Article

Bulletin of Science, Technology & Society
1–14

DOI: 10.1177/02704676211003808
journals.sagepub.com/home/bst
SAGE

Crowdsourcing as a Tool for Research: Methodological, Fair, and Political Considerations

Stephen C. Rea[1], Hanzelle Kleeman[1], Qin Zhu[1], Benjamin Gilbert[1], and Chuan Yue[1]

Abstract

Crowdsourcing platforms are powerful tools for academic researchers. Proponents claim that crowdsourcing helps researchers quickly and affordably recruit enough human subjects with diverse backgrounds to generate significant statistical power, while critics raise concerns about unreliable data quality, labor exploitation, and unequal power dynamics between researchers and workers. We examine these concerns along three dimensions: methods, fairness, and politics. We find that researchers offer vastly different compensation rates for crowdsourced tasks, and address potential concerns about data validity by using platform-specific tools and user verification methods. Additionally, workers depend upon crowdsourcing platforms for a significant portion of their income, are motivated more by fear of losing access to work than by specific compensation rates, and are frustrated by a lack of transparency and occasional unfair treatment from job requesters. Finally, we discuss critical computing scholars' proposals to address crowdsourcing's problems, challenges with implementing these resolutions, and potential avenues for future research.

Keywords

crowdsourcing, crowd workers, research ethics, methods, ethics of technology

Introduction

Crowdsourcing platforms (e.g., Amazon Mechanical Turk [MTurk], Prolific, Clickworker, etc.) have become powerful tools for researchers in the social, psychological, behavioral, and computer sciences. These platforms allow individuals—known as "crowd workers"—to complete short, online tasks, such as data labeling, transcription, and participation in research surveys or experiments, for pay. Proponents claim that they possess certain strengths compared with existing tools for data collection. For example, crowdsourcing platforms can help researchers quickly recruit enough human subjects from a diverse demographic pool at a low cost to generate significant statistical power. However, critical computing scholars have expressed serious concerns about the use of crowdsourcing platforms in research, such as labor exploitation and data quality.

To better understand these concerns, this article starts with a brief contextual overview of crowdsourcing as a research tool. It then conceptualizes concerns about crowdsourcing as a data collection tool with respect to three different dimensions: methods, fairness, and politics. Through a critical analysis of publications that have used crowdsourced data, a comprehensive literature review of research about crowdsourcing, and qualitative interviews with crowd workers, this article explores crowdsourcing's value propositions for academic research and common problems that arise for researchers and workers. Finally, this article examines proposals to address concerns about crowdsourcing as a research tool, and potential challenges with implementing these resolutions.

Crowdsourcing as a Research Tool: A Brief Contextual Overview

Former *WIRED* magazine editor Jeff Howe first coined the term "crowdsourcing"—a portmanteau of "crowd" and "outsource"—defining it as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" (2006). As computing power and telecommunications speed and stability have improved over the last two decades, many individuals and organizations have turned to digital, web-enabled crowdsourcing platforms to solicit on-demand contributions remotely from

[1]Colorado School of Mines, Golden, CO, USA

Corresponding Author:
Stephen C. Rea, Humanities, Arts, and Social Sciences, Colorado School of Mines, 412N Stratton Hall, Golden, CO 80401, USA.
Email: srea@mines.edu

anonymous crowd workers who have registered with the platform to make their services available to "job requesters." Requesters recruit workers for a range of activities at different levels of compensation. For example, some crowdsourcing platforms allow corporations to solicit design proposals from high-skilled, freelance creative workers who compete with one another as individuals or in teams for prize money (Schlagwein et al., 2019). Others afford workers the opportunity to negotiate fees, leveraging their particular skills and experience to meet requesters' needs (see Felstiner, 2011).

Typically, however, crowd work consists of "human intelligence tasks" (HITs), micro-tasks requiring certain knowledge and skills that cannot be replicated by machine intelligence. Requesters provide workers with instructions for completing HITs, set the price that they are willing to pay—usually between a few cents and a couple of dollars—and a time limit within which workers must finish the HITs once they have accepted them. Workers review lists of HITs that are currently available and accept those that they are interested in and for which they are qualified on a first come, first serve basis. Crowdsourcing platforms earn revenue by collecting a percentage from job requesters who use their services. For example, the world's largest and most widely used platform, MTurk, charges a base fee of 20% on the compensation that is paid to workers, an additional fee of 20% for batches that include ten or more individual HITs, and fees for using MTurk's qualifications system ranging from 5 cents to 1 dollar per task.[1]

While initially conceived as a way for organizations to outsource piecemeal work, academics quickly realized that they, too, could use crowdsourcing platforms for certain research projects. For instance, a psychologist could design an experiment to test respondents' reactions to a series of images and post it to MTurk. Or a political scientist could disseminate a survey about voter opinions on a number of issues or candidates in an upcoming election. Neither of these research studies necessitates face-to-face interaction, and so recruiting crowd workers to participate in them is a viable option. Crowdsourcing's value propositions for academia are obvious: it offers a convenient way to reach a vast number of potential participants without having to go through an external recruiter, schedule times for lab visits, or find ways to advertise and disseminate article surveys. Data can be collected quickly—often within a matter of hours or days—and at scale, and researchers can prescreen participants by restricting which workers can accept a HIT based on platforms' different qualification criteria. Most significantly, crowdsourcing costs far less than alternative methods, allowing researchers with minimal funding to design studies that would otherwise have been too expensive, and to allocate their resources more efficiently; one evaluation of crowdsourcing platforms for survey research in political science estimated that while "per subject costs for typical undergraduate samples are about $5 to $10,[2] for nonstudent campus samples about $30, and for temporary agency subjects between $15 and $20," paying crowd workers on MTurk at a rate of $6/hour resulted in "a per-respondent cost of $.55" (Berinsky et al., 2012, p. 353).

Although the extent to which crowdsourcing has affected expectations for academic research is unclear, in highly competitive institutional environments where "publish or perish" is the norm, being able to collect, analyze, and produce results about data more quickly poses an advantage. However, increased use of crowdsourcing for academic research also raises a number of concerns. Practitioners and critics—who are often one and the same—have questioned how representative crowd workers are in comparison to broader populations, and whether insights produced from the data that they provide are generalizable (Williamson, 2016a). Others point to unequal power dynamics between requesters and workers (Samuel, 2018), low compensation rates for completing HITs (Silberman et al., 2018), and the economically precarious circumstances that many crowd workers are in as reasons to question using crowdsourcing in the first place (Berg et al., 2018). And as in other areas of online-mediated research (e.g., Chee et al., 2012), questions of research ethics and the extent to which consent can be validated are not entirely resolved. Through a critical examination of crowdsourcing, this article offers a step toward improving conditions for requesters and workers, and addressing crowdsourcing's shortcomings as a research tool both methodologically and in terms of research ethics.

Method

Our analysis stitches together insights from a comprehensive literature review of research about crowdsourcing, with special attention to its use in academia; a critical analysis of journal articles that have used crowdsourced data; and semi-structured interviews with crowd workers. By employing a mixed-method approach, we were able to assess contemporary debates about crowdsourcing as a tool for academic research, how crowdsourcing is being used and written about by academics currently, and some of the common problems that workers encounter with crowdsourced research. Given the existing literature and our own recruitment strategy, our data are primarily drawn from MTurk. MTurk is not the only crowdsourcing platform that academics use for research, but it is the most popular and widely used. Moreover, among our interviewees who completed academic research tasks on multiple crowdsourcing platforms, they noted that MTurk was where they found the vast majority of their work. While there may be small differences with respect to other platforms, we are confident that focusing on MTurk affords insights about the most relevant issues affecting crowdsourced research.

For the critical analysis, we identified three disciplinary areas in which crowdsourcing has become a popular method for data collection: psychology, political science, and marketing/consumer research. While these are by no means the only academic disciplines in which researchers use crowdsourcing, researchers in all three have historically relied upon large survey instruments to evaluate population-level patterns, such as social conformity, voter behavior, and reactions to advertising campaigns. Therefore, it stands to reason that researchers accustomed to this method would find benefits in the cost and time savings that crowdsourcing platforms afford. For this reason, among others, MTurk specifically has been evaluated and recommended as a research tool in psychology (Bates & Lanza 2013), political science (Berinsky et al., 2012), and marketing/consumer research (Goodman & Paolacci 2017). Additionally, in our interviews—described below—crowd workers described completing academic surveys, identifying psychology studies, studies about political sentiment, and market research posted by business school researchers, specifically.

We used ScimagoJR—a publicly available web portal that calculates scientific journal indicators—to search for journals with the highest impact factors in each of these three disciplines. We then eliminated annual reviews and theory-focused publications in order to focus on crowdsourcing's used in empirical research. With those qualifications in place, we selected three journals from each discipline: from psychology, *Psychological Science*, *Journal of Experimental Psychology: General*, and *Journal of Personality & Social Psychology*; from political science, *Political Analysis*, *Political Psychology*, and *Political Behavior*; and from consumer research/marketing, *Journal of Marketing*, *Journal of Marketing Research*, and *Marketing Science*. Next, we performed searches on each journal's homepage for "crowdsource," "crowdsourcing," and "Mechanical Turk," and selected from among the top search results. Doing so allowed us to capture not only articles that had used MTurk—the most popular crowdsourcing platform for data collection—but also those that used crowdsourcing without specifying which platform(s) they used. We then verified that each article had used crowdsourcing as a method, discarding any articles that referenced crowdsourcing and/or MTurk as objects of analysis rather than as research tools. Some journals had more results than others. We limited the total number of articles from any single journal to 5. In total, we analyzed 38 different articles published between 2011 and 2019. Our analysis is not intended to be statistically robust or offer a generalizable conclusion; rather, we provide a comparative look of some of the ways in which crowdsourcing has been used as a tool for research in published work from different disciplinary approaches.

We recruited interviewees by posting calls for participation in two online forums that are popular with crowd workers—the *Turker Nation* Slack Channel and *MTurk Crowd*—and by creating a recruitment HIT on MTurk. Both our study and decision to use MTurk were approved by our university's Human Subjects Team beforehand. The interviews each lasted roughly 1 hour. Workers were paid $20 for a phone interview and $30 for a video interview. 19 respondents contacted us after seeing the forum posts and an additional 13 accepted our MTurk HIT, for a total of 32 interviewees. While most of our interviewees had participated in academic research studies that were posted to other platforms such as Prolific and Clickworker, all of them reported finding most of their HITs on MTurk. Again, the bias in favor of MTurk is symptomatic of our recruitment strategy, but should not invalidate our findings regarding the use of crowdsourcing in academic research more generally.

Interview participants had to be at least 18 years old, located in the United States, and a current crowd worker. Because our primary concern was collecting qualitative data about crowd working experiences and challenges with crowdsourcing platforms and were not seeking a statistically representative sample, we did not include demographic quotas in our research design. 19 participants were men and 13 were women, while the average participant's age was 40 with a high degree of variability (range = 51 years; standard deviation = 14.34 years). Compared with Ipeirotis's demographic data of Turkers (2010b), our sample skewed heavily male (Ipeirotis found that 70% of Turkers were women) and slightly older (Ipeirotis found that 54% of Turkers were between 21 and 35 years old, while 50% of our interviewees fell within this range). The average length of engagement with crowd work was 3.5 years, with a range between 5 weeks and 12 years. 16 participants were full-time crowd workers, while an additional 2 had previously done it full-time and 1 had become a full-time liaison helping job requesters post their tasks to different crowdsourcing platforms. However, all of the part-time crowd workers reported that it provided a necessary supplement to their other sources of income.

We did not ask about education or income level specifically. Based on their descriptions of previous or current jobs, explicit statements about being enrolled in a degree program, and references to holding specific degrees, we can confidently state that at least 5 participants were either college graduates or current college students, though it is likely that as many as 13 more had completed at least some college. Similarly, crowd workers were motivated by a range of different economic incentives: two had turned to crowd work after losing their jobs in construction during the Great Recession; one had recently lost his job and home; one had recently been fired; five were between jobs and crowd working while looking for full-time employment elsewhere; four had health issues or disabilities that prevented them from working outside the home; five were responsible for taking care of small children and could not afford childcare to be able to work outside the home; 4 were retired and used crowd work earnings to supplement retirement savings and/or Social Security payments; two had previously been involved in online resale activities and saw crowd work as a "natural" transition; eight reported that their crowd work earnings were necessary supplements to other income for meeting basic household expenses; and four did crowd work "just for fun" or to have something to occupy their free time.

Critical Analysis

In general, the authors of the articles that we analyzed did not appear to follow any sort of disciplinary or journal-specific standards for describing how they used crowdsourcing in their research. Some simply included a note in their methods section that they had, for example, posted a survey HIT on MTurk, but did not elaborate further. Others shared more details about their sampling strategies, participant demographics (e.g., age, gender, education level, income level, etc.), and research design. Despite these inconsistencies, there were a few observable patterns that help frame the considerations of methods, fairness, and politics that we explore in the following sections.

Details About Compensation and Completion Times for HITs

Indeed, 16 of 38 articles declared the specific dollar amounts that crowd workers were paid for their participation. An additional 5 articles mentioned that participants were paid, but did not specify how much each HIT had been worth. This does not necessarily mean that the authors of the other 17 articles offered no compensation whatsoever, but does indicate that they did not consider it necessary to disclose that information. Of the 21 total articles that did mention payments, 7 provided estimates for how long participation would take (i.e., the number advertised in their HIT posting), and 3 of those 7 either calculated the average time that it took participants to complete the task or enforced time limits.

Among the seven articles that shared both payment details and information about completion times, the accompanying compensation rates—real, in the case of those that measured average completion times, and theoretical for those that only shared estimates or time limits—varied greatly, from $3/hour to $22.50/hour. The average hourly rate was $9.69/hour with a standard deviation of $6.31, indicating a high degree of variance. Compensation rates in 4 of the 7 were greater than the U.S. federal minimum wage of $7.25/hour,[3] but the actual compensation-per-HIT ranged from 7 cents to $5, with an average of $3.46 (standard deviation = 95 cents; mode = $1), with estimated and actual completion times ranging between 2 and 10 minutes. To give some sense of researchers' perceived norms regarding payments, two articles explicitly stated that their compensation rates were higher than those for other HITs on MTurk, yet both were below federal minimum wage ($3.90/hour and $5.63/hour, respectively) and neither provided any evidence to support their claims. We detail the significance of wages for crowd workers in the Fairness Considerations section below.

Concerns About Validity

At least 23 articles addressed potential criticisms of crowdsourcing's validity as a research method. Whether these were preemptive or added in response to peer reviewers' comments is unknown, but does indicate the degree to which crowdsourcing is still novel in academic research communities and likely to provoke skepticism from some observers. While acknowledging potential limitations, 13 articles used complementary methods such as regression analysis, sample weighting, power analysis through pilot testing, reliability calculations (e.g., Cronbach's α), robustness checks, and comparing against nationally representative data sets (e.g., Census data) to assess their sample validity and the generalizability of results. Another strategy that appeared in 10 articles was referencing one or more other publications that had either used crowdsourcing in their own data collection or done critical analyses of crowdsourced research. We provide more details about common criticisms in the Methodological Considerations section below.

Quality Assurance and Participant Validation

In addition to concerns about statistical validity and generalization, crowdsourced data are often scrutinized for the authenticity of crowd workers' responses and their quality level. The authors in our critical analysis employed different strategies for mitigating inauthentic participation and poor-quality data. 14 used attention and/or manipulation checks—for example, inserting a simple math problem that is unrelated to the research design into a survey—as a means of verifying that workers were indeed "real" people; if a participant failed one of these checks, their contributions were typically excluded from the data set. Additionally, 20 articles used MTurk's qualifications system to restrict participation based on certain criteria. Commonly used qualifications included requiring participants to be based in the United States, requiring that participants be over the age of 18 years, eliminating anyone who had done a previous study from the same research team, and excluding anyone who was not a native English speaker. 7 of those 20 also used MTurk's "approval" ratings system to eliminate crowd workers who fell below a certain threshold. Whenever a requester approves or rejects a worker's HIT submission, it is calculated into their overall approval rating covering their time working on the platform. Four articles required an approval rating of 95% or better, one required 90% or better, one required 80% or better, and the last did not specify. All noted that they used the approval rating qualification in the hopes of collecting better quality data. We describe the significance of rejections and MTurk's approval ratings system in greater detail in the Political Considerations section below.

Research Ethics

Only five articles explicitly mentioned either having received institutional review board (IRB) approval through their university or obtaining informed consent from participants.[4] As with information about payments, this does not necessarily mean that the other 33 studies did not also abide by ethical standards for research. However, we contend that the fact that the majority of articles did not include anything about basic research ethics—for example, IRB protocols or statements about informed consent—reflects perceptions about crowd work and crowd workers that help reinforce unbalanced power dynamics between researchers and research participants.

Methodological Considerations

Compared with other data collection tools, crowdsourcing has certain noticeable advantages and disadvantages for academic researchers. One of its greatest strengths is that it may increase and diversify sample populations by making researchers less dependent on undergraduate populations, which are often criticized for their so-called "W.E.I.R.D. (White, Educated, Industrialized, Rich, and Democratic)" characteristics (Palmer & Strickland, 2016). Crowdsourcing also allows researchers to access large numbers of participants and process huge, unique data sets quickly (Stamm & Eklund, 2017), generating significant statistical power (Pittman & Sheehan, 2016). Researchers may also use crowdsourcing platforms' tools, such as MTurk's qualifications system, to more easily target specific sample populations. For example, participation can be restricted to a specific geographic area (e.g., only New Jersey residents), or certain demographic profiles (e.g., African American women over 40 years). And perhaps most significantly, crowdsourcing reduces monetary costs associated with data collection from large populations by, for example, making it possible to bypass third-party recruiters, avoid paying for survey canvassers, and save on overhead for transportation and scheduling for laboratory visits.

Nevertheless, scholars have also expressed concerns regarding crowdsourcing's methodological limitations. While crowd workers may be more diverse than undergraduate student populations, there are also a number of ways in which they are not representative of more general populations. For instance, the demographics of crowd workers on MTurk's platform skew younger, with smaller household sizes and lower incomes than the average American (Difallah et al., 2018), and tend to be more politically liberal (Berinsky et al., 2012). Participants must also have internet access to work on data collection surveys (Palmer & Strickland, 2016), effectively excluding those with limited or no access. And although tools like MTurk's qualifications system do make it easier to capture more granular population samples, researchers cannot control for the characteristics of crowd workers as a whole. Stamm and Eklund argue that this lack of control not only "challenges the scientific rules of representativeness" but also "leaves methodological designs vulnerable to researchers' implicit assumptions about the crowd" (2017). Furthermore, limitations on monitoring participation may impact a study's validity if "those who truly possess the characteristics of interest differ significantly from those who do not" (Sheehan, 2018, p. 147).

Another of crowdsourcing's limitations is the disparity between active workers and the total number who are registered with a given platform. For example, one of MTurk's greatest value propositions is that, in theory, it functions as an on-demand recruitment platform to over 500,000 crowd workers around the world, but that number includes many accounts that are dormant or used only sporadically. The actual number of workers who are active on MTurk at any given time is closer to 2,000 (Difallah et al., 2018). Some observers have called these workers "super-Turkers," that is, a subset of the overall MTurk population who complete the majority of the HITs that are posted (Samuel, 2018; Wazny, 2017).[5] Similarly, Gray and Suri (2019) have likened MTurk labor market dynamics to a Pareto distribution, whereby 80% of tasks are performed by 20% of the workers on the platform. Crowdsourcing platforms' Pareto distributions and phenomena like super-Turkers make it likely that different researchers are reaching many of the same individuals (Williamson, 2016a). Because many survey participants come from a smaller subpopulation within a given platform's worker community, and because different universities often use similar survey instruments (e.g., personality tests for psychology research), the data that they provide are more likely to duplicate what previous surveys have collected. Cole,[6] a 54-year-old full-time crowd worker near Dallas, said, "The thing is, every psychology department in every university uses the same standardized [question set]. So, I've seen the same questions—oh my god—a million times!" The potential for survey results to be skewed and for individual studies to have artificially high levels of correlation therefore also increases with the relative nonnaivete of crowd worker populations (Peer et al., 2017).

A related concern is the inherent difficulty of verifying research subjects' identities on crowdsourcing platforms and the risk of collecting poor quality data. While survey research is always vulnerable to inaccurate or outright falsified responses, that risk is magnified when participants are only identifiable by a pseudonymous string of numbers and letters, as MTurk workers are. Furthermore, crowdsourcing platforms are host to inauthentic users in the form of automated accounts known more commonly as "bots." In the summer of 2018, a psychology graduate student posted on his personal blog about how an unusually large number of responses to open-ended questions in a survey he had posted on MTurk appeared to be "random," and that many of them appeared to come from accounts with the same GPS coordinates.[7] His observations resonated with more widespread complaints about diminished quality of data in the psychology research community with respect to crowdsourced surveys (Bai, 2018). A short time later, articles in *New Scientist* (Stokel-Walker, 2018) and *WIRED* (Dreyfuss, 2018) described a "bot panic" among researchers who worried about their study results being compromised. Other observers argued that the bot problem was likely overstated and that some of the concerns that had been raised could be mitigated with better research design (Miele, 2018).

According to some of the crowd workers who we spoke with, there were some noticeable changes in the wake of the bot panic, such as more attention checks being included in surveys in an effort to trip up automated accounts. However, these efforts also had some unanticipated effects on workers, and according to some of our interviewees led to their work being rejected because they had missed an attention check. Anthony, a 31-year-old Houston area man who had been working part-time on MTurk for nearly 4 years, described the potential ripple effects of the bot panic on data quality:

> MTurk does have a bad reputation of having a lot of bot usage, and that has made many of the requesters paranoid. It hurts workers like myself who are the competent, honest workers who want to get in there and give good quality data. It starts making you wonder, "Why am I even taking my time to give such good data?" I can see why whoever it is that are using bots are cheating the system. I have a hunch that a lot of these are disgruntled ex-workers who once were giving quality data, and they've been through so much that they say, "You know what? I'm going to get back at the system."

Finally, the relatively much lower cost of recruiting participants on crowdsourcing platforms may potentially encourage compromises in research design, as some observers have argued that more expensive data collection incentivizes researchers to be more careful when designing experiments (Morrissey et al., 2019). Some researchers have also raised concerns that payments made to participants, no matter how small, could potentially threaten crowdsourced research studies' external validity (Hassell & Visalvanich, 2015).

Although crowd work is clearly transactional and earning money is the primary motivation for participating, at least 3 workers who we interviewed expressed genuine curiosity and excitement about opportunities to take part in research that could have broader impacts on the world. For example, Bernice, a 41-year-old New Jersey resident, told us that her favorite HITs were those where she felt that she was "contributing to science." She explained, "If you're just talking about academic research, these are the [people] that are doing the research of tomorrow, and we're the raw data!" Bernice's comments are evidence to support arguments that crowdsourced data are of comparable quality or better than what can be collected using other methods (Paolacci & Chandler, 2014). In fact, studies have found that compensation rates have no effect on data quality, leading some to argue,

> When it is possible to use non-financial rewards, such as harnessing intrinsic motivation, the quality of the work will be as good or better than using financial rewards, and therefore work can be accomplished as effectively for little to no cost. (Mason & Watts, 2009, p. 84)

While crowd workers' self-reported intrinsic interests in academic research may not assuage all fears about validity, they do offer evidence that workers understand well the gravity of their participation and the importance of providing honest, high-quality data.

Fairness Considerations

Crowdsourcing's most frequent object of criticism—both in the literature and in our interviews with crowd workers—is the relatively low rates of pay that requesters offer for their HITs. A 2016 Pew Research Center nonprobability survey of 3,370 MTurk workers found that 52% made less than $5/hour completing HITs, and that 89% of HITs posted to the platform paid $1 or less. Evidence suggests that a significant fraction of crowd workers depend upon crowdsourcing as their primary source of income, though they are not in the majority; 25% of workers in Pew's survey reported earning all or most of their income from doing HITs on MTurk (Hitlin, 2016; see also Ipeirotis, 2010a). While these data include HITs that are posted by requesters from a variety of organizations, not just academics, the issue of fair pay applies to crowdsourced research studies as well. The crowd workers who we interviewed reported that academic researchers tended to pay better than other requesters, but that there were exceptions to this rule. According to Victoria, a 46-year-old crowd worker in Austin, Texas who has been working on MTurk full-time off and on since 2015, "There're some academics that pay very poorly. It just varies wildly."

As stakeholders who are not primarily motivated by profit margins, academic researchers should, theoretically, be sympathetic to workers' roles and motivated to pay workers fairly for their participation, however they may define "fair." However, budget constraints or beliefs in the intrinsic utility of academic research may help some researchers justify paying less for their HITs. Filbert, a 70-year-old retiree in Indiana, hypothesized, "What I think happens is the requester, maybe some big university, has, say, only $5000, but they need 1000 survey takers to get a good sample of their target population. So, they can only pay $5 per worker." He said that for some surveys, $5 would be an above average rate of pay, but for others that take a long time to complete or are especially complex, it can feel like the requester is "trying to get free information out of you." Miranda, a 59-year-old full-time crowd worker from New York City, added,

> I understand that coming from your perspective, you're doing research, you're trying to get information, your concern is not how much the worker makes. You're working within a budget, you have a grant. But that doesn't mean you should exploit people.

As noted in the previous section, unfair pay may affect data quality if participants do not feel that their work is being properly valued. However, studies have shown that data quality does not diminish with compensation rates (Bohannon, 2011; Litman et al., 2015). All of our interview participants reported that they completed tasks honestly and accurately regardless of payment due to fear of rejection, which we address in the following section. Anthony described a hypothetical scenario:

> Sometimes you just have to get in there and test it, and then sometimes you'll spend all this time on the HIT, and midway through you discover, "Wow, this is taking much longer, I didn't think it would take this long!" Since you already invested that time, it's like there's no turning back now, so you have to [finish it].

At worst, crowd workers will simply avoid low-paying HITs, return them without completing them, or never accept HITs from low-paying requesters in the future. As Anastasia—a 41-year-old part-time crowd worker in Austin, Texas—explained, "A lot of these [HITs]'ll come across, you know, 30-to-45-minute surveys for a dollar. No, I am not about to [do those]. And then you into the ones that are just endless bubbles, and I'll throw that away."

Crowd workers' legal status as "independent contractors" who have voluntarily entered into agreements with individual requesters also helps rationalize low pay, such as in the following:

> There is an established contract between the requester and worker to do the work at the agreed wage independent of the time required to do the task. . . . The working conditions and hours are wholly determined by the worker. There is absolutely no direct or indirect obligation or constraint on the workers to do any work on Mechanical Turk. In other words, the decision to engage in the contract is completely at the worker's liberty, a situation that rarely, if ever, exists in other employment situations. (Mason & Suri, 2012, p. 16)

A few of our interviewees agreed with this sentiment. Matty, a 43-year-old Wisconsin man who has been crowd working part-time for the past 5 years, said,

> If you don't find [the pay] to be fair, don't do it and stop your bitching, is how I see it. If you want to sit there and work for 3 bucks an hour, and the HITs are fun, then why not do them?

D'Arcy, a 55-year-old part-time crowd worker from northern Ohio, said that she felt the compensation on MTurk was fair because "you really don't have to buy special clothes, you don't have to buy lunch at work, and you don't have to drive anywhere." She added, "Do I wish it paid more? Probably. But I will accept the pay for being able to work at home."

Critics, however, argue that crowd workers' time commitments and legal classification are irrelevant to the issue of fair pay (Silberman et al., 2018). There appears to be a divide between academics whose primary interest in crowdsourcing research is expediency—that is, collecting high-quality data quickly, cheaply, and at scale—and those more concerned with social and economic justice. To wit, compare this characterization of crowd work as "a tremendous amount of human computation power [that] is available for accomplishing jobs almost for free" (Allahbakhsh et al., 2013, p. 80) with the insight that "the very abstraction that lets human computation researchers access thousands of workers in a click also renders invisible the practical problems faced by people in the crowdworking workforce," including lack of recourse, exposure to security and privacy threats, and no standard wages (Silberman et al., 2010, p. 40). Although it does not put any restrictions on compensation for HITs, Amazon recommends that MTurk requesters pay around $6/hour (Sheehan, 2016). In the absence of any formal regulations governing crowdsourcing labor markets, the overall effect on wages is a "race to the bottom," so to speak. Moreover, some academics argue that they are incentivized against paying high wages if they want high-quality data; in a review of MTurk for *Science*, a political scientist remarked,

> If you offer more than a dollar, you attract the spammers who sort jobs by level of pay. . . . You have to find the sweet spot where the payment is not too high but still attractive enough for most Turkers. (quoted in Bohannon, 2011, p. 307)[8]

Pay is only one aspect of fairness in crowdsourcing, however. Even if platforms like MTurk did mandate that requesters pay fairly—however that may be defined—the substantial amount of time that crowd workers spend looking for HITs would still go uncompensated. As a result, crowd workers' effective hourly wages are lower than they might otherwise appear. A recent study that used a browser plugin to track workers' activities on MTurk found that their average hourly wages fell between $3.13 and $3.48/hour, even though HITs paid an average of $11.58/hour (Hara et al., 2018). Given that time is quite literally money for crowd workers, accounting for how long it will take to complete a HIT is critical when deciding whether or not to accept it. Requesters often include expected completion times when they post a HIT, but those expectations are not always accurate. Filbert noted,

> In the advertisement [for the HIT], when they're trying to get you in, they'll say it's going to take only ten minutes, but a lot of times it will take two or three times as long as they said it does.

All except two of the workers who we spoke with use community-created scripts—small lists of commands that execute automatically—to record how long they spend on

HITs and calculate and their hourly rates. Many also rely on feedback from other workers through community resources such as the rating and review site Turkopticon (discussed in greater detail in the next section) or the sub-Reddit r/HITsWorthTurkingFor in order to evaluate gaps between estimated and actual completion times for HITs. For the most part, inaccurate time estimates are unintentional consequences of requesters' unfamiliarity with their own materials and crowd worker community norms. However, there was also a widespread perception among our interviewees that a small number of requesters purposefully manipulate time estimates and time limits on the HITs that they post. As Filbert told us, "Some of them are real tricky: They'll narrow that time down, and then they know the average person can't possibly do that survey in that frame of time." If the HIT "times out," the worker typically does not get paid for their participation because they have failed to fulfill the terms of the contract. But the requester still has access to the data that the worker has provided.

A related obstacle that workers encounter is having their submission rejected because they completed a HIT "too quickly." Experienced workers are often able to do good quality work faster than the requester expects. Miranda explained, "When you've encountered the same thing so often, you tend to be very fast. So, when they reject you because they think it should have taken you 15 minutes and it took you 5, it's very unfair to me." As Matty pointed out, misunderstandings could easily be avoided if the requester was explicit about their time expectations in the HIT instructions, or programmed the HIT so that the worker could not progress or submit until a certain period has passed; "[A requester] released a bunch of HITs and they rejected a lot of mine because I didn't take long enough to them," he said. "If you want me to take 30 seconds on something, tell me you want me to take 30 seconds!" While they emphasized that such occurrences were rare, three of the workers we interviewed explicitly described experiences like these when they felt that their work had been stolen from them, including by academic requesters.

Filbert, Miranda, and Matty's experiences highlight another aspect of crowdsourcing where fairness dilemmas frequently arise: rejections. On MTurk, requesters have the right to reject any submissions that do not meet their expectations without having to pay the worker. Workers' approval rates are recorded and made visible to requesters, and as noted above, requesters can set qualifications for their HITs that prohibit anyone below a certain approval threshold from accepting them. As Cole explained, "If you're below 99% approval, the really, really juicy [HITs], you're not going to get them. You go below 95? You're going to get nothing but crap." New requesters are often unfamiliar with the devastating effects that rejections can have for workers. Victoria remembered, "One time I had somebody who had a broken study, and she just took everything off the platform, and she just rejected everybody: 'Sorry, I can't pay, it didn't work.'" Rejections were a bigger concern for our interviewees than even fair pay. As Cole put it, "I don't even want your money, I just don't want that rejection."

Political Considerations

Understanding what rejections entail for crowd workers and how to use them responsibly is crucial for researchers' ethical use of crowdsourcing. But rejections also demonstrate crowdsourcing's unbalanced power dynamics, from which a number of other, political problems arise. On its face, MTurk's rejection tool is a quality assurance mechanism: the threat of their work being rejected and future opportunities being jeopardized because of lower approval rates incentivizes workers to be careful and diligent. However, requesters have unilateral discretion over rejections, and as a result they can exercise significant control over workers' fates. Molly, a 69-year-old full-time crowd worker from New Jersey, noted that this imbalance applied to academic and nonacademic requesters in equal measure: "You could reject everybody and get all the data free, and if this is your only research project—like, you're going to get your PhD and you're out of there—why would you even care?"

Moreover, requesters on MTurk are not required to provide workers with any explanation about why their work was rejected. Because approval rates are so vital to their ability to work on the platform, workers do their best to avoid rejections, and explanations are crucial resources in this regard. Cameron, a 22-year-old full-time crowd worker in central Pennsylvania, told us,

> Usually whenever I get rejected, I can see the reason why they rejected me. But some people will just put an "X" and that's it. If you give me a reason for rejecting me, and I remember the HIT, I'm like, "Okay, that makes sense, I'm cool with it."

One consequence of the power imbalance over rejections is an information asymmetry that makes it more difficult for workers to learn from mistakes and apply that knowledge to submitting better quality work in the future.

Lack of explanations about rejections is part of a larger problem with communication on crowdsourcing platforms like MTurk. The one-way character of communication was the biggest source of frustration for our interviewees, and made some feel more resentful toward and less trusting of certain requesters. Workers may attempt to contact the requester directly to clarify a HIT's instructions, report glitches such as broken hyperlinks to third-party survey websites, or challenge decisions about rejections. According to Anastasia, academic researchers tend to be better at communicating with workers than other requesters. She told us, "Professors, psychologists, they are 99% more likely to respond and have a conversation with you." However, without any incentives to communicate, there is no guarantee that a requester will respond, whether they are an academic or not.

Communication issues and the power dynamics around rejections illustrate the general "hands off" approach that platform operators like Amazon take regarding disputes between stakeholders who use their service. According to the MTurk Participation Agreement, since Amazon does not set the terms of contracts between requesters and workers, it explicitly refuses to be involved in resolving any disagreements that arise from those transactions (Felstiner, 2011). Although this policy would appear to apply equally to both requesters and workers, since the MTurk power dynamic is tilted so heavily in favor of requesters to begin with, it ends up affecting workers more harshly. For example, MTurk does not provide workers with any tools to negotiate wages for HITs, and there are no mechanisms to ensure that payments are made. Whereas it may make good business sense for Amazon to remain neutral about disagreements between requesters and workers, that decision sends an implicit message to workers about their political status in the crowdsourcing ecosystem. Cameron lamented,

> It really doesn't even feel like Amazon helps at all. There's no support, nothing like that. If you get rejected, you can't really do anything about it but contact the requester and hope they are nice; Amazon won't do anything about it.

Anthony's criticism was even more forceful: "I've never seen any platform like [MTurk] where it's just like the customer service doesn't have any empathy. Like, no regard for the worker."

To address the information asymmetry between requesters and workers on MTurk, design activists and crowd workers collaborated to create Turkopticon, a system that allows workers to rate and review requesters on four variables—communicativity, generosity, fairness, and promptness—on a scale from 1 to 5 (Irani & Silberman, 2013; Silberman et al., 2010).[9] A web browser add-on lets workers see Turkopticon ratings on the MTurk user interface, affording them opportunities to make more informed decisions about the likelihood that a requester will reject their submissions or refuse to pay them. In the absence of any platform-initiated restrictions, Turkopticon also represents an important check on bad faith requesters. Victoria commented that tools like Turkopticon and the web forums where crowd workers congregate "are the closest thing that we have to a union."

However, Turkopticon is not a perfect system. Its vulnerability to manipulation was demonstrated when an academic researcher intentionally submitted fake ratings as part of a study to evaluate how Turkopticon data influenced workers' choice of HITs. The crowd worker community managed to identify the deception and track down the researcher, discovering that he had received IRB approval for the study from his university. Human subjects research conducted in offline communities typically requires that intentional deception may only be employed if it is a necessary part of the research design and that alternative methods that would not achieve the desired result (for discussion, see Cook & Yamagishi, 2008). While this specific IRB's decision-making process in this case is unknown, it could reflect a broader misunderstanding of internet-based research methods and some of the ethical dilemmas that arise when conducting human subjects research that is not face-to-face.[10] Moreover, the incident revealed a lack of understanding in academia about the important role that Turkopticon plays regarding MTurk's power imbalance, and the negative effects that deliberate manipulation of ratings would have on workers' trust in the system (Salehi et al., 2015).

Many of crowdsourcing's political dilemmas are made possible by laissez-faire regulations on platforms. Since crowd workers are technically independent contractors who enter voluntarily into agreements with requesters, they are not considered employees of the platform operator and thus do not enjoy any of the guarantees or protections extended to workers under statutes like the Fair Labor Standards Act and National Labor Relations Act (Felstiner, 2011). For some academic researchers, crowdsourcing's legal ambiguity is a net positive, as the virtual, on-demand, and functionally anonymous character of crowd workers makes it easier to view them as fundamentally different from other kinds of workers, and rationalize subminimum wage payments. Yet even if one believes that crowd workers' lack of employment law protections is justified, it is difficult to escape "the ethical quandary of employed, tenured academics exploiting low-wage work by people at the opposite end of the pay and job security scale" (Samuel, 2018).

In response to these political considerations, some critics have argued,

> Researchers should see the digital worker as a citizen, imbued with all the rights and protections the Fourteenth Amendment calls for. . . . If the digital worker is granted protections as a member of a global digital community, then . . . their crowdsourced labor can be seen as community service rather than a series of simple economic transactions. (Dietz, 2016, p. 264)

Crowd working communities themselves are divided over the best way to improve political equity on the platforms. Some favor traditional labor activism strategies such as collective action and seeking regulatory and/or legislative solutions, while others argue "that the nature of work changed so substantially in crowdsourcing that any 'imports' of norms and rules from traditional employment [are] no appropriate" (Schlagwein et al., 2019, pp. 825-826). For her part, Anastasia voiced a concern that is shared by many crowd workers about the possibility that crowdsourcing's labor market could become more regulated. "The fear is that it is just going to become impossible to get jobs," she said. "I don't know what the trend is of researchers coming to MTurk, or researchers going somewhere else." In the final section, we highlight some of the ways that crowdsourcing's different stakeholders have addressed the considerations of methods, fairness, and politics that we have outlined here, as well as others that have been proposed and the challenges to implementing them.

Proposed Resolutions and Their Potential Challenges

User scripts, web forums, and community-created resources like Turkopticon are invaluable tools for both workers and requesters, as they help improve information asymmetry, unfair practices, and quality of work. For academic requesters specifically, a few resources exist to help streamline their use of crowdsourcing platforms, especially for novices. Third-party intermediaries like MTurk Data provide consulting and publishing services for survey research, leveraging experience with platforms' complex—and somewhat opaque—features so that academics can avoid common mistakes that lead to bad quality data and mistreatment of workers. Sheehan and Pittman (2016) authored a user guide for social scientists interested in getting started with MTurk, including instructions for how to create a HIT and tips from more experienced researchers based on their experiences. And Dynamo, a MTurk community platform, created "Guidelines for Academic Researchers," a comprehensive, freely available list of best practices collectively authored by crowd workers and intended for use by researchers and university IRBs to help academic communities better understand the specifics of crowdsourcing labor markets (Zawacki 2017; Salehi et al. 2015).

While these pragmatic approaches represent initial steps toward improving conditions for crowdsourced research, there are also a number of proposals that aim for more structural reforms. With respect to compensation rates, some critics have suggested that academics could commit to paying research participants the minimum wage for their location (Silberman et al., 2018). Mandating fair wages—however "fair" may be defined—could lead to more equity for crowd workers, but it still would not resolve the issue of effective wage rates discussed above. Moreover, some crowd workers fear that inviting greater regulation of crowdsourcing, such as by imposing a minimum wage, could have unintended consequences that "[threaten] the whole system" (Salehi et al., 2015, p. 1627). Cole worried about an even more drastic scenario: "I do have the fear of, like, all of sudden losing it," he said. "You know, like one day, 'Hey, MTurk just shut down.'"

Others contend that "it should not be up to individual researchers to impose workplace standards" (Williamson, 2016b). A more effective strategy, they argue, would be for institutional stakeholders to incentivize ethical research practices through accountability enforcement. For example, journal editors could commit to publishing only studies that have paid respondents at a rate defined as fair. Research funding agencies could make fair payments for crowdsourced data part of their criteria for awarding grants (Williamson, 2016a). More generally, university IRBs and journals could hold researchers to certain standards of transparency whereby authors are required to disclose what they paid, how they established criteria for rejections, the extent to which they communicated with research participants, and how they established informed consent. Obstacles to realizing these changes could include disagreements about setting standards for fair pay, disciplinary differences with respect to research design, and limited awareness among editors and administrators about crowdsourcing as a research tool.

Ultimately, however, academic researchers represent only part of the community of requesters on crowdsourcing platforms, and it is unclear how effective they can be at convincing platform operators to make changes to their services. Perhaps the most productive things that academics who are interested in promoting a more balanced crowdsourcing power dynamic can do are support crowd workers in their efforts to create cooperative arrangements, and use their own professional networks to advocate for workers' rights. As platforms' terms of service typically prohibit crowd workers from any sort of labor organizing, any such efforts would need to be done responsibly so that workers' livelihoods are not endangered by, for example, having their accounts suspended. An alternative could be supporting shifting the ownership of crowdsourcing platforms into the hands of crowd workers themselves. As Kristy Milland (2016), founder of the popular MTurk community *Turker Nation*, argues,

> Forcing Amazon to set some arbitrary minimum rate may benefit workers, although I question whether it would increase pay or instead just increase the size and scope of each HIT posted. I think we would be better served to focus on one of two alternative options. The first is pressuring requesters of work to consider ethical pay and treatment of workers before they post jobs to be completed. . . . The second . . . is to create worker-run platforms, either along the lines of a cooperative or some other format where workers have a say in the structure of the labor market. (pp. 263-264)

The Stanford Crowd Research Collective (2015) has explored a similar possibility, designing a "crowd-built, self-governed crowdsourcing marketplace" called Daemo (p. 101; see also Whiting et al., 2017). Daemo enables negotiation on HIT specifics by requiring that each posted task go through an initial period of "prototyping" during which time workers can give feedback to requesters, and requesters can approve workers who apply to participate. The platform is governed by a democratically elected leadership board consisting of three workers, three requesters, and one researcher. Any competitor to the larger, more established platforms, however, will face an uphill battle with respect to scale. If they cannot guarantee a critical mass of requesters posting

HITs, then workers will not flock to their service. And if they cannot attract a large, relatively stable population of workers, then they would have no clear value proposition for requesters, especially academics looking for samples large enough to generate statistical significance.

Conclusion

As the number of publications that include crowdsourced data rises year over year, it is clear that crowdsourcing has changed the shape of academic research, for better or worse. The same things that make crowdsourcing exciting for research—especially the speed and scale at which new findings can be disseminated—are also what make addressing its shortcomings all the more urgent. In this article, we have highlighted three dimensions of concern: methods, fairness, and politics. Depending on the type of research being performed (e.g., a behavioral experiment vs. a demographic survey) and an individual researcher's personal and professional goals, one dimension may be more pertinent than another. However, it is critical to understand how crowdsourcing's interlocking features and its various stakeholders belong to a connected ecosystem where interventions in one area affect all of the others.

We are writing this article in the context of the global COVID-19 pandemic, which has affected academic research and economic productivity more broadly. It is too early to tell whether or not more researchers turned to crowdsourcing during the pandemic as it became more difficult if not impossible to use face-to-face methods, or what long-term effects—if any—such a shift could have on academic research and publishing. Moreover, future research could assess whether there was an uptick in the creation of new crowd worker accounts during the pandemic as jobs shifted to "work-from-home" models and workers were laid off from their jobs temporarily and permanently.[11] Even before the pandemic, crowd work belonged to a broader pattern of transition into precarious, flexible work arrangements, that is, the "gig economy" (see De Stefano 2015; Heeks, 2017; Duggan et al., 2020). If crowd working communities continue to grow in size and scope, there could be a positive spillover effect for academic researchers regarding the diversity of subject populations and the availability of treatment-naïve participants. At the same time, if the regulatory environment remains unchanged, the same issues regarding fairness and power dynamics that we have identified here are unlikely to be solved and may, in fact, worsen as the labor supply grows more quickly than demand. Future studies could usefully attend to whether a growing market for crowdsourcing becomes a catalyst for better platform design and/or regulation that can adequately address problems such as verifiability of workers and requesters, HIT compensation rates, and unilateral rejections of work without explanation.

What is perhaps most important for academics to realize, whether they have been using crowdsourcing for years or are just beginning to consider the possibility, is that platforms like MTurk, Prolific, or Clickworker are labor markets first and research populations second. Accepting that premise helps make sense of how and why information asymmetries, power imbalances, and poorly designed tasks are the objects of so much frustration. Resolutions that have been proposed and those that have already been implemented address some of these concerns, but more will need to be done to bring about an environment that is both respectful of crowd workers and useful for researchers.

Acknowledgments

We wish to thank Weiping Pei for her assistance scheduling interviews with crowd workers, as well as all of the workers who agreed to participate. This research was supported by the National Science Foundation as part of the Convergence Accelerator (C-ACCEL) Phase 1 research track focused on the Future of Work at the Human-Technology Frontier.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: **Grant number:** NSF OIA-1936968

ORCID iD

Stephen C. Rea https://orcid.org/0000-0002-2732-6014

Notes

1. Source: https://requester.mturk.com/pricing (accessed September 18, 2020). MTurk also charges an additional 5% fee on rewards for HITs that require a "Masters" qualification, a platform-specific designation that applies to a small fraction of workers.
2. Unless otherwise noted, all dollar amounts refer to U.S. dollars.
3. The current federal minimum wage rate was instituted in July, 2009. All of the articles that we analyzed drew upon crowdsourced research that had been conducted since that date.
4. Only 1 of 38 articles did not include at least one U.S.-based researcher as an author. This single-author article was written by a researcher working in the Netherlands, where human subjects research is vetted by institutional research ethics committees similar to a U.S. IRB.
5. Complementing super-Turkers are so-called "super-requesters." In an analysis of MTurk data from 2009 to 2010, Ipeirotis (2010a) found that over 30% of all HIT postings came from just .1% of requesters on the platform (p. 17).
6. All crowd worker names are pseudonyms.
7. CloudResearch, a liaison between academic researchers and crowdsourcing platforms, found that server farms (i.e., data centers that route internet from multiple actual-world locations through one site) were responsible for many of the

low-quality data that were collected on MTurk in 2018, and hypothesized that farms were a bigger threat to quality than bots (Moss, 2018).
8. IRBs may also place restrictions on what academics can pay because of worries about potential coercion (see Sheehan & Pittman, 2016).
9. As of July 2020, Turkopticon is transitioning to a fully "non-profit, worker run review site." Source: https://twitter.com/turkopticon/status/1288121585543475201?s=20.
10. These issues are part of a much larger discussion of social science research ethics that are beyond the scope of this article. For a comprehensive overview of deception in online research, see Boellstorff et al. (2012).
11. Spurk and Straub (2020) hypothesize that crowd workers "are less affected by political measures like lockdowns, because they already are used to work under high levels of flexibility regarding location and work schedule. . . . Due to this, we expect that even if the world of work will be different after the COVID-19 pandemic, it might be that gig work and flexible employment relationships in general will take another, maybe even more recognized position in the labor market."

References

Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H. R., Bertino, E., & Dustdar, S. (2013). Quality control in crowdsourcing systems. *IEEE Internet Computing*, *17*(2), 76-81.

Bai, H. (2018, August 8). Evidence that a large amount of low quality responses on MTurk can be detected with repeated GPS coordinates. *Maxhuibai.com*. https://www.maxhuibai.com/blog/evidence-that-responses-from-repeating-gps-are-random.

Bates, J. A., & Lanza, B. A. (2013). Conducting psychology student research via the Mechanical Turk crowdsourcing service. *North American Journal of Psychology*, *15*(2), 385-394.

Berg, J., Furrer, M., Harmon, E., Rani, U., & Silberman, M. S. (2018). *Digital labour platforms and the future of work: Towards decent work in the online world*. International Labour Office.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351-368. https://doi.org/10.1093/pan/mpr057

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). *Ethnography and virtual worlds: A handbook of method*. Princeton University Press.

Bohannon, J. (2011). Social science for pennies. *Science*, *334*(6054), 307. https://doi.org/10.1126/science.334.6054.307

Chee, F., Taylor, N., & de Castell, S. (2012). Re-mediating research ethics: End-user license agreements in online games. *Bulletin of Science, Technology & Society*, *32*(6), 497-506. https://doi.org/10.1177/0270467612469074

Cook, K. S., & Yamagishi, T. (2008). A defense of deception on scientific grounds. *Social Psychology Quarterly*, *71*(3), 215-221. https://doi.org/10.1177/019027250807100303

De Stefano, V. (2015). The rise of the "just-in-time workforce": On-demand work, crowdwork, and labor protection in the gig-economy. *Comparative Labor Law & Policy Journal*, *37*(3), 471-504. https://doi.org/10.2139/ssrn.2682602

Dietz, C. (2016). Pragmatism and Mechanical Turk: Citizenship and labor rights in digital communities of knowledge. *Journal of Media Ethics*, *31*(4), 264-266. https://doi.org/10.1080/23736992.2016.1228816

Dreyfuss, E. (2018, August 17). A bot panic hits Amazon's Mechanical Turk. *WIRED*. https://www.wired.com/story/amazon-mechanical-turk-bot-panic/.

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 (WSDM '18)* (pp. 135-143). Association for Computing Machinery. https://doi.org/10.1145/3159652.3159661

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resources Management Journal*, *30*(1), 114-132. https://doi.org/10.1111/1748-8583.12258

Felstiner, A. (2011). Working the crowd: Employment and labor law in the crowdsourcing industry. *Berkeley Journal of Employment & Labor Law*, *32*(1), 143-204. https://ssrn.com/abstract=1593853

Goodman, J. K., & Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research*, *44*(1), 196-210. https://doi.org/10.1093/jcr/ucx047

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

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 (CHI '18)* (pp. 1-14). Association for Computing Machinery.

Hassell, H. J. G., & Visalvanich, N. (2015). Call to (in)action: The effects of racial priming on grassroots mobilization. *Political Behavior*, *37*(4), 911-932. https://doi.org/10.1007/s11109-014-9297-x

Heeks, R. (2017). Digital economy and digital labour terminology: Making sense of the "gig economy," "online labour," "crowd work," "microwork," "platform labour," etc. (Development Informatics Working Paper No. 70). http://dx.doi.org/10.2139/ssrn.3431728

Hitlin, P. (2016, July 11). Research in the crowdsourcing age, a case study. *Pew Research Center*.

Howe, J. (2006, June 2). Crowdsourcing: A definition. *Crowdsourcing*. https://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html/

Ipeirotis, P. G. (2010a). Analyzing the Amazon Mechanical Turk marketplace. *XRDS*, *17*(2), 16-21. https://doi.org/10.1145/1869086.1869094

Ipeirotis, P. G. (2010b). Demographics of Mechanical Turk (NYU Working Paper No. CEDER-10-01). https://ssrn.com/abstract=1585030

Irani, L. C., & Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '13)* (pp. 611-620). Association for Computing Machinery.

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*(2), 519-528. https://doi.org/10.3758/s13428-014-0483-x

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1-23. https://doi.org/10.3758/s13428-011-0124-6

Mason, W., & Watts, D. J. (2009). Financial incentives and the "performance of crowds." In P. Bennett, R. Chandrasekar, M. Chickering, P. Ipeirotis, E. Law, A. Mityagin, F. Provost, & L. von Ahn (Eds.), In *Proceedings of the ACM SIGKDD workshop on human computation* (pp. 77-85). Association for Computing Machinery.

Miele, J. (2018, August 11). The BOT problem on Mturk. *Tips For Requesters On Mechanical Turk*. http://turkrequesters.blogspot.com/2018/08/the-bot-problem-on-mturk.html.

Milland, K. (2016). A Mechanical Turk worker's perspective. *Journal of Media Ethics*, *31*(4), 263-264. https://doi.org/10.1080/23736992.2016.1228813

Morrissey, C., Yamasaki, M., & Levitan, C. (2019). *Invisible participants: Recommendations for researchers using crowdsourcing platforms for human subjects research*. PsyArXiv Preprints. https://doi.org/10.31234/osf.io/85bfe

Moss, A. (2018, September 18). After the bot scare: Understanding what's been happening with data collection on MTurk and how to stop it. *CloudResearch*. https://www.cloudresearch.com/resources/blog/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it/

Palmer, J. C., & Strickland, J. (2016, June). A beginner's guide to crowdsourcing: Strengths, limitations and best practices for psychological research. *Psychological Science Agenda*. https://www.apa.org/science/about/psa/2016/06/changing-minds/

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184-188. https://doi.org/10.1177/0963721414531598

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

Pittman, M., & Sheehan, K. (2016). Amazon's Mechanical Turk a digital sweatshop? Transparency and accountability in crowdsourced online research. *Journal of Media Ethics*, *31*(4), 260-262. https://doi.org/10.1080/23736992.2016.1228811

Salehi, N., Irani, L. C., Bernstein, M. S., Alkhatib, A., Ogbe, E., & Milland, K., & Clickhappier. (2015). We are Dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems (CHI '15)* (pp. 1621-1630). Association for Computing Machinery.

Samuel, A. (2018, May 15). Amazon's Mechanical Turk has reinvented research. *JSTOR Daily*. https://daily.jstor.org/amazons-mechanical-turk-has-reinvented-research/

Schlagwein, D., Cecez-Kecmanovic, D., & Hanckel, B. (2019). Ethical norms and issues in crowdsourcing practices: A Habermasian analysis. *Information Systems Journal*, *29*(4), 811-837. https://doi.org/10.1111/isj.12227

Sheehan, K. (2016, August 24). MTurk and market economics. *MTurk for Academics*. https://mturk4academics.wordpress.com/2016/08/24/mturk-and-market-economics/

Sheehan, K. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs*, *85*(1), 140-156. https://doi.org/10.1080/03637751.2017.1342043

Sheehan, K., & Pittman, M. (2016). *Amazon's Mechanical Turk for academics: The HIT handbook for science research*. Melvin and Leigh.

Silberman, M. S., Irani, L., & Ross, J. (2010). Ethics and tactics of professional crowdwork. *XRDS 17*(2), 39-43. http://doi.acm.org/10.1145/1869086.1869100

Silberman, M. S., Tomlinson, B., LaPlante, R., Ross, J., Irani, L., & Zaldivar, A. (2018). Responsible research with crowds: Pay crowdworkers at least minimum wage. *Communications of the ACM*, *61*(3), 39-41. https://doi.org/10.1145/3180492

Spurk, D., & Straub, C. (2020). Flexible employment relationships and careers in times of the COVID-19 pandemic. *Journal of Vocational Behavior, 119*(June), 103435. https://doi.org/10.1016/j.jvb.2020.103435

Stamm, I., & Eklund, L. (2017, April 5). With great power comes great responsibility: Crowdsourcing raises methodological and ethical questions for academia. *LSE Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2017/04/05/crowdsourcing-raises-methodological-and-ethical-questions-for-academia/

Stanford Crowd Research Collective. (2015). Daemo: A self-governed crowdsourcing marketplace." In *Adjunct Proceedings of the 28th annual ACM symposium on user interface software & technology (UIST '15 Adjunct)* (pp. 101-102). Association for Computing Machinery.

Stokel-Walker, C. (2018, August 10). Bots on Amazon's Mechanical Turk are ruining psychology studies. *New Scientist*. https://www.newscientist.com/article/2176436-bots-on-amazons-mechanical-turk-are-ruining-psychology-studies/.

Wazny, K. (2017). "Crowdsourcing" ten years in: A review. *Journal of Global Health*, *7*(2), 1-13. https://doi.org/10.7189/jogh.07.020601

Whiting, M. E., Gamage, D., Gaikwad, S. S., Gilbee, A., Goyal, S., Ballav, A., Majeti, D., Chhibber, N., Richmond,-Fuller, A., Vargus, F., Sarma, T. S., Chandrakanthan, V., Moura, T., Salih, M. H., Kalejaiye, G. B. T., Ginzberg, A., Mullings, C. A., Dayan, Y., Milland, K., . . . Bernstein, M. S. (2017). Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing (CSCW '17)* (pp. 1902-1913). Association for Computing Machinery.

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

Williamson, V. (2016b, February 3). Can crowdsourcing be ethical? *Brookings Institution TechTank*. https://www.brookings.edu/blog/techtank/2016/02/03/can-crowdsourcing-be-ethical-2/.

Zawacki, K. (2017, April 14). Amazon's Turkers kick off the first crowdsourced labor guild. *The Daily Beast*. https://www.thedailybeast.com/amazons-turkers-kick-off-the-first-crowdsourced-labor-guild.

Author Biographies

Stephen C. Rea is an Adjunct Instructor and Research Assistant Professor in the Department of Humanities, Arts, and Social Sciences at the Colorado School of Mines. He is a cultural anthropologist whose research focuses on digital culture.

Hanzelle Kleeman is an undergraduate student in the Mechanical Engineering, Public Affairs, and Engineering Grand Challenges programs at the Colorado School of Mines. Her work deals with the ethics of emerging technologies and strives to improve the impact of technology by encouraging human-centered design.

Qin Zhu is an Assistant Professor in the Department of Humanities, Arts & Social Sciences with an affiliate faculty appointment in the Department of Engineering, Design & Society at the Colorado School of Mines. His research interests include engineering (ethics) education, ethical issues in human-robot interaction, and global engineering.

Benjamin Gilbert is an Assistant Professor in the Economics and Business Division at the Colorado School of Mines. He is an applied microeconomist who actively conducts survey-based and experimental research in environmental and natural resource economics, behavioral economics, and applied microeconomics.

Chuan Yue is an Associate Professor of Computer Science at the Colorado School of Mines. His current research includes usable security and privacy, trustworthy crowdsourcing, and cybersecurity education.