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] · Shalom N. Jaffe[1,4] · Leib Litman[1,4]

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, subject to abuse, and earn inhumane wages. We investigated these issues with two representative probability surveys of the U.S. MTurk population ($N = 4094$). The surveys revealed: (1) the financial situation of people on MTurk mirrors the general population, (2) most participants do not find MTurk stressful or requesters abusive, and (3) MTurk offers flexibility and benefits that most people value above other options for work. People reported it is possible to earn more than $10 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 research.

Keywords Mechanical Turk · Ethics · Representative sample · Online research

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 20th century, many researchers relied heavily on laboratory studies conducted with undergraduates (see Sears, 1986) before shifting to online data collection in the 2010s (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 most common 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 et al., 2018; Robinson et al., 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 et al., 2011; Paolacci et al., 2010). Dozens of papers were published examining data quality (e.g., Litman et al., 2015), comparing data from MTurk to that from other sources (Chandler et al., 2019; Hauser & Schwarz, 2016; Peer et al., 2017), outlining ways to maintain data quality (Peer et al., 2014), and describing the demographics of people on the platform (Ipeiriotis, 2010; Robinson et al., 2020a). After gaining 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

✉ Leib Litman
leib.litman@touro.edu

1 CloudResearch, Queens, NY, USA

2 Department of Clinical Psychology, Columbia University, New York, NY, USA

3 Department of Computer Science, Lander College, Flushing, NY, USA

4 Department of Psychology, Lander College, Flushing, NY, USA

affect sample composition (Arechar et al., 2017; Casey et al., 2017; Fordsham et al., 2019), how repeatedly exposing participants to the same or similar materials might affect research findings (Chandler et al., 2015; Hauser et al., 2019), how to recruit naive participants (Meyers et al., 2020; Robinson et al., 2019), and how well findings on MTurk replicate in nationally representative samples (Coppock, 2019; Mullinix et al., 2015; Yeager et al., 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, 2020a).

However, MTurk has done more than just contribute to social science research; it has been used in various ways that have an important impact on society, including influencing industry, the media, and politics. Corporate researchers, for example, routinely use MTurk to develop technology that yields cleaner search results, moderated website content, or a better user experience among other applications (e.g., Fair Crowd Work, n.d.; Mechanical Turk, n.d.). Major news organizations like *ProPublica* (Rao & Michel, 2010) and *The New York Times* (see Newman, 2019) have used MTurk to crowdsource reader sentiment. And in politics, at least one major Presidential campaign used Mechanical Turk to improve the effectiveness of its messaging (Federal Election Commission, 2019) while other polling organizations appear to draw from MTurk to supplement their samples. Together, these varied uses of Mechanical Turk demonstrate the potential the platform has to serve as a resource for large swaths of U.S. society into the future.

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 (e.g., Damer, 2019; Newman, 2019). At the heart of this argument are concerns about exploitation. Specifically, some researchers worry that the people on MTurk are financially insecure and are relying on MTurk to make basic ends meet (e.g., Fort et al., 2011; Gray & Suri, 2019; Williamson, 2016). In addition, a series of popular press articles have raised the possibility of widespread abuse, characterizing MTurk 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 hour (Newman, 2019). These descriptions may have given academic researchers pause as few people want to advance their careers by exploiting vulnerable people.

Although much has been written and said about the ethics of MTurk, there is surprisingly little empirical evidence to support the claims described above. All the popular press pieces and most of the academic work to date 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 MTurk by conducting the first representative probability surveys of people on

the platform within the United States. 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 it is possible to 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 (the people who post tasks to the platform), 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 on other platforms.

Even though our research was motivated by general criticisms made of MTurk, the issues we examine are also relevant to the practical application of research ethics. Respect for people's ability to voluntarily participate in research, a commitment to minimizing risk or harm within studies, and the idea of beneficence are all bedrock principles of research ethics that stem from federal regulations like the Belmont Report and the Common Rule. Therefore, the issues we investigate about MTurk may inform the ethical considerations of researchers and institutional review boards when deciding whether to use Mechanical Turk for behavioral research.

Criticisms of Mechanical Turk: Past claims and evidence

Critics have raised three concerns about MTurk. First, some have suggested that people on MTurk are disproportionately poor or disabled and working on MTurk because they have trouble making ends meet elsewhere in the traditional labor market. The second claim is that working on MTurk involves abusive and stressful interactions with requesters. Finally, the third claim concerns pay: critics argue that people on MTurk are paid pennies. 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. We review the evidence for each of these claims below.

Claim #1: People on MTurk are financially vulnerable

Many researchers and journalists have wondered: who is willing to complete small tasks for small pay on MTurk? Popular press articles and some academic discourse have contributed to a dominant narrative that people on the platform are disproportionately poor or disabled. The source of this concern stretches back to some of the first academic surveys in which a sizable portion (20%) of people reported relying on MTurk as their primary income (Ipeiriotis, 2010; Ross et al., 2010; see Fort et al., 2011). The concern was later amplified by interviews conducted with

people on MTurk (Williamson, 2016) and 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, p. 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). 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 MTurk. For example, Williamson (2016) conducted interviews with 49 self-selected people on MTurk. While such interviews yield a rich glimpse into the lives of a few people, it is unclear how representative those people are of everyone on MTurk. Similarly, journalists who speak to a handful of people may be able to write a compelling article, but when roughly 100,000 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 canvassing of people on MTurk, conducted by organizations like the Pew Research Center (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 surveys aimed to provide a better answer to the question of whether people on MTurk are disproportionately poor or disabled. To do so, we asked a randomly selected sample of MTurk workers about their financial situation and whether they have a disability. We also 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 could. Not only does our sampling method yield representative results of the U.S. MTurk population, but, where possible, we compare our results to the broader U.S. population. 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.

Claim #2: People on MTurk are subjected to abuse

A second criticism of MTurk is that people are routinely subjected to stress, psychological harm, and abuse. 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 HIT (human intelligence task) has been sufficiently completed and when it has not. When a requester decides 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, making people sensitive to rejections.

While the structural factors on MTurk provide potential for abuse, what is not known is how prevalent such abuse is. Therefore, we asked people in our surveys whether they are satisfied or dissatisfied with the amount of stress that taking HITs on MTurk entails. We also asked people what percentage of HITs they are rejected from and what percentage of HITs contain disturbing material. In our second survey, we sought to replicate Horton's (2011) finding that most MTurk workers see requesters on MTurk as fairer and more honest than employers outside of MTurk. At the time, Horton et al. (2011) concluded that MTurk requesters are rated as fairer than employers outside of MTurk. Since no rigorous empirical evidence to the contrary has been produced since then, we expected similar results.

Claim #3: People on MTurk are paid pennies

The third criticism routinely made about MTurk concerns wages. Past reports about how much people earn have varied, but an often-cited academic study reports wages near $2.00 an hour (Hara et al., 2018). Meanwhile, studies by both the Pew Research Center and the International Labour Office have indicated that people made around $5.00 per hour in 2015 (Berg, 2015; Hitlin, 2016), while another, more recent analysis based on a larger dataset indicated that wages have risen to at least $5.70 an hour (Litman et al., 2020a). Among the most prolific users of MTurk wages may be above $20 per hour (TurkerView., 2019). Yet, in the popular press, earnings on MTurk 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, or in extreme cases, suggestions by reviewers or editors that studies with data from MTurk should not be accepted for publication.

We took two approaches to assessing wages. We asked people to report how much a person can earn per hour and we estimated wages from a large database of MTurk activity. We asked how much people *can* earn rather than asking people how much they *do* earn because direct self-reports have been criticized as unreliable (e.g., Hara et al., 2018). Furthermore, asking how much people can earn rests on the presumption that people are committed to maximizing

earnings in their time on MTurk, a presumption that past research has shown is not always the case for all workers (e.g., Chandler et al., 2014). Given that asking how much people can earn may bias estimates upward, our analysis of wages based on actual MTurk activity provided a way to assess the reliability of people's estimates. We hoped to see convergence across methods.

In addition to estimating wages, we assessed what reasons other than wages people might have for spending time on MTurk. Past claims in the popular press have stated that people turn to MTurk because they are unable to find better options for work (e.g., Newman, 2019). Hence, we asked people whether they would take another job over MTurk if such an opportunity became available. We also asked people how much another job would need to pay for them to give up MTurk and take the other job. Finally, we asked people to list and rank several reasons why they spend time on MTurk. 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.

Overview

This research 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 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 contemptible. 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 Presidential primary contender Pete Buttigieg spent $20,000 conducting polling on MTurk 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 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," (Grim & Lacey, 2020). Although the sentiment about MTurk being an exploitative platform is often repeated, few studies have examined whether these claims are backed by solid empirical evidence; hence, the reason for our studies.

In Study 1, we asked people on MTurk in the U.S. questions that speak to each of the criticisms commonly made of MTurk. We created several of our own questions and adapted

others from Gallup. Asking people questions that Gallup has used allowed us to compare people on MTurk to the general U.S. population. After analyzing the data from Study 1, we realized the survey did not include enough questions to characterize what people's time on MTurk is like. Therefore, we created our second survey to assess people's feelings toward requesters, how often HITs are rejected or contain disturbing content, and how positive or negative people's experiences tend to be. Together, our surveys provide a variety of data points that inform the conversation about the ethics of using MTurk for behavioral research.

Disclosures

Our research was approved by IntegReview, an independent review board that evaluates research involving human subjects. We preregistered our materials and general predictions for Study 1 (see here: https://osf.io/cwde4) but not Study 2. Because our surveys examined multiple point estimates, the best predictions we could make were about the general pattern of results. We report all measures and all data exclusions and have made all materials, data, and analysis scripts available at: https://osf.io/8nhyz/. Finally, we note that in addition to holding various academic appointments, all authors are affiliated with CloudResearch—a private company that facilitates online research projects including those on Mechanical Turk. As academic researchers, each member of the authorship team has followed the conversation about MTurk and ethics with interest. As employees of CloudResearch, we have had access to data that caused us to question whether previous characterizations of MTurk were accurate (e.g., data on wages and rejections). The studies we report were an attempt to gather data with sound methods that may resolve some of these discrepancies we saw over time while also informing the conversation about the ethics of using MTurk for behavioral research.

Method

We conducted two probability surveys of the U.S. MTurk population using stratified random sampling. 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 over a period of two months. Then, we stratified the list based on participant experience and randomly selected participants from within each stratum.

Sampling procedure

To sample workers with different levels of experience, we queried the CloudResearch database for all unique workers who completed at least one HIT in October and November 2019; this query yielded 43,274 people. Next, we randomly sampled people within four different levels of participant experience: people with less than 100 HITs completed all time (35% of the MTurk population); people with between 100 HITs and 1000 HITs completed (37.4% of the population); people with between 1000 HITs and 5000 HITs completed (16.4% of the population); and people with more than 5000 HITs completed (11.3% of the population). Except for people with less than 100 HITs completed, these groups and their share of the MTurk population were taken from Robinson et al. (2019).

Stratifying the sample by prior experience on MTurk and sampling each group in proportion to its share of the MTurk population was critical to gathering a representative sample. Because we knew, before conducting the study, that more active and more experienced people would respond to the survey at a higher rate than people who are less active (a consequence of the superworker bias, see Litman, Robinson, & Rosenzweig, 2020b), stratifying the sample mitigated bias created by differences in worker activity. An important limitation of nearly all previous research with MTurk is that the way MTurk operates—leaving studies open to workers on a first-come, first-served basis—leads to a form of self-selection bias that favors the most active and experienced workers. Specifically, as shown by Robinson et al. (2019), when samples on MTurk are not stratified by experience, over 70% of active workers get completely left out of studies. In this study, we aimed to obtain point estimates that capture the opinions of the entire U.S. MTurk pool, as well as to present data separately for people of different experience levels. In all four groups, we recruited participants in proportion to their percentage of the U.S. MTurk population based on our target sample size of 2000. This means, for example, that we aimed to recruit 700 people with less than 100 prior HITs completed ($35\% \times 2000 = 700$).

We determined people's level of experience by looking at the CloudResearch database 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 5000 completed HITs. After generating the sampling frame, we randomly selected people and invited them to the study with e-mail invitations sent through MTurk's application programming interface (API). 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 wave of data collection, we invited people to the study when it was launched and then sent reminder e-mails on each of the following 2 days. After 3 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 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 5000 HITs completed. For comparison, the response rate in surveys that use random digit phone dialing—often considered the gold standard for representative surveys—is around 6% (Kennedy & Hartig, 2019; Marken, 2018). Tables S1 and S2 in the supplemental materials show the basic demographic information of (a) people who participated in our survey, (b) people who were e-mailed but did not respond, and (c) the sampling frame.

People who responded to our invitation 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 min on average to complete for an effective pay rate of ~$9.00 per hour.

Analytic approach

To estimate 95% confidence intervals for the full U.S. MTurk population, we weighed each stratum to match its proportion of the MTurk population. Then, we used the SPSS Complex Sample function to generate point estimates and 95% confidence intervals (CIs) for our samples. The 95% CIs were computed taking stratification and weighing into account. We present 95% CIs in all tables and figures. Finally, in an exploratory analysis suggested during the review process, we weighed the data not to each stratum's share of the MTurk population but to the percentage of HITs completed by workers in each stratum. Theoretically, this weighing better represents the population of workers who spend the most time completing HITs. We report these analyses in the supplementary materials and return to the issue of how best to represent the MTurk population in the discussion.

Evaluating the sampling frame

CloudResearch is an independent company that enables researchers to design and manage online studies, including studies on MTurk. Each time a researcher uses
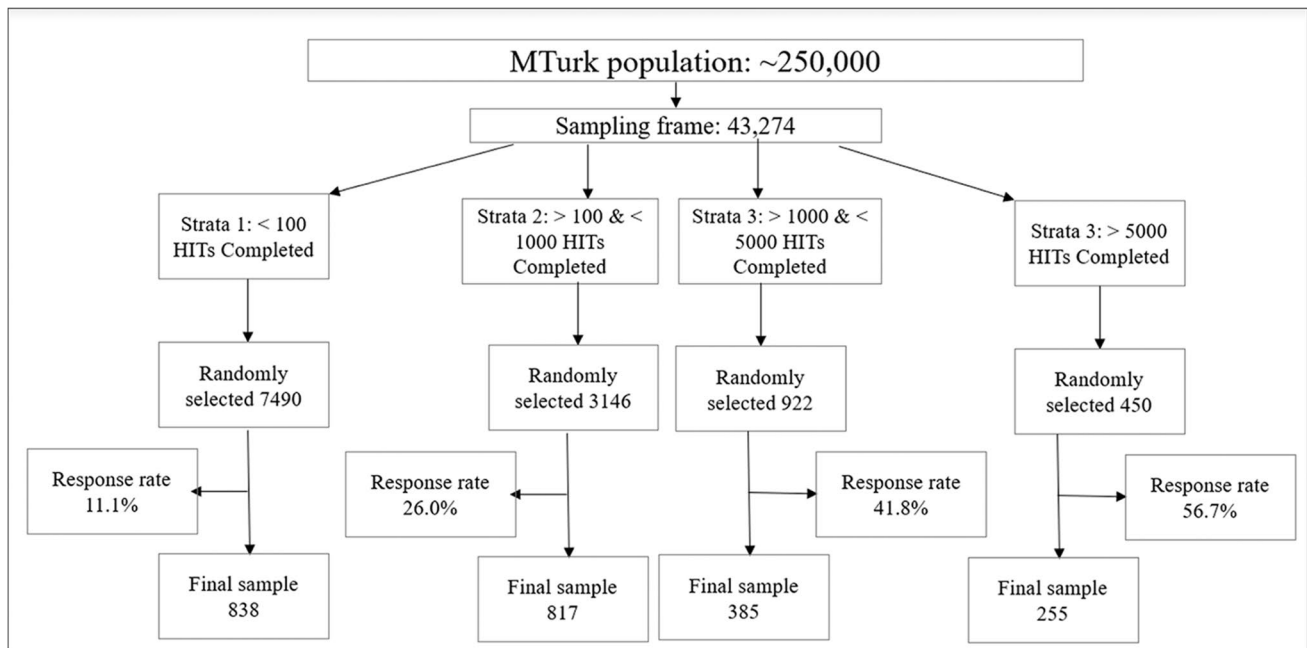
Fig. 1 Response rates in Study 1

CloudResearch to conduct an MTurk study, metadata from the study is captured and stored in the CloudResearch database. This metadata includes things like the study title, compensation, the Worker IDs of people who complete the study, and the requester's name.

In this study, we used the CloudResearch database as a sampling frame and to estimate participant wages. Assessing the suitability of this database as a sampling frame for MTurk requires knowing how many workers are on MTurk but not in the database. We had multiple reasons to believe that virtually all active MTurk workers are in the CloudResearch database. First, given the amount of activity on CloudResearch—more than 50,000 studies per year launched by over 5000 researchers and taken by ~100,000 unique participants in the U.S.—it is likely that anyone who completes tasks on MTurk for more than a few days will encounter at least one CloudResearch study and thus be in the database. Second, a review of the published literature reveals that the CloudResearch database contains more workers than the largest independent estimate for the size of the MTurk pool. Difallah et al. (2018) estimated MTurk to have around 100,000 active people. The CloudResearch database contained more than 250,000 workers all-time in 2019 (Robinson et al., 2019) and in August 2020 this number had risen to more than 320,000, suggesting substantial overlap between CloudResearch and MTurk. Finally, we conducted a direct examination of how many workers are on MTurk but not within the CloudResearch database using data from Arechar and Rand (2021). In pooled data from 16 studies with more than 7500 workers sampled directly from MTurk (without

using CloudResearch), we found just 17 people whose Worker IDs were not within the CloudResearch database. This means the CloudResearch database contained 99.8% of workers collected independently on MTurk. Altogether, this evidence indicates that the CloudResearch database provides a valid sampling frame for Studies 1 and 2 and a rich dataset from which to estimate hourly wages.

Study 1: A Representative survey of MTurk participants

Participants

We collected data from a total of 2026 U.S. adults in mid-December 2019[1] using the CloudResearch Toolkit (Litman et al., 2017). To prevent non-U.S. workers and fraudulent respondents from taking the survey, we employed tools to block suspicious geolocations and duplicate IP addresses. Tables 1 and 2 display complete demographic information. In general, the demographic profile of our sample converges with past reports about MTurk demographics based on population-level observations (Difallah et al., 2018; Litman, Robinson, & Rosenzweig, 2020b).

---

[1] Our data file contains responses from a total of 2277 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	Percent
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)
< $10,000	6.31	5.93
$10,000–$19,999	6.60	8.77
$20,000–$29,999	11.67	8.66
$30,000–$39,999	10.82	8.73
$40,000–$49,999	11.02	7.80
$50,000–$59,999	11.22	7.56
$60,000–$69,999	7.94	6.69
$70,000–$79,999	7.30	6.02
$80,000–$89,999	5.06	4.98
$90,000–$99,999	5.16	4.42
$100,000–$149,000	11.97	14.95
>$150,000	4.92	15.47

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

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, although people reported a range of household incomes (see Table 2), the overall distribution was within a few percentage points of the U.S. population for each income bracket except for households making more than $150,000 per year (U.S. Census Bureau, 2018).

Materials

The survey contained several questions that asked 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
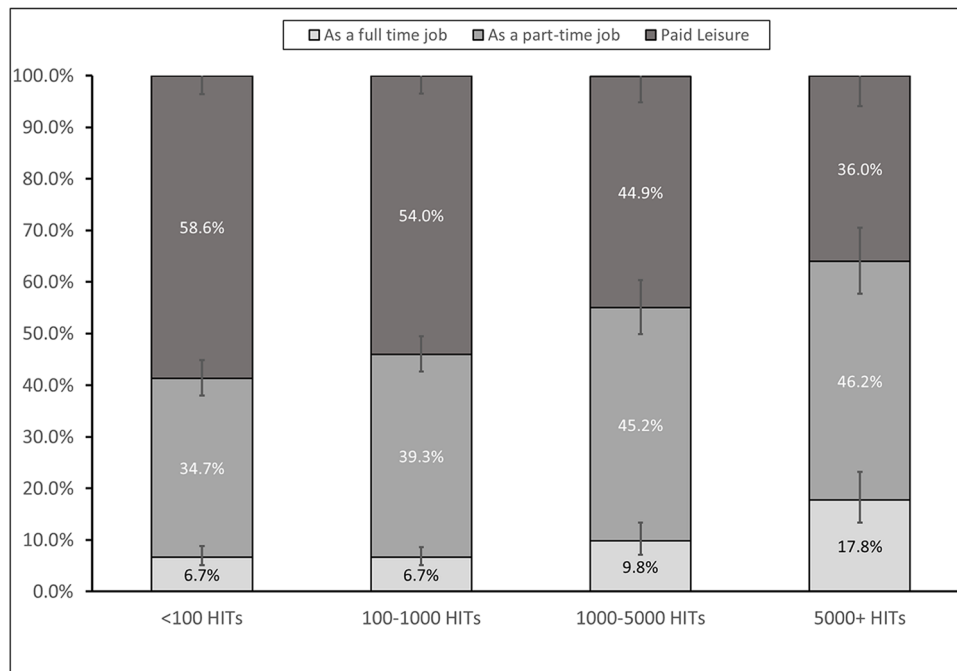
Fig. 2 Characterizations of MTurk based on prior HITs completed. *Note:* People with less than 100 HITs completed constitute 35% of the MTurk population; people with 100 to 1000 HITs are 37.5%; people with 1000–5000 HITs are 16.4%; and people with 5000 or more HITs are 11.3% of the population. These data are from Robinson et al. (2019) and are based on U.S. MTurk workers. *Error bars* represent 95% confidence intervals

adapted from nationally representative surveys, and the results are presented below with these comparisons when available.

Results

Because our goal was to answer questions about Mechanical Turk at the platform level, we present all data for the full sample and selectively report analyses based on worker experience level. Data disaggregated by level of worker experience are available in our supplemental materials.

Characterization of Mechanical Turk

We asked people if they characterize their time on MTurk as full-time work, part-time work, or a form of paid leisure. Most people (55.8%) characterized MTurk as a form of paid leisure. The next largest group (36.7%) characterized MTurk as a form of part-time work and 7.5% 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 on MTurk increased, people were less likely to describe MTurk as paid leisure and more likely to describe it as some form of work (see Fig. 2). At no level of experience, however, did more than half of people characterize MTurk as part-time work and the percentage of people characterizing MTurk as full-time work never surpassed 20%.

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 seven out of ten people (68.6%) said they work on MTurk to earn money for non-essential expenses. Another one-fifth of participants (19.9%) said they work to cover essential expenses, and the remaining 11.4% 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 eight provided answers that could be construed as covering essential things. The most common alternative responses were that MTurk provides extra money (41.4%), an amusing or pleasant distraction from other things (23.7%), money for savings (9.4%), money to supplement other income (6.3%), or money to cover purchases on Amazon (7.6%).

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 commitment: full-time employment, part-time employment, and as something to do in their leisure time.

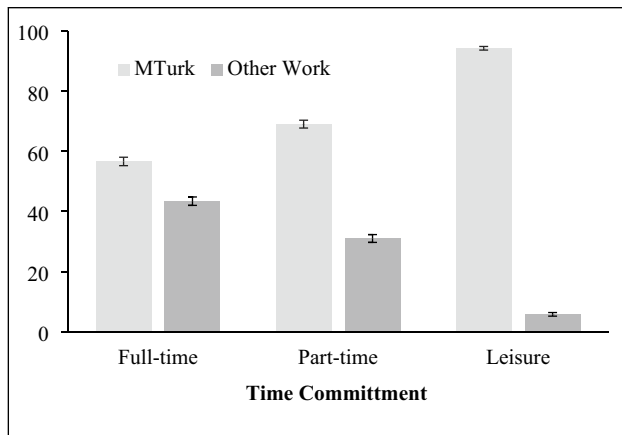The results from these questions are presented in Fig. 3. As shown, people preferred MTurk over other

Fig. 3 Percent of people choosing MTurk over other work. *Note: Error bars* are 95% confidence intervals

work regardless of the time committed. Furthermore, as the time commitment decreased people's preference for MTurk increased.

Reasons for working on MTurk

We asked why workers 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 3. As shown, the top reasons people chose involved flexibility. Specifically, people liked 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 gauge for what percentage of people might be on MTurk because of a disability. As Table 3 shows, disability was the least commonly selected reason for working on MTurk, with 15.4% of people in our sample selecting this reason. In the U.S. population 26% of people report living with a disability (Centers for Disease Control and Prevention, n.d.). Together, these data points suggest that people on MTurk are not more likely than people in the general population to have a disability that prevents them from working.

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 said they would require was $26.01 [24.89, 27.13] (*SD* = 18.90; median = 20.00). Because outliers may

Table 3 Reasons for spending time on MTurk

Reason	% Choosing	Average ranking
Work from home	94.6 [93.2, 95.7]	2.12
Flexible hours	91.4 [89.7, 92.8]	2.70
No commute	72.1 [69.5, 74.4]	5.05
Breaks between work	67.3 [64.6, 69.8]	5.30
No boss, supervisor, etc.	66.2 [63.6, 68.8]	4.91
No dress code	63.8 [61.1, 66.4]	6.30
I can work from wherever	61.1 [58.4, 63.7]	5.67
Less stressful	57.7 [55.0, 60.4]	5.80
I learn interesting things	45.9 [43.2, 48.6]	6.45
Flexibility for family	43.4 [40.7, 46.1]	4.32
Immediate pay	35.6 [33.0, 38.2]	5.79
I can make more money	25.4 [23.0, 27.9]	5.36
Physical or mental health constraints make it hard to find work elsewhere	14.1 [12.3, 16.1]	4.25

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

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 $26.18 [25.19, 27.17] per hour to take an alternative job over MTurk (*SD* = 17.52; median = $20.00).

As shown in Fig. 4, 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 to take another job over MTurk. People who characterized MTurk as full-time work reported that they would need to make $40.02 per hour [34.50, 45.54] (median = $24.12, winsorized mean = $38.96). People who characterized MTurk as part-time work or as paid leisure both said they would need to make more than $24 per hour (part-time M = $24.79 [23.08, 26.50], winsorized mean = $24.36, median = $18.77; leisure M = 24.93 [23.52, 26.34], winsorized mean = $24.59, median = $20.02). Altogether, these figures show that people assign a high value to their MTurk work. Even though people do not actually make as much as they said they would require to give up MTurk, nonmonetary factors such as convenience, lack of travel, flexibility, and low stress augment people's actual earnings.

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 $10.41 [9.57, 11.25] (*SD* = $13.35; median = $6.48). 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 $9.77 per hour [9.12, 10.42] (*SD* = $10.54; median = $6.48).
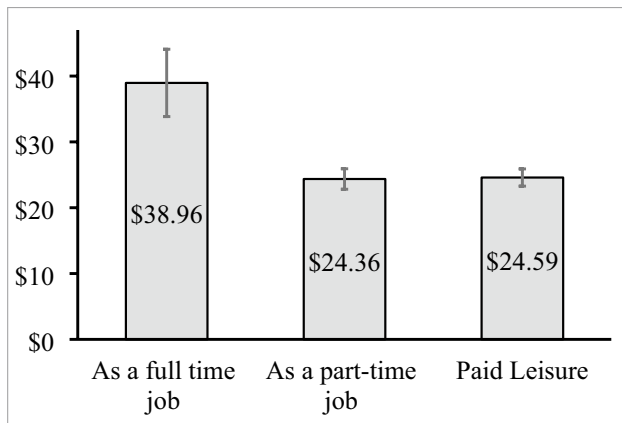
Fig. 4 Hourly wage people would require to trade MTurk for another job. *Note:* Values are based on winsorized distributions. *Error bars* represent 95% confidence intervals
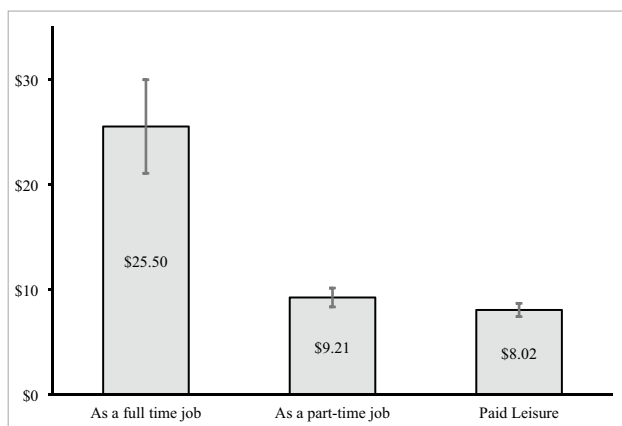

Fig. 5 Hourly wage people said someone can earn on MTurk. *Note:* Values are based on winsorized distributions. *Error bars* represent 95% confidence intervals

As shown in Fig. 5, 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 $30.05 per hour [23.93, 36.18] (winsorized mean = $25.50, median = $10.26). Meanwhile, people who characterized MTurk as part-time work or as paid leisure said it was possible to earn around $8 or $9 per hour (part-time M = $9.73 [8.54, 10.93], winsorized mean = $9.21, median = $6.94; leisure M = $8.21 [7.45, 8.97], winsorized mean = $8.02, median = $6.03).

Financial situation and satisfaction with MTurk

To assess people's current financial situation, we used polling questions from Gallup (see Brenan, 2019). The

results revealed that the financial situation of people on MTurk looks like the U.S. population (see Fig. 6). For example, while 15% of the general population describes their financial situation as "poor" the number on MTurk is 16.4% (see Fig. 5). While 36% of the general population describes their financial situation as "only fair" the number on MTurk is 36.9%. Most importantly, 47% of people on MTurk described their financial situation as being "Good" or "Excellent" compared to 49% in the U.S. population.

In another question borrowed from Gallup, we asked people to describe their household finances (Brenan, 2019; see Fig. 7). People on MTurk were less likely than those in the U.S. population to report they are "saving a lot" and more likely to say they are "running into debt," but across all other answer choices the distribution was fairly similar. For example, nearly 30% of people on MTurk said they are "saving a little" compared to 37% in the U.S. population, and less than 10% of people in both the U.S. population and on MTurk said they are "having to draw on savings." Thus, while answers to this question suggest some financial hardship among people on MTurk, people on MTurk within the U.S. look a lot like the general population.

Next, we asked people how they feel about pay on MTurk. According to Gallup, between 40 and 50% of Americans routinely report they are underpaid (Norman, 2018). On MTurk, this number was 65% [62.3, 67.6]. In addition, whereas 50% of Americans say they are paid about right, the number on MTurk was 35% [32.4, 37.6] (see Fig. 8).

We also 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 (Gallup, n.d.). As shown in Fig. 9, about 60% of people on MTurk reported being satisfied with the amount of money they earn. In the general population, this number is 78%.

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 (see Fig. 10). In the general population, 73% of people say they are satisfied with their job-related stress.

Finally, we asked people a direct question: is MTurk part of the problem or part of the solution to your financial needs? Approximately 87% of people said MTurk was part of the solution whereas just 3.2% said it was part of the problem. Ten percent said, "It's more complicated than that/A little bit of both."

Additional wage estimate

The second approach we used to estimate wages was to examine data from more than 26 million completed assignments in the CloudResearch database. These data come from studies stretching a span of over 4 years. We computed average hourly
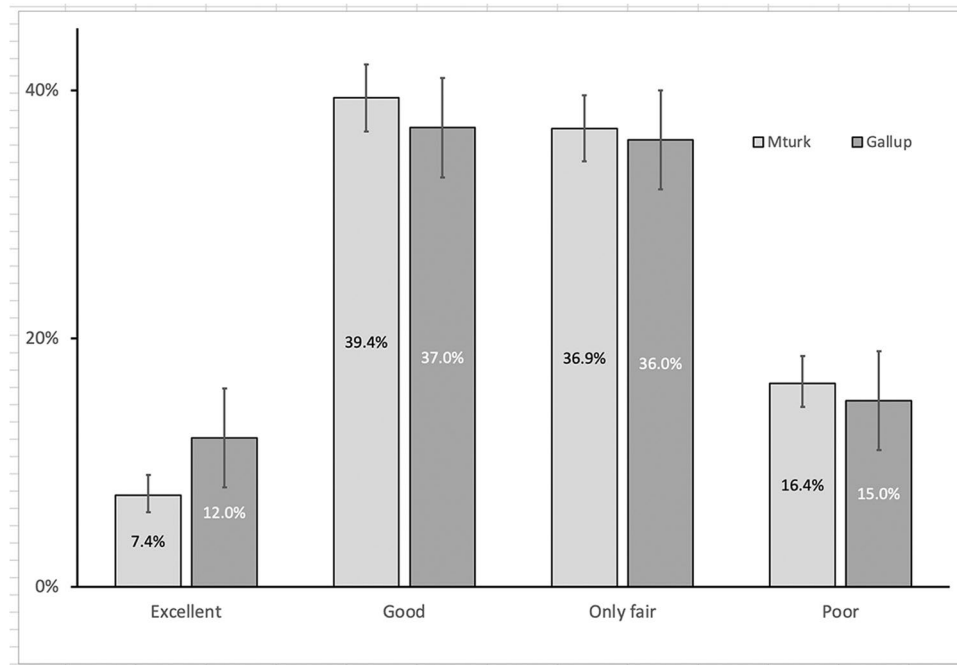
Fig. 6 The financial situation of people on MTurk compared to the U.S. population. *Note: Error bars* represent 95% confidence intervals. Error bars for the Gallup data are based on a maximum ±4% margin of error
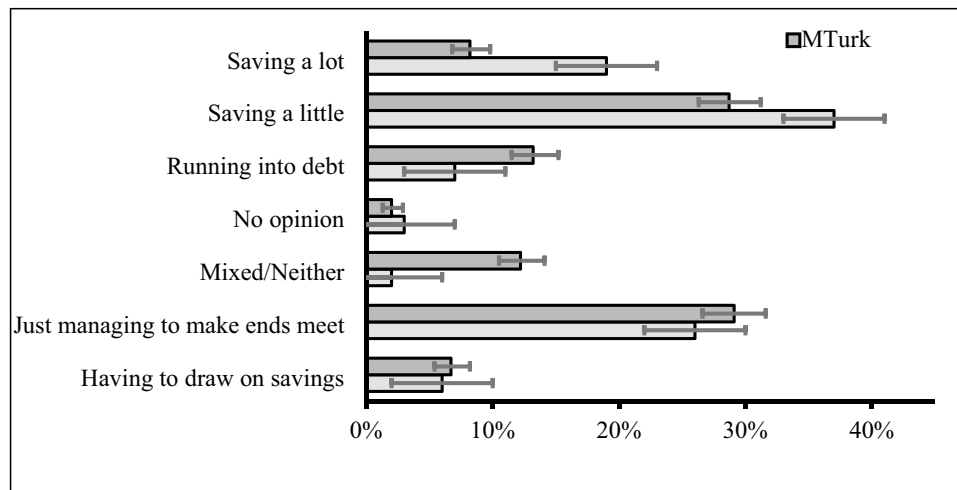

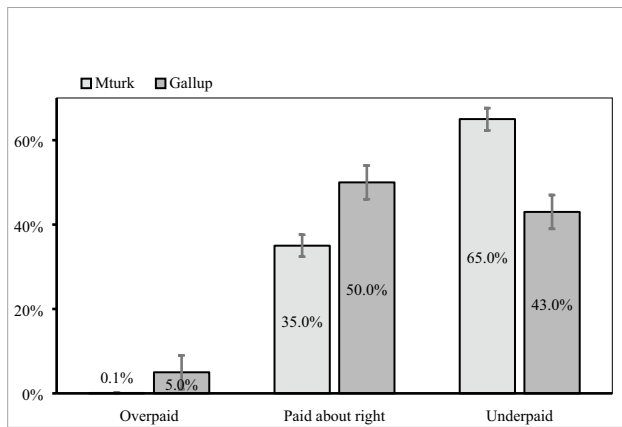**Fig. 7** Household finances for people on MTurk and in the U.S. population. *Note: Error bars* represent 95% confidence intervals. Error bars for the Gallup data are based on a maximum ±4% margin of error
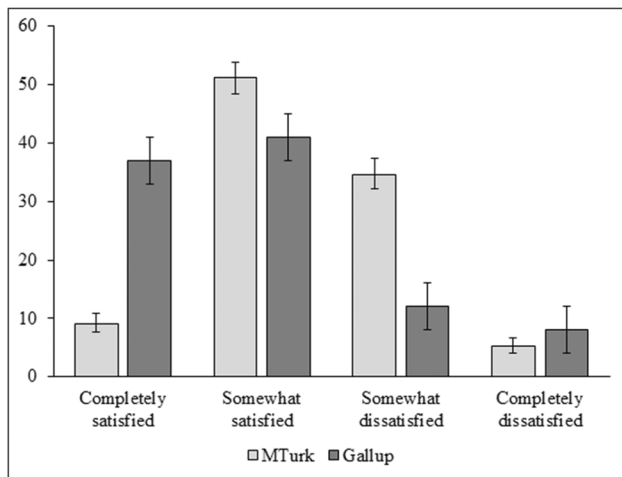
wages by subtracting the start time of each HIT from the end time and then dividing the total time required to complete each HIT by the payment offered. Complete details for our wage analysis are included in Supplementary Materials.

The mean hourly wage for each half-year interval from mid-2015 to 2019 is plotted in Fig. 11. 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 mean hourly wage had risen by 25% to $6.85 per hour.

Discussion

Overall, the results of our first survey do not support the common criticisms made of MTurk. More than half of people surveyed said they view MTurk as a form of paid leisure and

Fig. 8 Pay sentiment among people on MTurk and in the U.S. population. *Note: Error bars* represent 95% confidence intervals. Error bars for the Gallup data are based on a maximum ±4% margin of error

Fig. 10 Satisfaction with stress among people on MTurk and in the U.S. Population. *Note: Error bars* for MTurk represent 95% confidence intervals. Error bars for the Gallup data are based on a maximum ±4% margin of error

Fig. 9 Pay satisfaction among people on MTurk and in the U.S. population. *Note: Error bars* for MTurk represent 95% confidence intervals. Error bars for the Gallup data are based on a maximum ±4% margin of error

about 70% said they use the money they earn for nonessential spending. People in our survey also reported a financial situation that was similar to or only slightly worse than the general population. Also, counter to some popular press pieces, people reported that they are satisfied with the amount of stress that MTurk entails. Finally, people said it is possible to earn about $10 per hour and our wage analysis suggested that average wages are just under $7 per hour. The increase in wages over time may be the result of shifting norms among academic researchers (thanks to conversations about what is appropriate compensation for tasks), competition among requests using the site, or both. Either way, the picture painted by our data calls into question some of the ways MTurk is often portrayed.
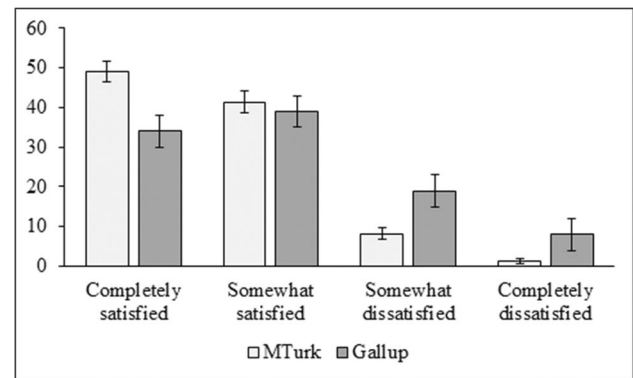
In our second survey, we sought to better assess how people feel about requesters on MTurk and whether people's time on MTurk is characterized by more positive or negative experiences. To these ends, we asked participants how fair and honest requesters are, how often HITs are rejected, and how often HITs contain disturbing content. We expected the data to show that unpleasant requesters and experiences were uncommon.

Study 2: Are requesters on MTurk fair?

Participants and procedure

The sampling process for Study 2 was like Study 1. We collected data from a total of 2068 people. Participants were recruited based on response rates across each level of experience observed in Study 1 (see Fig. 1). All data were collected over a 2-week period at the end of March 2020. During this 2-week period, we sent e-mail invitations until we filled our quotas, which as in Study 1, were set to reach a representative group of U.S. MTurkers stratified by experience level (700 people with <100 HITs, 748 with 100–1000 HITs, 328 with 1000–5000 HITs, and 226 people with 5000+ HITs). Response rates were similar to Study 1, allowing us to meet our quotas in three of the four groups. In the <100 HITs group, response rates in Study 2 were slightly lower (9.7% response rate), which led us to invite an additional 910 people to fill the remaining 82 slots. We omitted demographic measures to keep the survey short.

Materials

The survey contained seven questions asking people about the fairness of requesters on MTurk and how commonly they
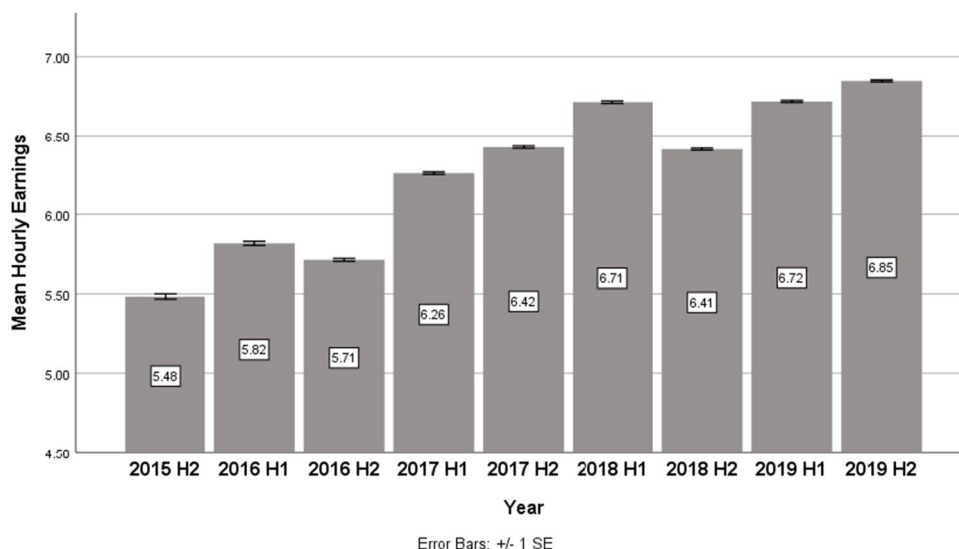
Fig. 11 Mean hourly wages over time. *Note:* H1 = the first half of the year, January through June. H2 = the second half of the year, July through December

experience rejections and disturbing content. The first two questions asked people to evaluate the fairness of requesters on MTurk and the fairness of employers outside of MTurk. While we hoped to replicate past work by randomly assigning the order people answered these two questions in (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 84.51% (*SD* = 18.93) of requesters on MTurk treat them fairly (median = 91.00%) while only 74.19% (*SD* = 20.85) of employers outside of MTurk do the same (median = 79.20%). These ratings favoring requesters on MTurk were consistent across levels of worker experience (see Fig. 12).

Next, we asked people if they have ever worked on a microtask platform other than MTurk; almost a quarter of people reported they had (23.9%). To this group, we asked where people are generally treated the most fairly. Half of people in this group said they are treated the most fairly on MTurk, while just under one-fifth said they are treated more fairly on other

microtask platforms and about one-third said their treatment is about the same regardless of platform (see Fig. 13).

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 reported average was 8.23 hours [7.81, 8.65] (SD = 10.04; median = 5.00). Hours per week varied by experience. People with less than 100 HITs completed reported spending an average of 6.91 hours per week while people with more than 5,000 HITs reported spending an average of 13.37 hours per week. In all strata,
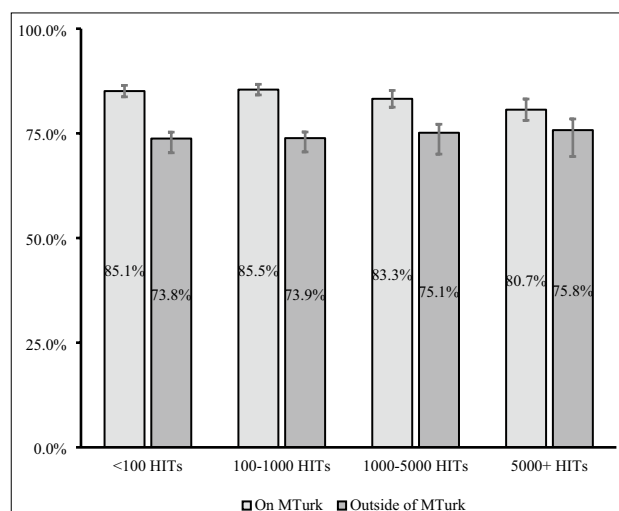

Fig. 12 Fairness of MTurk requesters compared to outside employers. *Note: Error bars* represent 95% confidence intervals
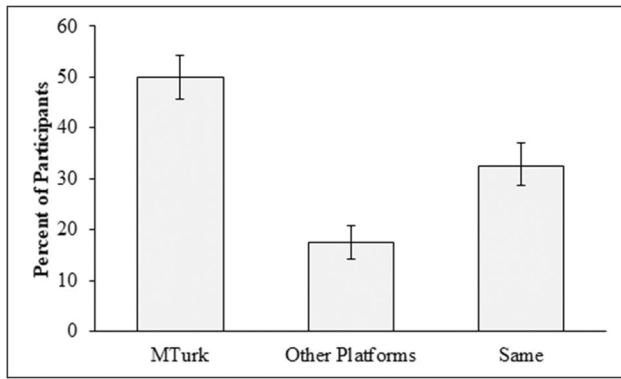
Fig. 13 Participants' reports of which online platforms are the fairest. *Note: Error bars* are 95% confidence intervals
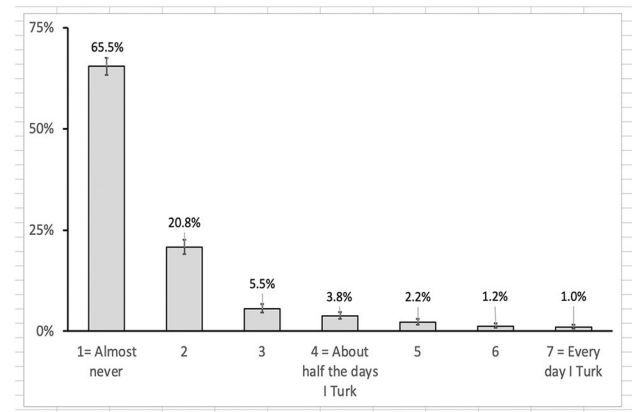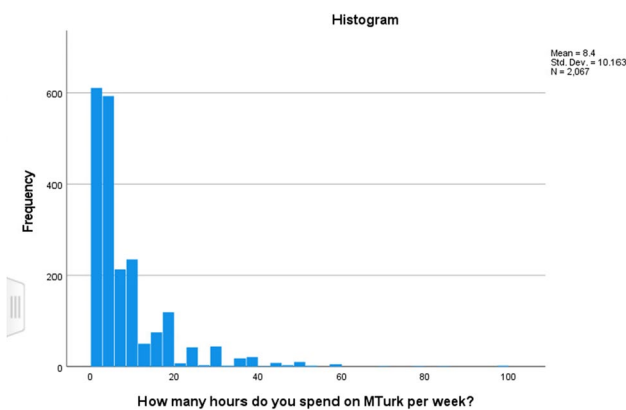

Fig. 14 Reported hours spent on MTurk per week

medians were lower, never reaching above 10 hours. A distribution of hours worked per week is displayed in Fig. 14.

In response to a question asking what percentage of HITs they have been rejected from, people reported a mean of 5.69% [5.07, 6.32] (SD = 14.29)[2] and a median of 1.00%. Interestingly, when we asked people what percentage of their rejections were unfair, the average was 36.37% [34.67, 38.07] (*SD* = 39.56; median = 13.80%), suggesting that people believe most of the rejections they receive are justified.

In response to a question asking what percentage of HITs contain disturbing content, people reported an average of 6.84% [6.16, 7.51] (*SD* = 15.54) and median of 1.00%.

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

---

[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%.

Fig. 15 Frequency of upsetting events

4 (*About half the days I Turk*). As shown in Fig. 15, people reported that upsetting experiences occur infrequently (*M* = 1.64 [1.59, 1.69], *SD* = 1.18).

General discussion

Mechanical Turk has often been described as an exploitative platform that is an outlier in the gig economy for its mistreatment of workers. For example, when criticizing a Presidential primary candidate who used MTurk for research, two journalists wrote, "While exploitation is rampant in the gig economy, MTurk has been identified as one of the worst offenders by journalists and researchers" (Grim & Lacey, 2020). Yet, the results of the present study stand in stark contrast to this characterization. Our representative surveys of the U.S. MTurk population revealed that people on MTurk: (1) are about as well-off financially as the general population, (2) complete tasks as a form of paid leisure and to supplement their primary incomes, (3) find MTurk to be significantly less stressful than other jobs, (4) find great value in the flexibility and benefits MTurk offers over more traditional work, (5) earn more than previous reports have indicated, (6) are substantially more likely to say MTurk is part of the solution to their current financial situation rather than part of the problem, (7) report that upsetting experiences on MTurk are rare, and (8) are inclined to say that requesters on MTurk are fairer and more honest than employers outside of MTurk. In fact, most people with experience on multiple microtask platforms reported that MTurk is where they are treated the most ethically. In addition, we reported wage data that suggests people currently earn about $6.50 per hour and would require more than $25 per hour to give up the flexibility MTurk allows and earn money in another way. In our view, these data provide a strong case that the criticisms of MTurk have been disproportionate to the reality.

At the same time, our data revealed that people on MTurk are more likely to say they are underpaid and unsatisfied

with their earnings than people in the general U.S. population. These data show that MTurk is not a perfect platform and that aspects of MTurk can and should be improved. Importantly, however, the percentage of people reporting dissatisfaction with wages is, arguably, lower than what might be expected based on common criticisms of MTurk. If U.S. MTurk workers really were earning "pennies per hour" (Newman, 2019) it is unlikely that 60% of them would report being either "somewhat" or "completely" satisfied with the wages. Indeed, more people in the U.S. population report being "completely dissatisfied" with their earnings (8%) than do people on MTurk (5%), and MTurk is not advertised or intended to serve as a primary source of income. Thus, our results show that while efforts should be made to improve the earnings of MTurk workers, describing MTurk as an "online hell" (Semuels, 2018) or a "digital sweatshop" (Graham, 2010) where people work on tasks for hours on end and earn just pennies per job is inaccurate.

Is Mechanical Turk exploitative?

Contrary to past characterizations of MTurk, people in our surveys described their financial picture in terms that generally matched the U.S. population. For example, only 16% of people on MTurk described their financial situation as poor compared to 15% in the general population. In another question, people on MTurk provided answers that demonstrated slightly less financial security than the general U.S. population (e.g., fewer people saying they were saving a lot), but still a distribution that mirrored the general population across answer choices (i.e., the number of people saying they were saving a little or running into debt). Importantly, how people described their finances remained relatively constant across all levels of worker experience (see Supplemental Materials). Also, people's description of their finances is buttressed by our demographic question asking about household income. The distribution of income on MTurk mirrored that of the U.S. population until household income exceeded $150,000 per year. Therefore, our data do not support the idea that people are on MTurk because they are disproportionately poor.

The second piece of data from our surveys concerns whether people say they spend time on MTurk because a disability precludes them from finding other work. We asked people in our first survey to rank all the reasons they work on MTurk. Out of the 13 available choices, "Physical or mental health constraints make it hard to find work elsewhere" was the least commonly selected reason with 15% of people citing this as a reason for spending time on MTurk. The percentage of people with a disability in the U.S. population, meanwhile, is 26% (Centers for Disease Control and Prevention, n.d.), which suggests MTurk does not attract a disproportionate number of disabled people. One limitation of this finding is that people on MTurk are younger than the U.S. population as a whole and people tend to incur more disabilities as they age. It is also possible that the wording of our question or people's reluctance to admit that a disability is the reason they spend time on MTurk may affect this estimate. Nevertheless, our data shows that the vast majority of people on MTurk do not report working on MTurk due to a disability, which runs contrary to what has previously been suggested (e.g., Newman, 2019).

Finally, we asked people several questions that assessed how they view MTurk and how they use the money they earn. Our data revealed that 56% of people do not characterize MTurk as a job but as a form of paid leisure. Another approximately 37% characterized MTurk as a part-time job, and less than 8% of people viewed MTurk as a full-time job. These numbers are important because much of the debate over MTurk seems to hinge on whether people perceive MTurk as a job. Even though MTurk bills itself as a place where people can 'earn money in their spare time,' many criticisms of MTurk, and the framework of proposed solutions (e.g., FairWork, 2021), rest on the assumption that people use MTurk for the bulk of their income or because they cannot find work elsewhere (e.g., Newman, 2019; Semuels, 2018). The strength and insistence of this narrative is inconsistent with our data showing that how people characterized their time on MTurk remained relatively consistent across all levels of experience (see Supplemental Materials). Although the percentage of people saying they use MTurk as a form of part time work increased from 35% to 46% as worker experience increased, the percentage of people saying they see MTurk as full-time work never exceeded 20% and the percentage classifying MTurk as a form of paid leisure never fell below 36%. Adding to this data, 78% of people said they use MTurk earnings for non-essential expenses. Thus, our data are largely consistent with the idea that most people on MTurk have primary jobs away from MTurk and that their financial situation is like people in the general population.

What is it like to spend time on MTurk?

We evaluated what it is like to spend time on MTurk by asking people how stressful they find their time on the platform and how often they have unsettling encounters with requesters such as being rejected for no reason or being exposed to offensive content without warning. Overall, people on MTurk reported less stress than those in the U.S. workforce. Specifically, 90% of people on MTurk reported either being "completely satisfied" or "somewhat satisfied" with the amount of stress they experience while just 69% in the U.S.

population reported the same thing. About 10% of people reported being either somewhat dissatisfied or completely dissatisfied with the amount of stress on MTurk, while 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 indicated that they perceive requesters on MTurk as 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 being that people rated requesters on MTurk even more positively in the current study. This may be 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.

Another question that queried people about their experiences on MTurk was on which microtask platform they experience the most ethical treatment. Of the people who had experience on multiple microtask platforms, most said MTurk is where they receive the most ethical treatment. Finally, we asked people how often they encounter rejections and disturbing content. Based on mean numbers, which may be inflated by outliers, people reported that less than 7% of HITs they complete are rejected or contain disturbing content; median numbers place these estimates below 1%. Thus, these data are inconsistent with the claim that there is widespread abuse on MTurk. 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 research tasks (Litman & Robinson, 2020b) and the time that requesters advertise for surveys typically ends up overestimating how long it takes people to complete the task, not underestimating it (Litman et al., 2020a). 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 4% said they find the content of tasks more distressing than activities in daily life (Litman & Robinson, 2020b) and most of these people (75%) said the benefits of MTurk outweigh the potential costs. All of these data points—stress, the behavior of researchers, and the frequency of negative experiences—are relevant to conversations about research ethics and are the types of concerns institutional review boards typically consider when evaluating research proposals.

In addition to the questions above, our survey provided another way to assess whether people are dissatisfied with their experience on MTurk: we asked people if they would take a job over MTurk if one were available. This query again speaks to the claim that people spend time on MTurk because they cannot find work elsewhere. 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. Perhaps most convincing, 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.

How much do people make?

MTurk has been criticized as a platform that offers extremely low wages—so low in fact that they render the entire platform unethical. When we asked people how much they can earn per hour the median response was $6.61 (mean = $10.41). People's answer to this question did, however, vary by how they characterized their time on MTurk (and worker experience). People who characterized MTurk as a full-time job reported that it is possible to earn a median wage of about $10 per hour (mean = $26.75) while people who reported working on MTurk part-time or as paid leisure both reported it is possible to earn between $6 (mean = $9.43) and $7 (mean = $8.02) per hour. Even though our question asked how much "people can make per hour," the numbers people provided are in line with past research asking people how much they actually make (e.g., Berg, 2016; Hitlin, 2016) and with the wage estimate we presented from the CloudResearch database. Furthermore, our data reinforce the idea that people with more experience on MTurk often earn more money (e.g., Kaplan et al., 2018). When considered altogether our data, and that from previous studies, seem to suggest that one of the most often cited statistics about wages on MTurk—that people earn $2.00 per hour (Hara et al., 2018)—is an outlier.

Nevertheless, it is important to note the difficulty of estimating hourly wages on Mechanical Turk. Past studies that ask people to report their earnings have been criticized on the grounds that people may have inaccurate memories and that these self-reports do not account for unpaid time between tasks (e.g., Hara et al., 2018). Further complicating estimations of wages is that average hourly wages likely show a bimodal if not multimodal distribution. This is because how much people earn per hour is often associated with their experience on the platform. People with more experience often use scripts, plug-ins, and other tools to quickly find and accept high-paying HITs (Kaplan et al., 2018), minimizing time between tasks. People who are new

to the platform seldom use such tools meaning they are often beaten out of high paying HIT opportunities or locked out by the reputation qualifications requesters use (see Robinson et al., 2019). Estimating wages requires deciding which of these two groups is a "better" representation of MTurk.

There are also differences in wages between workers at the task level. Two workers completing the same task are unlikely to earn the same pay rate unless they complete the task in the same amount of time. This is often not the case. Thus, within any 1 hour of work, workers are likely to complete dozens of HITs that each pay a different effective hourly wage based on how quickly the assignments are completed. Therefore, there is wide variability in how much people earn from completing tasks (Litman & Robinson, 2020b), making the job of calculating wages in hourly terms quite complicated.

Finally, estimating how much people earn on MTurk is complicated by disagreement about whether wages are better represented by a mean or median. Typically, when researchers look at income in a population or in the traditional economy median numbers are preferable over means to avoid the biasing effect of people who may earn several million dollars a year or more. However, when income is examined at the task-level, as is the case in the gig economy and in this report, medians cut out real wages people earn from high paying tasks. As a simple example, imagine if on a given day a worker completed nine tasks that paid $1.00 each and one task that paid $20. It would be misleading to characterize these earnings as $1 per task (median). Instead, it would be more accurate to report the mean wage per task which would be $2.90. Because there is a legitimate long tail to task-specific wages on MTurk, high paying tasks may be relatively infrequent (i.e., once a day or every few days), but they still represent actual money earned by people on the platform and these earnings are not concentrated in the hands of just a few individuals like in the traditional economy. Therefore, choosing the median as the best representation of task-level wages on MTurk risks underestimating how much people actually earn (e.g., Hara et al., 2018).

Beyond the particulars of 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 value 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. All these characteristics may be especially important for people who think of MTurk as a form of paid leisure or a part time job because few other avenues of making money offer the same flexibility. Interestingly, 23% of people said they work on MTurk because they can make more money than in a traditional job.

Aside from the issue of direct earnings, we reasoned that the things people value about MTurk may have hidden value that augments 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 median hourly wage people reported was just over $20 per hour. Among people who said they use MTurk as a form of full-time work, the number was greater than $26.50 per hour. Hence, our data provide a picture of people on MTurk and how much they earn that is very different from the common narrative.

Limitations

Among the limitations of our surveys is the possibility of non-response bias. Across strata, response rates ranged from 56 to 11%. While it would be ideal to obtain equally high response rates across all groups, this is impossible because more active workers will always be more likely to participate in studies than less active workers. Even though non-response bias is a concern, we believe it is mitigated by several factors. First, the response rates in all groups within our survey were higher than typical response rates in random digit dialing surveys, which are the gold standard for research in polling and public opinion. Second, a comparison of the people who participated in our survey to those who were contacted but did not respond and those in the overall sampling frame revealed little to no difference across variables such as household income, education, race, ethnicity, age, and gender (see Tables S1 and S2 in the Supplementary Materials). Third, on virtually all measures within the survey, participants in different strata provided similar responses (see Supplementary Materials). Finally, even if some sampling bias affected the results of our surveys, the current sampling strategy is a marked improvement over all previous surveys of MTurk workers. Because virtually all past surveys have made no attempt to control for the activity level of workers, past data characterizing MTurk oversamples

a relatively small group of highly active workers (see Robinson et al., 2019). Our study is the first to address the issue of sampling bias on MTurk by developing a novel method for gathering representative data based on probability sampling stratified by worker experience. Therefore, our data provide one of the most complete pictures of MTurk to date.

A second possible limitation of our survey was our decision to stratify the sample based on worker experience. We believe this stratification is crucial to capturing a representative view of MTurk and corrects for a form of sampling bias that has affected nearly all previous research. Some people may, however, argue that inexperienced workers are a poor representation of MTurk. People with little experience on the platform may have little understanding of how much people can earn, may have less interaction with abusive requesters, and may be less stressed by the work required to find and complete HITs. In other words, some people may argue that our sample contains far too many people who are responsible for completing just a small share of the overall HITs and not enough of the people who complete most of the work (see Robinson et al., 2019). In response to such critiques, we would point to the consistency of our results across levels of worker experience and to an analysis within our supplemental materials. In the analysis, we weighed our data to each group's share of completed HITs, drawing numbers from Robinson et al. (2019). Although the numbers in this analysis are slightly different from what we present here, the main message does not change: representative data show a disconnect between how MTurk is often described and how people actually experience it. In addition, the financial picture of people on MTurk actually improves when examining the people who are most active on the site.

Finally, a third limitation of our surveys may be our use of the U.S. workforce and polling data from Gallup as a point of comparison for MTurk. Comparing what people do on MTurk to the wide variety of jobs people perform to earn a full-time living in the U.S. economy may not be a good comparison. For example, a substantial number of households in the U.S. have earnings of more than $100,000 per year while people on MTurk appear to complete tasks for hourly earnings near $6.00 or $7.00 an hour. Given the greater variability in both wages and job-related stressors among the U.S. workforce, it may be expected that people on MTurk would report both less stress and less satisfaction with wages. Despite these limitations, we believe the nationally representative data provide some context for interpreting our MTurk data.

Conclusions

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, people 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 is ethical to conduct behavioral research on Mechanical Turk.

Author Contributions

	AJM	CR	JR	SNJ	LL
Conceptualization					
Data curation					
Formal analysis					
Investigation					
Methodology					
Project administration					
Software					
Writing – original draft					
Writing – review & editing					

Declarations

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

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

Arechar, A. A., & Rand, D. G. (2021). Turking in the time of COVID. *Behavior Research Methods.* https://doi.org/10.3758/s13428-021-01588-4

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.

Berg, J. (2016). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comparative Labor Law & Policy Journal, 37*(3).

Brenan, M. (2019). *Americans Feel Generally Positive About Their Own Finances*. Gallup. https://news.gallup.com/poll/249164/americans-feel-generally-positive-own-finances.aspx

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6*, 3–5. https://doi.org/10.1177/1745691610393980

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 (n.d.). *Disability impacts all of us.* https://www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html

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

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

Damer, E. (2019). Stop using MTurk for research. *Prolific Blog.* https://blog.prolific.co/stop-using-mturk-for-research/

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

Fair Crowd Work. (n.d.). Amazon Mechanical Turk - Review. Retrieved June 9, 2022, from http://faircrowd.work/platform/amazon-mechanical-turk/

FairWork, (2021). FairWork 2021 Annual Report. https://fair.work/wp-content/uploads/sites/131/2022/01/Fairwork-Annual-Report-2021.pdf

Federal Election Commission (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). Variation among Mechanical Turk Workers Across Time of Day Presents an Opportunity and a Challenge for Research. 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

Gallup. (n.d.). *Work and Workplace*. Gallup.Com. Retrieved August 10, 2022, from https://news.gallup.com/poll/1720/Work-Work-Place.aspx

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

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.

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/

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). 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*. 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

Ingraham, C. (2019). *Nine days on the road. Average commute time reached a new record last year*. The Washington Post. https://www.washingtonpost.com/business/2019/10/07/nine-days-road-average-commute-timereached-new-record-last-year/

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

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

Kennedy, C., & Hartig, H. (2019). Response rates in telephone surveys have resumed their decline. Pew Research Center. Available at: https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/

Litman, L., & Robinson, J. (2020a). Introduction. In L. Litman & J. Robinson (Eds.), *Conducting online research on Amazon Mechanical Turk and beyond* (pp. 1–26). Sage Academic Publishing.

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

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US and India based workers on Mechanical Turk. *Behavior Research Methods, 47*, 519–528. https://doi.org/10.3758/s13428-014-0483-x

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

Marken, S. (2018). Still listening: The state of telephone surveys. *Gallup* [blog post]. Available at: https://news.gallup.com/opinion/methodology/225143/listening-state-telephone-surveys.aspx

Meyers, E. A., Walker, A. C., Fugelsang, J. A., & Koehler, D. J. (2020). Reducing the number of non-naïve participants in Mechanical Turk samples. *Methods in Psychology, 3*, 100032. https://doi.org/10.1016/j.metip.2020.100032

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

Norman, J. (2018). *Four in 10 U.S. Workers Think They Are Underpaid.* Gallup. https://news.gallup.com/poll/241682/four-workers-think-underpaid.aspx

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

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

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/

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/

TurkerView. (2019). Writer who never learned to drive works for Uber. Makes $0.97/hr. https://blog.turkerview.com/writer-who-never-learned-to-drive-works-for-uber/

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

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. https://doi.org/10.1037/pspa0000171