November 2019; this query yielded a list of 43,274 people. We were confident that this list contained nearly all active MTurk participants because CloudResearch hosts over 1 million HITs per month and close to 90,000 unique workers take HITs through CloudResearch each year. Further, the CloudResearch database contains more than 250,000 unique workers (Robinson et al, 2019), a number that is higher than previous estimates for the total number of people on MTurk (Difallah, Filatova, & Ipeirotis, 2018). Thus, it is likely that virtually all MTurk workers have taken at least one HIT through CloudResearch, making our sampling frame a near complete representation of the MTurk population.

Next, we determined people's level of experience by looking for the previous task each person had completed with the highest required number of approved HITs. For example, people were eligible for recruitment in Group 4 if the database indicated they had previously taken a study that required at least 5,000 completed HITs. After generating a sampling frame, we randomly selected people and invited them to the study. The selection process took place in waves to account for non-response.

In the first wave, we opened the survey to only the number of participants we wanted in each stratum, anticipating that some people would not respond. In subsequent waves, we invited new participants to fill the remaining slots in each stratum. In each wave of data collection, we invited people to the study when it was launched and then reminded them of the study with an email on each of the following two days. After three full days we closed the study and launched

a new wave to fill the remaining spots. In the last wave of data collection there were only a few slots remaining, so we invited twice the number of people needed in order to fill each stratum and bring data collection to a close. Figure 1 depicts the total number of people who were invited to take our HIT within each level of experience and the total response rate across all five waves of data collection. Overall, response rates ranged from 11.1% in the group of people with less than 100 prior HITs completed to a high of 56.7% in the group with more than 5,000 HITs completed. For comparison, the response rate in surveys that use random digit dialing—often considered the gold standard for representative surveys—is around 9% (Keeter et al., 2017).

Figure 1. The number of people sampled within each stratum and the total response rate across all waves of data collection.



Materials

The survey contained several questions asking people their thoughts about taking HITs on MTurk. First, we asked people to characterize their time on MTurk and how they use the money they earn. Then, we asked people whether they prefer MTurk over alternative jobs. Next, we asked people to rank some of the reasons they choose to work on MTurk and to put a price on the conveniences of MTurk when compared to working more traditional jobs. We asked people to report how much it is possible to earn on MTurk per hour, and finally, we asked people questions about their satisfaction with the wages and stress of MTurk. Several of our questions were adapted from nationally representative surveys, and the results are presented below with these comparisons when available.

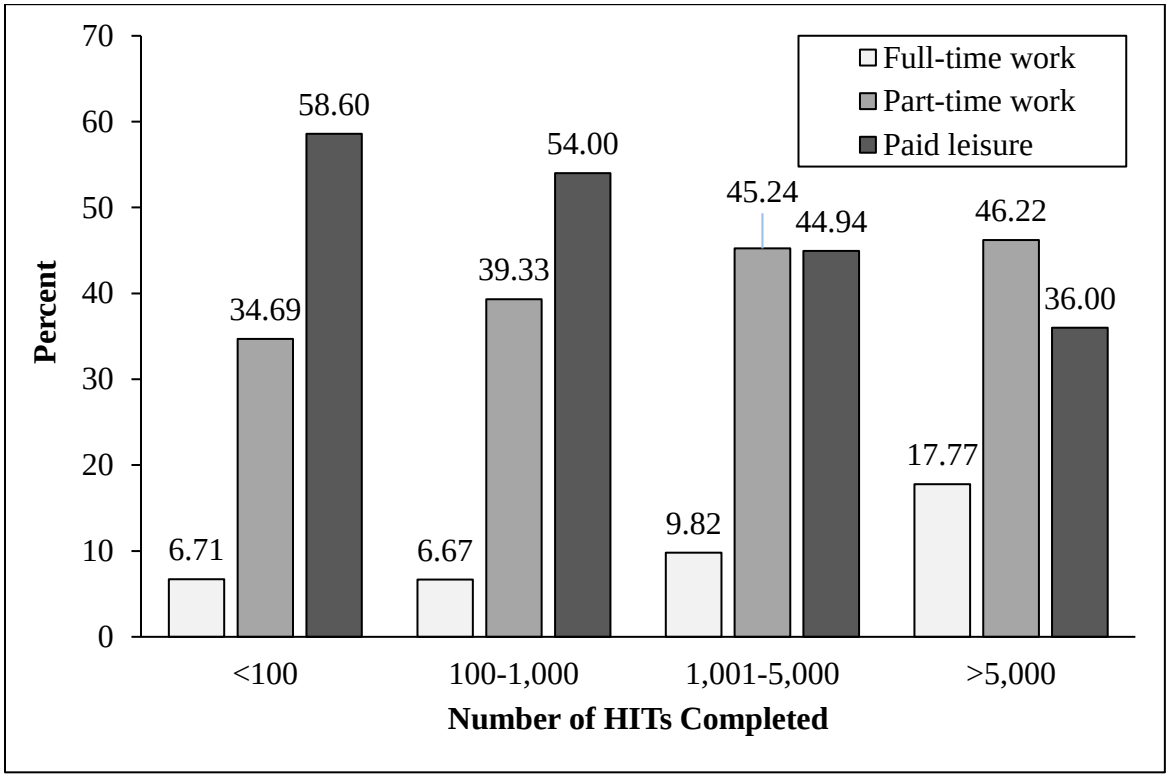
Results

Characterization of Mechanical Turk

We asked people to characterize their time on MTurk as either full-time work, part-time work, or a form of paid leisure. Most people (52.12%) characterized MTurk as a form of paid leisure. The next largest group (39.44%) characterized MTurk as a form of part-time work, and only 8.44% of people said MTurk was a form of full-time work.

As might be expected, people's experience (i.e., number of previous HITs completed), influenced how they characterized their time on MTurk. As experience increased, people were less likely to describe MTurk as paid leisure (see Figure 2). At no level of experience, however, did more than half of people characterize MTurk as part-time work and people characterizing MTurk as full-time work never surpassed 20%.

Figure 2. People's characterization of MTurk varied based on how many prior HITs they had completed on the platform.



The next question asked people why they work on MTurk. Answer choices were framed as earning money that covers "essential living expenses," "non-essential spending," or for "Other" reasons. Nearly 70% of people said they work on MTurk to earn money for non-essential expenses. Another 21.17% said they work to cover essential expenses. The remaining 10.51% of people provided alternative responses. Coding the alternative responses revealed a variety of reasons people work on MTurk. Importantly, however, of the 209 people who provided an alternative response just 2.4% provided answers that could be construed as covering essential things. The most common alternative responses were that MTurk provides extra money (38.8%), an amusing or pleasant distraction from other things (23.4%), money for savings

(12.9%), money to supplement other income (7.2%), or money to cover purchases on Amazon (6.7%).

Preferences for MTurk over Other Work

We asked people three questions measuring whether they would prefer to earn money on MTurk at their current wages or take an alternative job in retail or food service "earning typical pay for that job." We asked about this trade-off for three different levels of time commitment: full-time employment, part-time employment, and as something to do in their leisure time.

The results from these questions are presented in Table 3. As shown, people preferred MTurk over other work regardless of the time committed. Furthermore, as the time commitment decreased people's preference for MTurk increased.

Table 3. Percent of People Choosing MTurk over other Work

	Choice	
Time Commitment	MTurk	Other Work
Full-time	56.42	43.58
Part-time	68.66	31.34
Leisure	93.34	6.67

Reasons for Working on MTurk

We asked people to report reasons why they may prefer taking HITs on MTurk over working in a more traditional job. After selecting reasons, we asked people to rank each reason. The resulting data are presented in Table 4. As shown, the top reasons people chose involved flexibility. Specifically, people like the ability to work from home and to work flexible hours. Also supporting the idea that people like flexibility, people chose not having to deal with a boss, not having to commute, and flexibility for family as top reasons for making money on MTurk over other jobs.

One reason people could select for why they prefer to earn money on MTurk was "I have physical or mental health constraints that make it hard to work elsewhere." We used this question as a guage for what percentage of people might be on MTurk because of a disability. As Table 4 shows, 15.45% of people selected this reason. Although this represents a portion of people on MTurk, it is well below the 26% of the US population that report living with a disability (Centers for Disease Control and Prevention, 2019), suggesting, counter to some claims, that people on MTurk are not disproportionately likely to have a disability.

Table 4. Reasons for Spending Time on MTurk.

Reason	% Choosing	Average Ranking
Work from home	94.18	2.12
Flexible hours	92.25	2.70
No commute	71.72	5.05
No boss, supervisor, etc.	67.77	4.91
Breaks between work	66.49	5.30
No dress code	63.43	6.30
I can work from wherever	57.85	5.67
Less stressful	56.47	5.80
I learn interesting things	45.71	6.45
Flexibility for family	41.76	4.32
Immediate pay	35.34	5.79
I can make more money	22.85	5.36
Physical and mental constraints make it hard for me to find alternative work	15.45	4.25

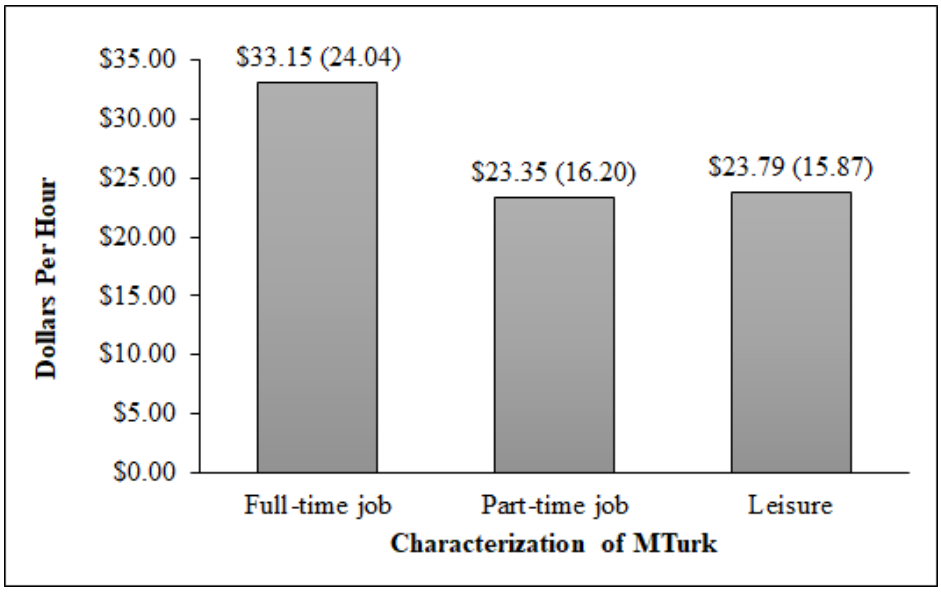
Note: Percent choosing represents the percentage of the sample that selected each reason. Average ranking reflects how important people who selected each reason felt it was. Participants only ranked reasons they selected. Numbers closer to 1 indicate more importance.

After people ranked reasons for being on MTurk, we asked two questions about wages. First, we asked people to consider the reasons they selected for spending time on MTurk and to report how much another job would have to pay for them to give up MTurk and take the other job. The average hourly wage people reported was $25.13 (*SD* = $18.68; median = $20.00). Because outliers may skew this number, we winsorized all values greater than three standard

deviations from the mean. Doing so had little effect on the results. In the winsorized data, people said they would need more than $24 per hour to take an alternative job over MTurk (*M* = $24.41, *SD* = 17.03).

As shown in Figure 3, whether people characterized their time on MTurk as a full-time job, part-time job, or a form of paid leisure affected how much they said they would need to earn in order to take another job over MTurk. People who characterized MTurk as full-time work reported that they would need to make more than $34.00 per hour (median = $22.25). People who characterized MTurk as part-time work or as paid leisure both said they would need to make more than $24.00 per hour (part-time median = $18.50; leisure median = $20.02). Altogether, these figures suggest that some of the reasons people have for spending time on MTurk may augment their actual earnings.

Figure 3. Hourly wage people said they would need to trade MTurk for another job.

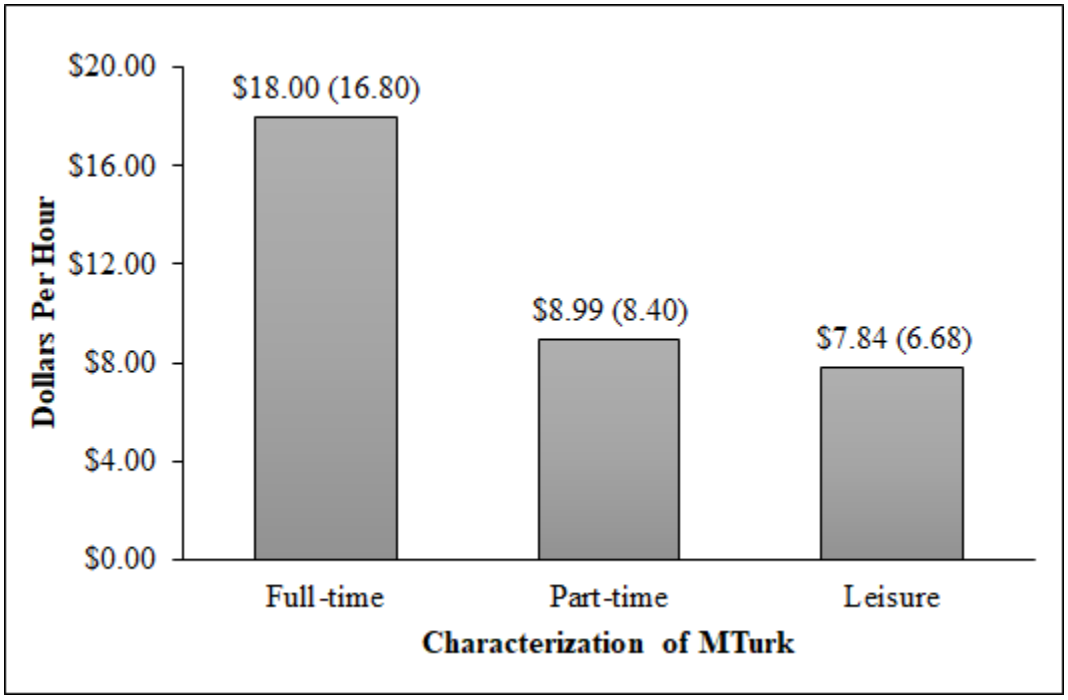


Note: Values are based on winsorized distributions. Numbers in parentheses are standard deviations.

The second wage question we asked people was, “Based on your experience, how much money would you estimate people can earn per hour through taking HITs on MTurk?” The average hourly wage people reported was $9.94 (*SD* = $13.01; median = $6.51). Once again, winsorizing values greater than three standard deviations from the mean had little effect on the results. In the winsorized data, people said it is possible to earn $9.15 (*SD* = $9.07) per hour.

As shown in Figure 4, whether people characterized MTurk as a full-time job, part-time job, or a form of paid leisure affected how much they said people could earn per hour. People who characterized MTurk as full-time work reported that it is possible to earn about $18.00 per hour (median = $10.14). Meanwhile, people who characterized MTurk as part-time work or as paid leisure said it was possible to earn around eight or nine dollars per hour (part-time median = $6.97; leisure median = $6.08). All these estimates are notably higher than past reports about hourly wages on MTurk.

Figure 4. Hourly wage people said it is possible to earn on MTurk.

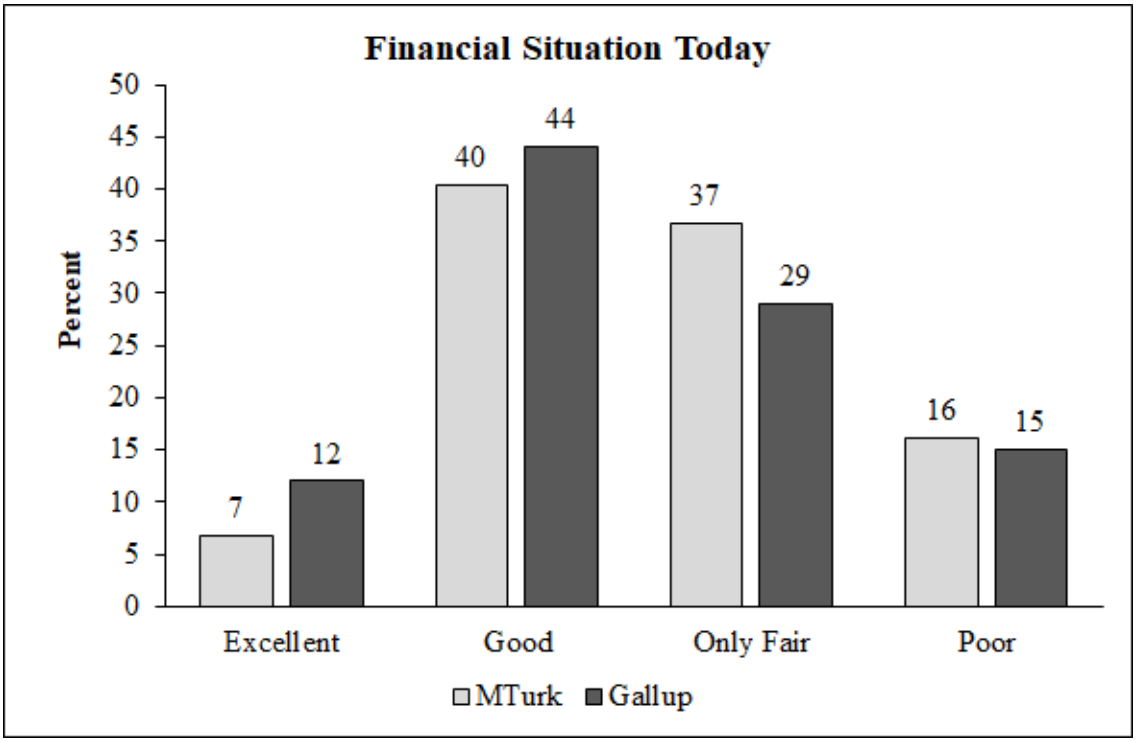


Note: Values are based on winsorized distributions. Numbers in parentheses are standard deviations.

Financial Situation and Satisfaction with MTurk

To assess people's current financial situation, we used polling questions from Gallup. The results revealed that the financial situation of people on MTurk looks quite similar to the US population. For example, while 15% of the general population describes their financial situation as "poor" the number on MTurk is 16% (see Figure 5). People on MTurk were slightly more likely to describe their financial situation today as "only fair" compared to those in the general population.

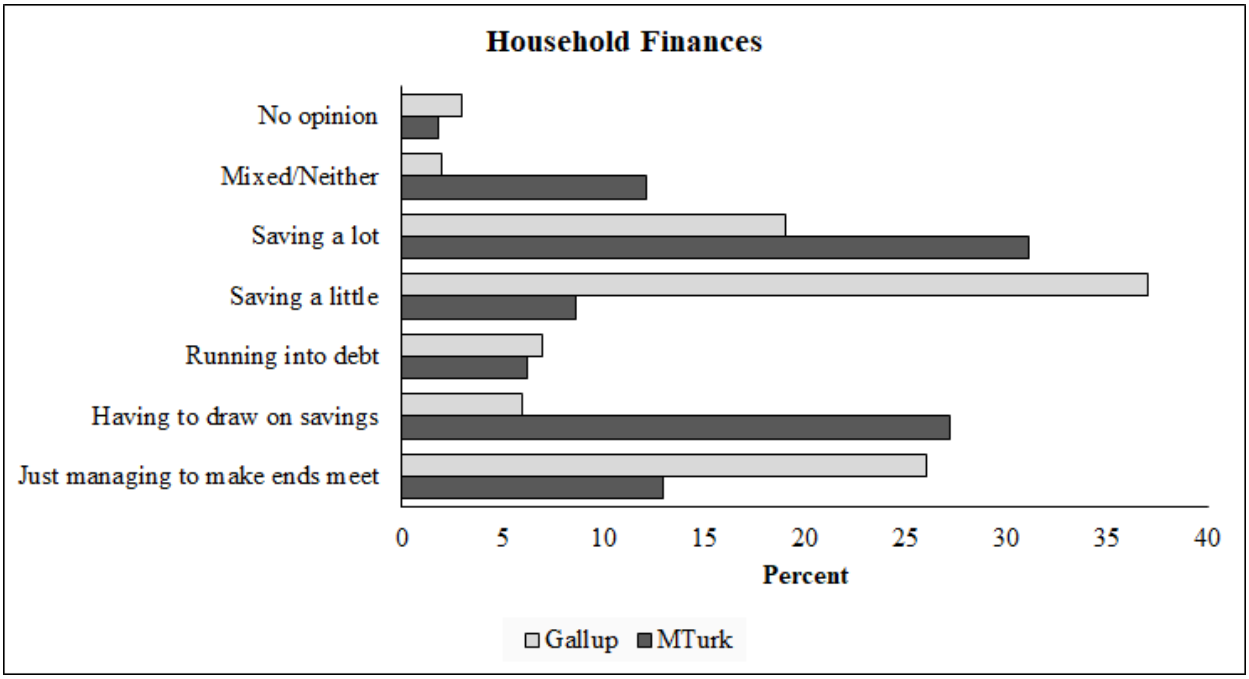
Figure 5. The financial situation of people on MTurk compared to the U.S. population.



In another question borrowed from Gallup, we asked people to describe their household finances. Although there were differences in the answer choices people used to describe their

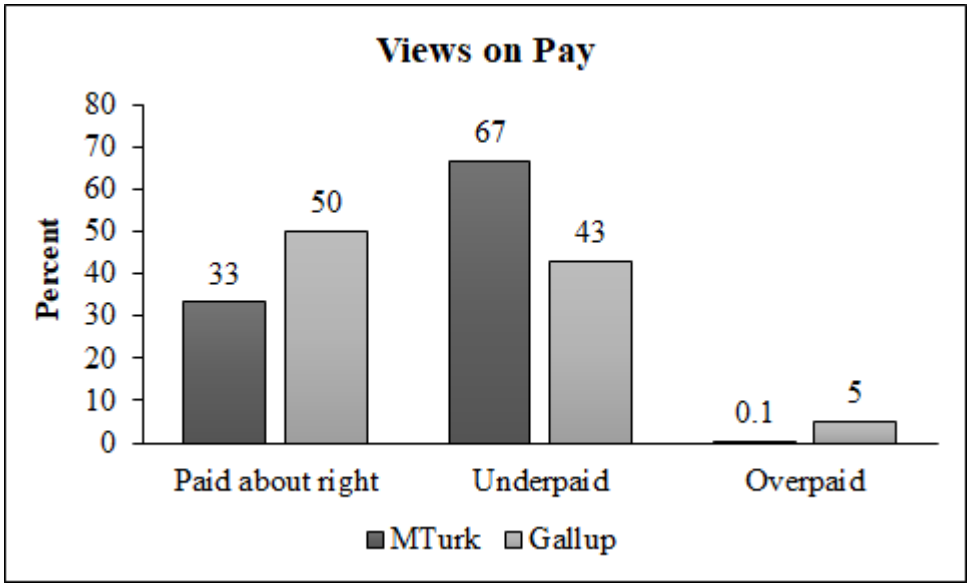
situation, Figure 6 reveals generally similar household finances for people on MTurk and those in the general population. For example, people on MTurk were more likely than the general population to say they were "saving a lot" but much less likely to say they were "saving a little." Similarly, while people on MTurk were more likely to say they were "having to draw on savings," they were only half as likely as the general population to say they were "just managing to make ends meet."

Figure 6. How people on MTurk and in the US population describe their household finances.



Next, we asked people how they feel about pay on MTurk. According to Gallup, between 40 and 50 percent of Americans routinely report they are underpaid. On Mechanical Turk, this number was 67%. In addition, whereas 50% of Americans say they are paid about right, the number on MTurk was 33% (see Figure 7).

Figure 7. How people on MTurk and in the U.S. population feel about their pay.



After the questions above, we asked people how satisfied they are with the amount of money they make on MTurk and the amount of stress that work entails. Both are job-related factors that Gallup has tracked in the U.S. population for more than two decades. As shown in Table 5, about 60% of people on MTurk reported being satisfied with the amount of money they earn. Although this number is less than the general population (78%), the idea that most people on MTurk are satisfied with their wages runs completely counter to the current conversation about wages and the ethics of MTurk.

For the question asking about stress, 90% of people said they were satisfied with the stress MTurk entails; only 1% of people said they were completely dissatisfied. In the general population, 73% of people say they are satisfied with their job-related stress. Even though comparing the stress of MTurk to the wide array of jobs Americans perform daily is an imprecise

comparison, it is worth noting that the vast majority of people on MTurk reported less stress than people in other jobs.

Table 5. Satisfaction with money earned and job-related stress. Numbers are percentages.

Variable	MTurk	Gallup
Money earned		
Completely satisfied	8	37
Somewhat satisfied	51	41
Somewhat dissatisfied	36	12
Completely dissatisfied	5	8
Job stress		
Completely satisfied	50	34
Somewhat satisfied	40	39
Somewhat dissatisfied	9	19
Completely dissatisfied	1	8

Finally, we asked people a direct question: is MTurk part of the problem or part of the solution to your financial needs? Approximately 88% of people said MTurk was part of the solution whereas just 3% said it was part of the problem. Nine percent said, “It’s more complicated than that/A little bit of both.”

The results from this survey present several empirical findings that are at odds with the main criticisms of Mechanical Turk.

Study 2: Are Requesters on MTurk Fair?

Participants and Procedure

The sampling process for Study 2 was similar to Study 1. We collected data from a total of 2,068 people. Participants were recruited based on response rates across each level of experience observed in Study 1. All data were collected over a two-week period at the end of March 2020. Response rates were similar to those in Study 1, except for the <100 HITs group. Response rates in this group were slightly lower than in Study 1 which led us to invite an additional 910 people to reach our quota in this group. We omitted demographic measures from Study 2 in order to keep the survey short.

Materials

The survey contained seven questions asking people about the fairness of requesters on MTurk and how commonly they experience rejections and disturbing content. The first two questions asked people to evaluate the fairness of requesters on Mechanical Turk and the fairness of employers outside of MTurk. While we hoped to replicate past work by randomly assigning the order in which people answered these two questions (Horton, 2011), a programming mistake led all participants to answer the question about MTurk requesters before the question about employers outside of MTurk. After the first two questions, we asked people if they have ever worked on other microtask platforms. For those who answered ‘Yes,’ we asked how MTurk compares to the other platforms in terms of fairness. Finally, we asked people what percentage of HITs that they submit on MTurk are rejected and what percentage of HITs contain disturbing content.

Results

We asked people what percentage of requesters on MTurk and what percentage of employers outside of MTurk have generally treated them honestly and fairly. Overall, people reported that requesters on MTurk generally treat them fairer and more honestly than employers outside of MTurk. On average, people reported that 84.44% (*SD* = 18.91) of requesters on MTurk treat them fairly (median = 90.80%) while 74.24% (*SD* = 20.80) of employers outside of MTurk do the same (median = 79.25%).

Next, we asked people if they have ever worked on a microtask platform other than MTurk; almost a quarter of people reported they had (24.6%). To this group, we asked where people are generally treated the most fairly. Fifty percent of people in this group said they are treated the most fairly on MTurk, while 17.5% said they are treated more fairly on other microtask platforms, and 32.5% said their treatment is about the same regardless of platform.

After asking questions about fairness, we asked people how many hours they spend on MTurk per week, how often they have work rejected, and how often they are exposed to disturbing content. After removing one person who said they spent 231,695 hours per week, the average people reported was 8.40 hours (SD = 10.16; median = 5.00). In response to a question asking what percentage of HITs they have rejected, people reported an average of 5.45% (SD = 14.03)[2]. This average, however, appears inflated by several outliers who said they have 100% of HITs rejected (an implausible outcome). Examining the median revealed that people reported just 1.00% of their HITs are rejected. Interestingly, when we asked people what percentage of

[2] For the first 562 participants, the sliding answer scale only allowed people to choose a whole number for their answer (e.g., 10%). After repeated messages from participants, however, we changed the scale to accommodate fractional values between whole numbers (e.g., 10.4%). This change was driven by participants who told us that the most accurate answer for the number of HITs they have rejected was a value between 0% and 1%.

their rejections were unfair, the average was just 37.02% (*SD* = 39.61; median = 16.65%), suggesting that people believe most of the few rejections they receive are justified.

In response to a question asking what percentage of HITs contain disturbing content, people reported an average of 6.85% (*SD* = 15.48). Once again, this average appears inflated by outliers; the median revealed that 1.00% of HITs contained disturbing content.

Finally, we asked people how often they have upsetting experiences on MTurk. The answer scale ranged from 1 (*Almost never)* to 7 (*Every day I Turk*) with a midpoint of 4 (*About half the days I Turk*). People reported that upsetting experiences occur infrequently (*M* = 1.63, *SD* = 1.16).

Overall, these results indicate that people on MTurk generally see requesters as fair and honest, and view requesters on MTurk as fairer and more honest than employers outside of MTurk. In addition, rejected HITs and disturbing content represent a small minority of the HITs people complete.

General Discussion

Social scientists need access to people and data in order to understand human behavior. In the last decade, Amazon's Mechanical Turk has made it easier than ever for academic researchers to collect such data. But recently, several people have raised concerns about the ethics of MTurk. Primary among these concerns is whether people on MTurk are financially vulnerable, what the experience of completing tasks on the platform is like, and the amount of money people earn. The data we reported speak directly to these concerns. Our probability-based surveys of the U.S. MTurk population revealed that most people do not view MTurk as work but as a form of paid leisure. In addition, people on MTurk are: 1) about as well off financially as the general population, 2) less stressed than people in the American workforce, 3) pleased with the

flexibility and benefits MTurk offers over more traditional work, 4) earning far more than previous reports have indicated, and 5) inclined to say that requesters on MTurk are fairer and more honest than employers outside of MTurk. In fact, a majority of people with experience on multiple microtask platforms report that MTurk is where they are treated the most ethically. In addition to this, we reported wage data that suggests people would require more than $25 per hour in order to give up the flexibility MTurk allows and take another job. Altogether, we believe, these findings provide important information for researchers to consider when deciding whether it is ethical to conduct behavioral research on Mechanical Turk.

Is Mechanical Turk Exploitative?

The suggestion that people on MTurk are exploited comes from several sources but has been made most prominently in popular press articles. In these articles, journalists have suggested that people on MTurk are disproportionately poor, disproportionately disabled, and working on MTurk because they have no other way to make ends meet. From this perspective, people on MTurk are a captive audience and the tasks that requesters post are coercive because they take advantage of people who have no choice but to accept the work regardless of how little it pays.

To examine whether people on MTurk are disproportionately poor or disabled, we asked several questions assessing people's financial situation and why they complete tasks on MTurk. In structuring these questions, we chose items from a 2018 Gallup poll conducted with a representative sample of the U.S. population so that we could compare the amount of financial struggle measured on MTurk to that in the general population. In the first question, we asked people whether their financial situation today is excellent, good, only fair, or poor. We reasoned that if people are driven to work on MTurk because they are financially disadvantaged, then

people on MTurk should be more likely to describe their financial situation as poor relative to the general population. This was not the case. Only 16% percent of people on MTurk described their financial situation as poor compared to 15% in the general population.

Another question we asked required people to choose the best way to describe their finances. The response option chosen by those struggling the most financially was "just managing to make ends meet." People on MTurk were less likely to select this response (13%) than those in the general population (26%). In addition, people on MTurk were more likely to report that they are "saving a lot" (31%) than the general population (19%). Finally, as another way to examine whether people are drawn to MTurk because they are struggling financially, we asked people whether they use money from MTurk for essential or non-essential expenses. Only a minority of people (21%) said they use MTurk earnings for essential expenses. Hence, these data all support the conclusion that people on MTurk are not disproportionately poor, but instead are average people who use MTurk to supplement their income for non-essential spending. These data converge with other work that shows people on MTurk are not more likely to be unemployed than the general U.S. population (see Robinson, Litmna et al., 2020).

Beyond finances, another claim about exploitation on MTurk is that people are disproportionately disabled. We asked people to select, from a list, all the reasons why they work on MTurk. One reason was because physical or mental disabilities make it hard to find work elsewhere. Fifteen percent of people said this was a reason for spending time on MTurk. However, the percentage of people with a disability in the U.S. population is 26% (Centers for Disease Control and Prevention, 2019), suggesting that MTurk does not attract a disproportionate number of disabled people.

Finally, another finding from our survey that speaks to concerns about exploitation is that 52% of people on MTurk do not characterize what they do as a job but as paid leisure. Another approximately 40% characterize MTurk as a sort of part-time job, and less than 10% of people view MTurk as a full-time job. Combined with the data on people's financial situation, this data converges with other evidence to show that people on MTurk are similar to the general population in most ways, and that, altogether, people on Mechanical Turk are simply people who choose to spend some of their free time making extra money online, not the poorest and most vulnerable members of society.

Are Requesters on MTurk Abusive?

The second argument made against MTurk is that it is abusive. Several writers have described the alleged abusive nature of the platform, chronicling stories of mistreatment and referring to MTurk as a "living hell" and an "online sweatshop" (Semuels, 2018; Graham, 2010). These concerns date back to at least 2010, when academic researchers began expressing concern about exploitation of people on MTurk at academic conferences (Horton, 2010). Similar concerns were reported in prominent places like *Newsweek* around the same time (Zittrain, 2009). One issue that has been at the center of claims about abuse is requesters' ability to reject work without giving a justification. Workers are sensitive to rejections because in addition to receiving no pay they also suffer some harm to their reputation. In addition to rejecting work, requesters can provide unclear or confusing instructions, post tasks with broken links, fail to properly warn people of disturbing content, and misrepresent how long a task should take. In all these cases, workers may not be properly compensated for their time and there are little to no consequences for the requester. Thus, the system is structurally designed with the potential for abuse.

While there is no question that abusive practices sometimes take place on MTurk, what is not known is at what scale. In our first survey we asked people about the overall stress they experience on MTurk. Like other questions in the survey, we borrowed the question wording from Gallup because it allowed us to compare the stress that people report on MTurk to the job-related stress experienced by people working in the U.S. population. The level of stress reported by people on MTurk was substantially lower than that reported by people in the U.S. workforce. Specifically, the largest segment of people on MTurk (50%) reported being "completely satisfied" with the amount of stress they experience. Comparatively, in the U.S. population, only 29% of people report being completely satisfied. The next largest segment of people on MTurk reported being somewhat satisfied (40%), and only 10% of people reported being either somewhat dissatisfied (8.6%) or completely dissatisfied (1.5%). In contrast, 33% of people in the U.S. population report being dissatisfied with the stress of their current job.

In our second survey, we asked people whether requesters on MTurk are fair and honest. People's responses revealed a positive view of requesters, with most people reporting that requesters on MTurk are fairer and more honest than employers outside of MTurk. This result closely replicates a study that was conducted almost ten years ago (Horton, 2011), with the exception that people rated requesters on MTurk even more positively in the current study. This result is somewhat surprising given that the original study was conducted before MTurk was widely adopted by academic or industry researchers. Given the larger and more diverse group of requesters today one might expect more abusive practices, but the opposite appears true.

The second thing we asked people was on which microtask platform they experience the most ethical treatment. A majority of people who worked on multiple microtask platforms said MTurk is where they receive the most ethical treatment. Finally, we asked people how often they

encounter rejections and disturbing content on MTurk. Based on median numbers, people reported that less than 1% of HITs they complete are rejected and less than 1% of HITs contain disturbing content. These data are inconsistent with the claim that there is widespread abuse on Mechanical Turk. In fact, these data converge with other lines of evidence to indicate that unscrupulous practices by requesters are relatively uncommon. For example, researchers appear to reject less than 1% of all submitted tasks (Litman & Robinson, 2020). Additionally, the time that requesters advertise for surveys and other tasks typically ends up overestimating how long it takes people to complete the task, not underestimating (Litman, Robinson, Rosen et al., 2020). And finally, contrary to some reports, it appears few people find the content of surveys they are asked to complete disturbing. In a poll of more than 10,000 people on MTurk, just four percent said they find the content of tasks more distressing than activities in daily life (Litman & Robinson, 2020). Importantly, however, in a follow up survey with the four percent of people who said MTurk is distressing, 75% of people said the benefits of MTurk outweigh the potential costs. This is the kind of information that is commonly used by institutional review boards when evaluating whether a research study is ethical.

In addition to asking people about stress and abusive requester practices, our survey provided another way to assess whether people are dissatisfied with their experience on MTurk: we asked people if they would take another job over MTurk if one were available. More specifically, we asked people whether they would trade their time on MTurk “earning what you currently earn” for a job in retail or food service “earning typical pay for that job.” Most people said they would prefer MTurk, regardless of whether the time committed was full-time, part-time, or something to do in their leisure time. This result runs counter to claims that people seek out MTurk as a way to make money because “there are few decent-paying jobs that can be done

at will" (Newman, 2019). Rather, the current data show that even when there is a job in retail or food service, people prefer MTurk. Perhaps most surprising, the pattern of results for this question revealed that the people who spend the most time on MTurk—the 8% of people who said MTurk is a full-time job—were also the most likely to prefer MTurk over alternative jobs. Almost 75% of people who say they use MTurk as a full-time job said they would prefer to work on MTurk over taking an alternative job in retail or food service. Like the data assessing whether MTurk is exploitative, the overall pattern of our data suggests that people's time on MTurk is not characterized by widespread abuse.

How Much Do People Make?

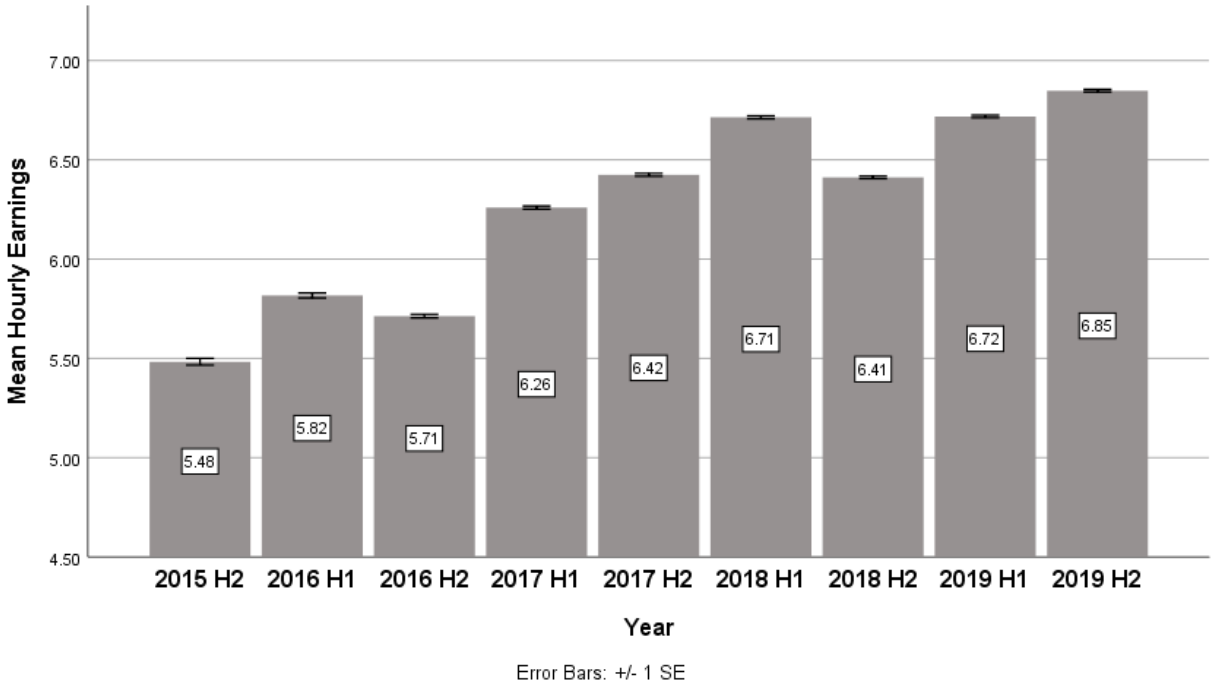
The final criticism often aimed at MTurk concerns how much people earn. In recent years, MTurk has been criticized as a platform that offers exceedingly low wages—so low in fact that they render the entire platform unethical. We estimated wages using two different methods. First, we asked people to report how much they think it is possible to earn per hour on MTurk. After removing outliers, the overall mean was greater than $9.00 per hour. In addition, examining estimated wages across people's self-characterization of how they use MTurk—full-time work, part-time work, paid leisure—revealed large differences. People who report working on MTurk full-time reported that it is possible to earn about $18.00 per hour while people who report working on MTurk part-time or as paid leisure both reported it is possible to earn about $8.00 or $9.00 per hour. All these estimates are notably higher than past research estimating wages on MTurk.

The second approach we used to estimate wages was to examine data from more than 26 million completed assignments in the CloudResearch database. This data comes from studies posted by more than 4,500 requesters, completed by virtually all workers on MTurk, and

stretching over a span of four years. We computed average hourly wages by subtracting the start time of each HIT from the end time, and then dividing the total time required to complete each HIT by the payment offered for the HIT. Complete details for our wage analysis, including the strengths and limitations of this estimation approach, are included in Supplementary Materials.

The mean hourly wage for each half-year interval from mid-2015 to 2019 is plotted in Figure 8. As shown, average wages tended to increase over time. In the second half of 2015, the mean hourly wage was $5.48 per hour. By the end of 2019, however, the average hourly wage had risen by 25% to $6.85 per hour.

Figure 8. Mean hourly wages over time.



Note: H1 = the first half of the year, January through June. H2 = the second half of the year, July through December.

Although the wage numbers we report are higher than what is commonly assumed in conversations about MTurk, they are actually in line with a number of past reports.For example, in a study conducted by Pew, about half of people on MTurk reported earning more than $5.00 per hour including 8% of people who reported earning more than $8.00 per hour (Hitlin, 2016). In another large, yet unrepresentative sample of people on MTurk, people reported an average hourly wage of $5.55 (Berg, 2016). And, in a paper exploring strategies people use to maximize their earnings on MTurk, the self-reported hourly wage was just over $5.00 per hour (Kaplan, Saito, Hara, & Bigham, 2018). When considering this evidence altogether, the most cited statistic about wages on MTurk—that people earn $2.00 per hour (Hara et al., 2018)—begins to look like an outlier.

It is not clear why the self-reported wages in our survey were higher than the wages from our database. One possibility is that the discrepancy arises from the way the question was framed. Specifically, our question asked people how much a typical person can make per hour, rather than asking people how much they personally make per hour. Questions about potential wages may rely on the assumption of maximal efficiency, meaning that people assume workers are trying to earn as much as possible. However, past research shows that many people treat MTurk tasks more like leisure, simultaneously watching TV, texting, and listening to music at least some of the time (Chandler, Mueller, & Paolacci, 2013). Because many people in our sample reported that MTurk is a form of paid leisure, they may have assumed that someone trying to maximize their efficiency could earn more. Another possibility is that the discrepancy stems from participants' practice of accepting multiple HITs at one time and then completing them as time allows, a practice that may deflate wage estimates (see Supplemental Materials for more details).

Beyond estimating wages, our survey sought to examine what people find valuable about MTurk when compared to more traditional jobs. Because gig economy work often entails trading the stability and rigidity of traditional jobs for the variability and flexibility of gig jobs, we expected people to report that flexibility was important. This was the case. Ninety four percent of people indicated that they prefer MTurk over other jobs because they can work from home. In addition, 92% of people said they work on MTurk because they can choose their own hours, 71% because they do not have to commute, and 68% because they do not have a boss or supervisor. Interestingly, 23% of people said they work on MTurk because they can make more money than in a traditional job.

Although past research has identified hidden factors that drag down people's wages—time spent searching for tasks, time spent on tasks that are returned or rejected (Hara et al., 2018)—we reasoned that the things people value about MTurk may actually have hidden value that augment their wages. For example, the U.S. Census Bureau estimates that in 2018 the average American commuted 28 minutes to work one-way (Ingraham, 2019). When people spend time on MTurk there is, obviously, no need to commute. In addition, another highly ranked reason people gave for completing tasks on MTurk was the freedom to set one's own schedule. Autonomy in choosing when to work is not a trivial issue. Major surveys of retail and food service employees indicate that hourly employees often face discrimination in the hours they are assigned and sometimes receive less than 72 hours notice about the schedule they are supposed to work (Schneider & Harknett, 2019). Such unpredictable and constantly changing schedules are a major burden for both employees and their families (Cain Miller, 2019), which may be why 42% of people in our survey listed flexibility for family as a reason they value MTurk.

To assess how important these factors are, we asked people how much another job would have to pay for them to quit MTurk and take a job that did not have these conveniences. The average hourly wage people reported was just over $25 per hour (median $20). Among people who said they use MTurk as a form of full-time work, the number was greater than $33.50 per hour. Overall,the responses to this question suggest that the vast majority of people may agree with the sentiment expressed by a woman who wrote the following on Reddit in response to an article in *The New York Times* that characterized MTurk as a place for extremely low wage work:

> "As someone who is able to earn (on a horrible day) $6 an hour with a 21 month old nursing on my tatas, MTurk works for someone like me. I'm able to earn $6 an hour, NOT have to pay another $10/hour for quality child care and then able to clean my house, keep up with teenagers, do laundry, not pay gas or transportation costs etc.; I feel like my meager $6 on a bad day is about $20/hour for any other working mother" (fairlybrash, 2019).

Hence, our data provide a picture of people on MTurk and how much they earn that is very different from the commonly accepted narrative. Rather than working on MTurk to make ends meet or because they have no other options, it appears most people chose to spend time on MTurk because it provides multiple advantages over other, more traditional ways of making money in their spare time. In addition, a majority of people on MTurk say they prefer MTurk over other jobs and would not switch from working on MTurk to another job unless it paid at least $20 per hour. Perhaps most telling, approximately 90% of people say MTurk is part of the solution to their financial situation, not part of the problem.

Implications

What we are not saying. We are not saying Mechanical Turk is a perfect platform with no room for improvement. In fact, we realize that for many people time on MTurk entails earning near minimum wage while sitting at a computer trying to quickly find and complete a variety of tasks. A common complaint from people on MTurk is that competition makes tasks hard to find. Mechanical Turk could do more to improve the user experience and to provide people with information about the reputation of requesters, something many worker-run groups provide (e.g., TurkerView, 2020). As a platform that seeks to connect requesters with people willing to complete tasks, improving the efficiency of these connections will result in a better experience for everyone who uses the platform.

Next, we are not saying that negative experiences do not occur on MTurk. Our data show that a small percentage of HITs are rejected, and a small percentage contain disturbing content. Even though these events occur infrequently, they can still be upsetting. People may find dealing with requesters frustrating or receiving a rejection that feels unwarranted aggravating.

Additionally, we are also not saying that what people currently earn on MTurk is all that they should earn. Requesters post a wide variety of tasks that pay various amounts of money. While two-thirds of people (67%) reported that they feel underpaid, close to the same number (59%) reported that they are satisfied with the money they earn. These seemingly discrepant numbers may reflect awareness on the part of workers that while many tasks could pay more money, the pay people do earn has some relationship to the time, effort, and skill required to complete the tasks. More importantly, our data show that market forces have caused wages on MTurk to increase over time. Part of this increase is likely due to the resourcefulness of the Turking community which has developed systems for identifying and avoiding bad requesters

and finding lucrative tasks (Robinson, Rosenzweig, & Litman, 2020). But, at least part of this increase may also be due to the influence of academic researchers—people who have spent a lot of time discussing, debating, and writing about wages on MTurk.

Finally, the goal of this paper was not to provide answers to ethical questions, and instead was to present rigorous empirical data that can inform the conversation about ethics. We are not saying that MTurk is or is not an ethical platform. We present empirical data that speaks to many ethical questions that have come up about MTurk, and believe it is important for researchers to see this data so they can make more informed decisions about whether or not it is ethical for them to use MTurk.

What we are saying. We are saying that the data show Mechanical Turk is a far better place for people to earn some money and for behavioral researchers to collect data than the current conversation about MTurk would suggest. The ethical concerns raised most often by researchers and journalists are not borne out in representative data. Furthermore, most people on MTurk seem pleased to have an avenue for making money in their spare time. People report that they work on MTurk because of its flexibility, that they make more money than commonly assumed, and that they would not trade MTurk for another job unless it paid near $25 per hour. One especially telling fact about MTurk is that a majority of people who report experience with multiple microtask platforms say MTurk is where they experience the most ethical treatment. With this in mind, we close with ways requesters can promote wage equity on MTurk.

Ways to Promote Wage Equity on MTurk

Although most conversations about wages on MTurk focus on how much people earn and what constitutes a fair wage, there is more to wage equity than dollars per hour. For example, even though people on MTurk are largely anonymous to requesters, there is a gender wage gap.

Men earn approximately 10.5% more per hour than women (Litman, Robinson, Rosen et al., 2020). The causes of the gender wage gap are numerous and complicated, but unlike the world outside MTurk requesters could virtually eliminate this gap by placing a gender quota onto all eligible HITs. Using quotas to balance the number of men and women who can complete each HIT would remove the advantage that superwokers, who tend to be men, have when quickly identifying and accepting high paying HITs. Hence, a small change in the practices of requesters could have a big effect on eliminating pay disparities on MTurk.

In addition to eliminating the gender wage gap, requesters who seek to raise compensation on MTurk can take a few simple steps to ensure that wage increases are equally spread among people. As previously mentioned, superworkers are the first people to find and accept high paying HITs because they are the most likely to rely on scripts and other tools to find such HITs (Kaplan, Saito, Hara, & Bigham, 2018; Litman & Robsinson, 2020). This means that if all a researcher interested in increasing wages did when posting his studies was increase the pay per survey to a minimum wage or a living wage, the people most likely to wind up in his study's sample—and to reap the benefits of higher wages—would be superworkers. Because superworkers are not naive to many common measures and because superworkers are the people already earning the most on MTurk, simply increasing wages is likely to have unintended and negative consequences. Instead, researchers could change their sampling practices to exclude the most experienced participants (see Robinson et al., 2019) or stratify their sample by worker experience. Stratifying the sample by opening multiple HITs to people with different levels of experience would ensure that higher wages are evenly spread among workers while also controlling the number of experienced participants who end up in a study's sample.

Finally, another way requesters can increase the wages people earn is to design better surveys. Most academic researchers receive extensive training in domains like statistics, data analysis, and the theories and knowledge of their field. Yet, few researchers receive advanced training in survey design. People working in industry may have even less training than academics which leads to a situation in which most MTurk tasks are designed by people who are technically proficient in many domains but have less experience when it comes to study design. The consequences of this situation may be manifold. People on MTurk may encounter task descriptions or instructions that are lacking important details or confusing. Study links and audio or visual media may not work. Research materials may contain words that people with less than a high school education do not understand. And, researchers who are pressed for time may not sufficiently pilot their study to know how long it will take and therefore what amount of pay is appropriate. Notably, all these negative consequences fall on the participant. Thus, designing more user-friendly surveys might be one way to facilitate higher wages on MTurk.

Conclusion

Since its introduction to behavioral researchers, Mechanical Turk has provided a fast, efficient, and affordable way to find research participants. Yet, almost as quickly as researchers began celebrating the benefits of MTurk, they also worried about its consequences. While some saw MTurk as an affordable place for research, others worried about exploiting people. While some sought to explore data quality and the range of tasks participants were willing to complete, others worried about ensuring researchers were using the platform ethically. In the years since, the voices claiming that MTurk is unethical have grown louder. Unfortunately, however, the claims made against MTurk have not been backed by solid empirical data. Our research reveals that the claims made most often—people on MTurk are poor and vulnerable, there is widespread

abuse, people earn just pennies—lack empirical evidence. While our research is not the final word about the ethics of MTurk, we believe our data offer important information for researchers to consider when deciding whether it's ethical to conduct behavioral research on MTurk.

References

Anderson, C. A., Allen, J. J., Plante, C., Quigley-McBride, A., Lovett, A., & Rokkum, J. N. (2019). The MTurkification of social and personality psychology. *Personality and Social Psychology Bulletin, 45,* 842-850. https://doi.org/10.1177/0146167218798821

Arch, J. J., & Carr, A. L. (2017). Using Mechanical Turk for research on cancer survivors. *Psycho-oncology, 26,* 1593-1603. https://doi.org/10.1002/pon.4173

Arechar, A.A., Gächter, S. & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics, 21,* 99–131. https://doi.org/10.1007/s10683-017-9527-2

Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. (2017). Turking overtime: How participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. *Journal of the Economic Science Association, 3,* 1–11. https://doi.org/10.1007/s40881017-0035-0

Bates, J. A., & Lanza, B. A. (2013). Conducting Psychology Student Research via the Mechanical Turk Crowdsourcing Service. *North American Journal of Psychology, 15,* 385-394.

Berg, J. (2015). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comparative Labor Law and Policy Journal, 37,* 543.

Buchheit, S., Doxey, M. M., Pollard, T., & Stinson, S. R. (2018). A technical guide to using Amazon's Mechanical Turk in behavioral accounting research. *Behavioral Research in Accounting, 30,* 111-122. https://doi.org/10.2308/bria-51977

Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6,* 3-5. DOI: 10.1177/1745691610393980

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science, 13,* 149-154. https://doi.org/10.1177/1745691617706516

Casey, L. S., Chandler, J., Levine, A. S., Proctor, A., & Strolovitch, D. Z. (2017). Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection. *SAGE Open, 7,* 1–15. https://doi.org/10.1177%2F2158244017712774

Cain Miller, C. (2019). *How unpredictable work hours turn families upside down.* The New York Times. https://www.nytimes.com/2019/10/16/upshot/unpredictable-job-hours.html

Centers for Disease Control and Prevention. *Disability impacts all of us.* https://www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html

Chan, C., & Holosko, M. J. (2016). An overview of the use of Mechanical Turk in behavioral sciences: Implications for social work. *Research on Social Work Practice, 26,* 441-448. https://doi.org/10.1177/1049731515594024

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46,* 112-130. https://doi.org/10.3758/s13428-013-0365-7

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science, 26,* 1131-1139. https://doi.org/10.1177/0956797615585115

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods, 51,* 2022-2038. https://doi.org/10.3758/s13428-019-01273-7

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology, 32,* 347-361. https://doi.org/10.1007/s10869-016-9458-5

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics.* https://doi.org/10.1177/2053168015622072

Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods, 7,* 613-628. https://doi.org/10.1017/psrm.2018.10

Crowston K. (2012) Amazon Mechanical Turk: A research tool for organizations and information systems scholars. In A. Bhattacherjee and B. Fitzgerald (Eds.), *Shaping the Future of ICT Research. Methods and Approaches. IFIP Advances in Information and Communication Technology,* vol 389. Springer, Berlin, Heidelberg

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE, 8,* e57410. 10.1371/journal.pone.0057410

Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February). Demographics and dynamics of Mechanical Turk workers. In Proceedings of the eleventh ACM international conference on web search and data mining (pp. 135-143).

Fairlybrash. (2019, November 18). *I'm Andy Newman. I write about jobs for The New York Times. I spent a few days on Mechanical Turk, a digital work marketplace run by Amazon where anyone can make small amounts of money (as little as 1 cent) doing digital jobs. I sucked at it and made 97 cents an hour. Ask me anything about turking.*

[Online forum response]. Reddit. https://www.reddit.com/r/IAmA/comments/dy5h2y/im_andy_newman_i_write_about_jobs_for_the_new/

Federal Election Commision (2019). *Browse disbursements, campaign finance data*. https://www.fec.gov/data/disbursements/?data_type=processed&recipient_name=Mechanical+Turk&two_year_transaction_period=2020

Fordsham, N., Moss, A. J., Krumholtz, S., Roggina, T., Jr, Robinson, J., & Litman, L. (2019, October 30). Variation among Mechanical Turk Workers Across Time of Day Presents an Opportunity and a Challenge for Research. https://doi.org/10.31234/osf.io/p8bns

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon Mechanical Turk: Gold mine or coal mine? *Computational Linguistics, 37,* 413-420. https://doi.org/10.1162/COLI_a_00057

Glass, M. (2020). Popular among Cornell professors, Mechanical Turk continues to face ethical questions. The Cornell Daily Sun. https://cornellsun.com/2020/02/21/popular-among-cornell-professors-mechanical-turk-continues-to-face-ethical-questions/

Graham, F. (2010). *Crowdsourcing work: Labour on demand or digital sweatshop?* BBC News. https://www.bbc.com/news/business-11600902

Grim, R., & Lacey, A. (2020). *Pete Buttigieg's campaign used notoriously low-paying gig-work platform for polling.* The Intercept. https://theintercept.com/2020/01/16/pete-buttigieg-amazon-mechanical-turk-gig-workers/

Grysman, A. (2015). Collecting narrative data on Amazon's Mechanical Turk. *Applied Cognitive Psychology, 29,* 573-583. https://doi.org/10.1002/acp.3140

Hall, M. P., Lewis, Jr., N. A., & Chandler, J. (2020). Conducting longitudinal research on Amazon Mechanical Turk. In L. Litman and J. Robinson (Eds)., *Conducting Research on Amazon Mechanical Turk and Beyond*. Sage Publishing. Thousand Oaks: CA

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018, April). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-14). https://doi.org/10.1145/3173574.3174023

Hauser, D. J., Paolacci, G., & Chandler, J. J. (2019). Common concerns with MTurk as a participant pool: Evidence and solutions. In F. R. Kardes, P. M. Herr, & N. Schwarz (Eds.), *Handbook of Research Methods in Consumer Psychology*. New York: Routledge.

Hauser, D.J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48,* 400–407. https://doi.org/10.3758/s13428-015-0578-z

Hitlin, P. (2016). Research in the crowdsourcing age, a case study. Pew Research Center. Available at: https://www.pewresearch.org/internet/2016/07/11/research-in-the-crowdsourcing-age-a-case-study/

Horton, J. J. (2011). The condition of the Turking class: Are online employers fair and honest? *Economics Letters, 111,* 10-12. https://doi.org/10.1016/j.econlet.2010.12.007

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14,* 399-425. https://doi.org/10.1007/s10683-011-9273-9

Ipeiriotis, P. (2010). *The demographics of Mechanical Turk* (NYU Working Paper No. CEDER-10-01). Retrieved from SSRN: https://ssrn.com/abstract=1585030

Johnson, D. R., & Borden, L. A. (2012). Participants at your fingertips: Using Amazon's Mechanical Turk to increase student–faculty collaborative research. *Teaching of Psychology, 39,* 245-251. https://doi.org/10.1177/0098628312456615

Kaplan, T., Saito, S., Hara, K., & Bigham, J. P. (2018, June). Striving to earn more: a survey of work strategies and tool use among crowd workers. In Sixth AAAI Conference on Human Computation and Crowdsourcing.

Keeter, S., Hatley, N., Kennedy, C., & Lau, A. (2017). What low response rates mean for telephone surveys. Pew Research Center. Available at: https://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/

Kittur, A., Chi, E. H., & Suh, B. (2008, April). Crowdsourcing user studies with Mechanical Turk. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 453-456).

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open*. https://doi.org/10.1177/2158244016636433

Litman, L., & Robinson, J. (2020). Conducting Online research on Amazon Mechanical Turk and Beyond. Sage Academic Publishing. Thousand Oaks: CA.

Litman, L., & Robinson, J. (2020). Conducting ethical online research: A data-driven approach. In L. Litman & J. Robinson (Eds.), *Conducting online research on Amazon Mechanical Turk and beyond* (234-263). Sage Academic Publishing. Thousand Oaks: CA.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49,* 433-442. https://doi.org/10.3758/s13428-016-0727-z

Litman, L., Robinson, J., Rosen, Z., Rosenzweig, C., Waxman, J., & Bates, L. M. (2020). The persistence of pay inequality: The gender wage gap in an anonymous online labor market. *PLOS ONE*, 15(2), e0229383. https://doi.org/10.1371/journal.pone.0229383

Litman, L., Robinson, J., & Rosenzweig, C. (2020). Sampling Mechanical Turk workers: Problems and solutions. In L. Litman & J. Robinson (Eds.), *Conducting online research on Amazon Mechanical Turk and beyond* (121-146). Sage Academic Publishing. Thousand Oaks: CA.

Marge, M., Banerjee, S., & Rudnicky, A. I. (2010, March). Using the Amazon Mechanical Turk for transcription of spoken language. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, (pp. 5270-5273). doi:10.1109/ICASSP.2010.5494979

Matsakis, L. (2018). *A window into how YouTube trains AI to moderate videos*. Wired. https://www.wired.com/story/youtube-mechanical-turk-content-moderation-ai/

Mehrotra, D. (2020). *Horror stories from inside Amazon's Mechanical Turk.* Gizmodo. https://gizmodo.com/horror-stories-from-inside-amazons-mechanical-turk-1840878041

Mortensen, K., & Hughes, T. L. (2018). Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. *Journal of General Internal Medicine, 33,* 533-538. https://doi.org/10.1007/s11606-017-4246-0

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science, 2,* 109-138. https://doi.org/10.1017/XPS.2015.19

Newman, A. (2019). *I found work on an Amazon website. I made 97 cents an hour.* The New York Times. https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411-419.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70,* 153-163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46,* 1023-1031. https://doi.org/10.3758/s13428-013-0434-y

Rao, S., & Michel, A. (2010). *ProPublica's Guide to Mechanical Turk*. ProPublica. https://www.propublica.org/article/propublicas-guide-to-mechanical-turk

Rashtchian, C., Young, P., Hodosh, M., & Hockenmaier, J. (2010, June). Collecting image annotations using Amazon's Mechanical Turk. In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (pp. 139-147). Association for Computational Linguistics.

Robinson, J., Litman, L., & Rosenzweig, C. (2020). Who are the Mechanical Turk workers? In L. Litman & J. Robinson (Eds.), *Conducting online research on Amazon Mechanical Turk and beyond* (121-146). Sage Academic Publishing. Thousand Oaks: CA.

Robinson, J., Rosenzweig, C., & Litman, L. (2020). The Mechanical Turk ecosystem. In L. Litman & J. Robinson (Eds.), *Conducting online research on Amazon Mechanical Turk and beyond* (121-146). Sage Academic Publishing. Thousand Oaks: CA.

Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. (2019). Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. PLoS ONE, 14(12): e0226394. https://doi.org/10.1371/journal.pone.0226394

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Mechanical Turk. In CHI'10 Extended Abstracts on Human Factors in Computing Systems (pp. 2863-2872). https://doi.org/10.1145/1753846.1753873

Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science, 2*, 107–114. https://doi.org/10.1177/2515245919838781

Schneider, D., & Harknett, K. (2019). *It's about time: How work schedule instability matters for workers, families, and racial inequality*. The Shift Project. https://shift.berkeley.edu/its-about-time-how-work-schedule-instability-matters-for-workers-families-and-racial-inequality/

Shank, D. B. (2016). Using crowdsourcing websites for sociological research: The case of Amazon Mechanical Turk. *The American Sociologist, 47*, 47-55. https://doi.org/10.1007/s12108-015-9266-9

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science, 1,* 213-220. https://doi.org/10.1177/2167702612469015

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515-530. https://doi.org/10.1037/0022-3514.51.3.515

Semuels, A. (2018). *The online hell of Amazon Mechanical Turk.* The Atlantic. https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods, 43,* 155-167. https://doi.org/10.3758/s13428-010-0039-7

Strickland, J. C., & Stoops, W. W. (2019). The use of crowdsourcing in addiction science research: Amazon Mechanical Turk. *Experimental and Clinical Psychopharmacology, 27,* 1-18. https://doi.org/10.1037/pha0000235.

Stritch, J. M., Pedersen, M. J., & Taggart, G. (2017). The opportunities and limitations of using Mechanical Turk (Mturk) in public administration and management scholarship. *International Public Management Journal, 20,* 489-511. https://doi.org/10.1080/10967494.2016.1276493

U.S. Census Bureau. (2018). *Current population survey, selected characteristics of households by total money income.* [Data set]. https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-01.html

Williamson, V. (2016). On the ethics of crowdsourced research. *PS: Political Science & Politics, 49(1),* 77-81. https://doi.org/10.1017/S104909651500116X

Woo, S. E., Keith, M., & Thornton, M. A. (2015). Amazon Mechanical Turk for industrial and organizational psychology: Advantages, challenges, and practical recommendations. *Industrial and Organizational Psychology, 8(2),* 171-179. DOI: https://doi.org/10.1017/iop.2015.21

Yeager, D. S., Krosnick, J. A., Visser, P. S., Holbrook, A. L., & Tahk, A. M. (2019). Moderation of classic social psychological effects by demographics in the U.S. adult population: New opportunities for theoretical advancement. *Journal of Personality and Social Psychology, 117*, e84-e99. http://dx.doi.org/10.1037/pspa0000171

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*, 493-504. https://doi.org/10.1037/pspa0000056

Zittrain, J. (2009). The Internet creates a new kind of sweatshop. *Newsweek.* https://www.newsweek.com/internet-creates-new-kind-sweatshop-75751

Competing Interests

The authors of this manuscript have the following potential competing interests: All authors are employed at Prime Research Solutions, the company that owns CloudResearch (formerly TurkPrime). CloudResearch's MTurk ToolKit was used to source Mechanical Turk participants, and the CloudResearch database was used to query some of the data.

Funding Statement

The funder, Prime Research Solutions, provided support in the form of salaries for authors [A.M, C.R., J.R., and L.L.], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the ‘author contributions’ section.

Author Contributions